SHARP

UPDATED

Application Notes

Integrated Circuits

RF Components

Liquid Crystal Displays

Optoelectronics

NOTICE

The specifications contained within these Sharp Application Notes are current as of the publication date.

The Sharp Application Notes are the proprietary product of Sharp and are copyrighted, with all rights reserved. Under the copyright laws, no part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, in whole or in part, without the express written permission of Sharp. Express written permission is also required before any use of this publication may be made by a third party.

Sharp reserves the right to make changes in the circuitry or specifications described herein at any time without notice in order to improve design or reliability. The information in these Application Notes have been carefully checked to be accurate, however, Sharp makes no warranty for any errors which may appear in this document. Contact Sharp to obtain the latest version of device specifications before placing your order.

Sharp makes no representations that circuitry described herein is free from infringement of patent or ther rights of third parties which may result from its use. No license is granted by implication under any patent rights or other rights of Sharp Corporation.

© SHARP Corporation, 1994. Printed and bound in the U.S.A.

Trademark Acknowledgments:

Intel is a trademark of Intel Corporation.

AMD is a trademark of Advanced Micro Devices.

Chips & Technology is a trademark of Chips and Technology, Inc.

TABLE OF CONTENTS

INTEGRATED CIRCUITS

DSP

FAST FOURIER TRANSFORM

1. INTRODUCTION

Fast Fourier transform (FFT) is a fundamental driving force for high-speed real-time digital signal processing. It consists of a wide variety of applications including radar, sonar, seismic, ultrasonic, medical, image, etc. Sharp Microelectronics Technology (SMT) provides two chip set solution with LH9124 and LH9320 to efficiently and powerfully implement these applications in real-time. 4K-point complex FFT can be finished within half of a millisecond by one LH9124. The LH9124 is a fixed-point high performance digital signal processor and its data path is optimized for the FFT radix-4 butterfly structure with peak performance of 6 multiplications and 11 additions per cycle. The LH9320 is a programmable address generator to provide the address patterns required by the LH9124. Both chips are operated at a 25 ns cycle time. The two chip set solution has the following advantages:

[a] Parallel or pipelined architecture can be easily configured to match the speed requirement.

[b] No constraint on the length of the FFT executed by the LH9124 is applied and the constraint on the length of the address sequence generated by one LH9320 is no more than 2^{20} points.

[c] The glue logic required by an application system is minimized due to flexible four I/O port configuration of the LH9124.

[d] The system design can be simplified because the user's system function block diagram can be easily mapped to the function instructions provided by the chips.

Both fast Fourier transform and discrete cosine transform (DCT) from one-dimension (1-D) to multi-dimension (M-D) can be efficiently implemented by the LH9124 and LH9320. The two-dimensional (2-D) FFT is suggested to be employed by the 1-D column-row or row-column approach because the LH9320 can provide the required address patterns for both data and twiddle factors. The implementation of the M-D FFT algorithm can be extended from the 2-D approach.

The benchmark of the FFT and DCT is shown in Table 1-1. The performance can be further improved by the paralleled or cascaded structures as shown in Table 1-2. The throughput may be improved by a factor of M when the chip set is configured into M cascaded stages. The latency due to computations can be reduced almost by a factor of M with M LH9124 chips executed in parallel for the same stage. Thus, the chip set can be easily configured to satisfy the user's application requirements.

TABLE OF CONTENTS

Table 1-1. Benchmark of the FFT and DCT

POINTS	COMPLEX DATA FFT		REAL DATA FFT		REAL DATA DCT	
	CYCLES	µSECS	CYCLES	µSECS	CYCLES	µSECS
8	52	1.30	40	1.00	54	1.35
16	68	1.70	52	1.30	70	1.75
32	150	3.75	101	2.53	127	3.18
64	214	5.35	149	3.73	191	4.78
256	648	16.20	462	11.55	600	15.00
512	1690	42.25	1111	27.78	1377	34.43
1K	3226	80.65	2135	53.38	2657	66.43
4K	12492	312.30	8304	207.60	10362	259.05
8 × 8 2-D	328	8.20	274	6.85	432	10.80
16 × 16 2-D	648	16.20	536	13.40	972	24.30
32 × 32 2-D	4268	106.70	2846	71.15	4468	111.70
64 × 64 2-D	16556	413.90	10622	265.55	16884	422.10
256 × 256 2-D	262416	6560.00	164898	4122.45	263512	6587.80
512 × 512 2-D	1573172	39329.30	919878	22996.95	1313148	32828.70
1K × 1K 2-D	6291764	157294.10	3674438	91860.95	5247356	131183.9 0
16 × 16 × 16 3-D	12492	312.30				

Table 1-2. Benchmark of the Parallel Complex FFT

FUNCTION	CYCLES	µSECS
1K (3 Stage Cascaded)	1024	25.6
4K (3 Stage Cascaded)	4096	102.4
64 × 64 (4 Stage Cascaded)	4096	102.4
256 × 256 (4 Stage Cascaded)	65536	1638.4
4K (2 Parallel)	6348	158.7
4K (2 Parallel with 3 Stage Cascaded)	2048	51.2
16K (4 Parallel)	16606	415.15
16K (8 Parallel)	8414	210.35
16K (16 Parallel)	4320	108.0

2. OVERVIEW OF FAST FOURIER TRANSFORM

Orthogonal transforms and transform properties play an important role for engineers to solve new problems. The great advantage of such a transform is that we are in a position of being able to analyze a signal with some knowledge of its constituent parts. In this note, the implementation of three fundamental orthogonal transforms – Fourier transform, cosine transform and Hartley transform – by the LH9124 and LH9320 is discussed.

2.1. Discrete Fourier Transform

For a long time, the Fourier transform has been a powerful and principle analysis tool in diverse fields such as linear systems, probability theory, boundary-valued problems, communications theory, signal processing, etc. The discrete Fourier transform (DFT) is the counterpart of the Fourier transform in the discrete time domain. The definition of the DFT is given by the expression:

$$X(k) = \sum_{n=0}^{N-1} x(n)\, W_N^{kn} \quad \text{for } k = 0, 1, ..., N-1 \qquad [2A]$$

and the inverse DFT, IDFT, is expressed as:

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k)\, W_N^{-kn} \quad \text{for } n = 0, 1, ..., N-1 \qquad [2B]$$

where $W_N^k = e^{-j2\pi k/N}$ is a sequence of twiddle factors of the DFT and is equally spaced around the unit cycle. In these equations x(n) is the sample value in the time domain and X(k) is the sample value in the frequency domain. If the sampling rate of a signal is F, the sequence of time-domain sampling locations becomes

$$0, 1/F, 2/F, 3/F, \ldots, (N-1)/F \qquad [2C]$$

Thus, corresponding to these time-domain samples, the sequence of frequency-domain sampling locations will be

$$0, F/N, 2F/N, 3F/N, \ldots, (N-1)F/N \qquad [2D]$$

The computation of the DFT and IDFT is basically a vector dot operation. The instructions BCFIR used for the FIR operation can be directly employed for the DFT or IDFT computation. The LH9320 provides an instruction ADECIM to fetch the required address patterns from the twiddle factor table. If the length of the sequence is N and the samples to be computed is M, the cycles required for the computation will be M * N + 18. Comparing the cycles required by the DFT with Table 1-1, it can be seen that the manipulation based on the FFT is generally more efficient if the number of frequency samples to be obtained is more than 3.

The DFT and IDFT both produce periodic results with period N. It can be seen from Equations [2A] and [2B] that some symmetric properties also exist. These properties are shown in Table 2-1. the computation of the real DFT will be more efficient by employing these symmetric properties.

Table 2-1. Symmetric Properties of the DFT

TIME-DOMAIN	FREQUENCY-DOMAIN
x(n) real	$X(k) = X^*(-k) = X^*(N-k)$
$x(n) = x^*(-n) = x^*(N-n)$	X(k) real
x(n) real and even	X(k) real and even
x(n) real and odd	X(k) imaginary and odd

2.2. Fast Fourier Transform

A direct computation of the DFT or IDFT requires N^2 complex multiplication is and N(N – 1) complex additions. The FFT is an efficient algorithm for computing the DFT and can be derived from the DFT. The discrete Fourier transform can be computed in $O(N \log_2 N)$ multiplications by the FFT. The data path of LH9124 is optimized for this algorithm. It can compute a decimation-in-time radix-2 butterfly in two cycles, radix-4 butterfly in four cycles, or radix-16 in 16 cycles. Therefore, a very large of important computational problems under the general rubric of Fourier transform methods or spectral methods can be efficiently solved.

2.2.1. Formation of the FFT

The fast Fourier transform algorithm achieves its computational efficiency through a divide and conquer strategy. The essential idea is a grouping of the time and frequency samples such that the DFT summation over N values can be expressed as a combination of DFT summations over N/2 samples. When N is a power of two, this process of grouping can be repeatedly applied until the DFT summation has been reduced to a combination of DFT summation over only two samples. For example, when N is a power of two, Equation [2A] can be decomposed as follows:

$$X(k) = \sum_{n=0}^{N/2-1} x(2n)\, W_N^{2nk} + \sum_{n=0}^{N/2-1} x(2n+1)\, W_N^{(2n+1)k} \qquad [2E]$$

$$= \sum_{n=0}^{N/2-1} x(2n)\, W_N^{2nk} + W_N^k \sum_{n=0}^{N/2-1} x(2n+1)\, W_N^{2nk}$$

Define two (N/2)-point sequences {h(n)} and {g(n)} as the even and odd elements of {x(n)}, respectively. Then,

$$h(n) = x(2n) \quad [2F]$$

$$g(n) = x(2n + 1) \quad [2G]$$

The discrete Fourier transform of the two (N/2)-point sequences can be defined as follows:

$$H(k) = \sum_{n=0}^{N/2-1} h(n)\, W_{N/2}^{kn} \quad [2H]$$

$$G(k) = \sum_{n=0}^{N/2-1} g(n)\, W_{N/2}^{kn} \quad [2I]$$

Thus, the discrete Fourier transform of x(n) can be expressed in terms of even and odd elements as:

$$X(k) = \sum_{n=0}^{N/2-1} h(n)\, W_{N/2}^{2nk} + W_N^k \sum_{n=0}^{N/2-1} g(n)\, W_{N/2}^{2nk} \quad [2J]$$

In terms of H(k) and G(k), we have:

$$X(k) = H(k) + W_N^k\, G(k) \quad [2K]$$

For the coefficient at (K + N/2), we obtain:

$$X(k + N/2) = H(k) - W_N^k\, G(k) \quad [2L]$$

Therefore, N-Point discrete Fourier transform can be represented by two (N/2)-point discrete Fourier transform. Since N is a power of two, the above partitioning scheme can be iteratively applied to the sequences {h(n)} and {g(n)} by N/4 elements. These partitions can be carried out until the two-point DFT is reached. The process is depicted in Figure 2-1 for N = 8. Figure 2-2 shows the flow graph of an 8-point FFT that denotes the results of the 8-point decomposed DFT.

2.2.2. Fundamental Computing Structures: Butterflies

The structure shown in Figure 2-2 is called decimation-in-time (DIT). Its basic module is a radix-2 butterfly shown in Figure 2-3 in which two points $x_k(a)$ and $x_k(b)$ are computed to give two output points $x_{k+1}(a)$ and $x_{k+1}(b)$ via the operations represented by Equations [2K] and [2L]. Each radix-2 butterfly requires one complex multiplication and two complex additions. Observing Figure 2-2, it can be seen that there are Log_2 N radix-2 butterfly stages for N-point FFT and each stage has N/2 radix-2 butterflies. Therefore, there are totally $(N/2)\log_2$ N radix-2 butterflies in an N-point FFT. In addition, the input is in bit-reverse order and output is in linear order.

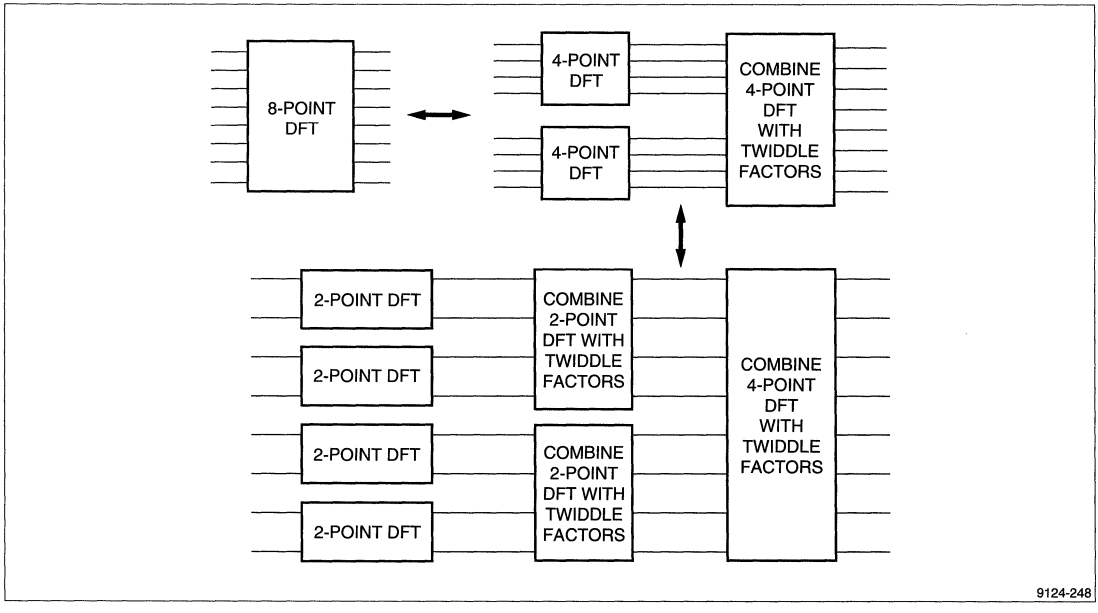


Figure 2-1. 8-Point DFT Decomposed into 8-Point 3-Stage Radix-2 FFT

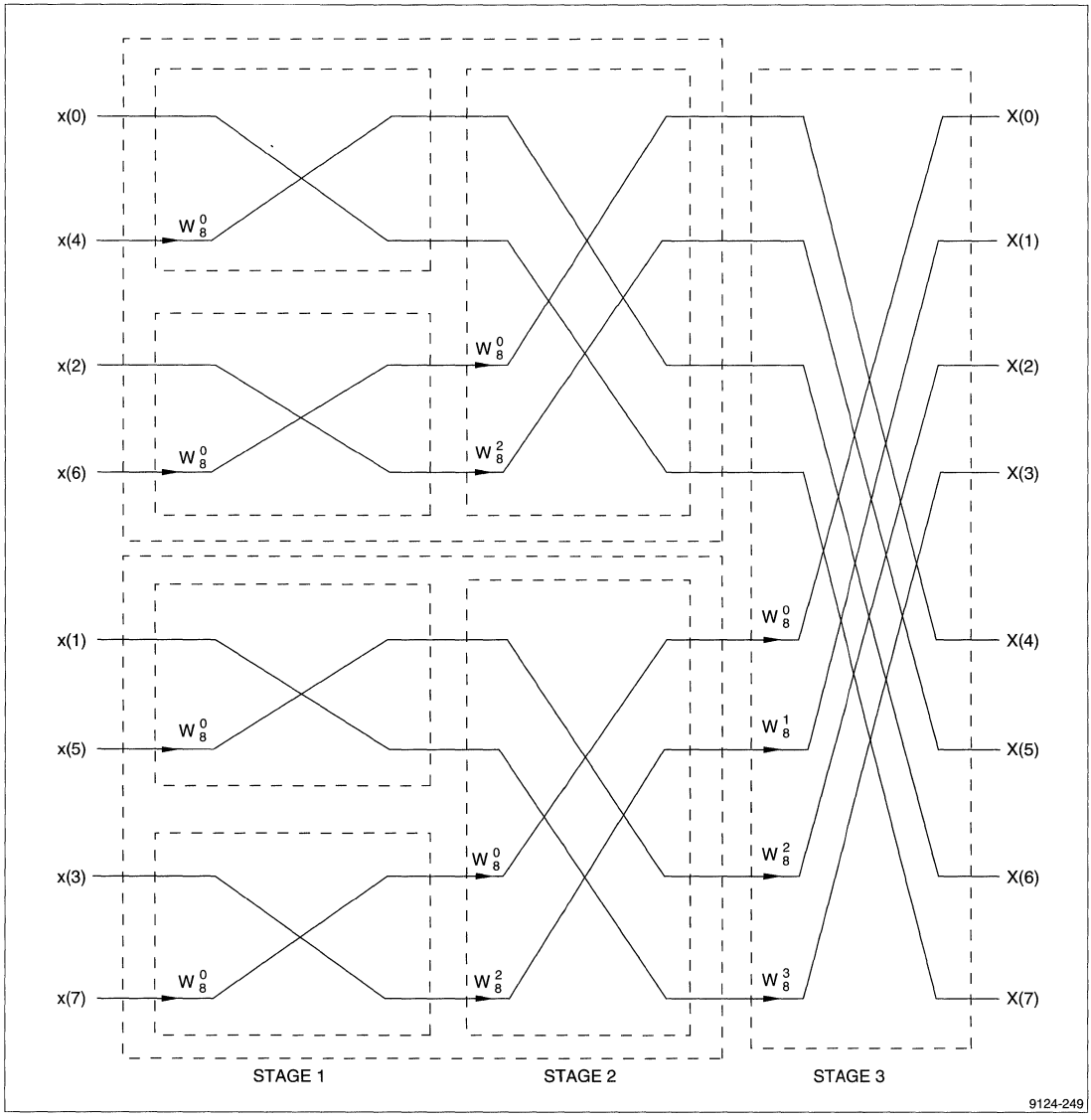


Figure 2-2. 8-Point FFT With 3-Stage Decimation-in-Time Radix-2 Structure

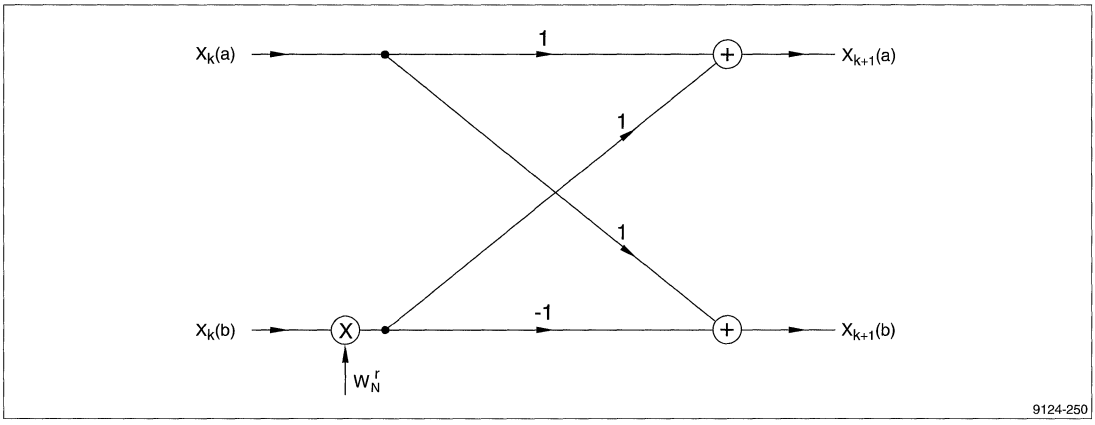


Figure 2-3. Radix-2 Butterfly Structure

From the flow graph shown in Figure 2-2, a lot of FFT structures can be derived by exchanging the position of the flow graph. One of these structures is shown in Figure 2-4 of which the input is linear order and output is in bit-reverse order. Both structures employ the in-place algorithm. In-place means that the computed outputs can be placed on the same storage as the inputs. Moreover, the addressing for the input and output data can be shared. One more equivalent structure is depicted in Figure 2-5. The four radix-2 butterflies enclosed in the dashed box can be implemented by a radix-4 butterfly. The structure has inputs in digit-reverse order and outputs in linear order. It is not an in-place algorithm for radix-2 but it is an in-place algorithm for radix-4. The notation of the DIT radix-4 butterfly is shown in Figure 2-6.

The LH9124 and LH9320 are designed to implement decimation-in-time FFT structures with digit-reverse inputs and linear outputs. The LH9124 provides three instructions BFLY2, BFLY4, and BFLY16 for computing radix-2, radix-4, and radix-16 butterflies. The radix-2 butterfly is a fundamental module for the FFT. The radix-4 butterfly implemented by BFLY4 is equivalent to the four radix-2 butterfly structure shown in Figure 2-7. The radix-16 butterfly implemented by BFLY16 is equivalent to the eight radix-4 butterfly structure shown in Figure 2-8 or to the 32 radix-2 butterfly structure shown in Figure 2-9. Thus, a radix-16 butterfly stage is equivalent to two radix-4 butterfly stages or four radix-2 butterfly stages with the same input-output map. If the instruction overhead is not counted, the computation time for one radix-16 stage is the same as that of one radix-4 stage or one radix-2 stage. Therefore, the rate of computing efficiency for radix-16, radix-4, and radix-2 is almost 4:2:1.

The N-point FFT can be implemented by the mixed radix operation as long as the following relation exists:

$$N = r_0 * r_1 * r_2 * \ldots * r_{R-1} \qquad [2M]$$

where r_i is 2 for a radix-2 stage, 4 for a radix-4 stage, and 16 for a radix-16 stage. The r_0 is the first stage and r_{R-1} is the last stage of the FFT. The order of the stages can be arbitrarily selected when implemented by the chip set. The structure of the N-point FFT with the order of r_i indicated by Equation [2M] is denoted by $r_0Xr_1X \ldots Xr_{R-1}$. Thus, the structures shown in Figure 2-2 and Figure 2-5 are denoted by $2 \times 2 \times 2$ and 4×2, respectively. If $N = 2^M$, there are M stages by the radix-2 structure. We may define that the FFT has M columns. For an arbitrary j-th column, we may have three ways to implement this column. First, it may be directly implemented as a radix-2 stage. Second, it and its next column may be implemented by a radix-4 stage. Third, it and its next three columns may be implemented by a radix-16 stage. The column defined here will be used to explain the FFT addressing instructions provided by the LH9320.

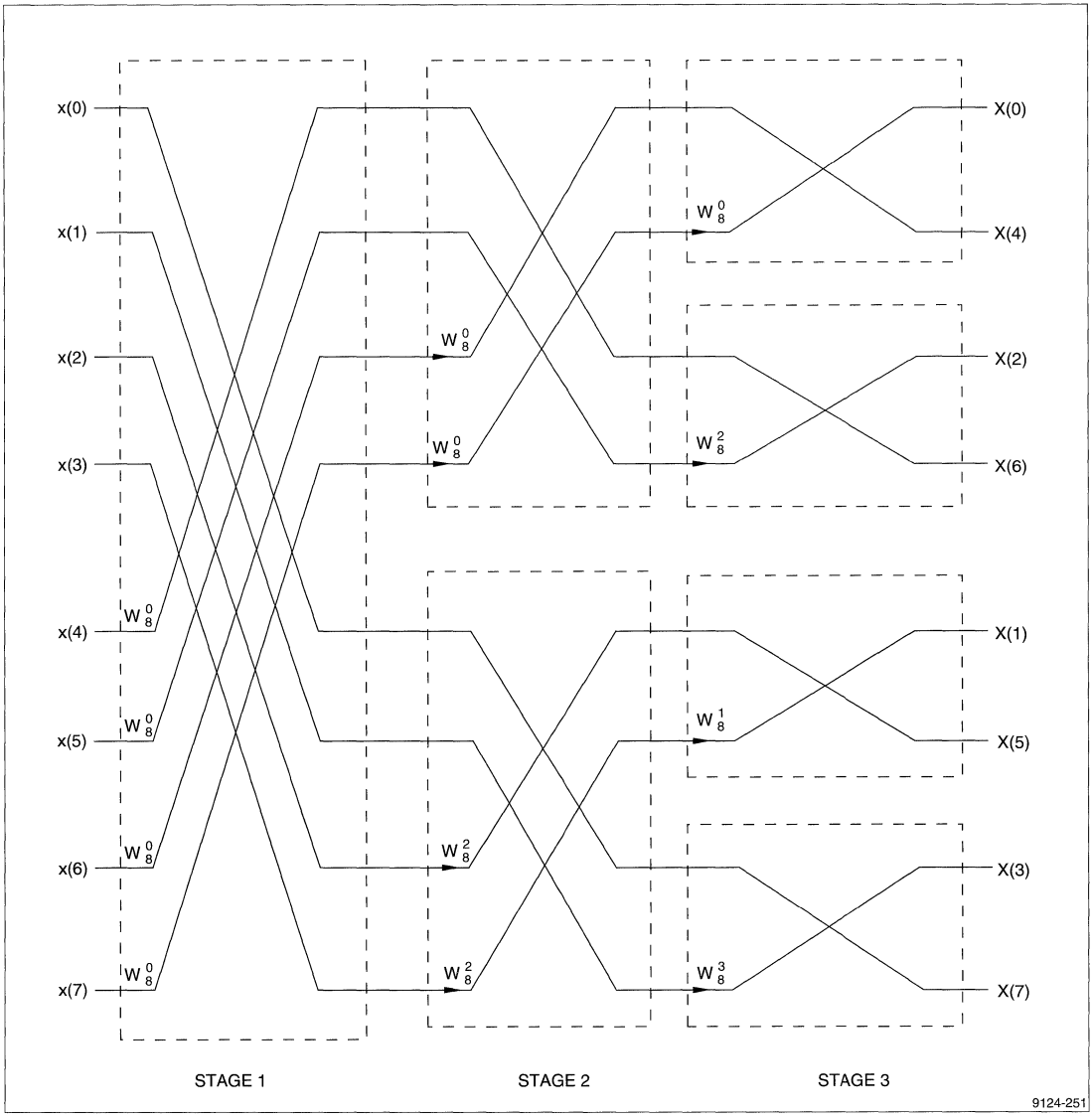


Figure 2-4. 8-Point FFT With Linear Input and Bit-Reverse Output

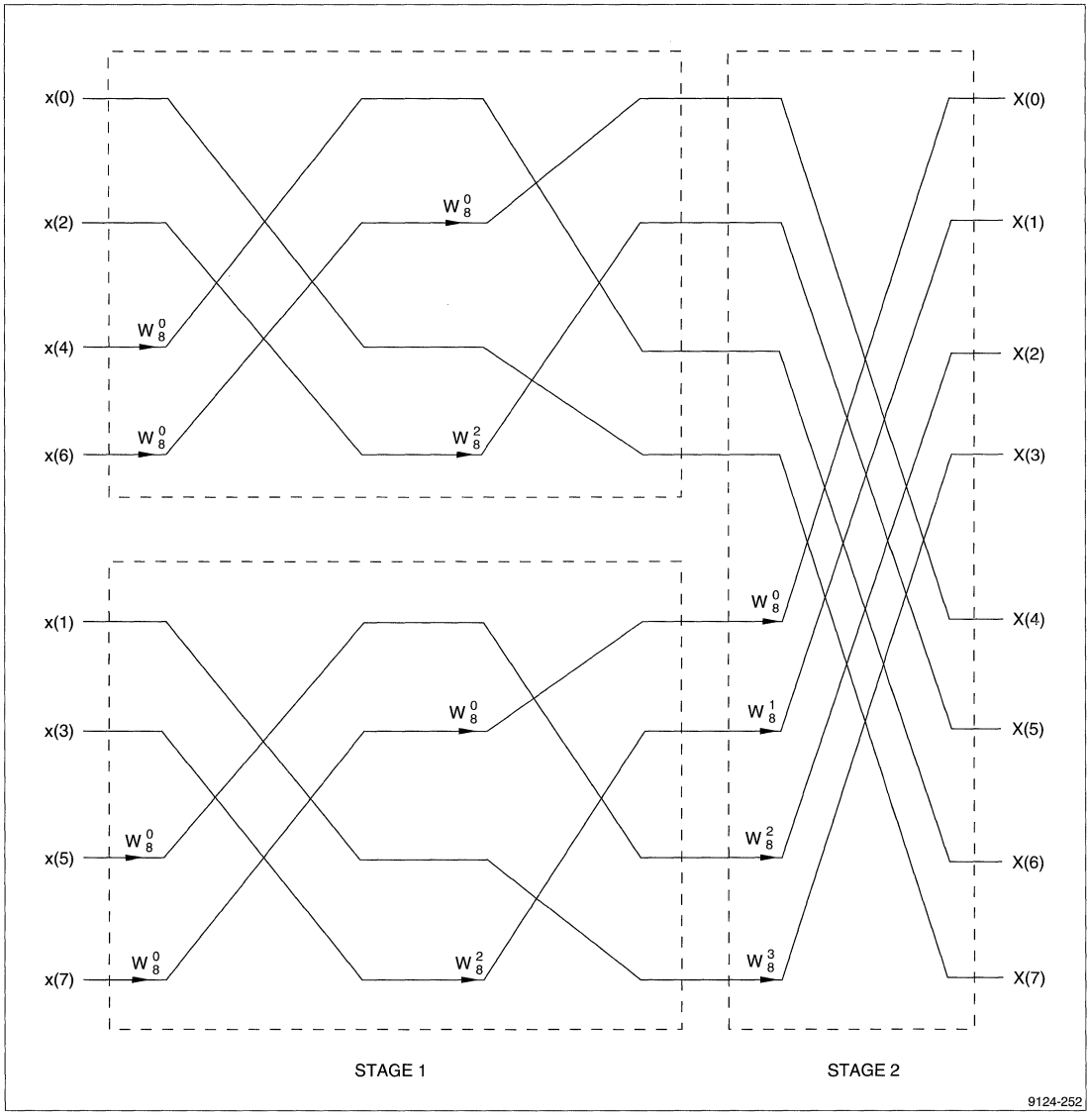


Figure 2-5. Digit-Reverse Input and Linear Output for 4 × 2 Stage FFT

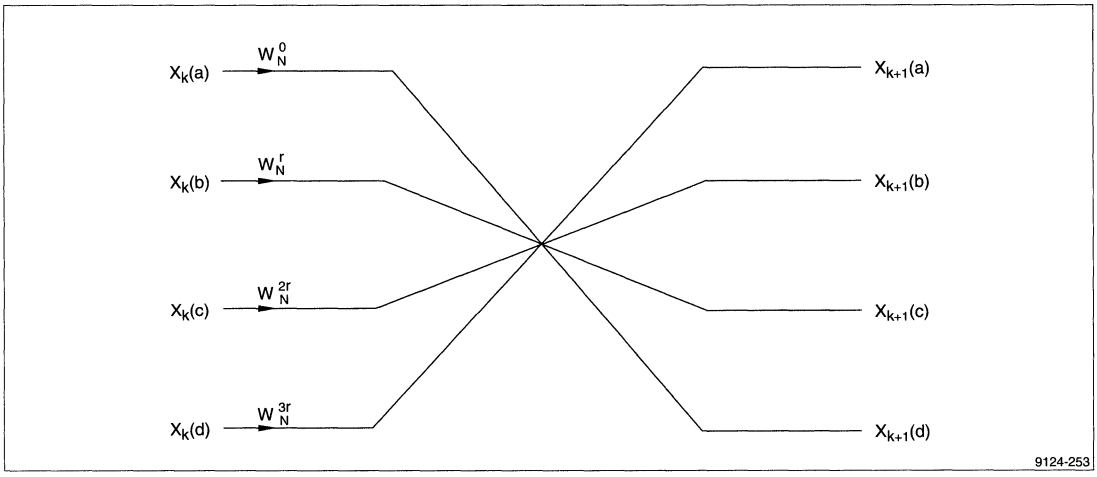


Figure 2-6. Notation of a Radix-4 Butterfly

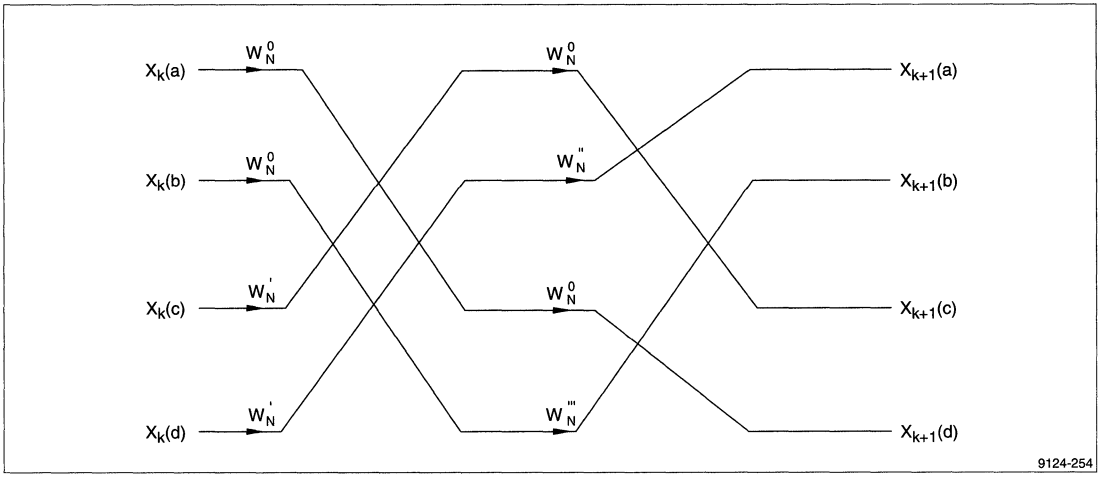


Figure 2-7. A Radix-4 Butterfly Depicted by its Equivalent Radix-2 Structure

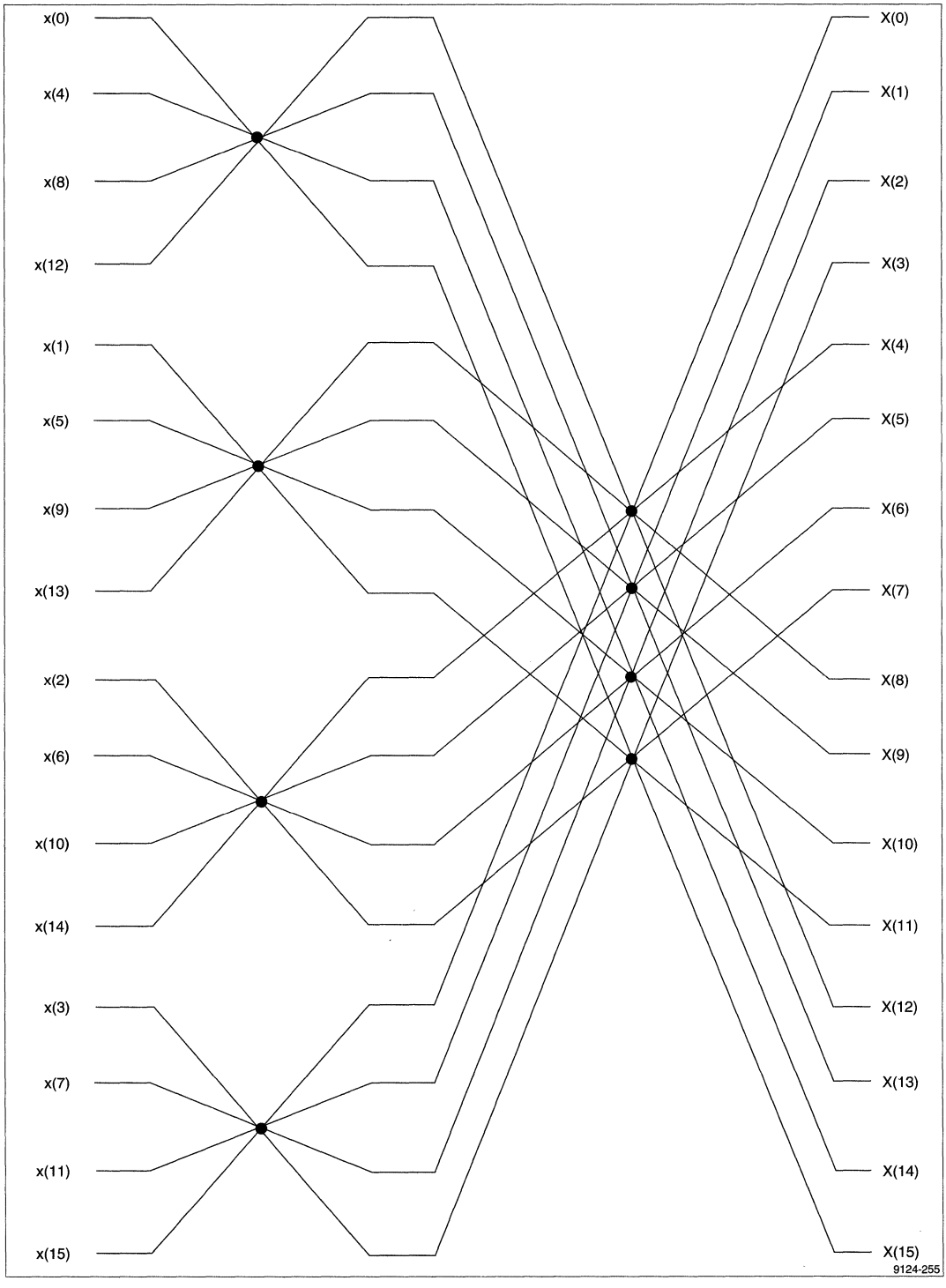


Figure 2-8. A Radix-16 Butterfly Depicted by its Equivalent Radix-4 Structure

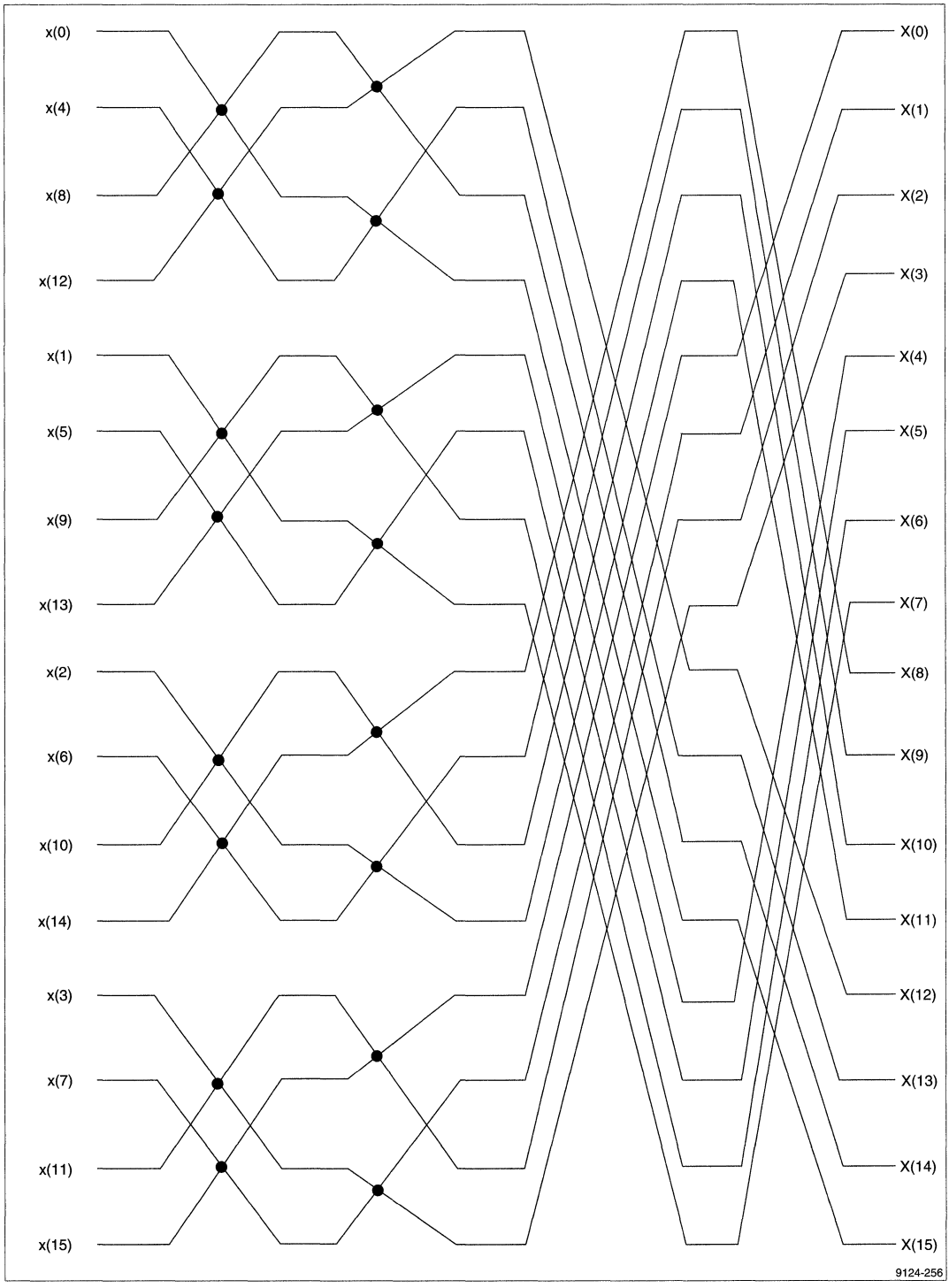


Figure 2-9. A Radix-16 Butterfly Depicted by its Equivalent Radix-2 Structure

The LH9320 will generate all the required data and twiddle factor address patterns according to the structure of the FFT. The FFT addressing instructions provided by the LH9320 are listed in Table 2-2. If the j-th column is implemented by a radix-2 stage, the instructions BF2j and TF2j are used to generate the required data and address patterns. If the j-th column is the beginning column of a radix-4 stage, the instructions BF4i and TF4i will be implemented when j = 2 * i or the instructions MXB24i and MXT24i will be implemented when j = 2 * i + 1. Similarly, we can select instructions for a radix-16 stage.

Table 2-2. The LH9320 Instructions for FFT Computations

INSTRUCTION	DESCRIPTION
BF2i	Data address sequence for i-th column radix-2
BF4i	Data address sequence for (2 * i)-th column radix-4
MXB24i	Data address sequence for (2 * i + 1)-th column radix-4
BF16i	Data address sequence for (4 * i)-th column radix-16
MXB216i	Data address sequence for (4 * i + 1)-th column radix-16
MXB416i	Data address sequence for (4 * i + 2)-th column radix-16
MXB2416i	Data address sequence for (4 * i + 3)-th column radix-16
TF2i	Twiddle factor address sequence for i-th column radix-2
TF4i	Twiddle factor address sequence for (2 * i)-th column radix-4
MXT24i	Twiddle factor address sequence for (2 * i + 1)-th column radix-4
TF16i	Twiddle factor address sequence for (4 * i)-th column radix-16
MXT216i	Twiddle factor address sequence for (4 * i + 1)-th column radix-16
MXT416i	Twiddle factor address sequence for (4 * i + 2)-th column radix-16
MXT2416i	Twiddle factor address sequence for (4 * i + 3)-th column radix-16
RBF0	Linear sequence to Digit-reverse sequence conversion

2.2.3. Data Sequence: Linear and Digit-Reverse

The decimation-in-time FFT algorithm with inputs in digit-reverse order and outputs in linear order is implemented by the chip set. Set a digit number $n = n_{N-1}n_{N-2} \ldots n_1n_0$ and assume the radius of n_i is r_i. A binary number is a special case of digit number with the radius of all the n_i being 2. The digit number implemented by the chip set for the FFT have the constraint that the radius of an arbitrary n_i is 2 or 4. The digit number used by the radix-16 is actually represented by two digits n_in_{i-1} with the radius of each digit being 4. The weighting factor of n_i is obtained by:

$$w_i = r_{i-1} * r_{i-2} * \ldots * r_1 * r_0 \qquad [2N]$$

The function of digit-reverse is then defined as follows:

$$dr(n) = dr(n_{N-1}n_{N-2} \ldots n_1n_0) = n_0n_1 \ldots n_{N-2}n_{N-1} \qquad [2O]$$

If the pattern of its associated radius $r_{N-1}r_{N-2} \ldots r_1r_0$ is symmetric, the function of digit-reverse is reversible. Thus,

$$dr(dr(n)) = n \qquad [2P]$$

The function of bit-reverse is always reversible because its radius pattern is symmetric.

The digit-reverse sequence can be generated in the same way of the linear sequence. Both can be recursive computed by adding a fixed addend as follows:

$$s_{i+1} = s_i + \text{addend} \qquad [2Q]$$

Two things are different in Equation [2Q] between the two sequence generators. First, the carry digit is propagated from right to left for the linear sequence and from left to right for the digit-reverse sequence. Second, the addends for the two are different. The addend of the linear sequence is 1 for the least significant digit and 0 for all the other digits. The addend of the digit-reverse sequence is 1 for the most significant digit and 0 for all the other digits.

Both the linear and digit-reverse 8-point sequences with the radius pattern $2 \times 2 \times 2$ and 4×2 are shown in Table 2-3. For the $2 \times 2 \times 2$ radius pattern, the addend is 001 for the linear sequence and 100 for the digit-reverse sequence. For the 4×2 radius pattern, the addend is 01 for the linear sequence and 10 for the digit-reverse sequence. It can be seen from the table that the digit-reverse sequence can also be obtained by reflecting the digit of the linear sequence according to the center digit. For example, the digit-reverse of 1234 and 12345 are 4321 and 54321, respectively.

The $2 \times 2 \times 2$ radius pattern is symmetric. Thus, the function of digit-reverse is reversible. The 4×2 radius pattern is not symmetric. Thus, Equation [2P] does not exist. The LH9320 provides one instruction RBF0 to generate a digit-reverse sequence based on the length of the sequence and the radius pattern defined by the user. The radius pattern actually shows the radix stage of the FFT operation. The radix-16 stage is represented by 4×4.

Table 2-3. Relation Between Linear and Digit-Reverse Order

2 × 2 × 2 PATTERN				4 × 2 PATTERN			
LINEAR		DIGIT-REVERSE		LINEAR		DIGIT-REVERSE	
DECIMAL	DIGIT	DIGIT	DECIMAL	DECIMAL	DIGIT	DIGIT	DECIMAL
0	000	000	0	0	00	00	0
1	001	100	4	1	01	10	2
2	010	010	2	2	02	20	4
3	011	110	6	3	03	30	6
4	100	001	1	4	10	01	1
5	101	101	5	5	11	11	3
6	110	011	3	6	12	21	5
7	111	111	7	7	13	31	7

2.2.4. Inverse Fast Fourier Transform

The inverse FFT (IFFT) defined by Equation [2B] can be changed to the following form:

$$x(n) = \frac{1}{N}\left[\sum_{k=0}^{N-1} X^{*}(k)\, W_N^{kn}\right]^{*} \quad \text{for } n = 0, 1, \ldots, N-1 \text{[2R]}$$

where the notation of the superscript * denotes the conjugated data.

If the input frequency samples are conjugated, the bracketed term in Equation [2R] is exactly an FFT operation. Thus, the computation of the IFFT can share all the instructions of the FFT. It can be seen from Equation [2R] that the final computed results have to be conjugated also. The LH9124 provides two pins DCI and DCO to control the input and output data conjugate. Thus, the computation of the IFFT is the same as the FFT except the input data sequence of the first stage and output data sequence of the last stage are conjugated.

The results computed by the FFT instructions are reference values and not exact values shown by Equation [2R]. If users like to get exact values, the results have to be multiplied by a factor. This factor can be derived from the length of data N and the total number of scaling for the block data. The total number of bits scaled for the input block data can be obtained from the block floating-point output pins BFPO[5:0] of the LH9124.

2.3. Two-Dimensional Fast Fourier Transform

Considering a complex function defined over a two-dimensional region $0 \le n_1 \le N_1 - 1$ and $0 \le n_2 \le N_2 - 1$, we can define two dimensional discrete Fourier transform over the same region as follows:

$$X(k_1,k_2) = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} x(n_1,n_2)\, W_{N_1}^{n_1k_1}\, W_{N_2}^{n_2k_2} \qquad \text{[2S]}$$

for $0 \le k_1 \le N_1 - 1$ and $0 \le k_2 \le N_2 - 1$

As in the 1-D DFT, the 2-D DFT also possesses some properties like linearity, symmetry, etc. The fast computation of the 2-D DFT is probably the most important means for realization of 2-D filters.

The direct computation of 2-D DFT is somewhat naive because there are $N_1 * N_2$ multiplications for each frequency sample. There are some different methods for fast computation of the 2-D DFT by the way of partitioning the Equation [2S]. The vector-radix method is not suggested because the address pattern is not suited to the LH9320, although the basic module of the vector-radix FFT is a radix-4 structure. The 1-D column-row or row-column approach is recommended because the LH9320 provides all the required address patterns for efficient computations.

2.3.1. Column-Row or Row-Column Decomposition Approach

Equation [2S] can be reordered into the column-row expression:

$$X(k_1,k_2) = \sum_{n_2=0}^{N_2-1} \left[\sum_{n_1=0}^{N_1-1} x(n_1, n_2)\, W_{N_1}^{n_1 k_1} \right] W_{N_2}^{n_2 k_2} \qquad [2T]$$

or into the row-column expression:

$$X(k_1, k_2) = \sum_{n_1=0}^{N_1-1} \left[\sum_{n_2=0}^{N_2-1} x(n_1, n_2)\, W_{N_2}^{n_2 k_2} \right] W_{N_1}^{n_1 k_1} \qquad [2U]$$

If the bracketed item in Equation [2T] is expressed by a 2-D function (k_1,n_2), then the 2-D FFT can be computed by two phases of 1-D FFT as follows:

$$F(k_1,n_2) = \sum_{n_1=0}^{N_1-1} x(n_1,n_2)\, W_{N_1}^{n_1 k_1} \quad \text{for } 0 \le k_1 \le N_1 - 1 \qquad [2V]$$

$$X(k_1,k_2) = \sum_{n_2=0}^{N_2-1} F(k_1,n_2)\, W_{N_2}^{n_2 k_2} \quad \text{for } 0 \le k_2 \le N_2 - 1 \qquad [2W]$$

Thus, the 2-D FFT can be obtained by computing 1-D FFT for each row of $x(n_1,n_2)$, putting the result into an intermediate array, and then computing 1-D FFT for each column of the intermediate array.

2.3.2. 2-D Data Sequence: Linear and Digit-Reverse

In the 1-D case, if the input sequence is in linear order, the output sequence after FFT operations will be in digit-reverse order and vice versa. The 2-D FFT can be obtained by doing two phases of 1-D FFT for each tuple. Therefore, the indices k_1 and k_2 of the output array after 2-D FFT are both in digit-reverse order with respect to the indices n_1 and n_2. For example, Table 2-4 shows the pattern of a (8,8) 2-D input array. Assume that each row is implemented by the 1-D FFT with a radix-4 stage followed by a radix-2 stage and each column is implemented by a radix-2 stage followed by a radix-4 stage. The function $F(k_1,n_2)$ of the intermediate stage 2-D array after row operations is shown in Table 2-5. It can be seen that the index k_1 is in digit-reverse order and n_2 is still in linear order. Table 2-6 shows the function $X(k_1,k_2)$ of the final 2-D array after two phases of 1-D FFT operations. Both indices k_1 and k_2 are in digit-reverse order. Conversely, If the indices of $x(n_1,n_2)$ are in digit-reverse order, the indices of the output array $X(k_1,k_2)$ will be in linear order.

Comparing Table 2-4 with Table 2-6, the function of the 2-D digit-reverse has the following relation:

$$dr(n_1, n_2) = (dr(n_1), (dr(n_2)) = (k_1, k_2) \qquad [2X]$$

The memory device is generally a 1-D array. In the 2-D FFT implementation, we will discuss how to use 1-D $N_1 * N_2$-point FFT to realize the 2-D FFT for the 2-D array, where N_1 and N_2 are the dimensions of the tuples.

2.4. Multi-Dimensional Fast Fourier Transform

The column-row or row-column approach of the 2-D FFT can be extended to the multi-dimensional (M-D) FFT case. The three-dimensional FFT is discussed here and the M-D FFT can be obtained by extending the tuples from 3 to M. A three-dimensional (3-D) discrete Fourier transform is represented by:

$$X(k_1,k_2,k_3) = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} \sum_{n_3=0}^{N_3-1} x(n_1,n_2,n_3)\, W_{N_1}^{n_1 k_1}\, W_{N_2}^{n_2 k_2}\, W_{N_3}^{n_3 k_3}$$

$$\text{for } 0 \le k_i \le N_i - 1 \qquad [2Y]$$

It can be reordered into the 1-D DFT expression as follows:

$$X(k_1,k_2,k_3) = \sum_{n_3=0}^{N_3-1} \left\{ \sum_{n_2=0}^{N_2-1} \left[\sum_{n_1=0}^{N_1-1} x(n_1,n_2,n_3)\, W_{N_1}^{n_1 k_1} \right] W_{N_2}^{n_2 k_2} \right\} W_{N_3}^{n_3 k_3} \qquad [2Z]$$

Thus, the 3-D FFT can be calculated by employing (1) 1-D N_1-point FFT $N_2 * N_3$ times first, (2) 1-D N_2-point FFT $N_1 * N_3$ times next, and (3) finally 1-D N_3-point FFT $N_1 * N_2$ times. After the 3-D FFT calculation, the indices n_i and k_i will be in digit-reverse order in each tuple as follows:

$$dr(n_1, n_2, n_3) = (dr(n_1), dr(n_2), dr(n_3)) = (k_1, k_2, k_3) \qquad [2AA]$$

The basic execution modules are unchanged for the 1-D to M-D FFT if the column-row or row-column approach is implemented. Thus, the radix execution instructions provided by the LH9124 can be applied to the FFT operation with an arbitrary dimension. The constraint for the LH9320 is that the number of points for a 1-D FFT operation cannot be larger than 2^{20}. If the addressing of the whole M-D array is seen as a 1-D array, the constraint will be $N_1 * N_2 * \ldots * N_M \le 2^{20}$, where N_i is the length of the i-th tuple.

Table 2-4. Pattern of an 8 × 8 2-D Linear Array

x(0,0)	x(1,0)	x(2,0)	x(3,0)	x(4,0)	x(5,0)	x(6,0)	x(7,0)
x(0,1)	x(1,1)	x(2,1)	x(3,1)	x(4,1)	x(5,1)	x(6,1)	x(7,1)
x(0,2)	x(1,2)	x(2,2)	x(3,2)	x(4,2)	x(5,2)	x(6,2)	x(7,2)
x(0,3)	x(1,3)	x(2,3)	x(3,3)	x(4,3)	x(5,3)	x(6,3)	x(7,3)
x(0,4)	x(1,4)	x(2,4)	x(3,4)	x(4,4)	x(5,4)	x(6,4)	x(7,4)
x(0,5)	x(1,5)	x(2,5)	x(3,5)	x(4,5)	x(5,5)	x(6,5)	x(7,5)
x(0,6)	x(1,6)	x(2,6)	x(3,6)	x(4,6)	x(5,6)	x(6,6)	x(7,6)
x(0,7)	x(1,7)	x(2,7)	x(3,7)	x(4,7)	x(5,7)	x(6,7)	x(7,7)

Table 2-5. Pattern of 2-D Array After 1-D FFT for Each Row

F(0,0)	F(2,0)	F(4,0)	F(6,0)	F(1,0)	F(3,0)	F(5,0)	F(7,0)
F(0,1)	F(2,1)	F(4,1)	F(6,1)	F(1,1)	F(3,1)	F(5,1)	F(7,1)
F(0,2)	F(2,2)	F(4,2)	F(6,2)	F(1,2)	F(3,2)	F(5,2)	F(7,2)
F(0,3)	F(2,3)	F(4,3)	F(6,3)	F(1,3)	F(3,3)	F(5,3)	F(7,3)
F(0,4)	F(2,4)	F(4,4)	F(6,4)	F(1,4)	F(3,4)	F(5,4)	F(7,4)
F(0,5)	F(2,5)	F(4,5)	F(6,5)	F(1,5)	F(3,5)	F(5,5)	F(7,5)
F(0,6)	F(2,6)	F(4,6)	F(6,6)	F(1,6)	F(3,6)	F(5,6)	F(7,6)
F(0,7)	F(2,7)	F(4,7)	F(6,7)	F(1,7)	F(3,7)	F(5,7)	F(7,7)

Table 2-6. Pattern of 2-D Array After 2-D FFT

X(0,0)	X(2,0)	X(4,0)	X(6,0)	X(1,0)	X(3,0)	X(5,0)	X(7,0)
X(0,4)	X(2,4)	X(4,4)	X(6,4)	X(1,4)	X(3,4)	X(5,4)	X(7,4)
X(0,1)	X(2,1)	X(4,1)	X(6,1)	X(1,1)	X(3,1)	X(5,1)	X(7,1)
X(0,5)	X(2,5)	X(4,5)	X(6,5)	X(1,5)	X(3,5)	X(5,5)	X(7,5)
X(0,2)	X(2,2)	X(4,2)	X(6,2)	X(1,2)	X(3,2)	X(5,2)	X(7,2)
X(0,6)	X(2,6)	X(4,6)	X(6,6)	X(1,6)	X(3,6)	X(5,6)	X(7,6)
X(0,3)	X(2,3)	X(4,3)	X(6,3)	X(1,3)	X(3,3)	X(5,3)	X(7,3)
X(0,7)	X(2,7)	X(4,7)	X(6,7)	X(1,7)	X(3,7)	X(5,7)	X(7,7)

2.5. Real Data Fast Fourier Transform

Three different methods of calculating the real data fast Fourier transform are discussed in this section. The LH9124 and LH9320 provide instructions to implement all these three methods. Therefore, users can flexibly select a method based on his application requirement.

The first and straightforward method is by padding the imaginary data with zero and employing the complex fast Fourier transform discussed in section 2.2. Thus, the performance of the N-point real FFT is the same as that of the N-point complex FFT. The other two methods employ the symmetric properties of the DFT shown in Table 2-1 to improve the performance. The second method discussed in section 2.2.1 uses the complex FFT to compute two real sequences at the same time. The last method discussed in section 2.2.2 handles 2N-point real FFT with the N-point complex FFT.

2.5.1. Fast Fourier Transform for Two Real Data Sequences

Given two real sequences h(n) and g(n), a complex sequence x(n) can be obtained by setting x(n) = h(n) + jg(n). The DFT of x(n) can be expressed in terms of the DFT of h(n) and g(n) as follows:

$$X(k) = \sum_{n=0}^{N-1} x(n)\, W_N^{nk}$$

$$= \sum_{n=0}^{N-1} h(n)\, W_N^{nk} + j \sum_{n=0}^{N-1} g(n)\, W_N^{nk}$$

$$= H(k) + jG(k)$$

$$= [H_r(k) - G_i(k)] + j[H_i(k) + G_r(k)] \qquad [2AB]$$

where $H_r(k)$ and $H_i(k)$ are the real and imaginary parts of H (k). Similarly, the frequency sample at the point (N – k) can be represented by:

$$X(N-k) = \sum_{n=0}^{N-1} x(n)\, W_N^{n(N-k)} = \sum_{n=0}^{N-1} x(n)\, W_N^{-nk}$$

$$= \sum_{n=0}^{N-1} h(n)\, W_N^{-nk} + j \sum_{n=0}^{N-1} g(n)\, W_N^{-nk}$$

$$= H*(k) + jG*(k)$$

$$= [H_r(k) + G_i(k)] + j[-H_i(k) + G_r(k)] \qquad [2AC]$$

It can be derived From Equation [2AB] and Equation [2AC] that the frequency sample H(k) and G(k) can be represented in terms of the frequency samples of X(k) at the points k and (N-k) as follows:

$$H(k) = H_r(k) + jH_i(k)$$

$$= \frac{1}{2}\{\mathrm{Re}[X(k) + X(N-k)] + j\mathrm{Im}[X(k) - X(N-k)]\} \qquad [2AD]$$

and

$$G(k) = G_r(k) + jG_i(k)$$

$$= \frac{1}{2}\{\mathrm{Im}[X(k) + X(N-k)] - j\mathrm{Re}[X(k) + X(N-k)]\} \qquad [2AE]$$

where Re[X(k)] and Im[X(k)] denote the real and imaginary parts of X(k).

The LH9124 provides one instruction BRFT to calculate the recombing equations Equations [2AD] and [2AE]. Therefore, two N-point real FFTs can be obtained by the N-point FFT followed by one pass of recombining process. The LH9320 provides the instructions BRFTL and BRFTU to support the required addressing for the equations.

Table 2-7 compares the computing efficiency of two approaches of implementing the FFT on two real sequences with length N apiece. The first approach employs the general FFT computation with one sequence at a time. The imaginary part of the time sample is padded with zero. The second approach employs the method discussed in this section to do the two real FFTs at the same time. The number indicates the required machine cycles to complete two real FFTs including the latency. It is obvious from the table that two real sequences computed at the same time has better performance.

From the symmetric properties of the DFT for a real sequence in time-domain, the real part are symmetric and the imaginary part are antisymmetric with respect to the center point N/2 of the transformed sequence in frequency-domain. Thus, the samples from 0 to N/2 provide enough information to describe the whole frequency spectrum. The address instructions for generating full or half length of the sequences are both provided.

2.5.2. 2N-Point Real FFT by N-Point Complex FFT

This section will shows how a 2N-point real FFT is computed from an N-point complex FFT. A frequency sample for a 2N-point real DFT can be expressed as:

$$X(k) = \sum_{n=0}^{2N-1} x(n)\, W_{2N}^{nk}$$

$$= \sum_{n=0}^{N-1} x(2n)\, W_N^{nk} + W_{2N}^{k} \sum_{n=0}^{N-1} x(2n+1)\, W_N^{nk} \qquad [2AF]$$

Set h(n) = x(2n) and g(n) = x(2n + 1). X(k) then becomes:

$$X(k) = \sum_{n=0}^{N-1} h(n)\, W_N^{nk} + W_{2N}^{k} \sum_{n=0}^{N-1} g(n)\, W_N^{nk} \qquad [2AG]$$

$$= H(k) + W_{2N}^{k}\, G(k)$$

Set y(n) = h(n) + jg(n). From Equations [2AB] and [2AC], Y(k) can be represented by:

$$Y(k) = [H_r(k) - G_i(k)] + j[H_i(k) + G_r(k)] \qquad [2AH]$$

Table 2-7. Performance Comparison for Two Real N-Point FFTs

N	8	16	32	64	128	256	512	1024	2048	4096
Direct	104	136	300	428	976	1296	3380	6452	16724	24984
Recom.	78	102	200	296	634	922	2220	4268	10430	16606

and Y(N – k) can be represented by:

$$Y(N-k) = [H_r(k) + G_i(k)] + j[-H_i(k) + G_r(k)] \quad [2AI]$$

Combining Equations [2AG], [2AH], and [2AI], the real part of X(k) can be represented by:

$$X_r(k) = \frac{1}{2}Re[Y(k) + Y(N-k)] + \frac{1}{2}\cos(\pi k/N)\ Im[Y(k) + Y(N-k)] - \frac{1}{2}\sin(\pi k/N)\ Re[Y(k) - Y(N-k)] \quad [2AJ]$$

and the imaginary part of X(k) can be represented by:

$$X_i(k) = \frac{1}{2}Im[Y(k) - Y(N-k)] + \frac{1}{2}\sin(\pi k/N)\ Im[Y(k) + Y(N-k)] - \frac{1}{2}\cos(\pi k/N)\ Re[Y(k) - Y(N-k)] \quad [2AK]$$

The LH9124 provides one instruction BFCT to calculate the recombing equations Equations [2AJ] and [2AK]. Therefore, one 2N-point real FFT can be obtained by the N-point FFT followed by one pass of recombining process. The LH9320 provides the instructions BFCTL and BFCTU to support the required addressing for the equations.

There is one dead point that cannot be obtained from the computation of the above recombination process. This point is the highest frequence sample located at N of the 2N frequency samples. The data at this point is real and can be obtained by the following calculation

$$X(N) = H(0) - G(0) \quad [2AL]$$

If the time sequence is symmetric, X(N) will be zero. Otherwise, user may assign a proper value to the sample.

Table 2-8 compares the computing efficiency of two approaches in implementing a 2N-point real FFT. The first approach directly employs the general 2N FFT computation. The imaginary part of the time sample is padded with zero. The second approach employs the method discussed in this section by dividing the 2N-point sequence into two N-point sequences. The number indicates the required machine cycles to complete the 2N real FFT including the latency. As the number of points increased, the 2N real FFT based on the recombination process is more efficient.

2.6. Discrete Cosine Transform by FFT

The Discrete Cosine Transform (DCT) is a fundamental image processing for many image and video compression systems because of its suboptimal property and easy implementation. This section will discuss three methods of employing the chip set to implement the DCT. The DCT is defined as follows:

$$X(0) = \left(\frac{1}{N}\right)^{1/2} \sum_{n=0}^{N-1} x(n) \quad [2AM]$$

$$X(k) = \left(\frac{2}{N}\right)^{1/2} \sum_{n=0}^{N-1} x(n) \cos\left(\frac{\pi(2n+1)k}{2N}\right)$$

for k = 1, 2, … , N – 1 [2AN]

and the Inverse Discrete Cosine Transform (IDCT) is defined as follows:

$$x(n) = \left(\frac{1}{N}\right)^{1/2} X(0) + \left(\frac{2}{N}\right)^{1/2} \sum_{k=1}^{N-1} X(k) \cos\left(\frac{\pi(2n+1)k}{2N}\right)$$

for n = 0, 1, … , N – 1 [2AO]

2.6.1. N-Point DCT by Zero Extended 2N-Point FFT

The first method of computing N-point DCT extends the sequence from N points to 2N points by padding zeros to the last N data. The DCT in Equations [2AM] and [2AN] can be expressed as:

$$X(k) = Re\left[c(k)\, e^{-j\pi k/2N} \sum_{n=0}^{2N-1} x(n)\, W_{2N}^{nk} \right]$$

for k = 0, 1, … , N – 1 [2AP]

where

$$c(k) = \begin{cases} \left(\frac{1}{N}\right)^{1/2} & k = 0 \\ \left(\frac{2}{N}\right)^{1/2} & \text{otherwise} \end{cases} \quad [2AQ]$$

Equation [2AP] means that the N-point DCT can be obtained by doing 2N-point real FFT first and then taking the real part of complex multiplications as shown in Figure 2-10. Thus, the DCT can be implemented by the LH9124 in three phases. The first computes the N-point FFT. The second computes the recombining process. The final phase computes complex vector multiplications.

Table 2-8. Performance Comparison for Real 2N-Point FFT

2N	16	32	64	128	256	512	1024	2048	4096	8192
Direct	68	150	214	488	648	1690	3226	8364	12492	32990
2N Recom.	86	118	232	360	762	1178	2732	5292	12478	20702

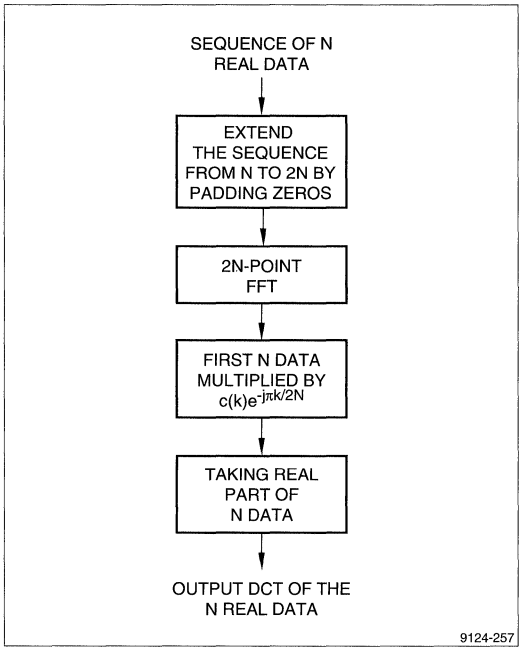


Figure 2-10. N-Point DCT by Zero Extended 2N-Point FFT

2.6.2. N-Point DCT by Symmetric Extended 2N-Point FFT

The second method of computing N-point DCT also extends the sequence from N points to 2N points. However, the extended 2N-point sequence is symmetric and defined as follows:

$$\hat{x}(n) = \begin{cases} x(n) & n = 0, 1, \ldots, N-1 \\ x(2N-1-n) & n = N, N+1, \ldots, 2N-1 \end{cases} \quad [2AR]$$

Then, the DFT of $\hat{x}(n)$ is given by:

$$\hat{X}(k) = \sum_{n=0}^{2N-1} \hat{x}(n)\, W_{2N}^{nk} \quad [2AS]$$

$$= 2e^{j\pi\frac{k}{2N}} \sum_{n=0}^{N-1} x(n) \cos\frac{\pi\,(2n+1)k}{2N}$$

Comparing Equation [2AS] with Equation [2AN], the DCT can be expressed as:

$$X(k) = \frac{c(k)}{2}\, e^{-j\pi k/2N}\, \hat{X}(k) \quad \text{for } k = 0, 1, \ldots, N-1 \quad [2AT]$$

Thus, the N-point DCT can be obtained by doing 2N-point real FFT first and then adjusting the results by complex multiplications as shown in Figure 2-11. As in the method 1, the DCT can also be implemented in the same three phases.

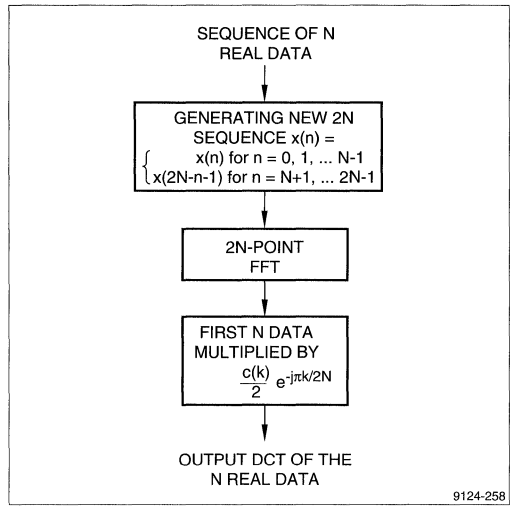


Figure 1-11. N-Point DCT by Symmetric Extended 2N-Point FFT

2.6.3 N-Point DCT by N-Point FFT

The first two methods introduce inefficiency in computation by extending the data sequence. The last method rearranges the data sequence instead of extending the data sequence. The new data sequence y(n) derived from the original sequence x(n) is represented by:

$$y(n) = \begin{cases} x(2n) & n = 0, 1, \ldots, N/2-1 \\ x(2N-2n-1) & n = N/2, N/2+1, \ldots, N-1 \end{cases} \quad [2AU]$$

Thus, the N-point DCT is obtained from the N-point FFT instead of the 2N-point FFT.

$$\hat{X}(k) = \sum_{n=0}^{N-1} x(n) \cos\frac{(2n+1)k\pi}{2N} \quad [2AV]$$

$$= \sum_{n=0}^{N/2-1} x(2n) \cos\frac{(4n+1)k\pi}{2N} +$$

$$\sum_{n=0}^{N/2-1} x(2n+1) \cos\frac{(4n+3)k\pi}{2N}$$

Setting $n = N-n-1$ in the second term of Equation [2AV], we may get:

$$\hat{X}(k) = \sum_{n=0}^{N/2-1} y(n) \cos\frac{(4n+1)k\pi}{2N} + \sum_{n=N/2}^{N-1} y(n) \cos\frac{(4n+1)k\pi}{2N} \quad [2AW]$$

$$= \sum_{n=0}^{N-1} y(n) \cos\frac{(4n+1)k\pi}{2N}$$

Thus, the DCT can be computed from the following equation:

$$X(k) = Re\left[c(k)\, e^{-j\pi k/2N} \sum_{n=0}^{N-1} y(n)\, W_N^{nk} \right] \quad [2AX]$$

It can also derived that:

$$X(N-k) = Im\left[c(k)\, e^{-j\pi k/2N} \sum_{n=0}^{N-1} y(n)\, W_N^{nk} \right]$$

$$\text{for } k = 0, 1, \ldots, N/2 \quad [2AY]$$

This means that we can only compute the first (N/2 + 1) points of the complex sequence. The first half of the DCT sequence can be obtained from the real part of the complex sequence and the second half of the DCT sequence can be obtained from the imaginary part of the complex sequence. The N-point DCT can be obtained by computing N-point real FFT first and then adjusting the results by complex multiplications as shown in Figure 2-12.

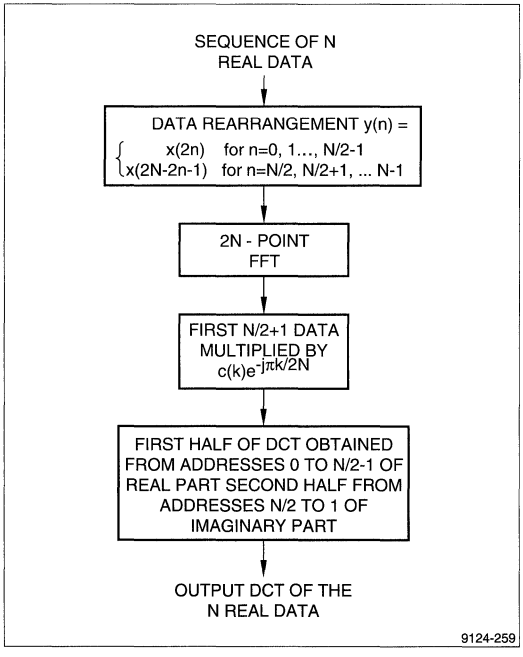


Figure 1-12. N-Point DCT of the N Real Data

2.6.4. Performance Comparison of the Three DCT Methods

Table 2-9 shows the computing efficiency of the DCT by three different methods discussed above. Method 1 extending the sequence to 2N by zero padding is discussed in section 2.6.1. Method 2 extending the sequence to a 2N symmetric sequence is discussed in section 2.6.2. Method 3 directly computing the rearranged N-sequence is discussed in section 2.6.3. The first two methods have the same performance by the chip set. The third method is obviously better than the first two. If two sequences are computed at the same time, the performance can be further improved. Method 3 is still the best among the three for 2 N-sequences. The figure in the table indicates the cycles required to finish the operation.

2.7. Two-Dimensional Discrete Cosine Transform

The discrete cosine transform has been widely recognized as the most effective technique among various transform coding methods for image and video signal compression. The 2-D DCT is defined as follows:

$$X(k_1,k_2) = c(k_1)c(k_2) \sum_{n_2=0}^{N_2-1} \sum_{n_1=0}^{N_1-1} x(n_1,n_2) \cos\frac{(2n_1+1)k_1\pi}{2N_1} \cos\frac{(2n_2+1)k_2\pi}{2N_2} \quad [2AZ]$$

where:

$$c(k) = \begin{cases} \left(\frac{1}{2}\right)^{1/2} & k=0 \\ 1 & \text{otherwise} \end{cases} \quad [2BA]$$

As in the 2-D DFT, the 2-D DCT can be computed by the 1-D column-row or row-column approach. Equation [2.52] can be decomposed into the following form for the column-row approach:

$$X(k_1,k_2) = c(k_2) \sum_{n_2=0}^{N_2-1} \left\{ c(k_1) \sum_{n_1=0}^{N_1-1} x(n_1,n_2) \cos\frac{(2n_1+1)k_1\pi}{2N_1} \right\} \cos\frac{(2n_2+1)k_2\pi}{2N_2} \quad [2BB]$$

Thus, the 2-D DCT can be computed as two phases of 1-D DCT. The benchmark of the 2-D DCT is shown in Table 1-1. It was derived by the method discussed in section 2.6.3.

2.8. Windowing

The consequence of the cycle nature of the DFT is the creation of an artificial jump at the endpoints of the data sequence. Direct truncation of the data sequence leads to the well-known Gibbs phenomenon which manifests itself as a modify the data sequence by window smoothing before applying the DFT.

Thus, given a data sequence x(n) and the window function w(n), the truncated data $x_w(n)$ sequence by the window effect is defined by:

$$x_w(n) = x(n) * w(n) \quad [2BC]$$

The straightforward window function is the rectangular window defined as:

$$w_R(n) = \begin{cases} 1 & \text{for } |n| \le \dfrac{N-1}{2} \\ 0 & \text{elsewhere} \end{cases} \quad [2BD]$$

The frequency response of the rectangular window is shown in Figure 2-13. The multiplication of the window function in the time-domain implies the convolution of the window function in the frequency-domain. In order to keep the shape of a original signal in the frequency-domain, desirable window characteristics are (1) reducing the width of the main lobe and (2) reducing the energy of the side lobes.

There are a lot of window functions proposed. User can select a proper window function for his application. One of the frequently used windows is the generalized Hamming window defined as:

$$w_H(n) = \begin{cases} \alpha + (1-\alpha)\cos(\dfrac{2\pi n}{N}) & \text{for } |n| \le \dfrac{N-1}{2} \\ 0 & \text{elsewhere} \end{cases} \quad [2BE]$$

where α is in the range between 0 and 1. If $\alpha = 0.54$, it is called a Hamming window. If $\alpha = 0.5$, it is called a Hanning window.

The window function can be combined with the first stage of radix-2 or radix-4 operation. Thus, no extra phase is required for the windowed FFT operation. The coefficients combining the twiddle factors and window coefficients are stored in the C port memory either in linear order or digit-reverse order. There are two instructions BWND2 and BWND4 provided by the LH9124 to support the first stage of the windowed FFT operation. The address pattern for twiddle factors will be changed for this stage. The twiddle factor address pattern may use the same as the input data address pattern.

Table 2-9. Performance Comparison of Three DCT Methods

N	ONE N-SEQUENCE			TWO N-SEQUENCES		
	METHOD 1	METHOD 2	METHOD 3	METHOD 1	METHOD 2	METHOD 3
8	94	94	75	152	152	108
16	152	152	95	250	250	140
32	264	264	185	378	378	254
64	442	442	265	780	780	382
256	1452	1452	795	2750	2750	1200
512	3262	3262	1965	5310	5310	2754
1K	6334	6334	3245	10430	10430	5314
4K	24816	24816	14559	49410	49410	20724

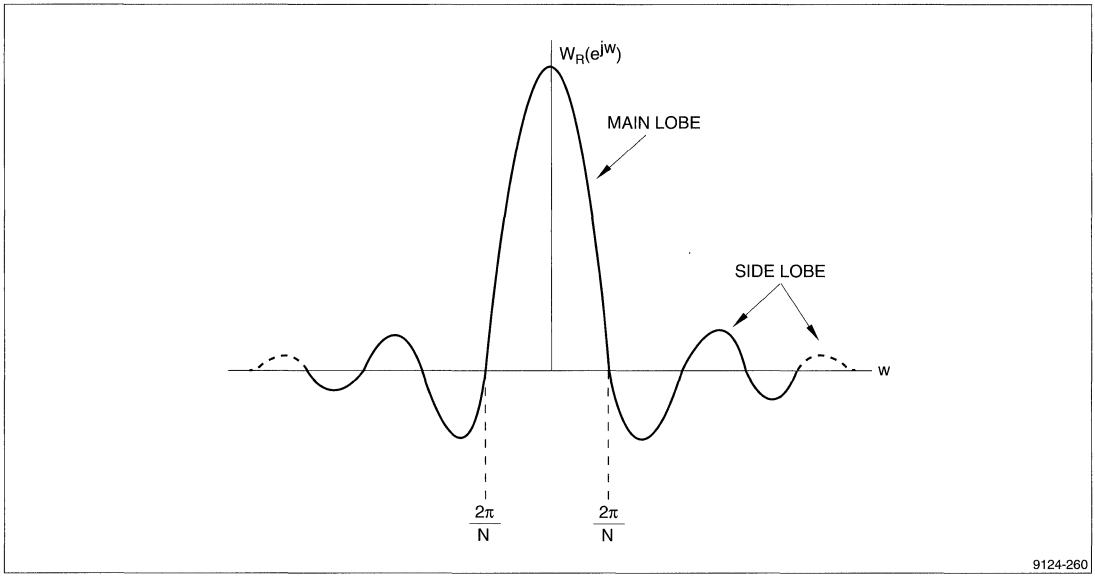


Figure 2-13. Frequency Response of a Rectangular Window

3. BLOCK FLOATING-POINT ARCHITECTURE

The LH9124 supports block floating-point arithmetic to improve the precision of computations for a fixed-point processor. The FFT algorithm for its inherent block processing structure is suited to block floating-point arithmetic. The block structure means that the FFT algorithm is decomposed into pipelined computing stages and that the block output data of the current stage is the block input data of the next stage as shown in Figure 2-1. The processor with block floating-point can have the advantages of both fixed-point and floating-point. First, the architecture is simple because both multiplications and additions are fixed-point. Second, the block floating-point can even get higher precision than the floating-point with the same word length.

Fundamentally, the block floating-point mechanism of the LH9124 consists of two parts. The output part is to calculate the maximum magnitude of the output data for the current stage of radix operations. Then, a scaling factor is derived as a reference value for the input scaling of the next stage. The input part receives the scaling factor generated by the previous stage. The number of bits to be shifted right for the current input data will be based on the scaling factor and the instruction to be executed. Therefore, the data overflow and the precision of integer operations are automatically controlled by the block floating-point mechanism.

3.1. Output Part of Block Floating-Point Mechanism

The input and output data for the FFT are complex numbers with real and imaginary part having 24 bits apiece. There is no loss of precision in the intermediate stage of data computations because the intermediate values can grow to 60 bits. The error occurs when the 60-bit data in accumulators are rounded to the 24-bit output data. This means that the error occurs at the output stage of data transfer. However, the block floating-point mechanism will make the output block data as precisely as possible.

The magnitude of the computed output data is compared to some threshold values to generate a scaling factor output (DSFO) parameter according to Table 3-1. The LH9124 will keep the maximum value of DSFO. If the new generated DSFO is larger than the maximum value, the maximum value will be updated. Table 3-1 lists the threshold values to be compared by the magnitude of complex data x. The value x_m is computed by the following equation:

$$x_m = (|x_r|^2 + |x_i|^2)^{1/2} \quad [3A]$$

where $|x_r|$ and $|x_i|$ denote the magnitudes of the most significant five bits of real and imaginary data x, respectively.

Table 3-1. The Scaling Factor Output DSFO Lookup Table

THRESHOLD MAGNITUDE	DSFO
$x_m \geq 1.000488$	5
$1.000488 > x_m \geq 0.5$	4
$0.5 > x_m \geq 0.25$	3
$0.25 > x_m \geq 0.125$	2
$0.125 > x_m \geq 0.0625$	1
$0.0625 > x_m$	0

3.2. Input Part of Block Floating-Point Mechanism

One input pin DSFISEL controls the way of scaling. If DSFISEL goes high, the number of bits to be shifted right for input data is directly controlled by the three scalar factor input pins DSFI[2:0]. In this user mode, the right shifting of the input data can be 0 to 7 bits. If DSFISEL goes low, the block floating-point automatic scaling will control the input data shifting. In this case, the three scalar factor input pins will be connected to the three scalar factor output pins. In the block automatic mode, the number of bits to be shifted is a function of the instruction to be executed and the scaling factor output generated by the previous stage as shown in Table 3-2. Moreover, the maximum number of bits for shifting is 5. In the FFT applications, the block automatic mode is suggested because the optimum performance is usually obtained by the automatic scaling. Thus, the overflow control and the precision for integer arithmetic are simplified and users only take care of the first stage of the input data.

Table 3-2. Input Data Auto-Scaling Lookup Table

DSFI	RADIX-2	RADIX-4	RADIX-16
5	2	3	5
4	1	2	4
3	0	1	3
2	0	0	2
1	0	0	1
0	0	0	0

3.3. Scaling Factor Accumulator

The block-floating point mechanism provides information not only for the bit growing of the current stage but also for the total bit growing of all the previous stages. The 6 block floating-point output pins BFPO[5:0] show the total number of bits that input data being shifted right from the first stage. There are 6 block floating-point input pins BFPI[5:0] also. The BFPI pins are usually connected to the BFPO pins. The output value of BFPO is obtained by adding the input value of BFPI with the number of bits that

the input data are shifted for the stage. This means that the number of input data shifting can be continuously accumulated. Therefore, the performance of the 24-bit block floating-point is comparable to that of the 32-bit floating point because the exponent part of the floating-point is implicitly implemented by BFPI and BFPO.

3.4 Performance Analysis of Block Floating-Point

This subsection will employ an example[10] to analyze the signal to roundoff noise ratio of the fixed-point, floating-point and block floating-point arithmetic. It can be seen that the performance of the block floating-point is very close to that of the floating-point. In addition, the performance of the FFT implemented by the radix-4 or radix-16 is similar to that implemented by the radix-2, even though the radix-4 or radix-16 implementation can have double or quadruple the speed than the radix-2 implementation.

The input time samples are defined as follows:

$$x(n) = \{\sin(2\pi * 0.1245 * n) + j\,2^{-15} * \sin(2\pi * 0.2505 * n)\} * w(n)$$

for n = 0, 1, 2, ..., 1023 [3B]

where w(n) is a window function.

3.4.1 1K FFT With Three-Term Black-Harris Window

The minimum three-term Black-Harris [9] window can achieve a sidelobe level of −67 dB. The window is defined as follows:

$$w(n) = 0.42323 - 0.49755 * \cos\left(\frac{2\pi}{1024}n\right) + 0.07922 * \cos\left(\frac{2\pi}{1024}2n\right)$$

for n = 0, 1, 2, ..., 1023 [3C]

The power spectrum of the input signal illustrated in Figure 3-1 clearly shows two harmonic signals and the spectral shape of the Black-Harris window. The power spectrum are obtained by executing five radix-4 stages of the 1K FFT with block floating-point arithmetic. The roundoff noise is hardly to be seen because the sidelobe effect of the window dominates. Therefore, the block floating-point has almost the same performance as the double precision floating-point.

3.4.2 1K FFT With Four-Term Black-Harris Window

The low level harmonic signal can be more clearly seen if the four-term Black-Harris window is employed. The minimum four-term window can achieve a sidelobe level of −92 dB and is defined by:

$$w(n) = 0.35875 - 0.48829 * \cos\left(\frac{2\pi}{1024}n\right) + 0.14127 * \cos\left(\frac{2\pi}{1024}2n\right) - 0.01168 * \cos\left(\frac{2\pi}{1024}3n\right)$$

for n = 0, 1, 2, ..., 1023 [3D]

3.4.2.1. Double Precision Floating-Point Computation

The power spectrum of the 4-term windowed harmonic signals shown in Figure 3-2 are generated by the software package *MathCAD.*[11] The computation is through double precision floating-point. The two harmonic signals and the spectral sharp of the four-term Black-Harris window can be clearly seen.

3.4.2.2. Block Floating-Point Computation

Figures 3-3, 3-4 and 3-5 show the power spectrum of the four-term windowed harmonic signal generated by the LH9124 and the LH9320. Figure 3-3 is generated by the 1K FFT with all radix-2 stages, Figure 3-4 with all radix-4 stages and Figure 3-5 with one radix-4 stage followed by two radix-16 stages. The total number of shifting for the radix-2 implementation is eight and for the other two implementations is nine. In this case, the signal level of the radix-2 implementation is higher. However, the radix-4 implementation may have less roundoff noise. Therefore, there is no obvious performance difference between the radix-2 and radix-4 implementation. It can be seen that the performance of these block floating-point implementations are very close to that of double precision floating-point implementation.

3.4.2.3. Fixed-Point Computation

Figure 3-6 shows the power spectrum of the four-term windowed harmonic signal generated by the LH9124 and LH9320. It is generated by 10 radix-2 stages with the block floating-point mechanism turned off. There are 10-bit shift in total with one bit shift per stage. It can be seen that the noise level of the fixed-point implementations are obviously higher than that of block floating-point implementation because the signal level is lower.

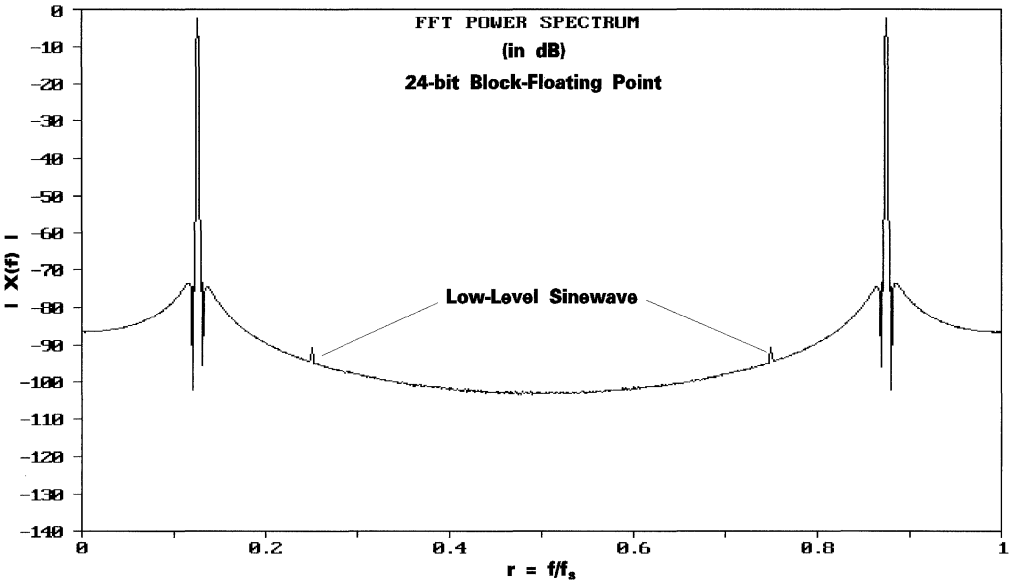


Figure 3-1. 1K FFT Using Radix-4 With Three-Term Black-Harris Window

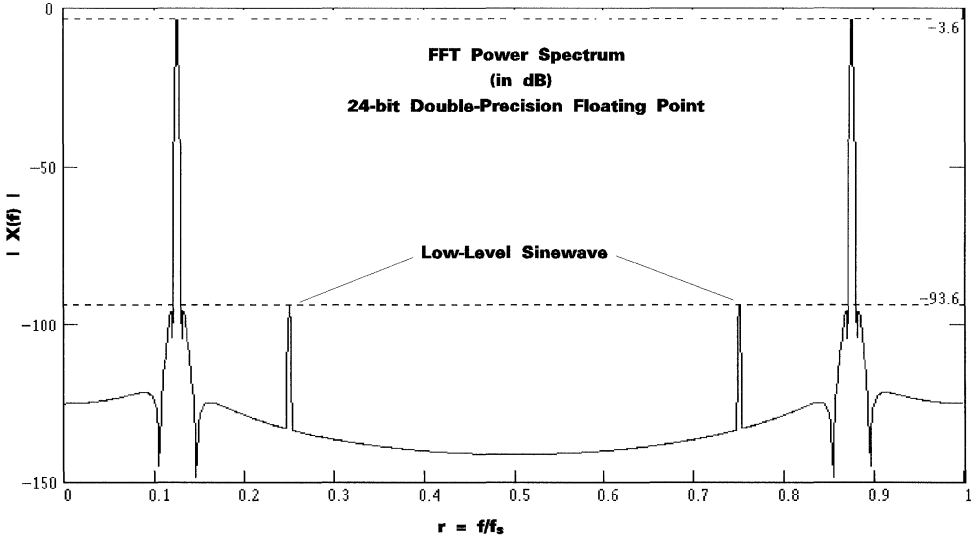


Figure 3-2. 1K Double-Precision Floating-Point FFT With Four-Term Black-Harris Window

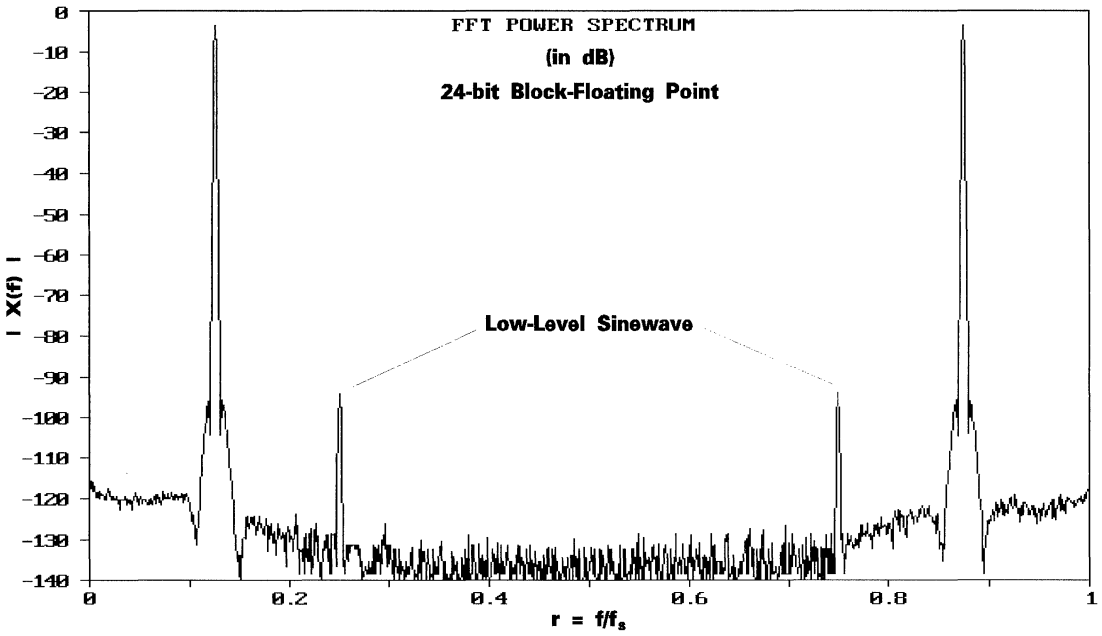


Figure 3-3. 1K FFT Using Radix-2 With Four-Term Black-Harris Window

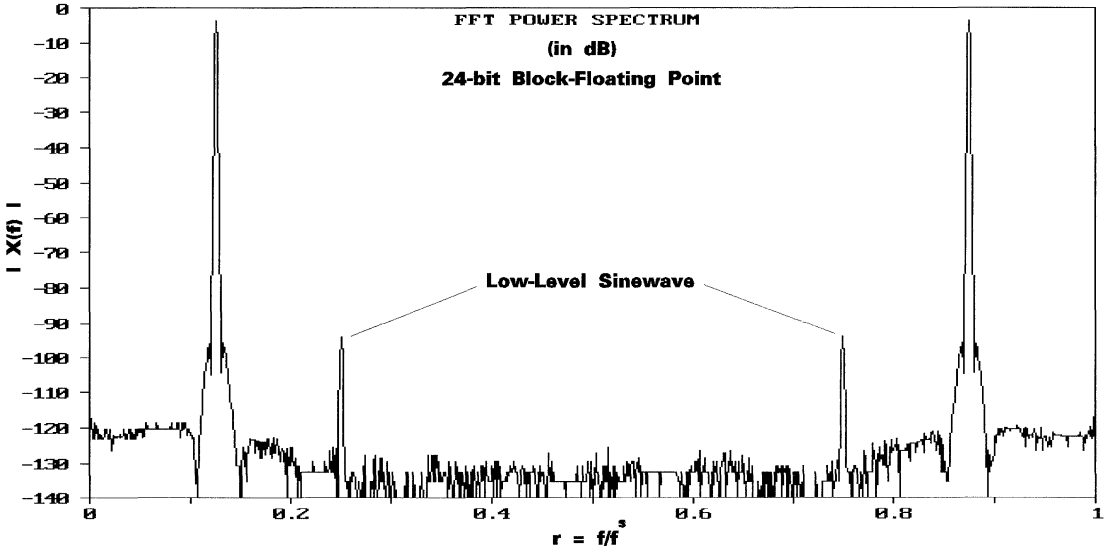


Figure 3-4. 1K FFT Using Radix-4 With Four-Term Black-Harris Window

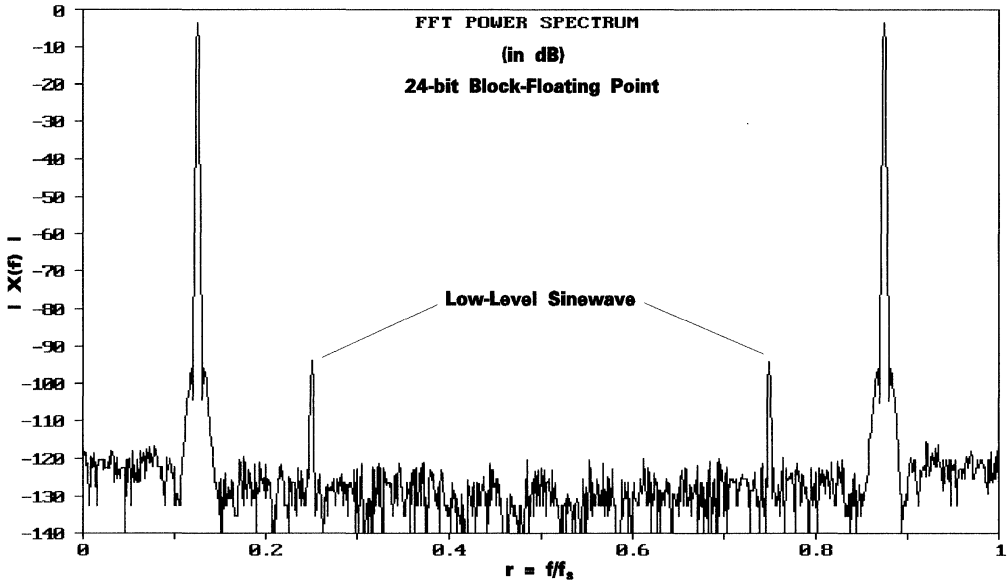


Figure 3-5. 1K FFT Using Mixed Radices 4 X 16 X 16
With Four-Term Black-Harris Window

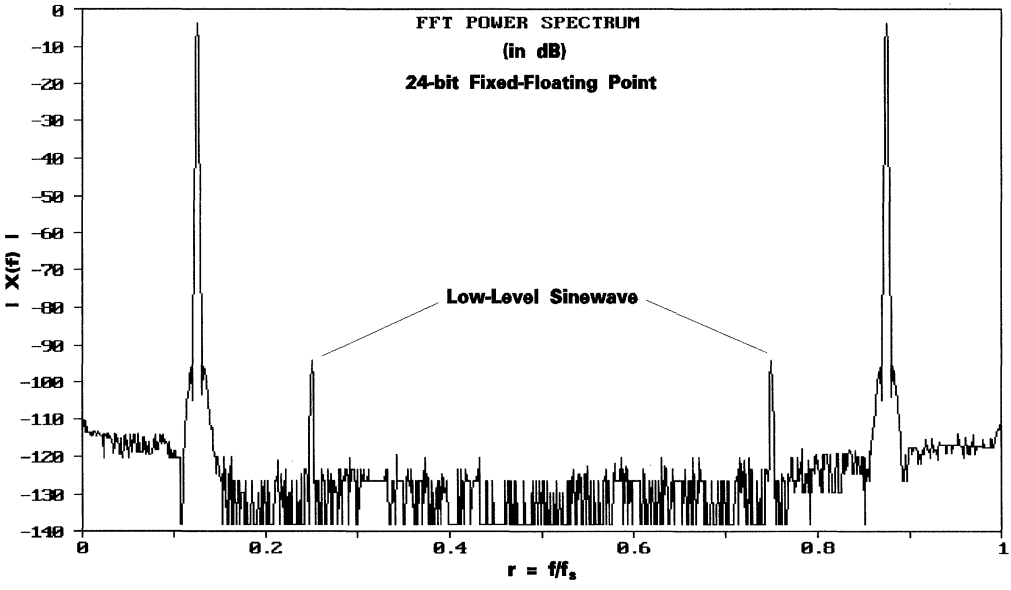


Figure 3-6. 1K FFT Using Radix-4 With Four-Term
Black-Harris Window Having Block Floating-Point Turned Off

4. SYSTEM CONFIGURATIONS

The LH9124 is a by-pass form digital signal processor. It is an execution unit without on-chip memory. Moreover, it provides a flexible four bidirection I/O port structure. Therefore, combined with a wide variety of memory structures, it can be easily configured into cascaded or paralleled architectures. The speed of a system is linearly increased with the number of the chips. The throughput and latency both can be improved by the parallel architectures to match the user's real-time application requirement. The benchmark of the FFT for parallel architectures is shown in Table 1-2. For more detailed information about system configurations, please refer to *LH9124 User's Guide.*

4.1. Single Chip Architectures

The LH9124 consists of four bidirection 48-pin I/O ports: A, B, C, and Q as shown in Figure 4-1. Data through these ports are complex. Twenty-four of the pins are for real data and the other 24 pins for imaginary data. The C port is usually employed as a coefficient port. Any two of the A, B, and Q data ports can be selected as input and output ports for the execution of the LH9124. These three ports provide the same function for implementation. The Q port may be employed as a acquisition port. Hence, Q port will be an input port in the first pass and an output port in the last pass of the FFT.

The LH9124 does not provide on-chip RAM and address generator. Thus, each port may be associated with one external RAM and address generator as shown in Figure 4-1. The LH9320 is a address generator for supporting the addressing required by the LH9124 instructions. Hence, it is a good candidate to be selected for implementation. The data flow of the system is controlled by three DF pins which specify the dedicated input and output ports as shown in Table 4-1.

Table 4-1. LH9124 Data Flow Function Set for FFT Instructions

MNEMONIC	DF[2:0]	DESCRIPTION
RAWB	111	Input data through A port and output data through B port
RAWQ	101	Input data through A port and output data through Q port
RBWA	000	Input data through B port and output data through A port
RBWQ	100	Input data through B port and output data through Q port
RQWA	001	Input data through Q port and output data through A port
RQWB	011	Input data through Q port and output data through B port

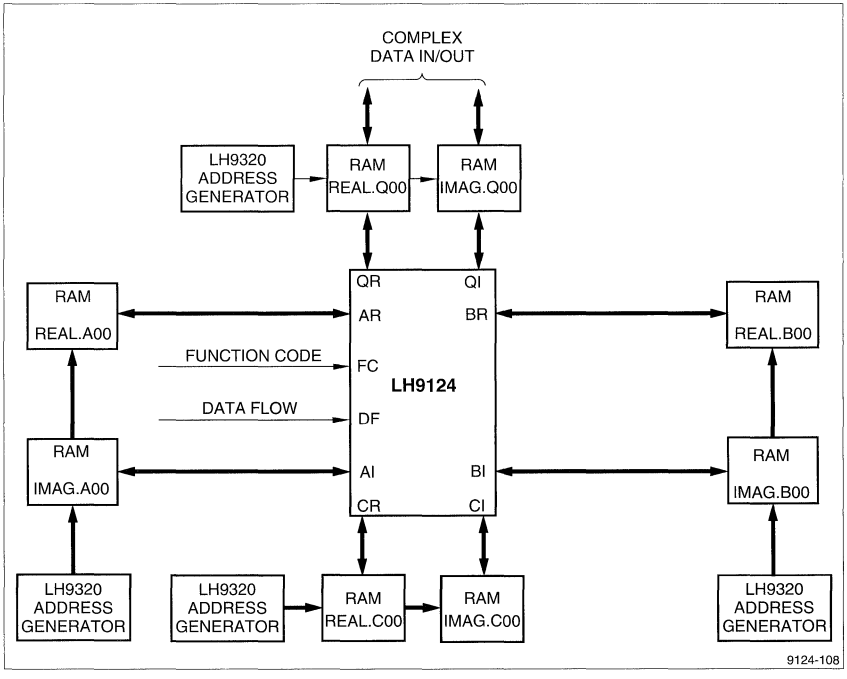


Figure 4-1. Single LH9124 System With Four Port Architecture

There are four input control pins that can change the values of the data and coefficients either for the input and output ports. The control pins DCI and DCO are used in conjunction with the data flow function. The data of the input sequence loaded to the LH9124 will be conjugated when DCI is set. Similarly, the data of the computed results transferred to the output buffer will be conjugated when DCO is set. The inverse FFT can be computed the same as the FFT if these two pins are properly controlled. The other two pins CCR and CCI are used by the coefficient port. The sign of the real part of the input coefficient will be changed when CCR is set. CCI will control the sign of the imaginary of the input coefficient. The memory size of the twiddle factors can be saved through the control of these two pins. The input data can also be scaled as discussed in Section 3.

The four port design of the LH9124 provides users the flexibility in system configurations. Users can use this flexibility either to optimize the system performance or to reduce the system cost. Two of system configurations are shown in Figure 4-1 and Figure 4-2. Figure 4-1 is a four port architecture and Figure 4-2 is a three port architecture. There are a lot of possible architectures depending on the user's application requirement.

4.2. Multi-Chip Architectures

The FFT algorithm is computing-intensive and its architecture has some form of regularity. Therefore, it is a good candidate to be implemented by the cascaded or parallel architecture. The throughput can be improved by the cascaded architecture and the latency can be improved by the parallel architecture. The LH9124 is more suitable in cascaded or parallel architectures than the contemporary FFT chips because of its by-pass form structure.

4.2.1. Cascaded Architectures

The decomposition process of the DFT shown in Figure 2-1 points out that the FFT inherently has the cascaded form. Each stage of the FFT can be implemented by one radix instruction of the LH9124. Hence, the N-stage FFT can be implemented by N cascaded LH9124 processors. The block output data of the i-th stage will be the block input data of the (i + 1)-th stage. The computation of the (i + 1)-th stage will wait for the completion of the i-th stage because of data dependency.

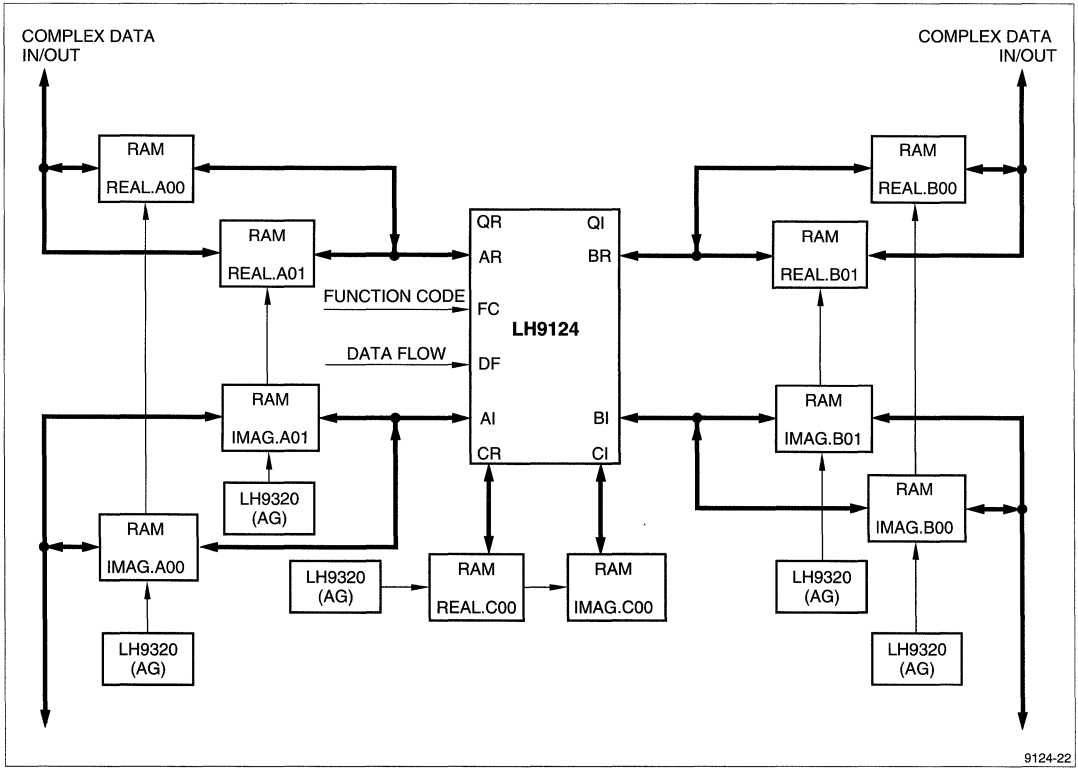


Figure 4-2. Single LH9124 System With Three Port Architecture

The computing efficiency will be very low if only one sequence of the FFT is to be computed. In this case, the performance of the cascaded structure will be the same as that of a single chip structure. If a lot of sequences with the same length are to be computed, the computing efficiency will be improved by a factor equal to the number of cascaded stages. In this way, the processor i is computing the n-th sequence instead of waiting for the completion of the (n + 1)-th sequence of the processor i – 1 as shown in Table 4-2. This means that double buffers have to be used in the interface between two stages as shown in Figure 4-3.

The input and output data rate in all the cascaded stages are the same because all the radix instructions use the same time to finish the same length of data. Figure 4-3 shows a three stage cascaded structure. The time required to compute the 4K FFT is 312.3 μsecs for a single chip structure. However, the 4K FFT can be completed in every 102.4 μsecs by the cascaded structure. This means that the speed is improved by more than triple.

4.2.2. Parallel Architectures

The alternative way to improve the throughput for the FFT is obtained by parallel processing. The advantage of the parallel architecture is that each processor can run an independent job or algorithm. For an N-point FFT with M stages, the time required for a single LH9124 to finish the operation is about N * M cycles. If the data transfer is the same as the machine cycle, the N-point data loaded to one processor will be N cycles. By the method of data multiplexing, the N * M points of data can be loaded to M processors with each processor having N points. Thus, the computing-bound and I/O bound will be balanced for the whole system. The throughput is improved by a factor of M. The schedule of data sequences for 3 parallel processors is listed in Table 4-2.

Figure 4-4 shows a common bus parallel system with three processors. A 4K-point FFT can be implemented by three stages of radix-16. It takes about 312.3 μsecs for a single chip system or 104.1 μsecs for the parallel system for every N-point FFT. Thus, the speedup factor is three for a three processor parallel system.

Table 4-2. Schedule of Sequences Processed by Cascaded and Parallel Structures

INPUT	CASCADED			PARALLEL			OUTPUT
SEQUENCE	PROC. 1	PROC. 2	PROC. 3	PROC. 1	PROC. 2	PROC. 3	SEQUENCE
n	n – 1	n – 2	n – 3	n – 3	n – 2	n – 1	n – 4
n + 1	n	n – 1	n – 2	n	n – 2	n – 1	n – 3
n + 2	n + 1	n	n – 1	n	n + 1	n – 1	n – 2
n + 3	n + 2	n + 1	n	n	n + 1	n + 2	n – 1
n + 4	n + 3	n + 2	n + 1	n + 3	n + 1	n + 2	n

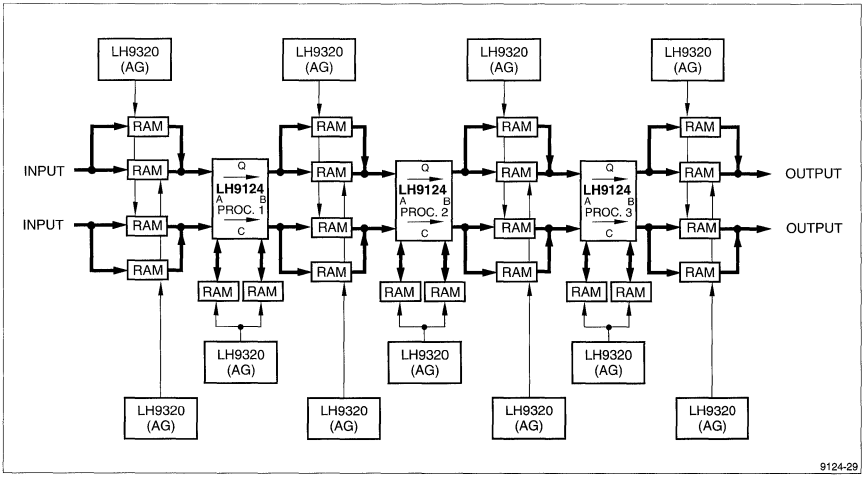


Figure 4-3. Three-Stage Cascaded System

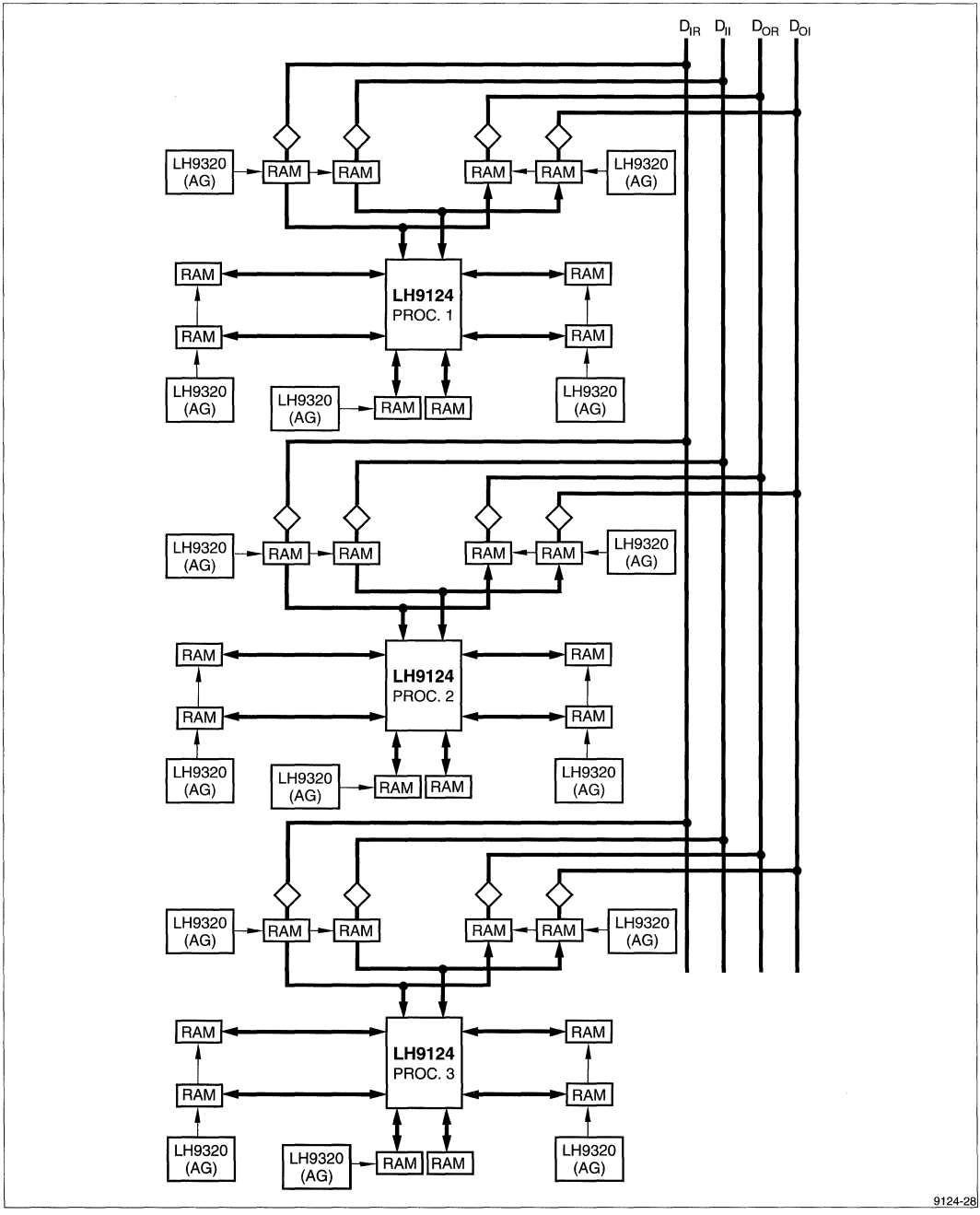


Figure 4-4. Common Bus Parallel System With Three Processors

4.2.3. Cascaded and Parallel Combined Architectures

The parallel systems discussed in the above two subsections do improve the throughput of the FFT. However, the latency time for an arbitrary N-point FFT is still the same. If the latency time is the main concern in the system design, the computing time in each stage has to be reduced. Since the structure in each stage is regular, it can be further partitioned by a factor k that is a power of two. Thus, the computing time for the stage will be reduced from N to N/k. The complexity increased in interface between stages is the cost for this improvement.

One of the architectures that improve both the throughput and latency is shown in Figure 4-5. It has three cascaded stages and two parallel processors in each stage. The latency for a 4K-point FFT is 312.3 µsecs for a single LH9124 system and 153.2 µsecs for the parallel system. The time required for a 4K-FFT is 312.3 µsecs for a single LH9124 system and 51.2 µsecs for the parallel system. Thus, the throughput is improved by a factor of 6 and the latency is improved by a factor of two for the proposed parallel system. For two processors in each stage, the dual-port memory is one of the methods to solve the interface problem.

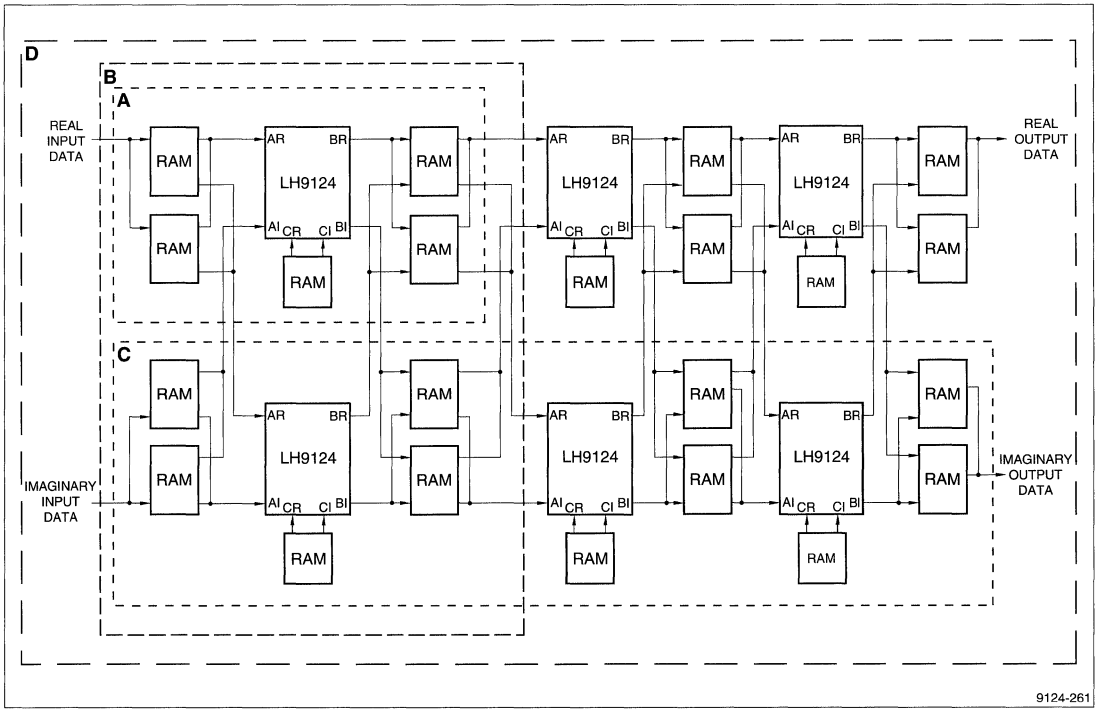


Figure 4-5. System With Three-Stage Cascaded and Two Parallel in Each Stage

5. LH9124'S FFT EXECUTION INSTRUCTIONS

One of the salient features of the LH9124 is its programming in function block level. Therefore, the instruction codes of the LH9124 will be ready as soon as the function block diagram is defined. For example, the DCT of two real 128-point sequences is computed by the method shown in section 2.6.1. The function block diagram to implement the example is shown in Figure 5-1. The two 128-point sequences is first zero-extended to two 256-point sequences and then form a 256-point complex sequence. The 256-point FFT is executed by two radix-16 stages. Then, the recombining process of transforming one complex FFT into two independent real FFTs is performed. Finally, the DCT can be obtained by taking the real part of two complex sequence multiplications. It can be seen that each function block is mapped into one instruction code. Therefore, it is easy for users to use the LH9124 to implement the DSP algorithms. Especially, the LH9320 is optimized to be used with the LH9124 to provide the required addressing sequences. However, users have the flexibility of selecting other devices such as PLD to replace the LH9320.

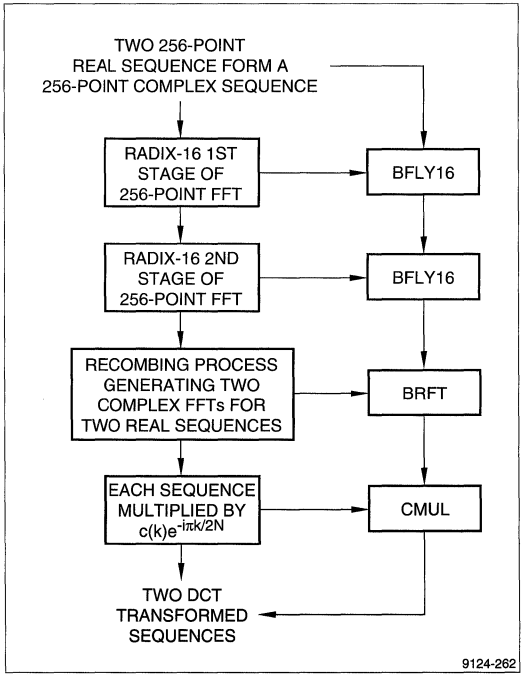


Figure 5-1. Mapping From Function Block Diagram to the LH9124 Instructions

The coding for the FFT is extremely easy by employing the LH9124 and LH9320. The instruction codes for the FFT algorithm are based on the stages of the FFT decomposition. The 8-point FFT shown in Figure 2-2 is decomposed into three radix-2 two stages. Thus, the FFT can be implemented by a series of three BFLY2 instructions. The 8-point FFT shown in Figure 2-5 is implemented by one radix-4 stage followed by one radix-2 stage. The corresponding instruction codes will be BFLY4 followed by BFLY2. Therefore, the LH9124 instruction codes will be defined once the stages of the FFT structure are decided.

The inverse FFT (IFFT) can be implemented almost the same as the FFT. The only difference is that the input data is conjugated in the first stage and the output data is conjugated at the last stage. The LH9124 contains two input pins DCI and DCO to control the conjugate operation of the input and output data, respectively.

5.1. Radix-2 Butterfly Instruction: BFLY2 or BWND2

The radix-2 butterfly is the fundamental computing module for FFT algorithms. The radix-2 butterfly is generally represented by the flow graph shown in Figure 2-3. The input-output relation of a butterfly operation is expressed as:

$$X_{k+1}(a) = X_k(a) + W_N^r X_k(b) \quad [5A]$$

$$X_{k+1}(b) = X_k(a) - W_N^r X_k(b) \quad [5B]$$

These two equations are executed by the instruction BFLY2 with throughput of two cycles per radix-2 butterfly. The instructions BF2i and TF2i provide the required data and twiddle factor address patterns.

If the radix-2 butterfly is combined with the window function in the first stage of the FFT operation, the input-output relation for this stage is represented by the following equations:

$$X_{k+1}(a) = X_k(a)\, w(a) + X_k(b)\, w(b) \quad [5C]$$

$$X_{k+1}(b) = X_k(a)\, w(a) - X_k(b)\, w(b) \quad [5D]$$

These two instructions are executed by the instruction BWND2 with throughput of two cycles per butterfly. It can be seen from Equations [5C] and [5D] that the address indices for both input and window functions are the same. Thus, the address pattern of windowed radix-2 coefficients can use the same as that of input data. The address instruction RBF0 or BF20 may be employed depending on the order of the coefficients. There are 18 cycle latency for BFLY2 and 20 cycle latency for BWND2 from loading the first data to the LH9124 to receiving the first computed result.

5.2. Radix-4 Butterfly Instruction: BFLY4 or BWND4

The LH9124 also supports a radix-4 butterfly instruction BFLY4 to improve the FFT computing efficiency. The input-output relation of the radix-4 butterfly with notation shown in Figure 2-6 are represented by:

$$X_{k+1}(a) = X_k(a)W_N^0 + X_k(b)W_N^r + X_k(c)W_N^{2r} + X_k(d)W_N^{3r} \quad [5E]$$

$$X_{k+1}(b) = X_k(a)W_N^0 - jX_k(b)W_N^r - X_k(c)W_N^{2r} + jX_k(d)W_N^{3r} \quad [5F]$$

$$X_{k+1}(c) = X_k(a)W_N^0 - X_k(b)W_N^r + X_k(c)W_N^{2r} - X_k(d)W_N^{3r} \quad [5G]$$

$$X_{k+1}(d) = X_k(a)W_N^0 + jX_k(b)W_N^r - X_k(c)W_N^{2r} - jX_k(d)W_N^{3r} \quad [5H]$$

The throughput is four cycles per radix-4 butterfly. With the same input-output map, a radix-4 butterfly is equivalent to the four radix-2 butterflies shown in Figure 2-7. The twiddle factor relationship between Figure 2-6 and Figure 2-7 is shown as follows:

$$W_N^r = W_N'' \quad [5I]$$

$$W_N^{2r} = W_N' \quad [5J]$$

$$W_N^{3r} = W_N'W_N'' = j\,W_N'W_N''' \quad [5K]$$

The instructions BF4i and MXB24i provide the required data address patterns. In addition, the instructions TF4i and MXT24i provides the required twiddle factor address patterns. If the radix-4 butterfly is combined with the window function in the first stage of the FFT operation, the input-output relation for this stage is represented by the following equations:

$$X_{k+1}(a) = X_k(a)w(a) + X_k(b)w(b) + X_k(c)w(c) + X_k(d)w(d) \quad [5L]$$

$$X_{k+1}(b) = X_k(a)w(a) - jX_k(b)w(b) - X_k(c)w(c) + jX_k(d)w(d) \quad [5M]$$

$$X_{k+1}(c) = X_k(a)w(a) - X_k(b)w(b) + X_k(c)w(c) - X_k(d)w(d) \quad [5N]$$

$$X_{k+1}(d) = X_k(a)w(a) + jX_k(b)w(b) - X_k(c)w(c) - jX_k(d)w(d) \quad [5O]$$

These four instructions are executed by the instruction BWND4 with throughput of four cycles per butterfly. It can be seen from Equations [5L] to [5O] that the address indices for both the input and window data are the same. Thus, the address pattern of windowed radix-4 coefficients can use the same as that of input data. The address instruction RBF0 or BF40 may be employed according to the order of the coefficients. There are 18 cycle latency for BFLY4 and 20 cycle latency for BWND4 from loading the first data to the LH9124 to receiving the first computed result.

5.3. Radix-16 Butterfly Instruction: BFLY16

The architecture of the LH9124 is optimized for this instruction. Thus, the peak performance can be obtained with 6 multiplications and 11 additions per cycle. I/O bound and computing bound are well-balanced. With the same input-output map, the structure of a radix-16 is equivalent to the eight radix-4 butterflies connected in the form as Figure 2-8 or to the thirty-two radix-2 butterflies connected in the form as Figure 2-9. The instructions BF16i, MXB216i, and MXB2416i provide the required data address patterns. In addition, the instructions TF16i, MXT216i, and MXT2416i provides the required twiddle factor address patterns.

The throughput is 16 cycles per radix-16 butterflies. There are 68 cycle latency from loading the first data to receiving the first computed result. The latency of the instruction BFLY16 is longer than that of the instructions BFLY2 and BFLY4. However, the computing efficiency of the instruction BFLY16 is two times higher than that of BFLY4 and four times higher than that of BFLY2.

5.4. Recombining Two Real FFT Instruction: BRFT

The two real FFTs can be computed by one complex FFT. Then, the complex FFT results can be decomposed and recombined into two complex sequences representing two real FFTs. The instruction BRFT performs the function of final decomposition and recombination as represented by Equations [2AD] and [2AE]. In the process, two complex data from the i-th and (N-i)-th elements of the complex FFT results are fetched and computed to get the results for the i-th element of each real FFT.

The LH9320 instructions BRFTL and BRFTU provide the input and output data addressing for BRFT. Since the spectrum for the real time samples has the symmetric property, only half of the frequency samples can provide all the information. To save time or storage, only half of frequency samples are computed. The LH9320 instructions BRFTLS and BRFTUS provide the input and output data addressing to compute the first half of frequency samples for each real input sequence. The instruction has 18 cycles of latency.

5.5. Recombining 2N Real FFT instruction: BFCT or BFCT2

The 2N-point real FFT can be computed by a N-point complex FFT. Then, the complex FFT results can be decomposed and recombined into the 2N-point real FFT results. The instruction BFCT or BFCT2 performs the function of final decomposition and recombination as represented by Equations [2AJ] and [2AK]. In the process, two complex data from the i-th and (N – i)-th elements of the complex FFT results are fetched and computed to get the results for the i-th element or the (2N – i)-th element of the 2N real FFT. The instruction BFCT generates the result for the i-th element only and BFCT2 generates the results for the i-th and (2N – i)-th elements both.

The LH9320 instruction BFCTL provides the input data addressing for BFCT and BFCT2 both. The instructions BFCTUS and BFCTU provides the output data addressing for BFCT and BFCT2, respectively. Since the spectrum for the real time samples has the symmetric property, the (2N – i)-th sample will be the complex conjugate of the i-th sample. The instruction BFCTT will provide the coefficient addressing for both instructions. Both instructions have 18 cycles of latency.

5.6. DFT or DCT Instruction: BCFIR or BDFIR

Both instructions basically compute the vector dot operation required for the FIR operation. BCFIR computes a complex vector dot operation with one tap per cycle and BDFIR computes two real vector dot operation with one tap per cycle. There are 18 cycles of latency for both instructions. The equations represented by the BCFIR for two complex sequences {h(n)} and {x(n)} are expressed as:

$$y_r = \sum_{j=0}^{\infty} [h_r(j) * x_r(j) - h_i(j) * x_i(j)] \quad [5P]$$

and:

$$y_i = \sum_{j=0}^{\infty} [h_r(j) * x_i(j) + h_i(j) * x_r(j)] \quad [5Q]$$

The equations represented by the BDFIR for two complex sequences {h(n)} and {x(n)} are expressed as:

$$y_r = \sum_{j=0}^{\infty} h_r(j) * x_r(j) \quad [5R]$$

and:

$$y_i = \sum_{j=0}^{\infty} h_i(j) * x_i(j) \quad [5S]$$

Thus, the instruction BDFIR sees the complex number as two independent real numbers and do multiplications for each part. The sum of product operation for both instructions will be infinitely executed as the function is continuously applied. The output data pins will be updated according to the data in the accumulator. Users can latch the output data at the proper time. The START signal will clear the accumulator and users use this signal to initiate a new sum of product operation and control the vector length for each output. No overhead occurs between two consecutive outputs.

5.7. General Complex Arithmetic Instruction: CADD, CMAG, CMUL, or CSUB

The instructions CADD, CMUL, and CSUB are complex operators which execute on two complex input data to get one complex output data. The instruction CMAG computes the magnitude square of a complex data to get one real data. If the function of the operator is continuously on, it can perform on the arbitrary length of arrays or vectors.

6. LH9320'S FFT ADDRESSING INSTRUCTIONS

The LH9320 provides the address patterns and matches the the I/O throughput required by the LH9124. It is programmable with 32 instruction buffers. The parameter PCSTART specifies the starting address and the parameter PCSTOP specifies the ending address of the executed codes. The instructions between these two addresses can be infinitely executed as a loop if the START pin is continuously triggered. The LH9320 provides over 150 address instructions. More than two thirds of these instructions are used by FFT operations. Each instruction associated with some definable parameters stored in the on-chip RAM or registers generates the desired address pattern at the proper time. The addressing modes required by general DSP algorithms such as digit-reverse, modulo, and linear are all provided.

The number of points for FFT operations has to be a power of two and is no larger than 2^{20}. This means that one LH9320 can handle up to 2^{20}-point FFT. We may define that the 2^M-point FFT has M columns. Each column can be implemented by one radix-2 stage. Hence, there will be M stages for the 2^M-point FFT. Moreover, any two consecutive columns can be implemented by one radix-4 stage or any four consecutive columns can be implemented by one radix-16 stage. The M columns will be denoted from column 0 to column M-1. The column 0 is always the first stage and can be a radix-2, radix-4 or radix-16 stage.

In the FFT operation, each stage not each column is to be specified by one LH9124 radix execution instruction in conjunction with three LH9320 instructions: first for input data addressing, second for twiddle factor addressing, and third for output data addressing. The input and output data have the same address patterns because the in-place algorithm is employed.

The addressing instruction for either data or twiddle factor is a function of two parameters: the radix of execution and the initial column of the stage. Table 6-1 shows the data addressing instruction and Table 6-2 shows the twiddle factor addressing as a function of the two parameters. For a 128-point FFT, there will be 7 columns from column 0 to column 6 because $128 = 2^7$. Assume the FFT is implemented by one radix-2 stage followed by one radix-4 stage followed by one radix-16 stage. The data and twiddle factor addressing instructions for the radix-2 stage are BF20 and TF20. These instructions can be selected from column 0 under radix-2 from Table 6-1 and 6-2. The radix-4 stage begins with column 1. The instructions from column 1 under radix-4 is to be selected from both tables. They will be MXB240 and MXT240. The radix-16 stage begins with column 3. Similarly, the instructions selected will be MXB24160 and MXT24160.

Table 6-1. LH9320 Instruction Selection for FFT Data Pattern

COLUMN	RADIX-2	RADIX-4	RADIX-16
0	BF20	BF40	BF160
1	BF21	MXB240	MXB2160
2	BF22	BF41	MXB4160
3	BF23	MXB241	MXB24160
4	BF24	BF42	BF161
5	BF25	MXB242	MXB2161
6	BF26	BF43	MXB4161
7	BF27	MXB243	MXB24161
8	BF28	BF44	BF162
9	BF29	MXB244	MXB2162
10	BF210	BF45	MXB4162
11	BF211	MXB245	MXB24162
12	BF212	BF46	BF163
13	BF213	MXB246	MXB2163
14	BF214	BF47	MXB4163
15	BF215	MXB247	MXB24163
16	BF216	BF48	BF164
17	BF217	MXB248	
18	BF218	BF49	
19	BF219		

Table 6-2. LH9320 Instruction Selection for FFT Twiddle Factors

COLUMN	RADIX-2	RADIX-4	RADIX-16
0	TF20	TF40	TF160
1	TF21	MXT240	MXT2160
2	TF22	TF41	MXT4160
3	TF23	MXT241	MXT24160
4	TF24	TF42	TF161
5	TF25	MXT242	MXT2161
6	TF26	TF43	MXT4161
7	TF27	MXT243	MXT24161
8	TF28	TF44	TF162
9	TF29	MXT244	MXT2162
10	TF210	TF45	MXT4162
11	TF211	MXT245	MXT24162
12	TF212	TF46	TF163
13	TF213	MXT246	MXT2163
14	TF214	TF47	MXT4163
15	TF215	MXT247	MXT24163
16	TF216	TF48	TF164
17	TF217	MXT248	
18	TF218	TF49	
19	TF219		

6.1. Digit-Reverse Input Data Addressing: RBF0

The FFT algorithm implemented by the LH9124 and LH9320 is decimation-in-time with digit-reverse inputs. The instruction RBF0 is used to generate a digit-reverse sequence from a linear sequence. The instruction consists of two parameters to be specified. The parameter N indicates the length of the FFT and the 20-bit parameter DIGITREV specifies the structure of the FFT.

For $N = 2^M$, the least significant M bits will be mapped by the structure of the FFT and all the other bits are set zeros. These M bits can also be mapped to the M columns of the FFT with column 0 mapped to the (M – 1)-th bit, column 1 mapped to the (M – 2)-th bit, and so on. If the column i is implemented by a radix-2 stage, the associated (M – i – 1)-th bit in DIGITREV will be zero. Otherwise, the (M – i – 1)-th bit will be one. For a 128-point FFT implemented by $2 \times 4 \times 16$, the least significant seven bits in DIGITREV will be specified by the structure. These seven bits will be 0111111 according to the specified rule. The first zero indicates the first stage of the FFT is a radix-2 stage. The following six bits are one. Thus, the stages followed the radix-2 stage could be three radix-4 stages, one radix-4 stage followed by one radix-16 stage, or one radix-16 stage followed by one radix-4 stage. In the LH9124, a radix-16 stage is actually implemented by two cascaded radix-4 stages. Thus, a radix-16 stage is seen as two radix-4 stages in digit-reverse sequence generation. The digit-reverse sequence will be the same for $2 \times 4 \times 16$, $2 \times 4 \times 4 \times 4$, or $2 \times 16 \times 4$, although they have different implementation structures.

In section 2.2.3, we have discussed that if the structure of the FFT have been defined, the digit-reverse sequence can be systematically generated. The radius pattern discussed in that section can be mapped into DIGITREV and vice versa. In the LH9320, if N and DIGITREV are defined, the digit-reverse sequence with length N will be generated.

6.2. FFT Data Addressing

The LH9320 provides instructions to generate data address sequences for the radix-2, radix-4, or radix-16 stage which can begin with an arbitrary column. The only parameter N specifies the number of points of the FFT and the number of addresses to be generated. This value has to be a power of two and is equal to or less than 2^{20}. The in-place FFT algorithm is employed. Thus, the input and output data address sequences will be the same for any stage of the FFT except the first stage. If the data stored in the input acquisition RAM is in linear order, the instruction of the first input stage will be RBF0 instead of BF20, BF40, or BF160.

6.2.1. Radix-2 Data Addressing: BF2i

The instruction in this group generates the data address sequence for the execution of a radix-2 stage. BF2i generates a data address sequence for the radix-2 butterfly stage at the i-th column. The radix-2 data address sequence of the i-th column can be generated by the following algorithm:

$$\text{for } (k = 0;\ k \le 2^i - 1;\ k{+}{+})$$
$$\quad \text{for } (j = 0;\ j \le \frac{N}{2^i} - 1;\ j{+}{+})$$
$$\qquad \{\text{Output} j * 2^i + k\}$$

6.2.2. Radix-4 Data Addressing: BF4i, MXB24i

The instruction in this group generates the data address sequence for the execution of a radix-4 stage. BF4i generates a data address sequence for the radix-4 stage beginning with the (2 * i)-th column. MXB24i generates a data address sequence for the radix-4 stage beginning with the (2 * i + 1)-th column. The radix-4 data address sequence of the (2 * i + m)-th column can be generated by the following algorithm:

$$\text{for } (k = 0;\ k \le 2^m 4^i - 1;\ k{+}{+})$$
$$\quad \text{for } (j = 0;\ j \le \frac{N}{2^m 4^i} - 1;\ j{+}{+})$$
$$\qquad \{\text{Output} j * 2^m 4^i + k\}$$

6.2.3. Radix-16 Data Addressing: BF16i, MXB216i, MXB416i, MXB2416i

The instruction in this group generates the data address sequence for the execution of a radix-16 stage. BF16i is to generate a data address sequence for the radix-16 stage beginning with the (4 * i)-th column. MXB216i generates a data address sequence for the radix-16 stage beginning with the (4 * i + 1)-th column. MXB416i generates a data address sequence for the radix-16 stage beginning with the (4 * i + 2)-th column. MXB2416i generates a data address sequence for the radix-16 stage beginning with the (4 * i + 3)-th column. The radix-16 data address sequence of the (4 * i + m)-th column can be generated by the following algorithm:

$$\text{for } (k = 0;\ k \le 2^m 16^i - 1;\ k{+}{+})$$
$$\text{for } (j = 0;\ j \le \frac{N}{2^m 16^i} - 1;\ j{+}{+})$$
$$\{\text{Output} j * 2^m 16^i + k\}$$

6.3. FFT Twiddle Factor Addressing

The LH9320 provides instructions to generate twiddle factor sequences for all the stages of the FFT. The stage can be a radix-2, radix-4, or radix-16. The instruction in this group with three specified parameters control the pattern of a generated address sequence. The parameter N denotes the number of points of the FFT and the number of addresses to be generated. The parameter MEMSIZE specifies the memory size of twiddle factors to be employed. The size of twiddle factors is equal to N or to N multiplied by a factor d. This factor has to be a power of two. The third parameter is the 4-th bit of the mode register Mode[4]. If zero, the memory stores 360 degrees of twiddle factors. Otherwise, it stores all the twiddle factors in the fourth quadrant and the point at 270 degree.

The algorithms generating twiddle factor sequences discussed later assume that N = MEMSIZE and the twiddle factors stored are 360 degrees. If the MEMSIZE is larger than N by a factor d, the effective address is obtained by the generated address Adr multiplied by this factor. If the twiddle factors stored are 90 degrees, the effective address is calculated from the generated address Adr by the following equation:

$$\text{Effective Address} = \begin{cases} \text{Adr} & \text{if Adr} \le \text{MEMSIZE}/4 \\ \text{MEMSIZE}/2 - \text{Adr} & \text{if MEMSIZE}/4 < \text{Adr} \le \text{MEMSIZE}/2 \\ \text{Adr} - \text{MEMSIZE}/2 & \text{if MEMSIZE}/2 < \text{Adr} \le 3{*}\text{MEMSIZE}/4 \\ \text{MEMSIZE} - \text{Adr} & \text{if } 3 * \text{MEMSIZE}/4 < \text{Adr} < \text{MEMSIZE} \end{cases} \quad [6A]$$

Thus, the memory size can be saved by a factor of four. Figure 6-1 shows the twiddle factors for MEMSIZE = 16, and N = 16. The twiddle factors are equally spaced on the unit circle in the clockwise order starting at position (1,0).

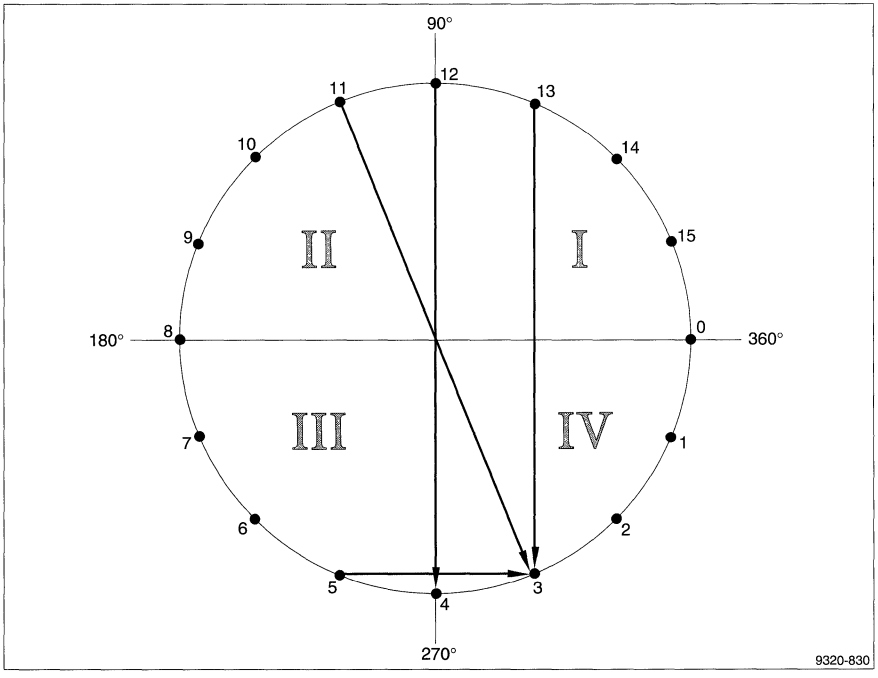


Figure 6-1. Address Mapping of 90 Degree Twiddle Factors

Any point that is not on the fourth quadrant or on the point of 270 degree can be mapped into the fourth quadrant or the point of 270 degree. For example, the 13th point at the first quadrant, the 11th point at the second quadrant, and the 5th point at the third quadrant will be mapped into the 3rd point of the fourth quadrant. Thus, only N/4 + 1 points instead of N points of twiddle factors are required for the N-point FFT.

The sign of real part or imaginary part of a complex coefficient might be changed when loaded to the LH9124 if Mode[4] = 1. There are two input coefficient sign control pins CCOMR and CCOMI. The LH9320 will generate the required control signals for these two pins. Table 6-3 shows how the control signals CCOMR and CCOMI are changed according to the generated address Adr. If CCOMR is one, the real part of the twiddle factor loaded to the LH9124 will be negated. Similarly, CCOMI will control the sign of the imaginary part of the twiddle factor. The twiddle factor addresses and control signals ccomr and ccomi are a function of Mode[4] and MEMSIZE. Table 6-4 shows some examples for a 16-point FFT.

Table 6-3. Relation Between Control Signals and Location of Twiddle Factor

TWIDDLE FACTOR LOCATION	CCOMR	CCOMI
Adr ≤ MEMSIZE/4	0	0
MEMSIZE/4 < Adr ≤ MEMSIZE/2	1	0
MEMSIZE/2 < Adr ≤ 3 * MEMSIZE/4	1	1
3 * MEMSIZE/4 < Adr < MEMSIZE	0	1

6.3.1. Radix-2 Twiddle Factor Addressing: TF2i

The instruction in this group generates the twiddle factor address sequence for the execution of a radix-2 stage. TF2i generates a twiddle factor sequence for the radix-2 stage at the i-th column. The algorithm generating the radix-2 twiddle factor address sequence at the i-th column is listed in the following:

$$Adr = \frac{N}{2^{i+1}}$$

for (k = 0; k ≤ 2^i − 1; k ++)

for (j = 0; j ≤ $\frac{N}{2^{i+1}}$ − 1; j ++)

{output 0

output k * Adr}

Table 6-4. Example of Twiddle Factor Addresses with MEMSIZE and Control Signals

4TH BIT OF MODE	N	MEMSIZE	DATA	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0	16	16	Adr	0	0	0	0	0	1	2	3	0	2	4	6	0	3	6	9
			CCOMR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			CCOMI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	16	32	Adr	0	0	0	0	0	2	4	6	0	4	8	12	0	6	12	18
			CCOMR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			CCOMI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	16	16	Adr	0	0	0	0	0	1	2	3	0	2	4	2	0	3	2	1
			CCOMR	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1
			CCOMI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1	16	32	Adr	0	0	0	0	0	2	4	6	0	4	8	4	0	6	4	2
			CCOMR	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1
			CCOMI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

6.3.2. Radix-4 Twiddle Factor Addressing: TF4i, MXT24i

The instruction in this group generates the twiddle factor address sequence for the execution of a radix-4 stage. TF4i generates a twiddle factor sequence for the radix-4 stage beginning with the (2 * i)-th column. MXT24i generates a twiddle factor sequence for the radix-4 stage beginning with the (2 * i + 1)-th column. The algorithm generating the radix-4 twiddle factor address sequence of the (2 * i + m)-th column is listed in the following:

$$Adr = \frac{N}{2^m 4^{i+1}}$$

for (k = 0; k ≤ $2^m 4^i$ − 1; k ++)
  for (j = 0; j ≤ $\frac{N}{2^m 4^{i+1}}$ − 1; j ++)
    {output 0
     output k * Adr
     output 2k * Adr
     output 3k * Adr}

6.3.3. Radix-16 Twiddle Factor Addressing: TF16i, MXT216i, MXT416i, MXT2416i

The instruction in this group generates the twiddle factor address sequence for the execution of a radix-16 stage. TF16i generates a twiddle factor sequence for the radix-16 stage beginning with the (4 * i)-th column. MXT216i generates a twiddle factor sequence for the radix-16 stage beginning with the (4 * i + 1)-th column. MXT416i generates a twiddle factor sequence for the radix-16 stage beginning with the (4 * i + 2)-th column. MXT2416i generates a twiddle factor sequence for the radix-16 stage beginning with the (4 * i + 3)-th column. The algorithm generating the radix-16 twiddle factor address sequence of the (4 * i + m)-th column is listed in the following:

$$Adr1 = \frac{N}{2^{2+m} 16^i}$$

$$Adr2 = \frac{N}{2^m 16^{i+1}}$$

for (k = 0; k ≤ $2^m 16^i$ − 1; k ++)
  for (j = 0; j ≤ $\frac{N}{2^m 16^{i+1}}$ − 1; j ++)
    {output 0}
    for (l = 1; l ≤ 3; l ++)
      {output l * k * Adr1}
    for (l = 1; l ≤ 3; l ++)
      {output l * k * Adr2}
    for (l = 1; l ≤ 3; l ++)
      {output l * ($2^m 16^i$ + k) * Adr2}
    for (l = 1; l ≤ 3; l ++)
      {output l * (2 * $2^m 16^i$ + k) * Adr2}
    for (l = 1; l ≤ 3; l ++)
      {output l * (3 * $2^m 16^i$ + k) * Adr2}

6.4. Recombining Two Real N-Point FFT Addressing

The instructions in this group generate address sequences to support the instruction BRFT of the LH9124. The frequency sample X(k) after the FFT operation is in linear order. We may assume that the sequence {X(k)} with length N stored in the buffer 2 of Figure 6-2 begins with address 0. Based on the recombination process implemented by Equations [2AD] and [2AE], we will generate two linear sequences {H(k)} and {G(k)} stored in the buffer 3 of Figure 6-2. The length for each sequence is N. {H(k)} begins with address 0 and {G(k)} begins with address N for the full length computation and address N/2 for the half length computation. The information for the other half can be obtained from the symmetric properties of DFT. It can be derived from Equations [2AD] and [2AE] that the input data at addresses i and N – i will generate two output data to be stored at addresses i and N + i for the full length sequence computation and at addresses i and N/2 + i for the half length sequence computation.

6.4.1. Input Data Addressing: BRFTL and BRFTLS

The instructions BRFTL and BRFTLS generate the input address sequences for the recombination process. The difference between the two is that BRFTL generates the full address sequences for {H(k)} and {G(k)} and BRFTLS only generates the first half of the address sequences. The address sequence generated by BRFTL is given by the following algorithm:

Output 0 and 0
for (k = 1; k < N; k ++)
  {Output k and N – k}

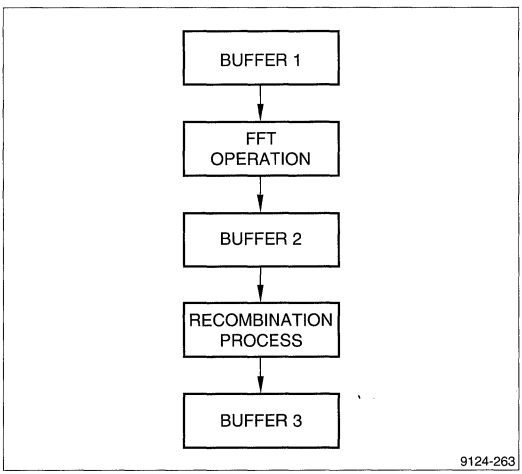


Figure 6-2. Block Diagram of Real FFT Recombination Process

and the sequence generated by BRFTLS is given by:

```
Output 0 and 0
for (k = 1; k < N/2; k ++)
  {Output k and N – k}
```

6.4.2. Output Data Addressing: BRFTU and BRFTUS

The instructions BRFTU and BRFTUS generate the output address sequence for the results of the recombination process. The instructions BRFTU and BRFTUS are used in conjunction with the instructions BRFTL and BRFTLS, respectively. The address sequence generated by BRFTU is given by the following algorithm:

```
for (k = 0; k < N; k ++)
  {Output k and N + k}
```

and the sequence generated by BRFTLS is given by:

```
for (k = 0; k < N/2; k ++)
  {Output k and N/2 + k}
```

6.5. Recombining Real 2N FFT Addressing

The instructions in this group generate address sequences to support the instructions BFCT and BFCT2 of the LH9124. Comparing Equations [2AJ] and [2AK] with Equations [2AD] and [2AE], it can be seen that the input data address sequences for both cases are the same. However, recombination process for 2N real FFT requires one coefficient address sequence to fetch twiddle factors used by the FFT operation. In addition, the results are a sequence with length 2N instead of two sequences with length N apiece. The generated 2N results have the symmetric properties of the DFT. The real data is symmetric and the imaginary data is anti-symmetric to the center point N. The information of the highest frequency sample at address N is lost. Based on Equations [2AJ] and [2AK] and the symmetric property of the DFT, the input data at addresses i and N – i will generate the output data to be stored at addresses i and 2N – i. Thus, the highest frequency sample at address N cannot be generated from the recombination process. From the symmetric property of the DFT, the first N points provide enough information to describe the whole 2N sequence except the point N.

6.5.1. Input Data Addressing: BFCTL

The address sequence generated by the instruction BFCTL is the same as that by the instruction BRFT. The address sequence is generated by the following algorithm:

```
Output 0 and 0
for (k = 1; k < N; k ++)
  {Output k and N – k}
```

6.5.2. Parameter Addressing: BFCTT

The instruction BFCTT generates the address sequence to fetch the parameter, W_{2N}^{k}, required by Equations [2AJ] and [2AK]. The factor d is defined the same as that used for the FFT instructions except the length being 2N instead of N. The bit Mode[4] can be set to save the buffer required to store the twiddle factors. The address sequence is generated by the following algorithm:

```
d = MEMSIZE/(N * 2)
for (k = 0; k < N; k ++)
  {Output k * d and k * d}
```

6.5.3. Output Data Addressing: BFCTU and BFCTUS

The instructions BFCTU and BFCTUS generate the output address sequences for the results of the recombination process. Both BFCTU and BFCTUS are used in conjunction with BFCTL. BFCTU generates the address sequence with 2N length and BFCTUS only generate the first half of the 2N addresses. The address sequence generated by BFCTU is given by the following algorithm:

```
Output 0 and index address
for (k = 1; k < N; k ++)
  {Output k and 2N – k}
```

and the sequence generated by BFCTUS is given by:

```
for (k = 0; k < N; k ++)
  {Output k and k}
```

6.6. DFT Addressing: ADECIM

The DFT operation can be represented by the following matrix form:

$$\begin{bmatrix} X(0) \\ X(1) \\ X(2) \\ \cdot \\ \cdot \\ X(N-1) \end{bmatrix} = \begin{bmatrix} W_N^0 & W_N^0 & \cdots & W_N^0 \\ W_N^0 & W_N^1 & \cdots & W_N^{N-1} \\ W_N^0 & W_N^2 & \cdots & W_N^{2N-2} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ W_N^0 & W_N^{N-1} & \cdots & W_N^{(N-1)^2} \end{bmatrix} \begin{bmatrix} x(0) \\ x(1) \\ x(2) \\ \cdot \\ \cdot \\ x(N-1) \end{bmatrix} \quad [6B]$$

where W_N^k is the twiddle factor of the FFT and is periodic with period N. Thus,

$$W_N^k = W_N^{k+m*N} \quad \text{for } m = \ldots, -2, -1, 0, 1, 2, \ldots \quad [6C]$$

To compute the i-th frequency sample, we have to fetch the twiddle factors from the i-th row of Equation [6B]. The instruction ADECIM providing five parameters is used to generate the address sequence to fetch the twiddle factors for the DFT. The parameter N specifies the length of the time or frequency samples and also indicates the length of addresses in each set of the address sequence. MEMSIZE defines the memory size of the twiddle factors.

The first set of the address sequence can be obtained by the following recursive equation:

$$Adr_{i+1} = Mod(Adr_i + DFACTOR_j)_{MEMSIZE} \quad [6D]$$

The initial value Adr_0 is always at address 0. The parameter DFACTOR denotes the address difference of two consecutive fetched addresses in the twiddle factor memory for the first set of N addresses. Thus, $DFACTOR_0$ = DFACTOR. Equation [6D] also indicates that the addressing of ADECIM is a modulo addressing with modulo length equal to MEMSIZE. The DFACTOR will be updated after one set of N addresses as follows:

$$DFACTOR_{j+1} = DFACTOR_j + LEAP \quad [6E]$$

LEAP denotes the value DFACTOR to be incremented between two consecutive set of addresses. NLEAP indicates how many sets of addresses to be generated. For example, given N = 8 and MEMSIZE = 16, the frequency samples X(1), X(3), X(5), and X(7) are to be computed. We will set DFACTOR = 2, LEAP = 4, and NLEAP = 4. Table 6-5 shows the twiddle factor address sequence generated for the DFT computations.

Table 6-5. Example of an ADECIM Address Sequence

ADDRESS	1	2	3	4	5	6	7	8
Set 1	0	2	4	6	8	10	12	14
Set 2	0	6	12	2	8	14	4	10
Set 3	0	10	4	14	8	2	12	6
Set 4	0	14	12	10	8	6	4	2

7. EXAMPLES OF FFT IMPLEMENTATION

The LH9124 and LH9320 system configuration with four I/O ports shown in Figure 4-1 will be employed for the following FFT application examples. Users have the flexibility in selecting either three or four port configuration for their FFT applications. The Q port is designed as an acquisition port and the C port is used as a coefficient port. We may assume that the input data are stored in linear order in the input acquisition Q port. The coefficients such as twiddle factors are always through the C port. The input and output ports can be any two of the three ports. There are three input pins, DF[2:0], to control the data flow among these four ports.

7.1. Comparison of Four Different 256-Point FFT Structures

FFT applications using the LH9124 and LH9320 are easy and powerful. This section gives one example to discuss how to implement the instructions of LH9124 and LH9320 and how to optimize the performance. The example is to implement a 1024-point FFT with four different structures: (1) $2\times2\times2\times2\times2\times2\times2\times2\times2\times2$, (2) $4\times4\times4\times4\times4$ and (3) $4\times16\times16$.

The general purpose DSP chip implementing the FFT algorithm with non-mixed radix mode usually requires writing a software program with three nested loops. The inner loop is to compute the fundamental module such as the radix-2 or radix-4 butterfly. The performance of the FFT basically depends on the speed of computing large volume of butterflies. The fundamental modules are built in the data path of the LH9124. Therefore, the performance of the LH9124 is tremendous improved by the hardware approach instead of the software programming.

The outer loop of the software program is to control the stage of the FFT operation. The instructions implemented by the LH9124 are in this level. Thus, the number of instruction codes depends on the number of stages of the FFT as shown in Table 7-1 to Table 7-3. The advantage of defining instructions at this level is easy for programming. The structure of the FFT can be directly mapped into and coded by the LH9124's instructions. In addition, both mixed and non-mixed radix mode of the FFT can be efficiently implemented. Furthermore, the time required for computing a stage is dependent on the length of block data and is regardless of the instructions implemented. Hence, users can easily find an optimum structure for their applications. For the general DSP chips, the length of instruction codes will be greatly increased if the FFT is implemented in mixed radix modes.

The function of the middle loop is to control the data transfer. It will generate the addresses for fetching data and twiddle factors executed by the inner loop within a stage and storing the computed data in the proper position for the next stage operation. This function is actually done by the LH9320. There are three instructions associated with the input data block, output data block and twiddle factors as shown in Table 7-1 to Table 7-4. No overhead occurs between the LH9124 and LH9320 because the functions are well-partitioned. Thus, the required cycles for the computation of a stage are equal to the number of points executed without overhead from the data or twiddle factor addressing. If the input data is in digit-reverse order in the acquisition port, the instruction RBF0 has to be replaced by BF20, BF40, or BF160. Table 7-1 and Table 7-3 show that the input data are in linear order and the digit-reverse instruction has to be applied. Table 7-2 shows that the input data are in digit-reverse order. Therefore, the don't care notation "X" is set for the parameter DIGITREV in the Q port of Table 7-5. However, the instruction at the output of stage 1 remains unchanged.

The instruction BFLY2, BFLY4, or BFLY16 will be employed to the first stage of execution with no window function applied to the input data or a rectangular window implicitly applied. When a window function other than a rectangular window is applied, the function can be executed in conjunction with the first stage of radix-2 or radix-4 operation. The instructions BFLY2 and BFLY4 will be replaced by BWND2 and BWND4 in the first windowed radix operation stage. A radix-16 stage cannot be combined with the window execution. Thus, the window function has to be performed by an extra stage.

Table 7-1 and Table 7-3 show that both the input and window data stored in the input and coefficient ports are in linear order. Thus, the digit-reverse address instruction is applied to both ports. In the example of Table 7-2, the input and window data both are stored in digit-reverse order. Therefore, the instructions BF40 and TF40 are performed. The final results are in linear order stored in the Q output acquisition port. For the input data sequence defined by Equation [3B], we may get the output power spectrum results shown in Figures 3-2 to 3-4.

Table 7-4 compares the performance of the four different FFT structures. It can be seen that the performance is inversely proportional to the number of stages of the structure. The higher the performance, the shorter the stages. Thus, the three stage $4\times16\times16$ structure has the best computing efficiency for the 1024-point FFT. We assume that the computing efficiency for the non-mixed mode radix-2 structure is 1. The computing efficiency of other structures will be compared with that of the radix-2 structure.

Table 7-1. Instructions for 1024-Point FFT by $2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2$

STAGE	EXECUTION	INPUT		OUTPUT		COEFFICIENT	
	CODE	PORT	CODE	PORT	CODE	PORT	CODE
1	BWND2	Q	RBF0	A	BF20	C	RBF0
2	BFLY2	A	BF21	B	BF21	C	TF21
3	BFLY2	B	BF22	A	BF22	C	TF22
4	BFLY2	A	BF23	B	BF23	C	TF23
5	BFLY2	B	BF24	A	BF24	C	TF24
6	BFLY2	A	BF25	B	BF25	C	TF25
7	BFLY2	B	BF26	A	BF26	C	TF26
8	BFLY2	A	BF27	B	BF27	C	TF27
9	BFLY2	B	BF28	A	BF28	C	TF28
10	BFLY2	A	BF29	Q	BF29	C	TF29

Table 7-2. Instructions for 1024-Point FFT by $4 \times 4 \times 4 \times 4 \times 4$

STAGE	EXECUTION	INPUT		OUTPUT		COEFFICIENT	
	CODE	PORt	CODE	PORT	CODE	PORT	CODE
1	BWND4	Q	BF40	A	BF40	C	TF40
2	BFLY4	A	BF41	B	BF41	C	TF41
3	BFLY4	B	BF42	A	BF42	C	TF42
4	BFLY4	A	BF43	B	BF43	C	TF43
5	BFLY4	B	BF44	Q	BF44	C	TF44

Table 7-3. Instructions for 1024-Point FFT by $4 \times 16 \times 16$

STAGE	EXECUTION	INPUT		OUTPUT		COEFFICIENT	
	CODE	PORT	CODE	PORT	CODE	PORT	CODE
1	BWND4	Q	RBF0	B	BF40	C	RBF0
2	BFLY16	B	MXB4160	A	MXB4160	C	MXT4160
3	BFLY16	A	MXB4161	B	MXB4161	C	MXT4161

Table 7-4. Performance Comparison for the Structures of 1024-Point FFT

STRUCTURE	$2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2$	$4 \times 4 \times 4 \times 4 \times 4$	$4 \times 16 \times 16$
Cycles	10420	5210	3276
No. of Stages	10	5	3
Efficiency	1	2	3.18

The LH9320 instruction codes can be stored either in the on-chip program memory or in the external memory. The bit Mode[0] controls where it is in the internal or external program memory mode. In the internal mode, two parameters PCSTART and PCEND specify the starting and ending addresses of a program. The parameter N specifies the number of points of the FFT and N = 1024 in the example. The parameter MEMSIZE specifies the size of twiddle factors and is equal to N multiplied by a factor, d, when $d = 2^j$, for j = 0,1, ... The parameter DIGITREV is specified by the structure of the FFT. The digit-reverse sequence is generated based on this parameter and N. A radix-16 stage is considered as two radix-4 stages in generating the digit-reverse sequence. Table 7-5 lists the parameters set in each port of the LH9320 for the above three examples.

The optimum performance is usually obtained by selecting a FFT structure that has a minimum number of stages. The 16-point FFT is the only exception to this rule because the latency time is higher than the computation time. Thus, a two-stage 4 × 4 structure is faster than a one-stage radix-16 structure. Table 7-6 lists the optimum structure of the FFT as a function of the number of points. The structure is arranged in ascending radix order. However, the order of radix operation can be exchanged without changing the performance.

Table 7-5. LH9320 Parameter Setting for Each Port of the LH9124

PARAMETER	N	MEMSIZE	DIGITREV
2 × 2 × 2 × 2 × 2 × 2 × 2 × 2 × 2 × 2			
Q Port	00000000010000000000	00000000010000000000	00000000000000000000
A Port	00000000010000000000	00000000010000000000	X
B Port	00000000010000000000	00000000010000000000	X
C Port	00000000010000000000	00000000010000000000	00000000000000000000
4 × 4 × 4 × 4 × 4			
Q Port	00000000010000000000	00000000010000000000	X
A Port	00000000010000000000	00000000010000000000	X
B Port	00000000010000000000	00000000010000000000	X
C Port	00000000010000000000	00000000010000000000	X
4 × 16 × 16			
Q Port	00000000010000000000	00000000010000000000	00000000001111111111
A Port	00000000010000000000	00000000010000000000	X
B Port	00000000010000000000	00000000010000000000	X

Table 7-6. Optimum N-Point FFT Structure by the LH9124

NO. OF POINTS	STRUCTURE
8	2 × 4
16	4 × 4
32	2 × 16
64	4 × 16
128	2 × 4 × 16
256	16 × 16
512	2 × 16 × 16
1024	4 × 16 × 16
2048	2 × 4 × 16 × 16
NO. OF POINTS	**STRUCTURE**
4096	16 × 16 × 16
8192	2 × 16 × 16 × 16
16382	4 × 16 × 16 × 16
32764	2 × 4 × 16 × 16 × 16
65528	16 × 16 × 16 × 16
131056	2 × 16 × 16 × 16 × 16
262102	4 × 16 × 16 × 16 × 16

7.2. 2-D FFT Implementation

The 2-D FFT implementation is usually obtained by employing 1-D FFT on the 2-D image column by column and row by row. The 2-D FFT implementation discussed in this section will assume the whole 2-D image as a long 1-D array. Then, the 1-D FFT implementation with a little modification can be directly and efficiently employed to the 2-D FFT.

7.2.1. 2-D Digit-Reverse Sequence Generation

For the FFT algorithm implemented by the LH9124 and LH9320, the input sequence either for 1-D or 2-D FFT is in digit-reverse order and the output sequence is in linear order. The straightforward way of obtaining a 2-D digit-reverse array is to do the 1-D digit-reverse for each column and then for each row of the 2-D array. In the following, a more efficient way is presented by considering the whole 2-D array together as a 1-D long sequence and doing the digit-reverse by the instruction RBF0 provided by the LH9320.

The general memory device is a 1-D array. Thus, the 2-D array with dimension (N_1,N_2) has to be mapped into the 1-D storage. Both N_1 and N_2 have to be a power of two. In the column-major order, an element (n_1,n_2) in a 2-D array (N_1,N_2) is mapped into a 1-D device with the address represented by:

$$n = N_2 * n_1 + n_2 \qquad [7A]$$

Table 7-7 shows an example of a 2-D array with dimension (8,8). The element (n_1,n_2) has the associated 1-D address depicted on the left of parenthesis. Assume that the 1-D array is implemented by the structure of 4×2 in each row and by the structure of 2×4 in each column. After the 2-D FFT, the function $X(k_1,k_2)$ has digit-reverse order on both indices k_1 and k_2. Table 7-8 shows the 2-D digit-reverse array with its associated address in the 1-D array depicted on the left of parenthesis. The digit-reverse sequence shown in the table has to be derived first because the input sequence of the FFT computed by the chip is in digit-reverse order.

Assume that the 2-D FFT is obtained by doing 1-D FFT for each row first and then for each column. Thus, we should get a 1-D sequence from Table 7-8 in the row-major order. It can be seen that the digit-reverse sequence can be obtained from the instruction RBF0 by setting N = 8 * 8 and the structure of $4 \times 2 \times 2 \times 4$. For the general array (N_1,N_2), the parameter N is obtained by N_1 multiplied by N_2 and the structure is obtained by the structure

Table 7-7. Memory Address for 8×8 2-D Linear Array

	n_1 →							
n_2 ↓	0(0,0)	8(1,0)	16(2,0)	24(3,0)	32(4,0)	40(5,0)	48(6,0)	56(7,0)
	1(0,1)	9(1,1)	17(2,1)	25(3,1)	33(4,1)	41(5,1)	49(6,1)	57(7,1)
	2(0,2)	10(1,2)	18(2,2)	26(3,2)	34(4,2)	42(5,2)	50(6,2)	58(7,2)
	3(0,3)	11(1,3)	19(2,3)	27(3,3)	35(4,3)	43(5,3)	51(6,3)	59(7,3)
	4(0,4)	12(1,4)	20(2,4)	28(3,4)	36(4,4)	44(5,4)	52(6,4)	60(7,4)
	5(0,5)	13(1,5)	21(2,5)	29(3,5)	37(4,5)	45(5,5)	53(6,5)	61(7,5)
	6(0,6)	14(1,6)	22(2,6)	30(3,6)	38(4,6)	46(5,6)	54(6,6)	62(7,6)
	7(0,7)	15(1,7)	23(2,7)	31(3,7)	39(4,7)	47(5,7)	55(6,7)	63(7,7)

Table 7-8. Memory Address for 8×8 2-D Digit-Reverse Array

	k_1 →							
k_2 ↓	0(0,0)	16(2,0)	32(4,0)	48(6,0)	8(1,0)	24(3,0)	40(5,0)	56(7,0)
	4(0,4)	20(2,4)	36(4,4)	52(6,4)	12(1,4)	28(3,4)	44(5,4)	60(7,4)
	1(0,1)	17(2,1)	33(4,1)	49(6,1)	9(1,1)	25(3,1)	41(5,1)	57(7,1)
	5(0,5)	21(2,5)	37(4,5)	53(6,5)	13(1,5)	29(3,5)	45(5,5)	61(7,5)
	2(0,2)	18(2,2)	34(4,2)	50(6,2)	10(1,2)	26(3,2)	42(5,2)	58(7,2)
	6(0,6)	22(2,6)	38(4,6)	54(6,6)	14(1,6)	30(3,6)	46(5,6)	62(7,6)
	3(0,3)	19(2,3)	35(4,3)	51(6,3)	11(1,3)	27(3,3)	43(5,3)	59(7,3)
	7(0,7)	23(2,7)	39(4,7)	55(6,7)	15(1,7)	31(3,7)	47(5,7)	63(7,7)

for the row array cascaded by the structure for the column array. The parameter DIGITREV is obtained by mapping the cascaded stages to the associated bits.

The 1-D sequence for the 2-D digit-reverse can be derived from the definition of 2-D to 1-D mapping shown in Equation [7B]. The digit-reverse sequence can be obtained by the following equation:

$$dr(n) = N_2 * dr(n_1) + dr(n_2) \quad [7B]$$

Since the 2-D FFT is obtained by doing the 1-D FFT for each row first, the 1-D digit-reverse sequence derived from the 2-D array is obtained by the following algorithm:

for (n_2; $n_2 < N_2$; n_2 ++)
  for (n_1; $n_1 < N_1$; n_1 ++)
    dr(n)

Given a 2-D digit-reverse array shown in Table 7-8, the output will be a 2-D linear array shown in Table 7-7 after 2-D FFT operations. However, the 2-D array mapped into the 1-D array is in the row-major order instead of the column-major order. The 2-D array mapped into the 1-D array can also be defined by the following equation:

$$n = N_1 * n_2 + n_1 \quad [7C]$$

In this case, the above discussion is still valid if row and column are exchanged.

7.2.2. Data and Twiddle Factor Sequence

Corresponding to the 2-D digit-reverse sequence generated in section 7.2.1, the 2-D FFT is done by the 1-D FFT with row-by-row first and then column-by-column. Assume that the stage executed is a radix-2 and its associated address instructions are BF2i and TF2i. If we do 1-D FFT for the same stage of all the rows, the data address sequence to be generated is given as the following algorithm:

for ($l = 0$; $l < N_2$; l ++)
  for ($k = 0$; $k \le 2^i - 1$; k ++)
    for ($j = 0$; $j \le \frac{N_1}{2^i} - 1$; j ++)
      {output $j * 2^i + k + l * N_1$}

and the twiddle factor address sequence to be generated is given as the following algorithm:

$$Adr = \frac{N_1}{2^{i+1}}$$

for ($l = 0$; $l < N_2$; l ++)
  for ($k = 0$; $k \le 2^i - 1$; k ++)
    for ($j = 0$; $j \le \frac{N_1}{2^{i+1}} - 1$; j ++)
      {output 0, k * Adr}

Since the in-place algorithm is used, changing the order of butterfly computations will not change the final results as long as the data and twiddle factors are changed accordingly. It can be seen from the twiddle factor algorithm that the butterflies with the same index k use the same twiddle factors. We may put all the butterflies with the same index k together. Thus, we can delete both the outer loops and change the length of the inner loop from N_1 to N in both algorithms, where $N = N_1 * N_2$. We still get the same results after changing the order of the address sequences. This means that the instructions BF2i and TF2i with $N = N_1 * N_2$ can implement the radix-2 stage for all the rows. The same discussion can also be applied to the radix-4 or radix-16 stage. For the example shown in Table 7-8, each row is implemented by the structure of 4×2. The FFT for all the row-by-row operation can be implemented by two stages with instructions shown in Table 7-9 and N = 64.

Next, we consider doing 1-D FFT for all the columns of the 2-D array. Similarly, assume that the stage executed is a radix-2 and its associated address instructions are BF2i and TF2i. If we do 1-D FFT for the same stage of all the columns, the data address sequence to be generated is given as the following algorithm:

for ($l = 0$; $l < N_1$; l ++)
  for ($k = 0$; $k \le 2^i - 1$; k ++)
    for ($j = 0$; $j \le \frac{N_2}{2^i} - 1$; j ++)
      {output $j * 2^i * N_1 + k * N_1 + l$}

and the twiddle factor address sequence to be generated is given as the following algorithm:

$$Adr = \frac{N_2}{2^{i+1}}$$

for ($l = 0$; $l < N_1$; l ++)
  for ($k = 0$; $k \le 2^i - 1$; k ++)
    for ($j = 0$; $j \le \frac{N_2}{2^{i+1}} - 1$; j ++)
      {output 0, k * Adr}

Similarly, we can put all the butterflies with the same twiddle factors together. Then, the data address sequence will be generated by the following algorithm:

for ($k = 0$; $k \le 2^i * N_1 - 1$; k ++)
  for ($j = 0$; $j \le \frac{N}{2^i * N_1} - 1$; j ++)
    {output $j * 2^i * N_1 + k$}

and the twiddle factor address sequence will generated by the following algorithm:

$$Adr = \frac{N}{2^{i+1}}$$

for ($k = 0$; $k \le 2^i - 1$; k ++)

for ($j = 0$; $j \le \frac{N}{2^{i+1}} - 1$; j ++)

{output 0, k * Adr}

Thus, the twiddle factor address instruction is the same as that for the 1-D column array but the length of the FFT is changed to N. The data address instruction will be changed from BF2i to BF2j, where $j = i + c_1$ and $N_1 = 2^{c_1}$. The same discussion can also be applied to the radix-4 or radix-16 stage. For the example shown in Table 7-8, each column is implemented by the structure of 2×4. The FFT for all the column-by-column operation can be implemented by two stages with instructions shown in the last two rows of Table 7-9 and N = 64.

Table 7-9 shows the LH9124 and LH9320 instructions to implement (8,8) 2-D FFT with its row structure by 4×2 and column structure by 2×4. Table 7-10 shows the LH9124 and LH9320 instructions to implement 64-point 1-D FFT with the structure by $4 \times 2 \times 2 \times 4$. Both cases will use the same time to complete the operation. It can be seen that both cases employ the same data address instructions. However, the instructions for the twiddle factors are different. The third stage of the 2-D FFT is actually the first stage of the 1-D FFT for the column array. Thus, the twiddle factor instruction in the 0-th column of Table 6-1 has to be used. Similarly, the fourth stage of the 2-D FFT is the second stage of the 1-D FFT for the column array.

There are some advantages of computing the 2-D FFT by the proposed method. The overhead occurred from one row to the other row or from one column to the other column can be eliminated. Moreover, the build-in block floating point mechanism can fully implement this approach as in the 1-D case without glue logic required. Furthermore, the LH9320 can provide all the required address pattern up to 2^{20} points, i.e., $n_1 * n_2 \le 2^{20}$.

7.2.3. Example of a 32 by 32 2-D FFT Implementation

This subsection will discuss the implementation of a windowed 2-D FFT with array size 32 by 32. The real part of the time sample is defined as:

$$x_r(n_1,n_2) = \sin(2\pi * 0.1255 * n_1)\sin(2\pi * 0.1255 * n_2) * w(n_1) * w(n_2) - 2^{-16}\sin(2\pi * 0.3755 * n_1) \sin(2\pi * 0.3755 * n_2) * w(n_1) * w(n_2) \quad [7D]$$

and the imaginary part of the time sample is defined as:

$$x_i(n_1,n_2) = 2^{-8}\sin(2\pi * 0.1255 * n_1)\sin(2\pi * 0.3755 * n_2) * w(n_1) * w(n_2) + 2^{-8}\sin(2\pi * 0.3755 * n_1) \sin(2\pi * 0.1255 * n_2) * w(n_1) * w(n_2) \quad [7E]$$

where n_1 and n_2 are from 0 to 31 and the window function $w(n_i)$ is a Black-Harris window defined as follows:

$$w(n_i) = 0.35875 - 0.48829 * \cos\left(\frac{2\pi}{32}n_i\right) + 0.14128 * \cos\left(\frac{2\pi}{32}2n_i\right) - 0.01168 * \cos\left(\frac{2\pi}{32}3n_i\right) \quad [7F]$$

for $n_i = 0, 1, 2, \ldots, 31$

Table 7-9. Instructions for 2-D (8,8) FFT With ($4 \times 2, 2 \times 4$) Structure

STAGE	EXECUTION	INPUT		OUTPUT		COEFFICIENT	
	CODE	PORT	CODE	PORT	CODE	PORT	CODE
1	BFLY4	Q	RBF0	A	BF40	C	TF40
2	BFLY2	A	BF22	B	BF22	C	TF22
3	BFLY2	B	BF23	A	BF23	C	TF20
4	BFLY2	A	BF42	Q	BF42	C	MXT240

Table 7-10. Instructions for 1-D 64-Point FFT With $4 \times 2 \times 2 \times 4$ Structure

STAGE	EXECUTION	INPUT		OUTPUT		COEFFICIENT	
	CODE	PORT	CODE	PORT	CODE	PORT	CODE
1	BFLY4	Q	RBF0	A	BF40	C	TF40
2	BFLY2	A	BF22	B	BF22	C	TF22
3	BFLY2	B	BF23	A	BF23	C	TF23

Assume that the row and column both are implemented by a 3-stage FFT with the $4 \times 4 \times 2$ structure. Thus, the 2-D FFT implementation will have six stages in total. Table 7-11 lists the LH9124 and LH9320 instructions implemented in those stages. As in the 1-D windowed FFT, the 2-D window function can be combined with the first stage of the 2-D FFT operation. Thus, the instruction BWND4 is performed instead of BFLY4. The input data and the window coefficients are assumed to be stored in the Q and C ports in linear order. Hence, the digit-reverse instruction RBF0 is used to generate the required address sequence in the first stage. The associated parameters of the LH9320 defined in each port are listed in Table 7-12. The power spectrum of the computed results is shown in Figure 7-1. The frequency pattern of the spectrum can be clearly seen, even though the resolution of the image is low and the difference of the signal levels is high.

7.3. 3-D FFT Implementation

The concept of the 2-D FFT implementation can be extended to the 3-D FFT implementation because they have the same form in mathematical representation. Assume that an element (n_1,n_2,n_3) in a 3-D array (N_1,N_2,N_3) is mapped into a 1-D device with the address represented by

$$n = N_2 * N_3 * n_1 + N_3 * n_2 + n_3 \quad [7D]$$

If $x(n_1,n_2,n_3)$ is a time sample, the frequency sample $X(k_1,k_2,k_3)$ is obtained from the 3-D FFT of the time sample. The sequence k_i is in a digit-reverse pattern of the sequence n_i. The input array is required to be in digit-reverse order to be implemented by the LH9124 and LH9320. As in the 2-D case, the required digit-reverse sequence can be obtained by employing the instruction RBF0. The parameter DIGITREV is mapped by the structure that cascades the structures of the 1-D FFT in three tuples. For a 3-D array (8,16,8), assume that the array is implemented by the structure of 4×2 for the first tuple, 16 for the second tuple, and 2×4 for the third tuple. Thus, the parameter DIGITREV is mapped from the structure $4 \times 2 \times 16 \times 2 \times 4$. The parameter N is equal to $N_1 * N_2 * N_3$.

Table 7-13 shows the instructions to compute the 3-D FFT for a 3-D array (8,16,8). There are five stages to implement the 3-D array. Table 7-14 shows the instructions to implement 1-D 1024-point FFT with the same structure used by the 3-D FFT. As in the 2-D case, the data address instructions used by the two cases are the same and the twiddle factor address instructions are different. The second and the fourth stages of the 3-D FFT are actually the first stage of the 1-D 8-point FFT. The third and fourth stages of the 3-D FFT are the first stage of the 1-D FFT. Thus, the twiddle factor address instruction in the 0-th column of Table 6-1 has to be used. Similarly, the fifth stage of the 3-D FFT is the second stage of the 1-D FFT in the third tuple.

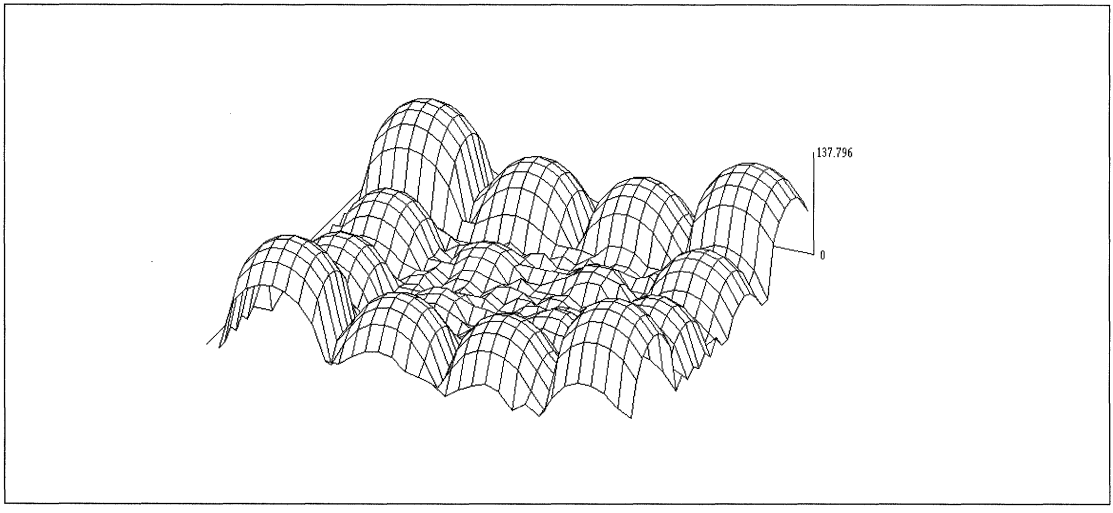


Figure 7-1. 2-D FFT of 1K Point (32×32) With Four-Term Black-Harris Window

Table 7-11. Instructions for 32 by 32 2-D FFT With 4 × 4 × 2 For Row and Column

STAGE	EXECUTION	INPUT		OUTPUT		COEFFICIENT	
	CODE	PORT	CODE	PORT	CODE	PORT	CODE
1	BWND4	Q	RBF0	A	BF40	C	RBF0
2	BFLY4	A	BF41	B	BF41	C	TF41
3	BFLY2	B	BF24	A	BF24	C	TF24
4	BFLY4	A	MXB242	B	MXB242	C	TF40
5	BFLY4	B	MXB243	A	MXB243	C	TF41
6	BFLY2	A	BF29	Q	BF29	C	TF24

Table 7-12. LH9320 Parameter Setting for Each Port of the LH9124

PARAMETER	N	MEMSIZE	DIGITREV
Q Port	00000000001000000000	00000000001000000000	00000000001111011110
A Port	00000000001000000000	00000000001000000000	X
B Port	00000000001000000000	00000000001000000000	X
C Port	00000000001000000000	00000000001000000000	00000000001111011110

Table 7-13. Instructions for 3-D (8,16,8) FFT With 4 × 2 × 16 × 2 × 4 Structure

STAGE	EXECUTION	INPUT		OUTPUT		COEFFICIENT	
	CODE	PORT	CODE	PORT	CODE	PORT	CODE
1	BFLY4	Q	RBF0	A	BF40	C	TF40
2	BFLY2	A	BF22	B	BF22	C	TF22
3	BFLY16	B	MXB24160	A	MXB24160	C	TF160
4	BFLY2	A	BF27	B	BF27	C	TF20
5	BFLY4	B	BF44	A	BF44	C	MXT240

Table 7-14. Instructions for 1-D 1024-Point FFT With 4 × 2 × 16 × 2 × 4 Structure

STAGE	EXECUTION	INPUT		OUTPUT		COEFFICIENT	
	CODE	PORT	CODE	PORT	CODE	PORT	CODE
1	BFLY4	Q	RBF0	A	BF40	C	TF40
2	BFLY2	A	BF22	B	BF22	C	TF22
3	BFLY16	B	MXB24160	A	MXB24160	C	MXT24160
4	BFLY2	A	BF27	B	BF27	C	TF27
5	BFLY4	B	BF44	A	BF44	C	TF44

7.4. Two Real 256-Point FFT

This method discussed in section 2.5.1 is used to implement this example. A 256-point complex FFT is used to compute two 256-point real FFTs. The two 256-point real sequences are arranged to form a 256-point complex sequence with one sequence placed in the real part and the other sequence placed in the imaginary part of the input memory. The 256-point complex sequence is stored in the Q input acquisition port in linear order.

The two real FFTs are obtained by two phase of computations. The first phase performs a 256-point complex FFT that is executed by two cascaded radix-16 stages as shown in Table 7-15. The results are stored in the B port in linear order. The second phase performs the recombining process represented by Equations [2AD] and [2AE]. The instructions executed for this phase are shown in the final stage of Table 7-15.

The final results are stored in the Q output acquisition port. The frequency samples in conjunction with the real part of the input acquisition port are stored in the first 256 memory addresses and the frequency samples in conjunction with the imaginary part are stored in the next 256 addresses of the output acquisition port in linear order. The LH9124 and LH9320 instructions can be straightforward applied to implement this example. The first half of the frequency sample and the sample at the highest frequency is enough to provide all the information in the frequency domain. If only these frequency samples are computed, the required instruction cycles will be 924 cycles.

7.5. A Real 512-Point FFT

This example is implemented by the method discussed in section 2.5.2. The 256-point complex FFT will be used to compute the 512-point real FFT. The 512-point real sequence {x(n)} is first transformed to the 256-point complex sequence {c(n)} with $c_r(n) = x(2n)$} and $c_i(n) = x(2n + 1)$. The 256-point complex sequence will be stored in the Q input acquisition port.

The real FFT is obtained by two phases of computations. The first phase performs a 256-point complex FFT that is executed by two cascaded radix-16 stages as shown in Table 7-16. The results are stored in the B port in linear order. The second phase performs the recombining process represented by Equations [2AJ] and [2AK]. The instructions executed for this phase are shown in the final stage of Table 7-16.

The final results are stored in the Q output acquisition port in linear order. The LH9124 and LH9320 instructions can be straightforward applied to execute this example. The highest frequency samples at the location 256 is not calculated in this example because the limitation of the recombining algorithm. Users may assign a proper value or zero to the location. 1178 cycles are required to complete the operation as shown in Table 2-8.

7.6. Two 256-Point Discrete Cosine Transforms

The two 256-point discrete Cosine transforms are implemented by the method discussed in section 2.6.3. The two 256-point real sequences are rearranged to form a 256-point complex sequence {x(n)} with one sequence placed in the real part and the other sequence placed in the imaginary part. The complex data sequence {x(n)} is first reordered by Equation [2AU] to generate a new complex sequence {y(n)}. This sequence will placed in the Q input acquisition port. Table 7-17 lists the instructions to execute the two real 256-point DCTs.

The procedure of computations is shown in Figure 2-12 and is executed with four stages by the LH9124 and LH9320. The first two stages perform the 256-point complex FFT. The frequency samples are stored in C port in linear order. Then, the two real FFT recombining instruction is employed to generate two complex frequency sequences for the two real time sequences. Both instructions BRFTL and BRFTU will generate 512-point addresses. Only the first 258-point addresses will be used. The break point register BREAKPOINT and the mode bit Mode[1] can control the number of addresses generated. The first 129 points of the frequency samples are generated for each sequence. The first complex sequence stored in the RAM begins with address 0. The second complex sequence stored in the RAM begins with address 256.

The final stage performs complex multiplication of two complex sequences. One complex sequence is the 129-point coefficient function $c(k)e = {}^{-j\pi k/2N}$ and will be multiplied by each of the two sequences generated from the recombining process. The final results are stored in the Q output acquisition ports. The first half of the first frequency sequence are stored in the real part of the RAM from address 0 to address 127. The second half of the first sequence are stored in the imaginary part of the RAM from address 128 to address 1. The first half of the second frequency sequence are stored in the real part of the RAM from address 256 to address 383. The second half of the second sequence are stored in the imaginary part of the RAM from address 384 to address 256. The instruction BFCTUS will generate address sequence in the form { 0, 0, 1, 1, 2, 2, ... }. The instruction BF28 will generate address sequence in the form { 0, 256, 1, 257, 2, 258, ... } and the parameter N = 512. The parameter BREAKPOINT is set to 258.

Table 7-15. Instructions for Two Real 256-Point FFTs by 16 × 16

STAGE	EXECUTION	INPUT		OUTPUT		COEFFICIENT	
	CODE	PORT	CODE	PORT	CODE	PORT	CODE
0	BFLY16	Q	RBF0	A	BF160	C	TF160
1	BFLY16	A	BF161	B	BF161	C	TF161
2	BRFT	B	BRFTL	Q	BRFTU		

Table 7-16. Instructions for Real 512-Point FFT by 16 × 16

STAGE	EXECUTION	INPUT		OUTPUT		COEFFICIENT	
	CODE	PORT	CODE	PORT	CODE	PORT	CODE
1	BFLY16	Q	RBF0	A	BF160	C	TF160
2	BFLY16	A	BF161	B	BF161	C	TF161
3	BFCT2	B	BFCTL	Q	BFCTU	C	BFCTT

Table 7-17. Instructions for Two 256-Point DCTs

STAGE	EXECUTION	INPUT		OUTPUT		COEFFICIENT	
	CODE	PORT	CODE	PORT	CODE	PORT	CODE
1	BFLY16	Q	RBF0	A	BF160	C	TF160
2	BFLY16	A	BF161	B	BF161	C	TF161
3	BRFT	B	BRFTL	A	BRFTU		
3	CMUL	A	BF28	Q	BF28	C	BFCTUS

8. REFERENCES

[1] A. Oppenheim and R. Schafer, *Discrete-Time Signal Processing*, Prentice-Hall, 1989.

[2] L. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*, Prentice-Hall, 1975.

[3] R. N. Bracewell, *The Fast Fourier Transform and Its Applications*, McGraw-Hill Company, 1986.

[4] E. O. Brigham, *Fast Fourier Transform and Its Applications*, Prentice-Hall, 1988.

[5] E. O. Brigham, *The Fast Fourier Transform*, Prentice-Hall, 1974.

[6] C. S. Burrus and T. W. Pakss, *DFT/FFT and Convolution Algorithms*, John Wiley and Sons, New York, 1984.

[7] H. J. Nussbaumer, *Fast Fourier Transform and Convolution Algorithms*, Spring-Verlag, New York, 1982.

[8] D. F. Elliott, ed., *Handbook of Digital Signal Processing Engineering Applications*, Academic Press Inc. 1987.

[9] F. J. Harris, *On the Use of Windows for Harmonic Analysis With the Discrete Fourier Transform*, Proceedings of the IEEE, Vol. 66, No. 1, pp 51-83, Jan. 1978.

[10] G. R. L. Sohie, *Implementation of Fast Fourier Transforms on Motorola's DSP 56000/DSP56001 and DSP6002 Digital Signal Processor.*

[11] Mathsoft, *MathCAD®* (Version 3.0), Mathsoft, Inc.

MEDICAL TOMOGRAPHY IMAGING USING THE LH9124/LH9320

INTRODUCTION

Tomographic imaging algorithms used in medical imaging applications are highly computation extensive. Even with current technology a significant amount of hardware is required. In the first part of this application note a brief description of the tomography algorithm is given. In the second part an efficient and simple implementation using the LH9124/LH9320 chip set will be presented. In particular, we will present the two most commonly used algorithms: the Filter-back propagation algorithm and the two-dimensional inverse Fourier transform algorithm.

Tomographic Imaging Algorithms

Tomography Imaging

Tomography refers to cross-sectional imaging of an object based on its transmission data. Tomography algorithms reconstruct the image of an object from its transmission data when illuminating from many directions. The major applications of these algorithms are Cat scan (X-ray) and Nuclear Magnetic Resonance (NMR). Figure 1 illustrates a typical measuring system, where f(x,y) is the two-dimensional-object parameter to be imaged.

The measurement made by each detector is a line integral of f(x,y) along the ray. The set of data received from the receiver array (a set of line integral) is called projection or view. The mathematical relation between the function f(x,y) and its projection data is given by:

$$P(t_1) = \int_{Ray} f(x,y)$$

where the line equation is:

$$t_1 = x\cos(\theta) + y\sin(\theta).$$

The function $P_\theta(t)$ for a given θ is a parallel projection of f(x,y) in angle θ. The two dimensional function $P_\theta(t)$ is called the Radon transform of f(x,y).

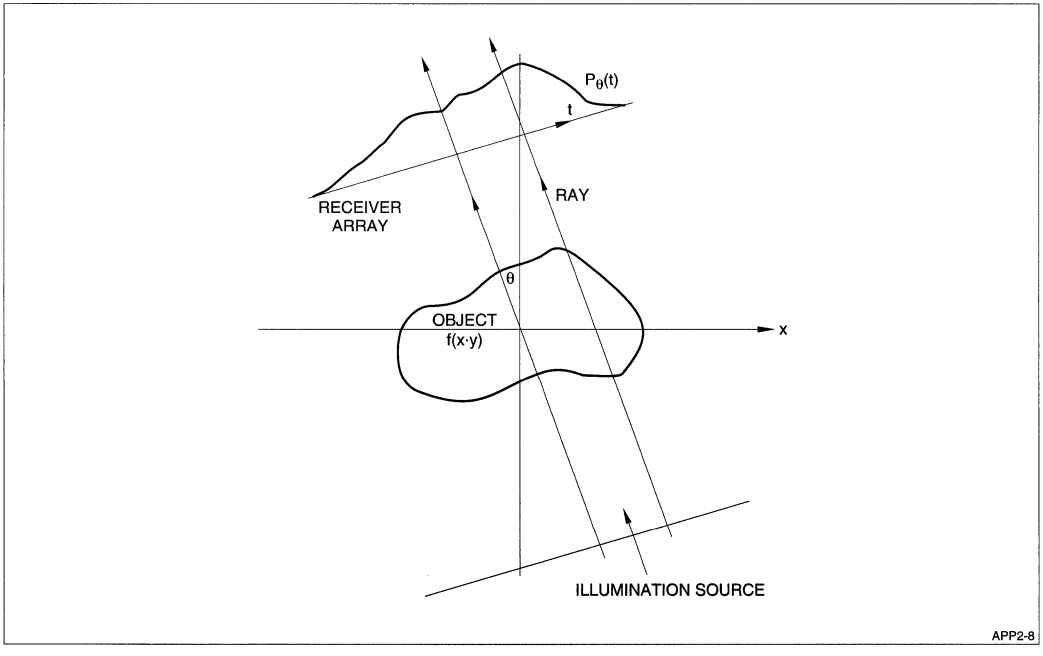


Figure 1. Typical Tomography Measuring System

The Fourier Slice Theorem

The Fourier Slice Theorem relates the one-dimensional Fourier transform of a projection of a two-dimensional function f(x,y) to its two-dimensional Fourier transform F(ω1,ω2) as follows:

The Fourier transform of a parallel projection of an image f(x,y) taken at angle of θ, gives a slice of the two dimensional transform F(ω1,ω2) along a line with angle θ to the axes ω1. Referring to Figure 2, the theorem claims that the Fourier transform $S_\theta(\Omega)$ of $P_\theta(t)$ gives the value of F(ω1,ω2) along line BB'.

The theory implies that if the object is illuminated from many directions then the two-dimensional frequency domain F(ω1,ω2) can be filled. The reconstruction is given by the two-dimensional Inverse Fourier Transform.

NOTE: In some NMR applications the collected data is the Fourier-domain data. In this case, the reconstructing algorithm is the two-dimensional Fourier transform.[2]

The Filter Back Propagation Algorithm

The filter back propagation algorithm is an image reconstructing algorithm based on the Fourier slice theorem. First the algorithm for parallel projection is presented, and in the second part the modification needed for FAN projecting (point illumination source) is described.

Let F(ω1,ω2) be the two dimensional Fourier transform of f(x,y) and let (Ω,θ) be the polar coordinate of the plane (ω1,ω2). The Inverse two dimensional Fourier transform can be written as:

$$f(x,y) = \int_0^{\pi} \int_{-\infty}^{\infty} F(\omega 1,\omega 2)\, \exp\{j(x\omega 1 + y\omega 2)\}\Omega d\Omega d\theta \qquad [1]$$

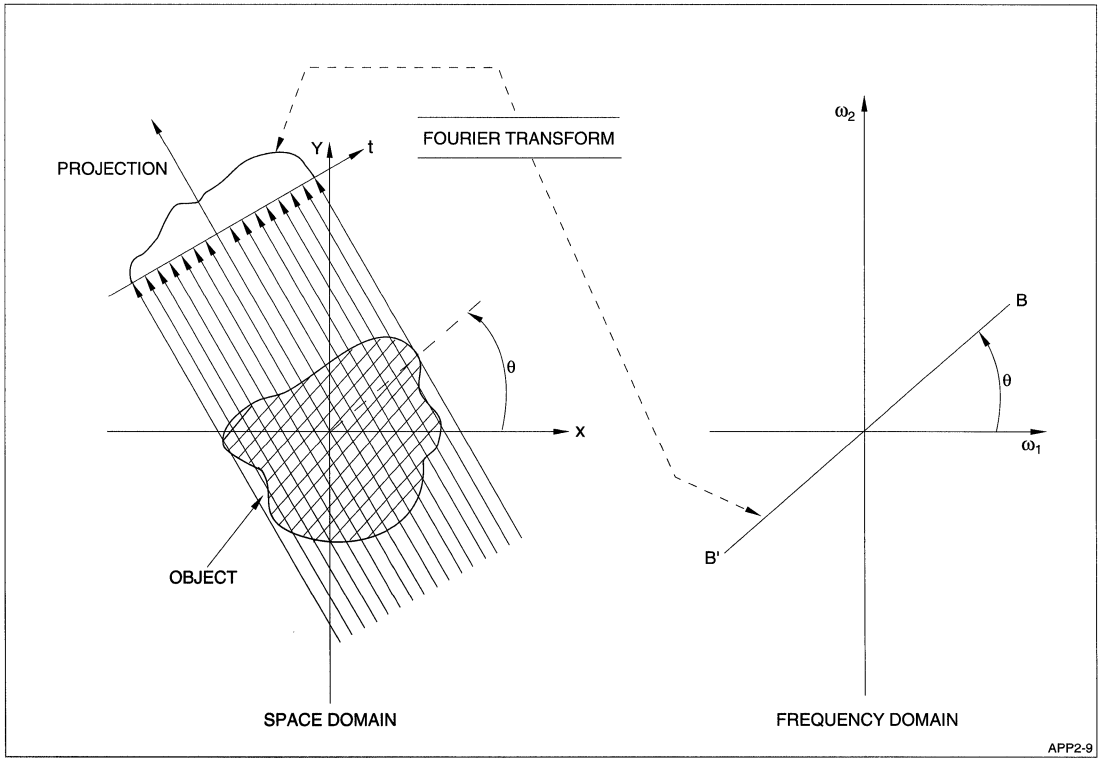


Figure 2. The Fourier Slice Theorem

Using the Fourier slice theorem (and some algebraic manipulation [1]) gives:

$$f(x,y) = \int_0^{\pi}\int_{-\infty}^{\infty} S_\theta(\omega)|\omega|\exp(j\omega t)d\omega d\theta \quad [2]$$

where:

$t = x\cos(\theta) + y\sin(\theta)$ is the ray equation

$S_\theta(\omega)$ is the one-dimensional Fourier transform of $P_\theta(t)$

Equation [2] can be rewritten as:

$$Q_\theta(t) = \int_{-\infty}^{\infty} S_\theta(\omega)|\omega|\exp(j\omega t)d(\omega) \quad [3.1]$$

$$= P_\theta(t) * F^{-1}\{|w|\}$$

where * is the mark for convolution

$$f(x,y) = \int_0^{\pi} Q\underline{\theta}(x\cos(\theta) + y\sin(\theta))d\theta \quad [3.2]$$

Using discrete notation we received:

$$Q_{\theta_i} = P_{\theta_i}(n\tau) * h(n\tau) \quad [4.1]$$

$$f(x_k,y_k) = \Sigma Q_{\theta_i}(x_k\cos(\theta_i) + y_k\cos(\theta_i)) \quad [4.2]$$

where $h(n\tau)$ is the reconstructing filter which is basically IFFT{$|\omega|$}.

The equation set [4] defines the filter back propagation algorithm:

First stage – All the projections $P_\theta(n\tau)$ are filtered with h(nt) to received $Q_\theta(n\tau)$.

Second stage – Back propagating all the filtered projections. This means that to each point (x_k,y_k) of the image the sum of the projection's contribution is calculated.

It should be noted that the point

$$t = x_k\cos(\theta_i) + y_k\cos(\theta_i)$$

is not necessarily an integer point of $n\tau$, and interpolation or up-sampling of the function $Q_\theta(n\tau)$ is needed.

The extensive part of the algorithm is the filtering operation. The filtering could be done directly in the time domain, but it is more efficient to implement in the frequency-domain (Fast Convolution), which is based on the Fourier transform. See equation [5].

$$Q_{\theta_i}(n) = P_{\theta_i}(n) * h(n) = \text{IFFT}\{\text{FFT}\{h(n)\text{FFT}\{P_\theta(n)\}\} \quad [5]$$

Another advantage of the frequency-domain implementation is that interpolating of the output sequence Q(n) could be done easily by zero padding the frequency-domain data prior to performing the IFFT When implementing frequency-domain filtering the input time-domain sequences should be zero-padded to length of at least of M + N to prevent circular convoluting where N and M are the original sequences length [3].

The Filter Back Propagation for FAN Projection

In practical, the illumination source of a Cat Scan system is not a plane-wave but a single-point source, which creates a fan-like source of illumination waves (Figure 3).

Derivation of the algorithm is beyond the scope of this paper. The derivation can be found in reference [1]. The back-propagation algorithm for FAN projection can be summarized by equation set [6].

$$Q_\theta(na) = \{w(na)R_\theta(na)\} * \{h(na)/2\}$$

$$f(xk,yj) = \Sigma Q_{\theta_i}(s)/U^2(x_k,y_k,\theta_i) \quad [6.1]$$

where $R_\theta(na)$ is the collected data.

$$w(na) = \frac{D}{\sqrt{D^2 + n^2a^2}}$$

a – the spacing between the receiver senores arrays

s – identifies the ray pass throw (x,y)

$U(xk,yj,\theta_i)$ is the ratio SP/D (related to Figure 3).

The equation set [6] implies that each view $R_\theta(na)$ has to multiply by a window w(na) and then filter by h(na). For each point (x_k,y_j) a weighted sum of the contribution of all the views is performed.

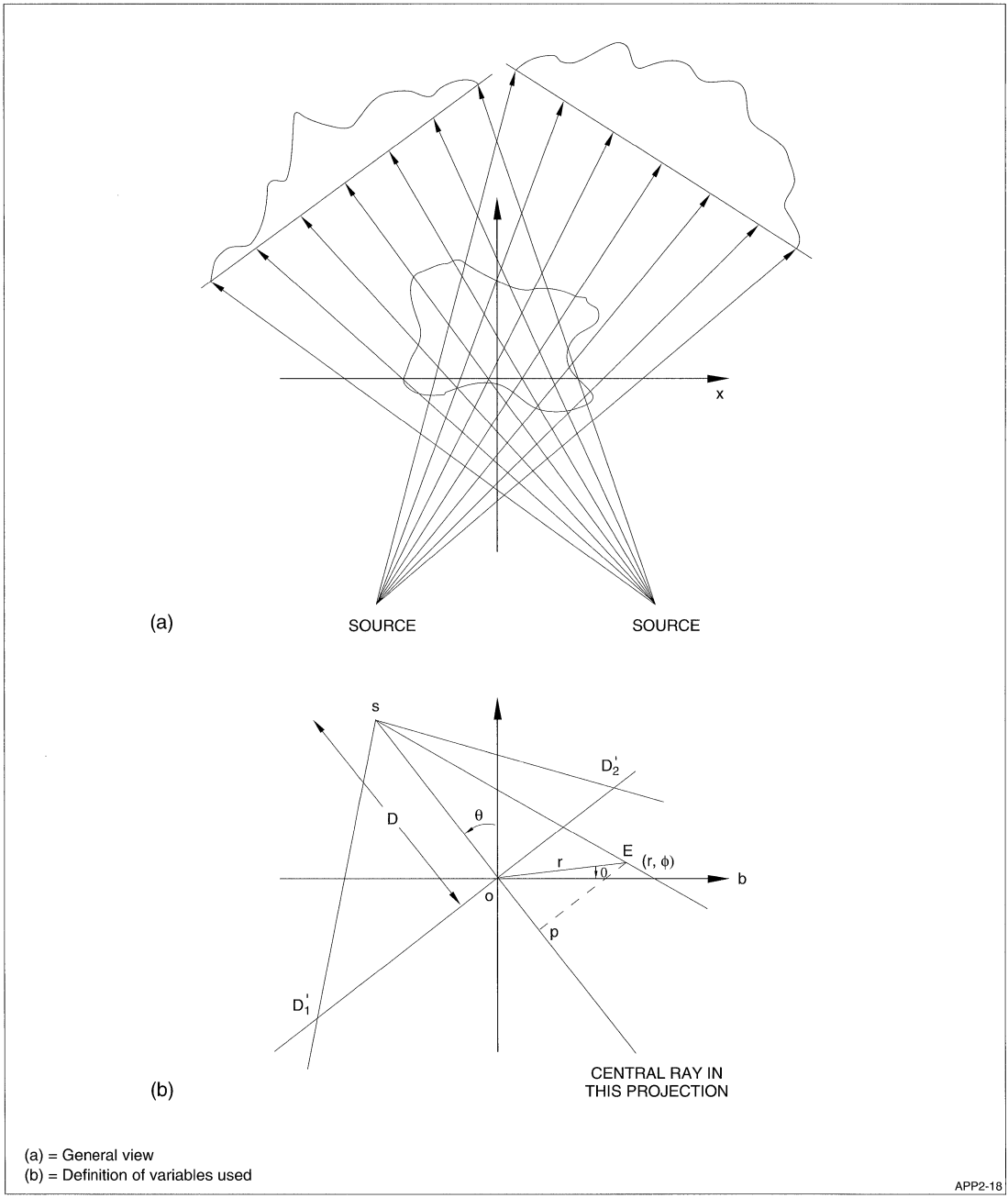


Figure 3. Point Source Illumination

Using the LH9124

The 2-D-IFFT and the Filter Back Propagation algorithms can be implemented efficiently and simply by using the LH9124/LH9320 chip set. A basic system configuration is illustrated in Figure 4. This system configuration can be referenced in the *LH9124 Data Sheet* [6].

Back Propagation Algorithm Using the LH9124

The extensive computation part of the algorithm is the filtering operation. The back propagation part involves mainly summation. In this application note only the filtering implementation will be considered. In the description and the computation load analysis the following system parameters are assumed:

N = 512 – Number of measured points in a view

M = 900 – Number of views (relates to 0.2 deg. resolution)

The algorithm will implement the frequency-domain filtering:

$$Q(n) = IFFT\{FFT\{P(n)w(n)\}H(k)\}$$

where:

P(n) is the measured data view

w(n) space-domain window

H(k) frequency-domain filter response

When implementing this algorithm using the LH9124, it should be noticed that it has a complex input and that two views can be processed simultaneously by loading one view to the real memory port and the other to the imaginary memory port. The real output of the algorithm will be the filtered first view and the imaginary part will be the filtered second view.

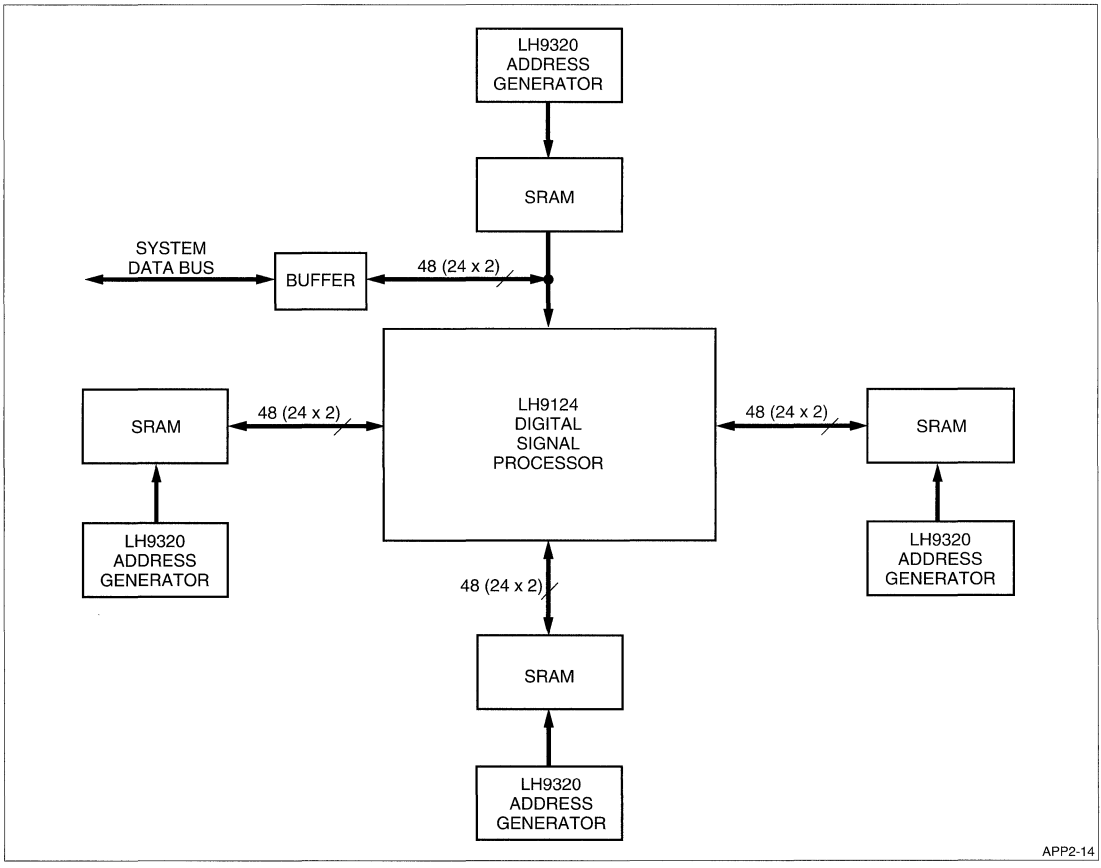


Figure 4. Basic System Configuration

The proof of the above statement is as follows:

let x(n) and y(n) be two views, connecting x to the real port and y to the imaginary port creates a complex input sequence:

$$z(n) = x(n) + jy(n)$$

The complex output sequence is:

$$zf(n) = IFFT\{H(k)FFT\{z(n)w(n)\}\}$$

$$zf(n) = IFFT\{H(k)FFT\{x(n)w(n)\} + jH(k)FFT\{y(n)w(n)\}\}$$

$$zf(n) = IFFT\{H(k)FFT\{x(n)w(n)\}\} + jIFFT\{H(k)FFT\{y(n)w(n)\}\}$$

$$zf(n) = xf(n) + jyf(n)$$

Due to the FFT properties [3] the sequences xf(n) and yf(n) are the convolution result of x(n) and y(n) with the filter h(n). Since the sequences x(n), y(n) and h(n) are all real sequences xf(n) and yf(n) are also real sequences, and the real output of zf(n) is xf(n) and the imaginary output is yf(n).

Programming the Algorithm

Two cases are presented in this paragraph. Case A needs no interpolation of the output. Case B needs an interpolation by a factor of 4.

Case A: Filtering With No Upsampling

The algorithm includes the following steps:

- zero-pad the time (space) domain data
- multiply by a shading window
- perform 1024 FFT
- multiply by the filter H(k) frequency response
- perform 1024 IFFT

The algorithm block diagram is shown in Figure 5. The system program is given in Tables 1 and 2.

Table 2. Address Generator Register Settings

REGISTER	VALUE	PORT
N	1024	A,B,C & Q
DIGITREV	0000000001111111111	A
ZEROPAD	512	Q
ADRLENGTH	1024	Q
ADRINC	1	Q,A
INDEXINC	zero-address	Q
ADRSTART	0	Q,A

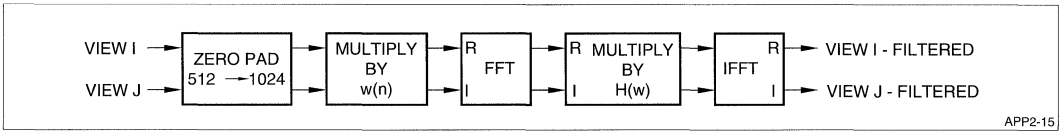


Figure 5. Frequency Domain Two-at-a-Time Filtering

Table 1. Frequency-Domain Filtering With No Up-Sampling

PASS	LH9124		PORT Q	PORT A	PORT B	PORT C	NUMBER OF CYCLES
	CODE	DATA FLOW					
1	MOVD	Q→A	PADHIGH	INC	–	–	1024 + 18
2	BWND4	A→B	–	RBF0	BF40	BRF0	1024 + 20
3	BFLY16	B→A	–	MXB4160	MXB4160	MXT4160	1024 + 68
4	BFLY16	A→B	–	MXB4161	MXB4161	MXT4161	1024 + 68
5	BWND4	B→A	–	RBF0	BF40	RBF0	1024 + 20
6	BFLY16	A→B	–	MXB4160	MXB4160	MXT4160	1024 + 68
7	BFLY16	B→Q	MXB4161	–	MXB4161	MXT4161	1024 + 68
					Total Number of Cycles		7498

Explanation

Pass Number 1

Loading the complex data from port Q to port A, zero padding the original input sequence from 512 point to 1024 point. The zero-padding is accomplished by setting a zero-padding sequence at the input (port Q) and a linear-sequence addressing (INC) at the output (port A).

Passes 2, 3 & 4

Performing 1024 FFT using mixed radix 4 × 16 × 16. The first FFT pass (pass 2) is radix-4 and incorporate multiply by the shading window stored in the coefficient memory.

Passes 5, 6 & 7

Performing IFFT operation using mixed radix 4 × 16 × 16 where the first IFFT pass (pass 5) includes multiply by the reconstruction filter frequency response. It should be noted that the total number of sequences is seven and the whole program can be stored in the Address Generator's program memory. When doing the IFFT it is possible to use the twiddle factor coefficient used for the FFT by setting the LH9124's **ci** and **co** control flags as follows:

ci = 1 during pass 5 (equals zero during all other cases)

co = 1 during pass 7 (equals zero during all other passes)

For further details see reference 5.

Computation Load

Table 1 shows that the number of cycles needed to complete two (since the system performs two at a time) views is 7500, which is equal to 187 μsec. The time to complete 900 views is 187 × 900/2 = 84 msec.

Memory Requirement

Ports A	1K complex word
Ports B	1K complex word
Ports Q	1K complex word
Ports C	1K complex word for twiddle factors
	1K complex word for the frequency domain filter response
	1K complex word for time (space) domain shading window

The word length is 24 bits.

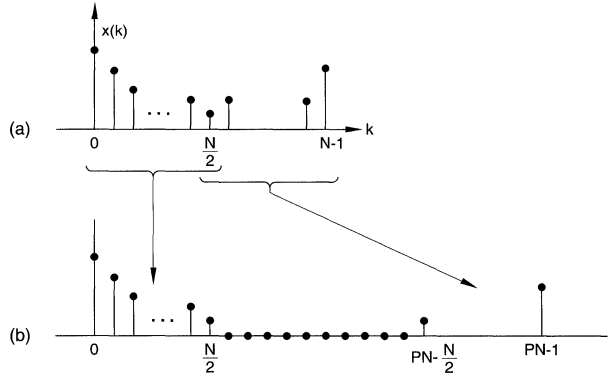
Case B: Filtering With Output Up Sampling Factor of 4

To perform an up-sampling or interpolation in the time (space) domain the data in the frequency domain should be zero-padded before doing the IFFT. When doing the zero-padding it should be noticed that the symmetric property of the frequency domain should be kept as illustrated in Figure 6.

The frequency domain zero padding should be as follows:

$$XP(k) = \begin{cases} X(k) & 0 \le k \le N/2 \\ 0 & N/2 < k < PN - N/2 \\ X(k - NP + N) & PN - N/2 \le k < PN \end{cases}$$

The algorithm block diagram for the frequency domain filter interpolator is illustrated in Figure 7. The programming for this algorithm is given in Tables 3 and 4.



(a) = Original frequency domain sequence
(b) = Zero padded frequency domain sequence

APP2-17

Figure 6. Frequency Domain Zero-Padding

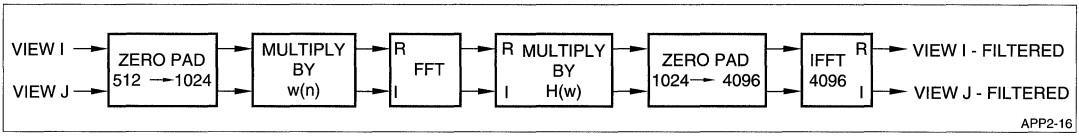


Figure 7. Two Channels Frequency Domain Filter Interpolator

Table 3. LH9124/LH9320 Programming for Frequency Domain Filter Interpolator

PASS	LH9124		PORT Q	PORT A	PORT B	PORT C	NUMBER OF CYCLES
	CODE	DATA FLOW					
1	MOVD	Q→A	PADHIGH	INC	–	–	1024 + 18
2	BWND4	A→B	–	RBF0	BF40	BRF0	1024 + 20
3	BFLY16	B→A	–	MXB4160	MXB4160	MXT4160	1024 + 68
4	BFLY16	A→B	–	MXB4161	MXB4161	MXT4161	1024 + 68
5.1	CMUL	B→A	–	PADHIGH	INC	PADHIGH	3585 + 18
5.2	CMUL	B→A	–	INC	INC	PADHIGH	512 + 18
6	BFLY16	A→B	–	RBF0	BF160	TF160	4096 + 68
7	BFLY16	B→A	–	BF161	BF161	TF161	4096 + 68
8	BFLY16	A→Q	BF162	BF162	–	TF162	4096 + 68
							20895

Table 4. Address Generator Register Settings

REGISTER	VALUE	PORT	PASS NO.
N	1024	A, B, C & Q	1, 2, 3 & 4
	4096	A, B, C & Q	5, 6, 7 & 8
DIGITREV	0000000001111111111	A	2
	0000000111111111111	A	6
ZEROPAD	512	Q	1
	513	B	5.1
ADRLENGTH	1024	Q	1
	4096 – 512 + 1 = 3585	A, B & C	5.1
	512	A, B & C	5.2
ADRINC	1	A, B, C & Q	All
INDEXINC	Zero-address	A, B, C & Q	All
ADRSTART	3585	A, C	5.2
	512	B	5.2
	0	All other cases and passes	

Explanation

Passes 1, 2, & 3

Identical to Case A.

Pass 5

Performs the frequency domain zero padding and multiply by the frequency domain filter response. The filter response should be saved in the coefficient memory in zero-padded form. The pass is composed of two subpasses.

Subpass 5.1 is with length of P N – N/2 + 1. Where N = 1024 and the up-sampling factor is P = 4. The Output and the coefficient sequences are linear sequences. The input sequence is a zero-pad where the first N/2 + 1 addresses are linear and the rest of the sequence is the '0' address (address to a location containing zero).

Sub-pass 5.2 of length N/2 is linear sequence in all the ports. The start address of the input sequence is N/2 (512), while the start address of the output and the coefficient sequences is PN-N/2.

Passes 6,7,8

Radix-16 passes.

Computation Load

From Table 3 we can see that the number of cycles needed to complete two (since the system performs two at a time) views is 20895, which equals 522 μsec. The time to complete 900 views is 522 × 900/2=235 msec.

Memory Requirement

Ports A 4K complex word
Ports B 4K complex word
Ports Q 4K complex word
Ports C 4K complex word – twiddle factors
4K complex word – frequency domain filter response
4K complex word – time (space) domain shading window.

The word length is 24 bits.

2-D-IFFT Reconstructing Algorithm Using LH9124/LH9320

The 2-D-IFFT Reconstructing algorithm is composed of three steps:

Step 1: Performing FFT_{512} to 900 views
Step 2: Building the 2-D Fourier domain using the Fourier slice theorem
Step 3: Performing 2-D-IFFT to reconstruct the image.

The algorithm block diagram is given in Figure 8.

Steps 1 and 3 are the computation intensive step of the algorithm. Implementation of this steps will be discussed below.

Step 1: Performing 900 time FFT_{512} .

The following options are available when performing a multi-channel one-dimensional FFT:

Option 1: Each FFT at a time using mixed radix 2 × 16 × 16
Option 2: Two-sequence-at-a-time method:
Loading the first view connected to the real port and the second view to the imaginary port.
FFT followed by separation pass (BRFT)
Option 3: Using the 2-N real FFT method:
Create a complex sequence with a length half that of the original sequence length by loading the even data point to the real port and the odd data point to the imaginary FFT, followed by a reconstruction pass (BFCT2)

For further details see reference 5.

The computation load for the three options are:

Option 1:	42 μsec for view	38 msec for 900 views
Option 2:	27 μsec for view	24 msec for 900 views
Option 3:	30 μsec for view	26 msec for 900 views

Step 2: Building the 2-D Fourier domain

Building the 2-D Fourier domain is mainly addressing and other processors (or dedicated hardware) are needed for this stage.

Step 3: 2-D FFT (512 × 512)

Implementing two-dimensional Fourier transform using the LH9124/LH9230 chip set is explained further in reference 5. The two-dimensional transform can be considered as a one-dimensional transform (of length N × M) with the proper addressing sequence settings.

For the case of (512) × (512), the following mixed radix can be selected:

(2 × 16 × 16) × (2 × 16 × 16)

The data addressing shall be as one-dimensional long transform, while the twiddle addressing shall be for each dimension separately. Tables 5 and 6 provide the algorithm programming.

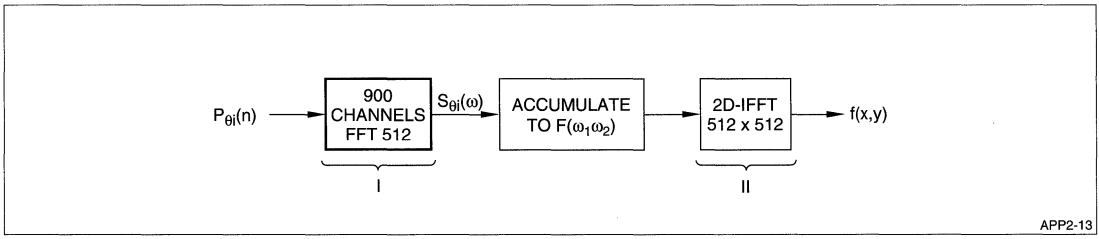


Figure 8. 2-D-IFFT Reconstruction Algorithm

Table 5. 2-D-IFFT for 512 × 512 Matrix

PASS	LH9124		PORT Q	PORT A	PORT B	PORT C	NUMBER OF CYCLES
	CODE	DATA FLOW					
1	BFLY2	Q→A	RBF0	BF20	–	TF20	512 × 512 + 20
2	BFLY16	A→B	–	MXB2160	MXB2160	MXT2160	512 × 512 + 68
3	BFLY16	B→A	–	MXB2161	MXB2161	MXT2161	512 × 512 + 68
4	BFLY2	A→B	–	BF212	BF212	TF20	512 × 512 + 20
5	BFLY16	B→A	–	MXB2162	MXB2162	MXT2160	512 × 512 + 67
6	BFLY16	A→Q	–	MXB2163	MXB2163	MXT2161	512 × 512 + 68

Table 6. Address Generator Register Settings

REGISTER	VALUE	PORT
N	512 × 512 = 262144	All
MEMSIZE	262144	C
DIGITREV	0 1111 1111 0 1111 1111	Q

Computation Load

The total number of cycles is 1.57 Mcycle = 39 msec (the clock rate is 40 MHz). The total time needed to complete steps 1 and 3 is 65-80 msec.

SUMMARY

This application note presented two major algorithms of medical imaging: the filter-back propagation algorithm and the 2-D-IFFT algorithm. It was shown that implementing the computation extensive part of the algorithms is straightforward and very efficient when using the LH9124/LH9320 chip set.

For the example that was demonstrated, the number of sensors equals 512 and the number of views equaled 900. The computation time for imaging is 80 ms. If higher performance is required, a multi-processor can be used in a cascade or a parallel configuration.

REFERENCES

[1] S. Hayking & Avi C. Kak. *Array Signal Processing.* Prentice-Hall Inc., Englewood Cliffs, New Jersey 07632.

[2] Z. H. Cho, H. S. Kim & James Cumming. *Fourier Transform Nuclear Magnetic Resonance Tomographic Imaging.* Proceeding of the IEEE Vol. 70 No. 10 October 1982 p.p. 1152-1173.

[3] Alan V. Oppenheim & Ronald W. Schafer. *Digital Signal Processing.* Englewood Cliffs, New Jersey

[4] *LH9124 Digital Signal Processor User's Guide*

[5] *LH9124 FFT Application Note*

[6] *LH9124 Data Sheet*

DFT FILTER-BANK USING THE LH9124/LH9320

INTRODUCTION

Digital filter-bank implementation arises in many applications, such as speech analysis bandwidth compression, communication, radar and sonar. Figure 1 illustrates a typical filter-bank scheme. The filter-bank structure can be interpreted as a short-time spectrum analyzer that decomposes the input signal into its instantaneous spectral components.

Usually the filter's output signals are decimated so that each filter's output rate is lower than the input rate:

$$f_{out} = f_{in}/M$$

where:

M is the decimation factor.

The extreme case is when K equals M, where K is the number of filters. Here, the output signal is critically sampled, since this rate is the minimal sampling rate without information loss.

This application note concentrates on the most important case of the filter bank – the DFT filter-bank – where the filter's frequency responses have equal spacing and equal bandwidth.

Historically, the algorithm to implement the DFT filter-bank was developed via two interpretations that were shown to be the realization of the same mathematical framework:

- Filter-bank interpretation using polyphase-filter decomposition
- Block-transform interpretation using weighted overlap-add structure

Filter-Bank Structure

Figure 2 illustrates the simplest interpretation of the filter-bank algorithm. The process for each channel is:

- Band-shifting (modulated by $e^{-j\omega_k t}$)
- Low-pass filtering by h(n)
- Decimation

where:

ω_k is the central frequency of filter k
h(n) is the prototype filter impulse response

The filter's output signals are given in equation [1].

$$X_k(m) = \sum_{n=-\infty}^{\infty} h(mM - n)x(n)W_k^{-kn} \qquad [1]$$

The design of the filter frequency response is application-dependent. For example, in communication applications where the input signal is a frequency multiplexing of several channels, the analyzer filter should be narrow to prevent overlapping in the frequency response of the filters, which appears as cross-talk between the received signals. For spectral analysis applications where the concern is a low side-lobe level and smooth estimate, the design usually allows an overlap between the filter's frequency response.

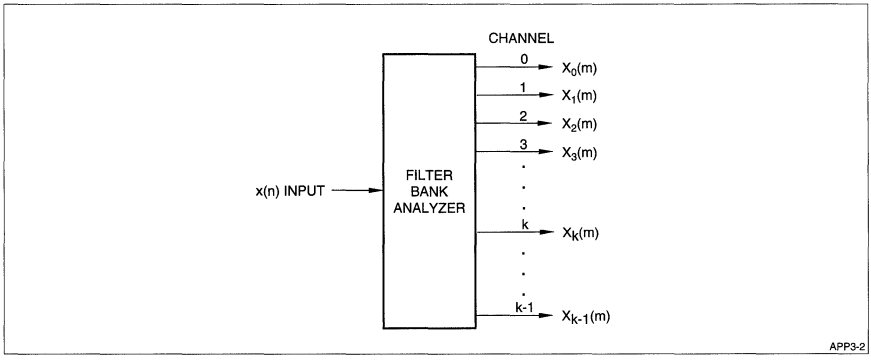


Figure 1. Typical Filter-Bank Scheme

Implementing the filter-bank analyzer according to the straightforward interpretation as formulated in equation [1] is highly inefficient. For a critical sampled DFT filter-bank, the polyphase-filter decomposition is an efficient implementation. The polyphase algorithm derivation found in [1] is summarized in equation set [2] and illustrated in Figure 3.

$$X_k(m) = \sum_{\rho=0}^{M-1} [P_\rho(m) * x_\rho(m)] \cdot W_k^{-k\rho} = \quad [2]$$

$$FFT_\rho\{P_\rho(m) * x_\rho(m)\}$$

where:

* is the convolution mark

P_ρ is the polyphase filter of branch ρ given by:

$P_\rho(m) = h(mM - \rho)$

$x_\rho(m) = x(mM + \rho)$

The polyphase structure assumes that the output is critically sampled, which means that the sampling rate of each filter is f_{in}/K, where K is the number of the filters. Increasing the output sampling rate by an integer factor I is possible by inserting a zero-padding stage of order I at the input of each polyphase filter branch.

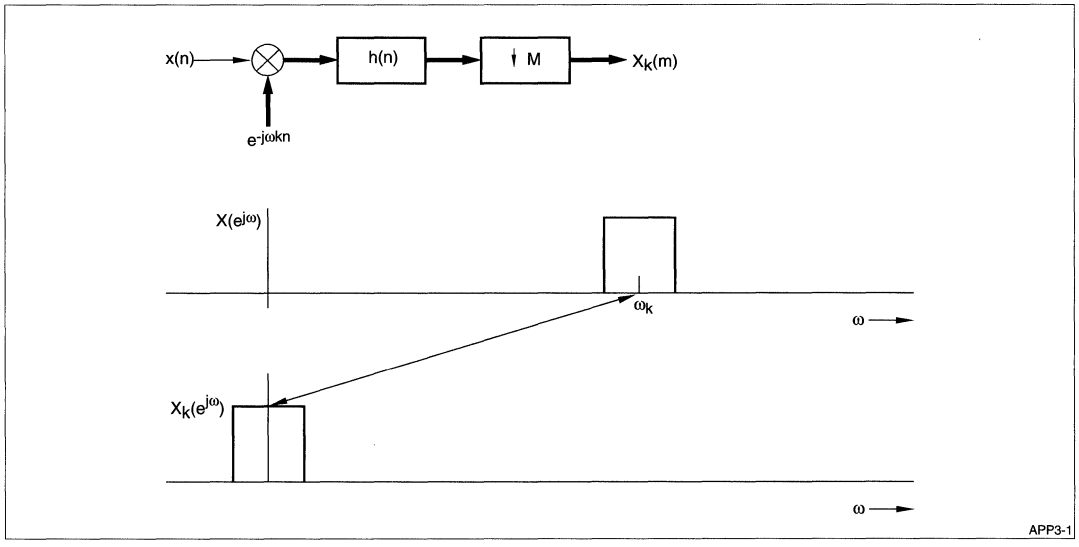


Figure 2. Simple Interpretation of Filter-Bank Algorithm

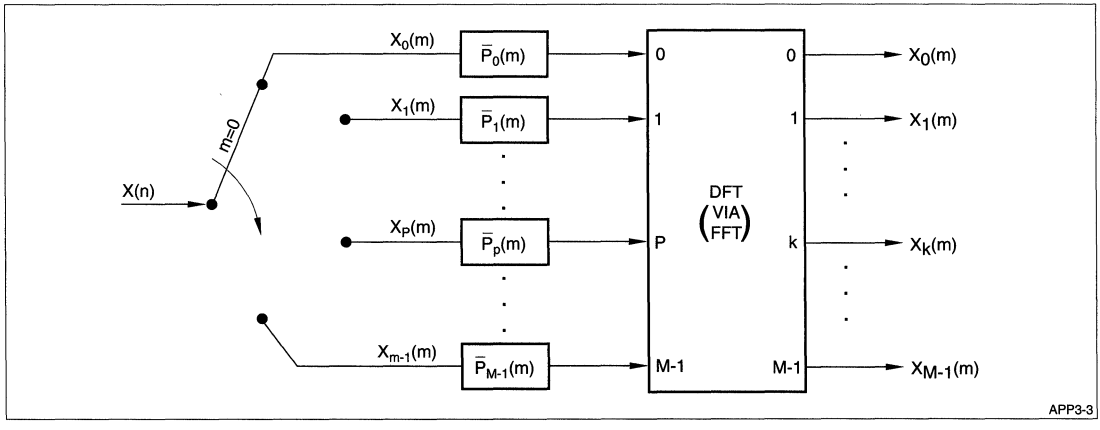


Figure 3. Polyphase DFT Filter-Bank Structure

Weighted Overlap-Add Structure

The weighted overlap-add structure implements the DFT filter-bank in terms of block-by-block analysis. This structure is more general than the Polyphase structure in the sense that it allows interpolating the output signals to any required sampling rate. As in the previous case, the output signals are given in equation [1]. The equation can be rewritten as:

$$y_m(n) = h(mM - n)x(n) \quad [3a]$$

$$x_k(m) = \sum_{n=-0}^{\infty} y_m(n)W_K^{-kn} \quad [3b]$$

Using this notation, the filter h(*) can be interpreted as a sliding analysis window that selects the short-time sequence block $y_m(n)$ (equation [3a]), and then the short-time sequence is a Fourier transform. In this interpretation, the decimated time index m is the block number, and $X_k(m)$ could be interpreted as the short-time spectrum of the signal at time n = mM. In equation [3b], the time reference point is constant n = 0. It is more convenient to have a sliding reference point always at the center of the analysis window n = mM by defining:

$$r = n - mM$$

The short-time transform is expressed as:

$$\underset{\sim}{X}_k(m) = \sum_{r=-\infty}^{\infty} h(-r)x(r + mM)W_K^{-kr} \quad [4a]$$

where:

$\underset{\sim}{X}_k(m)$ expresses the short-time Fourier transform referenced to the center of the sliding window:

$$X_k(m) = \underset{\sim}{X}_k(m)W_K^{-kmM} \quad [4b]$$

Equation [4a] is basically a Fourier transform, but it is not in a form that is compatible for the FFT algorithm, since the FFT algorithm requires that the number of the input points is equal to the number of the output point. Note that the length of the short-time input sequence $\underset{\sim}{y}_m(n)$, as determined by the analysis filter length, is generally greater than the number of the filters (the transform length). To bring equation [4] to the form that can be implemented by the FFT algorithm, the time-aliasing method is used. The sequence $\underset{\sim}{y}_m(n)$ is divided into segments of length K (K is the FFT length); the segments are summed to form the time-aliased sequence with a length of K.

The algorithm can be summarized by the following steps:

1. Multiplying the input signal x(n) by the analysis window h(–r) to form the short-time sequence:

$$\underset{\sim}{y}_m(r) = h(-r)x(r + mM)$$

2. Segmenting the short-time input sequence into blocks of length K, and summing them to form the time-aliased sequence:

$$\underset{\sim}{x}_m(r) = \sum_{l=-\infty}^{\infty} \underset{\sim}{y}_m(r + lK)$$

3. Performing a FFT to the time-aliased sequence to calculate the output:

$$\underset{\sim}{x}_k(m) = FFT\{\underset{\sim}{x}_m m(r)\}$$

4. If necessary, phase shifting the output to refer it to a fixed time reference point n = 0:

$$X_k(m) = \underset{\sim}{x}_k(m)W_K^{-kmM}$$

5. For the next block (m + 1), the data is shifted by M point and a new block of size M is loaded (the new data is overlapped by the old data by $N_h - M$, where N_h is the analyzer filter length).

Figure 4 illustrates the algorithm flow for the case of $N_h = 4 \cdot K$.

With this structure there is no constraint on the decimation factor M, while in the polyphase structure the decimation factor was limited to be an integer multiplicand of the filter bank length ($K = I \cdot M$).

For critical decimation M = K, it can be seen that the two structures, the polyphase filter and the weighted overlap-add, are identical. The only difference between the two is the order of the operations. The polyphase structure is point-by-point oriented – each new data is filtered by its corresponding polyphase filter. The output of the polyphase filters is Fourier transformed. The weighted overlap-add structure is block oriented – it first accumulates a block of data and then divides it into segments, adds the segment and performs Fourier transform.

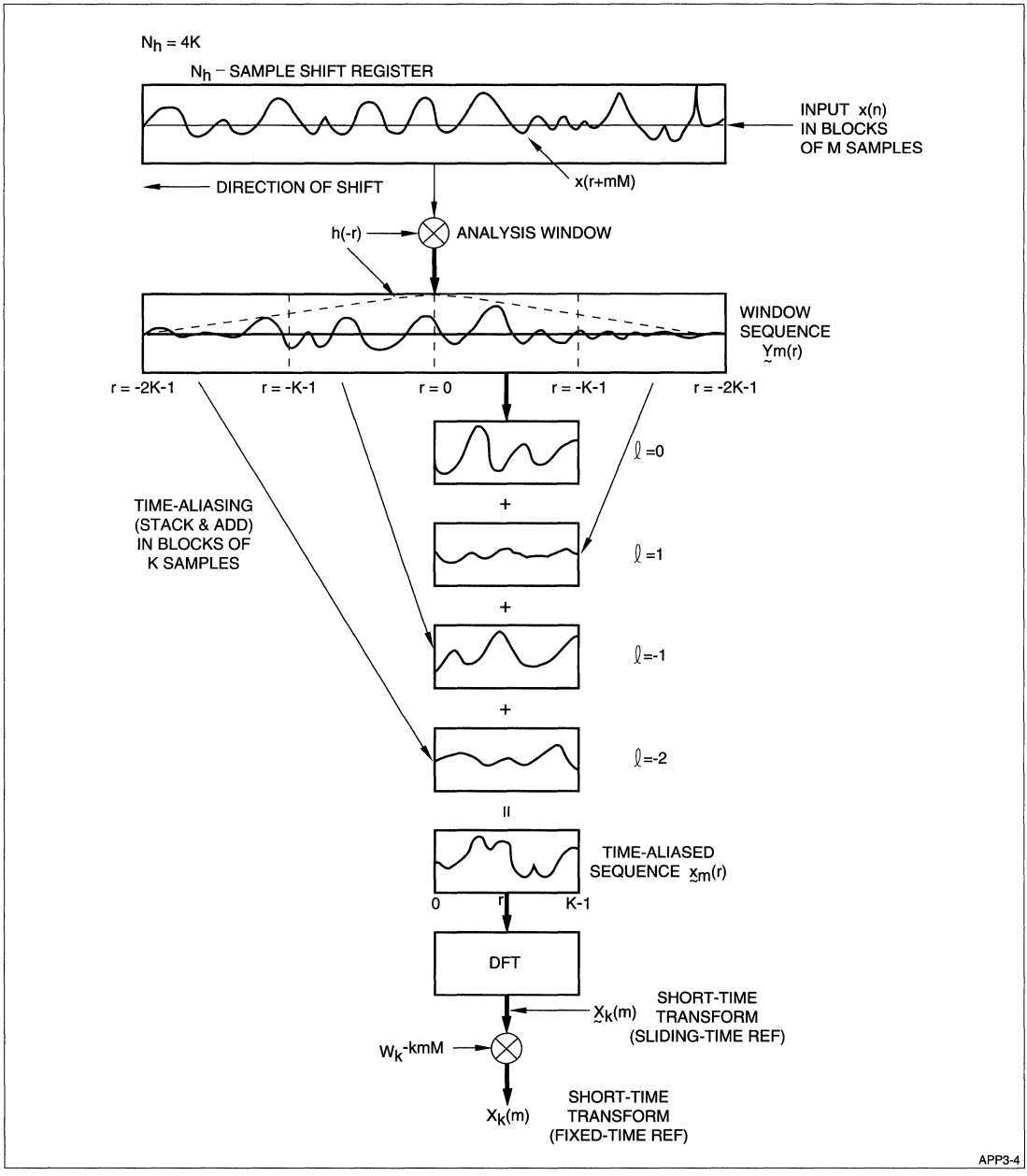


Figure 4. Weighted Overlap-Add Algorithm

Using the LH9124/LH9320

The LH9124 is a block-oriented processor; therefore, it is natural to select the weighted overlap-add method to implement the DFT filter-bank. In this section, the algorithm implementation for two cases with various FFT lengths are presented:

- Case 1: Processing two channels simultaneously
- Case 2: Processing one channel

In both cases, the standard system configuration is assumed as illustrated in Figure 5. The system includes:

- 1 DSP (LH9124)
- 4 AGs (LH9320)
- 4 complex-memory banks

Case 1: Processing Two Channels Simultaneously

Here, implementation of the algorithm for the following parameters is presented:

- Filter-bank length (FFT length) is in the range of N = 1K. 16K
- The analysis filter impulse response length is 2 · N
- Number of input channels is 2

NOTE: The case of N = 1K is presented first, then the modification needed for the other FFT lengths is given.

The input sequences {x} {y} are segmented into blocks of length N:

Xin1,Xin2,Xin3

Yin1,Yin2,Yin3

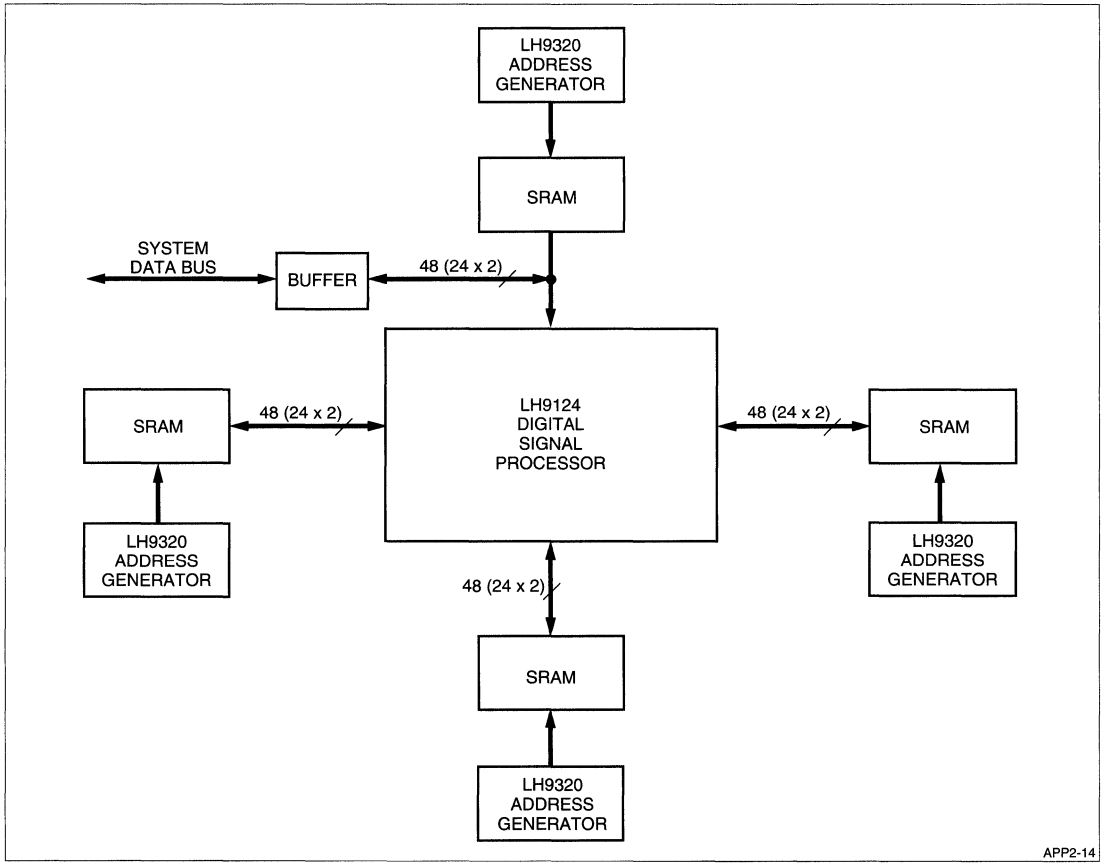


Figure 5. System Configuration

The output blocks are given by:

$$Xout1 = FFT\{Xin1 \cdot W1 + Xin2 \cdot W2\}$$

$$Yout1 = FFT\{Yin1 \cdot W1 + Yin2 \cdot W2\}$$

$$Xout2 = FFT\{Xin2 \cdot W1 + Xin3 \cdot W2\}$$

$$Yout2 = FFT\{Yin2 \cdot W1 + Yin3 \cdot W2\} \quad [5]$$

To achieve an efficient implementation with a minimum number of passes, the following key features of the LH9124 should be used:

- Performs two real FFTs at a time as follows:
  - Loading X input to the real input
  - Loading Y input to the imaginary input
  - Performing a separation pass after performing the FFT
- Performs any required mixed radix (2, 4, 16). The optimal radix for the 1K FFT case is $4 \times 16 \times 16$.
- Multiplies by a shading window during the first pass of the FFT. This option is available if the first pass of the FFT is radix-2 or radix-4.

Loading X input to the real input memory, and Y input to the imaginary input memory creates a complex input Q:

$$Q = X + jY$$

The algorithm then can be described by:

$$Out = Separation\{BFLY16\{BFLY16\{BFLY4\{Q1 \cdot W1 + Q2 \cdot W2\}\} \quad (6)$$

To gain the multiply by a window during the first pass, the linearity property of the radix-4 operation is used:

$$BFLY4(Q1 \cdot W1 + Q2 \cdot W2) = BFLY4(Q1 \cdot W1) + BFLY4(Q2 \cdot W2)$$

The algorithm as described in equation [6] can be written as:

$$Out = Separation\{BFLY16\{BFLY16\{BFLY4\{Q1 \cdot W1\} + BFLY4\{Q2 \cdot W2\}\}\}\} \quad (6a)$$

Based on equation [6a], mapping the algorithm to the LH9124/LH9320 requires seven passes as follows:

1. $BFLY4(Q1 \cdot W1) \rightarrow A$ (Q1 is X1 + jY1)
2. $BFLY4(Q2 \cdot W2) \rightarrow B$ (Q2 is X2 + jY2)
3. $A \rightarrow C$
4. $B + C \rightarrow A$
5. $BFLY16(A) \rightarrow B$
6. $BFLY16(B) \rightarrow A$
7. $Separation(A) \rightarrow Q$

Programming the algorithm is straightforward since each line of the above steps is mapped into one pass (one program line) of the chip set as detailed in Table 1.

Register Programming:

N	= 1024
MEMSIZE	= 1024
DIGITREV	= 11 1111 1111
ADRSTART	= 0
ADRLENGTH	= 1024
ADRINC	= 1

Performance: The computation time for one block of length N = 1K is:

$$T = (1024 \cdot 7 + 68 \cdot 2 + 20 \cdot 2 + 18 \cdot 3) \cdot 25\ nsec = 184\ \mu sec$$

The maximal sampling rate is 5.5 MHz.

Table 1. Algorithm Programming for Case 1

PASS	LH9124		ADDRESSING		
	OPCODE	DATA-FLOW	AGin	AGout	AGcoeff.
1	BWND4	RQWA	RBF0	BF40	RBF0
2	BWND4	RQWB	RBF0	BF40	RBF0
3	MOVD	RAWC	INC	–	INC
4	CADD	RBWA	INC	INC	INC
5	BFLY16	RAWB	MXB4160	MXB4160	MXT4160
6	BFLY16	RBWA	MXB4161	MXB4161	MXT4161
7	BRFT	RAWQ	BRFTLS	BRFTUS	–

NOTES:

1. The coefficient memory should contain three different buffers that are selectable by the system controller by paging operation, before starting a new pass.

   Buffer 1: contains the twiddle factor for the transform length of 1K

   Buffer 2: contain a real data window function (length 2K).

   Buffer 3: temporary buffer for passes 3 and 4

2. The separation pass creates for each channel only one-half of the points of the spectrum. Since the input sequences are real, this data contains all the required information (the Fourier transform of a real sequence is an anti-symmetrical function). However, if required, the full spectrum could be calculated by setting the address sequence of pass 7 to BRFTL and BRFTU instead of BRFLTS and BRFTUS. The computation time is:

   $T = (1024 \cdot 7 + 68 \cdot 2 + 20 \cdot 2 + 18 \cdot 3 + 1024) \cdot 25\ \text{nsec} = 210\ \mu\text{sec}$

Programming the algorithm for other filter-bank lengths (FFT lengths) is done by choosing the optimal radix for each FFT length. Table 2 summarizes the algorithm mapping for the various FFT lengths.

Register Programming:

N	The value of N should be the FFT length.
MEMSIZE	The value of MEMSIZE should be the maximal FFT length. See note 1.
ADRSTART	= 0
ADRLENGTH	= N
ADRINC	= 1
DIGITREV	The value of the register is determined by the radix structure: radix-2 is presented as 0 radix-4 is presented as 11 radix-2 is presented as 1111 The value of DIGITREV for the cases presented in Table 2 are:
0 11 1111 1111	for N = 2K ($2 \times 4 \times 16 \times 16$)
11 11 1111 1111	for 4K ($4 \times 4 \times 16 \times 16$)
0 1111 1111 1111	for 8K ($2 \times 16 \times 16 \times 16$)
11 1111 1111 1111	for 16K ($4 \times 16 \times 16 \times 16$)

NOTES:

1. The lookup table in Port C memory contains the twiddle factors for the maximal resolution (maximal FFT length). The value of the register MEMSIZE should reflect this value. When doing FFT with other lengths, the AG decimates the twiddle factor table.

2. The window function length should be equal to twice the current FFT length (the system controller should download them when changing the FFT length or switching between preloaded pages that contain the window functions with the various lengths).

Performance: The computation time for one block of size N is:

$T \cong N \cdot 8 \cdot 25$ nsec (neglecting the latency of each pass)

The maximal sampling rate is N/T = 5 MHz.

Table 2. Two Channels-at-a-Time Algorithm Mapping for Several FFT Lengths

	2K	4K	8K	16K
1	BFLY2(Q1 · W1) → A	BFLY4(Q1 · W1) → A	BFLY2(Q1 · W1) → A	BFLY4(Q1 · W1) → A
2	BFLY2(Q2 · W2) → B	BFLY4(Q2 · W2) → B	BFLY2(Q2 · W2) → B	BFLY4(Q2 · W2) → B
3	A → C	A → C	A → C	A → C
4	B + C → A	B + C → A	B + C → A	B + C → A
5	BFLY4(A) →B	BFLY4(A) → B	BFLY16(A) → B	BFLY16(A) → B
6	BFLY16(B) → A	BFLY16(B) → A	BFLY16(B) → A	BFLY16(B) → A
7	BFLY16(A) → B	BFLY16(A) → B	BFLY16(A) → B	BFLY16(A) → B
8	Separation(A) → Q	Separation(A) → Q	Separation(A) → Q	Separation(A) → Q

Case 2: Processing One Channel

The implementation of the algorithm is given when processing of one channel is required. The performance for this case is less than case 1, since the capability of the LH9124 to process two channels at a time is not fully used. The parameters for this case are the same as in case 1.

The algorithm for the DFT filter-bank for one channel can be described as:

$$Xout1 = FFT\{Xin1 \cdot W1 + Xin2 \cdot W2\} \quad [7]$$

The straightforward approach is to perform the same algorithm as in case 1 by setting the imaginary input to zero. The last pass (separation pass) is not required. The algorithm for N = 1K completes in six passes:

1. BFLY4(Q1 # W1) → A (Q1 is X1)
2. BFLY4(Q2 · W2) → B (Q2 is X2)
3. A → C
4. B + C → A
5. BFLY16(A) → B
6. BFLY16(B) → Q

Performance: One block computation time:

$$T = (1024 \cdot 6 + 68 \cdot 2 + 20 \exists 2 + 18 \cdot 2) \cdot 25 \text{ nsec} = 158 \mu\text{sec}$$

The maximal sampling frequency: F = 6.4 MHz.

A more efficient implementation is achieved by working with:

- Complex input data: Xin = Xin1 + jXin2
- Complex window function: W = W1 – jW2

The complex data input is received by loading the first data block, Xin1, to the real input memory, and the second data block, Xin2, to the imaginary input memory. The complex window is received by loading the first half of the analyzer window, W1, to the real coefficient memory, and the second half (negated), –W2, to the imaginary coefficient memory.

Multiplying the complex input by the complex window gives:

$$Xin \cdot W = (Xin1 + jXin2)(W1 - jW2) = (Xin1 \cdot W1 + Xin2 \cdot W2) + j(Xin1 \cdot W2 - Xin2 \cdot W1) \quad [8]$$

The real part of XinW is (Xin1 · W1 + Xin2 · W2). According to equation [7], Fourier transforming this sequence gives the required output. However, the time domain sequence Xin · W contains an imaginary part; therefore, a separation pass is required after Fourier transforming Xin · W, where only the first half of the data is relevant.

The algorithm can be formulated as:

$$Out = Separation\{BFLY16\{BFLY16\{BFLY4\{Xin \cdot W\}\}\}\} \quad [9]$$

where:

$$Xin = Xin1 + jXin2$$

$$W = W1 - jW2$$

Based on equation [9], mapping the algorithm to the LH9124/LH9320 system, for N = 1K, requires four passes as follows:

1. BFLY4(Q · W) → A
   (Q = Xin1 + jXin2; W = W1 (jW2)
2. BFLY16(A) → B
3. BFLY16(B) → A
4. Separation(A) → Q
   (only the first N/2 point is relevant)

Table 3 provides the algorithm programming for this case.

Register Programming:

N	= 1024
MEMSIZE	= 1024
DIGITREV	= 11 1111 1111

Performance: The computation time for N = 1K is:

$$T = (1024 \cdot 4 + 68 \cdot 2 + 18 \cdot 2) \cdot 25 \text{ nsec} = 107 \mu\text{sec}$$

The maximal sampling rate is 9.5 MHz.

Programming the algorithm for other filter-bank lengths (FFT lengths) is done by choosing the optimal radix for each FFT length. Table 4 summarizes the algorithm mapping for each FFT length.

Register Programming:

N	The value of N should be the FFT length.
MEMSIZE	The value of MEMSIZE should be the maximal FFT length. See note of Case 1.
DIGITREV	The radix structure determines the value of this register. See Case 1.

Performance: The computation time for one block with size N is:

$T \cong N \cdot 5 \cdot 25$ nsec (neglecting the latency of each instruction)

The maximal sampling rate is N/T = 8 MHz.

Table 3. Algorithm Programming for Case 2

PASS	LH9124		ADDRESSING		
	OPCODE	DATA-FLOW	AGin	AGout	AGcoeff.
1	BWND4	RQWA	RBF0	BF40	RBF0
2	BFLY16	RAWB	MXB4160	MXB4160	MXT4160
3	BFLY16	RBWA	MXB4161	MXB4161	MXT4161
4	BRFT	RAWQ	BRFTLS	BRFTUS	–

Table 4. One Channel, Algorithm Mapping for Several FFT Lengths

	2K	4K	8K	16K
1	BFLY2(Q1 · W1) → A	BFLY4(Q1 · W1) → A	BFLY2(Q1· W1) → A	BFLY4(Q1 · W1) → A
2	BFLY4(A) → B	BFLY4(A) → B	BFLY16(A) → B	BFLY16(A) → B
3	BFLY16(B) → A	BFLY16(B) → A	BFLY16(B) → A	BFLY16(B) → A
4	BFLY16(A) → B	BFLY16(A) → B	BFLY16(A) → B	BFLY16(A) → B
5	Separation(A) → Q	Separation(A) → Q	Separation(A) → Q	Separation(A) → Q

SUMMARY

This application note presents the DFT filter-bank. It is shown that by selecting the proper algorithm, weighted overlap-add, an efficient and simple implementation is achieved using the LH9124/LH9320 chip set.

Programming the algorithm is straightforward and requires between 4-8 lines of program (depending on the parameters).

Programming the algorithm for two cases is given. In the first case, two channels are processed simultaneously; in the second case, only one channel is processed. For both cases, the implementation for a wide range of FFT lengths is given. The maximal sampling rate that can be achieved by a single DSP system with recursive mode is 5 MHz for Case 1, when two channels are processed simultaneously, and 8 MHz for Case 2, when only one channel is processed. The system performance has a small sensitivity to the FFT length. If higher performance is needed, then a multi-DSP configuration can be used.

REFERENCES

[1] A. Oppenheim and R. Schafer, *Discrete-Time Signal Processing*, Prentice-Hall, 1989.

[2] R. Crochiere and L. Rabiner, *Multirate Digital Signal Processing*, Prentice-Hall.

[3] *LH9124 User's Guide.*

[4] *LH9320 User's Guide.*

SPECTRAL ANALYSIS

INTRODUCTION

Spectral Analysis is the analysis of the magnitude squared frequencies to determine the composition of a signal. The following description demonstrates how to perform a spectral analysis, and describes the LH9124 functions that allow you to compute the discrete Fourier transform of finite-duration sequences.

For this demonstration, spectral analysis will involve estimating a Fourier transform X(f) of the signal x(t). Within this relationship, the actual quantity of interest is the power density spectrum (PDS), $P_{xx}(f)$, which indicates how signal power is distributed over frequency.

As such, when a discrete-time signal x(n) is given, an FFT of the x(n), X(k) values is used for estimating X(f). $|X(k)|^2$ is a scaled estimate of the power density spectrum $P_{xx}(f)$.

If x(t) is a random signal, such that x(n) is a random sequence, the estimates cannot be used due to their inherent inaccuracy. In this case, the spectral analysis is performed by computing spectral estimates from the non-overlapping sections of x(n), then by averaging the results. This averaging operation improves the accuracy of the approximations.

PDS of a Stationary Signal

To determine the power density spectrum of a stationary signal, consider the stationary random process {xn}. For a zero initial condition, sampled sequence, x(n), the mean value m_x is defined as:

$$m_x = \lim_{n \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x(n) \qquad [1]$$

If N is large enough, the estimate of mx, $\hat{m}_x$, is often sufficiently accurate:

$$\hat{m}_x = \frac{1}{N} \sum_{n=0}^{N-1} x(n) \qquad [2]$$

The variance of this process {x(n)} is:

$$\delta^2 = E[(x_n - m_x)^2] \qquad [3]$$

When E is an operation for an expected value or mean value, the auto-covariance sequence of {x(n)} is:

$$\gamma_{xx}(m) = E[(x_n - m_x)(x^*_{n+m} - m^*_x)] \qquad [4]$$

where:

* denotes complex conjugation.

The Fourier transform of $\gamma_{xx}(m)$, $P_{xx}(\omega)$ is equal to the power density spectrum as follows:

$$P_{xx}(\omega) = \sum_{m=-\infty}^{\infty} \gamma_{xx}(m)\, e^{-j\omega m} \qquad [5]$$

The integral of $P_{xx}(\omega)$ over a band of frequencies is proportional to the power of each signal. In the band. $P_{xx}(\omega)$ is a symmetrical function and is non-negative.

When $m_x = 0$, the PDS can be defined as the Fourier transform of the auto-correlation sequence, rather than the auto-covariance. In most applications of spectral estimation, the 'white noise' or $m_x = 0$ is the case.

In the LH9124, the estimate of $P_{xx}(\omega)$ can be computed from the $I^2 + R^2$ of FFT output, with the estimate of $\gamma_{xx}(m)$ obtained before $P_{xx}(\omega)$ is computed. Because the auto-covariance and auto-correlation of a random process are the same when $m_x = 0$, the estimate of either is called the estimate of the auto-correlation sequence, denoted as

$$Cxx(m) = E[X(n)X^*(n+m)]$$

or:

$$C_{xx}(m) = \frac{1}{N} \sum_{n=0}^{N-|m|-1} x(n)\, x(n+m) \qquad [6]$$

$$0 \le |m| < N-1$$

Note that $C_{xx}(m)$ is a biased estimate of the auto-correlation sequence, although it is asymptotically unbiased.

Averaged Periodogram Spectral Estimation

Smoothing the Fourier transform of the variance estimate produces a good estimate of the power spectrum. Where exact expressions for the variance of spectral estimates become cumbersome, some approximate expressions can be easily interpreted. The following description defines a spectrum analysis that obtains approximate values for the PDS.

The Fourier transform of the biased auto-correlation estimate $C_{xx}(m)$ can be considered an estimate of the power density spectrum, $I_N(\omega)$.

$$I_N(\omega) = \sum_{m=-(N-1)}^{N-1} C_{xx}(m)\, e^{-j\omega m} \quad [7]$$

For a finite length real sequence x(n), the Fourier transform is:

$$X(e^{j\omega}) = \sum_{n=0}^{N-1} x(n)\, e^{-j\omega}$$

where:

$$0 \le n \le N-1$$

since:

$$C_{xx}(m) = \frac{1}{N}[x(m) \cdot x^*(-m)] \quad [8]$$

then:

$$I_N(\omega) = \frac{1}{N}[X(e^{j\omega})\, X^*(e^{j\omega})] = \frac{1}{N}|X(e^{j\omega})|^2 \quad [9]$$

The spectral estimate $I_N(\omega)$ is called a periodogram. This periodogram is the biased estimate of the power density spectrum $P_{xx}(\omega)$.

In a zero mean process:

$$E[C_{xx}(m)] = \frac{N-|m|}{N}\varphi_{xx}(m)$$

where:

$$|m| < N \quad [10]$$

then:

$\varphi_{xx}(m)$ is the auto-correlation of the random process:

$$E[I_N(\omega)] = \sum_{m=-(N-1)}^{N-1} \left(\frac{N-|m|}{N}\right) \varphi_{xx}(m)\, e^{-j\omega m} \quad [11]$$

The equation can be interpreted as the Fourier transform of a windowed auto-correlation sequence. The triangular window (Bartlett's window) is:

$$w(m) = \begin{cases} \dfrac{N-|m|}{N}, & |m| < N \\ 0, & |m| \ge N \end{cases}$$

An alternative rectangular window function is:

$$w(m) = \begin{cases} 1, & |m| < N \\ 0, & |m| \ge N \end{cases} \quad [12]$$

Since a periodogram is not a consistent estimate of the spectrum, modifications are necessary for more accurate results. Welch has introduced the Bartlett modification procedure which is particularly suited to the direct computation of a power spectrum estimate using an FFT.

In this procedure, the data record is segmented into K = N/M. N is the total length of the data and M is the sample number of each segment. There are K modified periodograms as indicated in the equation:

$$I^{(1)}M(\omega) = \frac{1}{MU}\left|\sum_{n=0}^{M-1} x^{(i)}(n)\, w(n)\, e^{-j\omega n}\right|^2$$

where:

$$i = 0, 1, \ldots, K \quad [13]$$

then:

the superscript (i) indicates the segment index number:

$$U = \frac{1}{M}\sum_{n=0}^{M-1} w^2(n) \quad [14]$$

This equation is a normalizing factor for window w(n).

The estimate of the power density spectrum is the average of $\left|{}^{(i)}_{M}\omega\right|$, defined as:

$$\beta_{xx}(\omega) = \frac{1}{K}\sum_{i=1}^{K} I^{(i)}_{M}(\omega) \quad [15]$$

This estimate of PDS is the Bartlett-Welch method which is computed using an FFT as discussed later.

Spectral Analysis Applications with Digital Signal Processors

The LH9124 uses a new architectural approach that allows a variety of FFT based DSP system configurations. This device can be used to estimate the power density spectrum. Using the complex arithmetic

instruction CMAG (the square of the magnitude of a complex number, for example, $I^2 + R^2$), the spectrum analysis is well supported using the following two applications.

Computation of the Bartlett-Welch Method

To compute a spectrum estimate at equally spaced frequencies by averaging periodograms, the frequency ω in Equations [13], [14] and [15] is substituted by $\frac{2\pi}{M}k$. The data record is sectioned into K = N/M segments of M samples each, changing Equation [15] to:

$$\beta_{xx}\left(\frac{2\pi}{M}k\right) = \frac{1}{K}\sum_{i=1}^{K} I_M^{(i)}\left(\frac{2\pi}{M}k\right)$$

where:

$$k = 0, 1, ..., M-1$$

then:

$$I_M^{(i)}\left(\frac{2\pi}{M}k\right) = \frac{1}{MU}\left|\sum_{n=0}^{M-1} x^{(i)}(n)\, w(n)\, e^{-j(2\pi/M)kn}\right|^2 \qquad [16]$$

$X_M^{(i)}$ is:

$$X_M^{(i)} = \sum_{n=0}^{M-1} x_n^{(i)}\, w(n)\, e^{-j(2\pi/M)kn}$$

and:

$$k = 0, 1, ..., M-1 \qquad [17]$$

Equation [17] can be computed by using FFT algorithm. Each data record segment can be transformed into $X_M^{(i)}$ and then, by using the CMAG function of the LH9124, $|X_M^{(i)}(k)|^2$ can be easily obtained and these results added together. When all K of the estimates have been accumulated, the result is divided by KMU.

Computation of the Correlation Estimate.

The correlation is a very important parameter in spectral analysis, especially when a signal has an infinite power spectrum. The correlation estimate is obtained using the following method.

The FFT is used to efficiently compute the auto-correlation estimate, which is defined as:

$$C_{xx}(m) = \frac{1}{N}\sum_{n=0}^{N-|m|-1} x(n)\, x(n+m) \qquad [18]$$

where:

$$0 \le |m| \le M-1$$

and:

$$M < N.$$

The convolution of x(m) and x(–m) is:

$$x(m) * x(-m) = \sum_{n=-\infty}^{\infty} x(n)\, x(n+m)$$

$$= \sum_{n=0}^{N-|M|-1} x(n)\, x(n+m) \qquad [19]$$

where:

$C_{xx}(m)$ is the convolution of x(m) and x(–m).

Now suppose that:

$$X(k) = FFT[x(m)],$$

then:

$$X^{*(k)} = FFT[x(-m)]$$

and:

$$X(k) \cdot X^*(k) = |X(k)|^2 = FFT[X(m) * X(-m)] \qquad [20]$$

The inverse FFT of $|X(k)|^2$ is the circular convolution of x(m) and x(–m). This means that by augmenting the sequence x(n) with (L – N) zero samples and computing an L-point FFT, the values of the circular convolution are correct in the interval $0 \le m \le M-1$. If $0 \le m \le M-1$, the number of padded zero samples is M – 1.

To obtain the correlation estimate, follow the steps:

1. Pad x(n) with (M – 1) zero samples to construct an L-point sequence.
2. Compute an L-point FFT:

   $$X(k) = \sum_{n=0}^{L-1} x(n)\, e^{-j(2\pi/L)kn}$$

   where:

   $$k = 0, 1, ..., L-1$$

3. Compute the magnitude-square of X(k):

   $$|X(k)|^2$$

4. Compute an L-point inverse FFT of $|X(k)|^2$:

   $$c(m) = \frac{1}{L}\sum_{k=0}^{L-1} |X(k)|^2\, e^{j(2\pi/L)km}$$

   where:

   $$m = 0, 1, ..., L-1$$

5. Divide the result from (d) by N to obtain the auto-correlation estimate:

$$C_{xx}(m) = \frac{1}{N} c(m)$$

Since the LH9124 is a very powerful device for FFT computation, the correlation estimate, $C_{xx}(m)$ can be easily and efficiently obtained.

An FFT system can be implemented with the LH9124 by using either a single or dual memory recursive system. If necessary, cascaded system architecture also can be used. In this case, the algorithm is defined by programming the LH9230 Address Generator instruction memory in the LH9124, and the hardware configuration is defined by appropriately programming the LH9124 control registers.

The FFT algorithm can be turned into a spectrum analysis algorithm by adding two additional passes (two operation codes), BWND (complex window pass) and CMAG, (square of magnitude of a complex number).

Note that memory for holding the window coefficients is required in addition to the trigonometric coefficient memory.

A pass means the transferring of a length of data array from one memory to another under control of the LH9230. (In a single memory system, a pass changes the data array to the same memory.)

Note that these transfers can be implemented on a recursive dual memory system, a recursive single memory system, a cascaded memory system or intermediate hybrid systems to achieve the desired performance.

REFERENCES

[1] W. B. Davenport, *Probability and Random Processor*, McGraw-Hill Book Company, New York, 1970.

[2] A. V. Oppenheim & R.W. Schater, *Digital Signal Processing*, Prentice-Hall, Inc., 1975.

[3] M. S. Bartlett, *An Introduction to Stochastic Processes with Special Reference to Methods & Applications*, Cambridge University Press, New York, 1975.

IMAGE CORRELATION

INTRODUCTION

Recognition of objects is needed in many applications. These include fields such as robotics, military targeting, and satellite surveillance.

The basis for many of these applications is the correlation of a sample signal with a desired signal. This produces a 'best fit' of where the desired signal is located in the sample, and of how good a fit it is. The application can then use some sort of artificial intelligence to interpret the results.

In robotics this can be used in determining location and camera angle information by comparing a video input from the camera to a list of pre-stored location images.

By using the LH9124 digital signal processor, XXXX number of 256x256 grey scale images can be correlated each second. Thus, by using the LH9124, low cost high end image recognition can be achieved.

Correlation Theory

$$\sum_{i=0}^{N-1} x(i)h(k+i) = \sum_{i=0}^{N-1}\left[\frac{1}{N}\sum_{n=0}^{N-1} X(n)e^{j2\pi ni/N}\right]\left[\frac{1}{N}\sum_{m=o}^{N-1} H(m)e^{j2\pi m(k+i)/N}\right] \quad [1]$$

From the Alternate Inversion Formula:

$$\frac{1}{N}\sum_{n=0}^{N-1} X(n)e^{j2\pi nj/N}$$

$$= \frac{1}{N}\sum_{n=0}^{N-1}\left(R_x(n)+ jI_x(n)\right)\left(\cos\frac{2\pi ni}{N} + j\sin\frac{2\pi ni}{N}\right)$$

$$= \frac{1}{N}\left(\sum_{n=0}^{N-1} R_x(n)\cos\frac{2\pi ni}{N} - I_x(n)\sin\frac{2\pi ni}{N}\right)$$

$$+ j\left(\sum_{n=0}^{N-1} R_x(n)\sin\frac{2\pi ni}{N} + I(n)\cos\frac{2\pi ni}{N}\right)$$

Conjugate:

$$= \frac{1}{N}\left(\left(\sum_{n=0}^{N-1} R_x(n)\cos\frac{2\pi ni}{N} - I_x(n)\sin\frac{2\pi ni}{N}\right)\right.$$

$$\left. -j\left(\left(\sum_{n=0}^{N-1} R_x(n)\sin\frac{2\pi ni}{N} + I_x(n)\cos\frac{2\pi ni}{N}\right)\right)^*\right)$$

Recombine:

$$= \frac{1}{N}\left(\left(\sum_{n=0}^{N-1}(R_x(n) - jI_x(n)\right)\left(\cos\frac{2\pi ni}{N} - j\sin\frac{2\pi ni}{N}\right)^*\right)$$

$$= \frac{1}{N}\left(\sum_{n=0}^{N-1}(R_x(n) - j\,I_x(n))\,e^{-j2\pi nj/N}\right)$$

$$= \frac{1}{N}\left(\sum_{n=o}^{N-1} X^*(n)\,e^{-j2\pi nj/N}\right)^* \quad [2]$$

By substituting [2] into [1]:

$$\sum_{i=0}^{N-1} x(i)h(k+ i)$$

$$= \sum_{i=o}^{N-1}\left(\frac{1}{N}\sum_{n=o}^{N-1} X*(n)e^{-2\pi nj/N}\right)^*\left(\frac{1}{N}\sum_{m=o}^{N-1} H(m)e^{j2\pi m(k+ i)/N}\right) \quad [3]$$

Since the second conjugation in [3] occurs only after calculating Rx an Ix, and since x(i) has only real values, the second conjugation can be ignored.

By doing this, and by rearranging, equation [3] becomes:

$$\sum_{i=0}^{N-1} x(i)\,h(k+ i) = \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} X^*(n)\,H(m)\,e^{j2\pi mk/N}\left(\frac{1}{N}\sum_{i=0}^{N-1} e^{-2\pi nj/N}\,e^{j2\pi mj/N}\right) \quad [4]$$

Since the right-hand side is periodic in N, it evaluates to zero for all cases of n and m except n = m where it equals N.

$$\sum_{i=0}^{N-1} e^{-j2\pi ni/N} e^{j2\pi mi/N} = \begin{cases} N, & \text{for } n = m \\ 0, & \text{otherwise} \end{cases}$$

Therefore, [4] evaluates to:

$$\sum_{i=0}^{N-1} x(c)\, h(k+i) = \frac{1}{N} \sum_{n=0}^{N-1} X^*(n)\, H(n)\, e^{j2\pi nk/N} \qquad [5]$$

Thus, the correlation can be done in the frequency domain: multiplying the conjugated frequency input by the frequency input of the correlation mask, and taking the inverse FFT (see Figure 1).

The boundaries required for correlation, as illustrated in Figure 2, need to be greater than the sum of the boundaries of X and H. This is because the correlation produces data on the correspondence between X and H for both positive and negative shifts of H.

Therefore, N > a + b – 1.

Also, since N points of each function have to be transformed by shifting x(k) into the right-most points as in Figure 3, the result will be contiguous with a known shift of N – a.

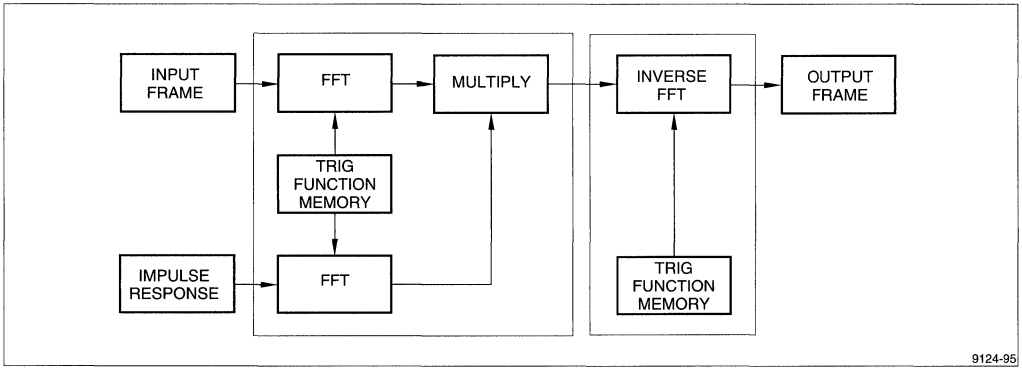


Figure 1. Correlation Flow Graph

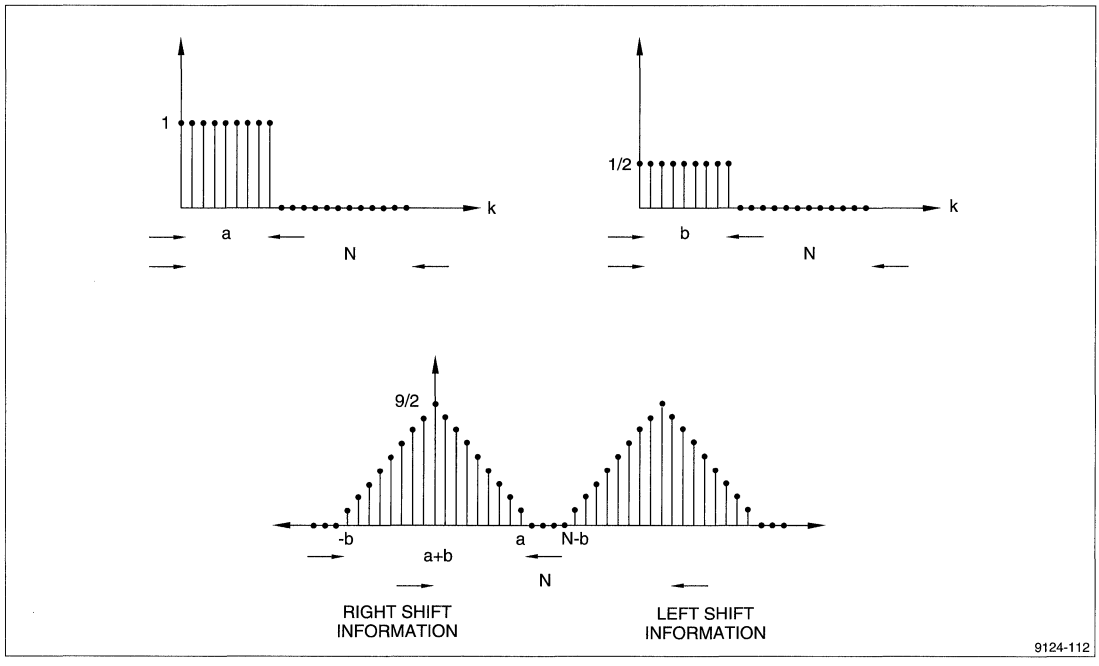


Figure 2. Non Shifted 1-D Correlation

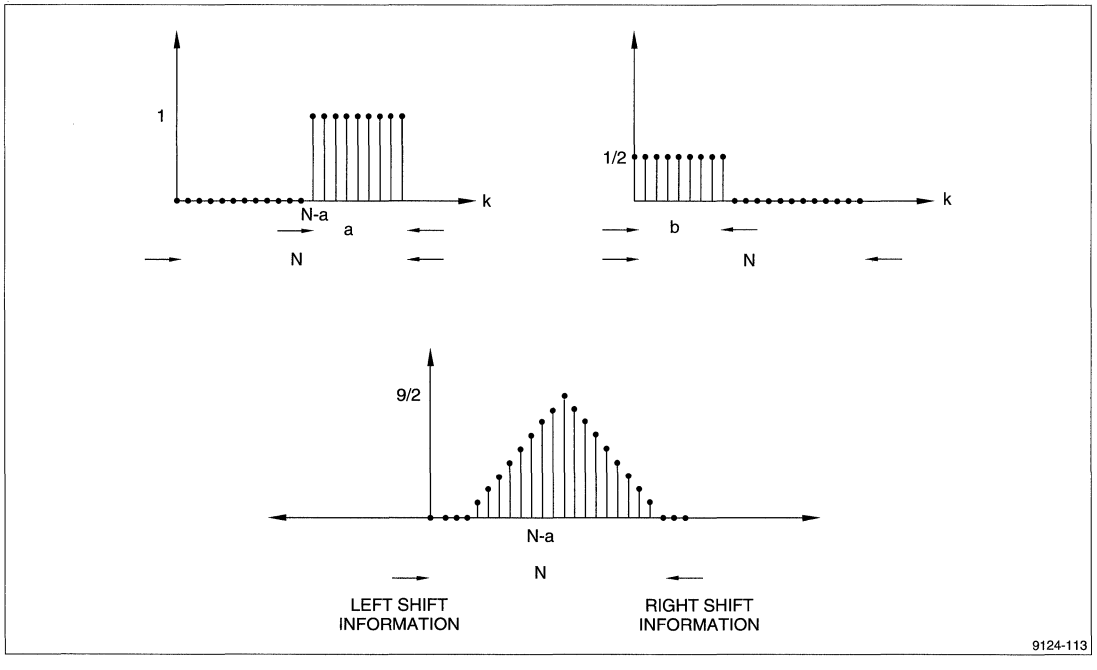


Figure 3. Shifted 1-D Correlation

FAST CONVOLUTION

INTRODUCTION

A linear convolution of two sequences is the process of modeling (filtering) the output response of a signal according to an input signal and the impulse response of the system that affects the signal.

A *Fast* Convolution is a convolution (filtering) process that is performed in the frequency domain. A fast convolution transforms the data into the frequency domain via a Fast Fourier Transform (FFT), the data is then multiplied and the result is transformed back into the time domain using the inverse FFT. Refer to Figure 1.

Because the Fast Convolution corresponds to a circular convolution of the two sequences, the key question is:

How is the result of a linear convolution of obtained using a circular convolution method?

The following description demonstrates how to implement the linear convolution of two sequences, and describes the LH9124 functions that allow you to compute the discrete Fourier transform of finite-duration sequences.

Linear & Circular Convolution

In a time-invariance system, suppose we have two discrete sequences, x(n) and h(n), with the linear convolution of these two sequences defined in the following equation:

$$y(n) = x(n) * h(n) = \sum_{k=-\infty}^{\infty} h(k)\, x(n-k) \qquad [1]$$

The order for convoluting these two sequences is unimportant, hence:

$$y(n) = x(n) * h(n) = h(n) * x(n) \qquad [2]$$

Note that the convolution sum is from $k = -\infty$ to ∞. In a practical signal case, the sum can be from $k = 0$ to ∞.

For a linear convolution, the basic operation involves multiplying x(n) by a reversed and linearly shifted version of h(n), and then by summing the values of the products. To obtain the successive values of the sequence representing the convolution, the two sequences are successively shifted relative to each other.

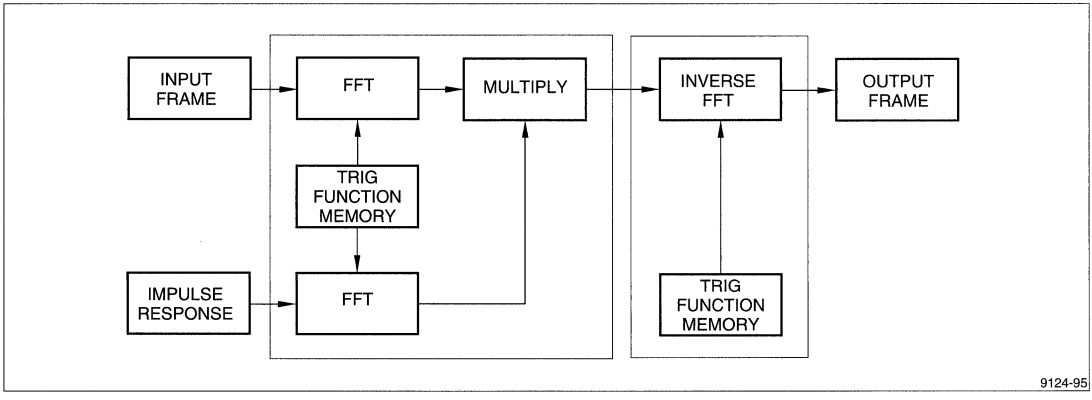


Figure 1. Fast Convolution Method

Consider two finite-duration sequences x(n) and h(n), both of length of N, with DFTs of X(k) and H(k) respectively. For example:

$$X(k) = \sum_{n=0}^{N-1} x(n)\, W_N^{nk} \quad \text{and} \quad H(k) = \sum_{n=0}^{N-1} h(n)\, W_N^{nk} \qquad [3]$$

Here, $W_N = e^{-j\frac{2\pi}{N}}$. To determine sequence y(n) with DFT coefficients, X(k) and H(k), assume that the two corresponding periodic sequences of period N (x(n) and h(n) are one period) are denoted by $\tilde{x}(n)$ and $\tilde{h}(n)$. Also, y(n) denotes one period of a periodic sequence of $\tilde{y}(n)$. Thus:

$$y(n) = \left[\sum_{k=0}^{N} -1\ \tilde{x}(k)\, \tilde{h}(n-k) \right] P_N$$

where:

$$P_N = \begin{cases} 1, & 0 \le n \le N \\ 0, & \text{otherwise} \end{cases} \qquad [4]$$

This interpretation of convolution is often referred to as *circular* convolution. For a circular convolution, one sequence is displayed around the circumference of a cylinder with a circumference of N points. The other sequence is displayed around the circumference of a second cylinder with a circumference of N points, *with the display reversed in time*. Then one of the cylinders is placed inside the other. The successive values of the convolution are obtained by multiplying the corresponding values of the two cylinders and forming the sum of these products.

To generate the successive results of the convolution, one cylinder must be rotated with respect to the other. This operation is equivalent to the convolution of two periodic sequences. The circular convolution requires that x(n), h(n) and y(n), the result of the circular convolution, are of N length. For a linear convolution of two finite duration sequences x(n) and h(n), both with length N:

$$y(n) = \sum_{k=0}^{N} x(k)\, h(n-k) \qquad [5]$$

The result of the linear convolution in Equation [5] is different from the circular convolution, both in the corresponding values and in the length of y(n), which is 2N – 1. If x(n) and h(n) have different lengths, say N and M respectively, the linear convolution of h(n) and x(n) has a length of N + M – 1.

Examples of Fast Convolution

Many applications require the implementation of a linear convolution of two sequences. For example, a filter impulse response (FIR) digital filter operation is a linear convolution of the finite duration FIR with the input signal.

To perform the filtering operation in the frequency domain using the FFT and IFFT algorithm, the input long sequence must be broken into data arrays of length N. Length N depends on the length of the FIR duration and the hardware/software system used with the LH9124.

Because the FIR is usually known, the respective FFT calculation H(k) can be executed beforehand and the H(k) data can be stored in the constant/coefficient RAM in the LH9124. The H(k) data is used to multiply the FFT data of the input signal. (The products of this multiplication are the next subject of the IFFT computation.)

This procedure is equal to the circular convolution of h(n) and x(n) as shown. In this case, if a linear convolution is to be obtained, the *circular* convolution has to have the effect of a *linear* convolution. For the LH9124, the preferable method for obtaining these results is the technique of overlap-discard.

Overlap-Discard Procedure

To illustrate the overlap-discard procedure in a practical application, assume that the signal and the FIR are given by an equation where the length of h(n) is equal to M = 4. Assuming that the FFT length is N = 8, the overlap-discard method is used to achieve a linear convolution. Refer to Figure 2 for block diagram of the overlap-discard procedure.

Notice the following equation:

$$h(n) = \begin{cases} 1 \text{ for } 0 \le n \le 3 \\ 0 \text{ for otherwise} \end{cases}$$

$$x(n) = \begin{cases} 0 \text{ for } n < 0,\ n = 3, 7 \\ 1 \text{ for } n = 0, 2 \\ -1 \text{ for } n = 4, 6 \\ 2 \text{ for } n = 1 \\ -2 \text{ for } n = 5 \end{cases}$$

$$x(n) = x(n+8) \quad \text{for } n \ge 0 \qquad [6]$$

The next subsection must be overlapped with the previous subsection by an amount equal to M-1 points (3 points in this example). To do this, the M-point h(n) must be extended to N-point by padding with (N – M) zeros. (In this case, four zeros.) In Figure 3, the first M – 1 samples of $x_1(n)$ are padded with zeros since the preceding section does not exist. That is, the initial conditions of the system are set to zero. Then each of the subsections x(n), $x_1(n)$, $x_2(n)$,and so on, perform the circular convolution with h(n) as illustrated by the multiplication of matrices.

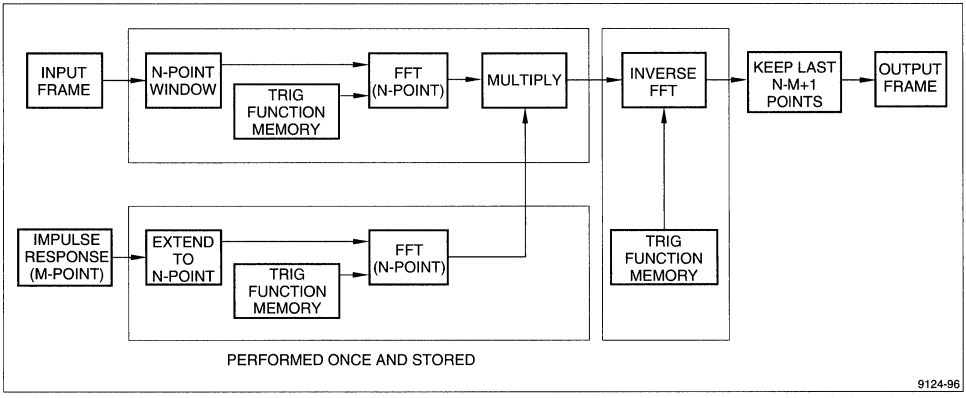


Figure 2. Overlap-Discard Method for Fast Convolution

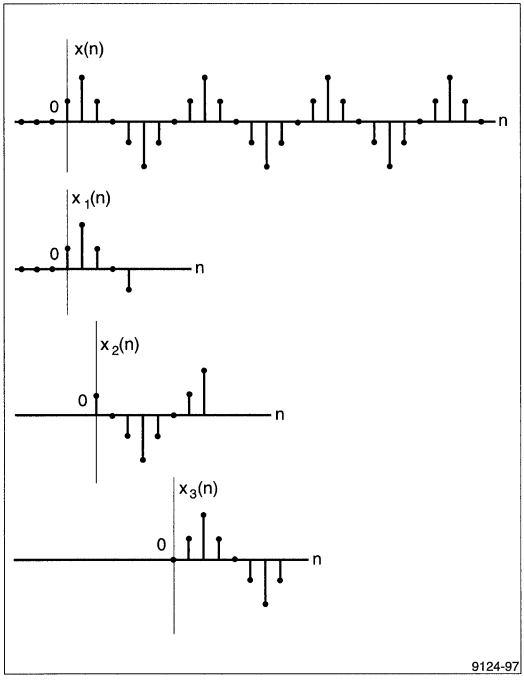


Figure 3. Example of Overlap-Discard Method

The circular convolution of subsection $x_1(n)$ with zero-padded h(n) yields:

$$\begin{bmatrix} y_1(0) \\ y_1(1) \\ y_1(2) \\ y_1(3) \\ y_1(4) \\ y_1(5) \\ y_1(6) \\ y_1(7) \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 2 & 1 \\ 1 & 0 & 0 & 0 & -1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 0 & 0 & -1 & 0 & 1 \\ 1 & 2 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 & 0 & -1 \\ -1 & 0 & 1 & 2 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ -1 \\ 1 \\ 3 \\ 4 \\ 4 \\ 2 \end{bmatrix}$$

[7]

The circular convolution of subsection $x_2(n)$ with zero-padded h(n) yields:

$$\begin{bmatrix} y_2(0) \\ y_2(1) \\ y_2(2) \\ y_2(3) \\ y_2(4) \\ y_2(5) \\ y_2(6) \\ y_2(7) \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 & 0 & -1 & -2 & -1 & 0 \\ 0 & 1 & 2 & 1 & 0 & -1 & -2 & -1 \\ -1 & 0 & 1 & 2 & 1 & 0 & -1 & -2 \\ -2 & -1 & 0 & 1 & 2 & 1 & -0 & -1 \\ -1 & -2 & -1 & 0 & 1 & 2 & 1 & 0 \\ 0 & -1 & -2 & -1 & 0 & 1 & 2 & 1 \\ 1 & 0 & -1 & -2 & -1 & 0 & 1 & 2 \\ 2 & 1 & 0 & -1 & -2 & -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 2 \\ -2 \\ -4 \\ -4 \\ -2 \\ 2 \end{bmatrix}$$

[8]

The circular convolution of subsection $x_3(n)$ with h(n) yields:

$$\begin{bmatrix} y_3(0) \\ y_3(1) \\ y_3(2) \\ y_3(3) \\ y_3(4) \\ y_3(5) \\ y_3(6) \\ y_3(7) \end{bmatrix} = \begin{bmatrix} 0 & -1 & -2 & -1 & 0 & 1 & 2 & 1 \\ 1 & 0 & -1 & -2 & -1 & 0 & 1 & 2 \\ 2 & 1 & 0 & -1 & -2 & -1 & 0 & 1 \\ 1 & 2 & 1 & 0 & -1 & -2 & -1 & 0 \\ 0 & 1 & 2 & 1 & 0 & -1 & -2 & -1 \\ -1 & 0 & 1 & 2 & 1 & 0 & -1 & -2 \\ -2 & -1 & 0 & 1 & 2 & 1 & 0 & -1 \\ -1 & -2 & -1 & 0 & 1 & 2 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -4 \\ -2 \\ 2 \\ 4 \\ 4 \\ 2 \\ -2 \\ -4 \end{bmatrix}$$

[9]

From the output of each circular convolution, the last N – M + 1 = 5 points are saved and the first M – 1 = 3 points are discarded. The saved sequences are concatenated successively, which yields the filtered output:

$$y(n) = y_1^1(n)\ y_2^1(n)\ y_3^1(n)$$

These are denoted as the sequences after the discard:

$$y(n) = y_1^1,\ (n)\ y_2^1,\ (n)\ y_3^1\ (n)$$

$$y(n) = y_1(n),\ y_2(n),\ y_3(n)$$
$$= 1, 3, 4, 4, 2, -2, -4, -4, -2, 2, 4, 4, 2, -2, -4 \quad [10]$$

Continuation of the overlap-discard process results in the indefinitely long filtered sequence given by:

$$y(n) = y_1^1(n),\ y_2^1(n),\ y_3^1(n),\ \ldots\ ,\ y_k^1(n) \quad [11]$$

The application of the fast convolution method with the overlap-discard method in the LH9124 is concluded as follows:

1. X(n) pad (m – 1) with zeros to the first ahead of the sequence.
2. Extend the filter's FIR h(n) from M-point to N-point. For example:

$$h(n) = \begin{cases} h(n), & n = 0, 1, \ldots M-1 \\ 0, & n = M, M+1, \ldots, N-1 \end{cases} \quad [12]$$

   Then perform N-point FFT for the h(n). Store the FFT output in the coefficient RAM for subsequent use.
3. Configure the LH9230 Address Generator for data acquisition and data overlap addressing to enable exact N points to be selected from the input signal sequence based on the following expression:

$$x_m(n) = x\,[n + (m-1)\ (N - M + 1)\,] \quad [13]$$

   Where N is the length of FFT, M is the length of filter impulse response and m is the subsection index for the input segments. Figure 4 shows the sectioning for the input signal sequence in Equation [13].
4. Execute the N-point FFT for the sectioned input segment.
5. Multiply the stored output in the coefficient RAM frequency response of the filter, such as H(k), by the FFT of the input segment obtained in step 3.
6. Execute the N-point IFFT for each pass of the output in step 4.
7. Discard the first M – 1 points from each successive output of step 5 and save the last N – M + 1 points. Append the last N – M + 1 points to y(n). See Figure 5.

LH9124 System Configuration

The overlap-discard method limits the performance of most DSPs. In contrast, the LH9124 accommodates this overlap-discard method by providing special sequences, control signals and a special control register to set the overlap amount. Figure 6 illustrates the data flow of this algorithm in the LH9124.

In general, the LH9124 is set to the recursive dual memory system for the overlap-discard algorithm. For example, the implementation of a recursive dual memory system is shown in Figure 7. The overlapping frames are created by writing the sampled data into the acquisition RAMs sequentially and by reading the data from the RAMs in frames. The frames are advanced by the overlap length for each frame read.

For example, if a 1 kilobyte frame with a 50 percent overlap is configured, the address pointer of this frame in the LH9230 would be advanced by 512 locations for each frame that is read into the LH9124. The discard is done by simply overwriting the discarded portions of the array with the data that follows. The coefficient RAMs store coefficients for windowing, the table and the filter frequency response trigger coefficients.

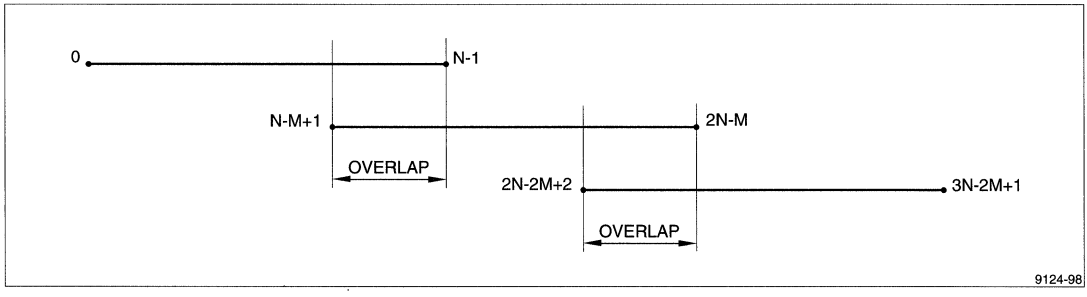


Figure 4. Data Sectioning of the Overlap-Discard Fast Convolution

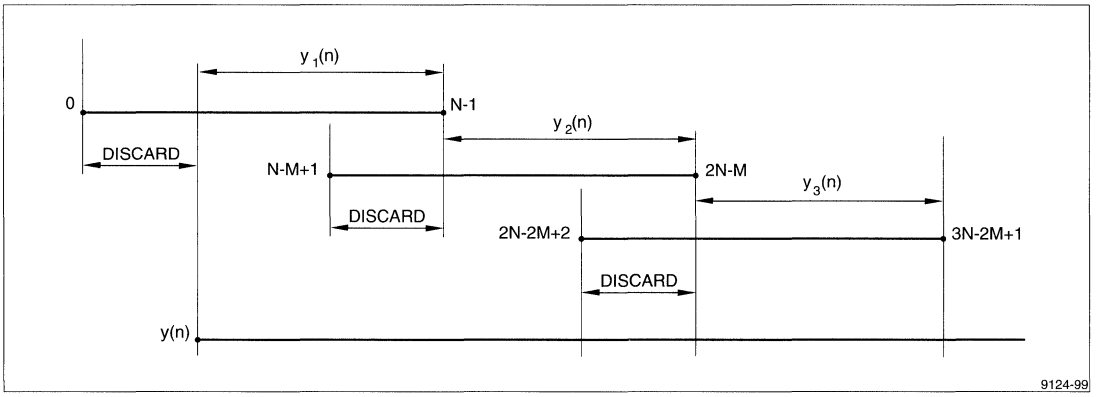


Figure 5. Discarding First M-1 Points from N-Point IFFT

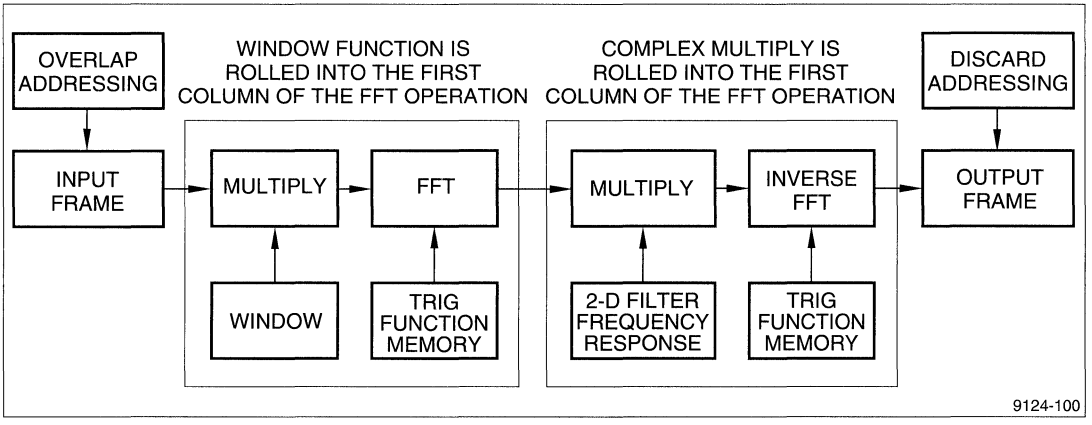


Figure 6. Fast Convolution Data Flow in the LH9124

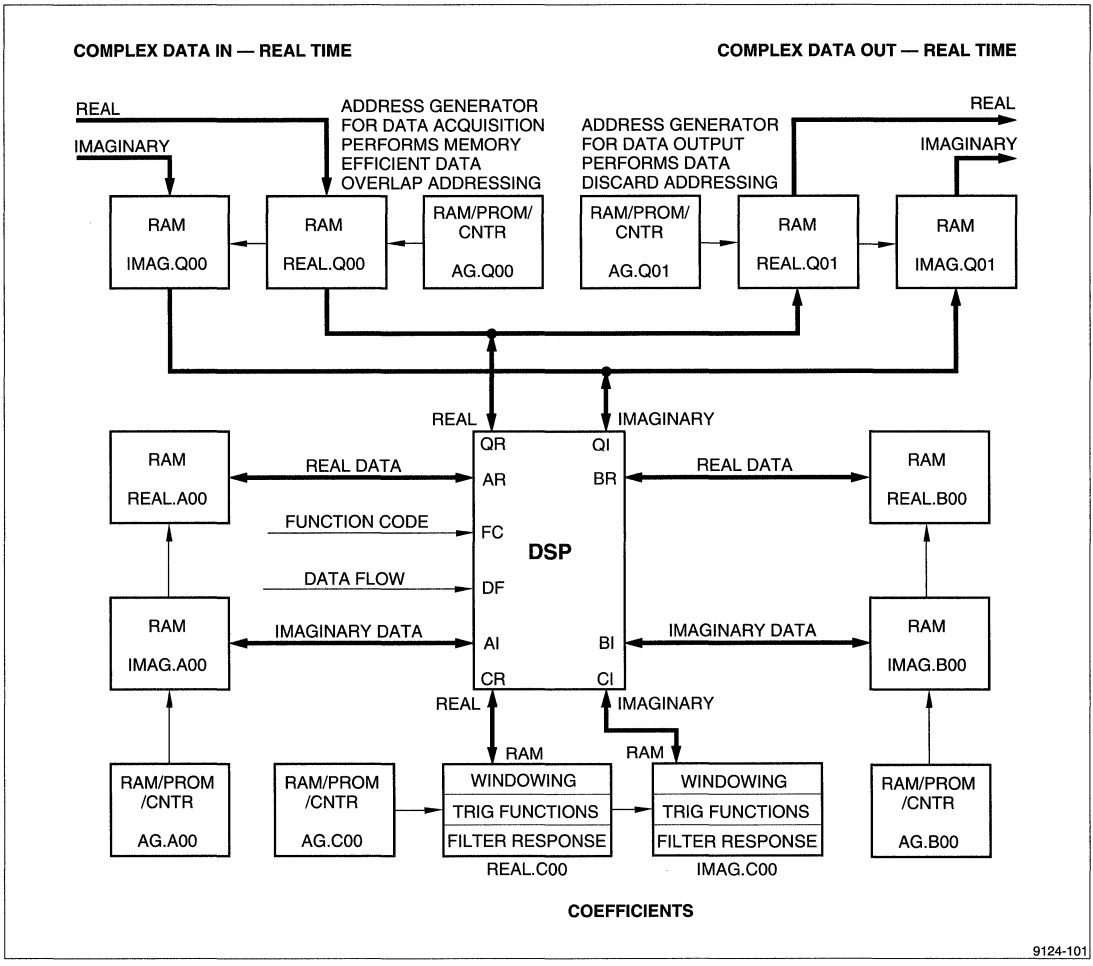


Figure 7. Overlap-Discard Recursive Dual Memory

2-D TIME DOMAIN CONVOLUTION

INTRODUCTION

The field of image processing ranges from simple bit twiddling of individual picture elements (pixels) to sophisticated analysis of picture objects (houses, people, etc.). Most low to mid level real time manipulation is done through the use of dedicated hardware. With a general purpose transform engine, the LH9124, a 24-bit 800 MHz data stream can be filtered to allow digital filtering in real time. For a frame size of 256 × 256 bits, 37.67 frames per second processing is feasible.

Faster rates may be obtained by decreasing the filter size (number of taps). This can be demonstrated by implementing a 2-D high pass finite impulse response (FIR) filter with a general purpose configuration of the LH9124 DSP transform chip.

This publication will present a hardware implementation of a 2-D FIR filter and the program code for FIR filter required for LH9124.

Time Domain Convolution: 2-D

A 2-dimensional (2-D) FIR filter is implemented by applying the theory of a 1-D FIR filter to two dimensions. For two-dimensional time domain convolution, filters can be causal, non-causal, recursive or non-recursive (Figures 1-4). The c and d terms represent the weighting values for the branch in which they are shown. The delay blocks represent a delay in the signal samples by a constant time or space increment.

The equations (equations [1]-[4]) that describe the 2-dimensional versions of these filters are described below.

In general (by eliminating the recursive nature of the digital filter) the filter design and implementation may be simplified while still preserving the convolution relation. This results in filters of Figures 2 and 4. Equations [2] and [4] represent the 2-dimensional versions of these filters.

A causal filter uses information about an event that has not yet occurred to modify a current event. This information can be obtained by making predictions about the future event with knowledge about the repeatability of the event, or the output can be delayed until the future event has occurred.

In the case of image data, an entire frame can be captured before the filtering of that frame begins, introducing a small delay between the input to the filter and the output. Therefore the non-causal, non-recursive filter of Figure 5, equation [4] can be used for digital filtering of image data.

In this Figure 5, the $C_{m1,m2}$ terms represents the filter weighting values, and n_1 and n_2 values represent the spacial coordinates.

Causal – Recursive:

$$y(n_1, n_2) = \sum_{m_1=0}^{M_1} \sum_{m_2=0}^{M_2} c_{m_1, m_2} \cdot x(n_1 - m_1, n_2 - m_2) - \sum_{m_1=0}^{M_1} \sum_{m_2=0}^{M_2} d_{m_1, m_2} \cdot y(n_1 - m_1, n_2 - m_2) \quad [1]$$

Causal – Non-Recursive:

$$y(n_1, n_2) = \sum_{m_1=0}^{M_1} \sum_{m_2=0}^{M_2} c_{m_1, m_2} \cdot x(n_1 - m_1, n_2 - m_2) \quad [2]$$

Non-Causal – Recursive:

$$y(n_1, n_2) = \sum_{m_1=-M_1}^{M_1} \sum_{m_2=-M_2}^{M_2} c_{m_1, m_2} \cdot x(n_1 - m_1, n_2 - m_2) - \sum_{m_1=-M_1}^{M_1} \sum_{m_2=-M_2}^{M_2} d_{m_1, m_2} \cdot y(n_1 - m_1, n_2 - m_2) \quad [3]$$

Non-Causal – Non-Recursive:

$$y(n_1, n_2) = \sum_{m_1=-M_1}^{M_1} \sum_{m_2=-M_2}^{M_2} c_{m_1, m_2} \cdot x(n_1 - m_1, n_2 - m_2)$$

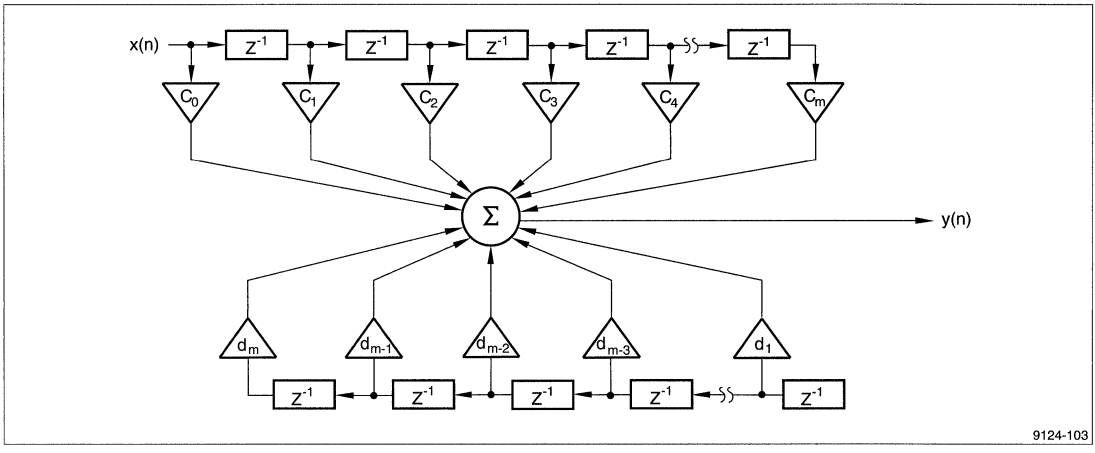


Figure 1. Causal - Recursive One-Dimensional Time Domain Convolution Filter

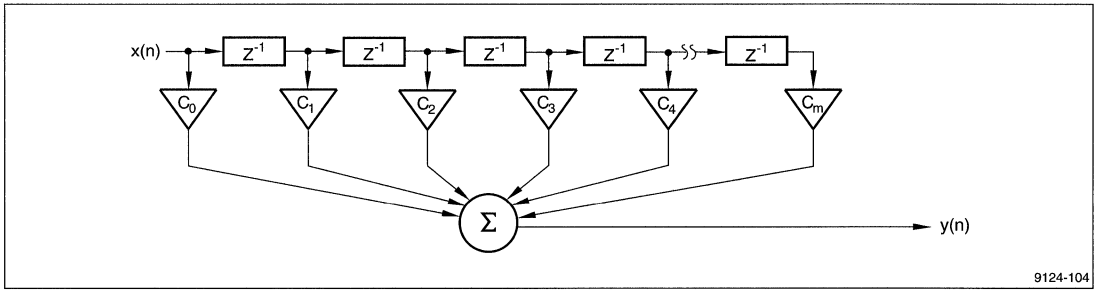


Figure 2. Causal - Non-Recursive One-Dimensional Time Domain Convolution Filter

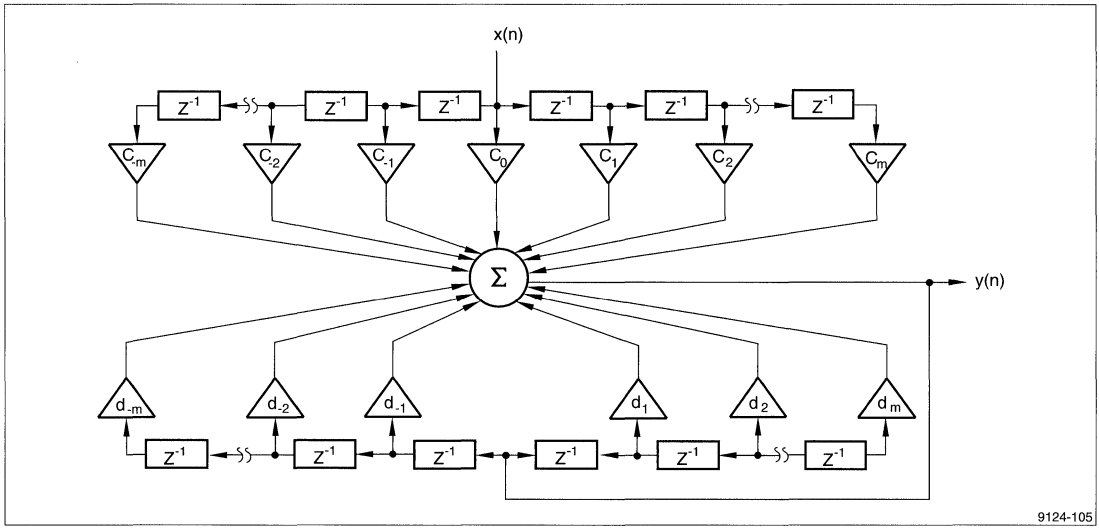


Figure 3. Non-Causal - Recursive One-Dimensional Time Domain Convolution Filter

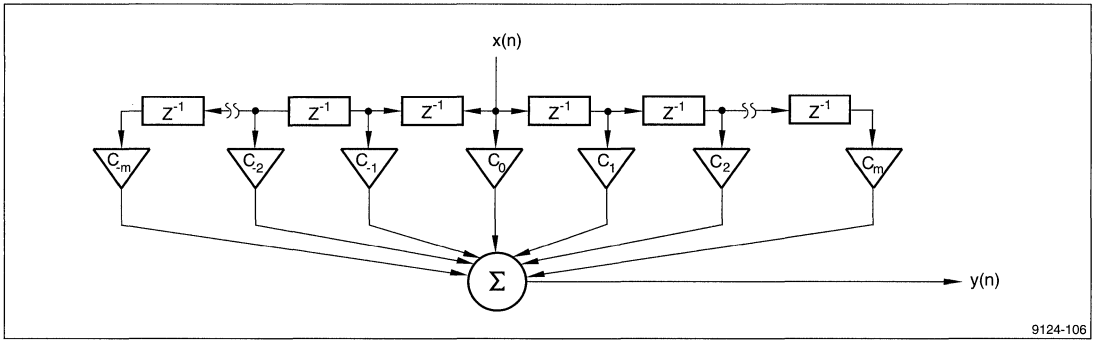


Figure 4. Non-Causal - Non-Recursive One-Dimensional Time Domain Convolution Filter

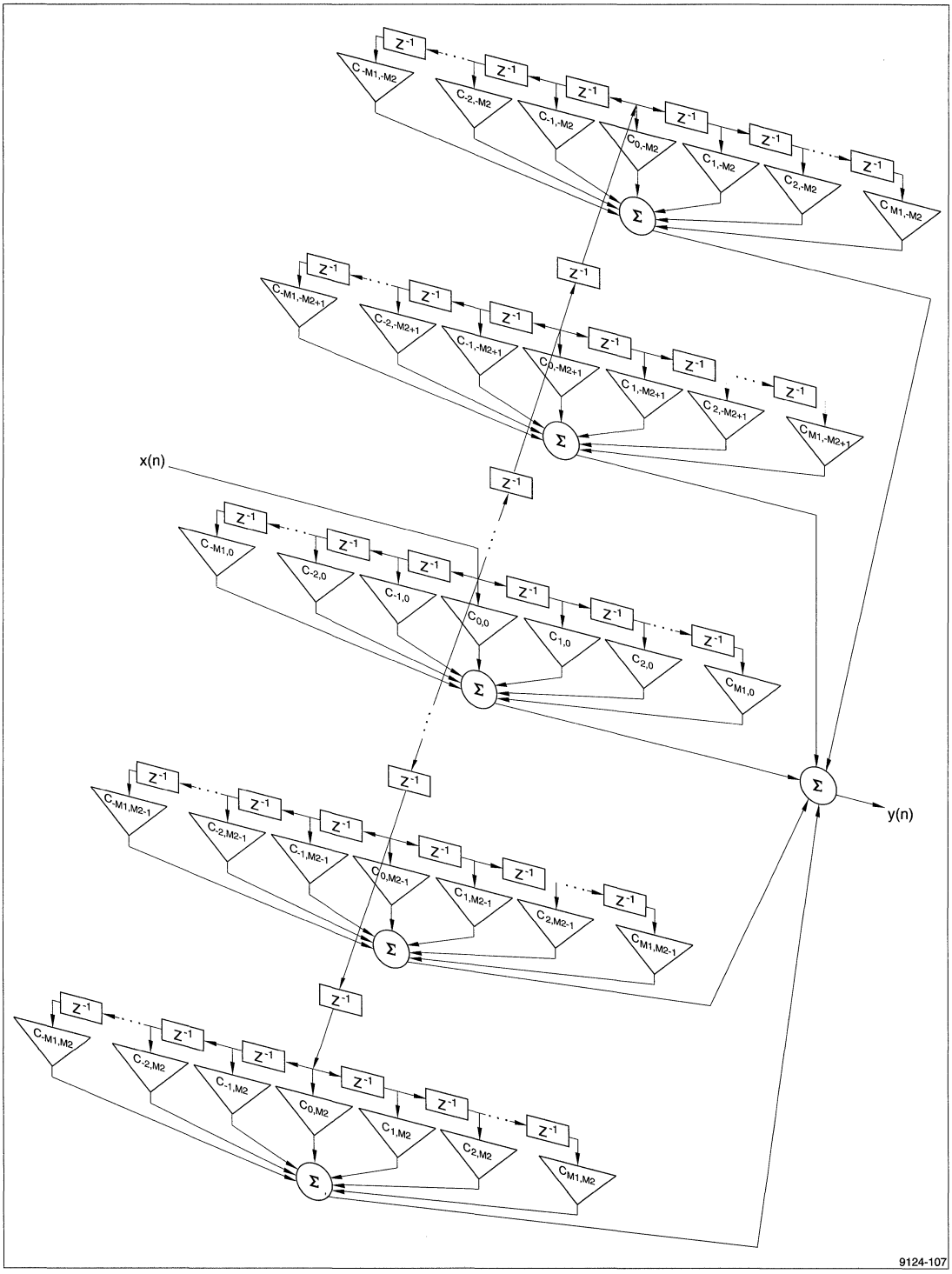


Figure 5. Non-Causal - Non-Recursive Two-Dimensional Time Domain Convolution Filter

Determining the Coefficients

By assuming that the input to the system represented by equation [4] is a complex sinusoid:

$$x(n_1, n_2) = A\, e^{j2\pi f_1 n_1}\, e^{j2\pi f_2 n_2} \quad [5]$$

where:

$$n_1 = \frac{x}{X\,max}\ , x = 0 \ldots X\,max$$

$$n_2 = \frac{y}{Y\,max}\ , y = 0 \ldots Y\,max$$

$$f_1 = 0 \ldots \frac{X\,max}{2}$$

$$f_2 = 0 \ldots \frac{Y\,max}{2}$$

Equation [4] becomes:

$$y(n_1, n_2) = \sum_{m_1=-M_1}^{M_1} \sum_{m_2=-M_2}^{M_2} c_{m_1, m_2} A e^{j2\pi f_1 (n_1 - m_1)} e^{j2\pi f_2 (n_2 - m_2)}$$

$$= A e^{j2\pi f_1 n_1} e^{j2\pi f_2 n_2} \sum_{m_1=-M_1}^{M_1} \sum_{m_2=-M_2}^{M_2} c_{m_1, m_2} e^{-j2\pi f_1 m_1} e^{-j2\pi f_2 m_2}$$

$$= x(n_1, n_2) \sum_{m_1=-M_1}^{M_1} \sum_{m_2=-M_2}^{M_2} c_{m_1, m_2} e^{-j2\pi f_1 m_1} e^{-j2\pi f_2 m_2} \quad [6]$$

The frequency response term for $y(n_1,n_2)$ is:

$$H(e^{j2\pi f_1}, e^{j2\pi f_2}) = \sum_{m_1=-M_1}^{M_1} \sum_{m_2=-M_2}^{M_2} c_{m_1, m_2} e^{-j2\pi f_1 m_1} e^{-j2\pi f_2 m_2} \quad [7]$$

Substituting a desired frequency response $A(f_1,f_2)$ for $H(e^{j2\pi f_1}, e^{j2\pi f_2})$:

$$A(f_1, f_2) = \sum_{m_1=-M_1}^{M_1} \sum_{m_2=-M_2}^{M_2} c_{m_1, m_2} e^{-j2\pi f_1 m_1} e^{-j2\pi f_2 m_2} \quad [8]$$

If we assume that the image data (input) is circularly symmetric, $c_{m1,m2} = c_{-m1,m2} = c_{-m1,-m2} = c_{m1,-m2}$:

$$A(f_1, f_2) = \sum_{m_1=-M_1}^{M_1} \sum_{m_2=-M_2}^{M_2} c_{m_1, m_2} e^{j2\pi f_1 m_1} e^{j2\pi f_2 m_2}$$

$$e^{-j2\pi f_1 n_1} e^{-j2\pi f_2 n_2} A(f_1, f_2) = \sum_{m_1=-M_1}^{M_1} \sum_{m_2=-M_2}^{M_2} c_{m_1, m_2}$$

$$e^{j2\pi f_1 m_1} e^{j2\pi f_2 m_2} e^{-j2\pi f_2 n_1} e^{-j2\pi f_2 n_2} \quad [9]$$

The maximum frequency of the data, as determined by Nyquist, is $\frac{f_{s1}}{2} = \frac{X_{max}}{2}$ for the horizontal data, and $\frac{f_{s2}}{2} = \frac{Y_{max}}{2}$ for the vertical data. Then integrating over these maximum frequency ranges:

$$\int_0^{\frac{f_{s1}}{2}} \int_0^{\frac{f_{s2}}{2}} A(f_1,f_2) e^{-j2\pi f_1 n_1} e^{-j2\pi f_2 n_2}\, df_1 df_2 = \int_0^{\frac{f_{s1}}{2}} \int_0^{\frac{f_{s2}}{2}} \sum_{m_1=-M_1}^{M_1}$$

$$\sum_{m_2=-M_2}^{M_2} c_{m_1, m_2} e^{j2\pi f_1 (m_1 - n_1)} e^{j2\pi f_2 (m_2 - n_2)}\, df_1 df_2$$

$$= \sum_{m_1=-M_1}^{M_1} \sum_{m_2=-M_2}^{M_2} c_{m_1, m_2} \int_0^{\frac{f_{s1}}{2}} \int_0^{\frac{f_{s2}}{2}} e^{j2\pi f_1 (m_1 - n_1)}$$

$$e^{j2\pi f_2 (m_2 - n_2)}\, df_1 df_2 \quad [10]$$

By using the Euler formula, the right hand side becomes:

$$\sum_{m_2=-M_2}^{M_2} \int_0^{\frac{f_{s2}}{2}} e^{j2\pi f_2 (m_2 - n_2)} \sum_{m_1=-M_1}^{M_1} c_{m_1, m_2} \cdot$$

$$\left(\int_0^{\frac{f_{s1}}{2}} \cos((m_1 - n_1) 2\pi f_1) df_1 + j \int_0^{\frac{f_{s1}}{2}} \sin((m_1 - n_1) 2\pi f_1) df_1 \right) df_2 \quad [11]$$

Since the integration is over an integer multiple of the sinusoids period, the integral evaluates to zero (area under a sinusoid), except for the case of n1=m1, where it evaluates to $\frac{f_{s1}}{2}$.

$$\int_0^{\frac{f_{s1}}{2}} e^{j(m_1 - n_1) 2\pi f_1}\, df_1 = \begin{cases} \frac{f_{s1}}{2}, & n_1 = m_1 \\ 0, & n_1 \neq m_1 \end{cases}$$

and:

$$\int_0^{\frac{f_{s2}}{2}} e^{j(m_2 - n_2) 2\pi f_2}\, df_2 = \begin{cases} \frac{f_{s2}}{2}, & n_2 = m_2 \\ 0, & n_2 \neq m_2 \end{cases}$$

Therefore the right hand side of equation [11] reduces to:

$$\sum_{m_2=-M_2}^{M_2} \int_0^{\frac{f_{s2}}{2}} e^{j2\pi f_2 (m_2 - n_2)} c_{n_1, m_2} \cdot \frac{f_{s1}}{2} df_2$$

and:

$$\frac{f_{s1}}{2}\cdot\sum_{m_2=-M_2}^{M_2}\int_0^{\frac{f_{s2}}{2}} e^{j2\pi f_2(m_2-n_2)}\,df_2\,c_{n_1,m_2}=\frac{f_{s1}f_{s2}}{4}\,c_{n_1,n_2}$$

By substituting this (right hand side) back into equation [10]:

$$\int_0^{\frac{f_{s1}}{2}}\int_0^{\frac{f_{s2}}{2}} A(f_1,f_2)\,e^{-j2\pi f_1 n_1}\,e^{-j2\pi f_2 n_2}\,df_1df_2=\frac{f_{s1}f_{s2}}{4}\,c_{n_1,n_2} \qquad [12]$$

Finally:

$$c_{n_1,n_2}=\frac{4}{f_{s1}f_{s2}}\int_0^{\frac{f_{s1}}{2}}\int_0^{\frac{f_{s2}}{2}} A(f_1,f_2)\,e^{-j2\pi f_1 n_1}e^{-j2\pi f_2 n_2}\,df_1df_2 \qquad [13]$$

Since the integral is over one period of a sinusoid, the integral boundaries can be changed to different endpoints of the same period.

$$c_{n_1,n_2}=\frac{4}{f_{s1}f_{s2}}\int_{\frac{-f_{s2}}{4}}^{\frac{f_{s2}}{4}}\int_{\frac{-f_{s1}}{4}}^{\frac{f_{s1}}{4}} A(f_1,f_2)\,e^{-j2\pi f_1 n_1}e^{-j2\pi f_2 n_2}\,df_1df_2 \qquad [14]$$

In order to simplify this further, the relationships $A_{f1,f2} = A_{-f1,f2} = A_{-f1,-f2} = A_{f1,-f2}$, from the assumption of a circularly symmetric impulse response, can be used. The coefficient equation then becomes:

$$c_{n_1,n_2}=\frac{4}{f_{s1}f_{s2}}\int_{\frac{-f_{s2}}{4}}^{\frac{f_{s2}}{4}}\int_{\frac{-f_{s1}}{4}}^{\frac{f_{s1}}{4}} A(f_1,f_2)\cos(2\pi f_1 n_1)\cdot\cos(2\pi f_2 n_2)\,df_1df_2 \qquad [15]$$

If the first sampling frequency, f_{s1}, is even:

$$c_{n_1,n_2}=\frac{4}{f_{s1}f_{s2}}\int_{\frac{-f_{s2}}{4}}^{\frac{f_{s2}}{4}}2\int_0^{\frac{f_{s1}}{4}} A(f_1,f_2)\cos(2\pi f_1 n_1)\cdot\cos(2\pi f_2 n_2)\,df_1df_2 \qquad [16]$$

And, if the second sampling frequency, f_{s2}, is even:

$$c_{n_1,n_2}=\frac{4\cdot 2\cdot 2}{f_{s1}f_{s2}}\int_0^{\frac{f_{s2}}{4}}\int_0^{\frac{f_{s1}}{4}} A(f_1,f_2)\cos(2\pi f_1 n_1)\cdot\cos(2\pi f_2 n_2)\,df_1df_2 \qquad [17]$$

At this point, if having a filter with a non-causal response, or output before input, is not acceptable, then further modifications of the final equations would be made, for example, time shifting the impulse response to make it causal. However, for 2-D time domain filtering of digitized images, non-causal filtering is acceptable, and the following two formulas can be used together for the design and implementation of a 2-D FIR filter.

$$c_{n_1,n_2}=\frac{16}{f_{s1}f_{s2}}\int_0^{\frac{f_{s2}}{4}}\int_0^{\frac{f_{s1}}{4}} A(f_1,f_2)\cos(2\pi f_1 n_1)\cdot\cos(2\pi f_2 n_2)\,df_1df_2 \qquad [18]$$

To find the filter coefficients, and:

$$y_{n_1,n_2}=\sum_{m_1=-M_1}^{M_1}\sum_{m_2=-M_2}^{M_2} h(m_1,m_2)\cdot x(n_1-m_1,n_2-m_2)$$

$$=\sum_{m_1=-M_1}^{M_1}\sum_{m_2=-M_2}^{M_2} c_{m_1,m_2}\cdot x(n_1-m_1,n_2-m_2) \qquad [19]$$

To produce the filtered output from the filter coefficients and the input.

An Example of a Digital Non-Causal Non-Recursive Filter Using an Image as Input

Choosing a desired frequency response, A(f1, f2), to be a high pass step function, with a threshold, $\theta = 90\%\frac{f_{s1}}{2} = 90\%\frac{f_{s2}}{2}$, produces:

$$A(f_1,f_2)=\begin{cases}0\ , 0\le f_1\le\theta,\ 0\le f_2\le\theta\\ 1\ ,\ \theta\le f_1\le\frac{f_{s1}}{2},\ \theta\le f_2\le\frac{f_{s2}}{2}\end{cases}$$

Since this is circularly symmetric, equation [18] may be used to find the coefficients.

$$c_{n_1,n_2}=\frac{16}{f_{s1}f_{s2}}\int_{\theta-\frac{f_{s2}}{4}}^{\frac{f_{s2}}{4}}\int_{\theta-\frac{f_{s1}}{4}}^{\frac{f_{s1}}{4}} 1\cdot\cos(2\pi f_1 n_1)\cdot\cos(2\pi f_2 n_2)\,df_1df_2$$

Where the lower limit is due to the limit shift of equation [14]. After integrating:

$$c_{n_1,n_2}=\frac{16}{f_{s1}f_{s2}}\left(\frac{\sin(2\pi\frac{f_{s1}}{4}n_1)-\sin(2\pi(\theta-\frac{f_{s1}}{4})n_1)}{2\pi n_1}\right)\cdot\left(\frac{\sin(2\pi\frac{f_{s2}}{4}n_2)-\sin(2\pi(\theta-\frac{f_{s2}}{4})n_2)}{2\pi n_2}\right) \qquad [20]$$

However, this equation is valid only for $n_1, n_2 \neq 0$. When either variable equals 0, the equation fails. Using La Hospital's rule, integrate the top and bottom of the 0/0 term.

For the case of $n_1=0$:

$$c_{n_1,n_2} = \frac{16}{f_{s1}f_{s2}}\left(\frac{\sin(2\pi\frac{f_{s2}}{4}n_2)-\sin(2\pi(\theta-\frac{f_{s2}}{4})n_2)}{2\pi n_2}\right)(-\theta) \quad [21]$$

For the case of $n_2=0$:

$$c_{n_1,n_2} = \frac{16}{f_{s1}f_{s2}}\left(\frac{\sin(2\pi\frac{f_{s1}}{4}n_1)-\sin(2\pi(\theta-\frac{f_{s1}}{4})n_1)}{2\pi n_1}\right)(-\theta) \quad [22]$$

For the case of $n_1=n_2=0$:

$$c_{n_1,n_2} = \frac{16}{f_{s1}f_{s2}}\cdot\theta^2 \quad [23]$$

By using equations [20], [21], [22], and [23], the coefficients can now be generated. For example, assuming the image is $256 \times 256 \times 8$ bits, $f_{s1} = f_{s2} = 256$, the coefficient $c_{-4,-4}$ can be calculated from equation [20], as shown in Example 1.

Table 1 presents the coefficients $c_{-4,-4}$ through $c_{4,4}$ for a 90% filter.

Figures 6 and 7 demonstrate the effect of 2-D FIR filter. The images in Figure 6 were passed through a 98% 2-D FIR filter. The filter rejected all low frequency components, passing only the frequencies in the upper 2% of the normalized range.

SUMMARY

Figure 8 demonstrates a system implementation of a 80 MHz data rate, 24-bit data, 2-D filter using the LH9124 Digital Signal Processor. In this configuration, two 24-bit (real) data lanes are input to the LH9124. At a clock rate of 40 MHz, an 80 MHz data rate is achieved. The performance of the system can be calculated using the following equation:

$$\frac{\text{clock rate} \cdot \text{pixels per clock cycle input}}{\text{pixels input per filtered pixels output} \cdot \text{pixels per frame}}$$

$$= \frac{40\ \text{MHz} \cdot 2}{81 \cdot 256 \cdot 256} = 15.07\ {}^{\text{frames}}/_{\text{second}}$$

A system implemented with an LH9124 can filter 256×256 frames of 24 bit real data in a 2-D time domain filter application in real time. This preserves the original signal to noise ratio and provides 60 bits of precision. To filter larger frames (e.g. $1K \times 1K$) at higher performance rates, the LH9124 can be used to implement 2-D frequency domain filtering, refer to application note # SMT89016.

$$c_{-4,-4} = \frac{16}{256\cdot256}\left(\frac{\sin\left(2\pi\frac{256(-4)}{4\ 256}\right)-\sin\left(2\pi\left(0.9\cdot\frac{256}{2}-\frac{256}{4}\right)\frac{(-4)}{256}\right)}{2\pi\frac{-4}{256}}\right)\cdot\left(\frac{\sin\left(2\pi\frac{256(-4)}{4\ 256}\right)-\sin\left(2\pi\left(0.9\cdot\frac{256}{2}-\frac{256}{4}\right)\frac{(-4)}{256}\right)}{2\pi\frac{-4}{256}}\right)$$

$$c_{-4,-4} = 0.0232$$

Example 1. $c_{-4,-4}$ for a $256 \times 256 \times 8$ Image, $M_1 = M_2 = 4$

Table 1. Coefficients for an Image $256 \times 256 \times 8$. $M_1 = M_2 = 4$

$c_{n1,n2}$	n1 = -4	n1 = -3	n1 = -2	n1 = -1	n1 = 0	n1 = 1	n1 = 2	n1 = 3	n1 = 4
n2 = -4	0.0232	-0.0137	-0.0289	0.0049	-0.1214	0.0049	-0.0289	-0.0137	0.0232
n2 = -3	-0.0137	0.0081	0.0171	-0.0029	0.0717	-0.0029	0.0171	0.0081	-0.0137
n2 = -2	-0.0289	0.0171	0.0360	-0.0061	0.1511	-0.0061	0.0360	0.0171	-0.0289
n2 = -1	0.0049	-0.0029	-0.0061	0.0010	-0.0256	0.0010	-0.0061	-0.0029	0.0049
n2 = 0	-0.1214	0.0717	0.1511	-0.0256	0.413	-0.0256	0.1511	0.0717	-0.1214
n2 = 1	0.0049	-0.0029	-0.0061	0.0010	-0.0256	0.0010	-0.0061	-0.0029	0.0049
n2 = 2	-0.0289	0.0171	0.0360	-0.0061	0.1511	-0.0061	0.0360	0.0171	-0.0289
n2 = 3	-0.0137	0.0081	0.0171	-0.0029	0.0717	-0.0029	0.0171	0.0081	-0.0137
n2 = 4	0.0232	-0.0137	-0.0289	0.0049	-0.1214	0.0049	-0.0289	-0.0137	0.0232

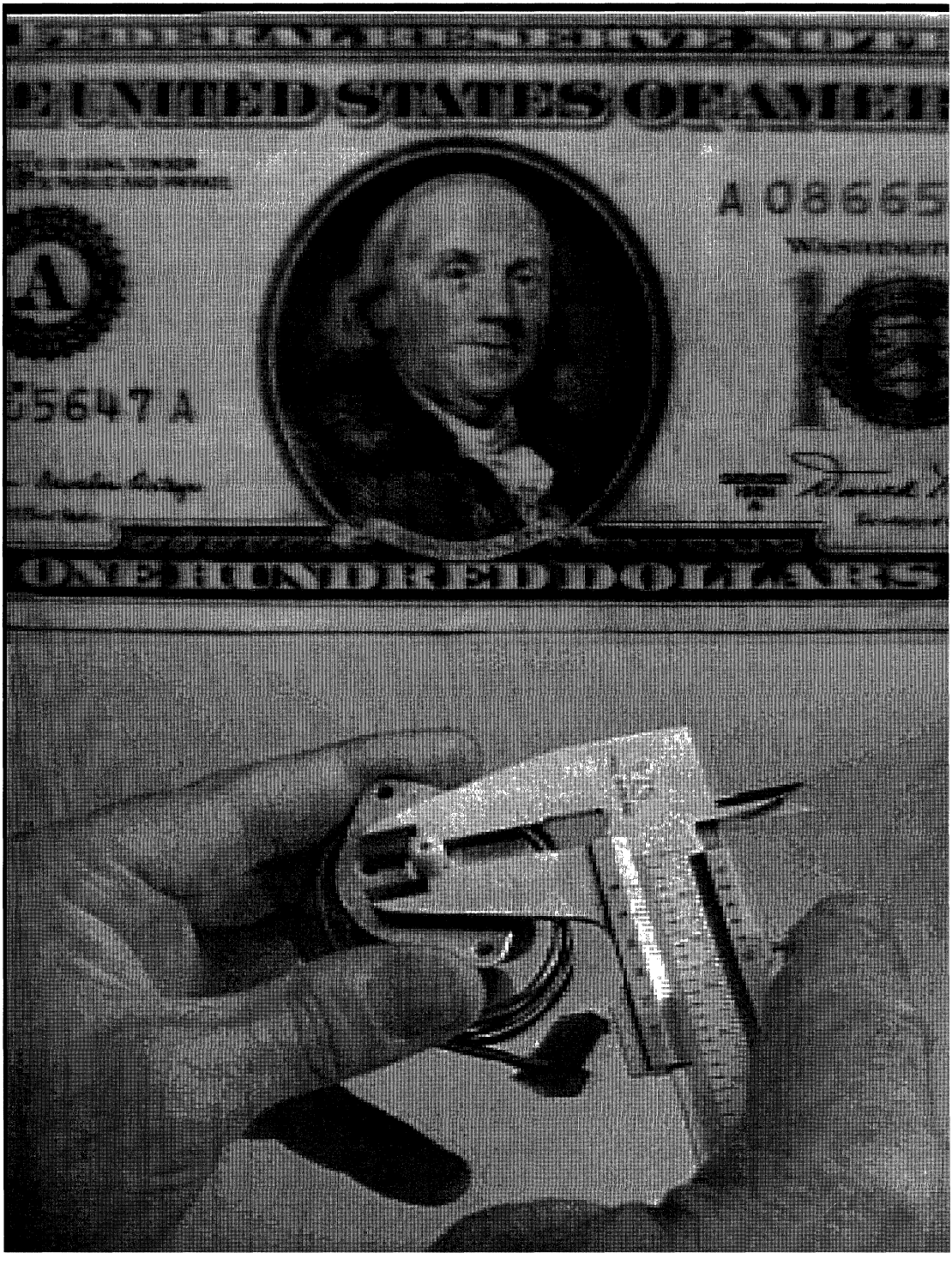


Figure 6. Input to 2-D Filter:
Raw Images Stored as 256 × 256 × 8

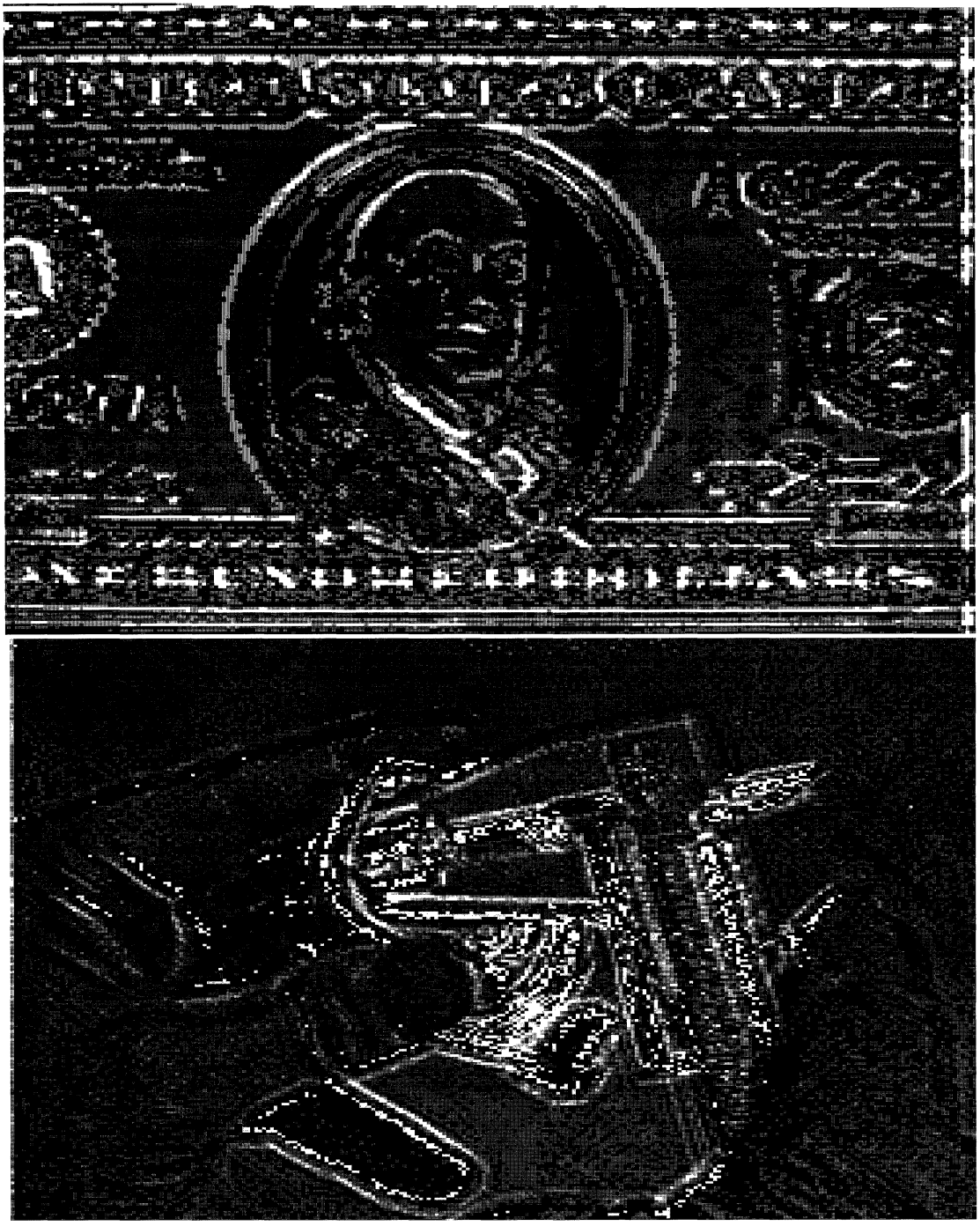


Figure 7. Output From 2-D Filter:

$$\theta = 98\% \frac{f_{s1}}{2} = 98\% \frac{f_{s2}}{2}, \text{ where } A(f_1,f_2) = \begin{cases} 0 & , 0 \le f_1 \le \theta, 0 \le f_2 \le \theta \\ 1 & , \theta \le f_1 \le \frac{f_{s1}}{2}, \theta \le f_2 \le \frac{f_{s2}}{2} \end{cases}$$

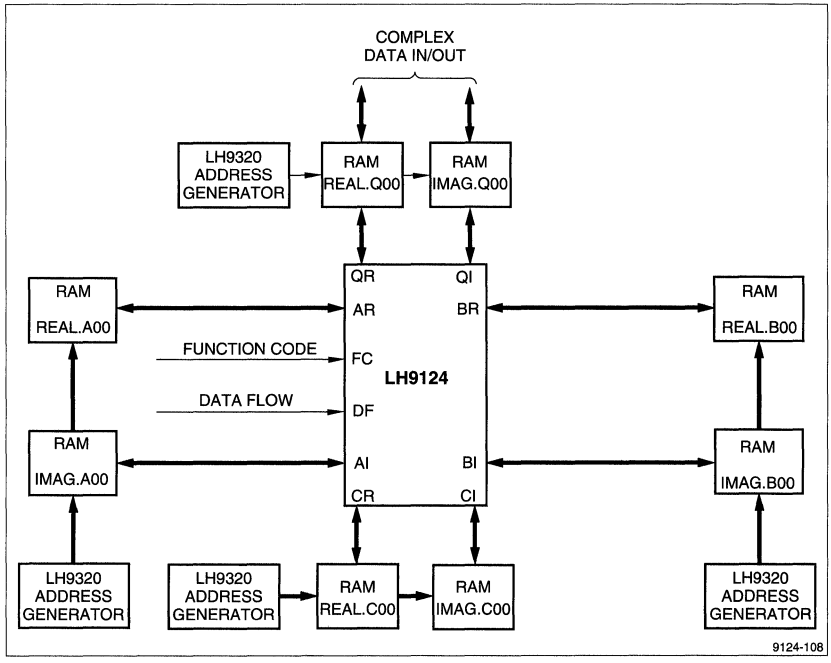


Figure 8. Hardware System Implementation of 200 MHz Data Rate, 24-Bit Data, 2-D Filter Using the LH9124

TIME DOMAIN CONVOLUTION

INTRODUCTION

A Time Domain Convolution is used to perform filtering for both real and complex filter impulse response (FIR) filters. There are two types of Time Domain Convolutions: recursive and non-recursive. Each is described as follows.

Recursive System

The general form of a recursive system is shown in Figure 1. Here x(n) represents a sampled input and y(n) represents a sampled output. The b_i and a_i terms represent weighting values for the respective branches. The delay is used to block the signal samples by a constant time increment.

$$y(n) = \sum_{i=0}^{m} b_i\, x(n-i) - \sum_{i=0}^{m} a_i\, y(n-i) \qquad [1]$$

Equation [1] describes a convolutional relationship between the input and the output. In this case, the system is also recursive because the current output y(n) is dependent on the previous outputs. If a simplification is made, the recursive system can be eliminated, with the convolutional relationship preserved.

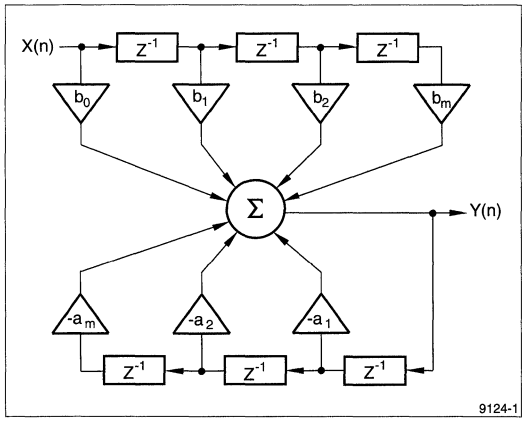


Figure 1. Recursive m^{th} Order Digital Filter

Nonrecursive System

In a nonrecursive system, the output depends on previous input values. A recursive system is a filter since it discriminates against certain frequencies and not others. A nonrecursive filter is also called a finite impulse response (FIR) digital filter. An FIR is stable and can be made to have exact linear phase characteristics as shown in Equation [2].

In this case, the frequency response of the filter equation depends on the values of b_i and m as shown:

$$y(n) = \sum_{i=0}^{m} b_i\, x(n-i) \qquad [2]$$

If the input is a sampled complex sinusoid, then:

$$x(n) = Ae^{j2\pi fn}$$

Consequently, the equation [2] becomes:

$$y(n) = \sum_{i=0}^{m} b_i A e^{j2\pi f(n-i)} \qquad [3]$$

This can be rewritten as:

$$y(n) = Ae^{j2\pi f} \sum_{i=0}^{m} b_i\, e^{-j2\pi fi} \qquad [4]$$

If so, the frequency response term for y(n) is now:

$$\sum_{i=0}^{m} b_i e^{-j2\pi fi} \qquad [5]$$

Figure 2 illustrates a system which the impulse response is causal and finite for finite m. The Fourier transform of the impulse response is:

$$H(e^{j2\pi f}) = \sum_{i=0}^{m} b_i e^{-j2\pi fi} \qquad [6]$$

The Fourier transform is periodic in the frequency domain with period 2π. The amplitude response of the Fourier transform is now:

$$\left| \sum_{i=0}^{m} b_i e^{-j2\pi fi} \right| \qquad [7]$$

With the proper choice of b_i and m, the amplitude response can be a bandpass, lowpass or highpass filter.

The realization of Figure 2 uses a method whereby the b_i values can be stored in a programmable random access memory. The filter characteristics can be modified for real time adaptive filtering programmatically, rather than via the hardware.

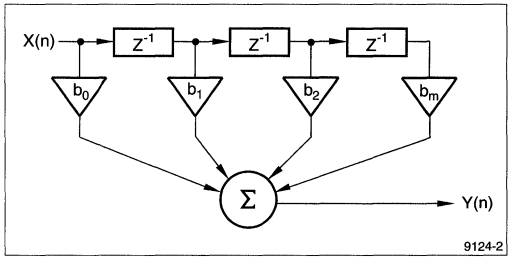


Figure 2. Nonrecursive m^{th} Order Digital Filter

Example of a Digital Nonrecursive Filter

Figure 3 illustrates the simple three branch nonrecursive filter. The input x(n) is a sampled signal composed of two sine waves:

$$\text{COS } \frac{\pi t}{4} \text{ and COS } \frac{3\pi t}{4}$$

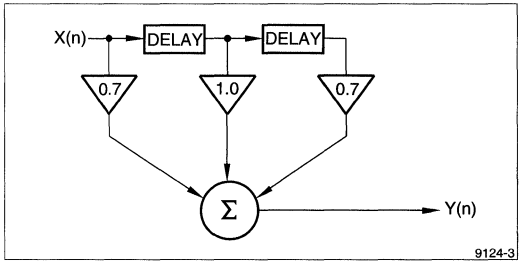


Figure 3. Three Stage Branch Nonrecursive Digital Filter

Figures 4(a), 4(b), and 4(c) graph x(n) and its components x(n) when expressed as:

$$x(n) = \text{COS } (\frac{3\pi n}{4}) + \text{COS } (\frac{\pi n}{4}) \quad [8]$$

Assuming $b_0 = 0.7$, $b_1 = 1.0$, $b_2 = 0.7$.

Using the expression for y(n):

$$y(n) = .7x(n) + 1\ x(n-1) + .7x(n-2) \quad [9]$$

The output sum of the two cosine waves may be calculated by applying the point by point values of Figure 4c to Equation [2] above, with the following results:

$$y(0) = .7 \times 2 + 1 \times 0 + .7 \times 0 = 1.4$$
$$y(1) = .7 \times 0 + 1 \times 2 + .7 \times 0 = 2.0$$
$$y(2) = .7 \times 0 + 1 \times 0 + .7 \times 2 = 1.4$$
$$y(3) = .7 \times 0 + 1 \times 0 + .7 \times 0 = 0$$
$$y(4) = .7 \times (-2) + 1 \times 0 + .7 \times 0 = -1.4$$
$$y(5) = .7 \times 0 + 1 \times (-2) + .7 \times 0 = -2.0$$
$$y(6) = .7 \times 0 + 1 \times 0 + .7 \times (-2) = -1.4$$
$$y(7) = .7 \times 0 + 1 \times 0 + .7 \times 0 = 0$$
$$y(8) = .7 \times 2 + 1 \times 0 + .7 \times 0 = 1.4 \quad [10]$$

Plotting the values of y(n) for the time domain output of the filter results in the waveform in Figure 5.

The output can be expressed as:

$$y(n) = 2 \text{ COS } (\frac{\pi n}{4} - \frac{\pi}{4}) \quad \text{or} \quad [11]$$

$$y(n) = 2 \text{ COS } ((n-1)\frac{\pi}{4}) \quad [12]$$

Equation [12] describes the effects of the filter. This equation has rejected the $\text{COS}\frac{3\pi n}{4}$ term and passed the $\text{COS}\left(\frac{\pi n}{4}\right)$ term. Compare Equations [8] and [12]. Further, the equation has a gain of two and has shifted the phase of the COS(n) term by $\frac{\pi}{4}$.

Figure 6 illustrates a hardware system using the LH9124 to implement a FIR filter. The sampled data is stored in a circular buffer.

For an n point FIR filter, the acquisition address generator provides addressing to the RAMs sequentially to access the sampled data from the 1^{st} sampled point to the n^{th} sampled point for each pass. Each sample acquired becomes the 1^{st} point and is placed on top of the queue dropping the previous n^{th} sample and initiating the next pass.

The coefficient address generator is nothing more than a circular counter with length n. For a FIR filter, the control inputs to the LH9124 are hardwired.

Figure 6 identifies the static states of the inputs using the acquisition port as the input port and the B port as the output port.

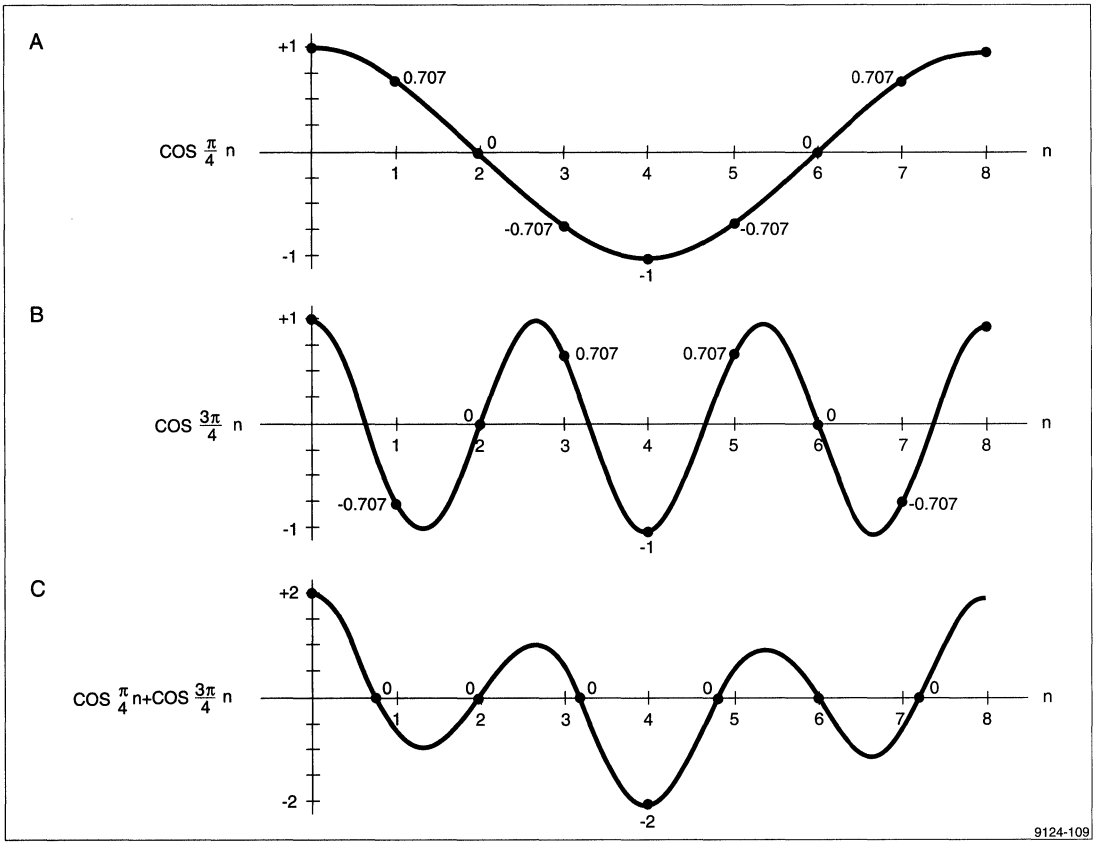


Figure 4. Digital Nonrecursive Filter Example

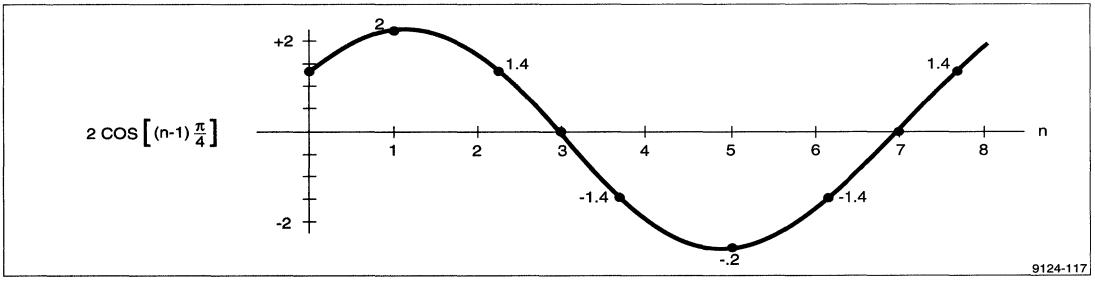


Figure 5. Filter Output: 2 COS $((n-1)\frac{\pi}{4})$

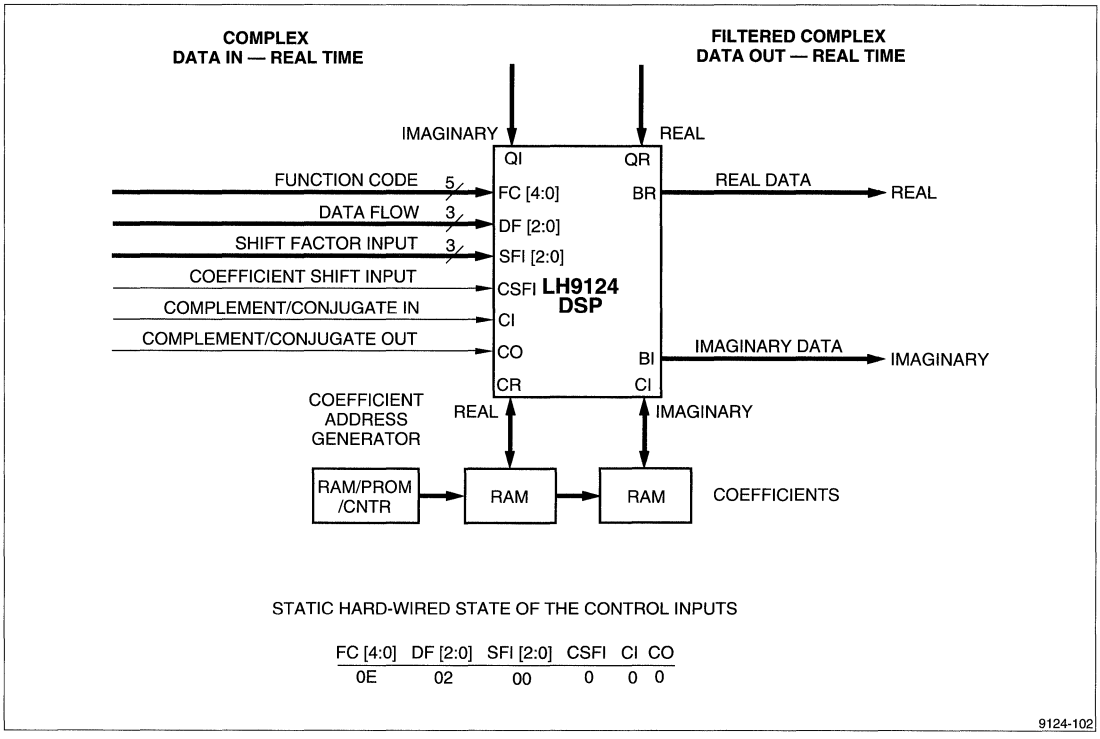


Figure 6. LH9124 System for FIR Filter Implementation

REFERENCES

The references for this application note are as follows:

[1] A. V. Oppenheim, R. W. Schafer, *Digital Signal Processing*. Prentice-Hall, Inc., 1975.

[2] A. K. Jaim, *Fundamentals of Digital Image Processing.*. Prentice-Hall, 1989.

[3] David J. DeFutta, Joseph G. Lucas, William S. Hodkiss, *Digital Signal Processing: A System Design Approach*. John Wiley & Sons, 1988.

[4] L. K. Rubiner, B. Gold, *Theory and Application of Digital Signal Processing*. Prentice-Hall, 1975.

[5] R. C. Gonzales, P. Wintz, *Digital Image Processing*. Addison-Wesly Publishing Co., 1987.

[6] R. W. Heinming. *Digital Filters*, 3rd edition. Prentice-Hall, 1988.

IMAGE COMPRESSION

INTRODUCTION

When images are digitized, the amount of required storage can become astronomical as the number of pictures grow. For example, LANDSAT satellites send images that actually consist of four digital images. In this case, each image represents the same area of land through different spectral windows.

The size of each image is 2340 × 3234 pixels, with each pixel being a 7-bit binary word corresponding to one of 128 different grey levels. Since the satellites send about 30 pictures in a day, the storage needed to record them is (2340)(3234)(7)(4)(30) ~ 6 Giga bits/day ~ 800 Meg bytes/day.

Image compression techniques allow this number to be halved while still allowing the image to be reconstructed error free. Other types of compression can reduce the storage requirements to as low as 0.5 – 0.3 bits per pixel. These compression schemes don't produce exact reproductions, but the results are visually acceptable.

These compression schemes usually use some type of function to transform the image into a domain consisting of orthogonal data types. Since images usually have high correlation between pixels, map onto a small region of the map space. Because of this, the picture can be encoded with fewer bits, due to the correlated data lumping into particular regions of the map space, and leaving the rest zero.

The optimal transform for this mapping of data onto a maximum set of uncorrelated eigenvectors, has been proved to be the Karhunen-Love transform. However, this transform has no fast method of being implemented and is rarely used. The Discrete Cosine Transform (DCT), on the other hand, produces results that closely approach the optimum. Plus, it can be implemented using a Fast Fourier Transform, thus giving it Order(NlogN) operations.

The Discrete Cosine Transform

The Discrete Cosine Transform, defined by [1] and [2]:

$$C(0) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} f_{(k)} \qquad [1]$$

$$C(n) = \sqrt{\frac{2}{N}} \sum_{k=0}^{N-1} f_{(x)} \cos\frac{2\pi nk}{N} \qquad [2]$$

can easily be found by making f(x) into an even function, $h_e(k)$, and then taking the Fourier transform to produce $H_e(k)$.

In doing this, let $h_e(k)$ be an even function of f(x), created by mirroring f(x) about its origin, and $R_e(n)$ be the real part of $h_e(k)$'s Fourier transform $H_e(n)$. Plus, since the summation is over an even number of cycles of an odd function, the imaginary term sums to zero.

$$H_e(n) = \sum_{k=0}^{2N-1} h_e(k) e^{-j2\pi nk/2N}$$

$$= \sum_{k=0}^{2N-1} h_e(k) \cos\frac{2\pi nk}{2N} + j \sum_{k=0}^{2N-1} h_e(k) \sin\frac{2\pi nk}{2N}$$

$$= \sum_{k=0}^{2N-1} h_e(k) \cos\frac{2\pi nk}{2N}$$

$$= R_e(n) \qquad [3]$$

Thus, the DCT could be done with only these two steps. However, since the imaginary terms are zero, this is a waste of their imaginary products, which are automatically calculated during the FFT.

A more efficient way in which to calculate the FFT of this 2N point even function, would be to use the imaginary terms in an N-point FFT and then later recombine the result to produce the correct output of a 2N point FFT.

By breaking up the 2N point real function into $m(k) = h_e(2k)$, and $n(k) = h_e(2k + 1)$ for $k = 0, 1, \ldots, N - 1$, which 'interlace' each other, the transform can be written as:

$$H(n) = \sum_{k=0}^{2N-1} h_e(k) e^{-j2\pi nk/2N}$$

$$= \sum_{k=0}^{N-1} h_e(2k) e^{-j2\pi n(2k)/2N} + \sum_{k=0}^{N-1} h_e(2k+1) e^{-j2\pi n(2k+1)/2N}$$

$$= \sum_{k=0}^{N-1} h_e(2k) e^{-j2\pi nk/N} + e^{-j\pi n/N} \sum_{k=0}^{N-1} h_e(2k+1) e^{-j2\pi nk/N}$$

$$= \sum_{k=0}^{N-1} m(k) e^{-j2\pi nk/N} + e^{-j\pi n/N} \sum_{k=0}^{N-1} n(k) e^{-j2\pi nk/N}$$

$$= M(n) + e^{-j\pi n/N} N(n) \quad [4]$$

This means that the original 2N point real even FFT can now be done with two N point FFTs, and a recombination phase with some twiddle factors. M(n) being the FFT of the even points (2k), and N(n) being the FFT of the odd points (2k + 1) for $k = 0,1, \ldots, N - 1$.

However, this still isn't the 2N point real transform, done with one N point complex FFT. To yield this, the principle of linearity, and the transform property are used.

If we let y(k) be the complex combination of the two desired functions to be transformed:

$$y(t) = m(t) + j\, n(t) \quad [5]$$

Then from the linearity property:

$$Y(n) = M(n) + jN(n)$$

$$= \left[M_r(n) + j\, M_i(n)\right] + j\left[N_r(n) + j\, N_i(n)\right]$$

$$= \left[M_r(n) - N_i(n)\right] + j\left[M_i(n) + N_r(n)\right]$$

$$= R(n) + j\, I(n) \quad [6]$$

Since the FFT of real input produces even real and odd imaginary frequency terms, and imaginary input produces odd real and even imaginary terms,

$$M(n) = R_e(n) + jI_o(n) \quad [7]$$

$$jN(n) = R_o(n) + jI_e(n)$$

$$N(n) = I_e(n) - jR_o(n) \quad [8]$$

Plus, since a general function g(x) can be decomposed into even and odd components by the following:

$$g(x) = \frac{g(x)}{2} + \frac{g(x)}{2}$$

$$= \left[\frac{g(x)}{2} + \frac{g(-x)}{2}\right] + \left[\frac{g(x)}{2} - \frac{g(-x)}{2}\right]$$

$$= g_e(x) + g_o(x) \quad [9]$$

and if the function is periodic in N (as are m(t) and n(t)), then:

$$g(-x) = g(N-x) \quad [10]$$

and:

$$g_e(x) = g\frac{(x)}{2} + g\frac{(N-x)}{2} \quad [11]$$

$$g_o(x) = g\frac{(x)}{2} - g\frac{(N-x)}{2} \quad [12]$$

M(n) and N(n) can now be found by creating even and odd functions of the real and imaginary terms of [6] by using [11] and [12].

$$M(n) = \left[\frac{R(n)}{2} + \frac{R(N-n)}{2}\right] + j\left[\frac{I(n)}{2} - \frac{I(N-n)}{2}\right] \quad [13]$$

$$N(n) = \left[\frac{I(n)}{2} + \frac{I(N-n)}{2}\right] - j\left[\frac{R(n)}{2} - \frac{R(N-n)}{2}\right] \quad [14]$$

Finally, by substituting [13] and [14] back into [4]:

$$H(n) = \left[\frac{R(n)}{2} + \frac{R(N-n)}{2}\right] + j\left[\frac{I(n)}{2} - \frac{I(N-n)}{2}\right] + e^{-j\pi n/N}\left[\left(\frac{I(n)}{2} + \frac{I(N-n)}{2}\right) - j\left(\frac{R(n)}{2} - \frac{R(N-n)}{2}\right)\right] \quad [15]$$

By rearranging terms:

$$H(n) = \left[\frac{R(n)}{2} + \frac{R(N-n)}{2}\right] + \cos\frac{\pi n}{N}\left[\frac{I(n)}{2} + \frac{I(N-n)}{2}\right] -$$

$$\sin\frac{\pi n}{N}\left[\frac{R(n)}{2} - \frac{R(N-n)}{2}\right] + j\left\{\left[\frac{I(n)}{2} - \frac{I(N-n)}{2}\right] -\right.$$

$$\left.\sin\frac{\pi n}{N}\left[\frac{I(n)}{2} + \frac{I(N-n)}{2}\right] - \cos\frac{\pi n}{N}\left[\frac{R(n)}{2} - \frac{R(N-n)}{2}\right]\right.$$

Finally, as in [3], the Real term is the desired DCT.

$$R_e(H(n)) = \left[\frac{R(n)}{2} + \frac{R(N-n)}{2}\right] + \cos\frac{\pi n}{N}\left[\frac{I(n)}{2} + \frac{I(N-n)}{2}\right] -$$

$$\sin\frac{\pi n}{N}\left[\frac{Rn}{2} - \frac{R(N-n)}{2}\right] = \sum_{k=0}^{2N-1} h_e(k)\cos\frac{2\pi nk}{2N}$$

Thus, the transform has been accomplished, using an N point FFT of the mirrored N point input to find R(n) + j I(n), and a final recombination pass with twiddle factors to generate the DCT coefficients.

Since the LH9124 includes this last pass as one of its instructions, the data rate of the DCT is that of an FFT of size N, plus one pass to recombine the data.

NOTE: If the output is required to be scaled by $\sqrt{\frac{2}{N}}$, then one more pass must be performed.

CONSTANT-FALSE-ALARM-RATE ANALYSIS USING THE LH9124/LH9320

INTRODUCTION

The Constant-False-Alarm-Rate (CFAR) process is an important post process for detection systems such as Sonar and Radar. The process is usually applied to preprocessed data, before it's displayed to an operator or transferred to an automatic detection and tracking system.

The input to the CFAR process varies from application to application, and could be one or two dimensions. A typical one-dimension input is a spectrum, where the input vector is the energy versus the frequency. In this case the signal is a tonal line and the background level is noise (Figure 1).

The detection criteria of CAR algorithm is:

$$E(i)/N(i) > R$$

where:

E(i) = input vector level

N(i) = background noise level

R = detection threshold

The above criteria means that a point on the input vector is detected as a signal if it is R time bigger than the background noise.

The advantages of this criteria rather than a constant level threshold are:

- The detection threshold is a normalized number and does not depend on the absolute level of the input level. Therefore, does not depend on the system and process gains and the calibration.
- The detection is done with a constant-false-alarm probability. The false-alarm rate is only a function of the parameter R.

The key issue in the CFAR processor is estimating the background noise level. Usually, the background noise level is not constant, therefore it is necessary to estimate it separately for each point on the input vector. The most popular algorithm to estimate the noise level is the Split Sliding Moving Window. For each point of the input vector the estimated noise level is a weighted average of the levels of the input vector around the point of interest, excluding the very-close neighborhood. The design of the window function usually gives more weight to the points that are close to the center of the window. A typical window function is illustrated in Figure 2.

The estimation of the noise level is given by:

$$N(i) = \sum_k w(k) \cdot E(i+k) = \sum_k w(-k) \cdot E(i-k) =$$

$$w(-i) * E(i)$$

since w(i) is usually symmetric function:

$$N(i) = w(i) * E(i)$$

Where * is the convolution mark.

The meaning of the above is that the noise-level-estimation process is a convolution operation between the input vector and the window function.

The CFAR algorithm is summarized in Figure 3.

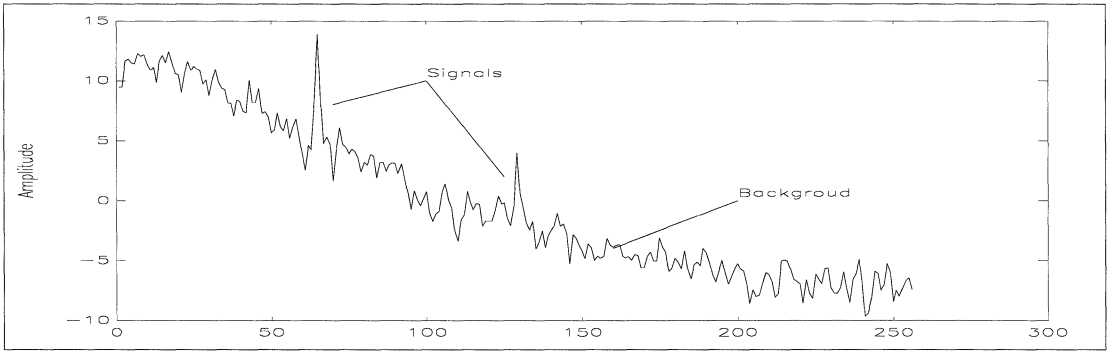


Figure 1. Spectrum Display

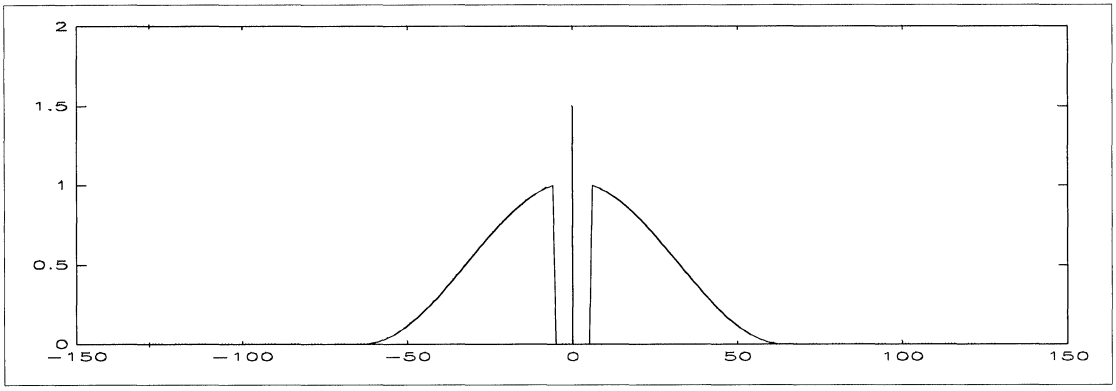


Figure 2. Typical Weighting Window for Noise Estimation

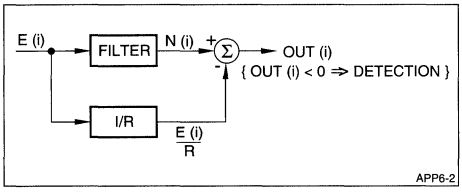


Figure 3. CFAR Process Block Diagram

The filter response should have a gain of one for constant level input, therefore the DC gain of the filter should be one, To obtain this requirement the filter coefficient should be normalized as follows:

$$w(0) = \sum w(n) = 1$$

The heavy number crunching part of the algorithm is the FIR operation. Basically, there are two options to implement the FIR operation in the time domain or in the frequency domain (Fast Convolution). Usually, it is more efficient to implement the frequency domain (Fast Convolution).

When using a trivial window function (i.e., a constant level with G zeros in the center). The moving average can be computed in a complexity of four ADD operations and one MUL operation for each input point. Sliding the averaging window by one step involves adding two new points; subtracting two points; and multiplying the result by the factor R.

This option is not considered in this application note for the following reasons:

1. It is not possible to handle a non-trivial window function.
2. The algorithm is numerically unhealthy, since the computation noise is accumulated in the background estimator with no fading mechanism. The computation-noise variance increase linearly with i, where i is the input-vector index.
3. The algorithm can not be described as a vector operation. Therefore, it is not well mapped to the LH9124/LH9320 based system.

USING THE LH9124/LH9320

One-Dimension Case

With this example the following parameters are assumed:

Length of the input vector N = 4096

Length of the averaging window D = 32.

Algorithm Steps:

1. Perform the frequency-domain convolution between the input vector E and the window function W to obtain the noise estimation N.
   - Perform FFT to the input vector using the optimal mixed radix. For the case of N = 4096, the most efficient radix is $16 \times 16 \times 16$
   - Multiply by the window-function-frequency response. This response should be pre-prepared at the C port. The window-function-frequency response is:

     $W = FFT_{N=4096}(w)$

NOTE: the window function, as defined, contains values for negative indexes ($i < 0$), while the input to the FFT assumes that $0 \leq i < N$, therefore the negative-index points should be reflected to positive-index points prior to performing the Fourier transform.

$w(-i) = w(N - i)$ (see Figure 4)

– Perform IFFT

2. Normalize the input vector by a factor of 1/R and transfer the result to the C port.
3. Compare the vectors (the result of steps 1 and 2). The comparison is done by subtracting the vectors and monitoring the sign bit.

Programming

In this one-dimension example a standard hardware configuration is assumed (Figure 5).

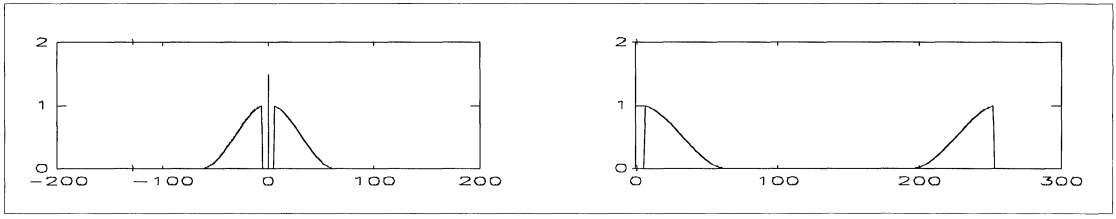


Fig. 4. Window Funct. Reflation Int. Positive Time Index

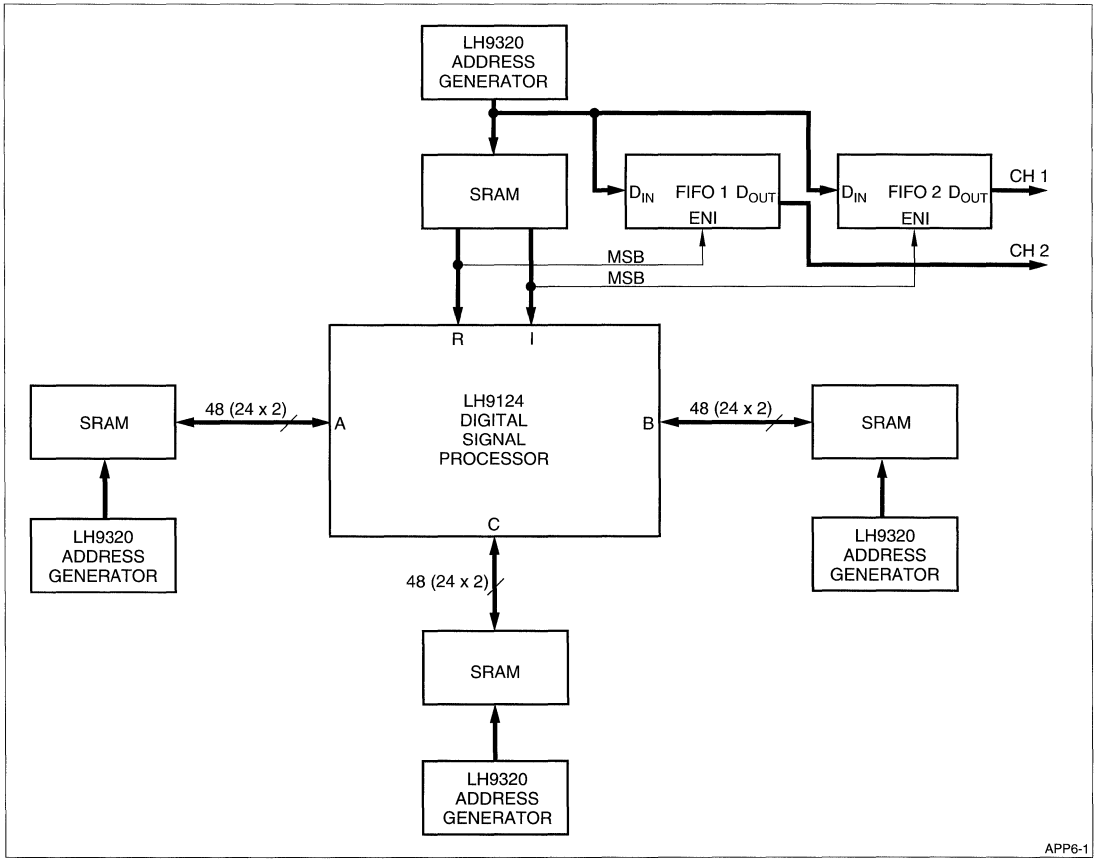


Figure 5. Hardware System Configuration Example

The registers programming is:

REGISTER	PROGRAMMING
N	4096
DIGITREV	0000 0000 1111 1111 1111
MEMSIZE	4096
ADRLENGTH	4096
ADRSTART	0
ADRINC	1

Explanation and Remarks

1. Steps 1-7 performs the FIR in the frequency domain to estimate the noise level.
2. When performing the IFFT (steps 5-7) it is possible to use the twiddle-factor-lookup table used for the FFT by controlling the LH9124 CI and CO control signals as follows:

   CI = 1 during pass 5, and 0 for all other passes.

   CO = 1 during pass 7, and 0 for all other passes. For details see reference 3.
3. Step 8 normalizes the input vector by 1/R
4. Step 10 performs (N(i) – E(i)/R). Whenever the result is negative (the sign bit set to one) the point is above the detection threshold and the address of this point should be stored. In this example, we propose a FIFO device to accumulate all the address of all the detected points (Figure 5).
5. It is important to notice that when subtracting the two vectors, N(i) and E(i)/R, they should have the same scale factor. One way to guarantee this is to use the BFPO outputs at the end of pass 7, as DSFI input in pass 9, and by setting manual scale factor mode-DSFISEL = 1.
6. In the implementation above, we assumed that the input is real, in a case where the input is complex a CMAG (magnitude square) pass should precede the algorithm steps described above.

As detailed in Table 1, the CFAR algorithm is completed in ten passes (for the case of N = 4096). However, it should be noted that two input vectors can be processed simultaneously with exactly the same algorithm (and the same number of passes) by loading one vector to the real input memory and the second channel to the imaginary-input memory. In this case, the output will be:

- The sign bit of the real output will indicate a detection of the first channel.
- The sign bit of the imaginary output will indicate the detection of the second channel. This process is correct since:

  $$Out = IFFT\{W \cdot FFT\{E\}\} - E/R\}$$

  Loading one vector to the real input and the second channel to the imaginary input creates a complex input.

  $$E = E1 + jE2$$

  $$Out = IFFT\{W \cdot FFT\{E\}\} - E/R\} = IFFT\{W \cdot FFT\{E1 + jE2\}\} - (E1 + jE2)/R\}$$

  $$Out = \{\{IFFT\{W \cdot FFT\{E1\}\} - E1/R\} + j\{\{IFFT\{W \cdot FFT\{E2\}\} - E2/R\}$$

  $$Out = Out1 + jOut2$$

  Since both Out1 and Out2 are real sequences:

  $$Real\{Out\} = Out1;$$

  $$Imaj(Out\} = Out2;$$

Table 1. CFAR Programming for One-Dimension Case

PASS	OP CODE	DATA FLOW	AGin	AGout	AG COEFF
1	BFLY16	Q→A	RBF0	BF160	TF160
2	BFLY16	A→B	BFF161	BF161	TF161
3	BFLY16	B→A	BF162	BF162	TF162
4	CMUL	A→B	INC	INC	INC
5	BFLY16	B→A	RBF0	BF160	TF160
6	BFLY16	A→B	BF161	BF161	TF161
7	BFLY16	B→A	BF162	BF162	TF162
8	CMUL	Q→B	INC	INC	INC
9	MOVD	B→C	INC	INC	INC
10	CSUB	A→Q	INC	INC	INC

Two-Dimension Case

The CFAR algorithm can appear as a two-dimension problem. For example, a waterfall display, where the spectrum as a function of time, is presented to the operator to increase the detection probability. Another example is a defect-detection system that looks for crakes on a homogeneous surface.

In the two-dimension case the background level is done by averaging over a two-dimension window around the point-of-interest that excludes the vary close neighborhood of the point itself. The design of the averaging window is application dependent on the case of a spectrum-waterfall display, where the system looks for a vertical line on the display. A typical averaging window is illustrated in Figure 6.

As in the one-dimension case the noise estimation is a convolution operation between the input (two-dimensional input) and the window function. Basically, the algorithm is the same as the one-dimensional case, but the FFT and the IFFT should be two-dimensional.

The method of performing a 2-D-FFT using the LH9124/LH9320 chip set is explained in the LH9124-FFT Application note. Basically, the two-dimension transform can be presented as a one-dimension transform with a length of $N \times M$ and with a proper addressing sequence setting.

In this application note the case of is 4096×256 is demonstrated. The transform length is 4096×256. Therefore, the optimal radix is $(16 \times 16 \times 16) \times (16 \times 16)$.

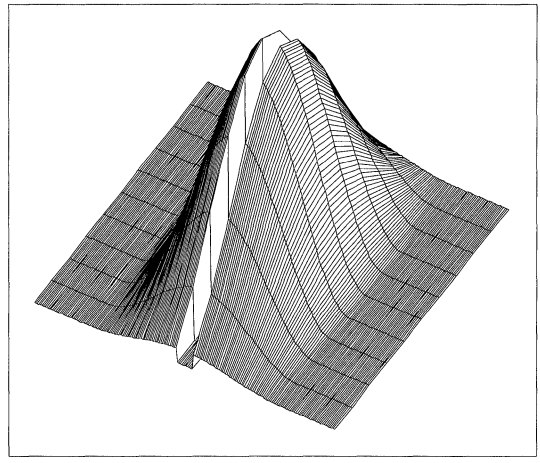


Figure 6. 2-D Window Function for Noise Estimation

The registers programming is:

REGISTER	PROGRAMMING
N	1M
DIGITREV	1111 1111 1111 1111 1111
MEMSIZE	1M
ADRLENGTH	1M
ADRSTART	0
ADRINC	1

NOTE: that the remarks stated for the one-dimension case are relevant to the two-dimension case.

Table 2. CFAR Programming for One-Dimension Case

PASS	OP CODE	DATA FLOW	AGin	AGout	AG COEFF
1	BFLY16	Q→A	RBF0	BF160	TF160
2	BFLY16	A→B	BFF161	BF161	TF161
3	BFLY16	B→A	BF162	BF162	TF162
4	BFLY16	A→B	BFF163	INC	INC
5	BFLY16	B→A	BF1624	BF160	TF160
6	BFLY16	A→B	BFF165	BF161	TF161
7	CMUL	B→A	INC	INC	INC
8	BFLY16	Q→B	RBF0	BF160	TF160
9	BFLY16	B→C	BF161	BF161	TF161
10	BFLY16	A→Q	BF162	BF162	TF162
11	BFLY16	B→A	BFF163	BF163	TF161
12	BFLY16	A→B	BF1624	BF160	TF162
13	BFLY16	B→A	BFF165	BF161	TF161
14	CMUL	A→B	INC	INC	INC
15	MOVD	B→C	INC	INC	INC

Synchronous and Asynchronous SRAM Memory for SHARP's LH9124/LH9320 DSP Chip Set

INTRODUCTION

The SHARP LH9124/LH9320 Digital Signal Processing chip set typically uses either a recursive ping-pong or a cascaded dataflow memory array configuration. This application note focuses on the timing for a recursive separate acquisition memory configuration as shown in Figure 1.

Separate Acquisition Memory Configuration

This system configuration provides five separate arrays of memory. Data enters the system through the Q-port acquisition memory; is processed in a ping-pong method (back and forth) between the A-port and the B-port memory arrays; and then exits the system through the Q-port output memory. The circuit timing for the memories and the FIFO (First In, First Out memory) control logic is discussed and illustrated later in this application note.

The specification for the DSP chip set describes setup and hold times that indicate an SRAM address to data valid time of ≤10 ns (in a 40 MHz system). In some applications the cost of this memory may be prohibitive, however, synchronous SRAM design in these cases is the recommended alternative. For systems operating slower than 40 MHz, less expensive and slower asynchronous SRAMs can be used. This application note discusses designs for both synchronous and asynchronous 40 MHz memory designs.

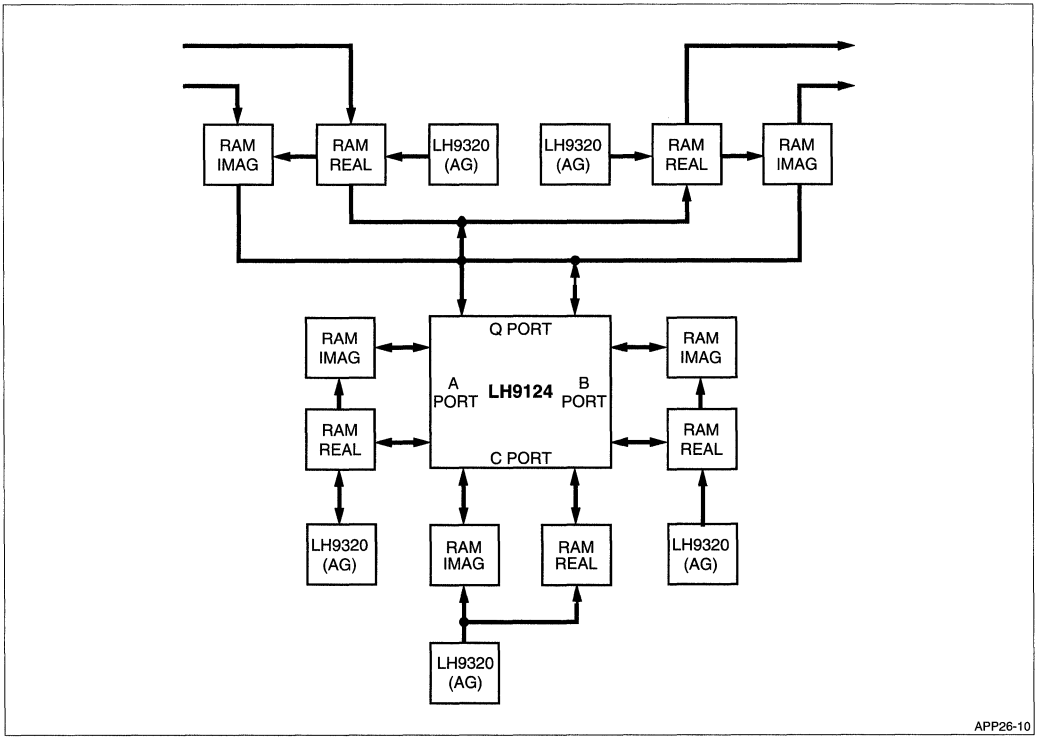


Figure 1. Separate Acquisition Memory Configuration

Ping-Pong Memory

The Ping-Pong Memory Array implementation consists of an A-port and B-port LH9320 Address Generator with a SRAM for each port (Figure 2). As the system processes, its algorithm data is passed back and forth between the A-port and the B-port memories. The timing and the control logic for this memory is straightforward and is shown in the Timing and Logic (TL) Figures 5 and 6.

Coefficient/Twiddle Memory

The Coefficient/ Twiddle Memory Array contains the DSP specific data that is used for either the function or the algorithm implemented (Figure 3). Also, this memory array is implemented as the host acquisition port. The host port can read and write this memory to load data into the system or read results from the system. Both the Q-port and C-port acquisition memories are separate I/O memories. Reading and writing the C-port memory is straightforward. The circuit timing is shown in Figures 5 and 6.

FIFO and Acquisition Memory

The FIFO and Acquisition Memory implementation consists of a LH9320 Address Generator, a SRAM, and the acquisition FIFO (Figure 4). This FIFO insures that data is not lost from the real-time data stream when the acquired data is passed through the LH9124.

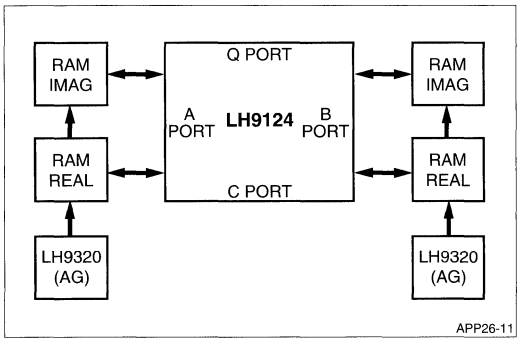


Figure 2. Ping-Pong Memory Configuration

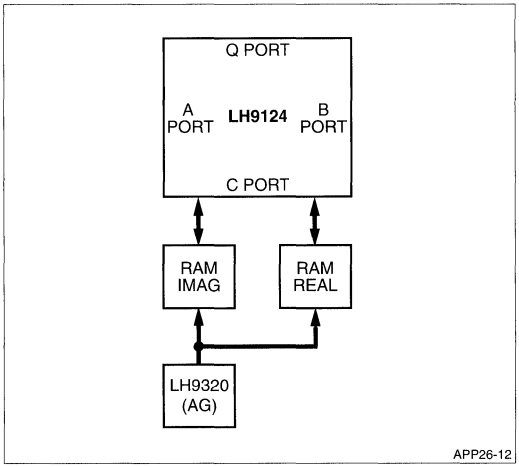


Figure 3. Coefficient/Twiddle Memory

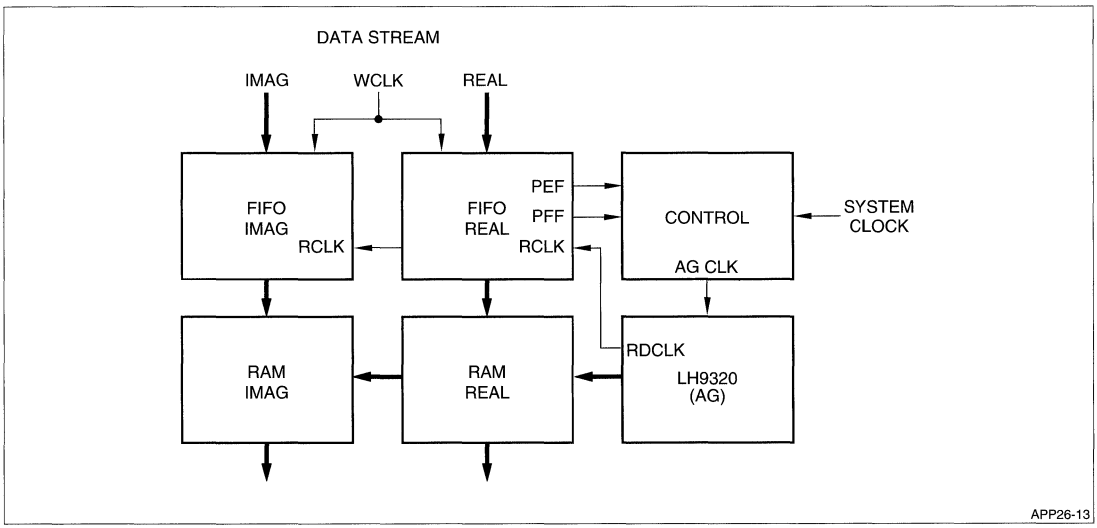


Figure 4. FIFO and Acquisition Memory Configuration

Data is collected in the FIFO and passed to the acquisition SRAM using the LH9320's address sequence. Since the real-time data rate is slower than the DSP system the FIFO indicates to the LH9320 when it is empty and that it needs to wait for more data. When enough data is acquired, the FIFO indicates to the LH9320 that it may resume loading the memory.

For the acquisition control, the FIFO turns off and on the LH9320 clock. When the DSP is using the acquisition memory, the LH9320 clock is the system clock and the FIFO is ignored by the LH9320. Figures 9 and 10 show the timing and control logic for the operation of the block.

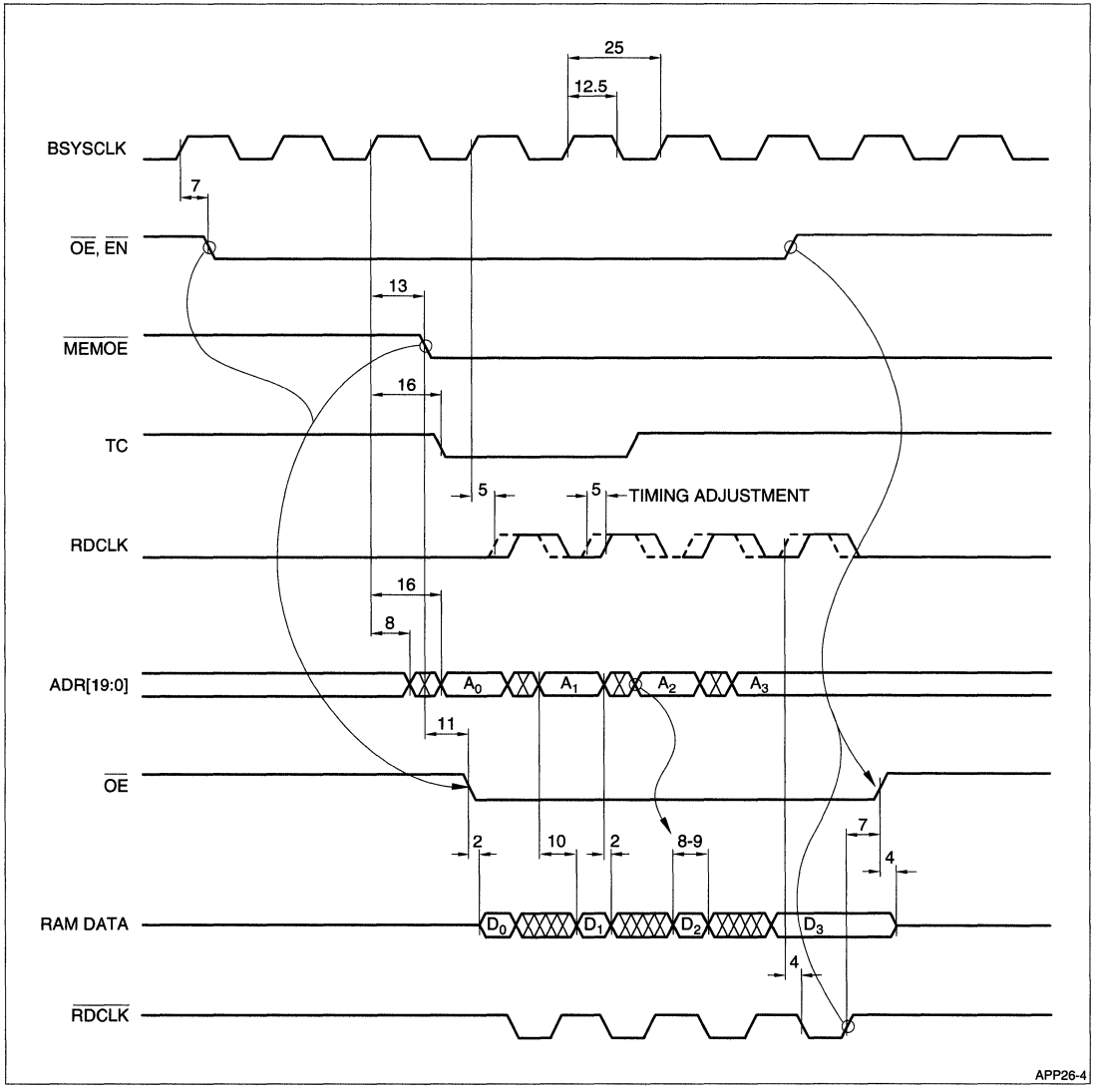


Figure 5. Ping-Pong/Coefficient Read Cycle

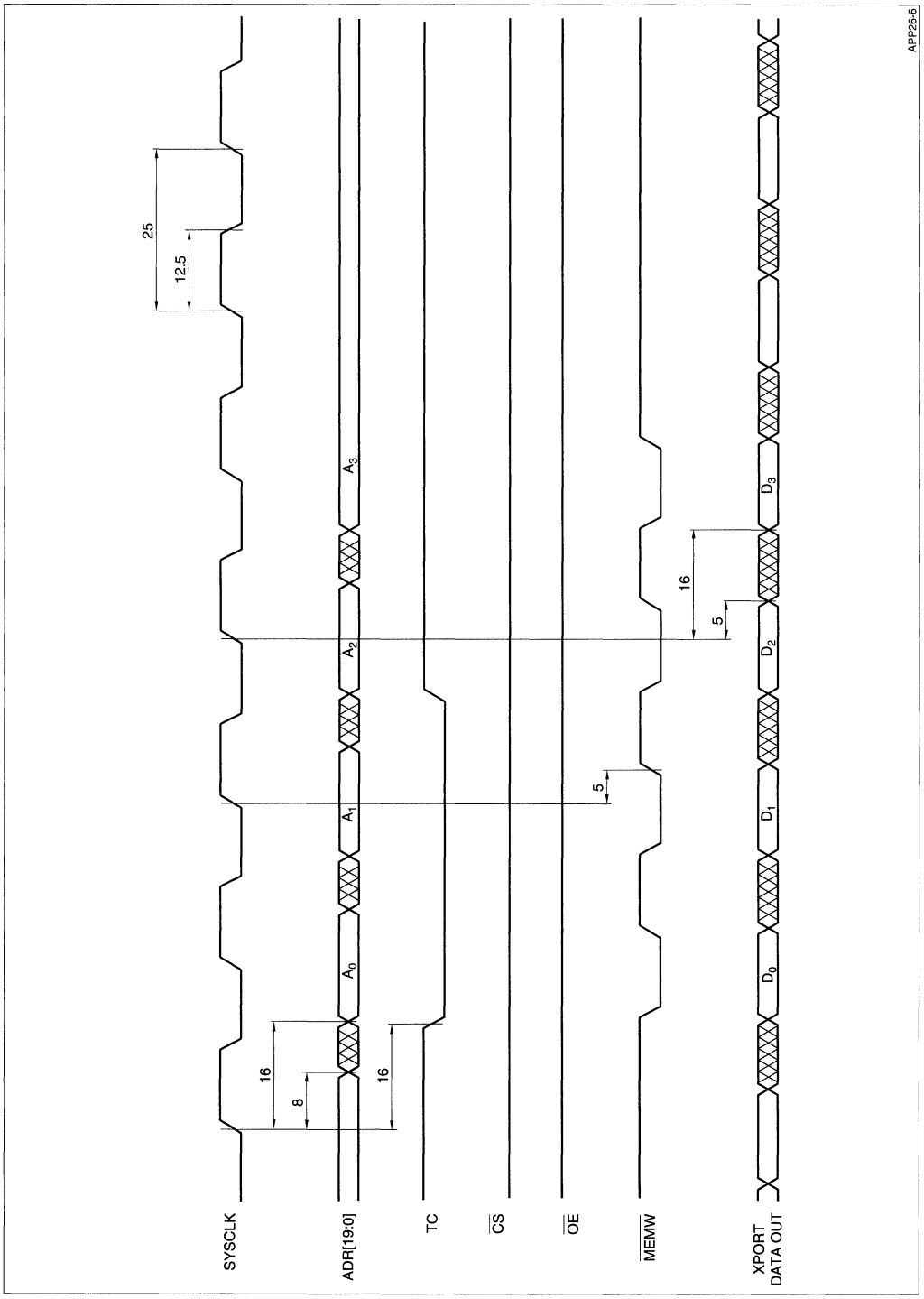


Figure 6. Write Cycle Timing for A, B, and C Ports (Typical)

Output Memory

The Output Memory is very similar in design to the Input Memory (Figure 7). The output data SRAM is filled by the last pass of the DSP algorithm being processed and then switched to the FIFO control mode for output. As the FIFO fills, it stops the LH9320 until it's almost empty. The LH9320 is resumed after a pause and then this pause/resume relationship continues until the data from the array is exhausted.

After the output data array is fully output, the output array is switched and reloaded from the DSP for the next pass of data. This unload/load processing continues until the system is either halted or reset.

Circuit Board Component Placement

Figure 8 illustrates a circuit board component placement configuration. The SRAMs are positioned as close as possible to the DSP ports that they are servicing. The LH9320s are placed as close as possible to the SRAMs. Miscellaneous components (host, etc.) are placed in the remaining spaces.

IMPORTANT DESIGN NOTE: The placement of the SRAMs and the LH9320s is critical to insure for short signal delays and more typical timing.

SUPPLEMENTARY INFORMATION

Generally, the control for this type of memory system would be either an external bus or a microprocessor/controller. In the application discussed here, a simple PLA controller was blocked out and implemented. The PLA controller sends a status to all LH9320s at once and switches the acquisition/output memories. As the LH9320s complete their sequences, they send a TC (terminal count) back to the scheduler so it can cycle to the next state and continue the processing. Since this application note is concerned with memory implementation, the scheduler is defocused and is used only for simple operation. Further reference information on scheduler applications can be found in the application note for microprocessor control for the LH9124/LH9320 DSP system.

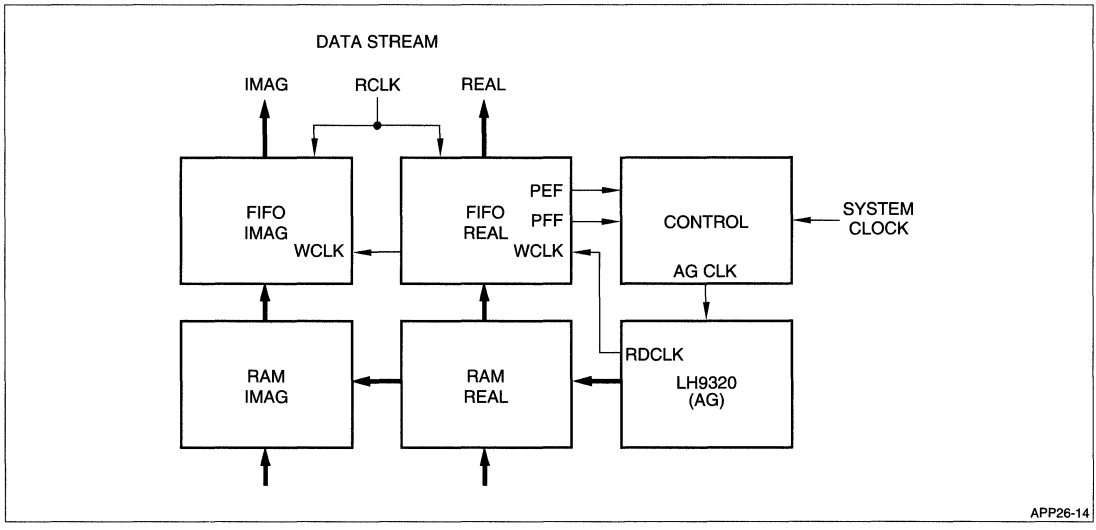


Figure 7. FIFO and Output Memory Configuration

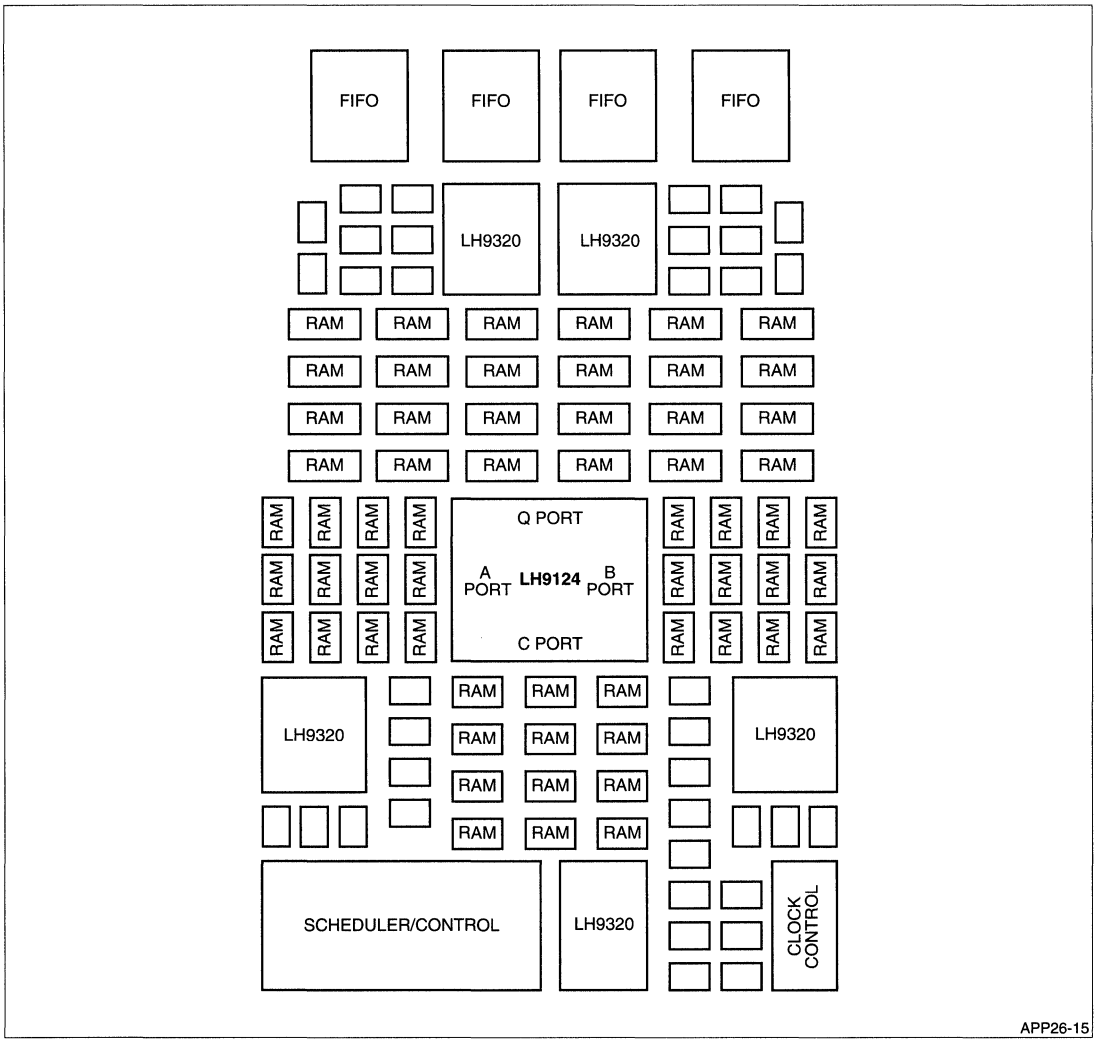


Figure 8. Circuit Board Component Placement

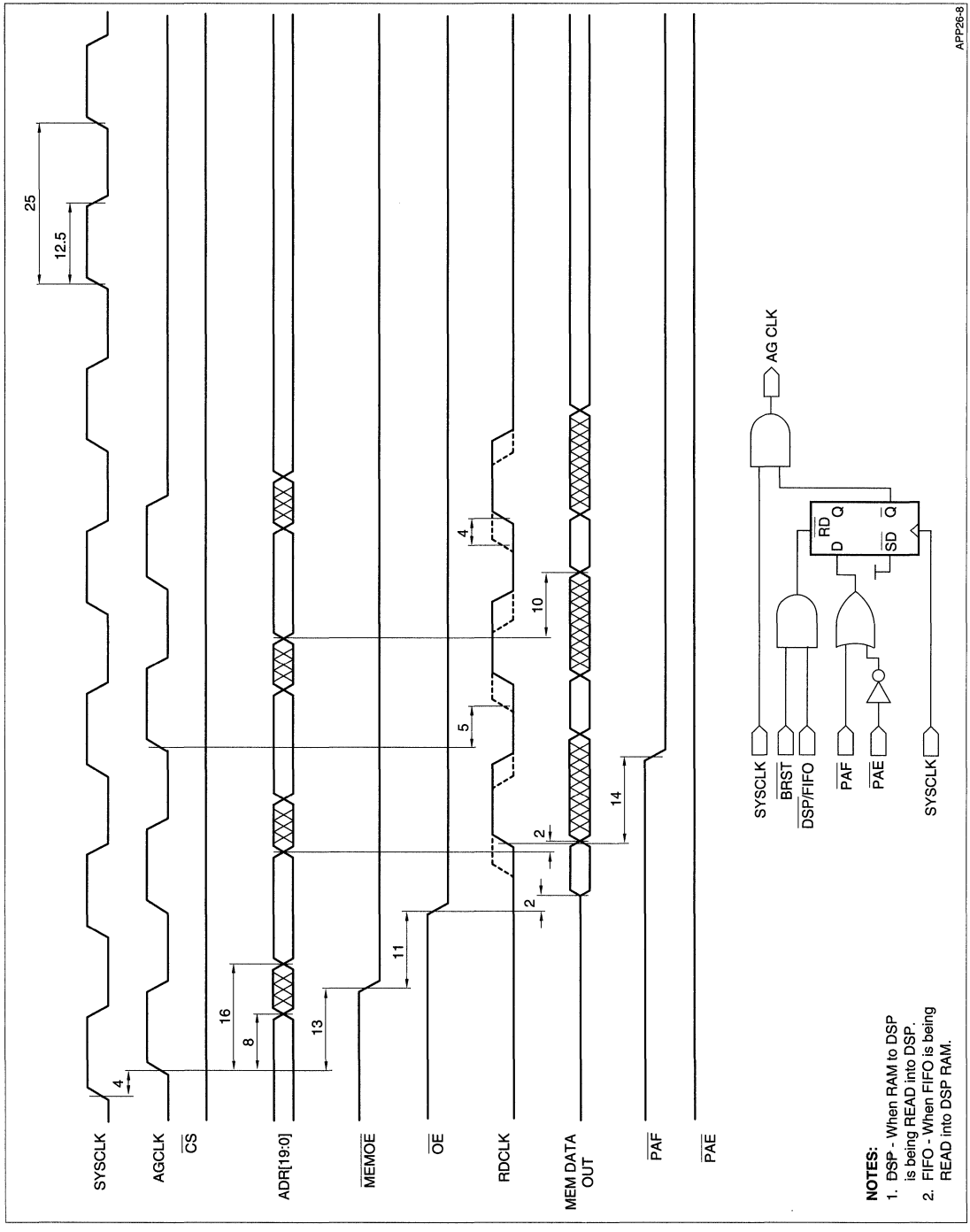


Figure 9. Write Pause AG When FIFO is PAF
AG Clock Circuit for FIFOs

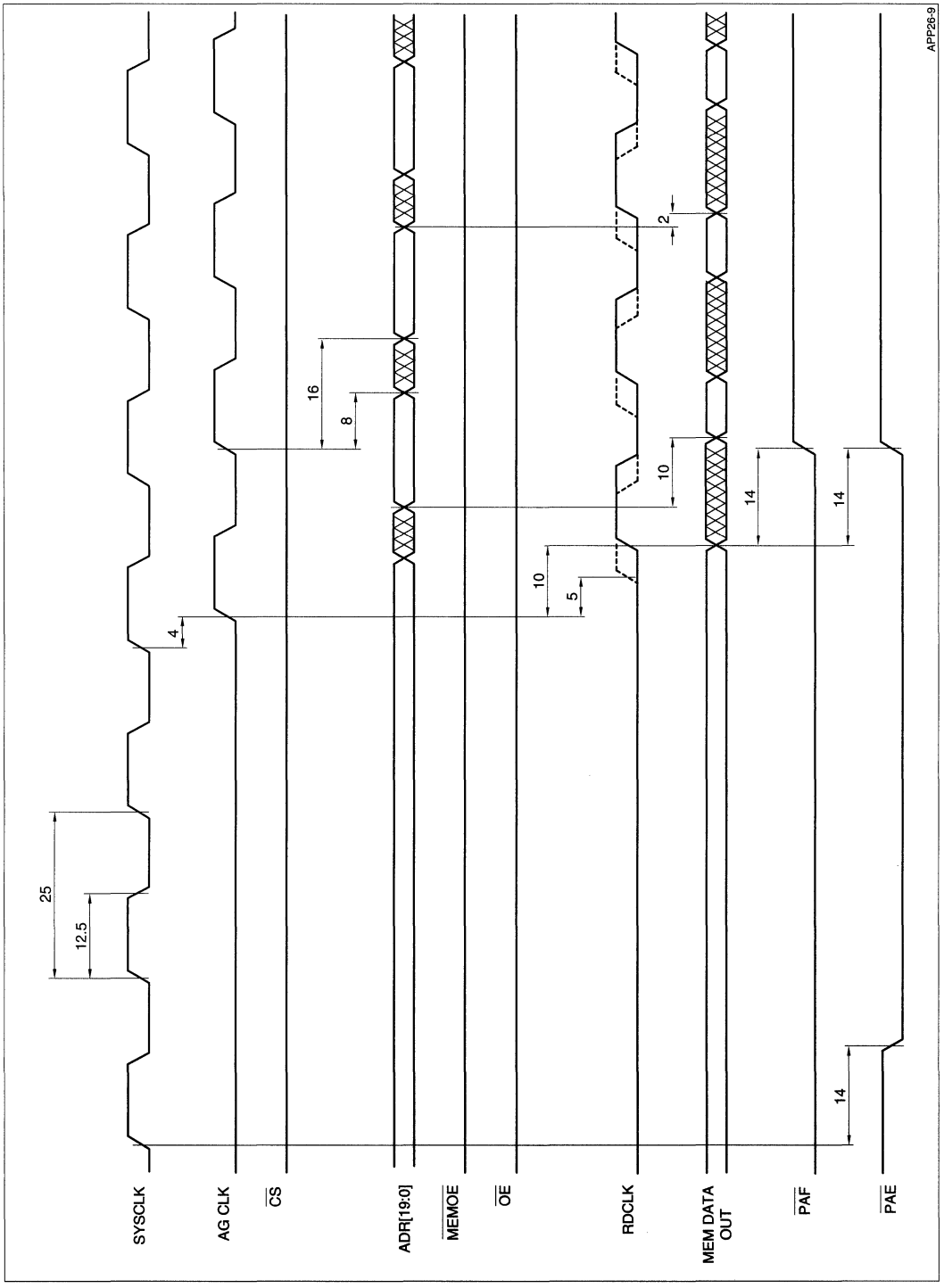


Figure 10. Paused AG When FIFO is PAF (cont'd)

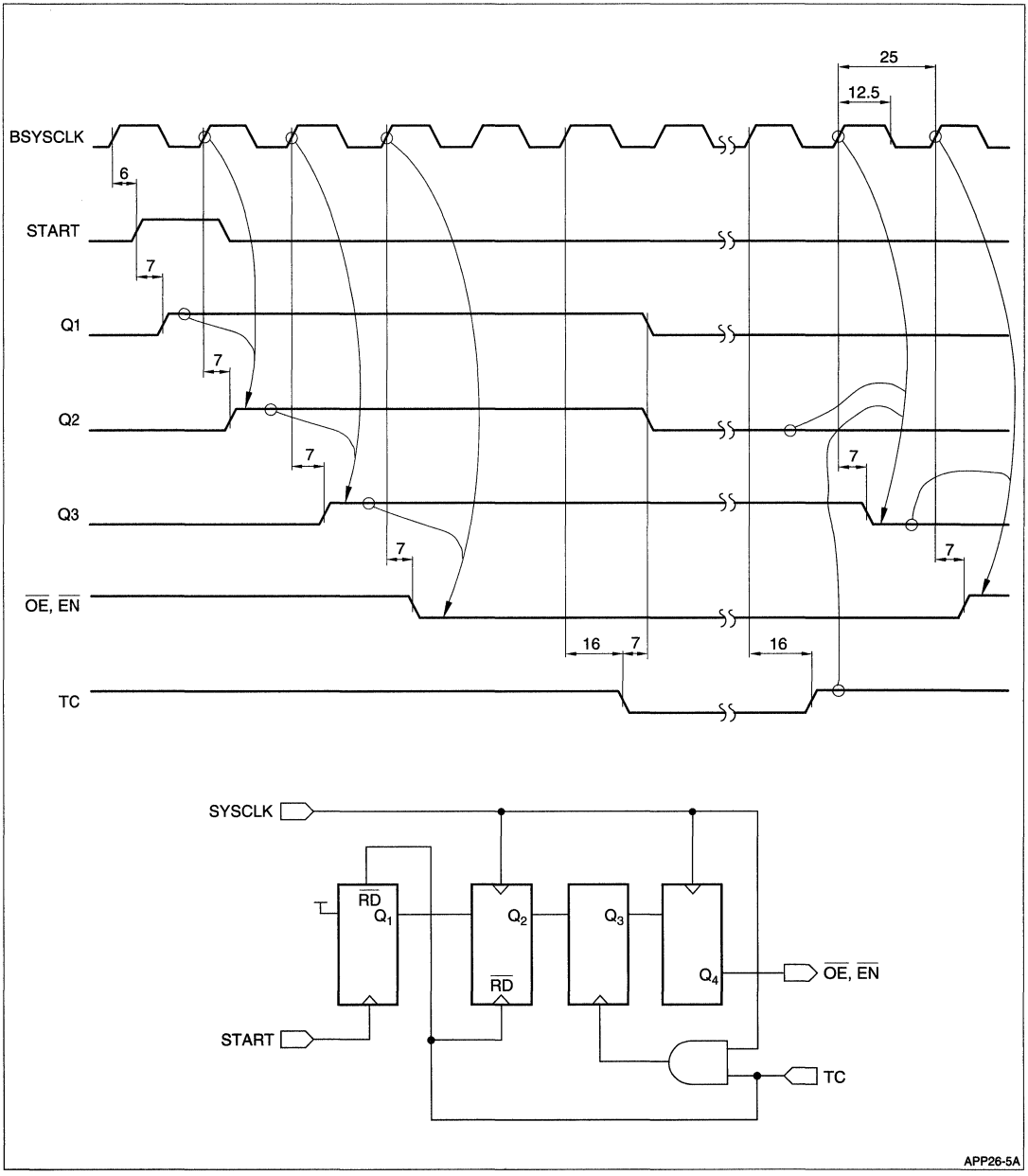


Figure 11. Start Control Circuit Timing

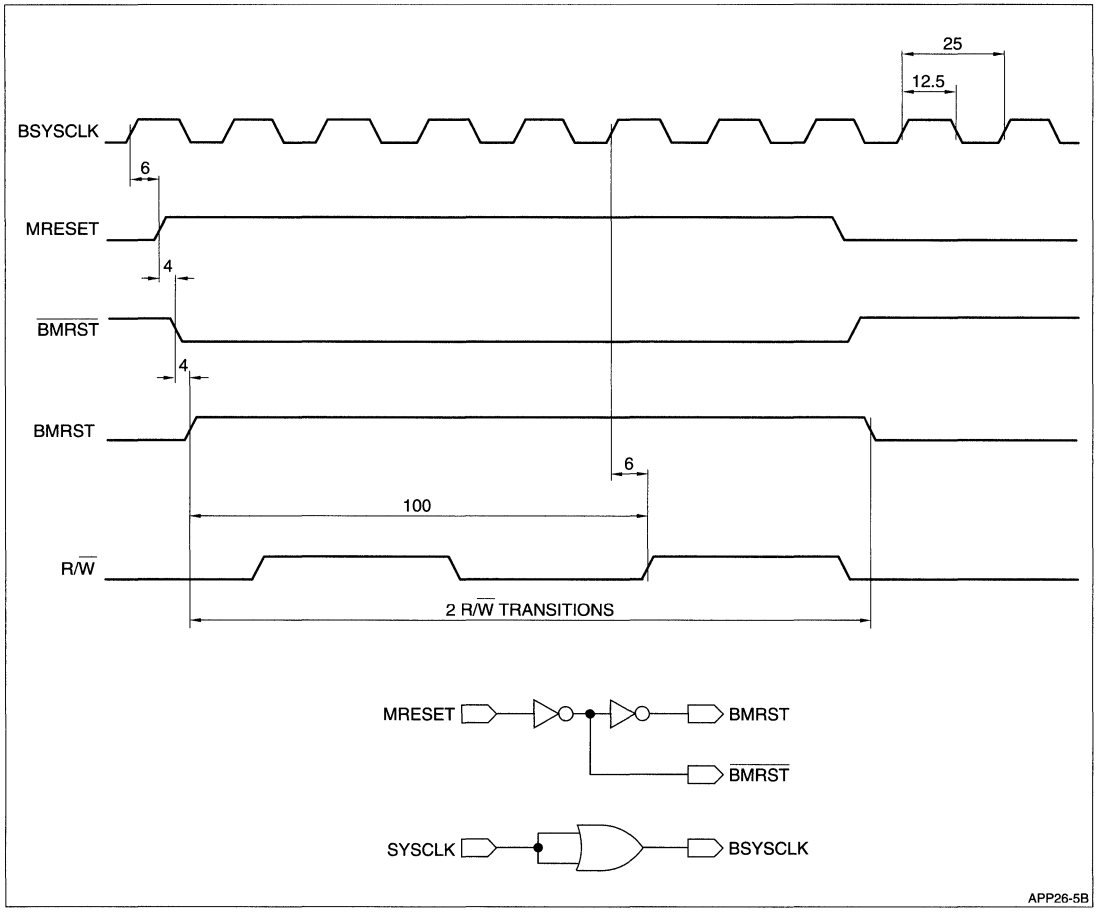


Figure 12. Reset and Clock Buffering

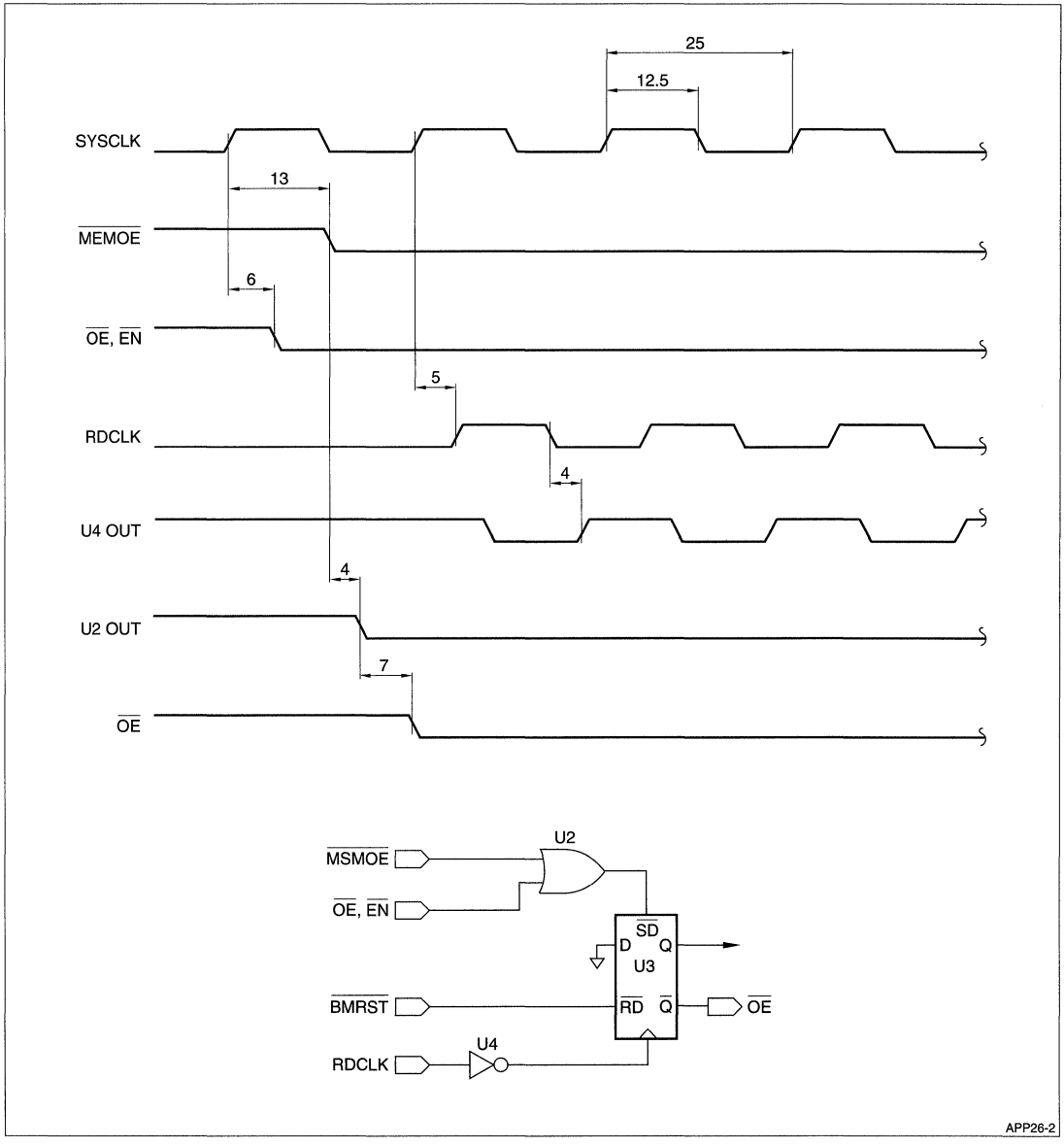


Figure 13. RAM OE Circuit Timing (OE Enabled)
OE Circuit for DSP

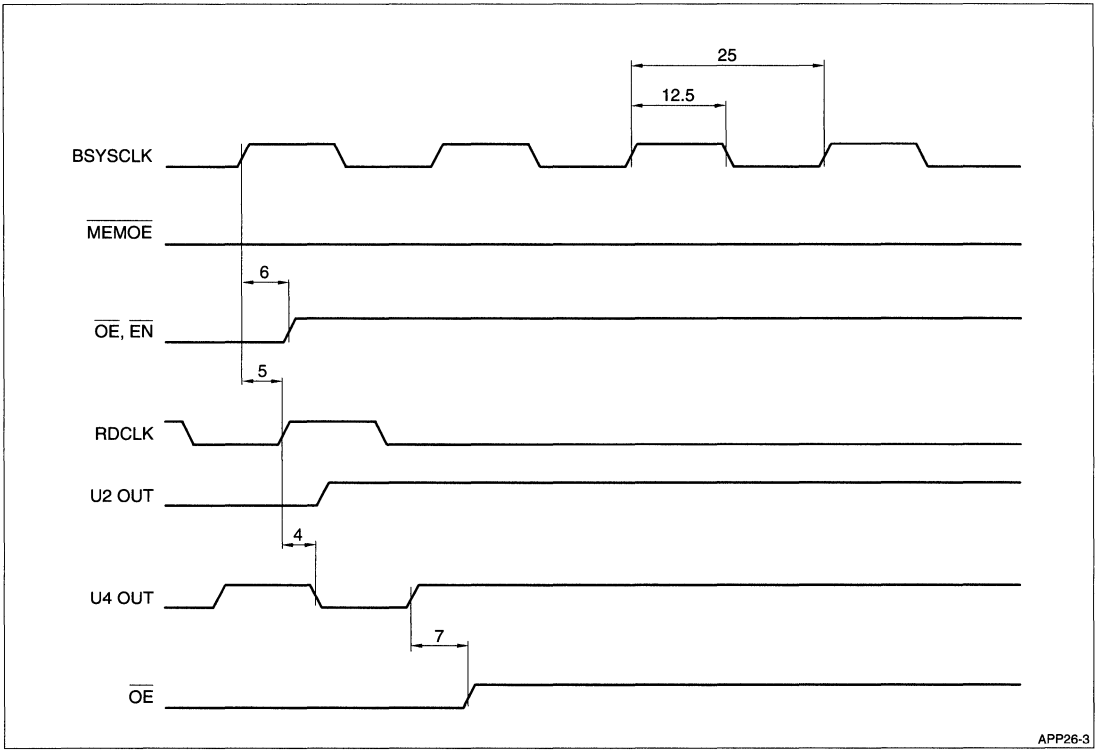


Figure 14. RAM OE Circuit Timing (OE Disabled)

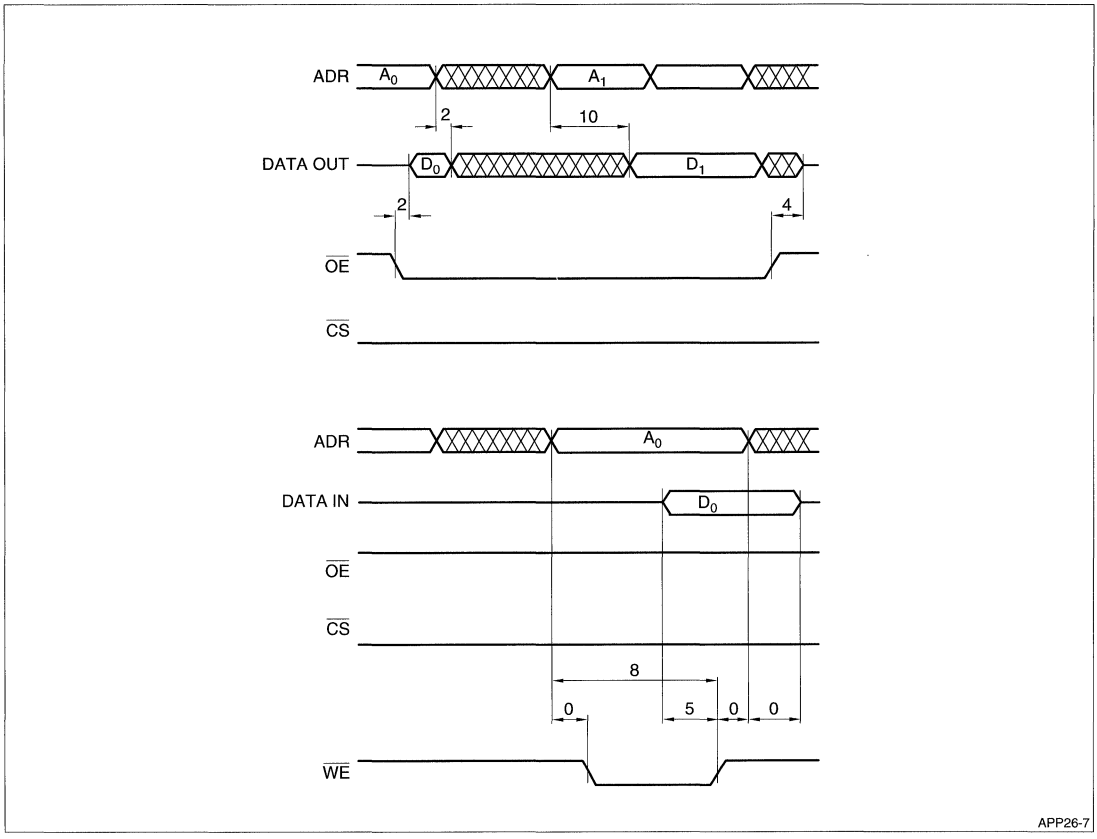


Figure 15. Basic RAM Read and Write Timing

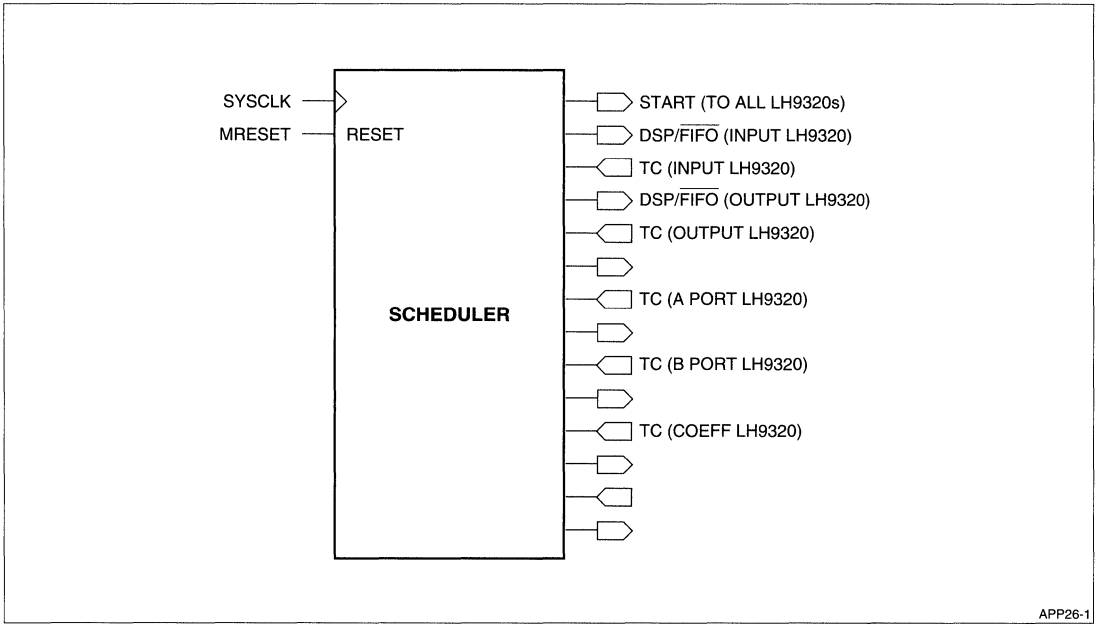


Figure 16. Scheduler

FREQUENCY DOMAIN LMS – DEMONSTRATED FOR MULTICHANNEL ECHO CANCELLING

INTRODUCTION

Adaptive filters are used in a variety of applications, such as channel equalizing, interference cancellation, adaptive antennas, spectral line enhancement, and echo cancelling. The most popular implementation is a taped delay filter whose coefficients are updated according to the LMS algorithm. This application note describes the implementation of the frequency domain LMS [2] using the the LH9124. For demonstration purposes, this application note describes in detail the echo cancellation problem. However, the principles presented here can be used to implement other Adaptive-Filter systems. Frequency domain adaptive filtering offers the following advantages over time-domain implementations:

1. Better Use of Computational Resources. With time-domain implementations of the LMS algorithm, computation time increases linearly with filter length. With the frequency domain implementation using the LH9124, increasing the filter length introduces a very moderate increase in the computation load. For example, increasing the filter length from 128 to 2048 (factor of 15) increases the computation load by 10-20%.
2. Better Stability. Since the frequency-domain implementation of the LMS algorithm is block implementation, there is an averaging effect, which makes the algorithm more stable than the point-by-point instantaneous gradient estimation of the time-domain approach.

However it should be noted that a frequency-domain implementation of the LMS algorithm is a block implementation, therefore it introduces a latency (delay) equal to the filter length.

The Echo-Cancellation Problem

True full-duplex telephony requires a four-wire system, but telephones typically have only two wires. The conversion from four-wire to two-wire is performed by a device called a 'hybrid,' which ideally has infinite isolation between talker and listener. In practice, hybrid attenuation between the talking and listening sides of a four-wire transmission medium is guaranteed to be more than 6 dB. This small attenuation means that when the talker at one end of a telephone circuit is speaking the leakage of the hybrid at the other end will return to the talker as an echo. For a long-distance circuit, the delay could be up to 500 ms, resulting in a very noticeable and disturbing echo. A simplified schematic of a long-distance connection is illustrated in Figure 1.

In the past, the solution to this problem was to use echo suppression, which identifies a single-connection talker situation by comparing the voltages on the transmission and return sides of the hybrid, and then switching off the return side to stop the retransmission of the hybrid-leakage signal which causes echoes. This is not a good solution for long-distance circuits, because the return-path relay switching time causes front-end clipping. Adaptive echo-cancellers were developed to stop echoes without introducing new artifacts.

The echo-causing mechanism can be modeled as illustrated in Figure 2.

The usual assumption is that h(t) is a finite impulse response. The basic idea behind adaptive echo-cancellation, as illustrated in Figure 3, is to estimate the echo-transfer function h(t) to reconstruct a replica of the echo, and to subtract it from the signal on the return path of the hybrid. Due to the quasi-stationary character of the medium, it is not sufficient to assume a constant known transfer function. Some form of adaptive filter is required.

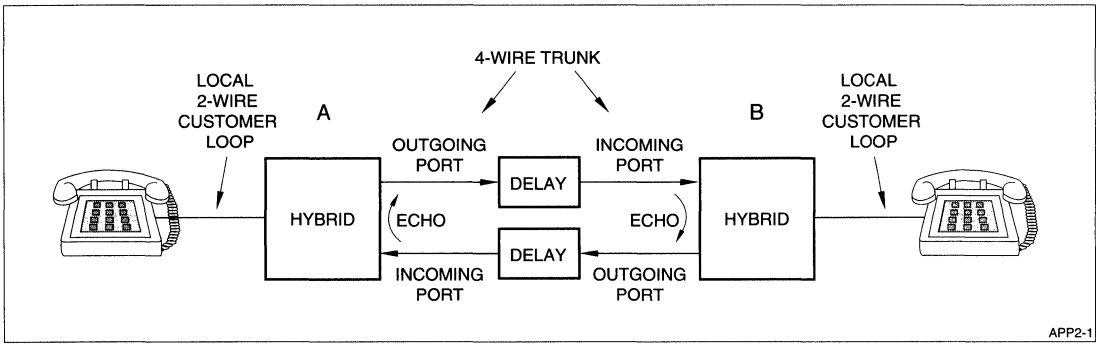


Figure 1. Simplified Schematic of Long Distance Connection

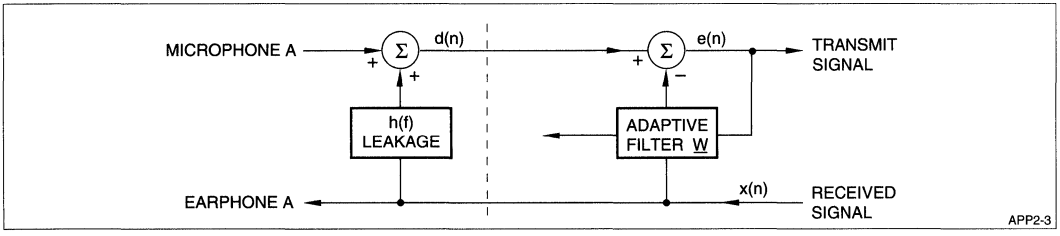


Figure 2. Adaptive Echo-Canceller Signals

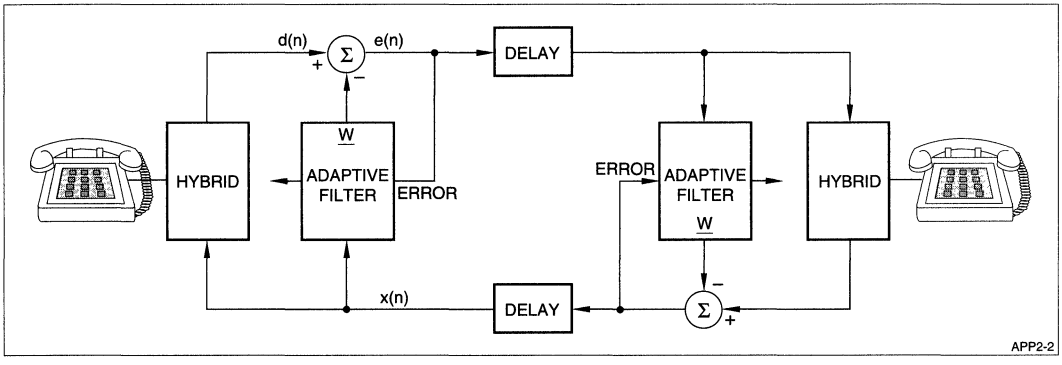


Figure 3. Adaptive Cancellation System Structure

Adaptive Filter Structure

The adaptive algorithm works by subtracting a filtered version of the received signal from the transmitted signal. The filter coefficients must be updated from time to time, in order to minimize the error energy, or echo signal. The optimal solution, called the Wiener Solution, is (refer to Figure 2):

$$e = d - \underline{X} \cdot \underline{W}^T$$

It can be shown that

$$Wopt = R^{-1} \cdot P$$

where:

$R = E(\underline{X} \cdot \underline{X}^T)$ and $P = E(\underline{X} \cdot d)$
$\underline{X}$ is the received signal
d is the transmitted signal

However, this solution has no practical use for our case, since R and P are not constant and not known.

The practical solution is to update the filter coefficients so that the error signal converges to the optimal solution, or close to it. The most common convergence algorithm is the time-domain Least Mean Square, or LMS, algorithm. The LMS algorithm updates the filter coefficients in a direction opposite to the gradient of the error. [4]

$$w_{k+1}(i) = w_k(i) + 2\mu E\{x(i-k) \cdot e(i)\}$$

NOTE: The term E{e(i)x(i-k)} is the cross-correlation between e and x. Normally, a good estimate of the cross-correlation is to average the instantaneous correlation over time.

The algorithm proposed by Widrow uses the instantaneous correlation as an estimate, yielding an equation for updating the coefficients:

$$w_{k+1}(i) = w_k(i) + 2\mu x(i-k) \cdot e(i)$$

Here, μ is the adaptation step size. The step size governs the behavior of the filter. Increasing μ increases the convergence rate, but at the expense of lower stability and higher residual noise. Since the convergence rate is also a function of the input power $\left(\sigma_x^2 = E\{x^2\}\right)$, a high-power signal will have the same effect as high μ. A normalized step size is usually defined as $\mu = \dfrac{\alpha}{\sigma_x^2 \cdot N}$

where:

N is the filter length
σ_x^2 is the input power
α is a pure number, usually selected to be 0.1

The time-domain LMS can be summarized by the following equations, for each data point:

$$e(i) = d(i) - \sum_{k=0}^{N-1} x(i-k)\, w_i(k) \qquad \text{(Filtering)}$$

$$\sigma_x^2(i) = (1-\lambda)\,\sigma_x^2(i-1) + x^2 \cdot \lambda \qquad \text{(x-Power estimate)}$$

$$w_{i+1}(k) = w_i(k) + \frac{\alpha}{\sigma_x^2 N} e(i)\, x(i-k)$$

In the system description, we did not assume that there is only a single talker. In practice, the algorithm converges much faster and to a better solution in the single-talker case than for the full-duplex situation. It is therefore recommended to identify the single-talker situation, and to update the filter coefficients only during this condition. The filtering operation and echo subtraction is performed continuously in both full-duplex and single-talker situations.

The single-talker situation is identified by comparing the estimated power at points x and d. A single talker is declared if $\sigma_x^2 > k\sigma_d^2$. k should be about four or more, since the hybrid attenuates the leakage by more than 6 dB.

Frequency-Domain Implementation

The frequency-domain implementation is equivalent to a block-adaptive filter, where the filter coefficients are changed only once for each block. In the frequency-domain implementation, both convolution (filtering) and correlation (coefficient updating) are done in the frequency domain, which is more efficient computationally than doing these tasks in the time domain.

Doing convolution in the frequency domain yields a circular convolution, not a linear one. To produce a linear convolution, an overlap and save (or overlap and discard) technique is used. [1,4] This means that for a filter of length N, a 2N Fourier transform is done, of which only the last N points are valid.

The following section presents the main points of an adaptive frequency-domain-based LMS algorithm. A detailed explanation and proof of convergence can be found in. [2]

The Adaptive Frequency-Domain Echo-Canceller

The algorithm is illustrated in Figure 4, and includes the following steps:

1. Read a block of new data (N points) from the signal x and signal d.
2. Form a 2N data block of x, by cascading the old data block with the new data block.
3. Calculate the estimated variance:

$$\sigma_x^2 = \frac{1}{N}\sum x_i^2$$

$$\sigma_d^2 = \frac{1}{N}\sum d_i^2$$

4. Do a 2N FFT on the block x:

   x(f) = FFT2N{x(t)}

5. Perform the filtering operation in the frequency domain:

   Y(f) = X(f) * W(f)

   then, transform the result back into the time domain:

   $y(t) = IFFT_{2N}\ (Y(f))$

6. Collect only the last N points of y(t), discarding the first N points.
7. Calculate the error data block, which is also the output:

   e(t) = d(t) – y(t)

8. Form a 2N-length error block by zero padding e(t) with N leading zeros, and perform a 2N-FFT.
9. If $\sigma_x^2 > 10\,\sigma_d^2$ (single-talker situation), then update the filter coefficients in the frequency domain:

$$W_{k+1} = W_k + \frac{\alpha\,\bar{x}(f)\,E(f)}{N\,\sigma_x^2}$$

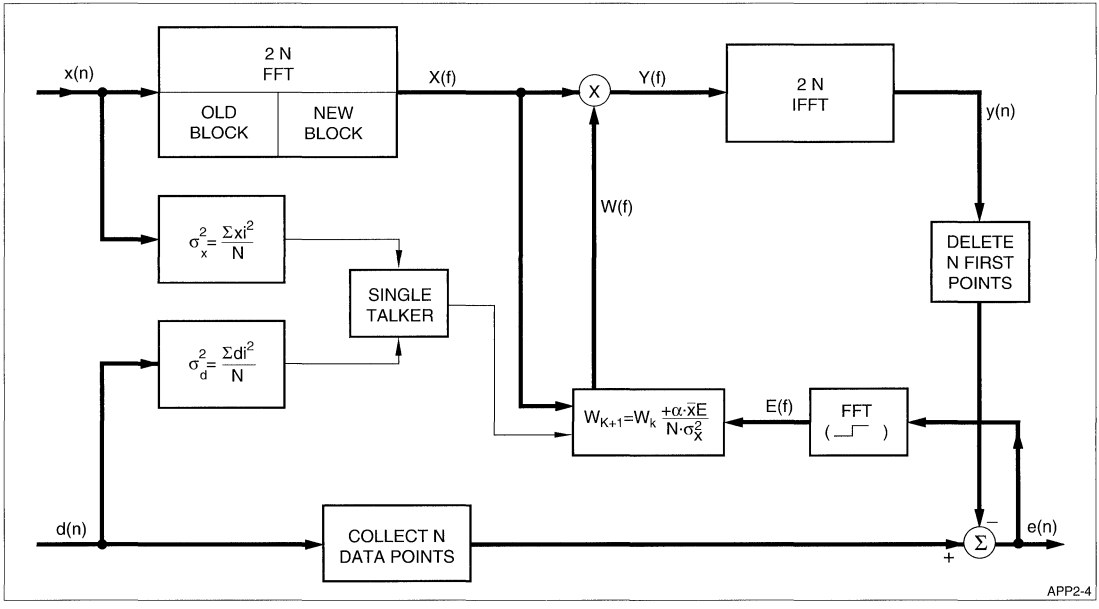


Figure 4. Adaptive Frequency-Domain Echo-Canceller

Hardware Implementation Using the LH9124 Digital Signal Processor

The LH9124 permits a straightforward implementation of the frequency-domain LMS adaptive echo-canceller, since each step of the algorithm corresponds to one or two LH9124 instructions (with proper addressing mode for the memories). Presented here is an implementation for 72 channels with a filter length of 128. Figure 5 illustrates a hardware system configuration with a filter length of 128.

Data Encoding and Decoding

Data coming from the T1 interface is PCM encoded in either A-Law or μ-Law, in order to reduce the bit rate on the line. Both A-Law and μ-Law are nonlinear functions which map 12 bits of linear dynamic range into eight bits. Since the echo-cancelling algorithm requires linear data, some form of translation is needed before the data is processed. This translation (or decoding and encoding) may be done simply with look-up tables.

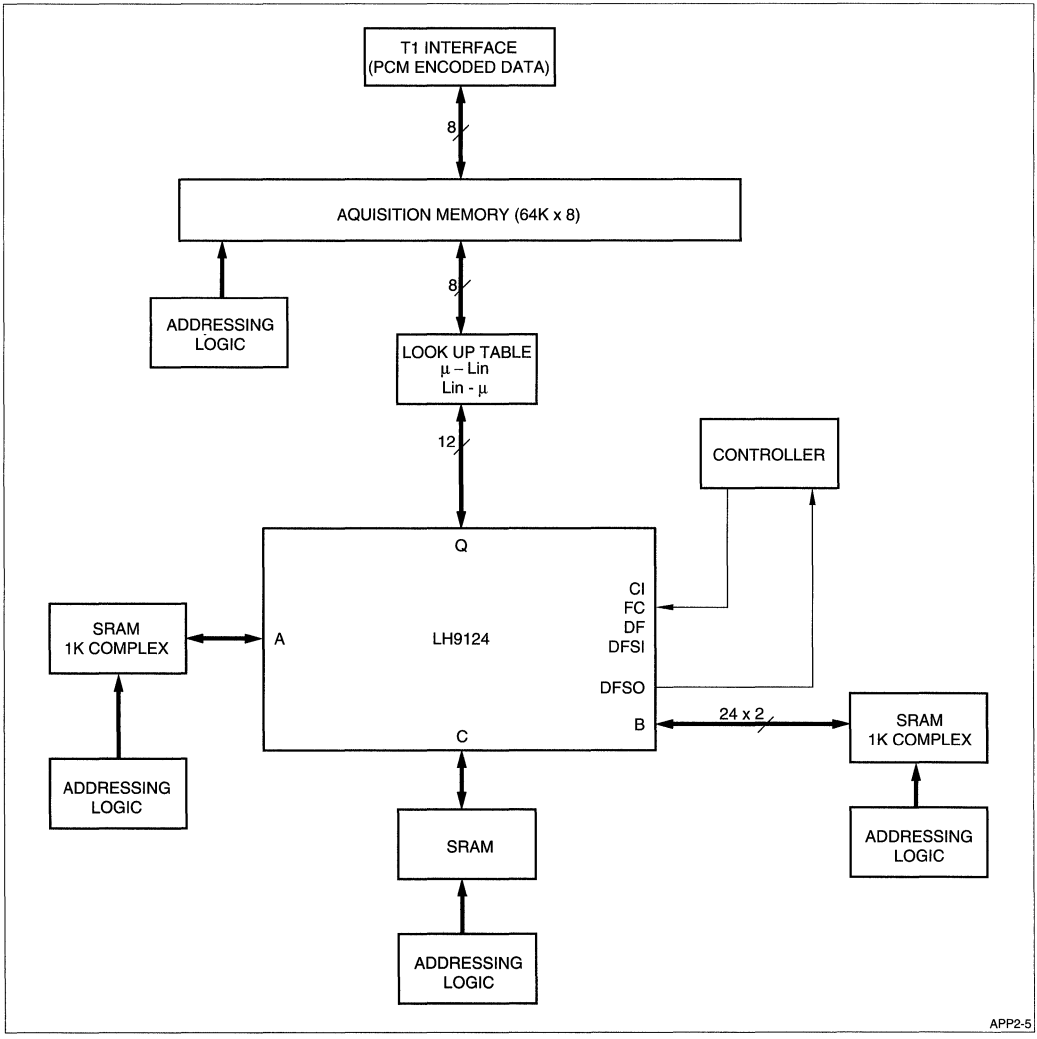


Figure 5. Hardware System Configuration

Addressing Logic

The frequency-domain adaptive-filter algorithm described above operates on blocks of data. Calculation is performed on blocks of 256 data points, while I/O is done on blocks 128 data points long.

Addressing within each block is not always in a consecutive sequence, as, for example, when calculating the FFT. The LH9320 Address Generator can be used to supply these special patterns with a lot of programming flexibility. However, since the algorithm requires non-trivial sequences for only a few limited cases, it is possible to calculate the required sequences off-line, storing them in SRAM for use in real-time operation. The SRAM can easily be loaded from a PROM or other non-volatile memory during the system initialization process. The LH9320 Real Time Simulator software can be used to calculate the required address sequences.

There are two additional requirements that the addressing logic must fulfill. First, it must compensate for the latency of the LH9124 when creating addresses for data output. Second, the addressing logic must handle the demands of multichannel operation: that is, each channel being processed has a different location in physical memory for its respective data block. In the present design, system memory is divided into blocks of 128 locations, corresponding to the smallest unit of block size used.

A functional block diagram of the addressing logic for the following options is shown in Figure 6:

- using the LH9320 AG
- using predefined sequences methods

The pre-calculated address patterns needed to run the algorithm are:

- Radix 16 – Pass 0, N = 256 with bit reverse
- Radix 16 – Pass 1, N = 256
- Continuous sequence 1-256
- Continuous sequence 1-128

The memory required to save the predefined sequences is less than 2K × 8.

Acquisition Memory

The LMS algorithm is the frequency-domain LMS operant on the reference input block of size 2N (2 × 128) with 50% overlap, which means that N points are old and N points are new. The memory space needed for each channel is 3 × N points, where 2N points are used as input to the algorithm, and N points are used to fill the new data for the next cycle.

The memory for the output and for d-input should be only 2N points for each channel, structured as a ping-pong buffer.

The multichannel memory management is the controller-task. However, it should be noted that the I/O is not done at high speed; the average I/O rate for 72 channels is 1.7 MHz. Thus, a simple controller can handle the task.

A conceptual structure for the acquisition memory is given in Figure 7.

The memory size for a 72-channel system is:

$(3 + 2 + 2) \times 72 \times 128 \times 8$ bit = 64K × 8 bit memory

This memory, in principle, is a dual-port memory; but, since the LH9124 is communicating with the memory in short bursts only, a SRAM with a FIFO on the interface side can be used by implementing simple arbitration.

Block Floating-Point Arithmetic

The algorithm is implemented using Block Floating-Point arithmetic. The controller is responsible for the management of the block exponents by reading the DSFO and by setting DSFI.

Port A and Port B Memories

These are small-size memories which contain temporary buffers. Each memory should be 1K × 24 complex. ('Complex' here implies two sets of memories.)

Port C

Port C contains three kinds of data:

Type a: Twiddle factor.
Window function.
(This data is loaded into this memory during the initialization process.)

Type b: The frequency-domain coefficient of the adaptive filters. Each channel has a 256 × complex coefficient.

Type c: Temporary buffers.

The memory size is 256 × 72 × 24 bit ≈ 36K × 24 bit.

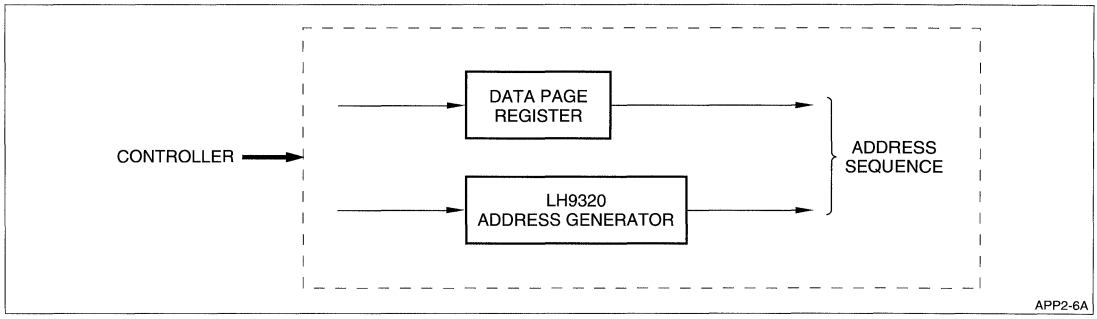


Figure 6A. Addressing Logic Using LH9320 AG

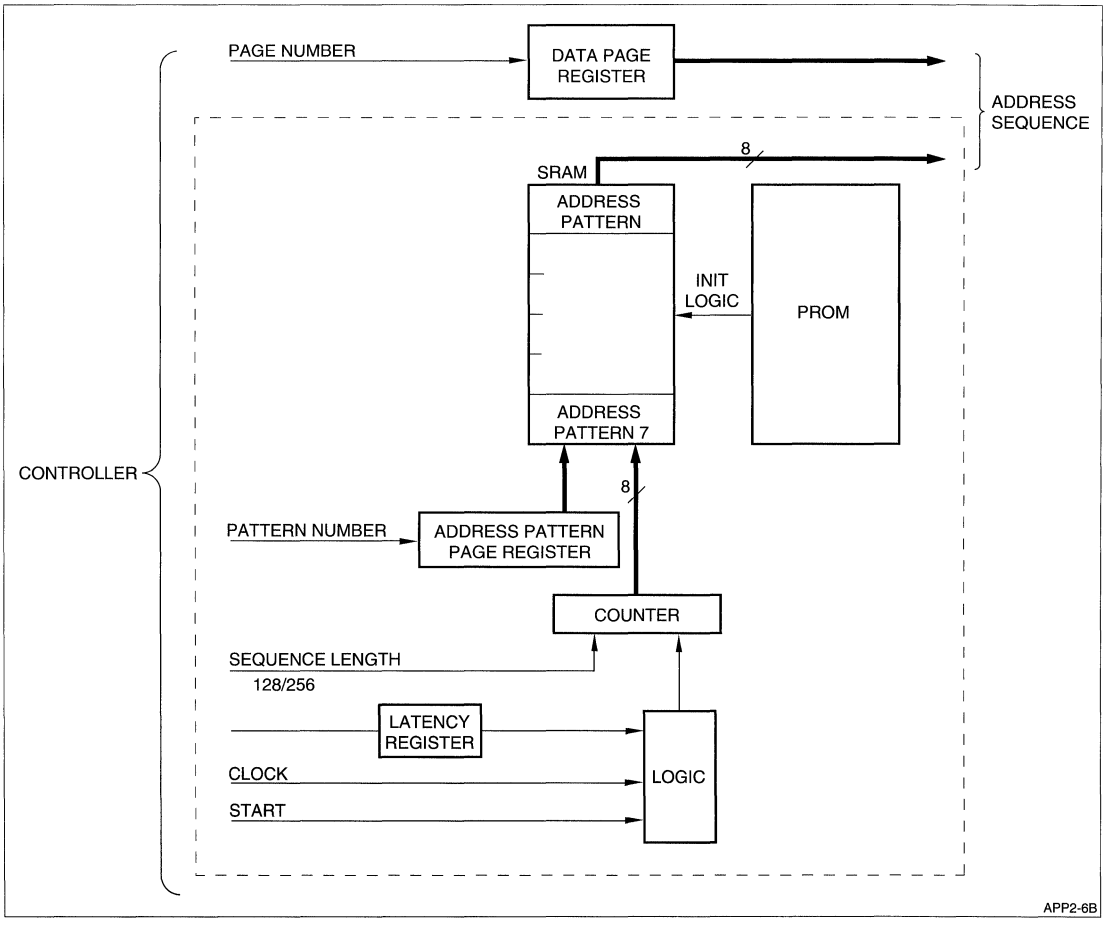


Figure 6B. Addressing Logic Using Predefined Stored Sequences

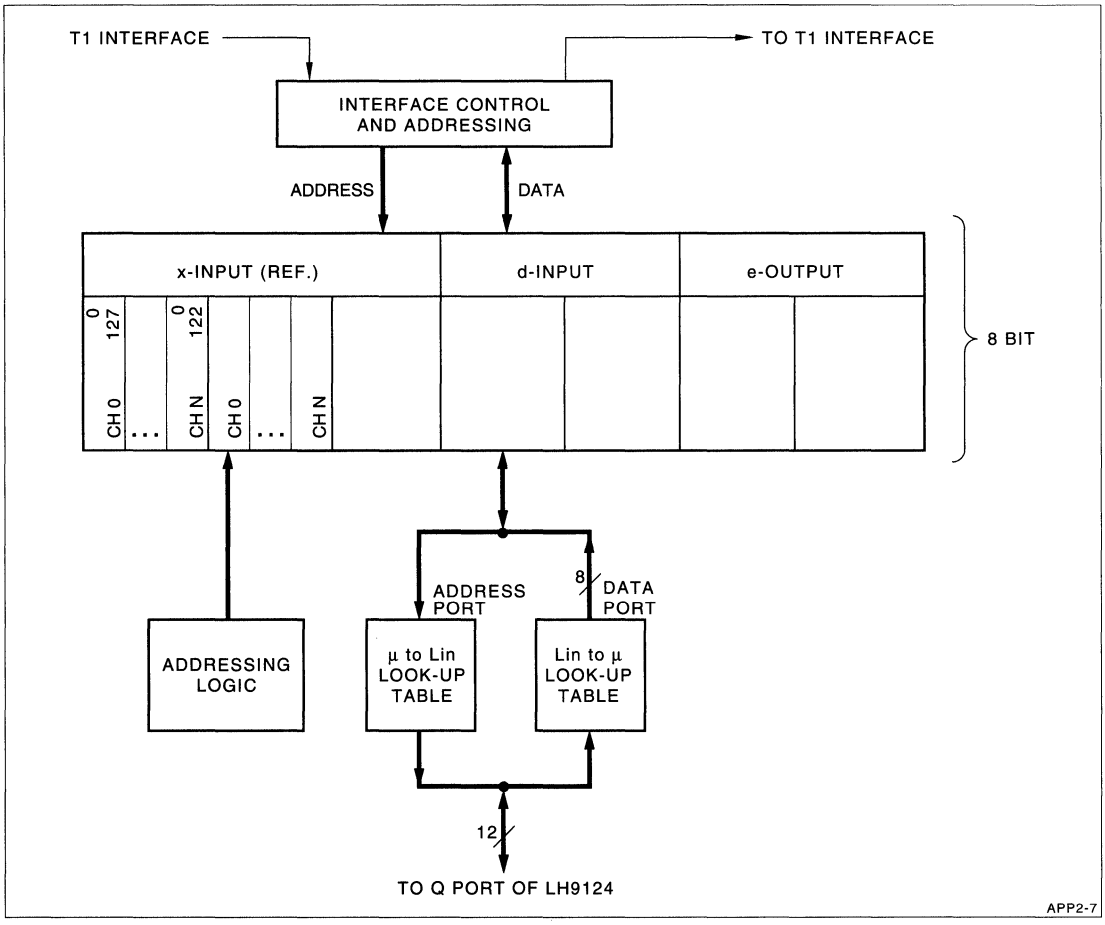


Figure 7. Acquisition Memory

Programming the Algorithm

Programming the algorithm using the LH9124 is straightforward, since each stage of the algorithm is mapped to just one or two instructions ('Pass-based' instruction). Table 1 shows the mapping of the algorithm to the LH9124-based system. Each line corresponds to one or two pass-based instructions of the LH9124, as explained below.

Line 1: Loading 2N points data x-input (received signal) from the acquisition memory to buffer A2 (in port A memory) using MOVD instruction.

Line 2: Calculating the variance of the new-data block variance. This calculation is done in two passes:

Pass 2.1: Move the data to port C memory (buffer C0).

Pass 2.2: Pass of length N of FIR operation using the BCFIR instruction, which gives

$$\sum_{i=0}^{N-1} a(n) \cdot c(n) = \sum x(m)\, x(m)$$

The result of this process is one number, which should be retained in the controller memory for later use.

Line 3: FFT transform of length 2N is done in two Radix 16 passes. The scaling of the first pass of the FFT is the DSFO value, as measured in the first stage. The second-pass scaling is the value of DSFO as measured in the first pass of the FFT. The accumulated scaling factor sx is saved in the controller memory.

Line 4: Complex vector multiplication pass, using the CMUL instruction. The vector W is the adaptive-filter frequency response, saved in the C port memory.

Line 5: IFFT operation is done by two Radix-16 passes. When doing the IFFT, manual scaling should be used. The total scale factor in the two passes should be 8-SH: the factor 8 is needed by the definition of the IFFT transform, and the -sx term is to compensate for the scaling that was applied during FFT(X).

Line 6: Loading d-input N-data points block from acquisition memory to port C memory (in buffer C2).

Line 7: Calculating the variance of d-input data block. (See the Line 2 explanation.)

Line 8: Calculating the error data block – which is the output data. This operation is done using the VSUB instruction.

Lines 9,10: This two-line MOVD instruction creates an error signal leading by N zeros in the buffer B2-B3.

Line 11: IFFT operation is done in two Radix-16 passes.

Line 12: Transferring buffer B2 (E(f)) into the C port memory – Buffer C0. This line is just a preparation to Line 13.

Line 13: Calculating $\frac{2\alpha}{\sigma_x^2 \cdot N} \bar{x}\, E$ is done by the CMUL instruction. The scale-factor multiplication is implemented using the DSFI signals. The controller should approximate σ_x^2 (calculated at line 2) to 2^{-KX}, so

$$\frac{2\alpha}{\sigma_x^2 \cdot N} = 2^{-10+KX}$$

The actual scale factor signal that should apply is K = sx + se + KX – 10,

where sx and se are the floating-point scale factors which were applied to X, and E during the FFTs.

Usually K is a negative number, which means that a scale-down operation is needed. In the case K > 0, no scale-down is done, which has the effect of stabilizing (and slowing the convergence rate) of the algorithm.

Computation Load

From Table 1, we can see that one channel computation time is about 100 µsec. All the channel computation times should not exceed the filter length $\frac{128}{8K} = 16\,\text{ms}$. Therefore, theoretically, one LH9124 can handle up to 160 channels. Practically, a design goal could be 100 channels.

If a longer filter length is needed, the computation load grows slightly. For example, if a filter length of 2048 points is needed, then the computation time of one channel is 1.8 ms, but the filter length is 256ms; therefore, the theoretical number of channels that can be handled by one chip is 140. It should be emphasized that the small sensitivity to the filter length is a feature of the LH9124 chip due to its increasing efficiency when doing higher-radix computations. With time-domain implementation using general-purpose DSP, increasing the filter length by a factor of 16 increases the computation load (or the hardware requirement) by a factor of 16.

As mentioned before, the filter introduces a delay in the telephone line. It is possible to reduce this delay, at the expense of increasing the computational load. The delay can be reduced by a factor of L, if the input sliding window is updated with every new $\frac{N}{L}$ point, and not with every N point. This modification increases the computational load by a factor of L.

Simulation

In order to demonstrate the performance of the frequency-domain algorithm, two simulation runs were performed using the MATLAB [1] language.

Case 1

The simulation condition:

- X-input – Normal random noise with variance of 1
- Channel leakage – Six reflections of different delays (a maximum delay of 10 ms)
- D-input is received by convolving X input with the channel transfer function

The simulation result and the program listing are given in Figure 8.

The plotted graph describes the transmitted signal. At a time close to time 0, we can see that the transmitted signal is high, which means a strong echo. At longer times, we see that the algorithms converge, and subtract the echo from the transmitted signal.

The time constant for this case is about 300 ms. An explanation of the program is given in Table 2.

For comparison, we implement the time-domain LMS, and the results are given in Figure 9. It may be seen that the algorithms give about the same results.

Case 2

The simulation condition:

- X-input – Sample of voice recorded from telephone line
- D-input received is as in the previous example

The simulation results are given in Figures 10 and 11.

Looking at the result shows that the algorithm performed very well under quite real conditions.

SUMMARY

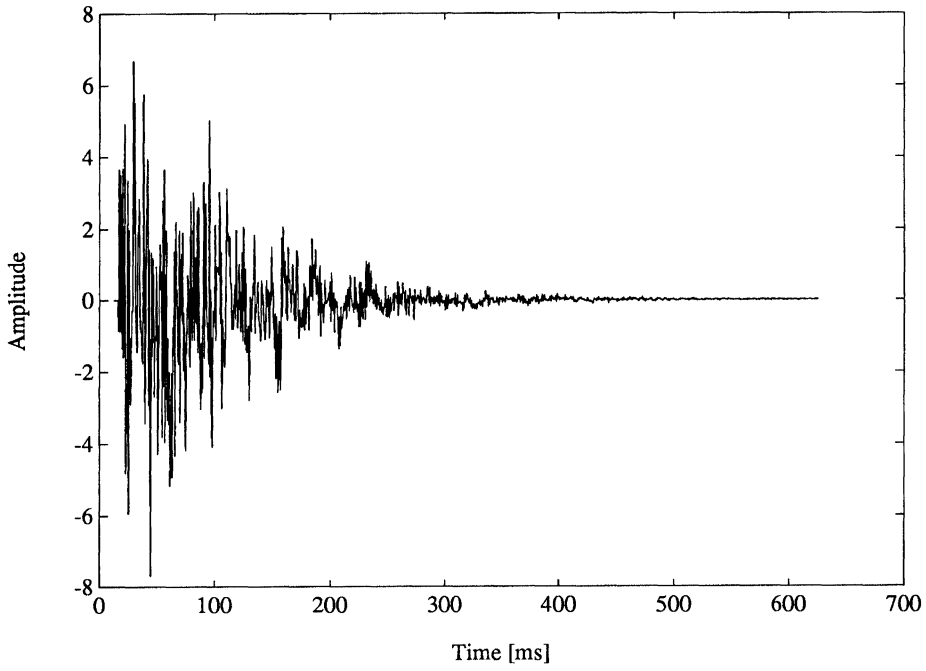
In this application note, we presented the frequency-domain LMS implementation based on the LH9124 Digital Signal Processor. The proposed solution is very hardware-efficient; it was shown that a practical design goal could be to implement 100 channels with one DSP. Furthermore, the hardware requirement has a very small sensitivity to the filter length.

The algorithm was tested using MATLAB simulation, to demonstrate the algorithm performance both for simulated white noise and for real speech data.

[1] MATLAB is a trademark of The Math Works, Inc.

Table 1. Mapping Frequency-Domain Echo-Cancelling to the LH9124

NO.	FUNCTION	BLOCK SIZE	TIME (μS)	REMARKS
1	In_x → A2	256	6.4	x(n) → A2 This operation is done in two stages: 1. Old block → A3 2. New block → A2 Save DSFO for proper scaling for the first pass of the next FFT.
2	$\sum A3^2/128 \rightarrow q\,1$	128	6.4	q1 is the variance of x = σ_x^2. The $\sum$ operation is accomplished by a FIR operation of x(m). By itself, the controller rounds $\sigma_x^2 \approx 2^{-kx}$.
3	FFT (A2) → A2, sx	256	14.8	FFT x(n) → A2 When doing the FFT, the controller sets the scale factor for the first pass of the FFT to be the value of DSFO as detected in stage 1 and uses DSFO for the second pass. sx is the accumulated scale factor.
4	A2 · W→ B2	256	6.4	y(f) = x(f) · w
5	IFFT (B2) → B2	256	14.8	y(n) → B2 When doing the IFFT, the controller should apply a scale factor of sx-8, to compensate for sx and 1/256 needed in the IFFT.
6	In_d → C2	128	3.2	d(n) → C2
7	$\sum C2/128 \rightarrow q2$	128	3.2	$\sigma_d^2 \rightarrow q2$
8	C2 – B3 → Out	128	3.2	e(n) = d(n) – y(n) This is the output signal.
9	'O' → B2	128	3.2	To form the error signal padded by N leading zeros. The controller should save DSFO for proper scaling of the FFT.
10	Out → B3	128	3.2	
11	FFT (B2) → B2, se	256	14.8	E(f) → B2, see remarks of step 4.
12	B → CO	256	6.4	$W_{k+1} = W_k + \frac{2\alpha \cdot X \cdot E}{\sigma_x^2 \cdot N}$
13	A2 · CO → BO			
14	W + BO → AO	256	6.4	k = sx + se + kx -10 where sx, se are to compensate for the shift in X and E during FFT, kx is the normalizing to σx^2. The step size factor $2\alpha/N = 2^{-10}$ for $\alpha = 0.1$ NOTE: Normally k is negative, which means scale-down. For the case k > O, no scale factor is implemented, which has the effect of stabilizing (and slowing) the algorithm.
15	if q1 > 1oq2 → → AO → W	256	6.4	If single talker, update coefficient. This decision is made in the controller.



SIMULATION
(Produces Input Data)

```
1. rand('normal');
2. rand('seed',0);
3. x=rand(1,10240);
4. e=zeros(1,128);
5. b=zeros(1,128);
6. b([4 10 20 44 54 78 ])=rand(1,6);
7. d=conv(x,b);
8. d=d(257:10240);
9. x=x(257:10240);
10. w=zeros(1,256);
```

MAIN PROGRAM

```
1. for k=2:20
2. xk=x((k-2)*128+1:k*128);
3. dk=d((k-1)*128+1:k*128);
4. xf=fft(xk);
5. yf=xf.*w;
6. yk=real(ifft(yf));
7. yk=yk(129:256);
8. ek=dk-yk;
9. e=[e,ek];
10. ef=fft([zeros(1,128),ek]);
11. w=w+0.2*conj(xf).*ef/128;
12. end
```

Figure 8. Frequency-Domain Echo-Canceller (Refer to Table 2 for Program Explanation)

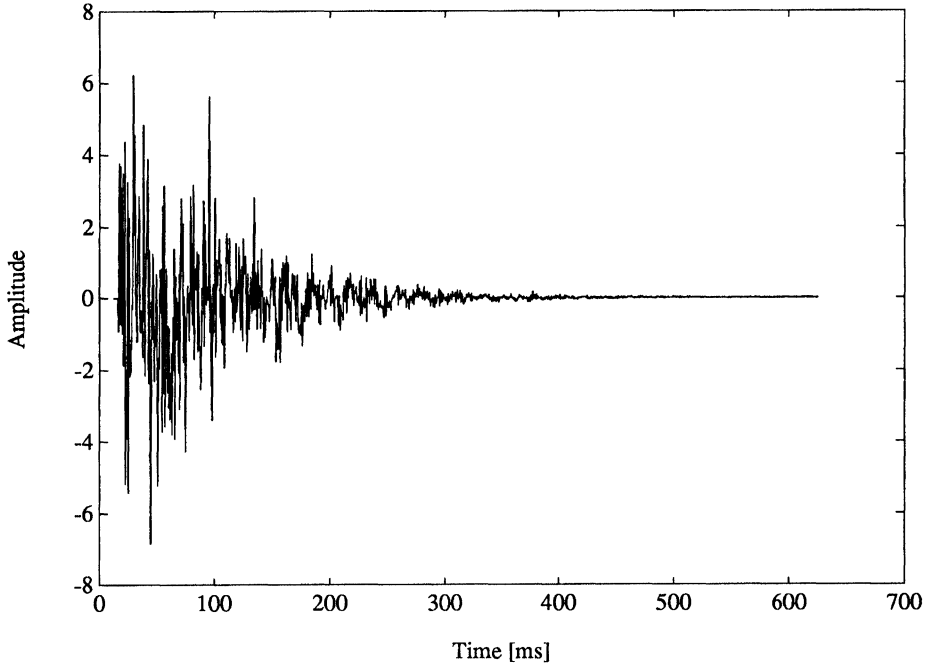
Table 2. Frequency-Domain Echo-Canceller

SIMULATION OF INPUT DATA

1. Specify normal probability distribution.
2. Initial condition for random function.
3. Fill vector with random numbers.
4. Set initial conditions.
5. Set initial conditions.
6. Select random echo-transfer function.
7. Convolve input with filter to produce echo.
8. Start after filter transient.
9. Start after filter transient.
10. Initial condition for adaptive filter.

MAIN PROGRAM

1. Set up loop to handle 19 blocks of data.
2. Read reference input data block of length 2N.
3. Read echo signal data block of length N.
4. FFT of input reference block.
5. Frequency-domain filtering (multiplication).
6. Do inverse FFT, transform back to time domain.
7. Select last N points of the filtered data block.
8. Calculate error (output).
9. Build output vector.
10. FFT of leading-zero-padded error signal.
11. Update filter coefficients.
12. Go back to Line 1 for block.



SIMULATION

```
1. l=10240+256;
2. rand('normal');
3. rand('seed',0);
4. x=rand(1,10240);
5. e=zeros(1,100);
6. b=zeros(1,128);
7. b([4 10 20 44 54 78 ])=rand(1,6);
8. d=conv(x,b);
9. d=d(257:10240);
10. x=x(257:10240);
11. w=zeros(1,128);
```

MAIN PROGRAM

```
1. for n=128:2500
2. xn=x(n-127:n);
3. yn=w*xn';
4. e(n)=d(n)-yn;
5. w=w+0.2*xn*e(n)/128;
6. end
```

Figure 9. Time-Domain Echo-Canceller

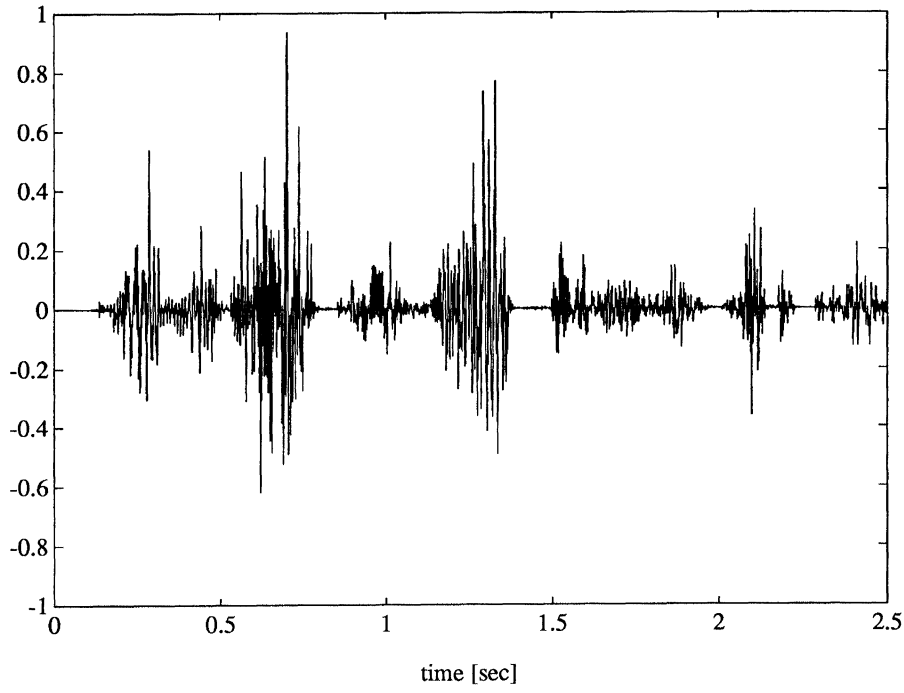


Figure 10. Echo Before Adaptive Cancellation

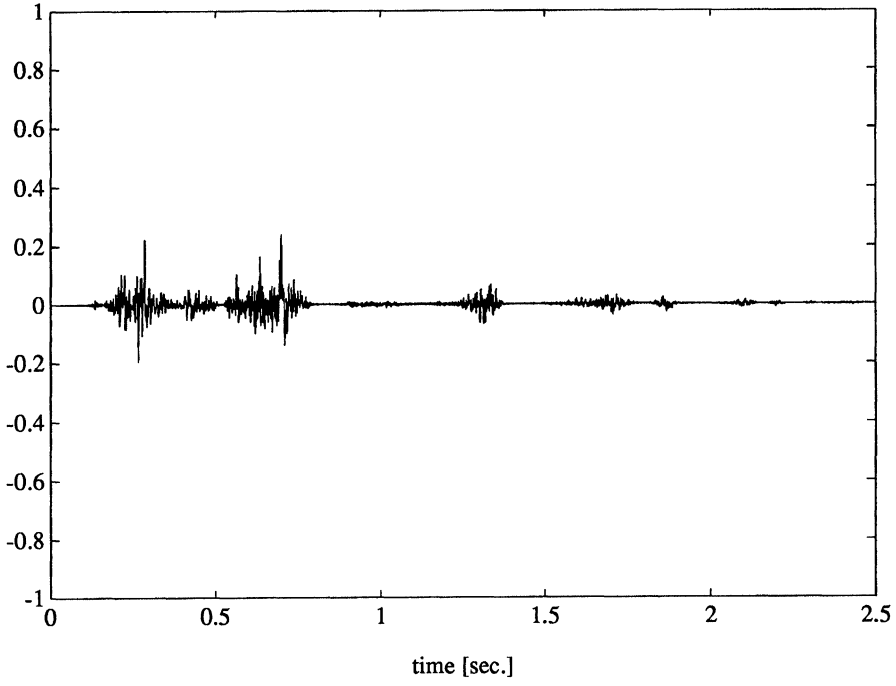


Figure 11. Residual Echo After Adaptive Cancellation

REFERENCES

1. Fast Implementation of LMS Adaptive Filters, Earl R. Ferrara. *IEEE Transactions on Acoustics and Signal Processing*, Vol ASSP-28, PP 474-475, August 1980.
2. Unconstrained Frequency Domain Adaptive Filters, David Mansour and Augustine H. Grays, *IEEE Transactions on Acoustics and Signal Processing*, Vol ASSP-30, No. 5, October 1982.
3. Adaptive Signal Processing, Bernard Widrow and Samuel D. Stearns, Prentice-Hall, Inc., Englewood Cliffs, NJ.
4. Digital Signal Processing, Alan V. Oppenheim and Ronald W. Schafer, Prentice-Hall, Inc., Englewood Cliffs, NJ.
5. Echo Cancellation in Speech and Data Transmission, David Messerschmit, *Journal on Selected Areas in Communications*, Vol SAC-2, No. 2, PP 283-297, March 1984.

Controlling the Falling Edge of $\overline{\text{MEMW}}$ With the Resistor Value of RPROG

INTRODUCTION

The LH9320 Address Generator has a built-in delay circuit feature that allows users to control the falling edge of the $\overline{\text{MEMW}}$ signal by changing RPROG's resistor value (external resistor). This feature saves board space and also reduces components by generating the WRITE enable signal for the memories.

Since the falling edge of SYSCLK has control over the falling edge of the $\overline{\text{MEMW}}$ signal, the maximum delay time, controlled by RPROG, is limited by the falling edge of SYSCLK plus 5 ns.

The $\overline{\text{MEMW}}$ signal is generated from both the SYSCLK and the delay circuit (Figure 1). The AND gate generates a falling edge of $\overline{\text{MEMW}}$ either from the SYSCLK driver or from the delay circuit, whichever arrives to an input first (Figure 2). NOTE: bit 7 in the pipeline/memory must be programmed to '0' to generate the $\overline{\text{MEMW}}$ pulse.

The delay times from the rising edge of SYSCLK to the falling edge of $\overline{\text{MEMW}}$ are proportional to RPROG's resistor value (Figure 3).

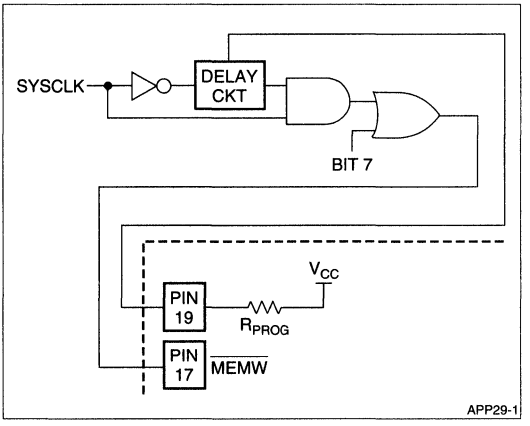


Figure 1. $\overline{\text{MEMW}}$ Signal Generation

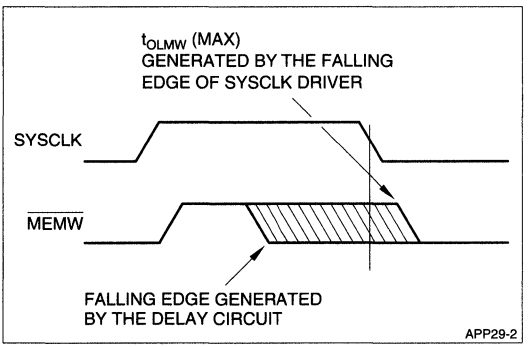


Figure 2. $\overline{\text{MEMW}}$ Pulse Generation

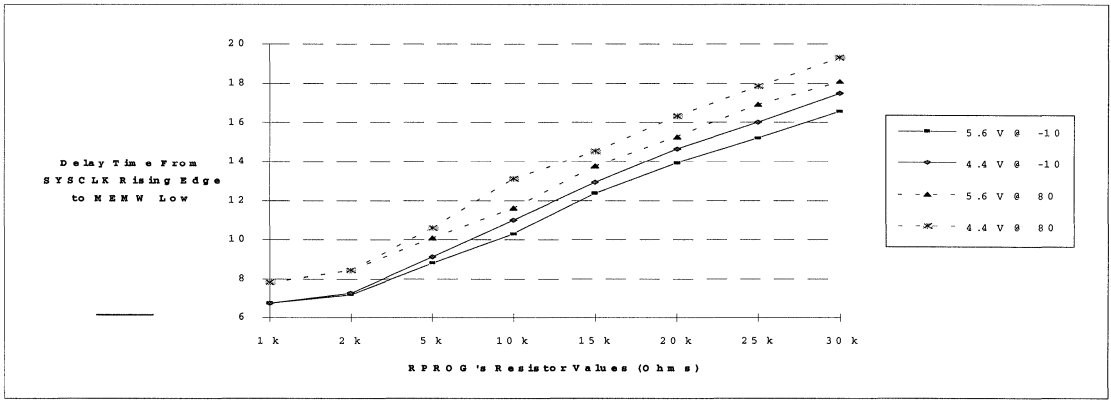


Figure 3. RPROG's Resistor Value vs. $\overline{\text{MEMW}}$ Low Time

An Ultra-High Speed DSP Chip Set For Real-Time Applications

Michael E. Fleming
SHARP Microelectronics Technology Inc.
5700 NW Pacific Rim Blvd.
Camas, Washington 98607

We are in the age of multi-million transistor integrated circuits with 10's of millions on the horizon. This ability to get a tremendous amount of functionality on one piece of silicon dictates that the silicon manufacturers begin to think in 'systems terms' instead of 'component terms' if they want to have a chance of being the company to introduce the next Weitek floating point, TI digital signal processor, or Intel microprocessor success.

The gate array and ASIC businesses attempt to place the silicon power in the hands of the system designer with ever increasing CAD and functional cell capability. The silicon vendor late to the ASIC game or not setup to compete needs to rely on bringing a cost-effective "system solution" that makes ultimate utilization of every transistor and internal bus structure.

SHARP's LH9124/LH9320 chip set applies this "the system is the chip" solution to a category of DSP applications that can be best described as the replacement of classical analog signal processing and conditioning with digital signal processing for real-time applications. A single LH9124 system performs a 1024 point complex FFT in 80.7 microseconds; three LH9124s cascaded in 25 microseconds.

The advantages of digital signal processing over analog signal processing are rapidly becoming accepted without examination. A few of these advantages include: no aging or temperature variation of components, exact repeatability in production without the usually required final adjustments, and the ability to apply mathematical finess to the signal processing that previously was very limited with analog techniques.

Most real time DSP algorithms can be reduced to a few high level functions such as:

- Digital filtering
- Time-to-frequency transformation
- Signal detection
- Frequency-to-time transformation
- Data buffering
- Signal comparison
- Signal modulation/demodulation
- Correlation

These functions, when combined in specific order and specific array sizes, can perform powerful DSP tasks such as:

- Machine vision
- Error correction
- Data communications
- Machine health monitoring
- Speech recognition
- Full motion video
- Spread spectrum communications

The industry benchmark for DSP performance is the 1024 point complex FFT execution time. Table 1 gives some relative perspective on how the new SHARP LH9124 performs against other possible solutions.

Table 1.
Benchmark 1024 Point Complex FFT

PROCESSOR	TIME	PRECISION	FACTOR
80386 (20MHz)	200 msec	16-bit fixed point	1X
68030 (20MHz)	150 msec	16-bit fixed point	1.5X
VAX 11/780	150 msec	16-bit fixed point	1.5X
TMS 320C25	15.8 msec	16-bit fixed point	12.6X
TMS 320C30	2.5 msec	32-bit floating point	80X
ADSP 2100 (8MHz)	7 msec	16-bit fixed point	28.6X
MOTO 56001	5 msec	24-bit fixed point	40X
MOTO 96001	2 msec	32-bit floating point	100X
CRAY X-MP	1 msec	64-bit floating point	200X
SHARP LH9124 (40MHz)	80 µsec	24-bit fixed point	2500X

Reprint: International Conference on DSP Applications and Technology – Berlin Oct 91

Table 2 shows execution times for both the single chip and the multi-chip (cascaded) LH9124 implementations for various common DSP tasks.

Table 2.
24-Bit Performance Benchmarks

FUNCTION	PERFORMANCE @ 40MHz		UNITS
	SINGLE	CASCADED	
1K Complex FFT	80.7	25.6 (3 Stages)	µs
4K Complex FFT	312.3	102.4 (3 Stages)	µs
64 × 64 2-D Complex FFT	413.9	102.4 (4 Stages)	µs
256-Tap Complex FIR	156	–	kHz
3 × 3 Tap FIR	8	–	MHz
10 × 10 Matrix Multiply	12.5	–	µs
Complex Integer Multiply	25	–	µs

A typical LH9124/LH9320 system level block diagram is shown in Figure 1. The main components to be discussed are:

- LH9320 Address Generator
- LH9124 Digital Signal Processor
- System Memory
- Scheduler Unit

LH9320 ADDRESS GENERATOR

The LH9320 AG shown in Figure 2 is designed for use with the LH9124 DSP and can be used with almost any pipelined processor capable of utilizing the set of over 150 generated address patterns.

Optimized for FFTs, FCTs, DFTs, DCTs, and FIRs the LH9320 generates addresses at a 40MHz rate with virtually no overhead. A summary of the address patterns is shown in Table 3. These patterns allow arbitrary mixing of radix-2, radix-4, and radix-16 multi-channel sample arbitration and circular buffering, real only transforms, etc.

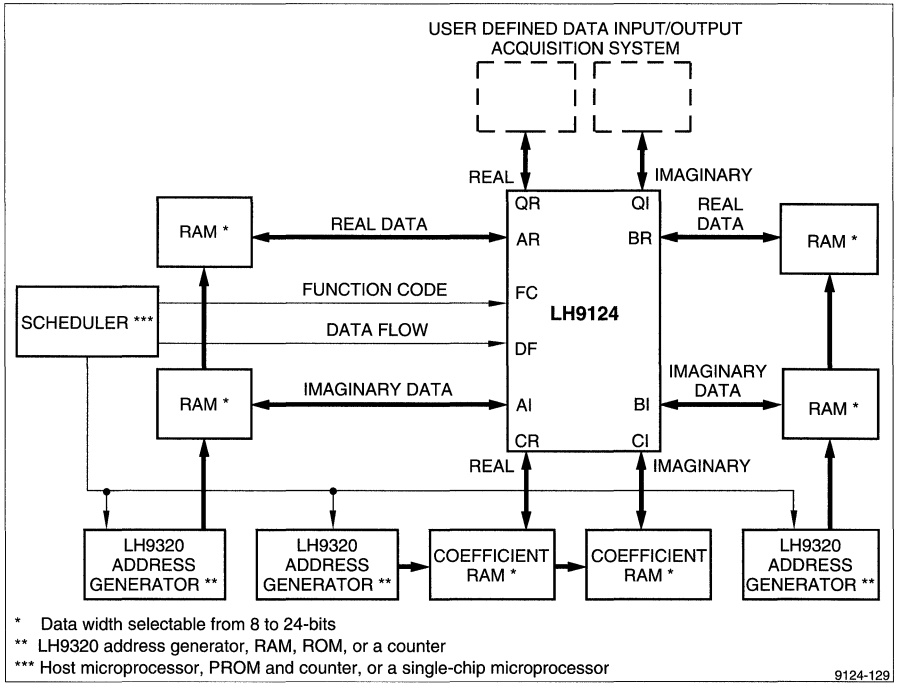


Figure 1. Typical LH9124/LH9320 System Block Diagram

Table 3.
LH9320 Address Pattern Set Summary

MNEMONIC	DESCRIPTION
FAST FOURIER TRANSFORMS (FFTs)	
BF2nn (0 to 19)	Radix-2 data addresses
BF4n (0 to 9)	Radix-4 data addresses
BF16n (0 to 4)	Radix-16 data addresses
TF2n (0 to 19)	Radix-2 twiddle factor addresses
TF4n (0 to 9)	Radix-4 twiddle factor addresses
TF16n (0 to 4)	Quasi radix-16 twiddle factor addresses
MXB24n (0 to 8)	Mixed radix (2,4) data addresses
MXB216n (0 to 3)	Mixed radix (2,16) data addresses
MXB416n (0 to 3)	Mixed radix (4,16) data addresses
MXB2416n (0 to 3)	Mixed radix (2,4,16) data addresses
MXT24n (0 to 8)	Mixed radix (2,4) twiddle addresses
MXT216n (0 to 3)	Mixed radix (2,16) twiddle addresses
MXT416n (0 to 3)	Mixed radix (4,16) twiddle addresses
MXT2416n (0 to 3)	Mixed radix (2,4,16) twiddle addresses
RBF0	Digit-reversed data address column 0
SEPARATION PASSES	
BRFTL	2-at-a-time real FFT separation pass, 2N, load
BRFTLS	2-at-a-time real FFT separation pass, N, load
BRFTU	2-at-a-time real FFT separation pass, 2N, unload
BRFTUS	2-at-a-time real FFT separation pass, N, unload
BFCTL	Fast cosine transform separation pass data addresses, 2N (long), load
BFCTT	Fast cosine transform separation pass twiddle addresses, 2N
BFCTUS	Fast cosine transform separation pass data addresses, N (short), load
BFCTU	Fast cosine transform separation pass data addresses, 2N, unload
BFCTUP	Fast cosine transform separation pass data addresses, 2N, unload using PO flag
BFCTUEP	Fast cosine transform separation pass data addresses, 2N, unload using early PO flag
BFCTULP	Fast cosine transform separation pass data addresses, 2N (long), unload using late PO flag
DECIMATION	
DECIM	Decimate
ADECIM	Auto decimate

MNEMONIC	DESCRIPTION
FINITE IMPULSE RESPONSE (FIR) FILTERS	
LPFIR1	Linear phase FIR, odd length, even symmetry
LPFIR2	Linear phase FIR, even length, even symmetry
LPFIR3	Linear phase FIR, odd length, odd symmetry
LPFIR4	Linear phase FIR, even length, odd symmetry
GENERAL PURPOSE ADDRESSING/UTILITIES	
MODINC	Modulo increment
MODDEC	Modulo decrement
INTER	Interpolate/index fill
INTERP	Interpolate/index fill using PO flag
INTEREP	Interpolate/index fill using early PO flag
INTERLP	Interpolate/index fill using late PO flag
PADHIGH	Pad at end of sequence
PADHIGHP	Pad at end of sequence using PO flag
PADHIGHEP	Pad at end of sequence using early PO flag
PADHIGHLP	Pad at end of sequence using late PO flag
PADLOW	Pad at start of sequence
PADLOWP	Pad at start of sequence using PO flag
PADLOWEP	Pad at start of sequence using early PO flag
PADLOWLP	Pad at start of sequence using late PO flag
OVERLAP	Overlap
DISCARD	Discard
DISCARDP	Discard using PO flag
DISCARDEP	Discard using early PO flag
DISCARDLP	Discard using late PO flag
CMAG	Square of magnitude of a complex number
INC	Index Increment
NOP	No operation
CIRCULAR BUFFER ADDRESSING	
CBUFFIR	Circular buffering for FIRs
CBUFFFT	Circular buffering for FFTs
OTHER	
CLRSIG	Clear signature
VIEWSIG	View signature

LH9320 ADDRESS GENERATOR (cont'd)

The LH9320 AG supplies a series of addresses to any array of memory for use in conjunction with an execution unit. The AG is programmed by the user and uses a minimum amount of glue logic, making it independent of the execution unit (LH9124), in most cases.

The AG has two modes, local pass mode and external pass mode. In local pass mode the user programs address patterns in the AG's internal memory (local memory), which can handle operation codes up to 32 patterns. The user must also set up the appropriate status values (other than the default) in the AG's twenty configuration registers. Once the scheduler gives the START signal, the AG begins to execute and generate arrays of addresses according to the address patterns programmed in the local memory. The generated address arrays are output through the 20-bit address bus.

For some sophisticated processing, where the number of address patterns exceed 32, local memory space may not be sufficient. The AG then switches to external program mode and accesses external memory space to read the address patterns. The AG accesses the address patterns in the external memory through data bus DB0-DB7. The address calculation logic is output as address arrays to Q, A, B, or C memory for the LH9124 execution unit.

The LH9320 also utilizes a powerful circular buffer technique for multiple-channel processing. Circular buffering is used extensively in real time DSP systems. Spectral accuracy can be improved if the incoming data is overlapped by a small percentage before the transform takes place, as in circular convolution, FIR structures, and time domain correlation. This circular buffer is applied usually when the system is used in a multiple-channel configuration.

Using the LH9320

The LH9320 can perform three different functions:

- Acquisition address generator
- Data address generator
- Coefficient address generator

Figure 2 shows a simple system and the related connections. The LH9320 is a small pin count addressing device that is initiated and controlled through a series of programmable commands that configure the AG and specify the address patterns to be generated over the entire length of the memory array. The AG then outputs a Terminal Count (TC) signal and enters an idle state until the next START signal is received. PO can go to the scheduler or the LH9124's CCI, CCR, ZEROOUT, or ZEROIN, depending on the DSP algorithm. If only one of these signals is needed for the whole program, they can all be tied together. Otherwise the scheduler must multiplex them into the correct location.

Acquisition Address Generator

As shown in Figure 2, the general configuration for an AG includes the following connections:

- Sample request input lines from the input channels or decode unit
- System clock
- Control signals from the scheduler
- Address register/data latch (A0, A1) control from the scheduler
- Data bus connection to the scheduler
- Terminal count (TC) connection to the scheduler or LH9124
- Program out (PO) connected to the LH9124 DSP
- Read clock (RDCLK) output connected to the LH9124
- Write enable ($\overline{WE}$) signal connected to RAM
- Address bus connected to RAM port

When an LH9320 is used as an acquisition AG for multiple-channel processing, the input lines ASG0-ASG4 are used for circular buffer input. When there are ≤ 5 multiple channel numbers, ASG0-ASG4 connect directly to each channel. If there are > 5 channel numbers, those 5 pins connect to the decode unit. This unit can have up to 32 input pins, each one connected to a separate channel, and five output pins connected to AGS0-AGS4.

Data Address Generator

The configuration of the data AG is generally the same as the acquisition and coefficient AG's, however, the input lines AGS0-AGS4 are always left idle.

Coefficient Address Generator

The configuration of the coefficient AG is similar to that of data AG, however, the input lines AGS0-AGS4 are idle.

Each AG has separate lines to the scheduler for $\overline{CS}$ and TC. Note that all of the AGs share a common data bus and R/$\overline{W}$ line. They also share SYSCLK and START signals with the LH9124 DSP.

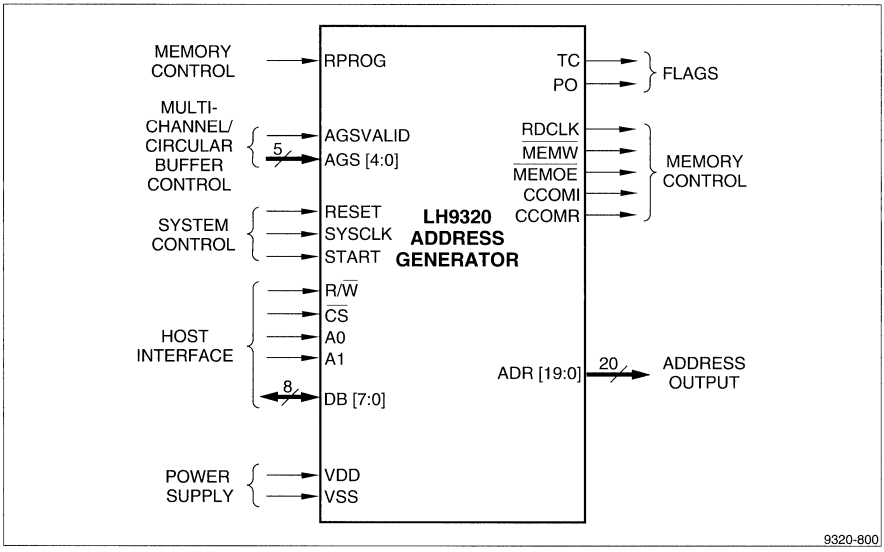


Figure 2. LH9320 DSP Address Generator

LH9124 DIGITAL SIGNAL PROCESSOR

The LH9124 is a high-performance DSP device that is suitable for most high-performance DSP algorithms.

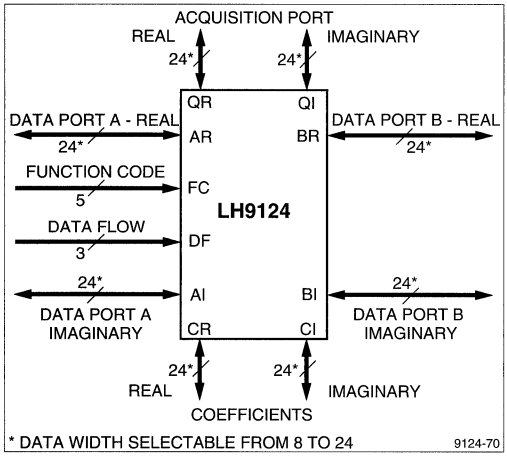


Figure 3. LH9124 Digital Signal Processor

The LH9124 performs the following types of functions as detailed in Table 4:

- Digital signal processing
- Complex arithmetic functions
- Vector arithmetic
- Vector logical
- Matrix arithmetic
- General purpose functions

Each of these functions requires an address pattern to select data from memory for use by the DSP. The scheduler unit outputs the necessary instruction codes to the LH9124.

The LH9124 uses a multi-port data flow structure that frees the user from externally multiplexing data, thereby increasing throughput and minimizing system component count. Each of the bidirectional LH9124 ports and their functions are listed as follows:

- AR – real data port
- AI – imaginary data port
- BR – real data port
- BI – imaginary data port
- QR – real acquisition port
- QI – imaginary acquisition port
- CR – real coefficient port
- CI – imaginary coefficient port

The 16- to 48-bit fixed point data (8- to 24-bit real and 8- to 24-bit imaginary) enters and exits the processor through the bidirectional data and acquisition ports.

Table 4.
Function Codes

FUNCTION MNEMONIC	FC OPCODE (HEX)	DESCRIPTION
		DSP FUNCTIONS
BFLY2	02	Radix-2 butterflies.
BFLY4	01	Radix-4 butterfly.
BFLY16	00	Radix-16 butterfly.
BWND2	05	Radix-2 complex window butterflies. Performs a radix-2 butterflies with a complex window function that is multiplied with the incoming data.
BWND4	04	Radix-4 complex window butterfly. Performs radix-4 butterfly with a complex window function that is multiplied with the incoming data.
BRFT	07	Dual real FFT separation. Performs the separation of two real data streams for a two-at-a-time FFT operation.
BFCT	06	Fast cosine transform/double length (N output). Performs the fast cosine transform (FCT) operation.
BFCT2	0E	Fast cosine transform/double length (2N output). Performs the fast cosine transform double length operation.
BCFIR	08	Complex finite impulse response (FIR) filter.
BDFIR	09	Double real finite impulse response (FIR) filter.
BRFIR	0A	Real finite impulse response (FIR) filter.
		COMPLEX ARITHMETIC FUNCTIONS
CADD	10	Complex add. Adds complex input data to complex coefficient data.
CMAG	0C	Complex magnitude squared. Performs the magnitude squared of an input.
CMUL	0D	Complex multiply. Multiplies complex input data by complex coefficient data.
CSUB	11	Complex subtract. Subtracts complex coefficient data from complex input data.
		VECTOR ARITHMETIC FUNCTIONS
VABS	13	Vector absolute value. Determines the absolute value of the input data.
VADD	10	Vector add. Adds the input data pairs to the coefficient data pairs.
VMUL	12	Vector multiply. Multiplies the input data by the coefficient data.
VMXM	1E	Vector maximum and minimum. Determines the maximum/minimum vector of the input data.
VSUB	11	Vector subtract. Subtracts the input data from the coefficient data.
		VECTOR LOGICAL FUNCTIONS
VNAND	18	Vector logical NAND. Performs a logical NAND between the input data and the coefficient data.
VNOR	1A	Vector logical NOR. Performs a logical NOR between the input data pairs and the coefficient pairs.
VPAS	19	Vector logical pass. Performs a VPAS on the input data.
VXNOR	1B	Vector logical exclusive NOR. Performs a logical exclusive NOR between the input data pairs and the coefficient pairs.
		GENERAL PURPOSE FUNCTIONS
MOVC	1D	Move coefficient data from the coefficient port to another (A,B,Q) port.
MOVD	1C	Move data from one port (A,B,C,Q) to another port (A,B,C,Q).

The acquisition ports and related memory provide communication with the external system. The signal data is mapped into the acquisition RAM before execution begins. Processing begins by reading the signal data through the QR and QI ports. During processing, the data recursively exchanges location between the AR, AI and BR, BI dual memory system. The data is arithmetically processed using the coefficient data which enters via the CR and CI ports. After processing is complete, the data is written to the acquisition RAM through the QR and QI ports, and is ready for output to the external devices or the system. Data can be routed through any one of 11 data flow paths.

These features give the LH9124 versatility in a variety of applications. Essentially, the LH9124 system can be configured to provide three levels of performance:

- a high-performance system using multiple parallel and/or cascaded LH9124 units;
- a medium-level system using dual memory and one LH9124 unit;
- a low-performance system using a single memory stage and one LH9124 unit.

Each LH9124 unit that is added to the system improves the performance of that system by a factor of one. For example, a 1K complex FFT performed using one LH9124 would take approximately 80µs while the same calculation could be done in approximately 25µs on a system utilizing three parallel LH9124 devices as shown in Figure 4, thus producing a three-fold improvement in performance. The efficiency of the system would show an even greater improvement in a cascaded configuration due to decreased latency.

A system based on the LH9124 unit can be configured using SHARP's LH9320 address generators which are companion devices to the LH9124. The address generators and programmable memory schedulers are vital for memory address generation and are responsible for managing system memory addresses. The control logic downloaded from ROM during initialization determines the management task that the address generators will perform. The setup/control logic also defines the algorithms to be executed by the DSP system.

System Bus Operation

Bus operation for the LH9124 DSP system can be divided into three main categories:

- a high-speed 48-bit complex data bus that routes signal data between the LH9124 and memory/acquisition hardware;
- several high speed 20-bit address buses;
- a low speed control bus.

The LH9124 performs all signal data routing between the different data ports. External signal data multiplexing to the LH9124 is confined to acquisition port Q. This permits the memories on ports A and B to be connected directly to the LH9124, thereby reducing output loads and allowing functions to proceed between the A and B ports at the maximum clock rate.

While processing between ports A and B, data is input and output asynchronously via the Q port memories and associated hardware. Input data, once gathered, can be loaded synchronously through the LH9124 to the high speed A and B ports. Similarly, processed data from ports A and B is output synchronously to Q port memory for subsequent output.

The address signals for all memories are loaded or generated locally to each memory group via the LH9320 address generator, or from counters, sequencers, PROMs, and/or the LH9320 address generator. Depend-

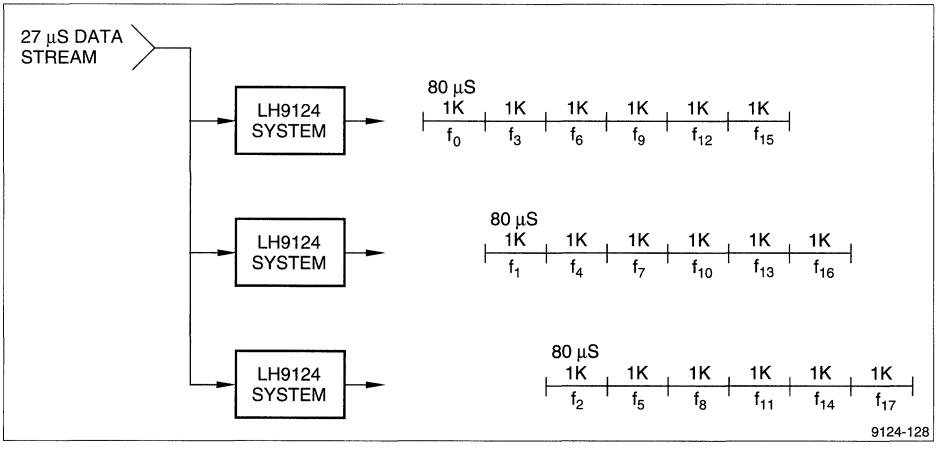


Figure 4. Parallel LH9124 System Processing

ing on the number of memory units to be addressed, buffering can also be required.

Control signals for the LH9124 system are low speed since each is changed on a per pass basis. Thus, when operating on N length data sets, the scheduler for the LH9124 function and data codes is required to operate only at 1/N times the frequency of the LH9124. For this reason, these signals can be set up by a microcontroller, or read from a memory unit.

Control signals to program the address generators will need to be clocked slightly faster than the function and data flow codes, depending on the number of address generators and the number of registers per address generator. These signals can be set up by a microcontroller or read from a memory unit.

MEMORY

The LH9124/LH9320 system requires three types of memory: Data, Coefficient, and Acquisition.

Data Memory

Each of the four data memory units (real and complex) is used as a temporary storage area for data as it moves recursively through the LH9124 during processing. In most cases, two of the units house real data and two house imaginary data. Each RAM buffer can have a maximum 24-bit width, depending on the precise memory configuration. Each set of two real/imaginary data memory units is connected to an adjacent address generator which outputs addresses to the data arrays.

Coefficient Memory

The two coefficient memory units use two 24-bit wide RAM units store the real and imaginary coefficients necessary for data processing. Different coefficients must be externally loaded for each type of calculation that the LH9124 performs. The units typically contain three types of coefficients: windowing coefficients, trigonometric functions (twiddle factors), and filter response coefficients.

Acquisition Memory

There are two acquisition memory units: one for real data and one for imaginary data. Each is a 24-bit wide RAM or FIFO. The acquisition memory holds input data until it is loaded into the system for processing. Input data can be from many different sources, including external devices such as a video camera or user-generated data. Processed data then returns to acquisition memory where it is accessed externally by the user.

MEMORY CONFIGURATIONS

In an LH9124 system, the DSP is surrounded by RAM buffers. The RAM for a maximum performance system must be static RAM (SRAM) and have a fast access time. The user also must consider the word size and the depth of the RAM buffer, which will have to be configured depending on the application.

Word Size and RAM Buffer Depth

The RAM buffer depth necessary for an application depends on the size of the processed data. A 1K RAM buffer for each array of real and imaginary data may be sufficient for a small segment of signal, however, for a 256 x 256 pixel digitized image signal, a RAM buffer 64K deep must be used. RAM buffers are addressed by the AG. The LH9320 AG has a 20-bit address bus output and can address up to 1M of RAM, which means that the maximum size for each RAM buffer is 1M x 24. The acquisition port Q can be configured with a FIFO and, in this case, no address generator is needed for the FIFO buffer and there is no maximum size limit for the Q RAM buffer.

The bit width of the RAM buffer depends on the memory configuration. Each of the I/O ports of the LH9124 have 48-bits; 24-bits for real data and 24-bits for imaginary data. This means that each RAM buffer has up to a 24-bit width. A 24-bit wide RAM buffer is not readily available, however, and the user may have to configure this buffer using three 8-bit RAM's. In some cases if only 8-bit or 16-bit data is needed for processing, just one 8-bit or 16-bit bus may be used.

NOTE

It is important that the 8-bit or 16-bit data bus from the RAM buffer be connected to the 8 or 16 most significant bits of the I/O port and that the rest of the I/O port bits are set to zero. This will alleviate the error/noise caused by truncation or round off which right-shifts the data during calculation.

Figure 5 is a static RAM Buffer and Figure 6 illustrates how to configure a 2K x 24 RAM Buffer using three 2K x 8 RAM devices.

The acquisition port can be equipped with a FIFO (first in, first out) dual port memory with internal addressing. The use of FIFOs for the acquisition port can simplify the system design in some applications. Since FIFO memory provides fully synchronous read/write operations and internal logic for unlimited expansion in both word size and depth, real-time processing in LH9124 is possible because the system will never be interrupted for loading or unloading data to/from the external device. Since in most FIFO's the address sequence is internally predefined, no external address information is required for operation and, thus, no address generator is needed.

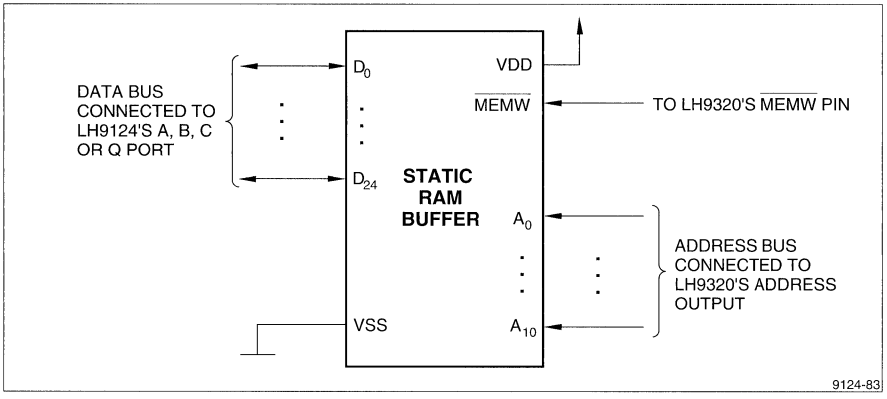


Figure 5. Connecting the RAM Buffer

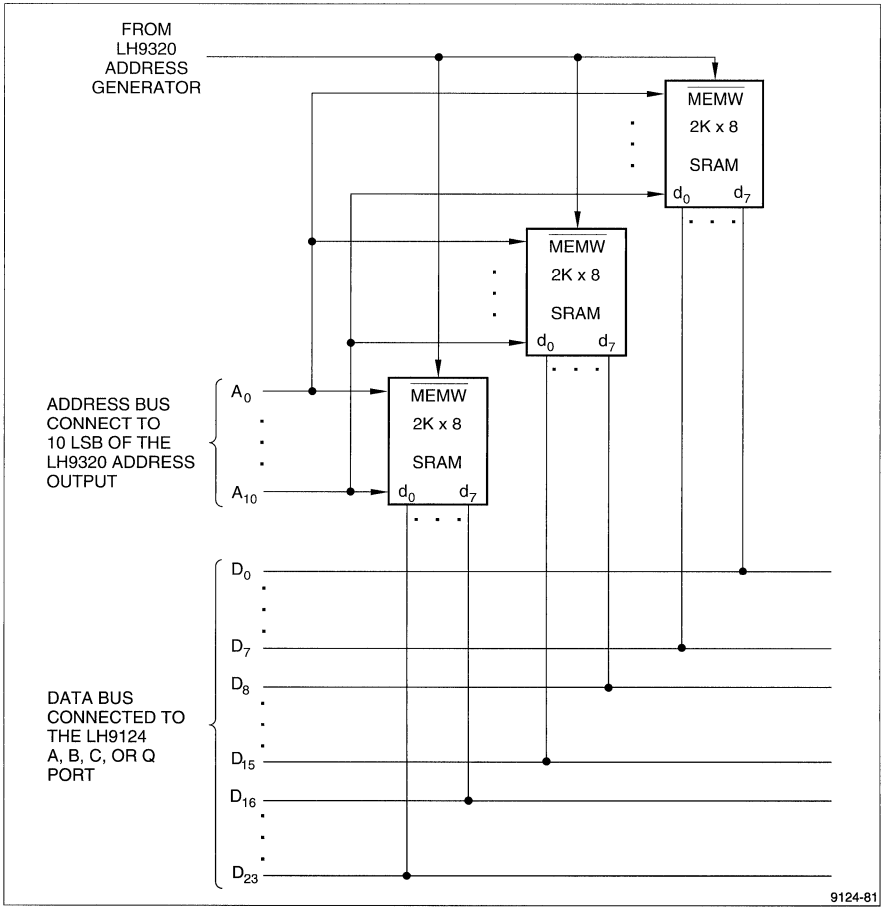


Figure 6. Word Width Expansion for the SRAM

The input port of the FIFO is connected to an external source device through the data-in bus and the output port is connected to the LH9124 acquisition port. In general, a FIFO memory is 8-bits wide and 512 to 4K deep and may need expansion to be configured with the LH9124 for digital signal processing tasks. Word width expansion is implemented by placing multiple devices in parallel. Each FIFO device should be configured for SINGLE mode. In this arrangement, the behavior of the status flags will be identical for all devices, so these flags may be derived from any one device. Depth expansion is implemented by configuring the required number of FIFO's in EXPANSION mode.

The $\overline{WE}$ and RS pin of the FIFO should be connected to the delay signal of the system clock (SYSCLK) or to the $\overline{WE}$ pin from the AG of the data RAM.

For both RAM's and FIFO's, the timing arrangement can be a delicate task. Refer to SHARP technical documentation for timing information.

SCHEDULER UNIT

The control device in the LH9124 system plays only a limited role and does not "control" the workings of the system as much as it "schedules" the sequence of events. The LH9124 system works with any general purpose control device or microcontroller. Minimally, the scheduler must have a sequencer (counter) connected to a memory device, which includes loop and interrupt or enable hardware.

Generally, when the control unit senses that the terminal count (TC) is low, it enables the device to output control sequences to the system. The START command and subsequent control codes and enables for the DSP system are output until the last instruction is reached, whereupon the control unit executes a loop and the process is repeated.

While more elaborate microcontrollers are available, the basic control device is a counter and RAM. The scheduler must be able to generate arbitrary control codes, be started after each pass, and be able to restart itself. When connected to a RAM or PROM, the addresses provided by the scheduler allow the memory unit to output the function and data flow codes for the LH9124 DSP. A more sophisticated microcontroller with internal memory reduces the total number of devices needed for the system and can perform some or all of the following functions:

- Self-programming
- Address generator programming
- Routing data on and off the board
- Post-processing of data

The device can be programmed using an EPROM or EEPROM as the memory unit, or with a direct connection to a PC in design environments.

The following are examples of control devices that can be used with the LH9124 system:

- Intel 8051
- Motorola MC68120
- Intel 8096
- Motorola MC146805
- Intel 8960
- Motorola MC68705

Connecting the Scheduler Unit

Any general purpose control device used for an LH9124 system will need to have common bus lines to all the AG's for data, R/$\overline{W}$, and START signals. In addition, each AG has a separate line to the scheduler for $\overline{CS}$ and TC.

The two control buses labeled Function Code (FC) and Data Flow (DF) are to be connected to the LH9124. The programmed control function and data flow mode signals are assigned to these two buses. Figure 7 shows the minimum hardware required to generate these signals.

A basic scheduler is composed of one EPROM, one counter and one gate. In this three-unit simple scheduler, the DF, FC, and extra control signals are tied to the LH9124 DSP. The TC pin signal is from the AG. Each AG has its own individual TC and $\overline{CS}$ line to the scheduler.

Refer to the manufacturer documentation for more information on connecting microcontrollers and other specific control units.

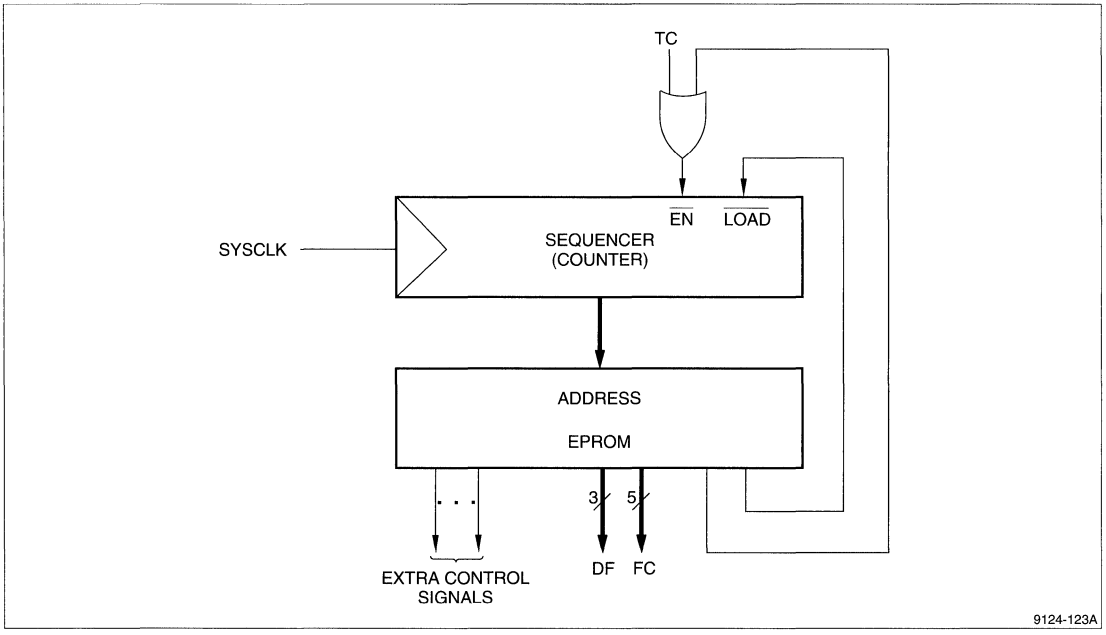


Figure 7. A Simple Scheduler Unit

GLOSSARY

Address Patterns. Lists of memory address locations which store data for specific DSP algorithms and general purpose applications.

Algorithm. A set of steps (procedure) that performs a specific calculation or process.

Array Processing. The process of performing a fixed set of operations on every element of an array.

Bidirectional Data Bus. A data bus that allows for a two way (I/O) transmission of data.

Block Floating-Point. The accumulated number of right shifts (divide by two) that are performed on an entire data array (block) to prevent overflowing.

Coefficient Data Arrays. Memory arrays that provide trigonomic terms for filtering and transforming input data via the coefficient port.

Complex Bus Architecture. Each of the LH9124 buses are composed of a tandem 24-bit real bus and a 24-bit imaginary bus.

Complex Number. Extends the real number system to provide solutions to equations like: X2 = –1. Complex numbers have 2 essential components: the real and the imaginary. Complex number representation is essential to the type of applications/problems DSP addresses and solves.

Complex Multiply. Multiplying 2 complex numbers requires 4 multiplications and 2 additions n.b. Sharp DSP performs a complex multiply in one cycle versus an average of 4-6 cycles for other DSPs.

Complex Window. A complex number sequence which, when multiplied against complex input data, smooths the effects of transforming a finite data set. A complex window can also be used to filter out unwanted frequencies.

Convolution. The process of modeling (filtering) the output response of a signal according to an input signal and the impulse response of the system that affects the signal.

Correlation. Process to find similarities between two signals.

> ***Example:*** High end copiers correlate the original image with the output image to verify high quality output.

Data Arrays. Sequences of numbers representing information that is usually separated in time or space (for DSP).

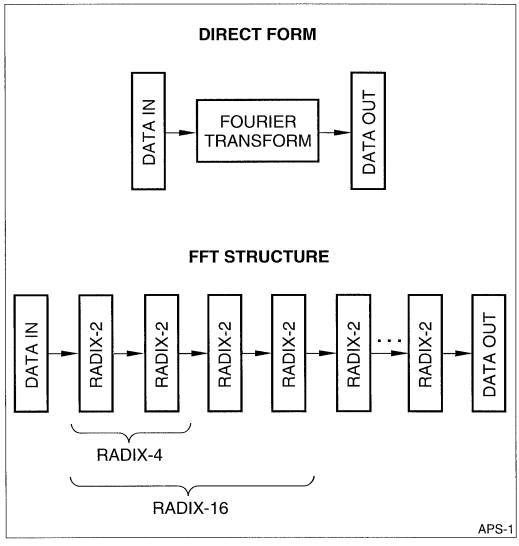
Data Flow. The directional path that an array follows in a unit or a system of units in order to perform a specific algorithm.

Digital Signal Processor. A unit that performs DSP algorithms on number sequences.

Digital Signal Processing (DSP). The practice of processing analog waveforms using digital techniques. Any signal can be converted into digital formats.

Fast Convolution. A fast convolution transforms the input and impulse response into the frequency domain via a Fast Fourier Transform (FFT). The transformed data is then multiplied by a filter coefficient and the result is transformed back into the time domain using the inverse FFT.

FFT. Fast algorithm to implement the Fourier Transform. The basic idea is to 'break' the computation into intermediate stages. The intermediate computations of the FFT are called radix passes.



Filtering. The process of removing or smoothing selected frequencies in data sequences.

Finite Impulse Response (FIR) Filter. The process of filtering data sequences by computing the impulse response of the desired filter and multiplying the finite data sequence in the time domain.

Fixed Point. Numbers represented as integer values: 1, 2, 3, etc. The advantage is that it is easy to implement the arithmetic in hardware.

Floating Point. Numbers represented by a mantissa and an exponent.

Example: 2.34 × 10 exp 2. The advantage is that you get a large dynamic range, but it is expensive to implement the arithmetic in hardware.

Fourier Series. A series of periodic functions that, when summed, produce a desired (arbitrary) waveform.

Fourier Transform. The algorithm for separating a signal into its component set of frequencies.

Function Codes. A set of numbers that trigger specific operations in a DSP algorithm.

Image Compression. The process of storing image data in less space using algorithms which find and notate repetitive data.

Image Correlation. The process of locating patterns in images using image templates (a set of predefined patterns).

Image Processing. The practice of manipulating images to produce, enhance or locate areas or items of interest.

Matrix Arithmetic. Mathematical operations performed on multiple rows and columns of numbers.

Precision. Accuracy. An example in DSP: The concept of 'round-off error' whereby numerical computations truncate data through many iterations, sometimes resulting in a very large 'round-off error' that can distort the results and, depending on the sort of information, must be compensated for somewhere in the analysis.

Example:

	Precision to 3 Decimal Places	Precision to 2 Decimal Places
addition of:	0.666	0.67
	+0.444	0.44
	+0.246	0.25
	+0.129	0.13
	1.485	1.49

The number 1.485 is more accurate than 1.49

Radix Considerations. A number base by which a Fast Fourier Transform (FFT) algorithm breaks down into an input signal.

RTS. Sharp's LH9124/LH9320 Real Time Simulators which simulate the capabilities of the LH9124 DSP and LH9320 AG.

Scaling. The process of dividing an entire data set by the same number (scaling factor) to prevent overflowing.

Spectrum Analysis. The analysis of the magnitude of squared frequencies which determines the composition of a signal.

Vector Dot Product. The vector result of multiplying the corresponding elements of two input vectors.

Windowing. A complex number sequence which, when multiplied against real input data, smooths the effects of transforming a finite data set. A complex window can also filter out unwanted frequencies when combined with an inverse fourier transform.

INTEGRATED CIRCUITS

FIFO

CONFERENCE PAPERS

LH5481/91/85/95
LH5496/97/98/99

HIGH SPEED/CONTROLLED OUTPUT FIFOS

INTRODUCTION

Sharp FIFOs include three product families:

1. Asynchronous
   - 64 × 8/9
   - 512 × 9
   - 1K × 9
   - 2K × 9
   - 4K × 9
2. Clocked
   - 4K × 9
   - 4K × 9 Parallel to Serial
   - 4K × 9 Serial to Parallel
3. Bidirectional
   - 256 × 36 × 2

These FIFOs boast high speed, low noise operation. System designers can use these parts to achieve cycle times of up to 40 MHz, without introducing the large current transients and ground bounce into the board environment typically associated with high performance CMOS devices. This achievement has been accomplished through the use of a unique architecture and current controlled output buffers and ensured by thorough evaluation.

Look Ahead Access Architecture

The Sharp FIFO memories are high-performance asynchronous dual-port memories that use a First-In First-Out (FIFO) addressing scheme. FIFO devices that incorporate random access memory architectures must be preconditioned within each cycle to allow an access to any location within the entire address space. All word-lines must be turned off and all bit-lines must be equilibrated every cycle. In the case of a FIFO memory, it is possible to take advantage of the sequential data access to hide or eliminate many of the overhead functions necessary for the operation of a conventional RAM array. Sharp has developed a unique architecture (patent pending) that uses these techniques to reduce access/cycle time and reduce power consumption. This 'Look-Ahead-Access' architecture incorporates the following:

- The use of multiple arrays to allow the equilibration of bit-lines to occur over multiple cycles.
- Selecting word-lines once for consecutive address locations mapped by columns.
- Shorting read bit-lines to write bit-lines when reading and writing occur on the same row within the same array. This prevents read pushout and/or write failure.

The 'Look-Ahead-Access' architecture allows Sharp to achieve extremely high-performance levels in the memory array. Sharp has also incorporated noise reducing techniques in the output buffers of the device.

Figure 1 illustrates a general block diagram of the Sharp FIFO architecture. The memory is segmented into two arrays A and B. Each array is organized into N rows, 8 words (9 bits per word) wide. The value of N is dependent on the depth of the FIFO. The arrays are dual-ported to support asynchronous read and write operations. Array A contains the even rows and B contains the odd rows. As consecutive read or write operations occur, the accessed location marches across a row to the array boundary. At the end of the row, the accessed location jumps to the alternate array. Subsequent accesses continue to march along the next row to the array boundary where the accesses jump back to the original array.

The use of multiple arrays allows equilibration of bit-lines to occur over multiple cycles. When consecutive accesses reach an array boundary and transition to the alternate array, equilibration of the bit-lines of the current array is initiated. Equilibration of the bit-lines of the passive array has a full 8 cycles to complete.

On the transition from one array to the other, the word-line in the newly accessed array is active for the accessed row during the recovery time between cycles. No additional delay is incurred in the transition between arrays, since the bit-lines are equilibrated in advance.

Managing Output Noise

The 'Look-Ahead-Access' architecture allows Sharp to achieve high performance without relying on high current, large transistor outputs. In addition, system level noise is minimized and output signal integrity is maximized by controlling output edge rates. The following techniques were used to further reduce noise:

- Current supplied to the output buffers is controlled to prevent large current spikes.
- On-chip capacitance has been incorporated between the two power planes to supply current in high demand situations and thereby reduce current transients through the supply pins.
- Lead inductance for the power pins has been minimized to reduce the voltage response to current transients through the supply pins.

Rigorous Component Evaluation

Thorough evaluation of FIFO memories is paramount to ensuring reliability. All DC electrical characteristics and AC switching characteristics were established through an intensive characterization process. These parameters are tested in the production environment for all modes of operation. An evaluation was conducted by testing the devices over a broad range of timing relationships between read and write operations. The timing relationship between the read and write operations may be asynchronous. Therefore, Sharp conducted a thorough evaluation of the FIFO memory architectures to ensure proper operation for all possible variations of asynchronous read/write timing conditions.

Asynchronous Evaluation

The Asynchronous test was modelled after experiments conducted at Xerox Corporation by M. Sheperd and D. Rogers as described in their paper, *Asynchronous FIFO's Require Special Attention*, IEEE International Test Conference, 1985. The FIFO is exercised with the following pattern:

a. 2 writes and 1 read until full

b. 1 write and 2 reads until empty

c. 1 read and 1 write to increment both read and write address pointers.

d. repeat steps a – c.

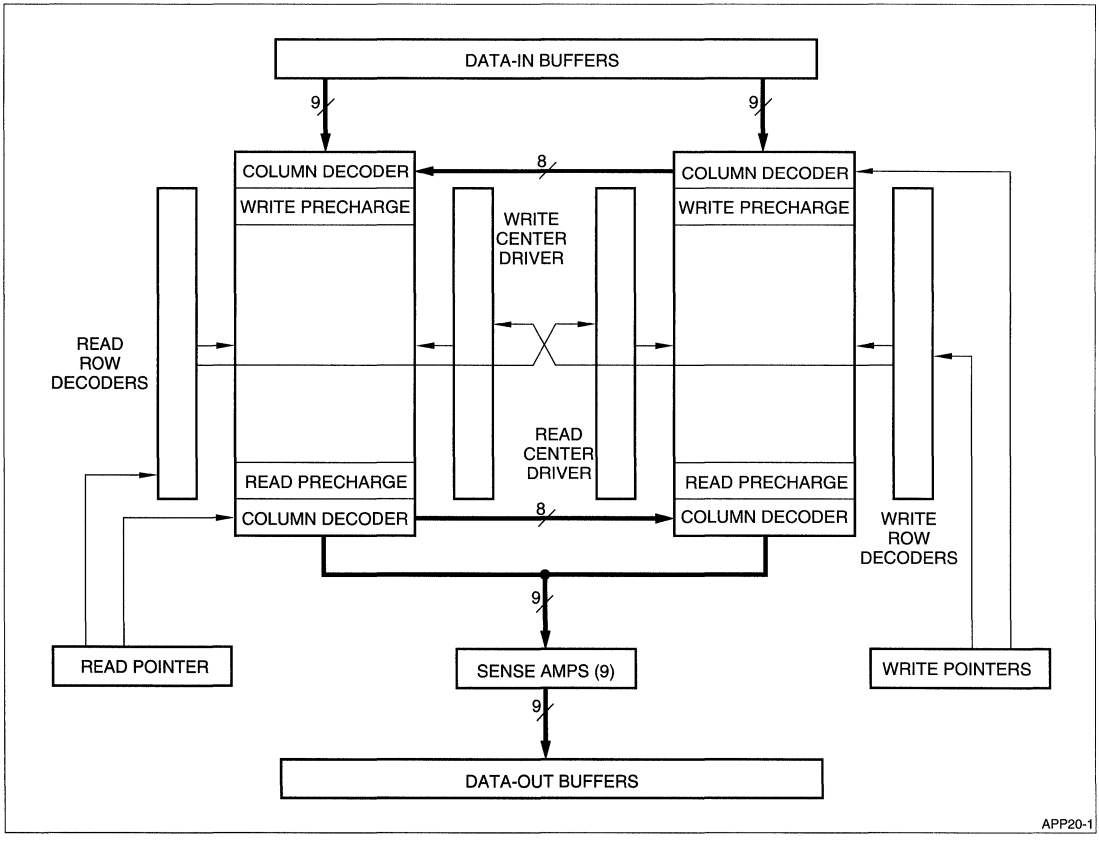


Figure 1. FIFO Architecture Block Diagram

The test pattern described above was used to exercise the FIFO for varying timing relationships between the read and write signals. Steps a – c constitute one pass. Figure 3 illustrates the timing relationships for sequential passes. A random delay generator setup was introduced between the tester and the write signal, see Figure 2. The random delay generator added a constantly varying 0 to 10 ns delay to the write signal within each pass of the test. The write signal is swept pass the read signal by making one nanosecond increments between passes. The test software was set up to loop continuously, see step d of the above test pattern.

The testing was conducted at room temperature over the specified voltage range. The continuous loop testing was conducted over a period of several hours. Data out and flags were tested to determine pass/fail. After several hours, no errors were recorded.

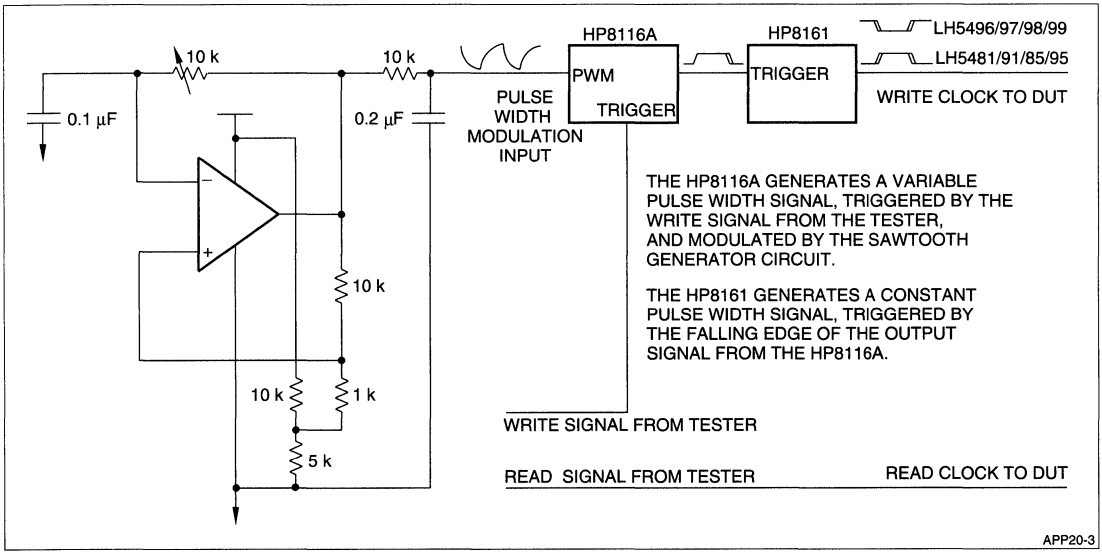


Figure 2. Write Signal Skew-Delay Generation Setup

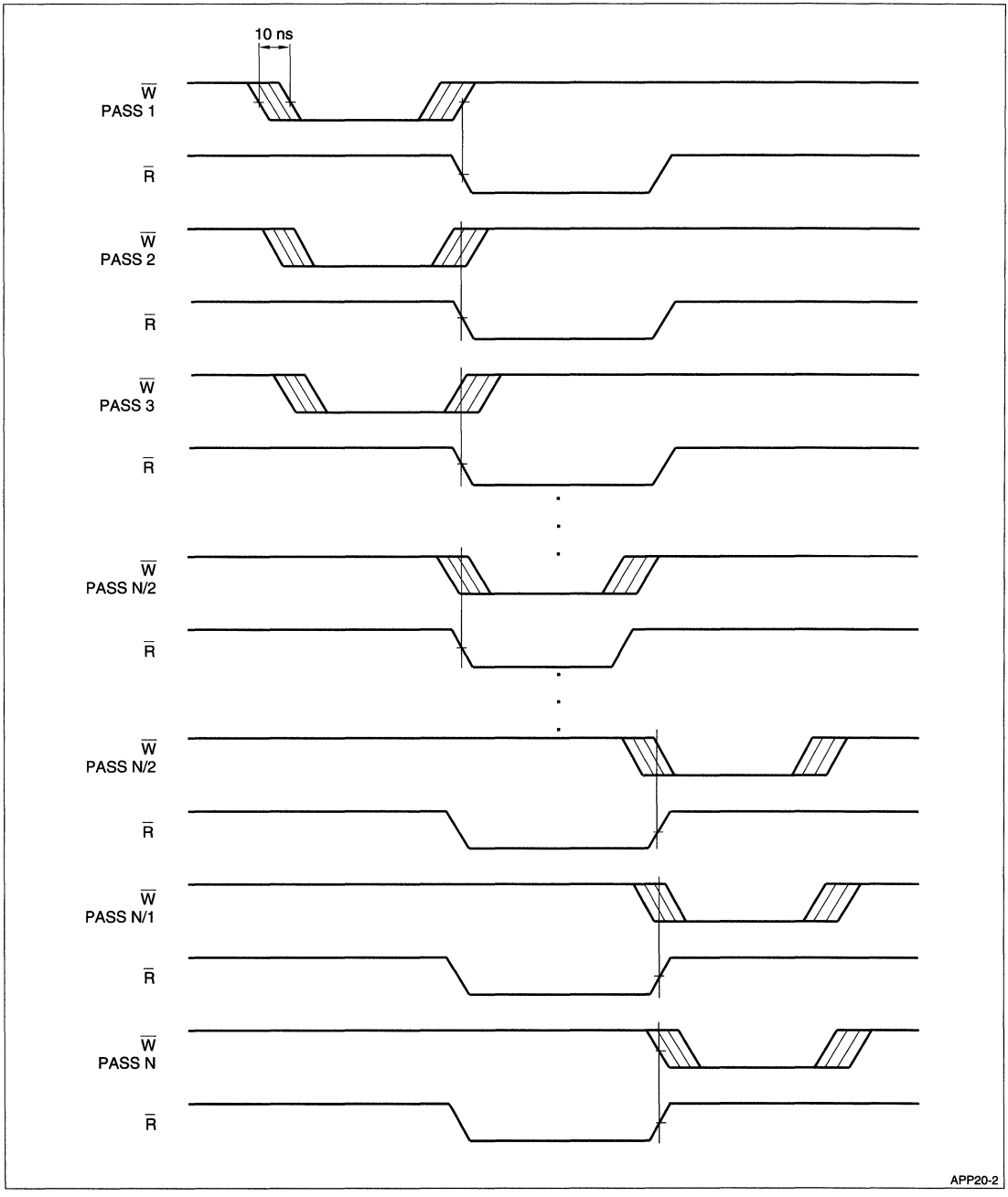


Figure 3. LH5496/97/98/99 Read/Write Timing for Sequential Passes

DATABUS FUNNELING MADE EASY

Mike Yee, Senior Applications Engineer

INTRODUCTION

The Sharp LH5420 256 × 36 × 2 CMOS Bidirectional FIFO is an innovative device which turns the difficult task of funneling and defunneling different-size databuses into an easy one-component solution. Funneling refers to a situation where data from a larger databus (e.g., 32-bits wide, 36-bits with parity) must be segmented (usually in increments of 8-bits, 9-bits with parity) and transferred to a smaller databus (e.g., 8-bits wide, 9-bits with parity). The funneling options available on the LH5420 are '36-bits to 9-bits' and '36-bits to 18-bits.' Defunneling refers to just the opposite of funneling. To defunnel, data from a smaller databus (e.g., 8-bits wide, 9-bits with parity) is combined together sequentially with other data from that databus, and transferred in parallel to a larger databus (e.g., 32-bit wide, 36-bit with parity). The defunneling options available are '9-bits to 36-bits' and '18-bits to 36-bits.' For wide word applications on both ports, '36-bit to 36-bit' buffering is also available.

A very important feature of the LH5420 is the ability to operate bidirectionally. The term Bidirectional refers to the LH5420s ability to funnel and defunnel between different sized databuses, allowing data to travel in both directions. Bidirectional operation is also available when the full width of both ports are used (e.g., 36-bit to 36-bit buffering).

The advantages of the LH5420 bidirectional FIFO to the system designer are: elimination of several conventional FIFOs and glue logic; significant reduction of board

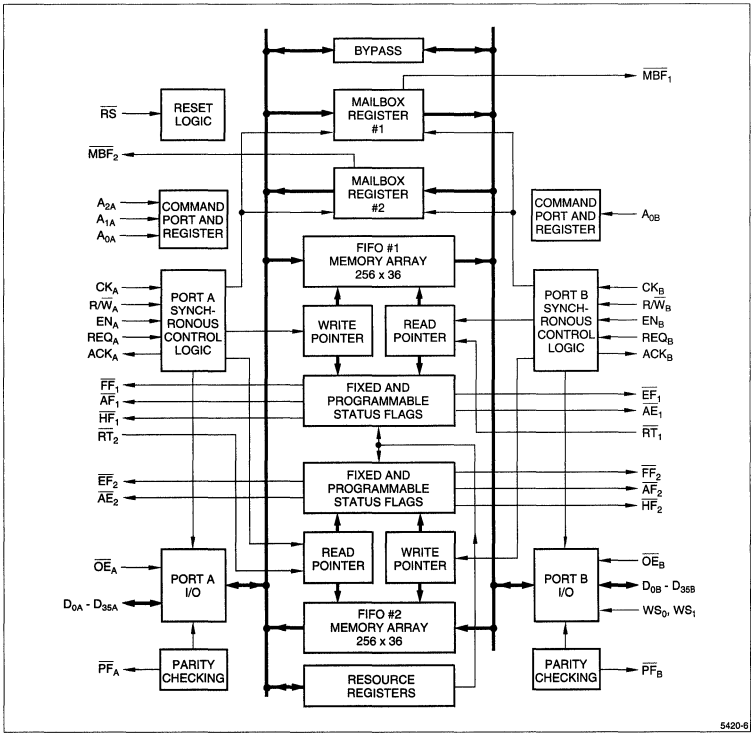


Figure 1. LH5420 Block Diagram

space; elimination of the complexities of handling bus contention; and improved system performance. But, most importantly, it makes databus funneling easy.

Conventional Databus Funneling Solutions Can Be Awkward

The rapid transfer of information between a databus of one size to a databus of a different size (funneling or defunneling) seems like a simple enough operation, when viewed on paper in block diagram form; but the block diagram must be transformed into a high-speed circuit design. Conventional solutions require many components, and considerable board area. Further, the timing required for reading, writing, and flag detection for multiple parts in parallel, places a heavy burden on reliable high speed operation.

Conventional Funneling Circuit Design

Figure 2 shows a bus-funneling circuit designed using conventional components. Figure 2a is an example of the timing required to use the circuit in Figure 2. Figure 3 shows the circuit which must accompany Figure 2 if the circuit were expected to operate bidirectionally (funnel and defunnel). Figure 3a is an example of the timing required to use the circuit in Figure 3. An obvious disadvantage of this conventional funneling circuit is the number of components required. One "Programmable Logic Device" (PLD) and four standard 256 × 9 FIFOs are required for one-way funneling. If bidirectional operation (funneling and defunneling) is important, two PLD's and eight 256 × 9 FIFOs are required. The combination of all these components results in very restrictive data setup (t_{DS}) and hold (t_{DH}) timings during a Write cycle, and restrictive access timings (t_A) due to the risk of databus contention during a Read cycle. In many cases, high speed operation would be out of the question. Tight controls on signal noise and signal skew might also be required to keep the four FIFOs synchronized. After all this, the circuit designer would do just about anything for a single-chip solution. Setup and Hold times for a single asynchronous 256 × 9 FIFO are typically 10 ns and 0 ns respectively for access times of 20 ns. Because this defunneling circuit is a combination of separate components (see Figure 3), setup and hold times would have to be increased significantly to ensure correct synchronization due to signal propagation delays of the control signals and data. In a conventional defunneling circuit, there could be as many as four 9-bit words waiting to be written sequentially into four different FIFOs. Each of the four

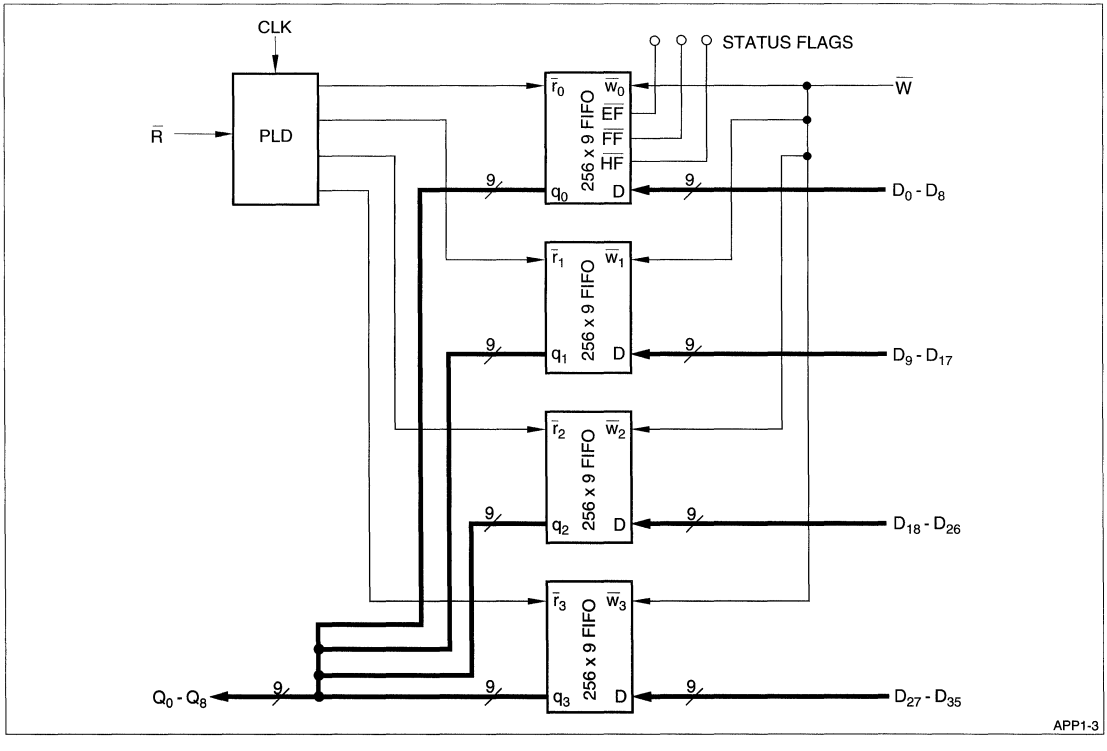


Figure 2. 36-Bit to 9-Bit Conventional Funneling FIFO Circuit

9-bit words requires its own setup (t_{DS}) and hold (t_{DH}) time (see Figure 3a). These restrictions will limit the maximum defunneling frequency of this circuit.

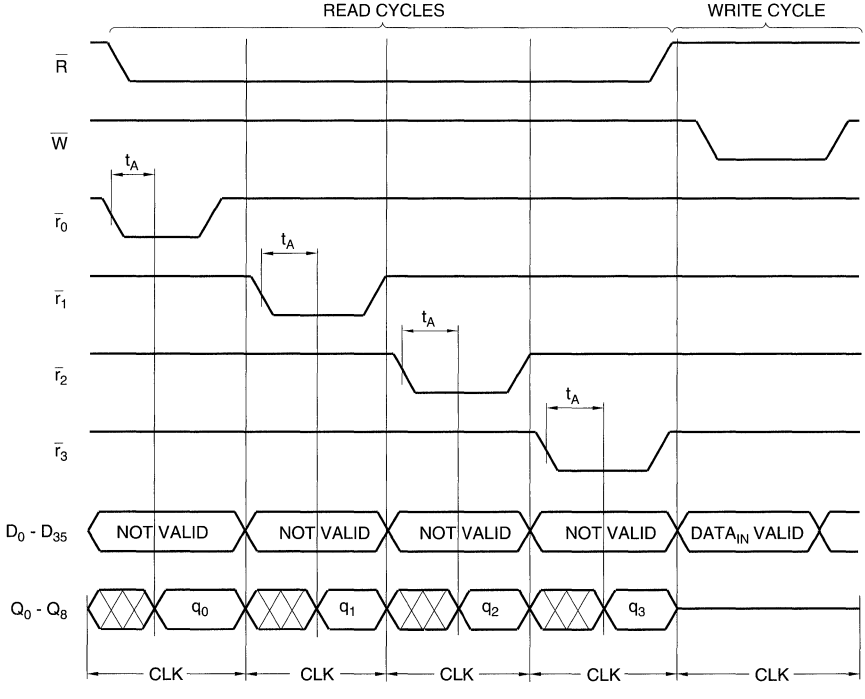
Databus contention is a common problem experienced when combining two or more output pins from different devices in parallel, on the same databus (see Figure 2). If during a Read cycle, at least two of the output pins happen to be momentarily on at the same time, the two output drivers potentially could fight against each other driving the data bus to opposite logic states (one driver pulling the bus to 0V, while the other driver is simultaneously pulling the bus to 5 V). Databus contention degrades system performance and increases the system operating current.

Another significant disadvantage with using the conventional component solution is handling the flags. Each 256 × 9 FIFO has 3 types of flags which can be used in the application to indicate the current FIFO status (Empty, Full, or Half Full). Most designers use a flag from only one of the four FIFOs. This flag-handling technique has a significant disadvantage. When a flag from only one of the four 9-bit wide FIFOs is used to represent the entire 36-bit word, there is no way to insure that the other three FIFO flags are synchronized (empty, full, or half full at the same time) with the first. There is the possibility that one, two, or all three of the other FIFOs may have become unsynchronized (due to signal noise, excessive signal skew, etc.) and are now contributing incorrect data to the 36-bit word.

Sharp's Single Chip Solution to the Complexities of Funneling

The LH5420 CMOS Bidirectional FIFO was designed specifically to simplify the handling of wide-word (up to 36-bits) data buffering. The notable features of this device relating to data bus funneling are:

- Selectable 36/18/9-bit Word Width on Port B
- Two 256 × 36-bit FIFO Buffers for Bidirectional Operation
- Synchronous operation on both Ports A and B
- Fully Asynchronous Communications between Port A and Port B
- Only One Set of Flags for the Entire 36-bit Wide Word
- Capable of 40 MHz operation



NOTE: $\overline{r}_0$, $\overline{r}_1$, $\overline{r}_2$, $\overline{r}_3$, = PLD generated READ signals.

APP1-4

Figure 2a. 36-Bit to 9-Bit Conventional Funneling Write and Read Timing Diagram

The LH5420 provides an easy one chip solution to the problems associated with funneling one size databus to a different size databus (see Figure 3). The LH5420 also provides a simple method of buffering wide word databusses up to 36 bits wide on each port. There are two ports on the LH5420, Port A and Port B. A Port is defined as an interface between the outside databus and the internal FIFO memory. Each port can be used as an input or an output depending on which direction the data will travel. The LH5420 allows Port B to be selectable in word widths from 36, 18 or 9 bits wide, while Port A is fixed at 36-bits wide.

Two separate 256 × 36-bit FIFO buffers work side-by-side to move data in opposite directions. This is what enables the LH5420 to operate bidirectionally. As an example, a 36-bit databus and a 9-bit databus can send and receive data back and forth, giving unrestricted communication privileges between an 8-bit microcontroller and a 32-bit microprocessor. Clock-frequency differences between the two busses are not an issue. Even though the individual ports are synchronous in nature, each port is controlled from separate system clocks (CKA and CKB). Each port operates independently from the other, so that port-to-port communication occurs asynchronously.

The LH5420 has five different types of flags available: Full Flag (FF), Empty Flag (EF), Half Full Flag (HF), Almost Full Flag (AF), and Almost Empty Flag (AE). The Almost Empty and Almost Full Flags are programmable. One set of these flags are available for each 256 × 36 FIFO buffer, to cover the status of data going in either direction. The low skew inherent in a single monolithic solution eliminates the risk that desynchronization will occur within the 36-bit wide word in the FIFO. Further protection is afforded because the flags cover the full 36-bit word width and not just the 9 bits that were used in the conventional funneling design mentioned above. The problems of designing a system around restrictive read and write timing constraints are no longer an issue, because the complexities of funneling timing synchronization are handled automatically within the LH5420 bidirectional FIFO.

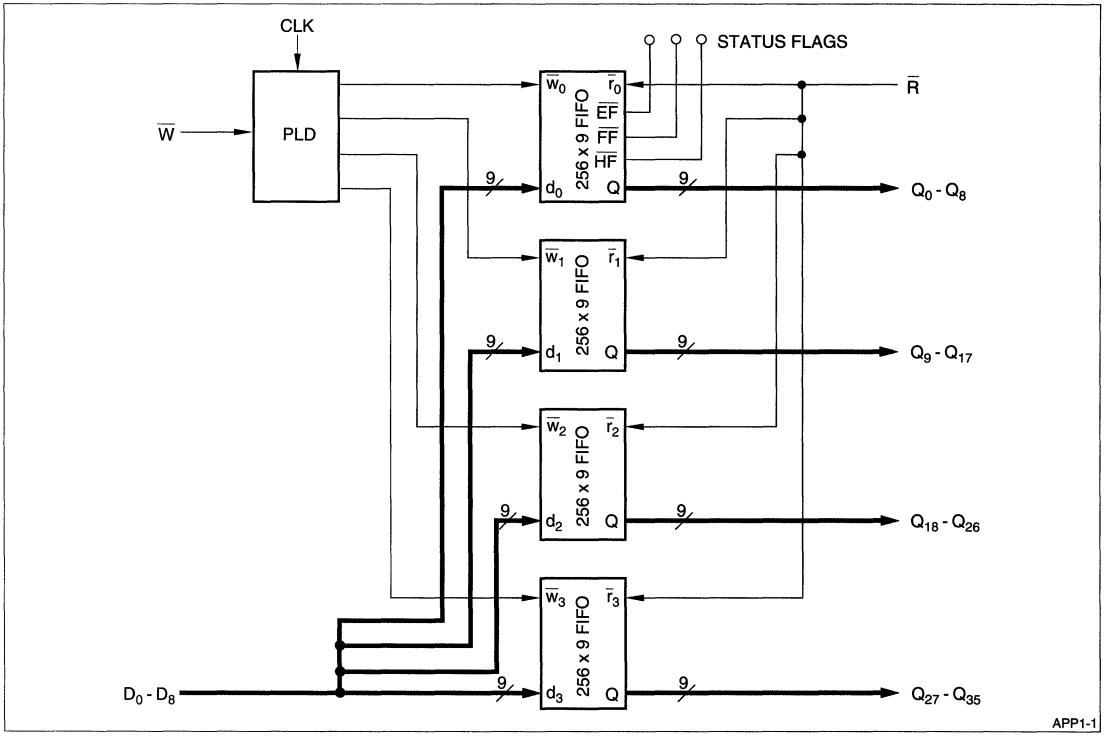


Figure 3. 9-Bit to 36-Bit Conventional Funneling FIFO Circuit

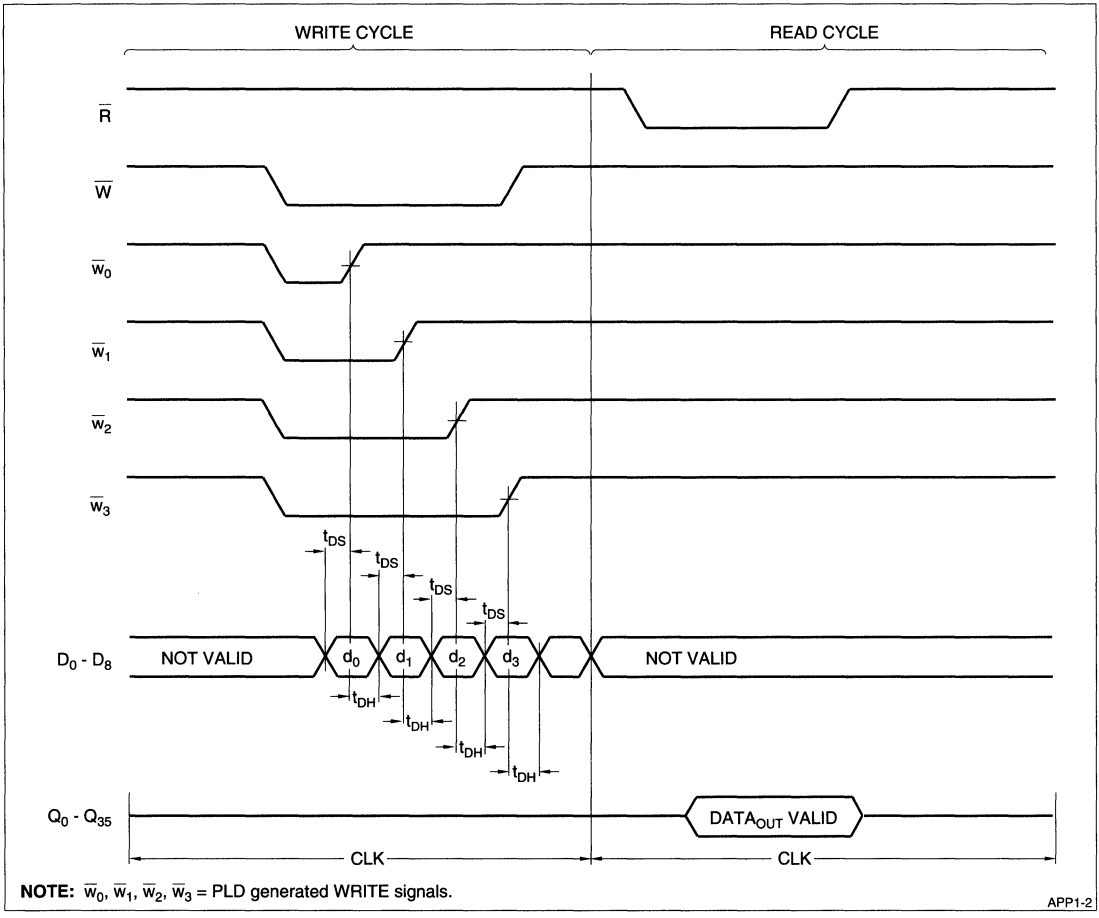


Figure 3a. 9-Bit to 36-Bit Conventional Defunneling Read and Write Timing Diagram

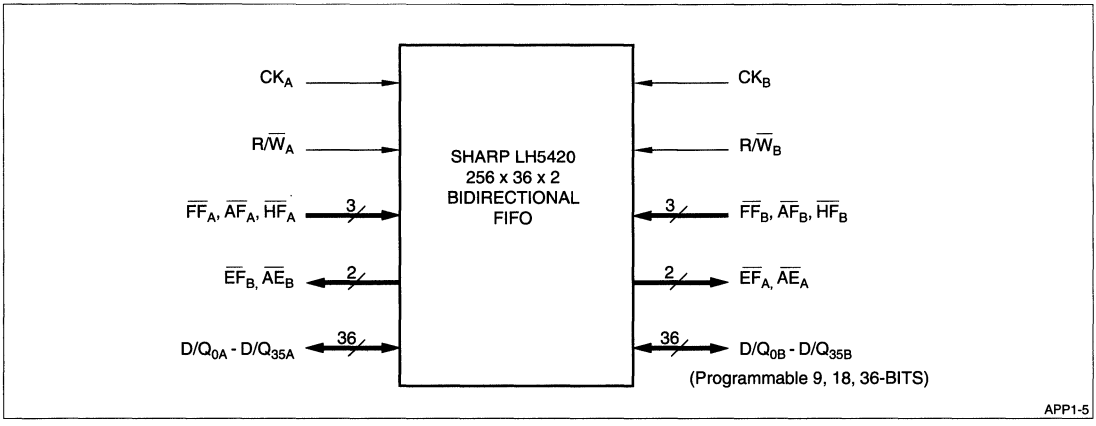


Figure 4. LH5420, the Single Chip Solution for Databus Funneling

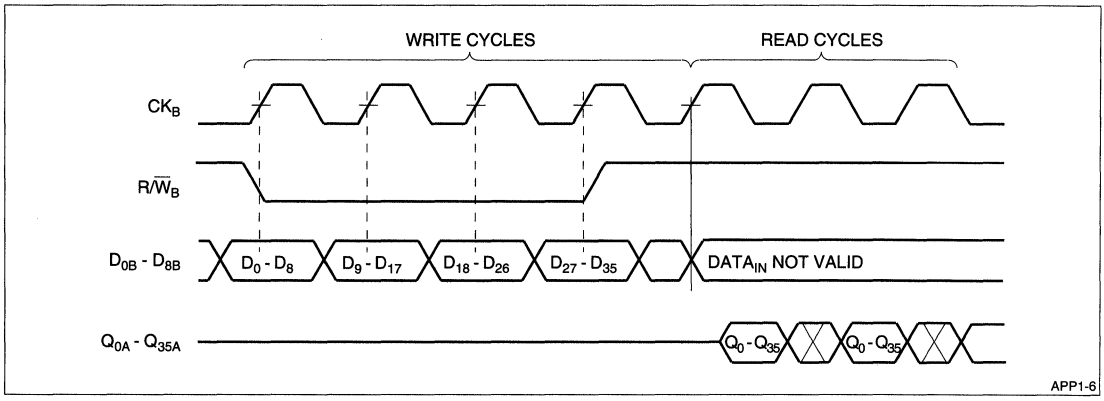


Figure 4a. LH5420 9-Bit to 36-Bit Funneling Write and Read Timing Diagram

SUMMARY

The Sharp LH5420 bidirectional FIFO provides many benefits to a system designer working on applications which use wide word databusses (36 bits wide), or applications which require funneling and defunneling between databusses of different widths (e.g., 8-bit to 32-bit, 18-bit to 36-bit, etc.). In comparison with conventional databus funneling methods, the LH5420 simplifies your circuit design, allows faster operating speeds, uses less board space, reduces component count, and provides bidirectional funneling with no additional circuitry. But best of all, it is easy to use.

'SMART RETRANSMIT' FIFOs

Baruch Berdugo, Applications Engineer II, Product/Applications

INTRODUCTION

Basically, a FIFO is a self-addressing two-port memory. Data is written into the FIFO from the input side, and is read by an independent process from the output side. Because the timing of the input process is decoupled from that of the output process, FIFOs are very useful for communication between independent processes.

The internal structure of most modern FIFOs includes a dual-port memory; a Write Pointer (WP) that defines the address of the next datum that will be written into the FIFO; and a Read Pointer (RP) that defines the address of the next datum that will be read out of the FIFO. Both pointers, WP and RP, are counters modulo in the FIFO depth. Writing into the FIFO and reading out from the FIFO are always done sequentially; thus data can be read out from the FIFO only once. Some recently-implemented FIFOs include the option to perform a 'retransmit' operation, which resets RP to address zero (the address of the first physical memory location).

However, this simple retransmit option provides only a limited solution. Many applications require a 'smarter' retransmit operation to control RP.

This Application Note defines the new retransmit mechanism used in the LH543620 1024 × 36 FIFO, which allows flexible control on RP. The flexibility and the potential of the new Retransmit are demonstrated by application examples from different areas: Digital Signal Processing (DSP), Computer Communications and Instrumentation.

THE LH543620 RETRANSMIT DEFINITION

Associated with the new retransmit feature are the following resources:

- Two registers: RBASE and ROFFSET.
- Three control signals. RTMD[1:0] defines the operation mode, $\overline{RT}$ enables the operation.

The resource registers are zeroed whenever the FIFO is reset. They may be reprogrammed in the same way as the Almost-Full and the Almost-Empty flag offset registers. The RBASE register also may receive new contents at any time, whenever the Retransmit control input (RT) is asserted. The new contents are derived from the current values of RP and the ROFSETT register.

Table 1. Retransmit Operation Modes

$RTMD_1$	$RTMD_0$	OPERATION	ACTION TAKEN
H	H	Mark	(RP) → (RBASE)
L	H1	Retransmit	(RBASE) + (ROFFSET) → RP
H	L	Retransmit and Mark	(RBASE) + (ROFFSET) → RP and (RBASE) + (ROFFSET) → RBASE
L	L	Cascade Mode	The Almost-Empty Flag is delayed by one CKO clock

NOTE: The cascade mode has no functional connection to the retransmit mechanism.

The Retransmit operation is acknowledge whether or not the read enable signals (ENO_1, ENO_2) are asserted. However, in order to read the new data word it is required to issue three enabled Read clocks (CKO). See Figure 1.

There is an inherent conflict between the Retransmit feature and the flags mechanism. The flags prediction logic is based on a state-machine, but the retransmit operation sets a new value to the Read Pointer; therefore, the 'state' of the state machine is incorrect at this moment. It is required to define a reasonable temporary state to enforce upon the state machines. The LH543620 flags indicate an 'Almost-Full State' when retransmit is asserted. For example:

- Empty flag and Almost-Empty flag are deasserted.
- Half-Full flag, Almost-Full flag, and Full flag are asserted.

The Flags' state machine is 'recovered' and shows the new state of the FIFO after three enabled read cycles (the same time a new data is read out of the FIFO). Actually, a complete recovery of the flags state machine from retransmit is guaranteed after two enabled read cycles with no enabled write.

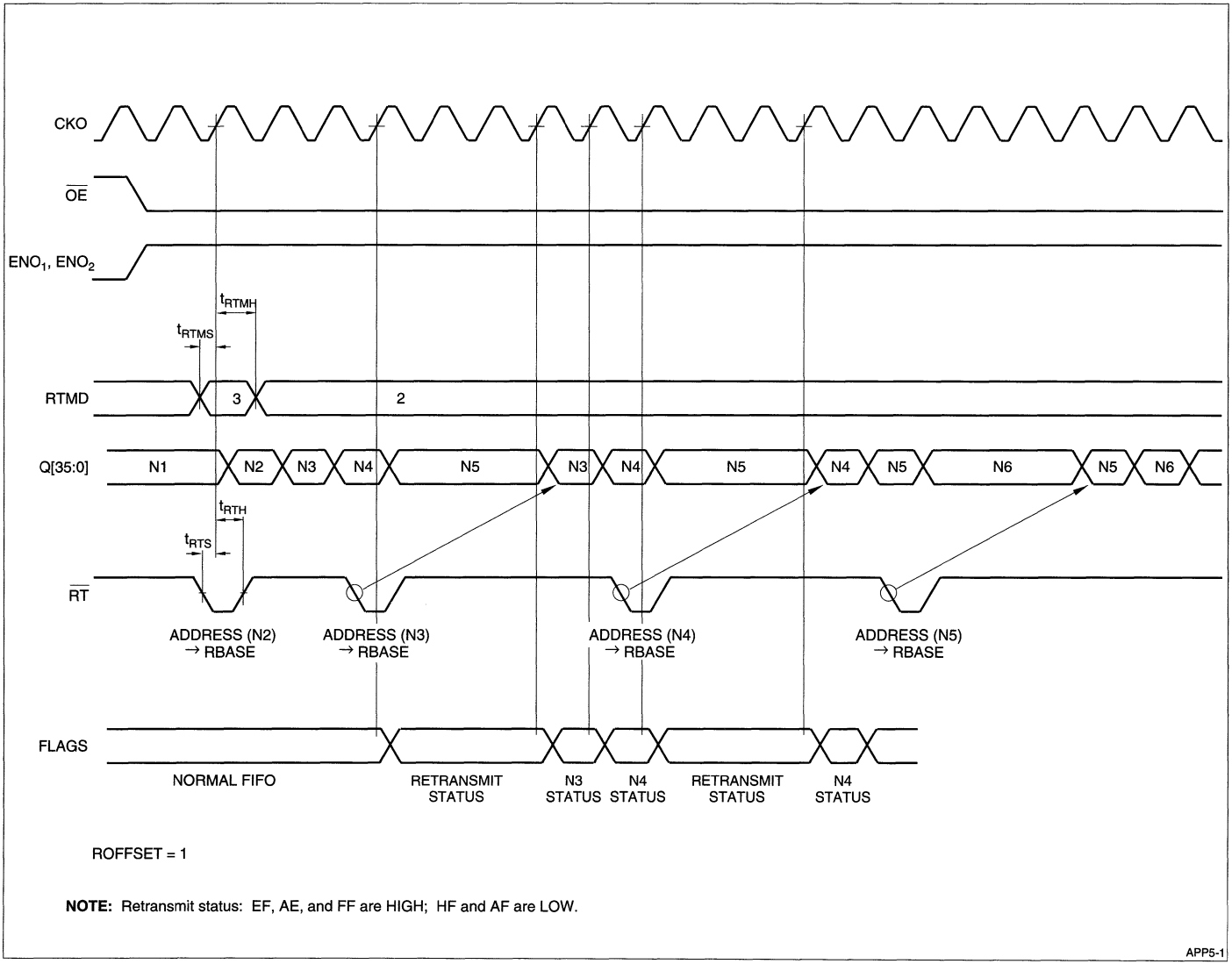


Figure 1. Retransmit Using Retransmit and Mark Mode

The flags ($\overline{AF}$ and $\overline{HF}$), when synchronized to the write clock, are further delayed to the next write clock.

During regular FIFO operation, the $\overline{FF}$ inhibits writing into the FIFO when it is full, and the $\overline{EF}$ inhibits reading when the FIFO is empty. This behavior inhibits a wrap-around scenario; i.e., the Read Pointer getting ahead of the Write pointer. However, when using retransmit, there is no self-protection against wraparound. It is the user's responsibility to avoid it. In the 1K × 36 FIFO, the user has to guarantee that the 'distance' between the marked point (RBASE) and WP is less than or equal to 1024.

APPLICATION EXAMPLES

This section presents some applications, which use the new Retransmit mechanism to gain a simple and efficient system implementation.

FIR Filter Using FIFOs

Finite Impulse Response (FIR) filters are widely used in digital signal processing (DSP) applications because of their well-behaved nature.

The FIR filter equation is.

$$Y(n) = \sum_{k=0}^{N-1} x(n-k)\, h(k) \qquad [1]$$

where:

x(n) is the input data
h(n) is the filter coefficients
y(n) is the output data

Equation [1] implies that, for every new input datum, the execution multiplies and accumulates the last N input points, against N coefficients.

Reference 1 describes a FIR filter implementation using FIFOs that are not equipped by the 'smart' retransmit capability. The idea in reference [1] is to rewrite N-1 points back to the FIFO every time a new datum is received. Since the FIFO's input is also used to rewrite the old data, the inherent decoupling between the input process and the output process is violated. To maintain this decoupling, the suggestion in [1] is to use an additional FIFO in front of the main FIFO. With this configuration the minimal hardware to sequence the data for the Multiply-Accumulate (MACC) is: three FIFOs, a multiplexer, a programmable logic device, and a counter.

By contrast, in the implementation described here, using the new retransmit, sequencing the input data for the FIR filter mechanism is implemented with two FIFOs with no additional logic (besides the initialization programming).

FIR Implementation Using the LH543620

This system includes two FIFOs as illustrated in Figure 2. FIFO0 is used for data acquisition, and FIFO1 is used to store the filter coefficients. FIFO0 is configured to Retransmit and Mark mode – RTMD[1:0] = 2; FIFO1 is configured to Retransmit mode – RTMD[1:0] = 1.

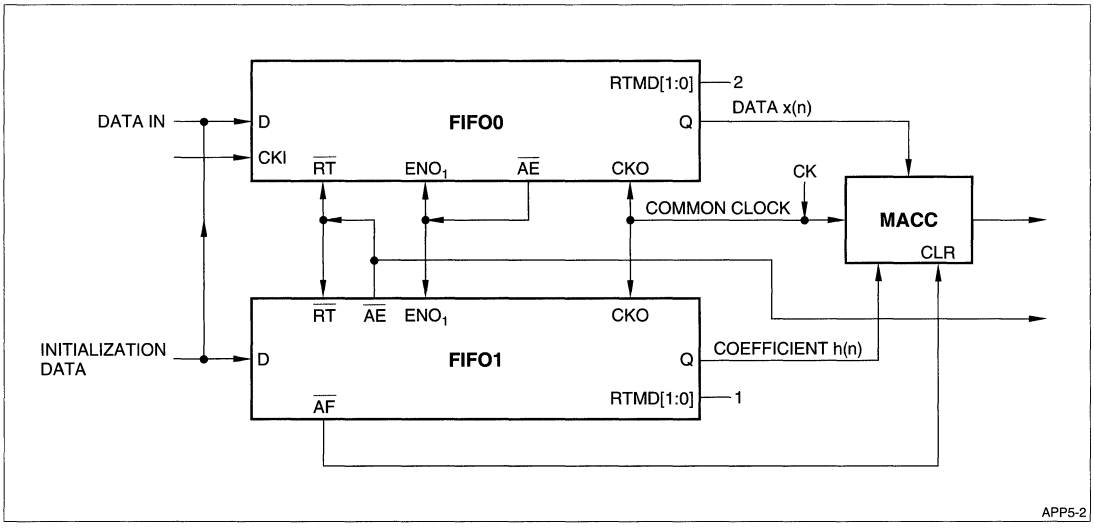


Figure 2. System Configuration for FIR Filtering

At the initialization stage the controller should do the following:

1. Write the coefficients, plus 9 zeros, to FIFO1.
2. Write 1 to the ROFFSET register of FIFO0.
3. Program the $\overline{AE}$ flag of both FIFOs to be synchronized to the output clock.

Figure 3 shows the process when the filter length is 5.

When $\overline{AE}$ of FIFO1 is asserted it asserts the $\overline{RT}$ inputs of both FIFOs which causes the following:

- The RP of FIFO1 is set back to address 0.
- The RP of FIFO0 is set to (RBASE), which contains the address of start of block.
- RBASE register is incremented by 1, to point the beginning of the next block.
- Data is read simultaneously from both FIFOs.

When N data points are read out of the FIFOs, FIFO1 became Almost-Empty and its $\overline{AE}$ flag is asserted, then the process is repeated. If there is not sufficient data in FIFO0, the $\overline{AE}$ flag of FIFO0 is asserted 'LOW' and disables reading data from the FIFOs, until A new samples are entered at the input side. The $\overline{AE}$ flag of FIFO1 can be used to indicate that data is ready at the output of the MACC. The $\overline{AF}$ flag can be used to clear the MACC so it is ready for the next filtering cycle.

Every word is written to the FIFO one time but is read out of the FIFO N times. If Fout is the output frequency of the common clock, the average input frequency should be smaller then N*Fout.

Overlap for FFT

The requirement for overlap addressing is common for FFT-based applications. For example, when performing the frequency-domain convolution using the overlap-and-save method, the input data is used in overlap, where the number of overlapping points is equal to the filter length. When performing spectral analysis using the Welch method, each processed block overlaps the previous block. The amount of the overlap depends on the window function. The used techniques to handle this overlap requirement normally is handled by means of a circular buffer, either internally in the processor, or externally using an Address Generator.

Figure 4 illustrates a system that implements overlap addressing using the LH543620 FIFO. The FIFO is set to Retransmit and Mark mode: RTMD[1:0] = 2. The $\overline{AF}$ offset register is programmed to N = FFT_Length, and ROFFSET register is set to (N-Overlap). The data is loaded to the FIFO each time CKin is triggered.

The DSP senses the $\overline{AF}$ (Almost-Full) flag of the FIFO. Whenever this flag is being asserted, a new block of data is available in the FIFO. The DSP then reads a block of data having the size of the FFT_length and then asserts the FIFO's RT signal, which causes the RP and RBASE registers to be set at the beginning of the new block. See Figure 4.

Computer Communication

FIFOs are widely used in computer-communications applications, so that each computer can send or receive a message at its own natural rate. Most systems perform some level of error-checking; whenever an error is detected, the receiver reads the last massage using retransmission. The simple Retransmit mechanism currently implemented in FIFOs sets the RP back to absolute

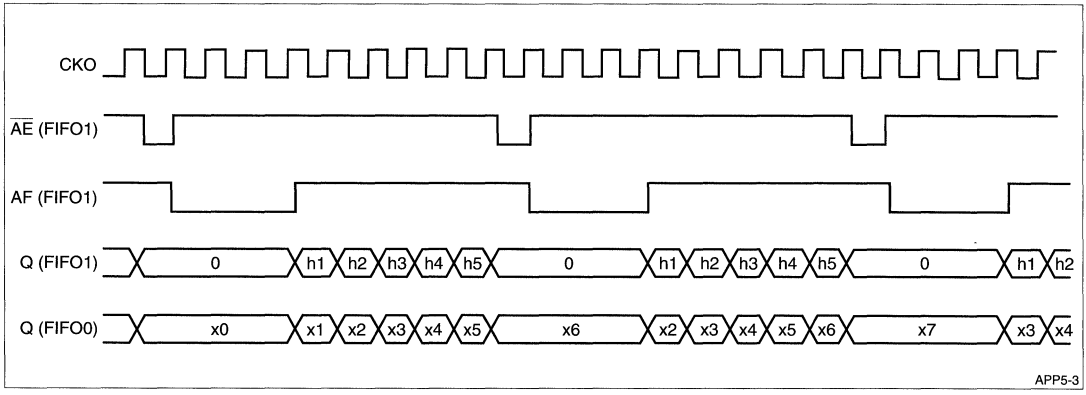


Figure 3. Sequence of Events for FIR Filtering

address zero. This mode of retransmission inherently assumes that the message must have started at absolute address zero. This constraint is burdensome since it implies that the sender cannot start sending any new message until it has received an acknowledge from the receiver. With the new Retransmit mechanism, this constraint no longer exists and a more efficient protocol can be implemented, as follows:

- The FIFO is set to mode 0 (MARK).
- The sender fills the FIFO, without being synchronized to the receiver.

When the receiver reads a block of data and finds no errors in the data block, it can mark the beginning of the new massage by setting the FIFO in Mark mode RTMD[1:0] = 3 and assert the $\overline{RT}$ signal for one clock cycle.

If the receiver finds an error in the data block, it can read the last massage again by setting the FIFO in Retransmit mode RTMD[1:0] = 1 and assert the $\overline{RT}$ signal for one clock cycle.

Waveform Generator

A typical implementation of a waveform generator instrument is to store one cycle of the waveform in a digital memory, and then to output it repetitively. The digital signals are converted into analog signals by a D/A followed by a low-pass filter. Some instrument manufacturers have found that FIFOs are attractive components for implementing the digital memory. The data for one cycle of the waveform is loaded into the FIFO during an initialization stage. The Empty (or Almost-Empty) FIFO flag signal is used to trigger Retransmit. Thus the same data block is read out from the FIFO in a cyclic pattern. The simple Retransmit mechanism always sets RP to address zero; therefore, only one waveform can be generated by the FIFO. With the new Retransmit mechanism, it is possible to store more then one waveform in the FIFO. Selection of a specific waveform is done by proper programming of the RBASE and $\overline{AE}$ registers. The system configuration is illustrated in Figure 4.

During the initialization process, the FIFO is loaded with all required waveforms. Playing a specific waveform is done by programming RBASE to the starting address of that waveform, and the AE register to the distance between the ending address of the selected waveform and the ending address of the last waveform. The FIFO is set to Mode 1, and the data is read out of the FIFO according to a constant clock when the $\overline{AE}$ flag is asserted, RT is asserted, and the RP is set to the value of RBASE, which is the beginning of the waveform

SUMMARY

This Application Note presented a new Retransmit mechanism that allows for a flexible control on RP of the FIFO. Basically the new Retransmit consists of four operations: Mark, Retransmit Read, Retransmit Read and Mark, and Retransmit Write. The capability of the new Retransmit was demonstrated by four examples from different engineering areas.

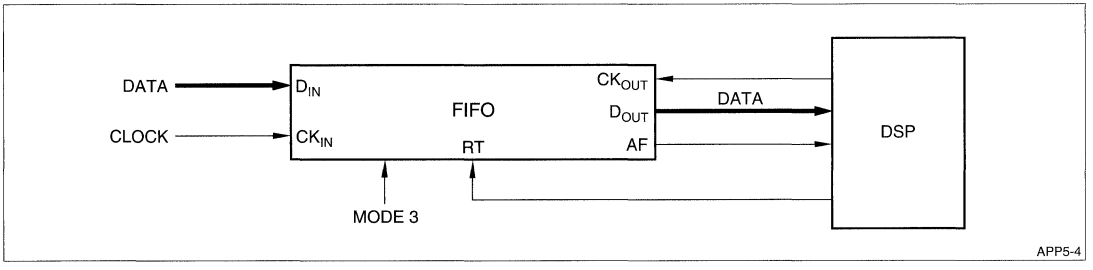


Figure 4. FIFO-DSP Connection For Overlapped Addressing

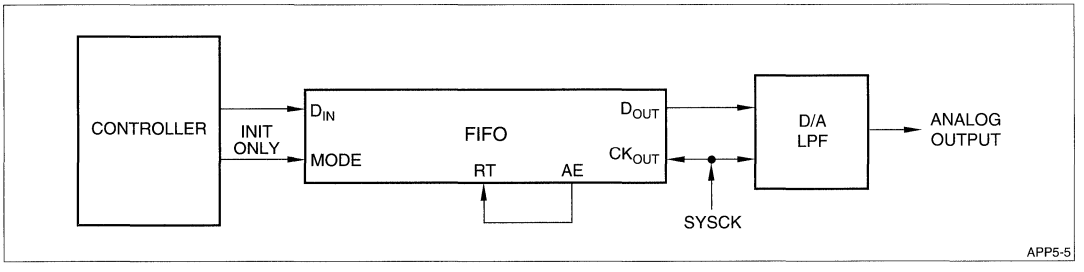


Figure 5. Waveform Generator

FIFO MEMORIES: EFFECTIVE, COMPACT, AND EASY TO USE

Chuck Hastings, Marketing/Applications Manager, FIFO and Specialty Memories

INTRODUCTION

FIFO memories ('FIFOs') are VLSI semiconductor memory devices which store information temporarily.

The word 'FIFO' is an 'acronym' in English. An acronym is a word formed from the first letters of other words, in this case 'First-In, First-Out.' This term was first used in the mathematical field of operations research to describe one particular 'queue discipline,' that is, priority rule for dealing with the members of some 'queue.' The other major possible queue discipline is 'Last-In, First-Out,' or 'LIFO.' The terms 'FIFO' and 'LIFO' also are used in the fields of cost accounting and inventory control, for the two major methods of assigning a current cost to an item which has been withdrawn from an inventory stock of items bought at varying times and at varying prices.

Data words may be pushed into a FIFO memory in sequence, one word at a time, without ever having to give the FIFO any 'address' to tell it exactly where to store a given word. Later on, the same sequence of data words may be pulled out of the FIFO at its other end, one word at a time, in the same order in which they entered the FIFO. Because the position of each word remains the same within the block of words, there is no need to 'address' the FIFO to tell it where a particular word is to be found.

Single-chip FIFO devices of several different word widths are available. The most common word width today is 9 bits. But there still are older FIFOs being sold with word widths of 4 or 5 or 8 bits, and newer FIFOs with word widths of 18 or even 36 bits. Sharp has been the first FIFO manufacturer to have brought 36-bit FIFOs to market.

FIFO memories are used for several diverse purposes in digital systems:

- Matching different data rates.
- Holding data temporarily, away from any main memory.
- Eliminating 'skew' between parallel data streams.
- Acquiring a time-sequential record of successive events.
- Reading out the same data over and over again repetitively.

Some uses of FIFOs perhaps may serve more than one of these purposes.

MATCHING DIFFERENT DATA RATES

Data-rate matching probably is the most common type of FIFO application. There are three common patterns for the timing of a stream of data words. The data words may arrive and/or depart

- at RANDOM times, individually.
- in BURSTS, that is, in blocks of some given number of words.
- at a CONSTANT rate.

A FIFO may be used to hold input-data-stream words, which are arriving according to any of the above timing patterns, for a variable and controllable duration, so that they depart according to the timing pattern needed for output-data-stream words. On the average and over the long term, however, exactly as many words must depart from the FIFO as previously arrived – no more, and no less. Otherwise, either the FIFO memory would become permanently full, or else it would become permanently empty.

Since there are three timing patterns for the arrival of data words, and also the same three timing patterns for their departure, there are nine possible timing patterns in all for a FIFO's complete data-transfer task. Probably the four most important of these are 'random-to-random,' 'burst-to-burst,' 'burst-to-constant,' and 'constant-to-burst.'

The 'random-to-random' case is referred to as the 'single-server-problem' in operations-research mathematics, where it is often described using the illustration of hungry customers arriving at a hamburger stand at random intervals, and departing with their hamburgers at different random intervals.

Often some electromechanical peripheral device is connected to one end of a FIFO, while a purely electronic device is connected to the other end. Electromechanical

devices usually must acquire or emit data at fixed times, which are a function of their own design; their timing can't be rearranged on command by an electronic system. On the other hand, most all-electronic devices are much more able to adjust their data-transfer timing to the needs of other devices with which they are communicating, although there are some major exceptions.

Even all-electronic devices may have their operating efficiency badly degraded, if they are subjected to constant 'overhead' interruptions by electromechanical devices. Thus, FIFOs often perform in the useful role of isolating all-electronic devices such as microprocessors from common electromechanical devices such as disk memories, laser and ink-jet printers, and bar-code readers.

As mentioned, there are some important all-electronic devices, without any moving parts, with timing characteristics just as inflexible as those of electromechanical devices. Many kinds of computer-graphic displays must be refreshed or changed at some predetermined rate. A telephone-line network or a local-area network likewise has an established data-transfer frequency, to which any external device connected to the network must adapt itself, often via a FIFO.

Also, industry-standard databus specifications restrict the timing allowed for many signals, so as to ensure reliable communications interfacing between hardware modules designed independently by different groups of people.

Most newer FIFO memory devices provide built-in status outputs, called 'flags,' which help the system's controlling logic implement efficient FIFO-memory operating strategies. Typically, there are five of these flags: 'Full,' 'Almost-Full,' 'Half-Full,' 'Almost-Empty,' and 'Empty.' The 'Almost' flags may be used as advance warnings to the system that some action must be taken soon. For, once the FIFO memory has reached the 'Full' condition or the 'Empty' condition, it may already be too late to avoid losing or scrambling data.

One simple but effective control strategy, for 'burst-to-burst' FIFO operation, is to select a FIFO memory having a depth of twice as many words as the usual system datablock size. A new block of data is brought in whenever the 'Half-Full' flag indicates that the memory is now less than half full, thus making efficient use of the FIFO's memory capacity.

There also are some more sophisticated strategies, which use the 'Almost' flags as well. In many FIFO devices, the 'offset' values for these 'Almost' flags are 'programmable.' This 'offset' value is a specification by the system of the number of FIFO-memory words by which 'Almost-Full' is to differ from 'Full,' and may be changed by the system during operation; likewise for 'Almost-Empty' and 'Empty.'

Occasionally there is a 'constant-to-constant' FIFO application. Here, the FIFO has to serve as a pure time-delay element, and the data rates at both ends of the FIFO must be exactly the same.

HOLDING DATA TEMPORARILY, AWAY FROM ANY MAIN MEMORY

The information-storage capacity of FIFO memories usually is smaller than that of random-access memories ('RAMs'). Nevertheless, today's FIFO memories are large enough to fulfill many needs for 'local' storage at various points in a system remote from any 'main memory.'

FIFOs don't require that the system create or monitor any 'addresses' for the information stored within them. Whatever goes in one end simply comes out the other end, in the same order. Thus, FIFOs provide a convenient form of 'data storage at a point.' In some cases, what is being stored is control information, such as commands or instructions; it is not necessarily always 'data.'

One other unique and valuable property of FIFOs is that, within one given 'family' of architecturally-compatible FIFO parts, FIFOs having different sizes of internal memory are DIRECTLY 'drop-in-compatible' with each other without ANY system redesign. Thus, a 'deeper' FIFO, having more memory-word capacity, often is used to replace a smaller or 'shallower' FIFO, which may have turned out not to have enough memory-word capacity for the application.

Sometimes, the need to replace smaller FIFOs with larger FIFOs may come about because of a deliberate system-engineering strategy, of upgrading the performance of a microprocessor-based system design by increasing the system datablock size. With larger datablocks, the microprocessor does not need to be interrupted as often in order to deal with FIFO data, and can be focused more efficiently on its main tasks.

ELIMINATING SKEW BETWEEN PARALLEL DATA STREAMS

Parallel data streams which are supposed to be mutually synchronized, but which get out of synchronization by varying amounts, are said to be 'skewed' with respect to one another. An interesting specialized FIFO application is that of 'deskewing' the individual bits read from the parallel bit-recording longitudinal data 'tracks' of a multiple-channel magnetic tape. The magnetic-tape skew problem arises primarily because a particular tape may first have been recorded on one physical tape drive, and later on moved to a different physical tape drive for reading. Similar applications arise when dealing with other not-too-precise synchronous devices.

The solution to this design problem, in its usual form, requires one FIFO memory device per magnetic-tape longitudinal data track. The FIFO devices are operated at a repetition rate which is much higher than the databits-per-unit-time rate for each magnetic-tape track. Using an algorithm which depends on the magnetic recording format in use, the system must continuously determine the current phase timing for each track, and then must advance or hold back the unloading of each FIFO to keep all of the FIFO outputs in phase together.

ACQUIRING A TIME-SEQUENTIAL RECORD OF SUCCESSIVE EVENTS

A FIFO, with its data inputs connected to some 'interesting' system point, can serve as an 'event recorder.' In such applications, the FIFO's 'Write Clock' or 'Shift-In Strobe' (or other signal which tells the FIFO when it is to write in a new data word) is controlled by system logic, which decides WHEN the information at that system point is 'interesting' and hence should be captured by writing it into the FIFO.

One 'event-recording' application for FIFOs is that of serving as the data-capture memory within a logic analyzer or a medical ultrasound scanner. A characteristic feature of such instrumentation applications for FIFOs is that the very same data, once captured, must be read out over and over again. Many FIFO devices have a 'Retransmit' facility which allows doing so.

A similar application, within a processor, is that of a 'jump-trace memory.' The FIFO's data inputs are connected to the processor's 'program counter' or 'microprogram counter' (or other control-sequencing register), and a FIFO input word is recorded every time that a jump out of the normal incremental program-step-addressing sequence occurs. Such a 'jump-trace memory' feature can be very helpful when debugging real-time operating systems; if the system crashes, having a 'trace' record of the last few dozen jump addresses can be very valuable in analyzing the cause of the crash, since in real-time applications there usually isn't any way to repeat the exact sequence of events which led up to the crash.

Another processor application for FIFOs is the queueing of 'interrupts,' which are high-speed requests for the processor's attention. In certain types of systems, such as digital-telephony processors, the number of real-time interrupts is extremely high; but a delay of a few milliseconds is permissible when responding to each of them. A useful technical approach in such cases is to form an encoded interrupt-information word by hardware means, for each interrupt as it is received, and store these interrupt-information words in time-sequenced order within a FIFO, to be unloaded and dealt with whenever the processor has the free time to do so.

READING OUT THE SAME DATA OVER AND OVER AGAIN REPETITIVELY

When implementing reliable data-communications hardware which must use an unreliable communications channel, it is customary for the receiving device to transmit an 'acknowledge' signal back to the sending device to indicate that the last datablock was received, and that it appeared to be correct. If no such 'acknowledge' signal gets received by the sending device, or if – worse yet! – a 'negative-acknowledge' signal gets received instead, then the sending device must transmit the erroneously-received datablock another time. A FIFO with a 'Retransmit' facility, as described above, is a convenient means of implementing this capability within the sending device.

A waveform generator is a laboratory instrument which can produce a repetitive waveform of any desired shape or frequency, up to its performance limits. Some waveform generators use digital electronics; the heart of one of these may be a large-but-slow read-only memory, which holds an extensive library of waveforms expressed as successive digital sample values, plus a FIFO. The FIFO loads up the desired waveform once, at the read-only memory's slow speed; then it reads out the waveform over and over again, using its 'Retransmit' facility, often at a much higher speed.

SUMMARY

Once, FIFOs were rather high-priced as compared to other digital semiconductor components. Now, competition and manufacturing improvements have reduced the price of FIFOs substantially. Today, more than ever, FIFOs should be considered for many applications for which they have always offered a superior design approach, but for which they have been considered too expensive. If your application can be described as 'data-rate matching,' 'local memory,' 'skew elimination,' 'sequential event recording,' or 'repetitive readout,' a FIFO is your best answer.

PARITY CHECKING IN THE LH5420 BiFIFO

Bob Laird, Field Applications Engineer

INTRODUCTION

Sharp's LH5420 256 × 36 × 2 bidirectional FIFO (BiFIFO) was the first monolithic FIFO from the semiconductor industry to offer customers a full 36-bit data word-width.

The LH5420's 36-bit wordwidth makes it the ideal choice for those 32-bit or 64-bit systems which use parity. It remains an ideal choice, even for those wide-word systems that don't use parity, because there are no available 32-bit FIFOs, and because systems frequently use an extra bit per byte for some type of 'tag' information anyway.

WHAT IS PARITY?

Many systems need to have error detection and correction (EDAC) implemented into the design, but cannot afford the added overhead – memory requirements increase dramatically. The most common compromise is to use just error detection.

By adding a single bit to each byte of data, the occurrence of single-bit errors may be monitored. This extra bit is referred to as the *parity bit.*

Parity is not a perfect means of detecting errors; but it is easy to implement. In fact, many off-the-shelf memory, microprocessor, and general-purpose-logic integrated circuits make provision for parity bits.

A system using parity requires a *parity-generation* circuit. The purpose of a parity-generation circuit is to examine each data byte, and set that byte's parity bit based on the number of ones (logic 'HIGH' levels) in the eight data bits.

Either 'odd parity' or 'even parity' may be chosen. These are defined as follows:

- *Odd Parity*: The parity bit is set to one if there is an even number of ones; otherwise, it is cleared (not set). Thus, the total number of ones, *including* the parity bit, is odd.
- *Even Parity*: The parity bit is set to one if there is an odd number of ones; otherwise, it is cleared (not set). Thus, the total number of ones, *including* the parity bit, is even.

In either case, a single-bit error in any bit position within a data byte, including an error in the parity bit itself, causes a parity-error indication.

The type of parity chosen may be dictated by the system design. However, neither type of parity offers any major economic or performance advantage over the other one. Some designers prefer odd parity, on the grounds that the correct parity bit is a one for an all-zero byte. Thus, a *completely*-blank byte (which presumably arose from some type of data-recording failure) shows up as having a parity error.

In any case, it is a good idea to maintain the same type of parity throughout the entire system.

LH5420 PARITY CHECKING

The LH5420 offers parity checking, which means that it examines each data byte to determine what its parity is, thereby computing an internal parity bit for that byte. This parity bit then is compared with the Parity Mode bit in the LH5420's Control Register. If they do not match, then there is the presumption of a parity error in that byte.

Since an LH5420 word is 36 bits wide, parity checking is done in parallel on the four bytes that make up a 36-bit word. If an error is detected in ANY of these four bytes, the parity flag is asserted LOW, indicating the parity error.

Figures 1 and 2 show the parity-flag logic for Port A. The ten-bit parity trees for the other three bytes are configured identically with the one shown in Figure 1, which is for the low-order byte. Also, the parity-flag logic for Port B is configured identically with that for Port A.

Since there is parity-flag circuitry associated with both LH5420 ports, data words may be checked both entering and exiting the BiFIFO. The parity-tree inputs are connected to the actual port input/output *pads*, through isolation transistors. Data words are monitored as they change on the system bus, which connects to the pads at that port.

It should be noted that the parity-flag circuitry at an LH5420 port is *not* controlled by the Output Enable ($\overline{OE}$) control signal at that port. Hence, parity checking *always* is active, regardless of whether or not the BiFIFO itself is driving the system bus. Therefore, as data constantly

changes on the bus, the parity flag may keep on changing states also. Some transitions on the data lines may make it appear that a parity error has occurred.

According to the LH5420 data sheet specification, the parity flag at a port is not guaranteed to be valid, until some given delay time *after* stable data is present at the input/output pads. For the LH5420P-25, for instance, this delay time is a maximum of 17 ns. Thus, to monitor the parity of data entering the LH5420, the sum of the data setup time and the data hold time must be sufficient to satisfy this 17 ns delay time.

Also, when monitoring the parity of data words being read from the LH5420, it is necessary to keep $\overline{OE}$ asserted long enough to satisfy this parity-flag delay time. If the LH5420 is being operated at the minimum cycle time of 25 ns, then $\overline{OE}$ would have to remain asserted 6 ns into the next cycle. In most designs, this requirement does not present a problem.

As it is, an LH5420 cannot perform a read operation and a write operation on *consecutive* cycles, when operating at the *minimum* cycle time. The reason is that the sum of the data-access time (for outbound data words) and the data-setup time (for inbound data words) already exceeds the minimum cycle time, for all three LH5420 speed grades.

The LH5420 Control Register Parity Mode bit is initialized for odd parity during a reset operation. However, it may be reprogrammed for even parity, and then back to odd parity, at will during system operation. Programming is performed by loading a full 36-bit word into the Control Register from Port A, with the code HLL at the LH5420's Address (resource-register-selection) inputs A_{2A}-A_{0A}. The least-significant bit of that 36-bit word, D_{0A}, selects the Parity Mode: HIGH for odd, and LOW for even. The Control Register is written into at the next rising edge of CK_A.

Note that the Control Register is a 'blind' or write-only register. The system cannot read it back.

The normal convention for parity-bit position is to use the most-significant bit of each byte; that is, D_{8A}, D_{17A}, D_{26A}, D_{35A}, D_{8B}, D_{17B}, D_{26B}, and D_{35B}. Now, the LH5420 is designed in such a way that *any* bit position within a data byte may be used as the parity bit. However, when programming the Almost-Full Flags and the Almost-Empty Flags, the offset values are written into the resource register using data pins D_{0A}-D_{7A}, D_{9A}-D_{16A}, D_{18A}-D_{25A}, and D_{27A}-D_{34A}. Thus, there is a presumption that the least-significant eight bits of a byte are 'the data bits.'

So, for systems using byte parity with the most-significant bit as the parity bit, which use the programmable-flag feature. The fields line up as they should. Otherwise, special consideration must be given when using the parity bit as a data bit for programming.

Figure 1 is a simple schematic for the ten-bit parity-checking logic circuit for one byte, in this case the least-significant byte at Port A. The tenth bit is the Control Register Parity Mode bit. Since this bit is HIGH for odd parity and LOW for even parity, the output of the ten-bit parity checker is LOW whenever the parity of the data byte being examined agrees with the LH5420's Parity Mode bit.

The parity-circuit outputs for the four bytes of Port A, designated in Figure 1 as PC_{0A}-PC_{3A}, are NORed together to compute the parity flag $\overline{PF}_A$. Similar logic is used at Port B, to compute $\overline{PF}_B$.

Note that the $\overline{PF}_A$ and $\overline{PF}_B$ NOR gates do not incorporate latches. The states of these two outputs depend directly on the data words present at Port A and at Port B respectively.

USING PARITY CHECKING TOGETHER WITH WORDWIDTH SELECTION

Another useful feature offered by the LH5420 BiFIFO is *wordwidth selection*, also known as *funneling/defunneling*, at Port B. Port A always assumes a wordwidth of 36 bits. But Port B may assume a 36-bit, 18-bit, or 9-bit wordwidth, according to the setting of two LH5420 control input signals, WS_0 and WS_1. In the latter two cases, there is an effect on parity checking at Port B.

For example, assume that Port B has been set for 9-bit wordwidth. As a byte is being written into D_{0B}-D_{8B}, the lowest-order Port B byte, the Port B parity checker is monitoring not only this byte, but also the three *unused* bytes D_{9B}-D_{35B}. If any of those three bytes has improper parity, according to the current Parity Mode setting, then $\overline{PF}_B$ is asserted LOW to indicate a parity error.

If the Port B 9-bit-wordwidth setting never changes during system operation, then it may be advantageous to tie pins D_{9B}-D_{35B} all HIGH for odd parity, or all LOW for even parity. Similar statements apply for pins D_{18B}-D_{35B}, given the use of 18-bit wordwidth at Port B.

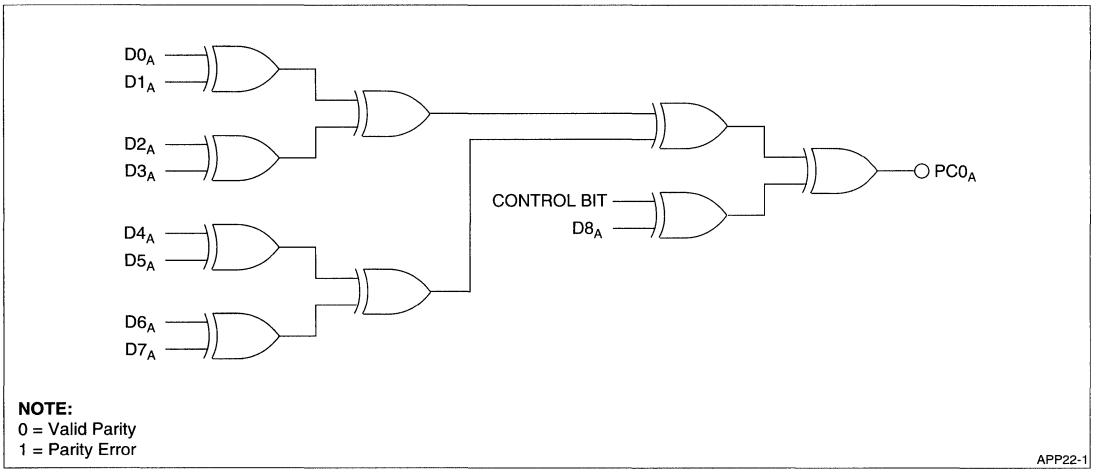


Figure 1. Example of Byte-Parity Gate: Port A, Low Order Byte

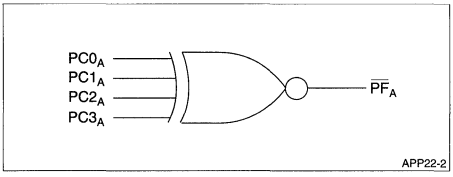


Figure 2. Example of Parity-Flag Logic: Port A

When outputting a byte from Port B with 9-bit word-width in effect, the LH5420 actually conveys a full 36-bit word to the output buffers; although, presumably, the other devices on the output bus are examining only D_{0B}-D_{8B}. Presenting the remaining bytes to the output bus is accomplished by circular-shifting operations, which are inherently available in the Port B output logic. The second byte of the word is read by shifting D_{9B}-D_{17B} down to D_{0B}-D_{8B}. Concurrently, D_{27B}-D_{35B} are shifted down to D_{18B}-D_{26B}. D_{18B}-D_{26B} are shifted down to D_{9B}-D_{17B}, and D_{0B}-D_{8B} are shifted circularly to D_{27B}-D_{35B}.

This circular-shifting procedure is repeated twice more, for the remaining two bytes of that data word. After that, another full 36-bit data word is again fetched, to output the next byte. Assuming that the fetched data bytes all have correct parity, then $\overline{PF}_B$ continues to indicate correct parity, after each successive circular shift of the data word.

SUMMARY

Sharp's LH5420 256 × 36 × 2 BiFIFO features two built-in parity checkers, one connected to each 36-bit data bus. The parity of each 9-bit data byte may be checked twice as it passes through an LH5420: once as it enters, and once as it leaves.

An LH5420 parity checker may be used to examine a data word which is present on either of the LH5420's two 36-bit data buses, even if that word never actually gets read into one of the LH5420's internal FIFO memories.

The LH5420 may be programmed to consider either odd parity or even parity as correct. A parity error in any byte, of a data word passing through an LH5420 port, is signaled by asserting the LH5420's parity flag at that port LOW, as long as that data word remains present there.

Proper use of this port parity flag can detect *any* single-bit error in *any* data byte passing through an LH5420, assuming that the parity flag gets read at the appropriate time. No additional hardware is required.

PQFP-to-PGA CONVERTER

Dan Holton, Sr. Product Engineer

A PQFP-to-PGA converter allows a system board designer the ability to use fine-pitch surface-mount PQFP packages with thru-hole board designs. For example, Sharp's LH543620P PQFP would be soldered onto the converter's printed circuit board PQFP land pattern/footprint, then the converter's PGA pins dropped into a thru-hole board design. The board designer need only layout the thru-hole board according to the LH543620P PGA converter pin assignment.

Sharp recommends converters from ITT Pomona Electronics, specifically ITT SMT/PGA Generic Converter Model No. 5853. This converter accepts a 132-pin PQFP (25-mil pitch) and maps all pins to a generic 13 × 13, 132-pin PGA (100-mil pitch). For ease of use with Sharp's LH543620P, a pin mapping table for the LH543620P is shown in Table 1; however, ITT's generic converter is device independent. Physical PQFP pin to PGA pin assignment layouts are shown in Figures 1-3. Figure 4 shows the specific dimensions of the ITT converter. Please note that the PQFP pin numbering scheme on the converter itself may differ from the numbering scheme of the LH543620P. The PQFP pin numbering labels on the ITT converter can simply be ignored.

To order converters, contact ITT directly at:

ITT Pomona Electronics
1500 East Ninth St.
Pomona, CA 91766
(909) 469-2900

Table 1. LH543620P Pin List and PQFP-to-PGA Pin Mapping

PIN NAME	PQFP PIN NO.	PGA CONVERTER PIN NO.
D_{14}	1	A7
D_{13}	2	C7
D_{12}	3	B6
D_{11}	4	C6
D_{10}	5	A6
D_9	6	D6
D_8	8	C5
D_7	9	A5
D_6	10	D5
D_5	11	B4
D_4	12	C4
D_3	13	A4
D_2	14	B3
D_1	15	A3
D_0	16	C3
$\overline{MEF}$	18	B1
$\overline{MFF}$	19	B2
$\overline{EF}$	20	C1
$\overline{AE}$	21	C2
$\overline{HF}$	23	D3
$\overline{AF}$	24	D2
$\overline{FF}$	25	E4
$\overline{PF}$	26	E1
CKO	27	E3
Q_{35}	29	F4
Q_{34}	30	F1
Q_{33}	32	F2
Q_{32}	33	G2
Q_{31}	35	G3
Q_{30}	36	H2
Q_{29}	38	H1
Q_{28}	39	H4
Q_{27}	41	J3
Q_{26}	42	J1
Q_{25}	44	K2
Q_{24}	45	K3
Q_{23}	47	L2
Q_{22}	48	L1
Q_{21}	52	M2
Q_{20}	53	N3
Q_{19}	55	N4
Q_{18}	56	L4
Q_{17}	58	K5
Q_{16}	59	N5

PIN NAME	PQFP PIN NO.	PGA CONVERTER PIN NO.
Q_{15}	61	M5
Q_{14}	62	K6
Q_{13}	64	L6
Q_{12}	65	M6
Q_{11}	67	N7
Q_{10}	68	L7
Q_9	70	L8
Q_8	71	N8
Q_7	73	M9
Q_6	74	L9
Q_5	76	K9
Q_4	77	M10
Q_3	79	N10
Q_2	80	M11
Q_1	82	L11
Q_0	83	N12
$\overline{OE}$	85	M12
$\overline{RT}$	86	L13
$RTMD_1$	87	L12
$RTMD_0$	88	K13
$\overline{RS}$	89	K11
WSO_1	90	K12
WSO_0	91	J10
ADO_2	93	J11
ADO_1	94	J12
ADO_0	95	H10
ENO_2	96	H13
$EN0_1$	97	H11
$\overline{BYE}$	98	H12
CAPR	99	G12
WSI_1	101	G11
WSI_0	102	F12
ADI_2	103	F11
ADI_1	104	F13
ADI_0	105	F10
ENI_2	106	E12
ENI_1	107	E11
D_{35}	109	E10
D_{34}	110	D12
D_{33}	111	D11
D_{32}	112	D13
D_{31}	113	C12
D_{30}	114	C13
D_{29}	115	C11

PIN NAME	PQFP PIN NO.	PGA CONVERTER PIN NO.
D_{28}	116	B13
D_{27}	117	A12
D_{26}	119	A11
D_{25}	120	B11
D_{24}	121	A10
D_{23}	122	C10
D_{22}	123	B10
D_{21}	124	D9
D_{20}	125	A9
D_{19}	126	C9
D_{18}	127	B9
CKI	128	D8
D_{17}	130	C8
D_{16}	131	B8
D_{15}	132	B7
V_{SS}	7	B5
V_{CC}	17	A2
V_{SS}	22	D1
V_{CC}	28	E2
V_{SS}	31	F3
V_{CC}	34	G1
V_{SS}	37	H3
V_{CC}	40	J2
V_{SS}	43	J4
V_{CC}	46	K1
V_{SS}	49	L3
V_{SS}	50	M1
V_{CC}	51	N2
V_{CC}	54	M3
V_{SS}	57	M4
V_{CC}	60	L5
V_{SS}	63	N6
V_{CC}	66	M7
V_{SS}	69	M8
V_{CC}	72	K8
V_{SS}	75	N9
V_{CC}	78	L10
V_{SS}	81	N11
V_{CC}	84	M13
V_{CC}	92	J13
V_{SS}	100	G13
V_{CC}	108	E13
V_{SS}	118	B12
V_{CC}	129	A8

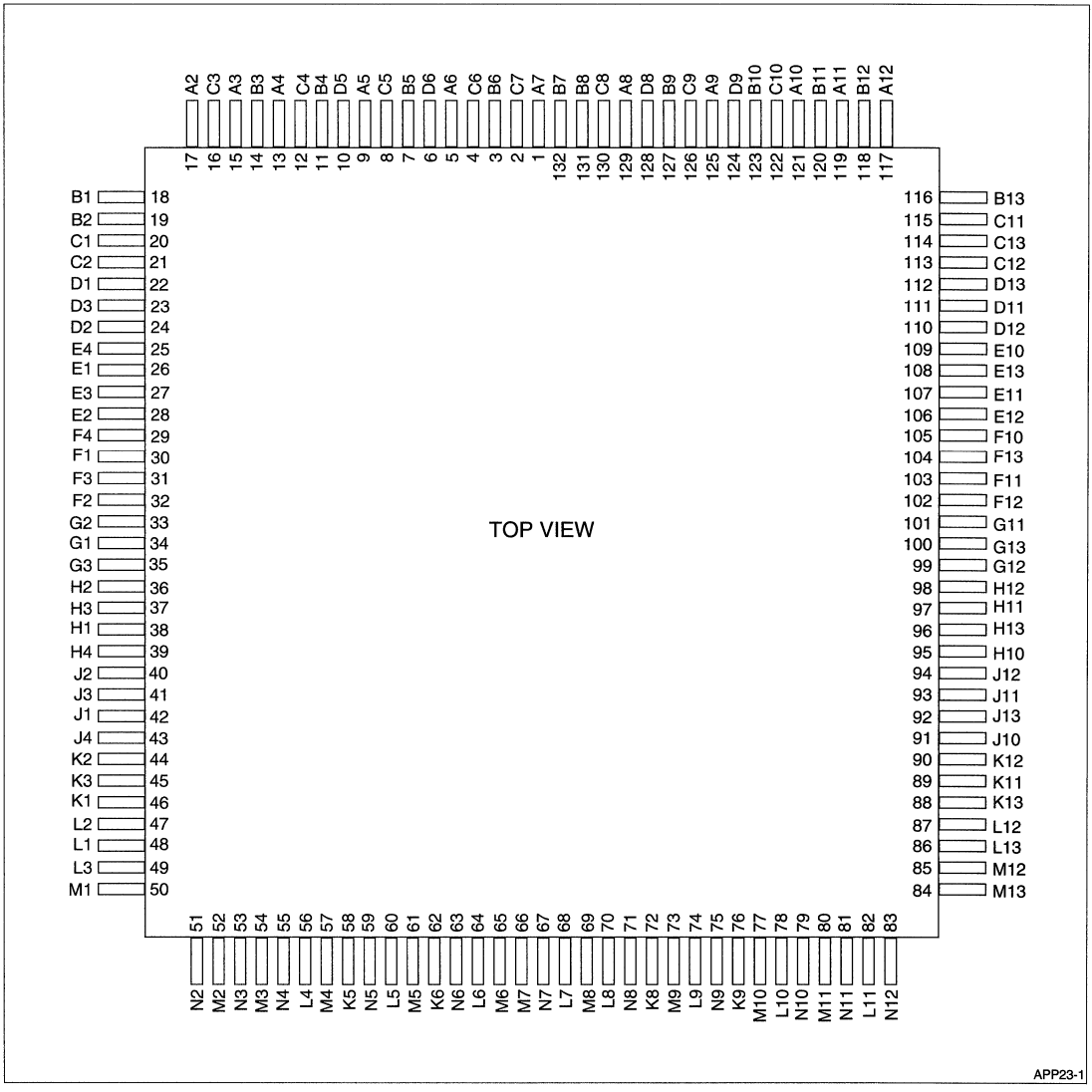


Figure 1. LH543620P PQFP Pin Assignment (Top View) Showing Converter [1] PGA Pin Labels

NOTE:

1. ITT Model No. 5853.

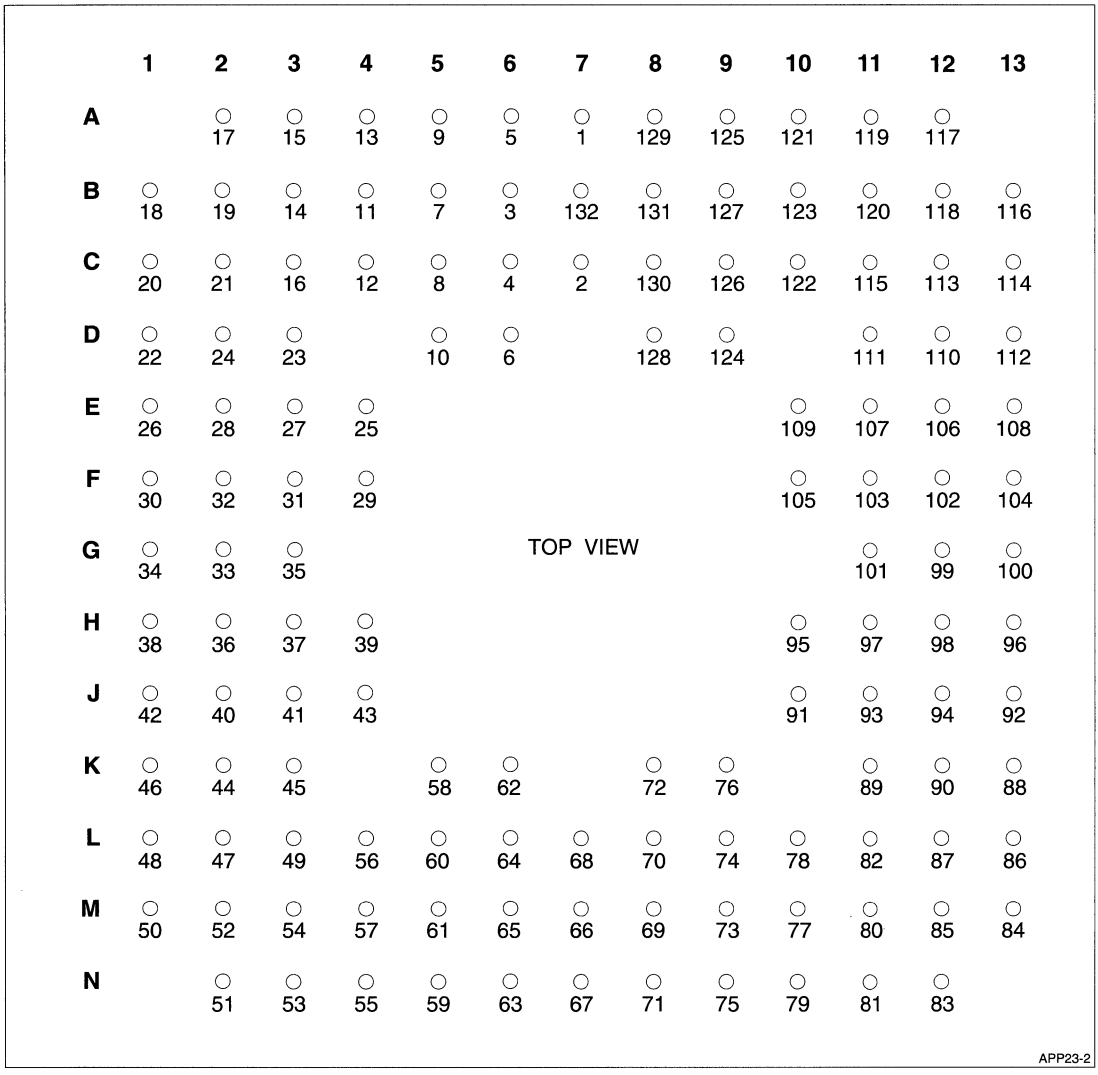


Figure 2. Converter [1] PGA Pin Assignment (Top View) [2] Showing LH543620P PQFP Pin Numbers

NOTES:

1. ITT Model No. 5853.
2. This Top View is equivalent to how the actual system board would have to be layed out for the converter pins.

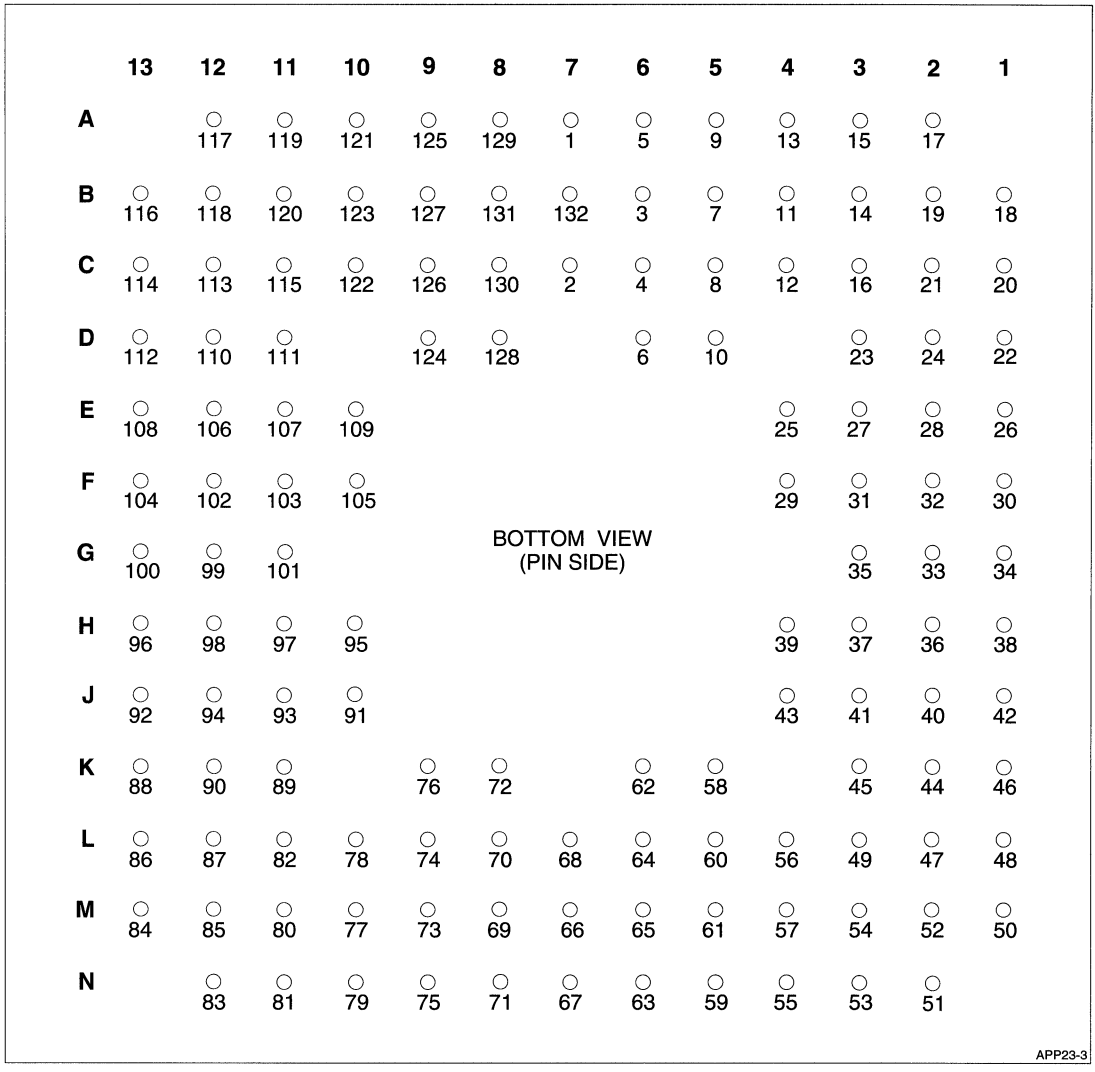


Figure 3. Converter [1] PGA Pin Assignment (Bottom View) Showing LH543620P PQFP Pin Numbers

NOTE:

1. ITT Model No. 5853.

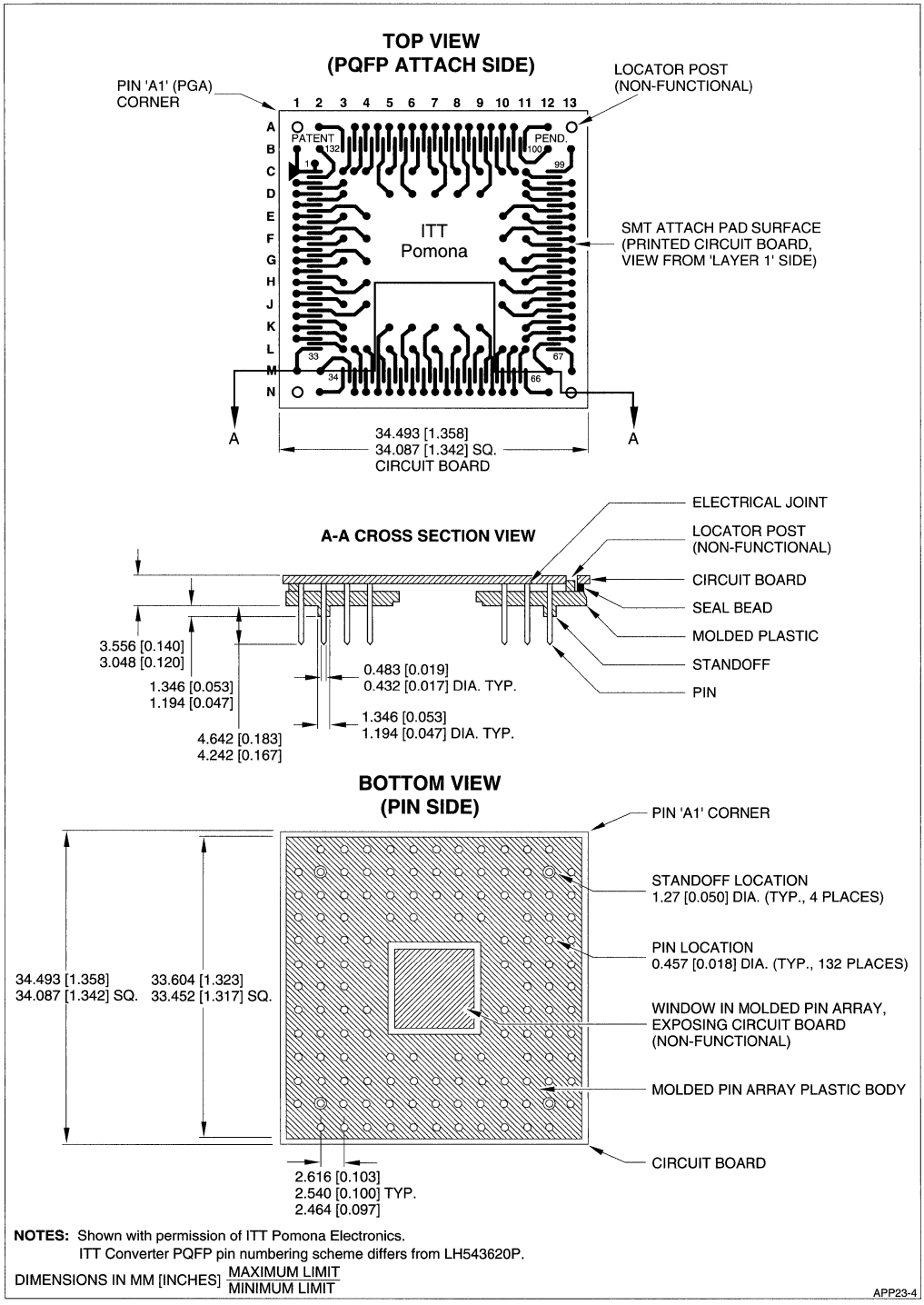


Figure 4. PQFP-to-PGA Converter Specifications (ITT Model No. 5853)

FIFO FLAG TIMING: MARCHING TO TWO DIFFERENT DRUMMERS

Chuck Hastings, Marketing/Applications Manager, FIFO and Specialty Memory

INTRODUCTION

A FIFO (First-In, First-Out memory device) can. be thought of as a 'passthrough window' between two *independently-clocked* digital subsystems – each marching to its own drummer, as it were. To adequately perform this role in practical digital systems, FIFOs are 'schizoid': they are memory devices, and they also are logic-synchronization circuits.

Increasingly, newer FIFOs have been made capable of recognizing commands from the overall digital system of which they are part, and also of providing that system with timely status information in the form of 'flag' values.

A FIFO can undergo a change in its status because of an event synchronized *either* to the subsystem at its input end, *or* to the subsystem at its output end. These two subsystems usually are independently clocked; and either of the two clock signals can affect a flag value. The external logic normally uses one or the other of these clock signals to synchronize reading the flag, which leads directly to a potential metastability hazard. The very newest FIFOs incorporate internal logic to minimize any practical metastability hazard; but, even then, their specifications have to include 'skew parameters,' which must be complied with by the external logic.

This paper includes a short review of FIFO basics, followed by a more detailed discussion of flag-synchronization and metastability issues. Finally, the flag-synchronization features of the new Sharp LH540215/25 18-bit synchronous FIFOs are described, as a case in point.

WHAT IS IT THAT FIFOs DO, IN THE FIRST PLACE?

Conceptually, using a FIFO as a 'passthrough window' accomplishes two related but different system functions:

- *Providing 'rubber-band memory'* between Subsystem A and Subsystem B.
- *Decoupling the clocking* of each subsystem from that of the other subsystem.

As long as each subsystem deals with the other subsystem through a FIFO, the two subsystems don't have to be synchronized or coordinated with each other in *any* way, at least for the purpose of passing a data stream from one to the other. This is a powerful *simplifying assumption* in digital-system design and integration. FIFOs often are used to save design effort and leadtime, and to reduce the complexity of timing logic, when interfacing two independent subsystems together.

Subsystem A, functionally upstream, stuffs data into a FIFO at whatever rate Subsystem A likes best; and Subsystem B, functionally downstream, unloads the data from the FIFO at whatever rate Subsystem B likes best. These rates may be derived from oscillator-driven clock circuits; or, they may not even be constant, if the 'clock' actually is an aperiodic 'demand' signal.

To cope with the requirements of this usage, a FIFO must include an *input register* which can be synchronized or 'clocked' by Subsystem A, and an *output register* which can be synchronized or 'clocked' by Subsystem B. In between, there needs to be the 'rubber-band memory.' (Also known as 'elastic storage,' if you prefer more formal-type technobabble.) This rubber-band memory should appear both to Subsystem A and to Subsystem B as having unlimited depth; that is, as containing enough words that neither subsystem ever bumps up against finite FIFO-memory capacity as a bothersome constraint. If you've designed with FIFOs previously, then maybe you knew all that.

Historically, FIFOs have been perceived by system-design engineers as *premium* semiconductor parts which Make Problems Go Away. That is, FIFOs are supposed to be *liberating*. It follows, of course, that these same customers can become severely annoyed with FIFOs, if they ever begin to feel that using FIFOs gets them entangled in lots of timing hassles. But, sometimes, they fail to distinguish between those timing issues which arise from the use of a particular part, and those timing issues which arise *intrinsically* from the characteristics of their application.

REAL-WORLD MATHEMATICAL ISSUES WHEN USING FIFOs

The term 'FIFO' comes from operations research, a branch of mathematics. It stands for a *queue discipline*, that of 'First In, First Out.' The term FIFO also turns up in cost-accounting methodology, for reckoning the cost of items being withdrawn from an inventory after having been bought at varying prices.

The operations-research 'single-server' model fits FIFOs being used to buffer 'randomly-arriving' data, as well as fitting the classic example of hamburger stands serving 'randomly-arriving' customers. In this model, the arrival of a piece of data at the FIFO's input end ('a customer') and the unloading of a piece of data from the output end of the FIFO ('serving a customer') are random events. In between the input end and the output end, there is a *queue* of customers waiting to be served – here, words being temporarily stored within the FIFO's internal memory array. If you like mathematical lingo, the entire scene is a 'stochastic process,' meaning that the events which are occurring are 'probabilistic' and not 'determinate.'

Why is this operations research model of interest? In the 'single-server' model, the *queue length* is a 'random walk,' and – if it is given enough time to fluctuate – is *unbounded*, meaning that no FIFO used to queue up *truly* random data can *ever* be quite big enough, if it has to operate for an infinitely-long time. Thus, this operations-research model *in principle* contradicts the common-sense notion that there is some 'large-enough' size for every FIFO application.

So, it's *not* just that semiconductor manufacturers don't *choose* to develop FIFOs having deep-enough internal memory-array capacities. It's that there *isn't* any finite depth which can provide theoretically-*absolute* protection, in the *truly*-random case.

DESIGN STRATEGIES USING FIFOs

But, *in practical systems*, there always *is* a 'large-enough' FIFO size. The probability of any queue length growing past a certain value, before the next ice age, usually can be made *very* small without making the FIFO's queue-length capacity uneconomically large.

If a catastrophic increase in queue length starts to occur, often because of some sudden-onset system problem, the designer's line of defense is the FIFO's 'fullness flags.' If the system is monitoring these, they can provide an adequate early warning of impending trouble, so that the system can take timely and effective countermeasures.

The key point here is that *finite*-depth FIFOs – and the semiconductor industry hasn't built any other kind of FIFOs yet! – *must*, for basic *mathematical* reasons, incorporate some kind of bulletproof logic for dealing with 'full' and 'empty' queue conditions.

Of course, there are also 'block-oriented' FIFO applications, in which Subsystem A dumps another block of data into the FIFO only when the FIFO is less than half full, and the FIFO size has been chosen a priori to be large enough to hold *two* blocks of data. Here, Subsystem A is doing some of the 'elastic' adaptation to fluctuations in the data rate, and is no longer depending upon the downstream FIFO to do all of that adaptation. Consequently, not all of the pessimistic assumptions of the 'single-server' model still have to be satisfied. But limiting the queue length in this manner can exact a price, in some other system-level performance characteristic.

Prudent digital-system designers usually try to avoid *ever* allowing their FIFOs to get *either* completely 'full' *or* completely 'empty' during normal system operation. Following this rule means that many fascinating 'gotchas,' which in principle can raise their ugly heads at these boundary conditions, never get to disturb the serenity of the system's operation, or of the system designer's working day. However, the Product Engineering and Test Engineering staffs of semiconductor manufacturers often must spend major time and effort dealing with these boundary-condition 'gotchas.'

FIFO 'FULLNESS' STATUS FLAGS

Contemporary FIFOs commonly provide *several* status flags, to inform the system logic external to the FIFO as to the FIFO-memory array's relative state of fullness or emptiness. Often, there are five such flags: 'Full,' 'Almost Full,' 'Half Full,' 'Almost Empty,' and 'Empty.' These 'Almost' flags originated some years ago, as a semiconductor-piece-part adaptation of the 'Yellow Warning' interrupt-flag scheme used in DEC's PDP-11 family of minicomputers.

Sometimes, when there aren't quite enough pins to provide every FIFO feature desired by the semiconductor manufacturer and its customers, the 'Almost-Full Flag' signal and the 'Almost-Empty Flag' signal are combined as an 'inclusive-ORed' 'Almost Full/Empty Flag' signal on one pin. The external logic can distinguish these conditions one from another by examining the 'Half-Full Flag,' on the assumption that a FIFO which is 'almost full' can be presumed to be more than 'half full,' and likewise that a FIFO which is 'almost empty' can be presumed to be less than 'half full.'

This last assumption sounds obvious to the point of silliness. Still, it needs to be *rechecked* when designing with FIFOs which feature 'programmable' 'Almost-Full Flag' and 'Almost-Empty Flag' signals, whenever *large* 'offset' values are being 'programmed' into the 'offset-value registers' which are associated with each of these two flags. These 'offset' values define the 'almost-full' condition as some exact number of words away from 'full,'

and the 'almost-empty' condition as some exact number of words away from 'empty.' Obviously, the rules are going to change, if ever one or both of these offset values exceeds half of the number of words in the FIFO-memory array.

THE FLAG-SYNCHRONIZATION DILEMMA

Back to Subsystem A, which is being synchronized by Clock A, and Subsystem B, which is being synchronized by Clock B. Subsystem A is 'synchronously' dumping data destined for Subsystem B into a FIFO, which is there to make everything come out OK so that Subsystem B can 'synchronously' read these same data back out.

Whenever Subsystem A stuffs in one more word, the FIFO can suddenly become 'full,' 'almost full,' or 'half full'; or it can suddenly cease to be 'almost empty' or 'empty.' Similarly, Subsystem B unloading one word can cause the FIFO now suddenly to become 'almost empty' or 'empty,' or now suddenly to cease to be 'full,' 'almost full,' or 'half full.' Thus, each flag signal can be affected for one transition by a 'write event,' and can be affected for the other transition by a 'read event.' Therefore,

> Events which can cause the state of any 'fullness' flag to change can occur at either end of the FIFO.

So, *each* of the five flags can change state *either* in synchronism with Clock A, *or* else with Clock B. So all five flags are 'synchronized to both clocks,' which is the same as saying that they aren't *really* synchronized to either clock – that is, that they are *asynchronous*. These statements, of course, apply strictly to the *internal* values of the five flags, as these values are computed by the FIFO's control circuits.

Incidentally, if this operational description seems to be stacked lopsidedly in favor of fullness, it's because the 'Half-Full Flag' customarily is called just that, rather than being called the 'Half-Empty Flag.' Probably, the term 'Half-Empty Flag' just sounded too negativistic for the tastes of positive-thinking FIFO marketeers! Anyway, the logic which controls the 'Half-Full Flag' behaves exactly in accordance with what one would expect from that name.

Now, in a 'synchronous' FIFO having 'enable' inputs as well as 'clock' inputs for both writing and reading, a 'write event' is synchronized to the 'write clock' input, and is made to occur by asserting the 'write enable' input. Likewise, a 'read event' is synchronized to the 'read clock' input, and is made to occur by asserting the 'read enable' input.

In principle, the 'write clock' and the 'read clock' both may be free-running periodic waveforms derived from crystal-oscillator frequency sources. However, most contemporary FIFOs avoid the use of any internal circuit techniques which would *require* that these 'clock' signals must be periodic.

So, now, what happens when either Subsystem A or Subsystem B wants to *read* the value of one of these flags? If Subsystem A and Subsystem B are connected by a 'synchronous' FIFO, then the logic outside the FIFO which is trying to read the FIFO's flag(s) very likely *also* is synchronized, probably either to the 'write clock' (here, Clock A) or to the 'read clock' (here, Clock B). Which is fine, as long as the flag value is stable, or as long as the most recent event which is capable of affecting this flag value was synchronized to the *same* 'clock' signal to which flag *reading* also is being synchronized.

But what, then, if the most recent event which has affected this flag value was synchronized to the *other* 'clock' signal? Then, we have a signal getting changed in synchronism with *one* 'clock' signal, but being read in synchronism with a second, *different* 'clock' signal which is not necessarily in any way synchronized or coordinated with the first 'clock' signal. Thus, the FIFO's flag output is most unlikely *always* to be meeting the setup-time and hold-time specifications for the downstream semiconductor-device input which it is driving.

METASTABILITY

This latter situation is, of course, the recipe for instant *metastability*. Metastability is a digital-system form of 'bad vibes.' It arises as the result of trying to read a digital signal while that signal is *changing*.

A potential 'metastability hazard' exists whenever the signal being read isn't a stable HIGH or a stable LOW, as of the exact instant when the clock-signal transition edge comes along to synchronize the reading process. If there are many ongoing attempts to read the signal, and the timing of the signal is unrelated to that of the clock signal, then what we have here is another example of a 'stochastic process.' And, eventually, there is bound to be an attempt to read the signal at *exactly* the wrong time.

When attempting to read such an unstable signal, three different outcomes are possible:

- The signal can be read as a HIGH.
- The signal can be read as a LOW.
- The reading circuit element itself can just get confused and 'hang up' in a 'metastable' state, outputting a signal close to its own input 'trip point' or 'transition threshold' for a while, before settling back either into a stable HIGH condition or else into a stable LOW condition.

This 'metastable' state can, in some circuit technologies, last for as long as *several* additional clock periods. Now, the probability of 'going metastable' remains very small as long as the reading circuit isn't being operated at a frequency really close to its maximum; but it increases, quite dramatically, as this maximum is

approached. The potentially-ruinous impacts on system reliability are all too obvious.

Notice that we got into this metastability issue as a direct consequence of the burden which FIFOs are expected to take on in systems, of Making Synchronization Problems Go Away. Ultimately, metastability is not intrinsically just a circuit-design problem; it arises mathematically as an *operations-research* problem, when deterministic digital circuits are used – whether wittingly or not! – to try to deal with 'stochastic processes.'

RESYNCHRONIZING FLAG SIGNALS

FIFO customers, like customers for any other products, develop new Wants over time, as suppliers learn to satisfy their previous Wants. So customers *now* Want to be able to perform flag-reading operations from *whichever* end of the FIFO they wish, for *any* of the five usual 'fullness flags,' without having to spend time and circuit-board real estate protecting themselves against metastability hazards.

FIFO manufacturers first began to address this customer Want by attempting to handle the potential metastability problems internally *within* the FIFO, but for just two out of the five flags – the 'Full Flag,' and the 'Empty Flag.' The approach was to resynchronize each of these two flag signals, for *both* of its possible state transitions, to the clock signal for that end of the FIFO where there usually is the greatest interest in reading that flag: the 'Full Flag' to the 'write clock,' and the 'Empty Flag' to the 'read clock.'

The premise for these choices was, of course, that it's essential for Subsystem A to know if a FIFO which it's trying to write into is 'Full'; if it is, then the write operation can't be completed. But there's no similarly-urgent reason why Subsystem A needs to know if the FIFO which it's trying to write into is 'Empty.' Analogous reasoning also applies at the other end of the FIFO, where words are being read out.

USING PROGRAMMABLE FLAGS TO DO END-OF-BLOCK DETECTION

Certainly, this resynchronization of the external values for the 'Full Flag' and the 'Empty Flag' was a step in the right direction. It costs some on-chip resynchronizing flipflops; but it provides system designers with an easier-to-use FIFO part. But it didn't go quite far enough, since customers *also* turned out to be intensely concerned about getting absolutely dead-on-accurate readings from the 'Almost-Full Flag' and from the 'Almost-Empty Flag' – not just readings accurate to within a word or two.

A puzzling attitude, if one assumes that these latter flags are being used *only* as warnings to the system that it needs to do something soon, before the FIFO becomes either *completely* full or else *completely* empty as the case may be. But, as often happens, customers found a major use for the 'Almost-Full Flag' feature and the 'Almost Empty Flag' feature which was utterly different from the type of use planned for by the semiconductor manufacturers. In this case, the unanticipated major usage was as *cheap block-length counters*.

The semiconductor industry originally opened the door to this block-length-counter trick, while trying simultaneously to satisfy all of the different customers who could not agree on just what 'almost full' and 'almost empty' ought to mean – that is, how many words away from 'full' should be defined to mean 'almost full,' and how many words away from 'empty' should be defined to mean 'almost empty.'

The industry's response was to say, in effect, "Hey, you guys decide," by including *programmable registers* within FIFOs to hold 'offset values' which could be loaded during the operation of a system. These 'offset values' allowed the system to *specify* the 'how-many-words-away' parameters, during operation.

In older-architecture FIFOs which date back to when 'almost' flags first were introduced, these 'offset values' were *fixed*, 'wired-in,' architectural-design parameters. There are even certain newer FIFOs which have wired-in offset values, generally because they come in small packages without any extra pins to spare which could be used for controlling the necessary programmable registers, if their offset values were to be specifiable. Eight has been one very common wired-in offset value.

However, the very newest FIFOs generally have architectural provisions for reading in these 'offset values' from the external logic. If the external logic never gets around to imposing its own choice of 'offset values' on the FIFO, the FIFO's 'default values' for these parameters remain in effect during system operation. Again, eight is a very common 'default value.' Another frequent choice is one-eighth of the FIFO depth.

As for block-length counting, to make a long story short, if the sum of the two 'offset values' and the desired block length equals the depth of the FIFO, then the system can use a state change by one of the 'almost' flags as an 'end-of-block' signal. The details are left as an Exercise For The Reader.

A CONTEMPORARY EXAMPLE OF SYNCHRONOUS-FIFO ARCHITECTURE

Consider now Sharp's new 18-bits-wide synchronous FIFOs, LH540215 (512 × 18) and LH540225 (1024 ×18). These new ×18 FIFOs are pin-compatible drop-in replacements for the IDT72215B and IDT72225B ×18 synchronous FIFOs. But they also have some useful enhanced features, which go some ways beyond the architecture of the original IDT parts.

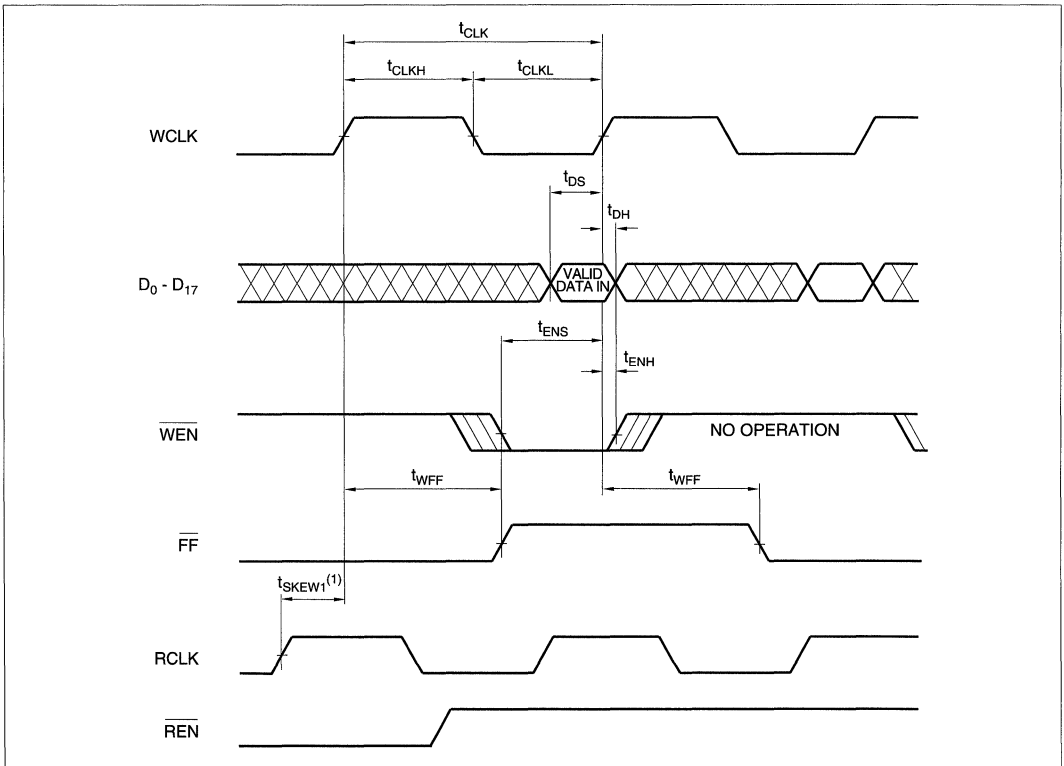
Naturally, they offer programmable offset-value registers for the Almost-Full Flag ($\overline{PAF}$) and the Almost-Empty Flag ($\overline{PAE}$). The default values for the contents of these registers are one-eighth of the FIFO depth, minus one.

Also, the LH540215/25's Full Flag ($\overline{FF}$) and Empty Flag ($\overline{EF}$) signals have been made to behave fully 'synchronously,' as far as the outside world is concerned. In order to avoid metastability problems, these flags have been *resynchronized*, to the clock at that end of the FIFO where the external logic is most likely to be trying to read them.

To review the rationale for this resynchronization, Subsystem A (the writing/input side) normally won't need to read $\overline{EF}$, since writing can occur at any time when the FIFO isn't actually full. Likewise, Subsystem B (the reading/output side) normally won't need to read $\overline{FF}$, since reading can occur at any time when the FIFO isn't actually empty. Therefore, $\overline{FF}$ has been resynchronized to change *only* as the result of a rising edge of WCLK (the writing-side clock); and $\overline{EF}$ has been resynchronized to change *only* as the result of a rising edge of RCLK (the reading-side clock).

This flag-signal behavior is probably optimum, from the viewpoint of a digital-system designer trying to make use of these FIFOs. However, for the resynchronization circuits within the FIFO to have time to function properly, it has proven necessary to specify two new 'skew' timing parameters, t_{SKEW1} and t_{SKEW2}; see Figure 1 and Figure 2.

t_{SKEW1} is the minimum delay for a rising edge of WCLK to occur after a rising edge of RCLK, in order to guarantee that the value of $\overline{FF}$ will be accurate as of *this* rising edge of WCLK, and won't get delayed until the *next* rising edge of WCLK. Similarly, t_{SKEW2} is the minimum delay for a rising edge of RCLK to occur after a rising edge of WCLK, in order to guarantee that the value of $\overline{EF}$ will be accurate as of *this* rising edge of RCLK, and will not get delayed until the *next* rising edge of RCLK.



NOTE:

1. t_{SKEW1} is the minimum time between a rising RCLK edge and a rising WCLK edge for $\overline{FF}$ to change predictably during the current clock cycle. If the time between the rising edge of RCLK and the rising edge of WCLK is less than t_{SKEW1}, then it is not guaranteed that $\overline{FF}$ will change state until the next following WCLK edge.

540215-6

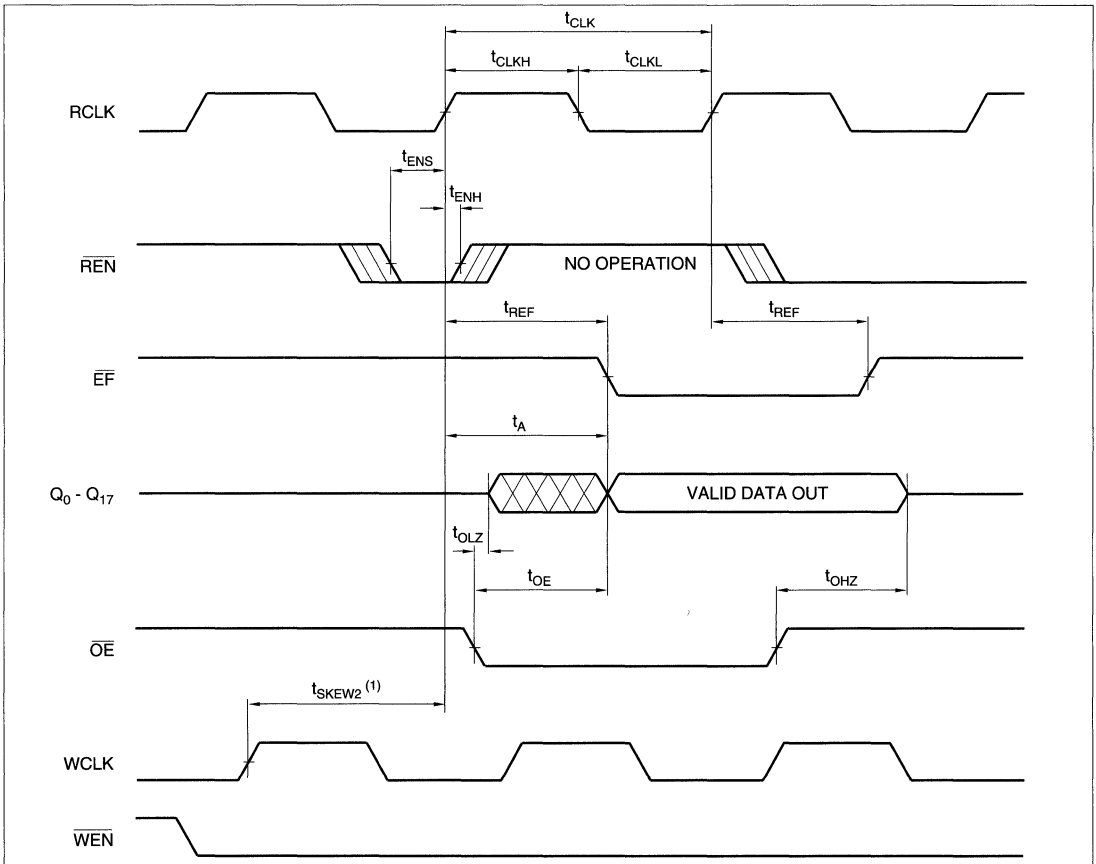
Figure 1. Synchronous Write Operation

Impatient digital-system designers sometimes are annoyed by having to observe the timing restrictions implied by these 'skew parameters.' However, even the fastest digital logic can't respond instantaneously. It appears that their objections really are to restrictions which are *inherent* in what FIFOs must be, if they are to solve the system timing problems which they are expected to solve. The basic flag-synchronization problem can be moved around, and changed in form, in all sorts of different ways; but it can't ever be totally *eliminated.*

Because a FIFO is inherently a pass-through window, between two logic subsystems which are synchronized *differently*, at *some* point there is always going to be a *boundary* between these two subsystems. And, at this boundary, the system designer inevitably is going to have to make allowances for some 'boundary effects.'

THE 'MIDDLE' FLAG SIGNALS

In the IDT ×18 FIFOs, the three 'middle' flags – the 'Programmable Almost-Full Flag' ($\overline{PAF}$), the 'Half-Full Flag' ($\overline{HF}$), and the 'Programmable Almost-Empty Flag' ($\overline{PAE}$) – have been passed straight through to the outside world in the exact same form in which, as we have seen, the FIFO's internal logic inevitably must generate them. That is, these flags are 'synchronized to both clocks,' which is to say that they are 'asynchronous.'



NOTE:

1. t_{SKEW2} is the minimum time between a rising WCLK edge and a rising RCLK edge for $\overline{EF}$ to change predictably during the current clock cycle. If the time between the rising edge of WCLK and the rising edge of RCLK is less than t_{SKEW2}, then it is not guaranteed that $\overline{EF}$ will change state until the next following RCLK edge.

540215-7

Figure 2. Synchronous Read Operation

But the Sharp ×18 replacements for these IDT parts go one step further. They incorporate a programmable 'Command Register,' by means of which the system can specify 'synchronous' behavior for these three 'middle' flags *also*.

$\overline{PAF}$, like $\overline{FF}$, normally gets read at the writing side of the FIFO; and so there is a Command-Register bit which can be used to program $\overline{PAF}$ to be resynchronized to WCLK. $\overline{PAE}$, like $\overline{EF}$, normally gets read at the reading side of the FIFO; and so there is a Command-Register bit which can be used to program $\overline{PAE}$ to be resynchronized to RCLK.

However, a priori, $\overline{HF}$ is equally likely to be needed at *either* end of the FIFO. Therefore, *two* Command-Register bits have been allocated for $\overline{HF}$ resynchronization. The system can select either WCLK or RCLK as the resynchronizing clock for $\overline{HF}$. Or, it can leave $\overline{HF}$ 'synchronized to both ends of the FIFO' – i.e., as 'asynchronous.'

These Sharp ×18 FIFOs incorporate an 'Enhanced MODE' control signal ($\overline{EMODE}$), which allows the system to select between two different sets of Command-Register 'default' conditions. If $\overline{EMODE}$ is held HIGH during a reset operation, the FIFO is initialized to function exactly like the same-depth IDT ×18 FIFO. But, if $\overline{EMODE}$ is held LOW during a reset operation, a selection is made from the Sharp ×18-FIFO menu of enhanced features; see Table 1. Also, the Command Register then becomes accessible for further modification of these selections.

The pin used for the $\overline{EMODE}$ control input is a Vcc pin in the IDT72215B and IDT72225B ×18 FIFOs. So, whenever the Sharp LH540215 and LH540225 ×18 FIFOs are used in pre-existing IDT ×18-FIFO sockets, they automatically get 'strapped' (pin-programmed) to behave exactly like the IDT FIFOs which they are replacing.

SUMMARY

A FIFO's mission is to serve as a 'pass-through window' between two conceptually-independent subsystems, which should not need to be coordinated or synchronized together in any way. To perform this mission well, the FIFO must provide *'rubber-band memory'* between the two subsystems, and also must *decouple* their clocking synchronizations one from another.

In its 'pass-through window' role, the FIFO itself has its input side and its output side 'marching to two different drummers.' The values of the FIFO's status flags, by which the FIFO informs the outside world of its state of relative fullness or emptiness, arise from computations which unavoidably are synchronized first to one subsystem, and then to the other.

These status-flag values must be *resynchronized* to one or the other of the two subsystem clocks. Otherwise, the external logic must contend with the hazards of *metastability*, which can give rise to profound system processing errors if no countermeasures are taken.

Two new Sharp ×18 synchronous FIFOs, the LH540215 (512 × 18) and LH540225 (1024 × 18), offer advanced new functionality for controlling status-flag synchronization. For practical purposes, when reading status-flag values, fully-synchronous status flags can reduce the risk of metastability to negligible levels.

Table 1. Command-Register Format for Sharp LH540215/25 ×18-Bit FIFOs

COMMAND REGISTER BITS	CODE	VALUE AFTER RESET		FLAG AFFECTED, IF ANY	DESCRIPTION	NOTES
		$\overline{EMODE}$ = H	$\overline{EMODE}$ = L			
00	L	L	H	–	Deassertion of $\overline{LD}$ does not reset the programmable-register write pointer and read pointer.	IDT-compatible addressing of programmable registers.
	H				Deassertion of $\overline{LD}$ resets the programmable-register write pointer and read pointer to address Word 0, the Programmable-Almost-Empty-Flag-Offset Register. The change takes effect after a valid write operation or a valid read operation, respectively.	Non-ambiguous addressing of programmable registers.
01	L	L	H	$\overline{PAE}$	Set by ↑RCLK, reset by ↑WCLK.	Asynchronous flag clocking.
	H				Set and reset by ↑RCLK.	Synchronous flag clocking.
03, 02	LL	LL	HH	$\overline{HF}$	Set by ↑WCLK, reset by ↑RCLK.	Asynchronous flag clocking.
	LH				Set and reset by ↑RCLK.	Synchronous flag clocking at output port.
	HL, HH				Set and reset by ↑WCLK.	Synchronous flag clocking at input port.
04	L	L	H	$\overline{PAF}$	Set by ↑WCLK, reset by ↑RCLK.	Asynchronous flag clocking.
	H				Set and reset by ↑WCLK.	Synchronous flag clocking.
05	L	L	H	–	$\overline{OE}$ has no effect on an internal read operation, apart from disabling the outputs.	Allows the read-address pointer to advance even when $Q_0 - Q_{17}$ are not driving the output bus.
	H				$\overline{OE}$ inhibits a read operation whenever the data outputs $Q_0 - Q_{17}$ are in the high-Z state.	Inhibits the read-address pointer from advancing when $Q_0 - Q_{17}$ are not driving the output bus; thus, guards against data loss.
06	L	L	L	–	Reserved.	Future use to control depth cascading and interlocked paralleling.
	H					
07, 08, 09, 10,11	LLLLL	LLLLL	LLLLL	–	Reserved.	Reserved.

NOTES:

1. *When $\overline{EMODE}$ is HIGH, and Command Register bits 00-05 are LOW, the FIFO behaves in a manner functionally equivalent to the IDT72215B/25B FIFO of similar depth and speed grade. Under these conditions, the Command Register is not visible or accessible to the external system which includes the FIFO.*
2. *If $\overline{EMODE}$ is not asserted (is HIGH), Command Register bits 00-05 remain LOW after a reset operation. However, if $\overline{EMODE}$ is asserted (is LOW) during a reset operation, Command Register bits 00-05 are forced HIGH, and remain HIGH until changed. Command Register bits 06-11 are unaffected by $\overline{EMODE}$.*

BOLD ITALIC = Enhanced Operating Mode

THE SHARP LH5402X5 FAMILY: MAINSTREAM, DESIGNER-FRIENDLY, 18-BIT-WIDE SYNCHRONOUS FIFOs *

Chuck Hastings, Applications Consultant

INTRODUCTION

The new Sharp LH5402X5-family 18-bit-wide First-In, First-Out memories (FIFOs) are high-speed, synchronous digital specialty-memory devices, useful for local temporary storage and for data-rate-matching applications in high-performance digital systems. [1]

These FIFOs are implemented in recent-vintage 0.8μ/0.7μ CMOS static-RAM technology, and feature many conveniences and useful options for digital-system designers. One of them can replace two industry-standard 9-bit-wide asynchronous FIFOs of the same depth, and at the same time provide higher-speed operation and more convenient timing characteristics. They also can replace other existing 18-bit-wide FIFOs, in many cases without design changes to the system.

The LH5402X5 family includes four pin-compatible FIFOs, differing in depth: LH540215 (512 × 18, meaning 512 18-bit words); LH540225 (1024 × 18); LH540235 (2048 × 18); and LH540245 (4096 × 18). As of this writing, the first two of these FIFOs are in production; the other two are in design. These FIFOs are available in speed grades up to 50 MHz (20 nsec-cycle-time).

DESIGN ADVANTAGES OF LH5402X5-FAMILY FIFOs

The flexible LH5402X5 control scheme allows selecting a wide variety of useful operating configurations. Some of these implement Sharp-proprietary functionality; the rest either are industry-standard, or else support emulating the behavior of other FIFOs.

These Sharp FIFOs are pin- and functionally-compatible with the existing IDT722X5B industry-standard 18-bit-wide FIFO family. And they are functionally-similar, although not quite fully 100% pin-compatible, with the existing TI SN74ACT7801/11/81 18-bit-wide FIFO family.

Figure 1 is the common block diagram for all Sharp LH5402X5-family FIFOs. Figure 2 is the pinout common to all LH5402X5 family members, for the industry-standard 68-pin PLCC (Plastic Leadless Chip Carrier) package. Table 1 summarizes the meanings of the pin names.

Bold-italic typeface is used in LH5402X5-family-related figures, tables, and text to indicate Sharp-proprietary functionality and/or signals.

Synchronous Operation

In LH5402X5-family FIFOs, all input and output data-transfer processes and most input control signals are fully synchronized, with respect to one or the other of the two port clocks. If their function relates to the input port, it's to the write clock WCLK. If their function relates to the output port, then it's to the read clock RCLK.

Some of the output status signals ('flags') are fully synchronized by design. Others may be programmed, at any time, to behave either synchronously or asynchronously. When flags are synchronized, it's always to the port clock most likely to be used by the system to synchronize reading them.

In addition to being respectively synchronized by the write clock and the read clock, input and output data-transfer steps likewise must be *enabled* by write-enable and read-enable control signals. These control signals are themselves synchronized to the respective port clocks.

* A slightly-modified version of this paper was presented at Northcon/94. It appeared in the *Northcon/94 Conference Record*, Session 1, paper 5.

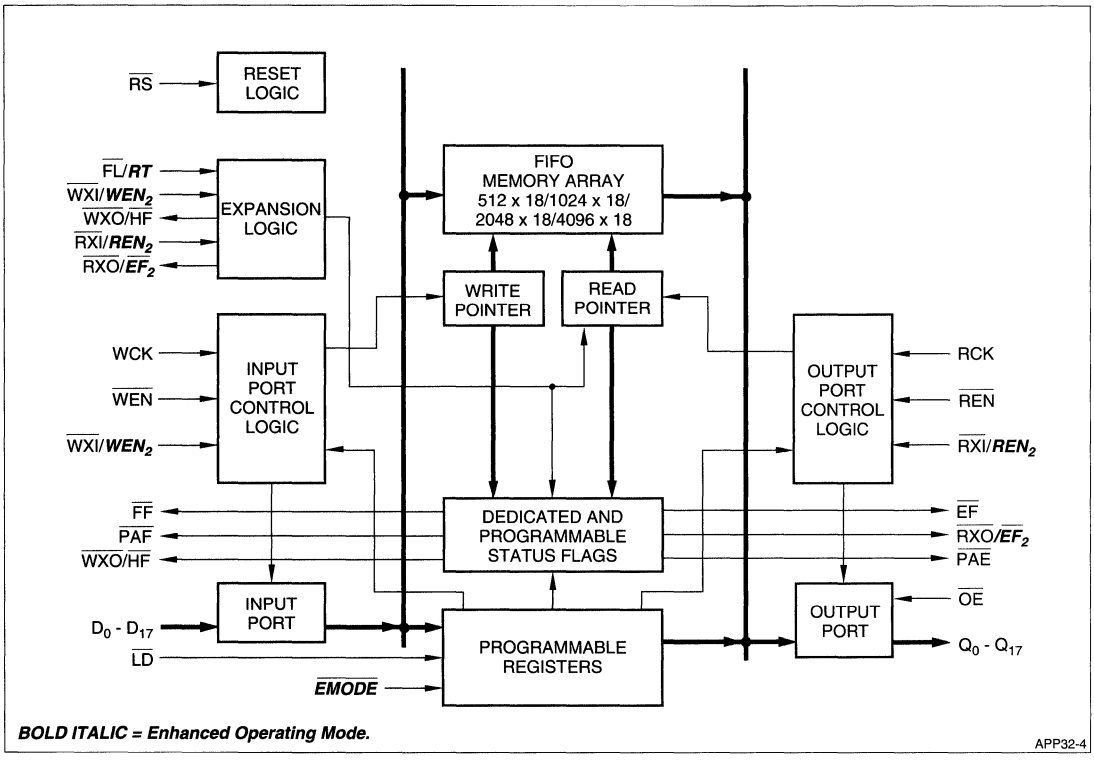


Figure 1. LH5402X5-Family Common Block Diagram

Table 1. LH5402X5-Family Common Pin Designations

PIN	NAME
$D_0 - D_{17}$	Data Inputs
$\overline{RS}$	Reset
$\overline{EMODE}$	***Enhanced Operating Mode***
WCLK	Write Clock
$\overline{WEN}$	Write Enable
RCLK	Read Clock
$\overline{REN}$	Read Enable
$\overline{OE}$	Output Enable
$\overline{LD}$	Load
$\overline{FL}$/RT	First Load/***Retransmit***
$\overline{WXI}$/WEN2	Write Expansion Input/***Write Enable 2***

PIN	NAME
$\overline{RXI}$/REN2	Read Expansion Input/***Read Enable 2***
$\overline{FF}$	Full Flag
$\overline{PAF}$	Programmable Almost-Full Flag
$\overline{WXO}/\overline{HF}$	Write Expansion Output/Half-Full Flag
$\overline{PAE}$	Programmable Almost-Empty Flag
$\overline{EF}$	Empty Flag
$\overline{RXO}$/EF2	Read Expansion Output/***Empty Flag 2***
$Q_0 - Q_{17}$	Data Outputs
V_{CC}	Power
V_{SS}	Ground

BOLD ITALIC = Enhanced Operating Mode

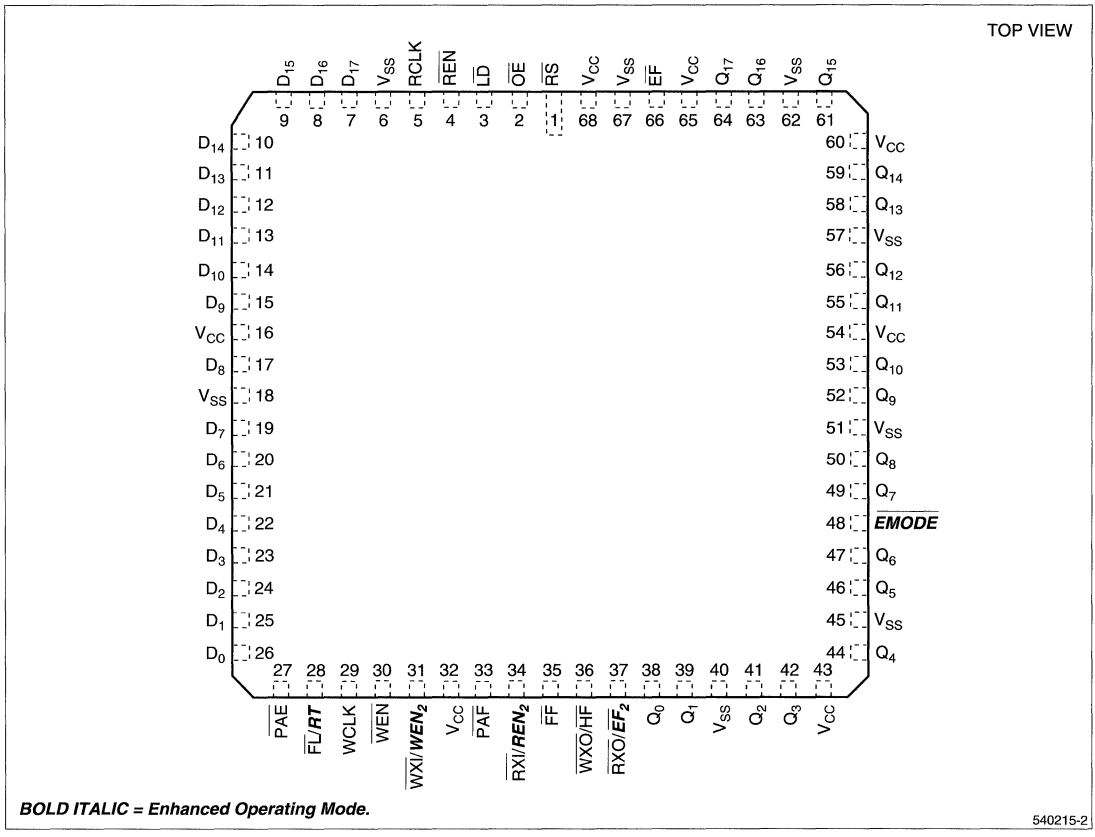


Figure 2. LH5402X5-Family Common PLCC Pinout

High-speed state machines, implemented with registered PLDs (Programmable-Logic Devices), often are used to control the operation of fast synchronous FIFOs such as these LH5402X5-family devices. A PLD which is controlling the input side of a synchronous FIFO may itself be synchronized to the same waveform which is being supplied as that FIFO's write clock. Likewise, a PLD which is controlling the output side of a synchronous FIFO may be synchronized to that FIFO's read clock. Simple, easily-generated, periodic symmetric square waves may be used for both of these FIFO clocks, as well as for PLD clocks.

Such a PLD-FIFO combination may be operated at fast cycle times – as small as the worst-case sum, of a clock-to-output time of one of the two parts added to a setup time of the other part, presumably with some small wiring-delay time factor also added in.

Asynchronous FIFOs aren't nearly as amenable to this same state-machine, symmetric-clock control strategy. Most such parts lack the separate write-enable and read-enable control signals, distinct from the write clock and the read clock. Rather, they have write-demand and read-demand control signals, often loosely referred to as 'clocks,' which *combine* within the same edge-sensitive signal both the clocking and the enabling functionality – in a manner which isn't too convenient for control of the FIFO by the system. Carefully-shaped, gated, *asymmetric* clock waveforms may in fact be needed to drive these demand signals, at least with optimized timing, whenever asynchronous FIFOs are run at data-transfer repetition rates faster than approximately 20-30 MHz.

BOLD ITALIC = Enhanced Operating Mode

Compatibility with Other FIFO Families

The Sharp LH5402X5-family FIFOs have a pinout identical to that of the Integrated Device Technology IDT722X5B-family FIFOs, and can directly replace them in existing applications with no system-design changes whatever. The pinout shown in Figure 2 also is the pinout for IDT722X5B-family FIFOs, with only the following changes: *all* signal names written in bold-italic type are disregarded, and V_{CC} is substituted for ***EMODE*** (pin 48).

However, the Sharp FIFOs have functionality and flexibility which go beyond that of the IDT FIFOs. An LH5402X5 master-control pin, ***$\overline{EMODE}$ (Enhanced Operating Mode)***, selects strict IDT-compatible operation when tied HIGH, but enables additional non-IDT Sharp ***'Enhanced-Operating-Mode'*** features when tied LOW. Since the ***$\overline{EMODE}$*** pin was a V_{CC} pin in the original IDT pinout, it *always* will have been tied HIGH in any design originally based on IDT722X5B FIFOs.

Several of these ***Enhanced-Operating-Mode*** features are directed at emulating the functionality of the Texas Instruments SN74ACT7801/11/81 FIFOs, which have a pinout very similar to – although not fully identical to – the Sharp and IDT FIFOs. In particular, all input-data and output-data pins are in the same locations, and likewise for many power and ground voltage-supply pins. Refer to Figure 3a and Figure 3b for the exact pinout similarities and differences.

Thus, the Sharp FIFOs can replace the IDT FIFOs, with *no* wiring or control-logic changes; and they also can replace the TI FIFOs, with minor wiring and control-logic changes. When operated in an appropriately-chosen mode, the Sharp FIFOs exhibit no major behavioral differences which would affect a *system-level* block diagram, for a system design originally based either on IDT722X5B FIFOs or on TI SN74ACT7801/11/81 FIFOs.

One proprietary Sharp feature, ***Retransmit*** capability, is available regardless of the state of ***$\overline{EMODE}$***. However, the ***Retransmit*** capability never becomes activated when a Sharp FIFO is used in a socket already prewired for an IDT FIFO, which has the ***$\overline{FL}$/RT*** pin grounded.

Configurability

As already mentioned, the Sharp LH5402X5 FIFOs use one IDT722X5B V_{CC} pin for their ***$\overline{EMODE}$*** control signal; otherwise, they strictly follow the IDT722X5B pinout. When ***$\overline{EMODE}$*** is tied HIGH, as it always would be in any design originally based on IDT722X5B FIFOs, IDT-compatible behavior is selected. However, when ***$\overline{EMODE}$*** is tied LOW, Sharp's proprietary ***Enhanced-Operating-Mode*** features are activated.

Switching this ***$\overline{EMODE}$*** signal dynamically during FIFO operation is *not* recommended. ***$\overline{EMODE}$*** should remain in the same state following any FIFO reset operation, during subsequent system operation, until another FIFO reset operation occurs.

Tying ***$\overline{EMODE}$*** LOW in and of itself changes the functionality of three LH5402X5 pins. In IDT-Compatible Operating Mode, these three pins are used to link together successive FIFOs being cascaded according to the IDT 'two-wire-token-passing' principle [2]. Tying ***$\overline{EMODE}$*** LOW also makes the Sharp-proprietary ***Control Register*** visible and accessible to the system, via the normal FIFO input and output data ports.

When ***$\overline{EMODE}$*** is LOW during a reset operation, the ***Control Register*** is initialized to activate *all* of Sharp's proprietary programmable ***Enhanced-Operating-Mode*** features. If any of these features are not desired in a given application, they may be disabled individually by writing different contents into bits 00-05 of the ***Control Register***. They also may be subsequently reenabled, at any time, in the same manner.

The programmable LH5402X5 features include: synchronizing the three 'middle' flags (Almost-Full, Half-Full, Almost-Empty) to the appropriate port clocks; suppressing the advancement of the internal read-address pointer whenever the data outputs are disabled ('data-loss prevention'); and simplifying the procedure for programming, or verification readback, of the FIFO's loadable registers.

When the ***Control Register*** is active, the Almost-Full Flag may be synchronized to the write clock; the Almost-Empty Flag may be synchronized to the read clock; and the Half-Full Flag may be synchronized to either one of these clocks. The Full Flag and the Empty Flag aren't programmable in this manner; they always are synchronized respectively to the write clock and to the read clock. This scheme assigns the synchronization of each flag to that clock most likely to be used for synchronizing the reading of the flag by the system, in order to minimize metastability hazards. [3] The Almost-Empty, Half, and Almost-Full Flags, when chosen to be synchronous, utilize two-stage synchronizers for a high level of metastability protection.

However, it is possible to program any or all of the three 'middle' flags to operate asynchronously, even in the ***Enhanced Operating Mode***. In the IDT-Compatible Operating Mode, they *always* operate asynchronously. In *either* mode, they operate faster than the corresponding flags of the same speed grade of the equivalent IDT device.

The LH5402X5 architecture also includes two other configuration-control registers, the 'offset' registers. These offset registers define the system meaning of the Almost-Full Flag and the Almost-Empty Flag respectively, and are fully equivalent to the similar registers in the IDT722X5B architecture. The number of bits actually implemented in these registers, and their default contents after a reset operation, increases somewhat for the deeper LH5402X5-family members.

BOLD ITALIC = Enhanced Operating Mode

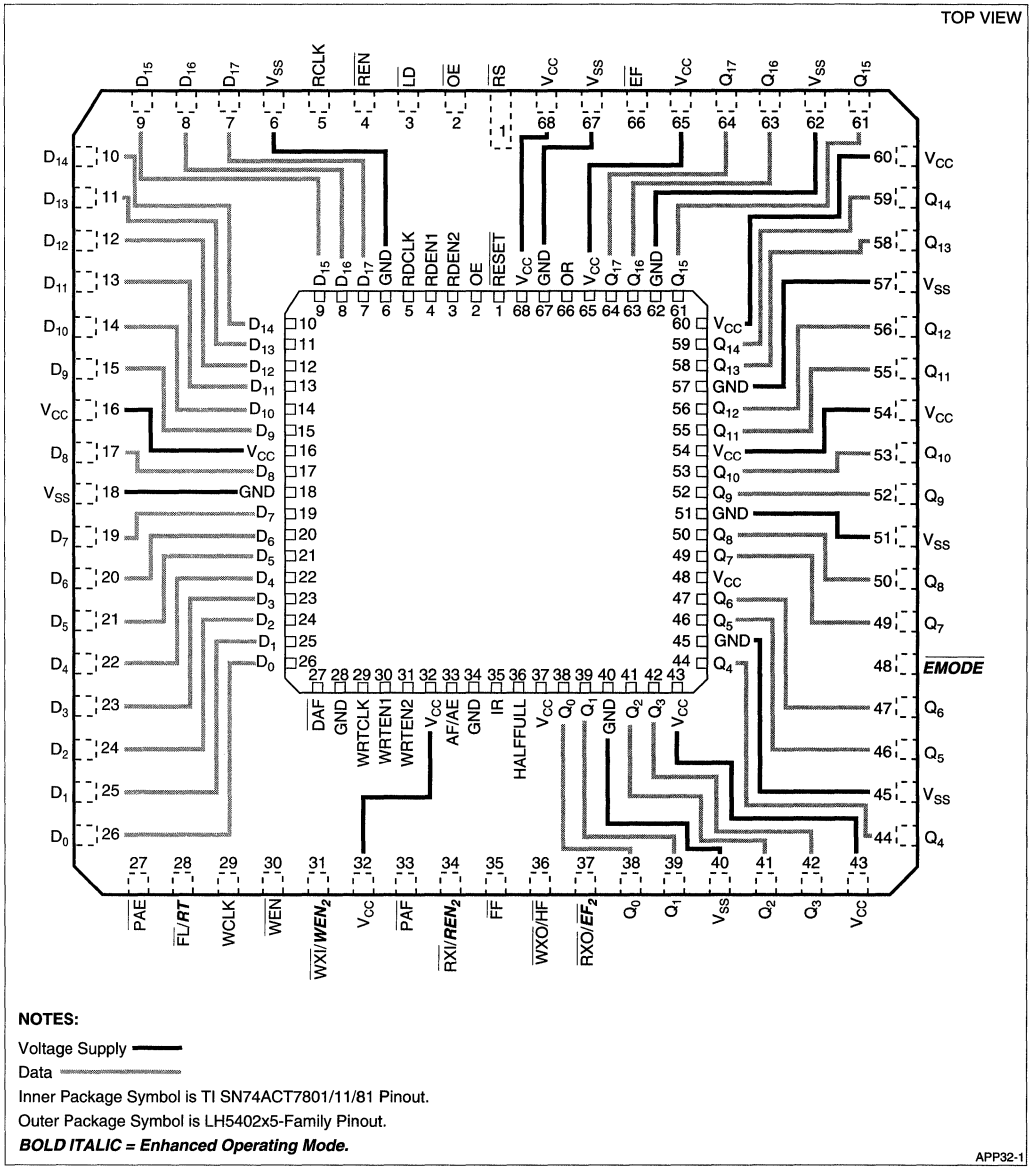
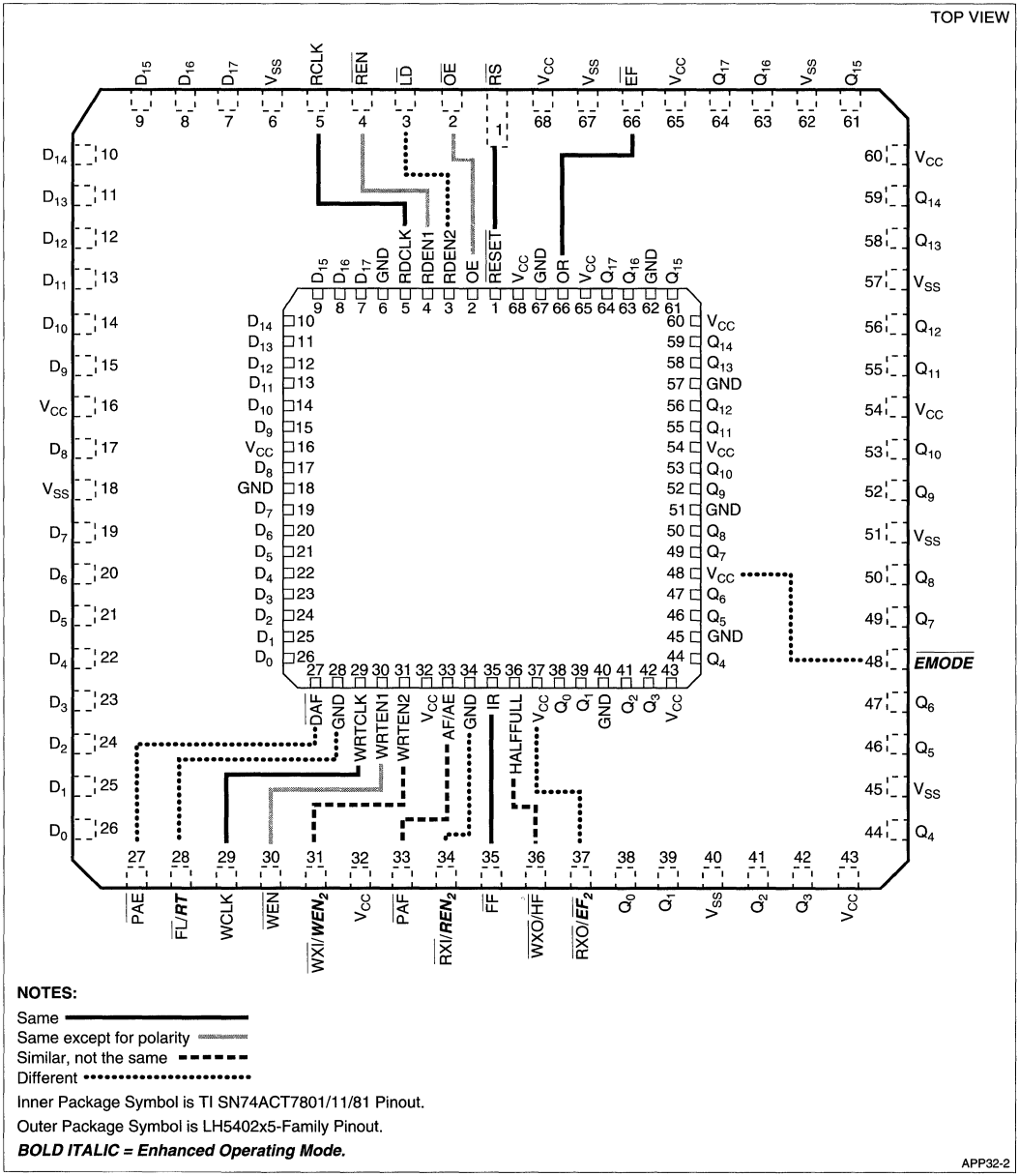


Figure 3a. Common Pin Locations for Voltage-Supply and Data Pins in Sharp and TI 18-Bit-Wide FIFOs



NOTES:

Same ——————

Same except for polarity ——————

Similar, not the same - - - - - -

Different ••••••••••••••••••••••••

Inner Package Symbol is TI SN74ACT7801/11/81 Pinout.

Outer Package Symbol is LH5402x5-Family Pinout.

BOLD ITALIC = Enhanced Operating Mode.

APP32-2

Figure 3b. Control-Pin Similarities and Differences Between Sharp and TI 18-Bit-Wide FIFOs

The prewired default values contained in these offset registers vary somewhat according to FIFO depth. For the smaller LH5402X5-family members, these default values are one-eighth of the FIFO depth, minus one – respectively, 63_{10} for the LH540215, and 127_{10} for the LH540225. For both the LH540235 and LH540245, the default values are again 127_{10}. Whenever the prewired default values are acceptable for an application, no reprogramming is necessary.

The offset registers and the Control Register may be programmed with new contents from the input-data bus, and their contents may be read back out for verification on the output-data bus. Asserting a special control signal, $\overline{LD}$ (LoaD), distinguishes configuration-register accesses from normal FIFO-memory accesses. Two simple state machines, one for each data bus, select the register to be written into or to be read back from, according to a fixed sequence.

In ***Enhanced Operating Mode***, each of these two state machines has *three* states apiece, so that all three registers are accessible and visible. However, in IDT-Compatible Operating Mode, the two state machines have only *two* states apiece, and the Control Register is neither accessible nor visible.

CASCADING AND PARALLELING LH5402X5-FAMILY FIFOs

For FIFOs, as for other semiconductor memories, there's always a memory-capacity gap – between the largest FIFOs currently in production, and the largest FIFOs which customers need for the new state-of-the-art digital systems which they are attempting to design. Over time, *both* the immediately-available FIFOs and the needed-but-not-yet-available FIFOs get larger and larger; but the gap between them always remains!

An approach which semiconductor manufacturers have taken, trying to address customers' perceived needs for ever-larger FIFOs, is to provide support within FIFO architectures for combining individual FIFO devices into larger 'effective FIFOs.' This support takes the form of extra control inputs and status outputs ('hooks') for control connections between FIFOs, and sometimes also of other additional on-chip resources.

There are two fundamental ways in which FIFO devices may thus be combined: *cascading*, or 'depth-cascading,' in which the effective FIFO has *more* memory words than do each of the individual FIFO devices; and *paralleling*, sometimes rather confusingly called 'width-cascading,' in which the effective FIFO has *wider* memory words than do each of the individual FIFO devices. Sometimes, both cascading and paralleling are used within a single FIFO application, to implement an effective FIFO which is *both* deeper and wider than one individual FIFO device.

LH5402X5-family FIFOs feature architectural support for *two* entirely different FIFO-cascading schemes: 'two-wire token passing,' which is the scheme used by IDT722X5B-family FIFOs; and 'handclasp' or 'pipelining,' which is the scheme used by TI SN74ACT7801/11/81 FIFOs. [2] They also support one 'interlocked' paralleling scheme, in which two FIFOs side-by-side operate together in strictly-synchronized lockstep. [2]

In principle, the two-wire token-passing cascading scheme also 'parallels' FIFOs, but in a fundamentally-different manner – as *alternates*, with a common input bus and a common output bus, arranged like a bunch of bananas. (See Figure 4a.) The pipelining cascading scheme, on the other hand, arranges FIFOs one after another in series, connected like a string of sausages. (See Figure 4b.)

Cascading by Two-Wire Token Passing

In the two-wire token-passing scheme, two or more individual LH5402X5 FIFO devices are connected in parallel. (See Figure 5.) Their input data buses (pins D_0-D_{17}) are all tied together; and their output data buses (pins Q_0-Q_{17}) also are all tied together. However, only one of the paralleled FIFOs may be written into at any given time. Likewise, only one of the paralleled FIFOs may be read from at any given time. [2] A data word passing through the effective FIFO passes through only one single physical FIFO device, and gets handled just once.

The FIFO which is currently activated for writing is not necessarily the same one which is currently activated for reading, except under system-startup conditions – that is, when all FIFOs have just completed being reset, and FIFO operation has only recently commenced.

BOLD ITALIC = Enhanced Operating Mode

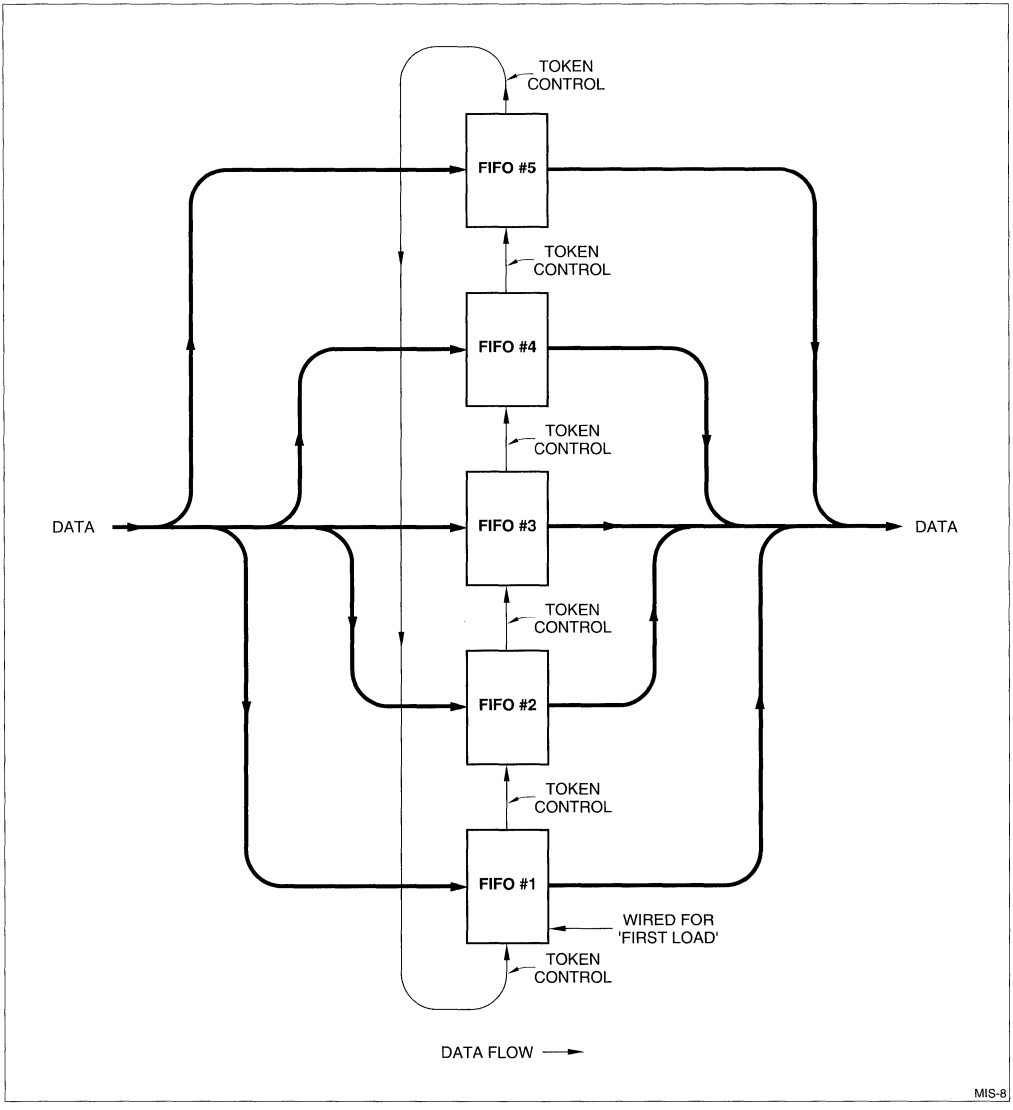


Figure 4a. Parallel Connection of Depth-Cascaded FIFOs

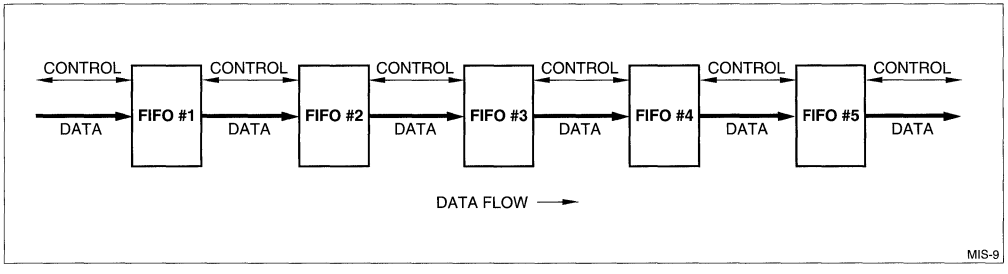
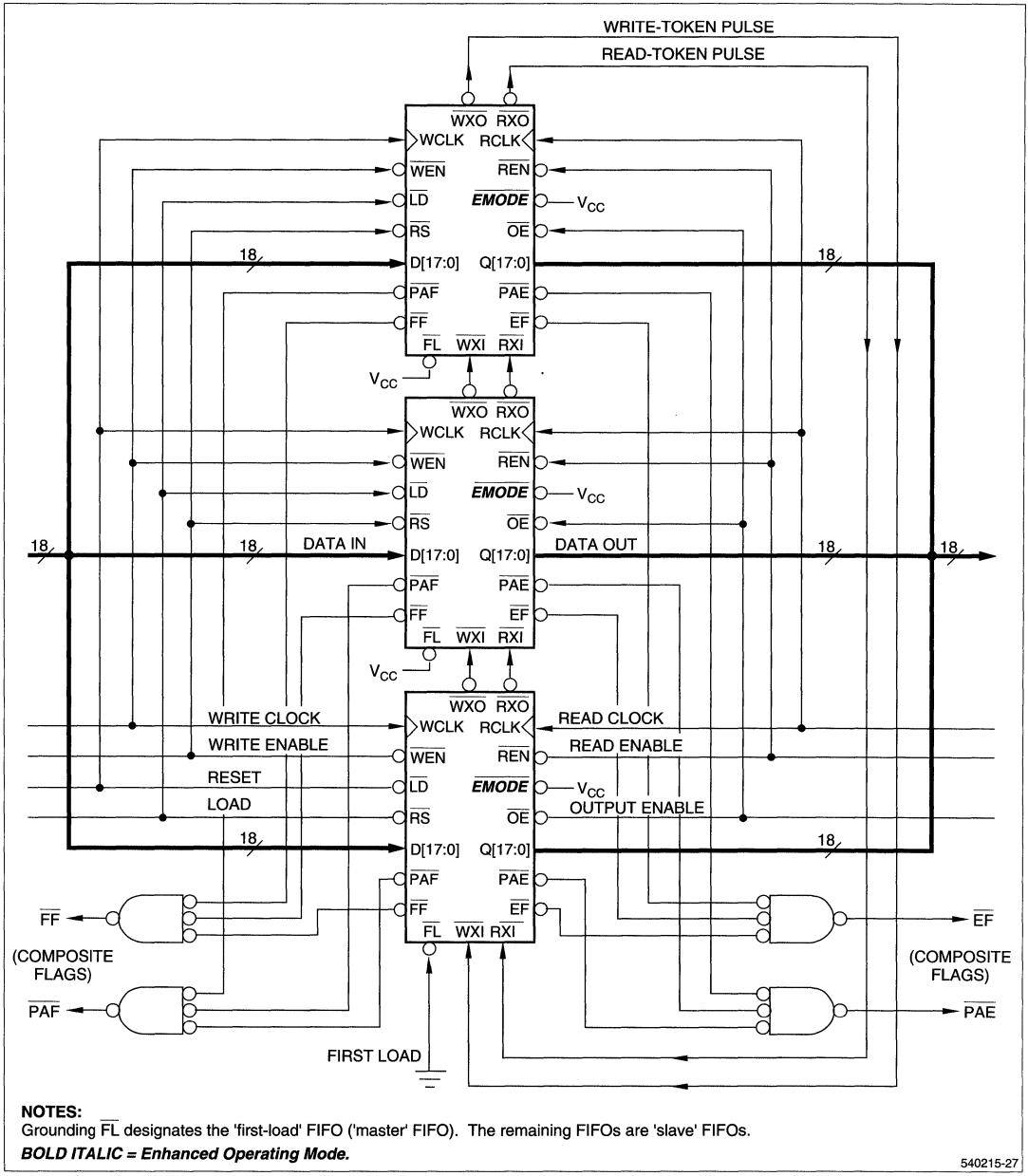


Figure 4b. Series Connection of Depth-Cascaded FIFOs



NOTES:
Grounding $\overline{FL}$ designates the 'first-load' FIFO ('master' FIFO). The remaining FIFOs are 'slave' FIFOs.
BOLD ITALIC = Enhanced Operating Mode.

540215-27

Figure 5. LH5402X5-Family FIFO Depth Cascading Using IDT-Compatible 'Token-Passing' Scheme

LH5402X5 FIFOs must have their ***EMODE*** control pins held at V_{CC}, whenever they are being cascaded according to the two-wire token-passing scheme.

Exactly one of the LH5402X5 FIFOs in the cascade must have its $\overline{FL}$/***RT*** pin grounded; the $\overline{FL}$/***RT*** pins of the remaining FIFOs are tied to V_{CC}. In this operating mode, the $\overline{FL}$/***RT*** pin is being used to designate one of the FIFOs as the 'first-load' or 'master' device; the remaining FIFOs are sometimes referred to as 'slave' devices. After a reset operation has initialized the entire FIFO cascade, each FIFO has figured out whether it is serving as a master device, or as a slave device; and both the 'write token' and the 'read token' reside within the first-load FIFO. [4] (See in particular Tables 1 and 2 of [4].)

Within a FIFO, these tokens are *implicit*. When a FIFO 'has one of the tokens,' that simply means that it is the FIFO device which is activated to respond and perform the requested operation, whenever a write-enable or read-enable control signal is broadcast to all the FIFOs within the cascade. Obviously, one – and *only* one – FIFO device should respond, for each requested operation. One consequence is that any FIFO which 'does not have the read token' has its outputs *disabled*, in a high-impedance state, *even if* its $\overline{OE}$ (Output Enable) control signal currently is being asserted (LOW).

However, when the 'token' (associated with the counter within the FIFO) has traversed all of the FIFO's memory locations, it does *not* return to physical location zero of that FIFO device, even though the counter does thus return. Rather, the token briefly becomes explicit; it takes the form of a narrow LOW-going pulse on one of the two Expansion Out outputs, $\overline{WXO}$ and $\overline{RXO}$. These are connected respectively to the Expansion In inputs, $\overline{WXI}$ and $\overline{RXI}$, of the next FIFO device in the cascade. After the (write or read) token pulse has been transmitted and received, the sending FIFO becomes deactivated for writing or reading as the case may be, and the receiving FIFO immediately becomes activated in its place, so that there never is any hiatus in the effective FIFOs' capability of performing a required operation.

The Half-Full Flag status output, $\overline{HF}$, shares a pin with the $\overline{WXO}$ token output. Hence, $\overline{HF}$ is not available from a cascaded LH5402X5-family FIFO, which has been initialized by a reset operation to be either a master or a slave. $\overline{HF}$ is available only from a FIFO which has been initialized to be a 'standalone.' [4] In any case, this signal does not have any really useful meaning for FIFOs within a cascade.

The two-wire token-passing cascading scheme was developed to eliminate one difficulty with the much-older one-wire token-passing cascading scheme – an occasional, intermittent, serious extra logic delay, in the event that the read pointer catches up with the write pointer in a FIFO, just as that FIFO is about to emit a write or a read Expansion Out token pulse. Of course, that combination of circumstances is infrequent; but it is by no means impossible! In cascaded FIFOs which use the one-wire scheme, if this delay is not explicitly allowed for in the system timing, then data may be lost unpredictably.

Now, in the one-wire token-passing architecture, *both* the write token pulse and the read token pulse must travel along to the next FIFO over the very same wire. Thus, every so often, *both* pulses must pass over that wire within one single word time, with enough of a gap in between them that they are recognized clearly at the other end as *two* distinct pulses. So, once in a great while, and at unforeseeable intervals, the system must halt briefly between two words, in order to head off the occurrence of an unrecoverable system error – or else the data rate must be slowed down greatly, in order to allow the time for this occasional extra pulse during *every* word time.

Because the two-wire token-passing cascading scheme avoids this troublesome flaw of the older one-wire scheme, it is much more appropriate for state-of-the-art high-performance synchronous FIFOs.

Cascading by Pipelining

In the pipelined cascading scheme, two or more individual LH5402X5 FIFO devices are connected in series; the output data bus (pins Q_0-Q_{17}) of the (n-1)st FIFO connects to the input data bus (pins D_0-D_{17}) of the (n)th FIFO.

A crosscoupled 'handclasp' signaling scheme is used between successive FIFOs in the cascade, so that they coordinate their operations with each other properly. (See Figure 6.) The handclasp ensures that the upstream FIFO device does not attempt to force another data word upon the downstream FIFO device when the latter is full – and, conversely, that the downstream device does not attempt to read another word from the upstream device when the latter is empty.

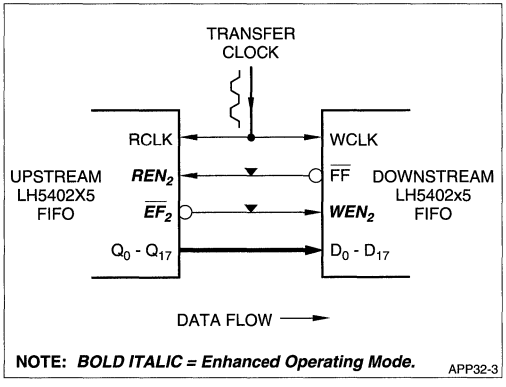


Figure 6. 'Handclasp' Depth-Cascading Interface Between LH5402X5-Family FIFOs

BOLD ITALIC = Enhanced Operating Mode

In this scheme, all of the LH5402X5 FIFOs must have their ***EMODE*** and $\overline{OE}$ control inputs grounded. A data word passing through the effective FIFO passes through every physical FIFO device in the cascade, and gets handled once per device.

A 'transfer clock' signal must be provided, at each FIFO-to-FIFO interface, to synchronize passing data from the (n-1)st FIFO to the (n)th FIFO. To optimize smoothness of operation, the same transfer-clock waveform may be used everywhere within a given FIFO cascade. Now, the write clock at the input end of the first FIFO in the cascade, and the read clock at the output end of the last FIFO in the cascade, may or may not be synchronous periodic waveforms. If both of these clocks are periodic waveforms, the faster one is an excellent choice for use also as the transfer-clock signal.

Grounding the ***EMODE*** control input of an LH5402X5-family FIFO changes the functionality of several dual-purpose control pins: $\overline{WXI}$/***WEN$_2$***, $\overline{RXI}$/***REN$_2$***, and $\overline{RXO}$/***$\overline{EF}_2$***. The Enhanced-Operating-Mode functionality is in each case that which is indicated by the second signal name, the one which is written in bold-italic type.

WEN$_2$ and ***REN$_2$*** are alternative, assertive-HIGH write-enable and read-enable control inputs respectively. They have the correct polarity, for the handclasp signaling scheme to operate properly with the assertive-LOW Full Flag and Empty Flag status signals.

$\overline{EF}_2$ is an exact duplicate of the usual Empty Flag $\overline{EF}$, except that it is delayed by one full read-clock interval with respect to $\overline{EF}$. This extra-clock-interval delay is necessary, in order for the handclasp signaling scheme to operate with the proper timing. [5]

For an LH5402X5 FIFO-to-FIFO interface connected as shown in Figure 6, whenever the upstream FIFO device is not empty and the downstream FIFO device is not full, a data word is transferred from the upstream device to the downstream device after *every* transfer-clock rising edge.

If some interval of time passes without the entire FIFO cascade either being written into or being read from, the meaningful data within the cascade will accumulate in the FIFOs furthest downstream – 'fall to the bottom of the hopper,' as it were. In fact, 'hopper' is one of the many technojargon synonyms for 'FIFO.'

Paralleling

The LH5402X5-family architecture supports 'interlocked' paralleling of two identical side-by-side 18-bit-wide physical FIFO devices, so that they function as a 36-bit-wide effective FIFO of the same depth. (See Figure 7.) The LH5402X5 interlocking scheme is the same one used with TI SN74ACT7801/11/81 FIFOs; however, the pinout and the signal nomenclature are somewhat different.

The purpose of 'interlocking' is to ensure that the two side-by-side FIFO devices never get out of step with each other: one cannot write a word unless the other one also is ready to write a word, and similarly for reading a word. Without interlocking, very minor timing differences between the two physical devices, or circuit-board wire-length differences affecting them, occasionally might result in one FIFO having half of some word present on its outputs, while the other FIFO has the other half of the *previous* word in the data block present on its outputs!

Such a mistake may occur, for instance, when the two FIFO devices have been entirely emptied out, and another full 36-bit word is written in, at such a time that t_{FRL} (First-Read-Latency time) isn't quite met – so that one device responds properly with the next word, but the other device doesn't quite make it. A similar and logically-dual possibility exists when the two FIFO devices have been entirely filled, and one word is read out at exactly the worst-case time – and one of the two devices takes slightly longer, to become ready to receive another word, than does the other one.

This type of problem may arise either with synchronous FIFOs or with asynchronous FIFOs, although the latter are much more vulnerable to it. Thus, the interlocking scheme of Figure 7 is recommended whenever LH5402X5-family FIFOs are used two abreast, to handle 36-bit data words. For wider data words, variations on this same scheme are possible; but they require external logic in addition to the FIFO devices themselves, since now more than two write enables and two read enables per physical device become necessary.

The architecture of LH5402X5-family FIFOs supports combining interlocked paralleling with pipelined cascading, which is the recommended approach to implementing large effective FIFOs using arrays of LH5402X5 devices. [4] (See Figure 8.) Unfortunately, these devices can't offer any similar support for combining interlocked paralleling with two-wire-token-passing cascading; the additional pins which would be required aren't available.

BOLD ITALIC = Enhanced Operating Mode

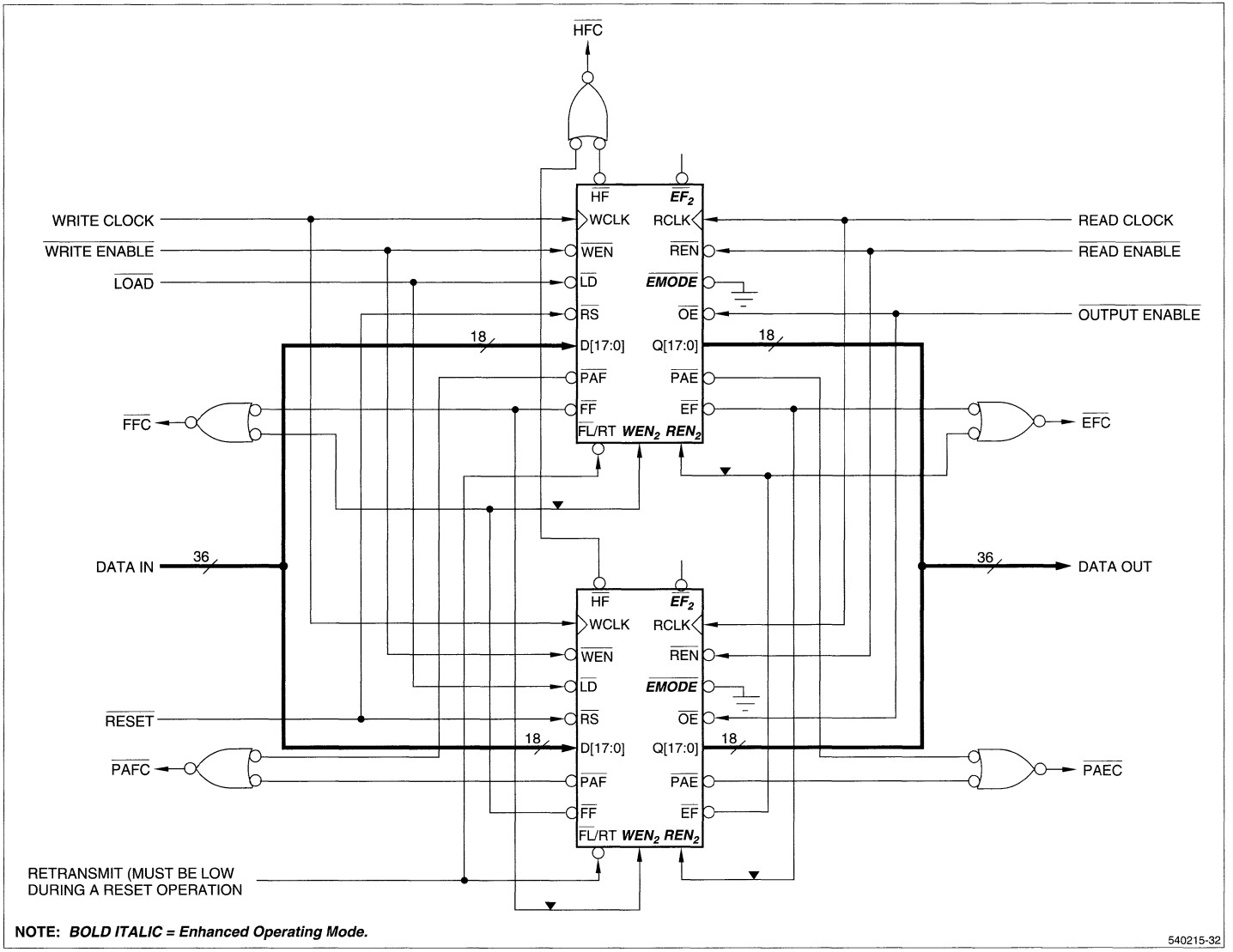


Figure 7. Interlocked-Paralleled 36-Bit-Wide Configuration for Two LH5402X5-Family FIFOs

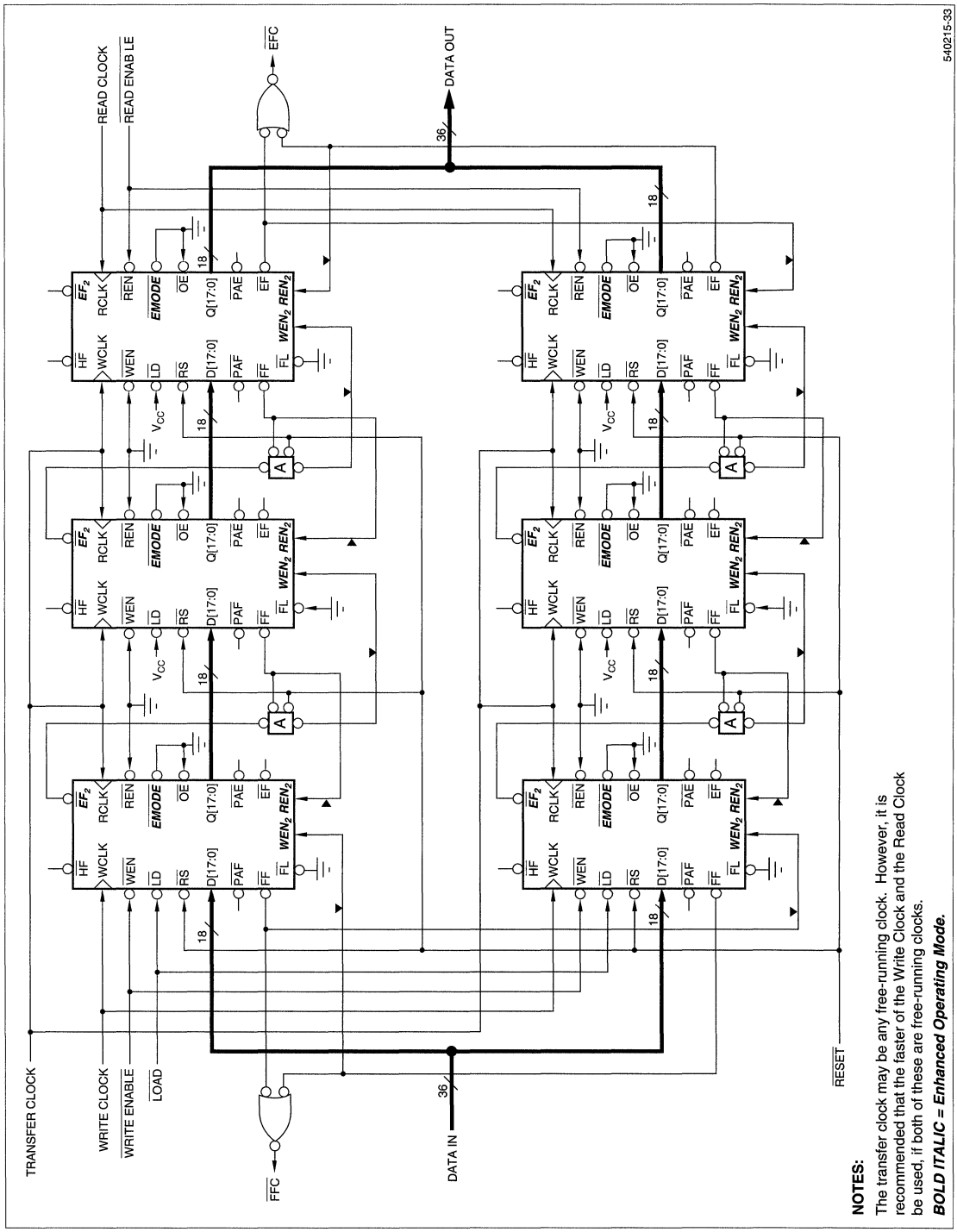


Figure 8. Interlocked-Paralleled and Pipelined-Depth-Cascaded 2 × 3 Array of LH5402X5-Family FIFOs

SUMMARY

The new Sharp LH5402X5-family 18-bit-wide FIFOs provide attractive alternatives for local-temporary-storage and data-rate-matching applications in contemporary high-speed digital systems. Members of this family range in capacity from 512 to 4096 18-bit words, with programmable choices between synchronous and asynchronous operation for three status-flag signals.

LH5402X5-family FIFOs are eminently suitable for use in new high-performance designs. A single such FIFO can replace two paralleled conventional 9-bit-wide asynchronous FIFOs, in many applications. Also, these FIFOs can seamlessly replace IDT722X5B-family FIFOs in existing 18-bit-wide designs, without any rewiring. Moreover, they have certain architectural features which support emulating TI SN74ACT7801/11/81 FIFOs, although their compatibility with these TI FIFOs is not quite 100%.

LH5402X5 FIFOs may be used individually, in 'standalone' mode. Also, they have many architectural features which support 'cascading' – putting them together to create deeper 'effective FIFOs.' Still other features support 'interlocked paralleling,' to create wider 'effective FIFOs.'

Check out these high-performance, full-featured, designer-friendly FIFOs for your next digital-system design. You'll be glad you did!

REFERENCES

[1] Chuck Hastings, *FIFO Memories: Effective, Compact, and Easy to Use*; Northcon/92 Conference Record, Paper M2/3. Available from The Institute of Electrical and Electronics Engineers, Inc., 445 Hoes Lane, P. O. Box 1331, Piscataway, NJ 08855-1331. Also reprinted in the *1993 Sharp Application Notes Book*, pages 1B-16 to 1B-18.

[2] Chuck Hastings, *Different Strokes: How to Cascade and Parallel FIFOs of Various Architectures*; Northcon/93 Conference Record, Paper 9/1. Available from The Institute of Electrical and Electronics Engineers, Inc., 445 Hoes Lane, P. O. Box 1331, Piscataway, NJ 08855-1331. Also reprinted in the *1993/1994 Sharp Memory Data Book*, pages 6-16 to 6-29.

[3] Chuck Hastings, *FIFO Flag Timing: Marching to Two Different Drummers*; Northcon/93 Conference Record, Paper 12/3; also Wescon/93 Conference Record, Paper 26/3. Available from The Institute of Electrical and Electronics Engineers, Inc., 445 Hoes Lane, P. O. Box 1331, Piscataway, NJ 08855-1331. Also reprinted in the *1993/1994 Sharp Memory Data Book*, pages 6-8 to 6-15.

[4] Preliminary Datasheet, *Sharp LH540215/25 512 × 18/1024 × 18 Synchronous FIFOs*, *1993/1994 Sharp Memory Data Book*, pages 5-193 to 5-230. (NOTE: The 'Preliminary' status no longer applies. Since this databook was published, these parts have reached full production status, and a revised and expanded data sheet has been prepared.)

[5] The first-production-version LH540215 and LH540225 FIFOs are prone to one anomalous behavior ***only*** during pipelined depth cascading: In the special case where an upstream FIFO becomes totally empty at the same time as a downstream FIFO becomes totally full (so that the data-transfer process momentarily halts), one extra copy of the last word transferred out of the upstream FIFO is passed downstream at the time that the data-transfer process resumes. This anomalous behavior is being corrected in the LH540235 and LH540245 FIFOs. A simple external-logic fix using programmable-logic devices (that preserves full 50 MHz operation) is described in the new revised version of the LH540215/25 datasheet. [4]

BOLD ITALIC = Enhanced Operating Mode

A ONE-CHIP TWO-WAY STREET FOR MICROPROCESSOR COMMUNICATIONS: THE SHARP LH5420 36-BIT BIDIRECTIONAL FIFO

Chuck Hastings
Marketing/Applications Manager, FIFO and Specialty Memories

Sharp Microelectronics Technology, Inc.
5700 N. W. Pacific Rim Boulevard
Camas, WA 98607
206/834-8615

INTRODUCTION

New integrated circuits often evolve as single-chip embodiments of groups of lower-complexity parts. When the same multiple-device configuration starts turning up in many new designs, a semiconductor manufacturer may get inspired to develop a one-chip-does-all replacement just by listening to its customers. Bidirectional FIFOs, wide enough to hold an entire word of data, are one such frequently-occurring combination. Perhaps one out of every five system applications for FIFOs fits this description. Usually, the role of a bidirectional FIFO is to provide convenient two-way communication between two processors or microprocessors.

In the past, an effective bidirectional FIFO for communication back and forth between two 32-bit-processors has needed to consist of at least *eight* industry-standard byte-wide unidirectional FIFO devices, arranged into two 'back-to-back' ranks of four paralleled FIFOs each. When parity checking is implemented, the data path between processors becomes 36-bit. Sometimes only one of the two processors is 32-bit, and the other one is 16-bit or 8-bit. In this event, even more devices must be added, to implement multiplexing, demultiplexing, and control functions at the narrower end of the bidirectional data path.

The LH5420 256 $\times$ 36 $\times$ 2 bidirectional FIFO, now available from Sharp, is a 'one-chip-does-all' solution to such system requirements for two-way interprocessor communication. One LH5420 can provide either a convenient fully-parallel two-way connection from one 36-bit bus to another such bus, or it can provide a two-way 'funneling/defunneling' connection from a 36-bit bus to an 18-bit bus, or to a 9-bit bus. Thus, the LH5420 supports all of the usual microprocessor word widths, and accommodates the extra bit per byte for parity or marker-bit usage. It operates at up to 40 MHz, and is available either in a 120-pin PGA package or in a 132-pin PQFP package.

LH5420 ARCHITECTURE AND OPERATION

The LH5420 includes several enhancements, aimed at making a system designer's life easier. The LH5420 itself can check the parity of all bytes passing through it in either direction. And it features programmable almost-full and almost-empty flags, retransmission capability in either direction, 'mailbox' capability in either direction, a limited form of transceiver-mode oper-ation, and a synchronous request/acknowledge capability which is useful in burst-mode communications.

Conceptually, an LH5420 is organized as two 36-bit-wide bidirectional ports, Port A and Port B. Two full-width 256-word FIFOs, FIFO # 1 and FIFO # 2, are connected between the two ports, one transmitting in each direction. (See Figure 1.) There are also two full-width one-word mailboxes between the two ports, one likewise transmitting in each direction. And there is a full-width bidirectional data bypass path, which functions during a reset operation. Two asynchronous control inputs set the data width of Port B at 36 bits, at 18 bits, or at 9 bits.

Each port has its own clock input. In typical applications, a port's clock input is connected to a periodic free-running clock signal, which may or may not be derived from the same frequency source as the other port's clock input. Each port also has three control inputs which are sampled at the rising edge (LOW-to-HIGH transition) of its clock: read/write, enable, and request. Each port also has an 'Acknowledge' output which is synchronized to its clock, a parity flag output, and asynchronous control inputs for initiating data re-transmission and for enabling/disabling its data outputs.

FIFO # 1 and FIFO # 2 each have five status flags to indicate relative fullness: Full, Almost-Full, Half-Full, Almost-Empty, and Empty. The Full, Half-Full, and Empty flags are hard-wired to signal exactly what their names

indicate. But there are programmable 'offsets' controlling the operation of the Almost-Full and Almost-Empty flags, to numerically define the boundaries of the 'Almost-Full' region and the 'Almost-Empty' region. These offset values are both initialized to eight during a reset operation; but either one may be changed under system control, independently of the other one, to any value from zero to 255.

During a data transfer, the port's Acknowledge output repeats the same information as either the Almost-Full flag or the Almost-Empty flag, depending on the current direction of data transfer – Almost-Full when writing, and Almost-Empty when reading.

The five relative-fullness status flags may change state either in response to a write event clocked at one port, or else in response to a read event clocked at the other port. The port's Acknowledge output signal, however, is totally synchronous with the clock input signal at that port; except, that it gets deasserted immediately if at any time the Request input signal is deasserted.

Both the Request control input and the Enable control input of a port must be asserted, in order for that port to carry out a read operation or a write operation. The Read/Write control input determines which type of operation gets performed.

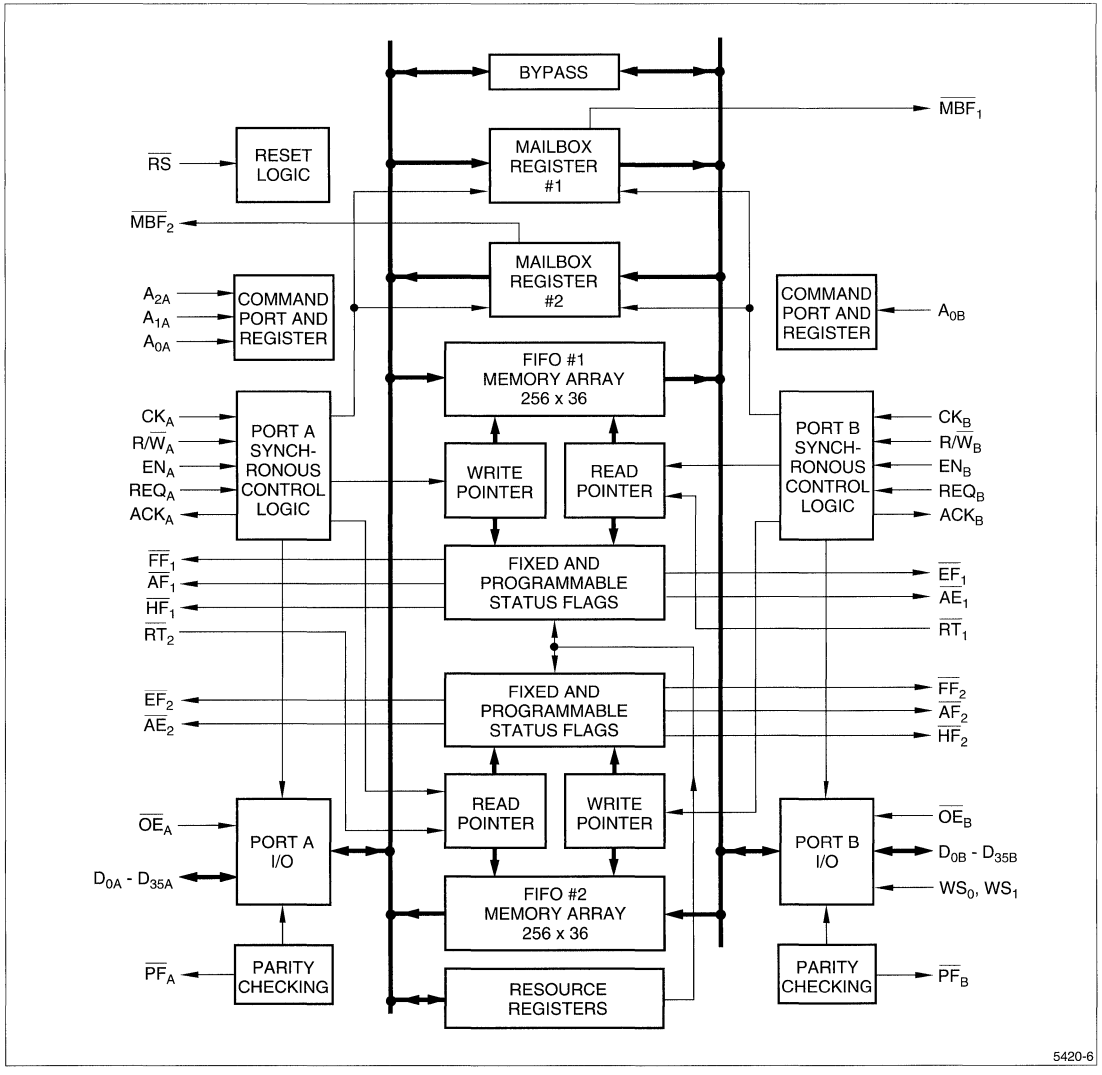


Figure 1. LH5420 Block Diagram

The action of the Request and Enable signals within the LH5420 are generally similar; but their detailed timing is different. The Enable signal is presumed to be originating as a synchronous signal referenced to the same clock signal used by the port. On the other hand, the Request signal may arise asynchronously, elsewhere in the system; the LH5420 contains resynchronizing circuits, which reference the Request signal to the port clock internally within the LH5420.

Either port may place a full 36-bit word in the other port's mailbox register. Doing so sets a mailbox flag, which is synchronized to the receiving port's clock. This flag is reset whenever the receiving port has read the word in the mailbox register. Both ports have the ability to select either their outgoing FIFO or their outgoing mailbox for writing, or either their incoming FIFO or their incoming mailbox for reading.

Although Port A and Port B both have the capability to send and receive 36-bit data words, each port has one major function unique to it. Port A is the master port for purposes of resource-allocation and control functions, such as changing the value of the offsets for the Almost-Full and Almost-Empty flags, or changing the byte parity scheme from odd parity to even parity. Port B, on the other hand, is the port which is capable of setting its effective data width at 36 bits, 18 bits, or 9 bits.

Two asynchronous inputs control the data width of Port B. Changing this data width does not require any reset operation. However, sufficient time must be allowed for the LH5420's internal byte-shifting and demultiplexing circuits to settle; waiting for two full Port B clock cycles is recommended.

'SYNCHRONOUS' FIFOs AND 'ASYNCHRONOUS' FIFOs

The antonyms 'synchronous' and 'asynchronous' each have taken on two very different meanings in FIFO applications literature. The first meaning has to do with the timing of the FIFO's data and control inputs, and of its data and status outputs. The second meaning has to do with the capability of the FIFO to adjust itself to different and unrelated timing requirements at each of its two ends.

According to the first meaning of these terms, a 'synchronous' FIFO operates with a free-running clock input, but performs operations such as writing or reading only when these operations are 'enabled.' Data inputs, and control inputs such as enable signals and mode-control signals, must all meet setup time and hold time requirements with respect to the free-running clock. Data outputs and status outputs are presumed valid after some specified delay time has elapsed, following a transition of the free-running clock.

FIFOs which are 'asynchronous,' according to this meaning of 'asynchronous,' do not use any such free-running clock. Some older-architecture 'asynchro-nous' FIFOs even use edge-sensitive, rather than level-sensitive, control inputs. 'Synchronous' FIFOs sometimes may be made to behave as 'asynchronous' FIFOs, if desired, by connecting their 'enable' inputs to be permanently asserted, and using their free-running clock inputs as asynchronous edge-sensitive 'demand' control input signals.

According to the second meaning of the terms 'synchronous' and 'asynchronous,' however, a 'syn-chronous' FIFO would be a FIFO having both its input port and its output port always synchronized to the same 'clock' signal; in other words, a glorified shift register. An 'asynchronous' FIFO, on the other hand, can operate with its input port synchronized to one timing signal, and its output port synchronized to a second timing signal having no necessary relation to the first one; and neither timing signal needs to be regular or periodic.

The LH5420 has a free-running-clock-plus-enable control structure; and so its two internal FIFOs are 'synchronous' FIFOs in the first sense of this term, except that the behavior of the five relative-fullness flags is not entirely 'synchronous.' However, they are completely 'asynchronous' FIFOs in the second sense; there is no necessary synchronization relation between the Port A clock and the Port B clock, nor is either of these clocks required to be strictly periodic. This type of behavior is usually considered to be useful, system-friendly, and what FIFOs are all about.

DESIGNING WITH THE LH5420

In some applications, data bursts get pushed through a FIFO at or close to the FIFO's maximum word rate; but the system must take some immediate action if the FIFO ever becomes completely full or completely empty. The LH5420's Request/Acknowledge feature supports such a mode of operation. The Acknowledge output signal meets the setup time and hold time requirements for the Enable input, and may simply be tied back to it, in order to prevent complete filling or complete emptying of the active FIFO. This mode of operation slightly decreases the maximum data rate.

In essence, the Acknowledge signal is a synchronous 'proxy' or 'predictor' for whichever 'Almost' flag is pertinent to the current data-transfer operation. Because synchronous predictive logic is used to determine the state of this signal, it is actually faster than the corresponding flag.

Assume now that a port's Request input is being continuously asserted, say for writing into the outbound FIFO for that port. As long as the FIFO does not get into the 'Almost-Full' region, that is, the number of vacant FIFO physical words never falls below the 'Almost-Full' offset value, then the Acknowledge output is continuously asserted by the LH5420 control logic, and a word gets written into the FIFO as a result of every write-clock pulse. However, if the FIFO does become 'Almost Full,' then the Acknowledge output gets asserted only on every *third*

write-clock pulse, rather than continuously. Thus, if the Acknowledge output has been tied back to the Enable input, the wide-open data rate then gets slowed down immediately, so that the writing of each word can be handled on a full-handshake basis. This operational technique allows achieving the maximum data rate much of the time, and yet protects the system against data loss caused by overrunning the FIFO boundaries.

When the system is operating an LH5420 in block-transfer mode, where a full block gets loaded at one port and then gets unloaded at the other port, the Acknowledge signals may be used to locate the end of a block, in lieu of having to implement an external block-length counter. As a simple example, say that the system block length is 193 words. The sending port loads in one complete block, and 55 words from the next block, in burst mode. At this point, its Acknowledge signal gets deasserted, indicating that the FIFO is 'Almost Full.' The Acknowledge signal does behave exactly in this manner, provided that the corresponding 'Almost-Full' flag offset still remains at its default value of eight. The receiving port then unloads the block. If its 'Almost-Empty' offset value has been set to 55, its Acknowledge signal will get deasserted exactly at the end of the block. Since this indication occurs within a clock period, it is fast enough to be accurate without any uncertainty.

The LH5420's parity-checking facilities treat all nine bits alike, of each byte passing through one of the two FIFOs; the 'parity bit' may be in any position within a byte. A ten-input parity gate scans each group of nine bits in the output register of each port; the tenth input of each parity gate is from the even/odd-parity control flipflop, which may be programmed from Port A. This flipflop is set for odd parity when the LH5420 is reset; but it may be reprogrammed to even, or back to odd, at any time subsequently. If any of the four parity gates at a port ever detects an odd number of 'ones' in a byte, *including* the control flipflop in the 'ones' count, then the port's parity flag is asserted as long as the word containing the erroneous byte remains in the output register.

SUMMARY

The LH5420 36-bit bidirectional synchronous FIFO, available now from Sharp, is a system-oriented 'one-chip-does-all' part, intended to simplify back-and-forth communications between two microprocessors, microcontrollers, or similar devices.

The LH5420 offers several sophisticated features: on-the-fly parity checking, word-width matching of a 36-bit bus to an 18-bit bus or to a 9-bit bus, two-way mailbox communications, and synchronous Acknowledge signals which can be used to give a quick and accurate end-of-block indication or an advance warning of FIFO fullness or emptiness.

In most bidirectional-FIFO applications, one LH5420 replaces many lower-level and discrete parts, and simplifies system design. It offers high performance for burst operations; it can transfer a 36-bit word in each direction every 25 nanoseconds.

COPYRIGHT INFORMATION: This paper is a slightly modified version of the paper with the same title which appeared in the *Northcon/91 Conference Record*, paper D6/1; 1-3 October 1991. Also, in the *Wescon/91 Conference Record*, paper 7/4; 19-21 November 1991.

FIFO ENHANCEMENTS, AND HOW YOUR SYSTEM CAN USE THEM

Chuck Hastings
Marketing/Applications Manager, FIFO and Specialty Memory
Sharp Microelectronics Technology, Inc.
5700 N.W. Pacific Rim Boulevard
Camas, WA 98607
206/834-8615

and

Matthew K. Walsh
Eastern Field Applications Engineering Manager
Sharp Microelectronics Technology, Inc.
14 Second Avenue, Building 1, N.W. Park
Burlington, MA 01803
617/270-7979

INTRODUCTION

Single-chip First-In, First-Out memories ('FIFOs') have been available since 1969, when Fairchild introduced the 64 × 4, 800-kHz, type 3341. Today's production-model FIFOs are often 40 MHz, with even faster ones being sampled. Although 64 × 4 FIFOs, pin-compatible with the venerable 3341, are still sold, the most common size today is 1024 × 9. And other parts as deep as 16384 words and as wide as 36 bits are in production.

More recently, there has been a virtual explosion in the *functionality* of FIFOs. They still handle data words on a First-In, First-Out basis. But they also, now,

- Provide the system with useful information as to how many data words are being stored.
- May be cascaded (to create a deeper 'effective FIFO'), or paralleled (to create a wider 'effective FIFO'), in several different ways.
- May check the parity of data words passing through them.
- May include 'funneling/defunneling,' to convert data back and forth between different word- lengths.

'Serialization/deserialization' is the one-bit-wide limiting special case of 'funneling/defunneling.'

Even more enhancements, beyond these, are being urged upon semiconductor manufacturers by today's demanding FIFO customers.

This paper provides a perspective on several of these FIFO enhancements, along with the applications rationale for them, and describes how some of them fit into the architecture of the new Sharp LH5420 36-bit bidirectional FIFO.

FIFO ENHANCEMENTS

There are two main types of FIFO enhancements:

(1) Enhancements which simply improve the usefulness and controllability of FIFOs in carrying out their basic 'rubber-band-memory' or 'FIFOing' task.

(2) Enhancements which go beyond plain ordinary 'FIFOing,' to provide a convenient means of performing various useful tests and modifications on data words as they pass through FIFOs.

Historically, Type (1) enhancements came first. The first two, a decade or more ago, were three-state outputs (with an Output Enable control signal) and a Half-Full Flag. Prior to that time, FIFOs needed eight-bit three-state 'interface' buffers, after their outputs, whenever they were required to be able to connect to/disconnect from a bus. And numerous customers were using two 'depth-cascaded' FIFOs in series, even when they didn't need that many words of FIFO capacity, just to derive a Half-Full Flag for the 'effective FIFO.' (An 'effective FIFO' is an interconnected group of FIFO devices, behaving overall as one single larger FIFO). The usual system purpose of the Half-Full Flag was to serve as a 'set point,' in the control-theory sense of that term, about which the FIFO was constrained to operate.

The next major enhancement, 'Almost' Flags, originally was introduced by Monolithic Memories as a product-definition *afterthought* – to use up all twenty pins of the type 67413 64 × 5 bipolar FIFO! The product goal was to provide an 'early warning' of impending FIFO fullness or emptiness, before one of those conditions actually came to pass, analogous to the 'yellow warning' indicator provided in the architecture of the Digital Equipment

PDP-11® series of computers. But there was only one uncommitted pin left in the 67413 pinout; hence, the flag feature as implemented signified 'Almost Full or Almost Empty' (AFE). Because the 67413 also had a Half-Full Flag, external logic could easily tell the 'Almost-Full' condition from the 'Almost-Empty' condition. 'Almost Full' was defined as fifty-six or more words filled, and 'Almost Empty' was defined as eight or fewer words filled.

Cypress®, IDT, Sharp, and other FIFO manufacturers adapted and improved this AFE-Flag feature. Higher-pin-count packages allowed the Almost-Full Flag and the Almost-Empty Flag each to have their own separate pin. Even more important, the definition of what these conditions meant was implemented in logic, with *loadable* 'offset' registers, so that the system could *program* how many words away from totally full (the 'offset') the Almost-Full Flag first would be asserted, and likewise for the Almost-Empty Flag.

Although this 'programmable-flag' feature was introduced originally as a more-flexible form of 'early warning,' customers soon found an entirely different use for it – as a low-cost means of counting the length of a datablock passing through a FIFO, by programming the offset to be equal to the FIFO depth (its word capacity) minus the desired block length. Which meant that block counts now had to be precise, and not just accurate to within one word or so, more or less. So the next customer demand was that the flags all must be FAST, meaning that the correct updated flag values had to be available within one word time; or, better yet, within the FIFO's data-access time! By this time, most newer FIFOs featured five flags: Full, Almost-Full, Half-Full, Almost-Empty, and Empty.

A FIFO is intended to be a bridge between two systems or subsystems which don't necessarily operate on the same clock, or perhaps don't even operate synchronously. Now, a Write operation at the input port adds one word to the FIFO's internal memory, which moves the evaluation of each of these five flags forward one step towards fullness. Likewise, a Read operation at the output port subtracts one word from the FIFO's internal memory, which moves the evaluation of each of these five flags back one step towards emptiness. If a Write operation and a Read operation occur simultaneously, all flag values are supposed to stay the same. The important point here is that each flag may be affected by an operation synchronized to the input-port clock, and also may be affected by an operation synchronized to the output-port clock – and that these clocks probably aren't even directly related to each other.

So, what happens when the external logic tries to read the value of a flag? Instant *metastability hazard*, since this external logic almost certainly is associated with the clock at one of the two ports, but not with the clock at the other port. The way to resolve this potential metastability problem is to make all five of the flags be 'synchronous' – that is, to allow each flag to change state only when triggered by the clock from that port where it is most likely to be read.

The Full Flag and the Almost-Full Flag, for instance, are of most interest to the input-port logic, which needs to be warned when there's no longer room within the FIFO memory for many more words. Conversely, the Almost-Empty Flag and the Empty Flag are of most interest to the output-port logic, which needs to be warned when there are only a very few words left within the FIFO memory. The Half-Full Flag may be up for grabs; often it is assigned to the input port for synchronization purposes.

Anyway, making all of the five standard flags synchronous entails some additional synchronization logic within the FIFOs. But doing so makes FIFOs much easier to use in systems, and so FIFO customers want this flag behavior very much. Thus, a rapid industry-wide shift to 'synchronous' flags is going on at present.

MORE COMPLEX FIFOs

A common situation is for FIFOs to be used to implement two-way data communication between two microprocessors or microcontrollers, which leads to two FIFOs being interconnected 'back-to-back' – the input port of each is tied directly to the output port of the other. This 'bidirectional-FIFO' or 'BiFIFO' configuration has been turned into full- duplex dedicated BiFIFO parts, by AMD, by IDT, and by Sharp.

Full-duplex BiFIFOs, such as Sharp's 40-MHz full-word-width, 36-bit LH5420, contain two physical FIFO macrocells on one single piece of silicon, connected 'back-to-back' between the two ports.

There are, however, such things as half-duplex BiFIFOs. One of these can carry data messages in just one direction at a time, and contains only one single physical FIFO macrocell. But it can reverse the direction in which its internal physical FIFO is interconnected between its two ports.

Data communication between two processors relies on software protocols. Thus, BiFIFOs typically include a 'mailbox' feature – a single-word register, in parallel with each main physical FIFO, between the BiFIFO's two ports. Also between the two ports, in parallel with the FIFO and with the mailbox, there sometimes is a 'bypass' data path, which allows for direct high-priority communications without disturbing the contents either of the FIFO or of the mailbox. The LH5420, for instance, features a mailbox in each direction, with a 'new-mail-alert' flag to indicate when there is updated mail to be read; and there is also a bidirectional data-bypass facility.

Implementing data communication between processors or buses having different wordlengths is an aggravating design problem for systems companies. Increasingly, however, BiFIFOs implement 'funneling/defunneling' – word-width matching. The Sharp LH5420 BiFIFO, for instance, has a 36-bit Port A, and a *variable-*

width Port B, which can operate as a 36-bit port, as an 18-bit port, or as a 9-bit port. Note that we've now moved beyond simple Type (1) 'FIFOing' enhancements; 'funneling/defunneling' actually *does* something to the data words as they pass through the FIFO, and hence may be considered a Type (2) enhancement.

The extreme case of 'funneling/defunneling,' of course, is where the narrower bus is just one bit wide. Several manufacturers produce FIFOs which serialize or deserialize, or even do both. Sharp's serializing LH5493 and deserializing LH5494 are examples. Both of these FIFOs are 4096 × 9, and operate at 40 MHz.

Another Type (2) enhancement, which the LH5420 also incorporates, is parity checking. Each new data word which enters the LH5420 from either port, bound for the other port, gets parity-checked before it is stored into the respective internal FIFO memory going in the appropriate direction. One LH5420 36-bit word is examined as four 9-bit bytes, and each byte is individually parity-checked. A parity flag indicates whenever a byte with erroneous parity gets detected; each port has its own parity flag. The parity scheme may be selected, via a programmable control register, to be either 'even parity' or 'odd parity.'

Although early FIFOs made use of 'bucket-brigade' shift-register technology for data storage, almost all FIFOs in use today incorporate internal static random-access memory (SRAM) for data storage; a few large, slow, low-cost FIFOs use internal dynamic random-access memory (DRAM) instead. SRAM-based FIFOs all have the property that, after a datablock has been read out via the output port, it still is present within the FIFO device, if only the FIFO logic is smart enough to allow it to be reread. Since there often is a system purpose for reading out the same datablock more than once, many recent SRAM-based FIFOs have a 'retransmit' or 'reread' facility, which moves the FIFO's internal readout pointer back either to the first internal physical-memory location, or to a location addressed by a special internal register. System applications for such FIFOs include data communications, signal-processing memory (for data samples, and also for coefficients), display refreshment, and repetitive waveform generation. The LH5420 provides a 'retransmit' facility in both directions.

Other Type (2) enhancements now being studied by FIFO manufacturers for future products include limited forms of data reformatting/format conversion, matching against a test word, and parity generation.

CONCLUSION

The constant pressure for progress in digital electronics is leading to more and more functionality within newer FIFOs, which makes them easier to use in systems. Today, these functionality enhancements have gone beyond mere 'FIFOing,' and are aimed at eliminating other external logic from customers' boards by absorbing additional functions into FIFOs. A new generation of 'system-level FIFOs' is starting to appear, beginning with the Sharp LH5420 36-bit bidirectional FIFO.

COPYRIGHT INFORMATION: This paper is a slightly modified version of the paper with the same title which appeared in the *Electro92 Conference Record*, paper 58/1; 11-14 May 1992.

FUTURE TRENDS IN FIFO ARCHITECTURES

Chuck Hastings
Marketing/Applications Manager, FIFO and Specialty Memories
206/834-8615
and
Steven B. Sidman
Manager, Strategic Marketing and Product Planning
206/834-8711

Sharp Microelectronics Technology, Inc.
5700 N.W. Pacific Rim Boulevard
Camas, WA 98607

INTRODUCTION

First-In, First-Out memories ('FIFOs') are VLSI semiconductor memory devices which store information sequentially. FIFOs have a wide variety of applications. [1]

As the name 'FIFO' implies, data words may be pushed into a FIFO memory in sequence, one word at a time, without ever having to give the FIFO any 'address' to tell it precisely where to store a given word. Later on, the exact same sequence of data words may be pulled out of the FIFO at its other end, one word at a time, in the same order in which these words entered the FIFO. Because the position of each word is known within the sequence, there is no need to 'address' the FIFO to tell it where a particular word is to be found.

Single-chip LSI FIFOs have been available since 1969, when Fairchild introduced the four-bits-wide 64 × 4 800-kHz type 3341. Now, single-chip FIFO devices of several different word widths are available. The most common FIFO word width today is 9 bits. But there still are older FIFOs being sold with word widths of 4 or 5 or 8 bits, and newer FIFOs with word widths of 18 or even 36 bits. Sharp is the first FIFO manufacturer to have brought 36-bit FIFOs to market.

As semiconductor products, FIFOs are in some ways like SRAMs (Static Random-Access Memories). But, in some other ways, they are like logic devices. Like an SRAM, a FIFO includes a memory array. In most FIFOs, this memory array is very similar to an SRAM memory array; although, a few semiconductor manufacturers have used DRAM-type (dynamic, rather than static) memory cells within FIFOs.

But, unlike an SRAM, a FIFO also must include sophisticated counting and comparison logic for generating internal-memory-array addresses automatically, and for determining the relative 'fullness' of its internal-memory array. And, because in most FIFOs *all* memory addressing is sequential, the circuit-design techniques and tricks used within a FIFO memory array often exploit this sequentiality, in ways which would not be allowable within the design of a conventional SRAM device. In particular, FIFO memory arrays often are extensively 'interleaved' and 'pipelined.' [2]

Recently, there has been a virtual explosion in the *functionality* of FIFOs. They still handle data words on a First-In, First-Out basis. But today, FIFO functionality enhancements have gone far beyond mere 'FIFOing,' and are aimed at eliminating other external logic from customers' boards by absorbing additional functions into FIFOs. FIFOs already, now,

- Provide the system with useful information as to how many data words are being stored.
- May be cascaded (to create a deeper 'effective FIFO'), or paralleled (to create a wider 'effective FIFO'), in several different ways.
- May check the parity of data words passing through them.
- May include 'funneling/defunneling,' to convert data back and forth between different wordwidths. When one of these wordwidths is just one-bit, then 'funneling/defunneling' becomes 'serialization/deserialization.'

Even more enhanced features, beyond these, are being urged upon semiconductor manufacturers by today's demanding FIFO customers.

This paper discusses several of these FIFO enhancements, along with the applications rationale for them, and describes their use in two new synchronous, 36-bit-word-width, 'system-level' Sharp FIFO products: the LH5420 dual-256-word bidirectional FIFO, which is now in full production; and the LH543620 1024-word unidirectional FIFO, which is now in development.

FIFO ENHANCEMENTS

There are two main types of FIFO enhancements:

① Enhancements that simply improve the usefulness and controllability of FIFOs in carrying out their basic 'rubber-band-memory' or 'FIFOing' task.

② Enhancements that go beyond plain ordinary 'FIFOing,' to provide a convenient means of performing various useful tests and modifications on data words as they pass through FIFOs.

Table 1 lists the Type ① enhancements and Type ② enhancements which are considered in this paper.

Historically, Type ① enhancements came first. The first two, a decade or more ago, were three-state outputs (with an Output Enable control signal) and a Half-Full Flag. Prior to that time, FIFOs needed eight-bit three-state 'interface' buffers after their outputs, whenever it was required that the FIFO outputs must be able to connect to/disconnect from a bus.

And numerous customers were using two 'depth-cascaded' FIFOs in series, even when they actually didn't need that many words of FIFO capacity, just to derive a Half-Full Flag for the 'effective FIFO.' (An 'effective FIFO' is an interconnected group of individual FIFO devices, behaving overall as one single larger FIFO). The usual system purpose of the Half-Full Flag was to serve as a 'set point,' in the control-theory sense of that term, about which the effective FIFO was constrained to operate, in the same manner as a room heater operating about the 'set point' on a thermostat.

The next major enhancement, 'Almost' Flags, originally was introduced by Monolithic Memories as a product-definition *afterthought* – to use up all twenty pins of the 67413 64 × 5 bipolar FIFO! The product goal was to provide an 'early warning' of impending FIFO fullness or emptiness, *before* one of those conditions actually came to pass, analogous to the 'yellow warning' indicator provided in the architecture of the Digital Equipment PDP-11® series of computers.

But there was only one uncommitted pin left in the 67413 pinout. Hence, as implemented, the 67413 flag feature signified 'Almost Full *or* Almost Empty' (AFE). Because the 67413 also had a Half-Full Flag, external logic could tell these two conditions apart quite easily. 'Almost Full' was defined as fifty-six or more words filled; and 'Almost Empty' was defined as eight or fewer words filled.

Cypress,® IDT, Sharp, and other FIFO manufacturers adapted and improved this AFE-Flag feature. Higher-pin-count packages allowed the Almost-Full Flag and the Almost-Empty Flag each to have their own separate pin, rather than having to be 'ORed' together on the same pin. Even more important, the definition of what these conditions meant was implemented in logic, with *loadable* 'offset' registers, so that the system could *program* a value (the 'offset') for how many words away from totally full the Almost-Full Flag first would be asserted, and likewise for the Almost-Empty Flag.

Although this 'programmable-flag' feature was introduced originally as a more-flexible form of 'early warning,' customers soon found an entirely different use for it – as a low-cost means of counting the length of a datablock passing through a FIFO, by programming the offsets so that their sum is equal to the FIFO depth (its word capacity) minus the desired block length. For instance, if the desired block length was 193 words (which is used in digital telephony), and the FIFO was 256 words deep, one offset could be set to eight and the other one to 55, and the FIFO could be operated between 'Almost Full' and 'Almost Empty.'

Which meant that block counts now had to be *precise* – not just accurate to within one word or so, more or less. So the next customer demand was that the flags all must

Table 1. FIFO Enhancements

TYPE ① ENHANCEMENTS	TYPE ② ENHANCEMENTS
• Three-State Outputs	• 'Funneling/Defunneling' (Word-Width Matching)
• Half-Full Flag	• 'Big-Endian' ↔ 'Little-Endian' Conversion (Byte-Order Reversal)
• 'Almost' Flag(s)	• Serialization/Deserialization
• Programmable 'Almost'-Flag Offsets	• Parity Checking
• Synchronous Data Transfer	• Parity Generation
• Synchronous Flags	• Retransmit: Basic (to 'Location 0'), 'Smart'
• Single-Chip Bidirectional FIFOs: Full-Duplex, Half-Duplex	• JTAG Test Port
• Mailbox	
• Bypass	

be *fast*, meaning that the correct updated flag values had to be available within one word time; or, better yet, within the FIFO's data-access time! By this time, most newer FIFOs featured five flags: Full, Almost-Full, Half-Full, Almost-Empty, and Empty.

In the 64-words-deep 67413, eight words was one-eighth of the depth of the entire FIFO memory array. Nowadays, virtually all newer FIFOs are much deeper than 64 words. Still, almost all of them use either eight words, or else one-eighth of the number of words in the entire FIFO memory array, as their 'default' value for the contents of the loadable 'Almost-Full' and 'Almost-Empty' programmable-flag-offset-value registers.

FIFOs usually are marketed in product *families*, with the depths of the family members differing one from another by powers of two. Choosing eight words as the default offset value implies that all family members use the same default offset value. Whereas, choosing one-eighth of the depth as the default offset value implies that, for the deeper family members, these default offset values increase with depth.

A FIFO is intended to be a bridge between two systems or subsystems which don't necessarily operate according to the same clock, or perhaps don't even operate synchronously. Now, a Write operation at the input port adds one word to the FIFO's internal memory, which moves the evaluation of each of these five flags forward one step towards fullness. Likewise, a Read operation at the output port subtracts one word from the FIFO's internal memory, which moves the evaluation of each of these five flags back one step towards emptiness.

If a Write operation and a Read operation occur simultaneously, all flag values are supposed to stay the same. The important point here is that each flag may be affected by an operation synchronized to the input-port clock, and also may be affected by an operation synchronized to the output-port clock – *and*, that these clocks probably aren't even directly related to each other.

So, what happens when the external logic tries to read the value of a flag? Instant *metastability hazard*, since this external logic almost certainly is associated with the clock at one of the two ports, but not with the clock at the other port. Thus, it is all too possible that the external logic may attempt to read the flag value, at the exact instant while it is in the process of changing.

The way to resolve this potential metastability problem is to make all five of the flags be 'synchronous' – that is, to allow each flag output to change state only when triggered by the clock from that port where it is most likely to be read. The Full Flag and the Almost-Full Flag, for instance, are of most interest to the input-port logic, which needs to be warned when there's no longer room within the FIFO memory for very many more words. Conversely, the Almost-Empty Flag and the Empty Flag are of most interest to the output-port logic, which needs to be warned when there are only a very few words left within the FIFO memory. The Half-Full Flag may be up for grabs; often, for synchronization purposes, it is assigned to the input port.

Anyway, making all of the five standard flags 'synchronous' entails some additional synchronization logic within the FIFOs. But doing so makes FIFOs much easier to use in systems, and so FIFO customers (at least, those with some sophistication about metastability!) very much prefer this flag behavior. Thus, a rapid industry-wide shift to 'synchronous' flags is going on at the present time.

MORE COMPLEX FIFOs

A common situation is for FIFOs to be used to implement two-way data communications between two microprocessors or microcontrollers, which leads to two FIFOs being interconnected 'back-to-back' – the input port of each is tied directly to the output port of the other. This 'bidirectional-FIFO' or 'BiFIFO' configuration has been turned into full-duplex dedicated BiFIFO parts, by AMD and IDT, and more recently by Sharp.

Full-duplex BiFIFOs, such as Sharp's full-word-width 36-bit LH5420 [3], contain two physical FIFO macrocells on one single piece of silicon. (See Figure 1.) These are connected 'back-to-back,' between the two ports. At any given time, such a full-duplex BiFIFO can be performing two simultaneous write operations or two simultaneous read operations, with both physical FIFOs active. Or, it can be performing a write operation and a read operation simultaneously, with one physical FIFO active and one physical FIFO idle.

There are, however, also such things as *half-duplex* BiFIFOs. One of these can carry data messages in just one direction at a time, and contains only one single physical FIFO macrocell. But it can at least reverse the direction in which its internal physical FIFO is interconnected, between its two ports. So it can perform a write operation and a read operation simultaneously, in either direction, like a full-duplex BiFIFO. But, unlike a full-duplex BiFIFO, it cannot perform two simultaneous write operations, or two simultaneous read operations. And it cannot begin a data transfer in one direction, until it has completed any data transfer which may have been in progress in the other direction, and its internal FIFO-memory array has been completely emptied out.

Data communication between two processors relies on software protocols. Thus, BiFIFOs typically include a 'mailbox' feature – a single-word register, in parallel with each main physical FIFO, between the BiFIFO's two ports, to convey 'mail.' Also between the two ports, in parallel with the FIFO and with the mailbox, there sometimes is a 'bypass' data path, which allows for direct high-priority communications without disturbing the contents either of the FIFO or of the mailbox. The LH5420, for instance, features a mailbox in each direction, with a 'new-mail-alert' flag to indicate when there is updated mail to be read; and there is also a bidirectional data-bypass facility.

Implementing data communication between processors or buses having different wordwidths is an aggravating design problem for systems companies. Increasingly, however, BiFIFOs implement 'funneling/defunneling' – word-width matching. The Sharp LH5420 BiFIFO, for instance, has a 36-bit Port A, and a *variable*-width Port B. Port B can operate as a 36-bit port, as an 18-bit port, or as a 9-bit port. [4]

The Sharp LH543620 1024 × 36 unidirectional FIFO has a different form of this word-width-matching feature, adapted to its unidirectional architecture. *Both* its Input Port and its Output Port have similar variable-width capabilities, with somewhat differing implementations due to the different structures of these two ports.

Although the LH543620 itself is unidirectional, it is intended to support very deep BiFIFO applications, with two parts – or, perhaps, even two *cascaded strings* of parts – connected 'back-to-back.' Accordingly, it too has mailbox and bypass facilities. (See Figure 2.)

Because, in the past, competing major computer-systems vendors made different architectural choices regarding byte ordering within words, an almost-universal data-formatting nuisance in processor-to-processor communications these days is 'Big-Endian' ↔ 'Little-Endian' conversion. (This terminology is a computer-technology adaptation of a phrase used in Jonathan Swift's satirical literary classic, *Gulliver's Travels*.) The LH543620 can perform this conversion as a 'byte-order-reversal' operation, on the fly, on each 36-bit data word passing through it.

Note that we've now moved beyond simple Type ① 'FIFOing' enhancements. 'Funneling/defunneling' and 'byte-order-reversal' actually *do* something to the data words as they pass through the FIFO, and hence are clearly Type ② enhancements.

The extreme case of 'funneling/defunneling,' of course, is where the narrower bus is just one bit wide. Several manufacturers produce FIFOs that serialize or deserialize, or even do both. Sharp's serializing LH5493 and deserializing LH5494 are examples. Both of these FIFOs are 4096 × 9, and operate at 40 MHz.

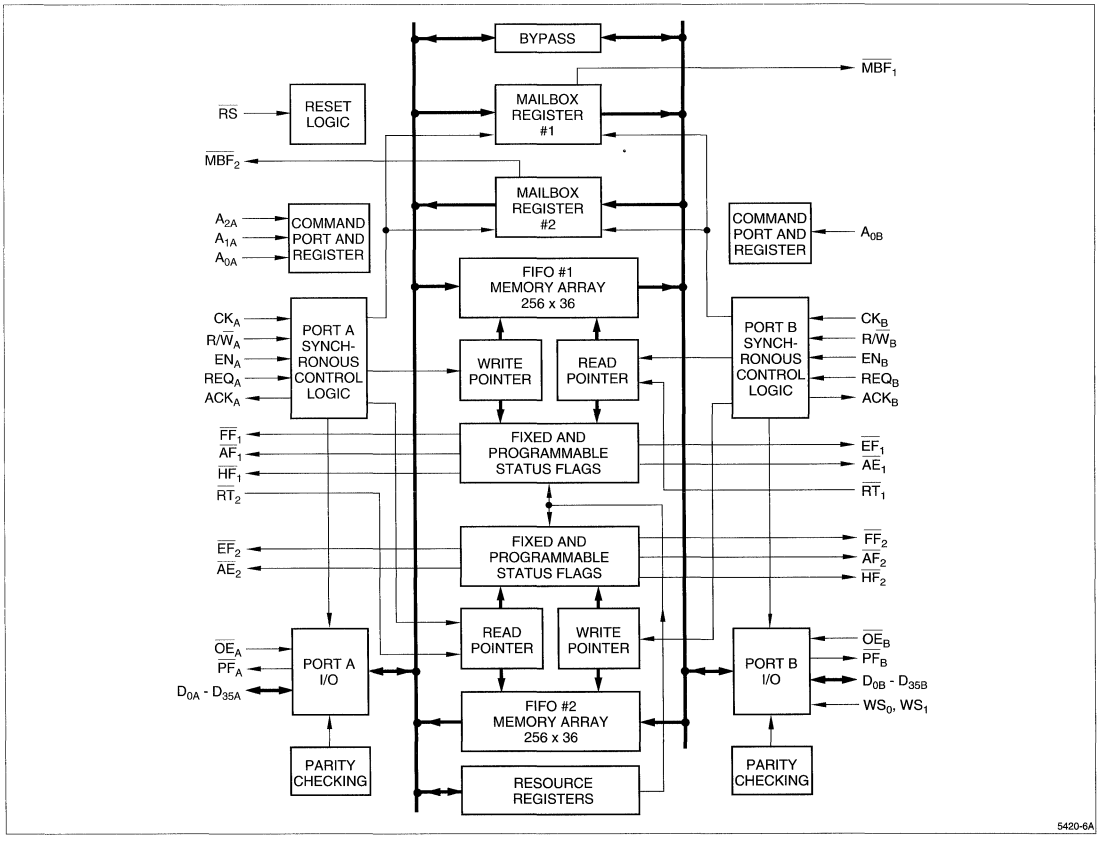


Figure 1. LH5420 256 × 36 × 2 Synchronous Bidirectional FIFO

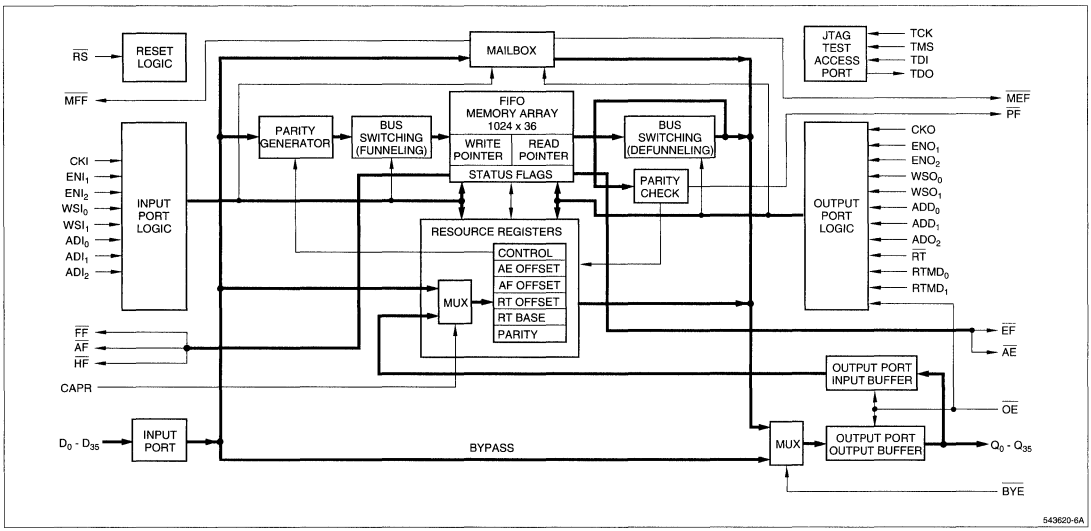


Figure 2. LH543620 1024 × 36 Synchronous Unidirectional FIFO

Another Type ② enhancement, which the LH5420 and LH543620 also incorporate, is parity checking. Each new data word which enters the LH5420 from either port, bound for the other port, gets parity-checked twice; once before it gets stored, into whichever of the two internal FIFO-memory array goes in the appropriate direction; and again when it gets read out of the internal FIFO-memory array, and arrives at the opposite port. The inputs to the parity-checking circuits for each port are connected at the data input/output bonding pads for that port.

One LH5420 36-bit word is examined as four nine-bit bytes; each byte is individually parity-checked. A parity flag indicates whenever a byte with erroneous parity gets detected. Each port has its own parity flag. The parity scheme may be selected, via a programmable control register, to be either 'even parity' or else 'odd parity,' and may be changed subsequently at any time.

Besides checking parity, the LH543620 has the additional capability of *generating* parity, 'on the fly,' from the eight least-significant bits of each of the four nine-bit bytes of a 36-bit word, and of storing the generated parity bit for each byte into its most-significant bit.

Although early FIFOs made use of 'bucket-brigade' shift-register technology for data storage, almost all FIFOs in use today incorporate internal static random-access memory (SRAM) for data storage. However, there are a few large, slow, low-cost FIFOs which use internal dynamic random-access memory (DRAM) instead. SRAM-based FIFOs all have the property that, after a datablock has been read out via the output port, it still is present within the FIFO device, if only the FIFO logic is smart enough to allow it to be reread.

Since there often is a system purpose for reading out the same datablock more than once, many recent SRAM-based FIFOs have a 'retransmit' or 'reread' facility, which moves the FIFO's internal readout pointer back to the first internal physical-memory location ('Location 0'). The LH543620 has a new 'smart retransmit' feature, with several operating modes, which goes even further, and permits resetting the read pointer to *any* prestored internal-FIFO-memory-array address value. System applications for FIFOs with such retransmit capabilities include data communications, DSP (digital-signal- processing) memory for data samples and also for coefficients, display refreshment, and repetitive waveform generation. The LH5420 'retransmit' facility can operate in either direction.

One more new LH543620 feature, applicable to system reliability and self-test design practice, is an IEEE1149.1-compliant ('JTAG') test port, believed to be the first such use of a JTAG port in a FIFO.

CONCLUSION

The constant pressure for progress in digital electronics is leading to more and more functionality within newer FIFOs, which is making them easier and easier to use in systems. Today, these functionality enhancements have gone far beyond mere 'FIFOing,' and are aimed at eliminating other external logic from customers' boards by absorbing additional functions into FIFOs. A new generation of 'system-level FIFOs' is starting to appear, beginning with two Sharp 36-bit synchronous FIFOs: the LH5420 bidirectional FIFO, and the LH543620 unidirectional FIFO.

REFERENCES

[1] Chuck Hastings, 'FIFO Memories: Effective, Compact, and Easy to Use,' *Northcon/92 Conference Record;* Session M2, Paper 2. Available from Electronic Conventions Management (ECM), 8110 Airport Boulevard, Los Angeles, CA 90045-3194. Also reprinted in the *Sharp 1992 Memory Data Book.*

[2] Steven Sidman, Dieter Spaderna, Jeffrey Miller, and David Jenkins, 'FIFOs – Innovation Through Architecture,' *Electro/91 Conference Record,* pages 142-143; Session 9, Paper 2. Available from ECM as above, or from Technical Publications Group, Sharp Microelectronics Technology, Inc., 5700 N.W. Pacific Rim Boulevard, Camas, WA 98607.

[3] Chuck Hastings, 'A One-Chip Two-Way Street for Microprocessor Communications: The Sharp LH5420 36-Bit Bidirectional FIFO,' *Northcon/91 Conference Record,* pages 220-225; Session D6, Paper 1. Also, *Wescon/91 Conference Record,* pages 130-133; Session 7, Paper 4. Available from ECM as above; also reprinted in the *Sharp 1992 Memory Data Book.*

[4] Mike Yee, 'Databus Funneling Made Easy,' reprinted in the *Sharp 1992 Memory Data Book.*

INTEGRATED CIRCUITS

SRAM, Pseudo SRAM

AUTOMATIC POWER DOWN

INTRODUCTION

This application note describes a feature available in some CMOS static RAMs known as Automatic Power Down. Automatic Power Down is a circuit design technique that reduces the supply current required by the static RAM during long Read and/or Write cycles. Since many systems using static RAMs operate at cycle times significantly exceeding the minimum cycle time allowable, Automatic Power Down is a very useful feature for many systems with strict power budgets.

Equivalent System Level Implementation

The circuit in Figure 1 represents a common system-level technique used for the reduction of supply current in systems with long Read cycles. The static RAM in the circuit has a much faster access time than the system cycle time, along with a transparent latch used in the Data-out path. The static RAM is enabled only for a short interval following the beginning of the Read cycle, long enough for valid data to appear at the Data bus. The latch is transparent while the static RAM is enabled, latching the accessed data when the static RAM is disabled. The time where the static RAM is enabled is controlled either by a delay line or another similar circuit technique. This circuit operates by reducing the duty cycle for the static RAM, thereby reducing the average DC supply current. The two implementations result in the average instantaneous current levels shown in Figure 2. Note that for the timing shown, the implementation using the fast static RAM requires less current from the power supply, even though both Active and Standby current specifications are higher than the slow SRAM. The values used are typical values, and may vary significantly for specific devices and system timing values. Figure 3 contains the system timing waveforms.

On-Chip Implementation

Automatic Power Down circuit is implemented on-chip with circuitry similar to the system level circuitry (see Figure 4). In addition to the long Chip Enable controlled Read cycles, the on-chip implementation also reduces power following an extended Address access cycle. Since a standard SRAM is often accessed with the Address changing after the falling edge of Chip Enable, the SRAM must time the access from both Chip Enable and Address bus transitions. Changes in the state of the Address bus, while the Chip is enabled, are detected by a frequently used (and well named) circuit: the Address Transition Detector. Further details of this circuit are the subject of another application note.

Some static RAM's Automatic Power Down circuits may also reduce supply current during extended Write cycles. The fundamental difference between Read and Write cycles is that all input signals are valid at the beginning of a Read cycle. Chip Enable and Write Enable are LOW and the Address bus is valid at the beginning of the Write cycle, but the Data-in bus may change during almost any time during the entire Write cycle. The Automatic Power Down circuit must detect

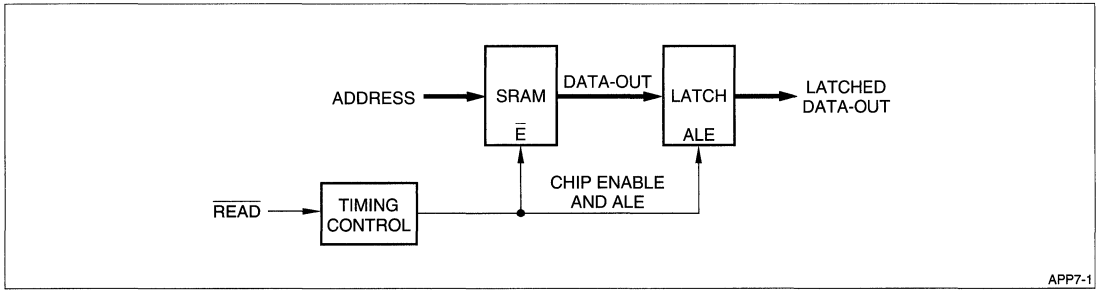


Figure 1. System Level Implementation

changes in the Data-in bus in order to minimize supply current during long Write cycles. This circuit is effective, provided that the Data-in bus is stable during the Write cycle. The circuit used to detect Data-in bus transitions is similar to the circuit used to detect Address bus transitions. Static RAMs that do not reduce current during extended Write cycles will not benefit in the lower supply current requirement.

The block diagram in Figure 4 describes the method used for implementing Automatic Power Down on-chip for a static RAM. Note the dashed line indicating the optional used of Automatic Power Down for Write cycles.

Supply Current Specifications

When the Automatic Power Down circuit is activated, reducing the supply current to a level near the Standby value, the instantaneous current level is reduced. The measurement of instantaneous supply current is very difficult and requires very sophisticated instrumentation. The system designer requires the specification of the static RAM's average DC supply current, under the conditions of system operation. The system variables involved are supply voltage, ambient temperature, static RAM cycle time and, to some extent, the input voltage levels for the static RAM. The Automatic Power Down advantage appears in the specification of significantly reduced supply current at longer cycle times. Since many static RAMs operate at cycle times much lower than the minimum allowed cycle time, the Automatic Power Down feature provides a significant reduction in actual supply current.

SUMMARY

Static RAMs are used in many applications, some with stringent power supply requirements. Those applications with relatively long Read and/or Write cycles will benefit from reduced power supply current with the use of static RAMs with Automatic Power Down.

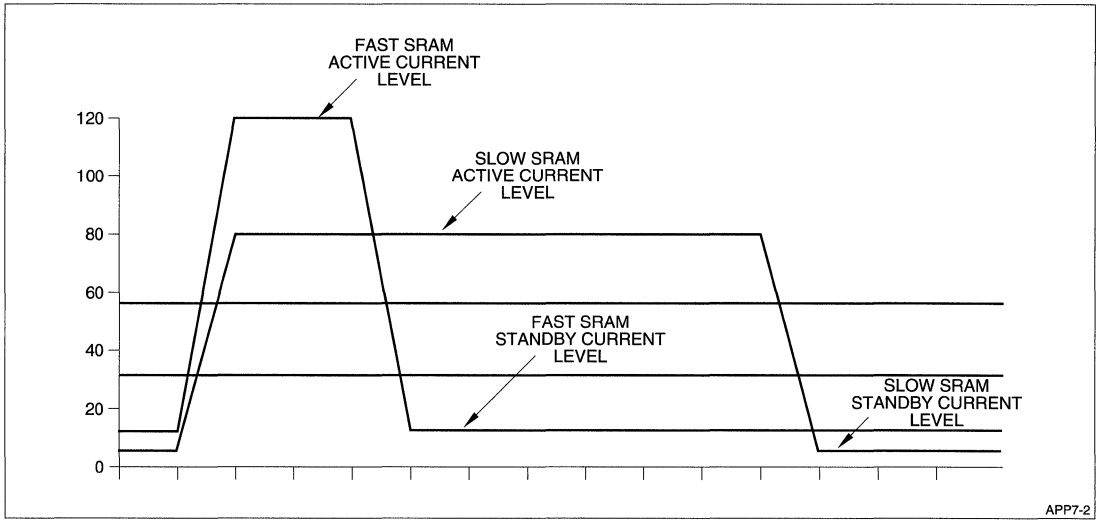


Figure 2. Current Waveforms

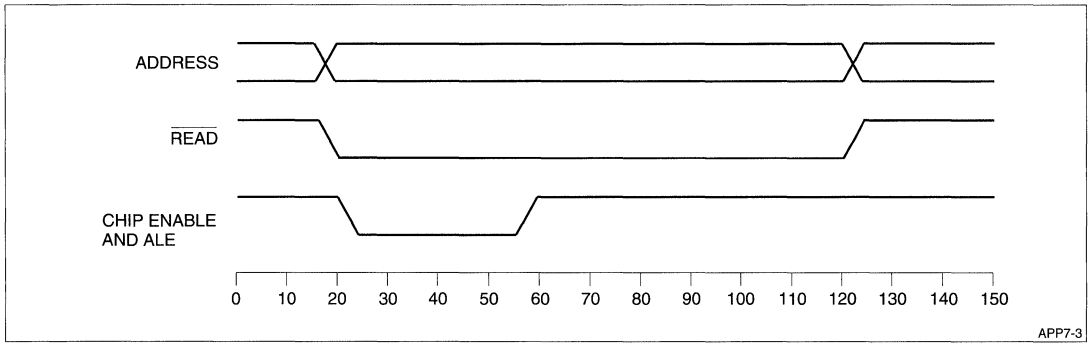


Figure 3. Timing Diagram

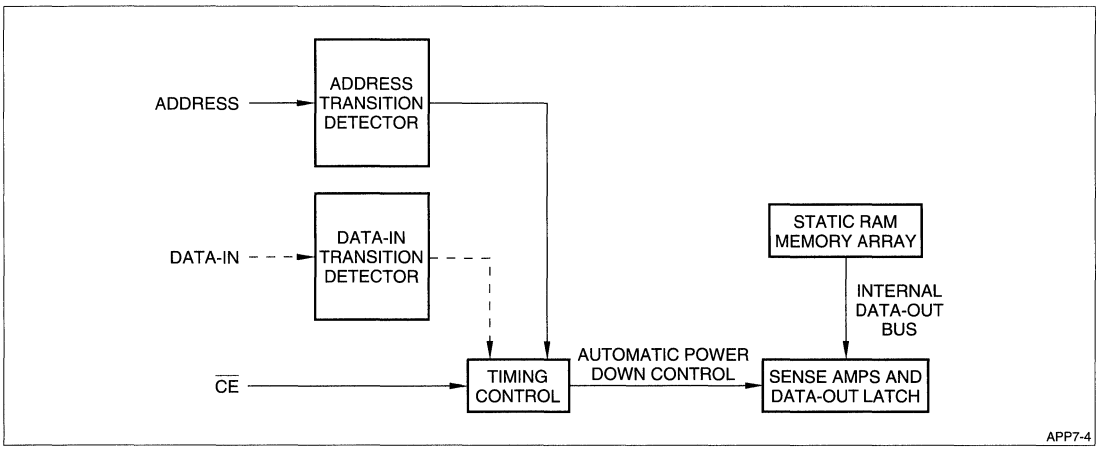


Figure 4. On-Chip Automatic Power Down

POWER GATING

INTRODUCTION

Power Gating is a design technique used in CMOS integrated circuits that reduces supply current under certain specific conditions. Slow CMOS memories frequently use Power Gating to reduce standby current. Slow CMOS static RAMs often provide the Power Gating feature, allowing the Address and possibly the Control input signals (those other than Chip Enable) to float during Standby or Data Retention modes. Power Gating is provided on many Sharp CMOS static RAMs, including several with fast access times.

The Power Gating Advantage

CMOS integrated circuits use CMOS switching levels internally, with TTL-to-CMOS or CMOS-to-TTL level conversion circuits for Input/Output where TTL voltage level compatibility is required. The TTL-to-CMOS voltage level conversion circuits consume significant power. Unlike standard CMOS logic gates, TTL-to-CMOS converters, or Input buffers, consume power while switching and while at static TTL HIGH or LOW voltage levels. Power Gating eliminates both the static and switching power consumed by the Input buffers.

Those static RAMs with the feature typically use Power Gating for the Address, Write Enable and Output Enable (if available) Input buffers. Chip Enable controls the Power Gating function. When any static RAM is in Standby mode (Chip Enable is HIGH), the other inputs are logically Don't Care. Those static RAMs that do not have Power Gating will still dissipate power in the Input buffers if the voltage is not near either V_{CC} or V_{SS}, while the input voltage is either static or switching. Figure 1 has an example of this standard input buffer. The same inputs on static RAMs with Power Gating are also logically Don't Care during Standby mode, but do not dissipate any power. The voltage at these terminals may vary at any frequency or voltage level and the power will not increase.

The Power Gated Input buffer, as shown in Figure 2, typically uses a NOR type logic gate instead of a CMOS inverter. NAND gates are used in some cases. Figure 3 is an example for the two circuits, with the current graphed as a function of input voltage. The graph indicates that some current is required at the worst case TTL input levels (0.8 and 2.2 V), while the current increases significantly when the voltage is near the actual switch point. The current flow is from V_{CC} to V_{SS} as shown, not into the input terminal.

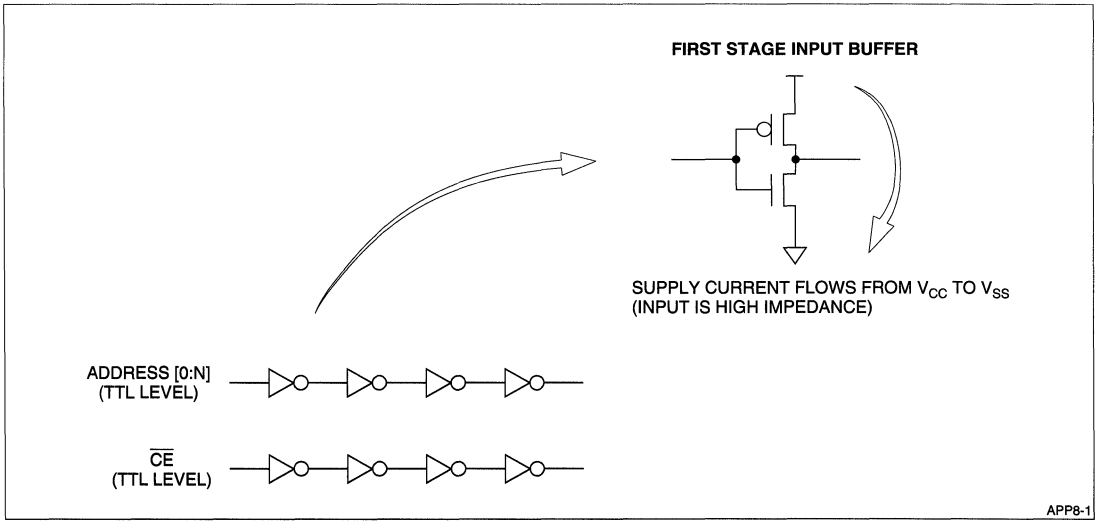


Figure 1. Standard Input Buffer Without Power Gating

The Power Gated input, while in Standby mode, does not allow current to flow when the input is at any voltage level. In Active mode (the SRAM's Chip Enable is LOW), the circuit displays characteristics similar to a circuit that is not Power Gated.

Advantages of Power Gating in Static RAMs

There are two advantages in the use of Power Gating in static RAMs. Both Standby and Data Retention modes see improvement with Power Gating. Figure 3 contains a rough comparison of the typical I_{CC} waveforms experienced by both the Power Gated input and the standard Input buffers.

The first benefit is during Standby mode, a Power Gated device's Input buffers do not draw current at any input voltage level. Input buffers that are not Power Gated can consume considerable current, especially at the worst case V_{IH} level (2.2 V). Also, there is no current required when the input level transitions HIGH-to-LOW or LOW-to-HIGH.

Also, every transition on these inputs will induce additional current flow through the input circuits as the input voltage level nears the switch point (typically 1.5 V). In addition to increased power, the static RAM that does not have Power Gated inputs will contribute more switching noise, in the form of di/dt, every time the inputs switch in Standby mode.

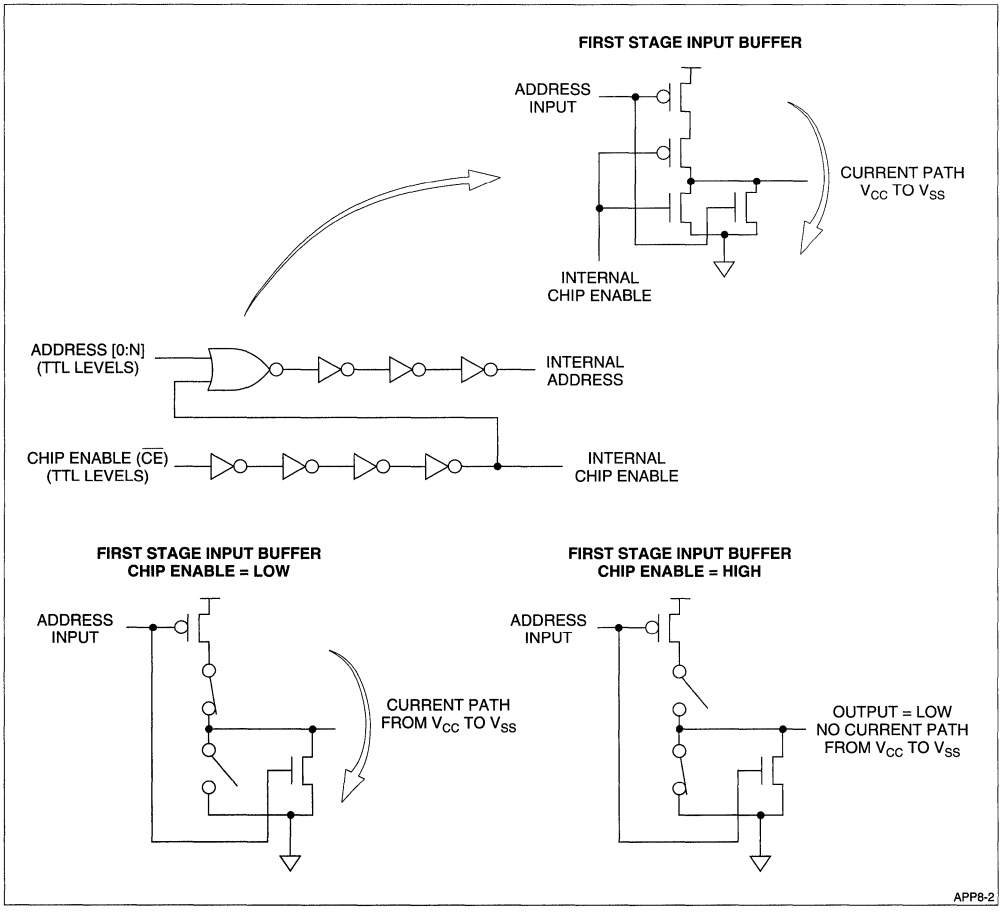


Figure 2. Input Buffer With Power Gating

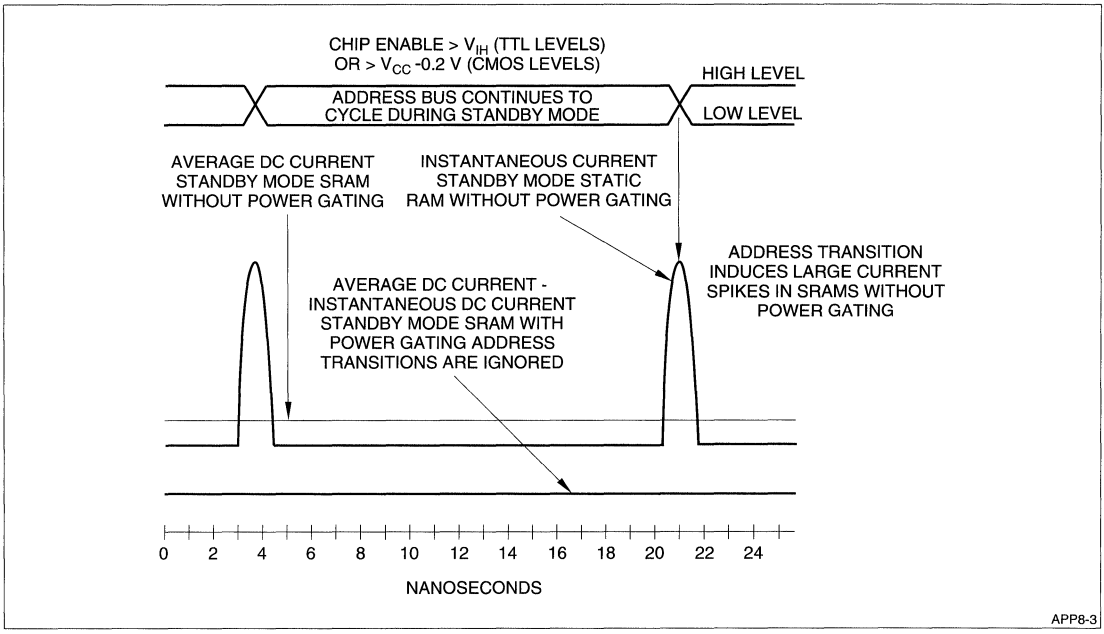


Figure 3. Comparison of I_{CC} Waveforms

The typical static RAM application using Standby mode will have the Address bus toggling during Standby mode. The static RAM with Power Gated inputs provide lower power and reduced noise in any system application using Standby mode.

The second advantage occurs when the static RAM is operated in a Data Retention mode. Static RAMs usually receive their supply current from batteries while in Data Retention mode. In order to maximize the operating life of the batteries, there is a need for a small Data Retention mode supply current.

Power Gating assists in the minimization of Data Retention current. For Power Gated static RAMs, the inputs other than Chip Enable may float during Data Retention mode. Static RAMs that are not Power Gated require controlled voltages at all input terminals. These input terminals must be driven usually within 0.2 V of either V_{CC} or V_{SS}. Generally, this requires an additional CMOS logic device that is an additional power drain on the battery.

Analysis of Product Specifications

The use of Power Gating primarily impacts two product specifications: Data Retention current and Standby current. Typical examples of the Data Retention and Standby Current specifications are shown in Tables 1 and 2.

The Data Retention current value does not change due to Power Gating, but the specification and system implementation do change for static RAMs without Power Gating. A note indicates that all inputs, including Chip Enable, requires input voltage levels usually within 0.2 V of either V_{CC} or V_{SS}. This means that the Address and Control terminals must not float. Static RAMs with Power Gating have this restriction only on the Chip Enable terminal – the others may float.

Note the significant difference in the static RAM without Power Gating requires a controlled voltage at all input pins, including the Data bus, while the static RAM without Power Gating requires a controlled voltage only on the Chip Enable pin.

Power Gating also provides improvement in the Standby Current specification. CMOS static RAMs typically have two Standby Current specifications, with either CMOS or TTL input voltage levels. The TTL specification covers the worst case TTL input conditions: 2.2 V for a HIGH level and 0.8 V for a LOW level. The CMOS specification requires that input signals are at 0.2 V for a LOW level and V_{CC} – 0.2 V for a HIGH level (usually 5.3 V). Static RAMs operate in Standby mode with Chip Enable and all other inputs at the appropriate voltage level, but all other inputs (i.e., the Address bus) toggle. Static RAM manufacturers test Standby Current with the Address toggling at the minimum cycle time.

The Standby current levels provided by Power Gated SRAMs are significantly smaller than those without this feature. The dominant factor in this major difference is the switching on the Address bus. Figure 1 shows the large current peaks exhibited on a static RAM without Power Gating, along with the higher average DC value. Power Gated static RAMs are not affected by transitions on the Address bus.

Table 1. Comparison of Supply Current Specifications

SYMBOL	PARAMETER	MIN	MAX	UNITS
Static RAM With Power Gating				
I_{CC2}	Standby Current TTL Levels $\overline{CE} > V_{IH}$		5	mA
I_{CC3}	Standby Current CMOS Levels $\overline{CE} > V_{CC} - 0.2$ V		1	mA
Static RAM Without Power Gating				
I_{CC2}	Standby Current TTL Levels $\overline{CE} > V_{IH}$		30	mA
I_{CC3}	Standby Current CMOS Levels $\overline{CE} > V_{CC} - 0.2$ V		25	mA

Table 2. Comparison of Data Retention Current Specifications

SYMBOL	PARAMETER	MIN	MAX	UNITS
Static RAM With Power Gating				
I_{DR}	Data Retention Current $\overline{CE} > V_{CC} - 0.2$ V		100	mA
Static RAM Without Power Gating				
I_{DR}	Data Retention Current $\overline{CE} > V_{CC} - 0.2$ V All Other Inputs: $V_{IN} < 0.2$ V or $V_{IN} > V_{CC} - 0.2$ V		100	mA

SUMMARY

The use of Power Gating, primarily on the Address bus, reduces Standby current, eases the system implementation of Data Retention mode and reduces system noise due to the constant switching of the Address bus. With power and noise as two of the most troublesome system problems, the SHARP fast static RAMs with Power Gating provide tremendous benefit to the system designer encountering these problems. Table 3 describes some of the SHARP fast static RAMs that offer Power Gating.

Table 3. SHARP Fast Static RAMs With Power Gating

PART NUMBER	ORGANIZATION	SPEEDS
LH52252A	64K × 4	25, 35, 45 ns
LH52253	64K × 4 (OE)	15, 20, 25 ns
LH52258A	32K × 8	15, 20, 25 ns
LH521002	256K × 4	20, 25, 35 ns
LH521007	128K × 8	20, 25, 35 ns
LH521008	128K × 8	20, 25, 35 ns

1M STATIC RAM CHIP ENABLES

INTRODUCTION

There are pinout and functional differences among the various 128K × 8 SRAMs available in the market. This document is a description of these differences and why they exist.

How Do Fast and Slow SRAMs Differ?

Historically, SRAMs have been characterized as either Fast or Slow. SRAMs with access times less than 70 ns were considered Fast SRAMs. SRAMs slower than 70 ns were considered Slow SRAMs. This distinction is no longer accurate. Fast 128K × 8 SRAMs must now have access times of 35 ns or less, while Slow SRAMs now have access times as fast as 35 ns.

Fast SRAMs are optimized for performance. Since faster SRAMs sell at a premium, Fast SRAM manufacturers focus on increasing speed rather than minimizing die size or power. The circuit design techniques used for Fast SRAMs result in die which are larger and consume more power than Slow SRAM dies of the same configuration and density. Larger die size translates to increased cost.

Fast SRAMs are designed with the memory organized into several smaller arrays to reduce capacitive loading, permitting faster access. When Fast SRAMs are enabled, only the small array containing the desired word is active. This technique, among others, provides the very fast access times seen today. The multiplicity of arrays, and the added complexity of the address decoders needed to access them, increase die size and cost.

Slow SRAMs are designed with low cost and low power as the primary objectives. Applications which use Slow SRAMs are typically cost sensitive. Low power operation is also important, because many Slow SRAM applications are powered by batteries. Some systems also require the data in the SRAM remain valid while the supply voltage is reduced to 2-3 V.

Slow SRAMs have in the past been designed with all of the memory arranged as one large array, with addresses divided into Row and Column to select one word in the array. This organization allowed the use of simple and efficient address decoding schemes which minimize die size. When this type of Slow SRAM is enabled, the entire chip is active and consumes power.

As Slow SRAMs increase in size, the capacitance of a single large memory array also increases. Proper bit cell operation requires sufficient current to drive this array capacitance. In order to reduce current requirements to a reasonable level at 1 Megabit densities, the memory array must be divided into several smaller arrays, so that the entire chip is not enabled during every access.

This change in architecture, along with other circuit features, results in Slow SRAMs that are almost as fast as Fast SRAMs. Today, a Slow 128K × 8 SRAM provides access times of 45 ns, with some chips achieving 35 ns. 1 Megabit Slow SRAMs exhibit slightly lower active power and significantly lower standby power than Fast SRAMs of the same size and organization, with only a 10 to 15 ns difference in speed. The key functional and pinout difference between Slow and Fast 1 Megabit SRAMs lies in the way that the Data Retention mode is controlled.

Controlling Data Retention on Slow 128K × 8 SRAMs

Most slow 128K × 8 SRAMs have an E_2 signal on pin 30 that controls Data Retention mode. This signal 'Power Gates' all Address, Data-In and other Control signals, including $\overline{E}_1$. 'Power Gating' means that, internal to the device, input signals are forced to a known state, under any input conditions, when E_2 is LOW. This gating is done by using either a NAND or NOR gate, instead of a CMOS inverter at the input (see Figure 1).

On Slow 128K × 8 SRAMs, when E_2 is inactive, all other inputs are Don't Care. With the input NAND/NOR gates disabled under the control of the E_2 signal, the inputs may be held at any voltage level (even float) without DC current flowing from V_{CC} to V_{SS}. If the inputs were inverters (see Figure 2) rather than gates, significant current could flow from V_{CC} to V_{SS} if the inputs were allowed to float. Because of the extra level of gating introduced by E_2, it usually has the worst case access time.

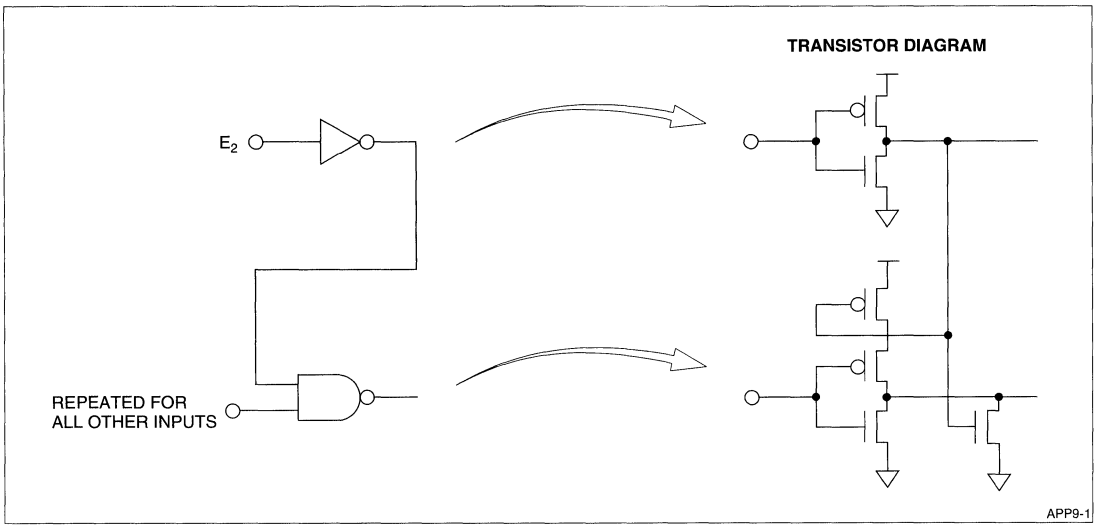


Figure 1. Power Gated Input Buffer

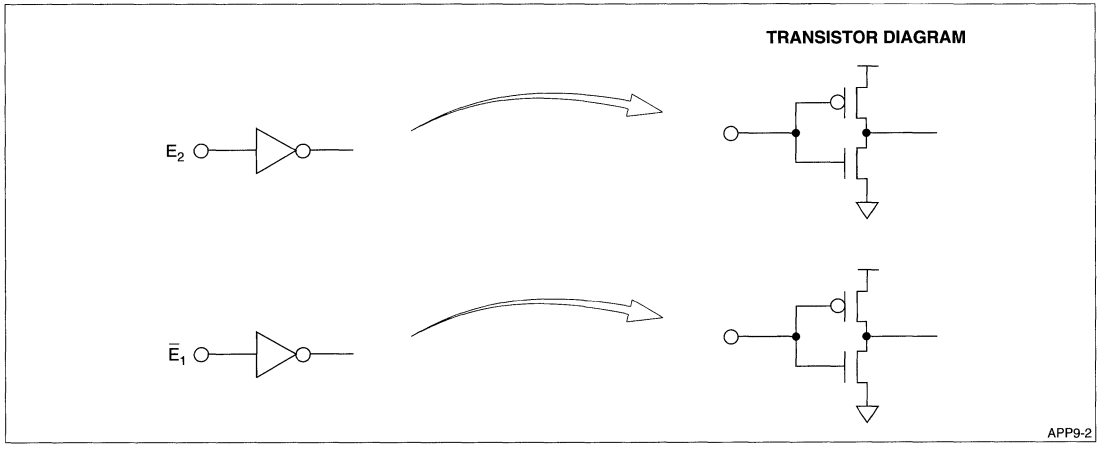


Figure 2. Non Power Gated Input Buffer

If the Chip Enable E_2 is active HIGH, Data Retention mode is entered by forcing E_2 LOW. This is very convenient, since a CMOS output driving a CMOS input (with no DC load) will result in a valid LOW level under any acceptable V_{CC} conditions. No matter what the V_{CC} is, ground is always zero.

If an active LOW Chip Enable is used, Data Retention mode is entered by asserting a HIGH level on E_2, where a logic HIGH is in the range of less than 0.5 V above V_{CC} and greater than 0.2 V below V_{CC}. This condition can be difficult to achieve when V_{CC} varies. Figure 3 displays the range of valid HIGH and LOW logic levels as a function of V_{CC}.

Slow SRAMs are available in 600-mil DIP, as well as SOP and other packages. The JEDEC standard for the 128K × 8 SRAM in 600-mil DIP is shown in Figure 4. Note the availability of E_2 as an option on Pin 30.

How Does E_2 Function on a Fast 128K × 8 SRAM (if the SRAM has E_2)?

Even though Fast SRAMs with E_2 may have the same pinout as Slow SRAMs, the E_2 signal does not serve the same function on the two devices. On Slow SRAMs, the E_2 signal is logically ANDed with the $\overline{E}_1$ signal and power gates the other inputs. On Fast SRAMs, the enables are logically separate and the

system must control both $\overline{E}_1$ and E_2 at all times. This places a difficult burden on systems using battery backup, since both an active HIGH Chip Enable and an active LOW Chip Enable must be controlled at different voltage levels. The EIA JEDEC standard for 128K × 8 SRAM in 400-mil SOJ (the package for Fast SRAM) does not allow E_2 for this reason. Since E_2 is required in some system applications that need more speed than a Slow SRAM can offer, there are two unique Fast SRAM product variations (in addition to the Slow SRAM). SHARP offers all versions, as shown in Table 1.

The most common package for Fast 128K × 8 SRAMs is the SOJ. The JEDEC standard for 128K × 8 SRAM in the SOJ package is shown in Figure 5.

NOTE: Pin 30 is No Connect (NC).

The $\overline{E}$ signal on many Fast SRAMs, especially those with an Output Enable ($\overline{OE}$ or $\overline{G}$), finds little use in many applications. The E_2 signal has been omitted from the SOJ JEDEC standard pinout. Some 128K × 8 Fast SRAMs (20-25 ns) will have the E_2 signal on Pin 30 to allow compatibility with slower SRAMs. The availability of the E_2 signal on non-standard Fast SRAMs permits their use in systems designed for the standard Slow pinout by adding a pull-up resistor on Pin 30. Additionally, some systems may take advantage of the E_2 signal to simplify address decoding.

NOTE: if there is no E_2, then $\overline{E}_1$ is simply $\overline{E}$.

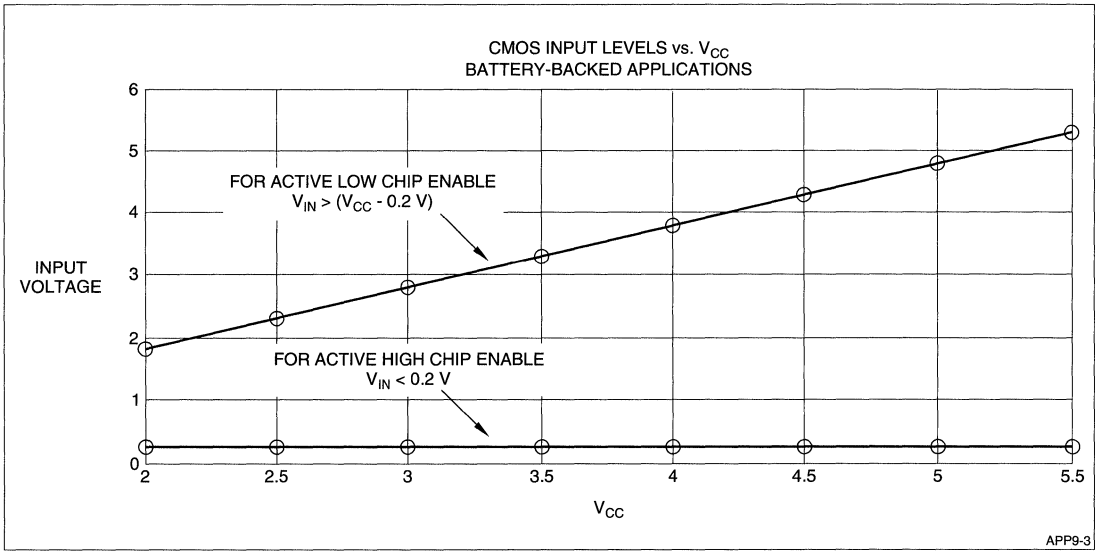


Figure 3. CMOS Input Levels vs. V_{CC} Battery-Backed Applications

Table 1. 128K × 8 SRAM Selector Guide

PART NUMBER	DESCRIPTION	FOOTPRINT	PIN 30	PIN 30 FUNCTION
LH521007	128K × 8	Figure 4	E_2	Active HIGH Chip Enable Both $\overline{E}_1$ and E_2 are controlled for Data Retention
LH521008	128K × 8	Figure 4	NC	$\overline{E}$ controls Data Retention Mode (All other inputs are Don't Care)

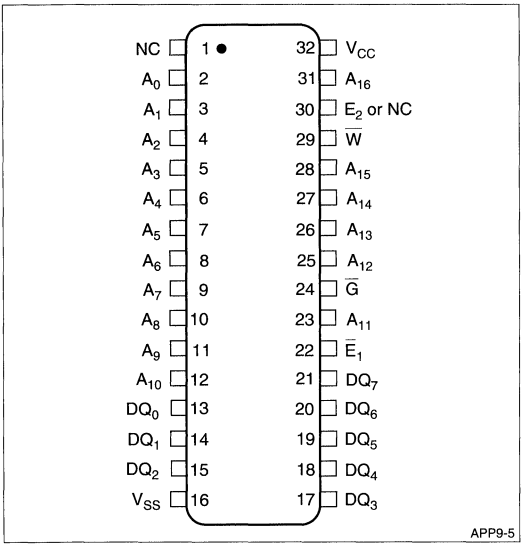


Figure 4. JEDEC Standard 600-mil DIP (Slow SRAM)

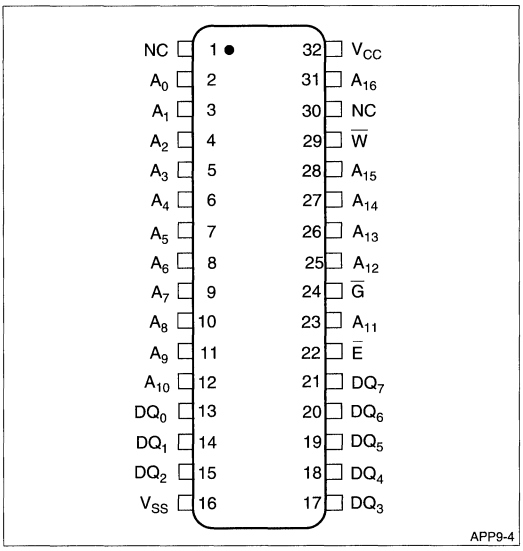


Figure 5. JEDEC Standard 400-mil SOJ (Fast SRAM)

CMOS Input Buffer Characteristics

A CMOS inverter has a DC path from V_{CC} to V_{SS}, as shown in Figure 2. When used as an input buffer, the inverter becomes linear when the input voltage is near the trip point (~1.5 V, nominal), permitting current to flow from V_{CC} to V_{SS}. The input itself, of course, draws no DC current. The supply current from V_{CC} to V_{SS} through the inverter is significant when the input is at the trip point, but becomes small when the input is near the supply voltages. See Figure 6.

To minimize supply current at the inputs, Slow SRAMs use a Power Gated input buffer, which has no DC path when the buffer is disabled. Even when the address bus is cycling, no DC current will flow through the Slow SRAM's input buffers (except for the active HIGH Chip Enable, which cannot control itself). The reduction in standby current is obtained at the expense of increased access time, due to the extra level of gating.

Fast SRAMs cannot tolerate any penalty on access time, so they typically use inverters as input buffers. Even when the chip is not logically enabled, the input buffers can consume several milliamps of standby current. As input signals pass through the inverter's trip point, a current flows from V_{CC} to V_{SS}.

How to Determine E_2's Function

The exact function of a 1M SRAM's E_2 may be determined by a careful analysis of the vendor's data sheet. If the Data Retention Current with V_{CC} at 3.0 V is guaranteed to have a low value when only E_2 is held LOW (usually at 0.2 V or less), then E_2 controls (Power Gates) all other inputs. Otherwise, if both $\overline{E}_1$ and E_2 must meet restrictions in order to guarantee the Data Retention current, one must assume that E_2 does not gate $\overline{E}_1$ and that both Chip Enables must be controlled during Data Retention.

SUMMARY

There are three distinct categories of 128K × 8 SRAMs. Sharp's product line includes all three variations. Each product serves a particular need in the market, with all needs matched by one of the available products.

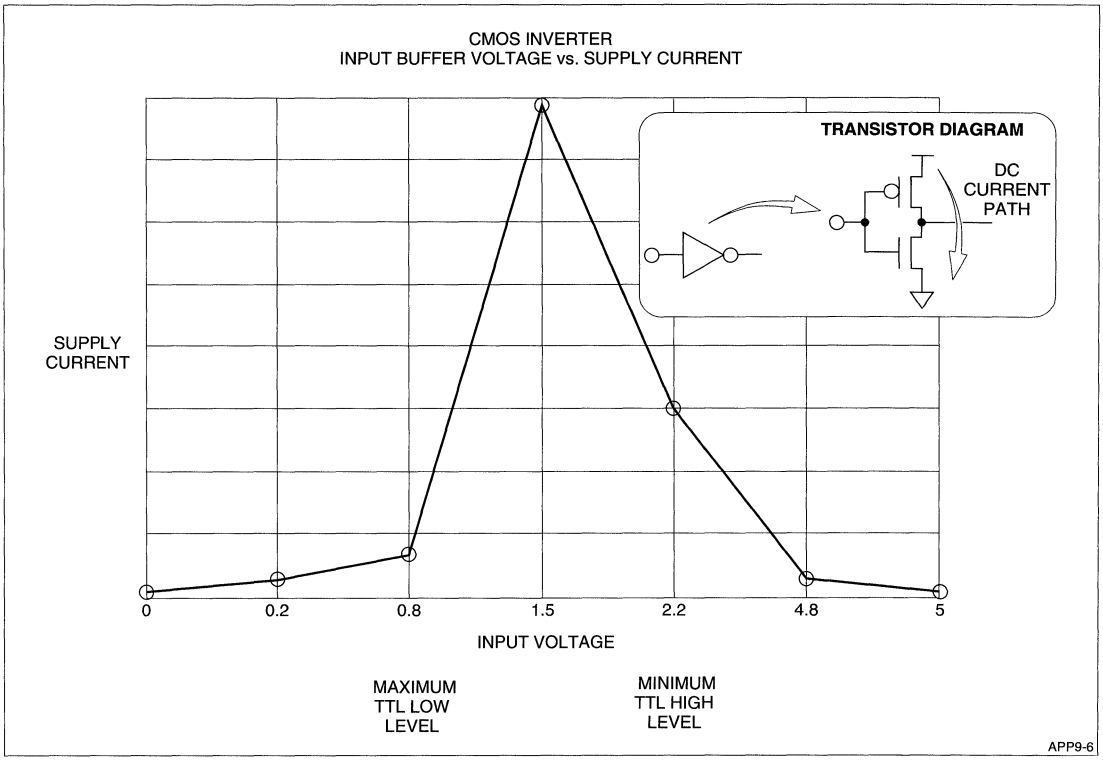


Figure 6. CMOS Inverter
Input Buffer Voltage vs. Supply Current

LH521028
CMOS 64K × 18 Fast Static RAM

WIDE-WORD STATIC RAM

INTRODUCTION

The Sharp LH521028 is a JEDEC Standard 64K × 18 CMOS static RAM. This device offers a wide word, very high density memory in a compact package. A total of 1,179,648 bits (1.125M) are produced in a 52-pin Plastic Leaded Chip Carrier (PLCC). The wide word and features of the LH521028 offer superior performance, while minimizing the required component count in many system applications. The large number of power supply terminals support the needs of the 18-bit data bus (see Figure 1).

Chip Architecture

The Block Diagram for the LH521028 is shown in Figure 2. Figure 3 contains an equivalent circuit based on older technology components. The LH521028 is functionally equivalent to these eight components:

DEVICE	QUANTITY	NOTE
64K × 4 SRAM	4	16 data bits
64K × 1 SRAM	2	For parity
LS74AS373 Latch	2	16-bit address

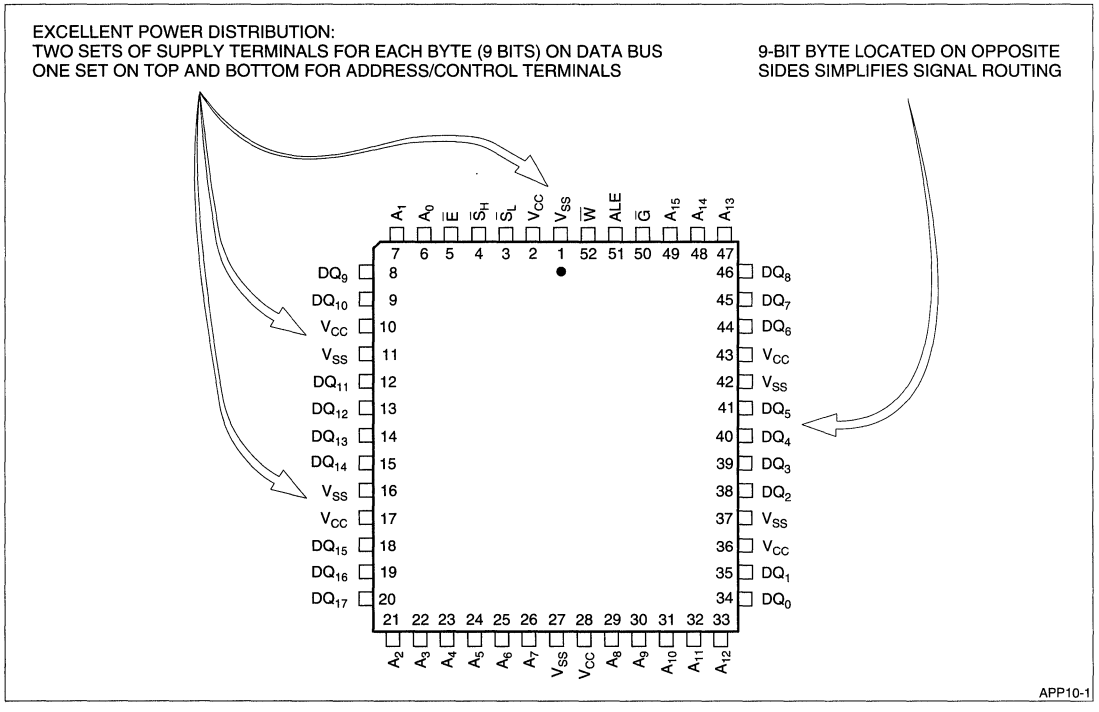


Figure 1. 64K × 18 Footprint

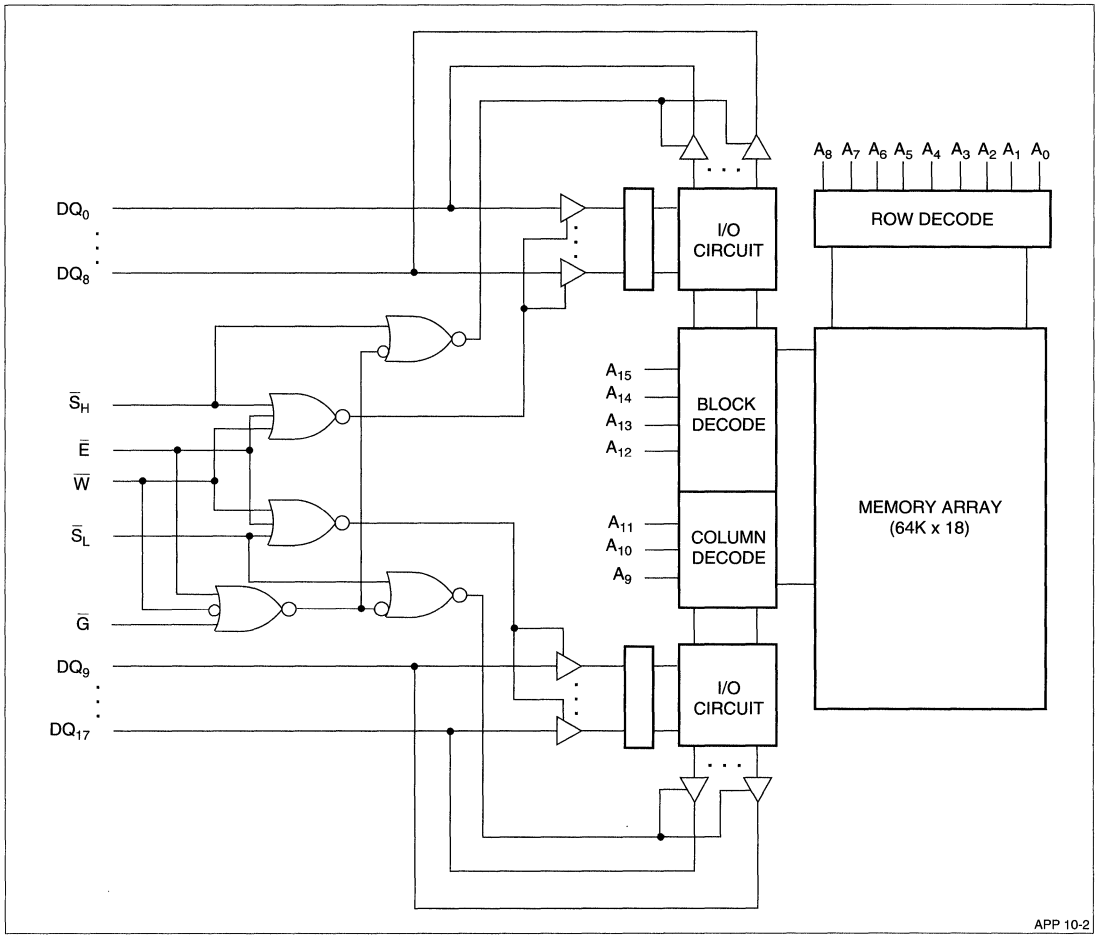


Figure 2. LH521028 Block Diagram

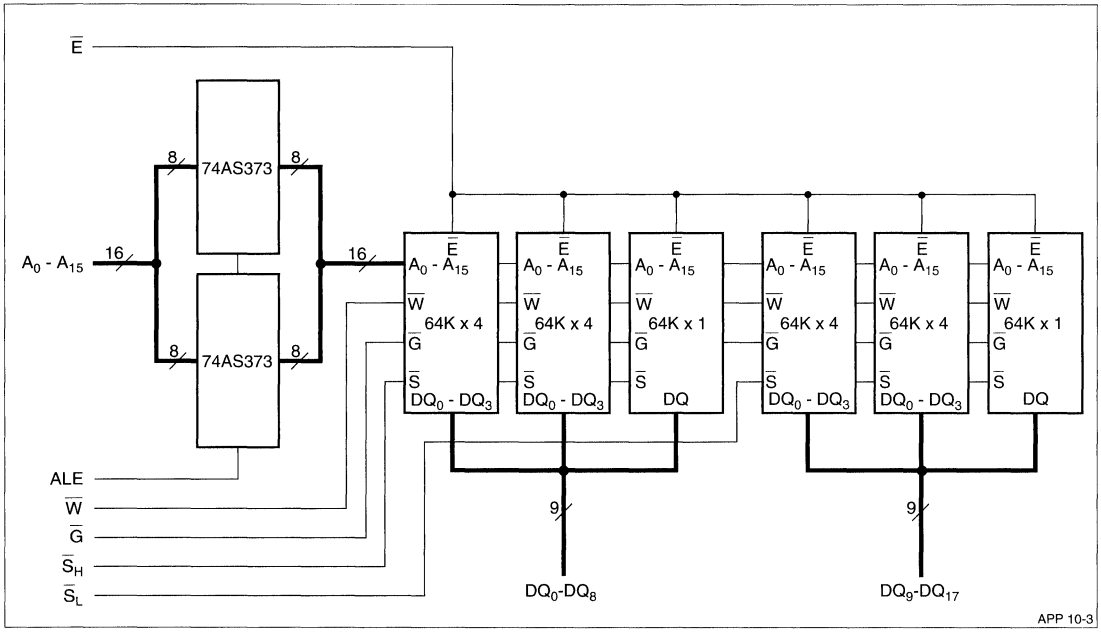


Figure 3. 64K × 18 Equivalent Circuit

Control Functions

The High and Low Byte Select controls ($\overline{S}_H$,$\overline{S}_L$) serve only to control Read and Write operations and do not place the chip in Standby mode when inactive. For Write cycles, these controls provide individual Byte-write capability, typically leaving the critical timing to the single Write Enable control signal. During Read cycles, the Byte Select controls provide a very fast access time. The Address Latch Enable operates independently of the other controls. The state of the Address bus is always latched when ALE transitions HIGH to LOW, even if Chip Enable is HIGH. ALE tied HIGH allows the Address bus to remain transparent, for those applications not requiring an address latch. The Chip, Write, and Output Enable controls ($\overline{E}$, $\overline{W}$, $\overline{G}$) function in a manner similar to those on conventional static RAMs.

Applications Example – TMS320C30 Memory

A Texas Instruments TMS320C30 Digital Signal Processor is shown in Figure 4, along with two LH521032s attached to the Primary bus. In this example, the memory is configured to occupy the lowest 64K words of the address space (partially decoded). The connections are very simple. The R/$\overline{W}$ control is inverted to control the SRAM's Output Enable signal. The SRAM's Write Enable terminal is directly connected to R/$\overline{W}$. The most significant address, A_{23}, is connected directly to both High and Low Byte Select terminals. The SRAM's Chip Enable is tied to $\overline{STRB}$. In this example, only a single Inverter gate is required in the memory interface. Note that the two unused Data pins are tied with a pull-up resistor to V_{CC}. This prevents the signals from floating during Write cycles. Also, the Address latch is always transparent, with ALE tied HIGH.

Static RAM Worst Case Timing Calculations – Read Cycle

SRAM Address Access Time	$t_{AA} \le t_{10}$ (min) – t_{14} (max) – t_{15} (min) ≤ 60 – 10 – 15 ≤ 35 ns
SRAM Chip Enable Access Time	$t_{EA} \le t_{10}$ (min) – t_{11} (max) – t_{15} (min) ≤ 60 – 10 – 15 ≤ 35 ns
SRAM Byte Select Access Time	$t_{SA} \le t_{10}$ (min) – t_{14} (max) – t_{15} (min) ≤ 60 – 10 – 15 ≤ 35 ns
Output Enable Access Time	$t_{GA} \le t_{10}$ (min) + t_6 (min) + t_8 (min) – t_1 (max) – Inverter t_{PD} (max) ≤ 60 + 1 + 23 – 10 – 8 ≤ 66 ns

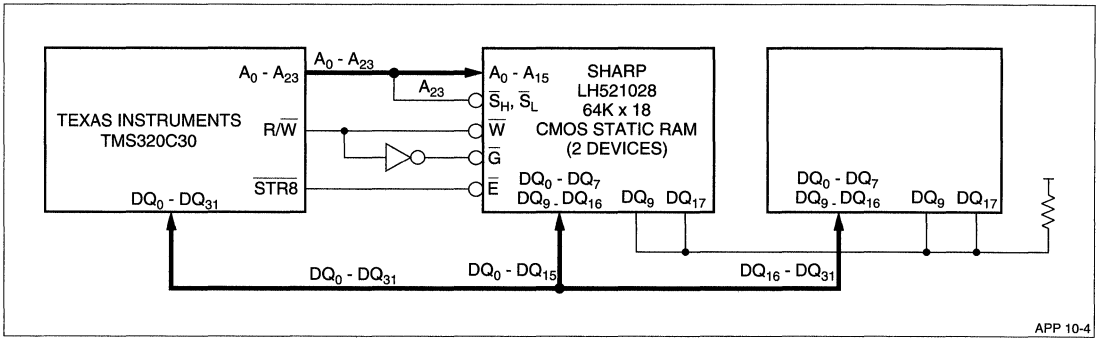


Figure 4. TMS320C30 With 64K Word Memory on Primary Bus

Static RAM Worst Case Timing Calculations – Write Cycle

Address Setup to Beginning of Write	$t_{AS} \leq (t_7\ (min) + t_9\ (min) + t_{13}\ (min)) - t_{14}\ (max) \leq (24 + 2 + 0) - 10 \leq 16$ ns
Address Hold From End of Write	$t_{AH} \leq t_{10}\ (min) - t_{14}\ (max) + t_{11}\ (min) \leq 60 - 10 + 0 \leq 50$ ns
Write Pulse Width	$t_{WP} \leq t_{10}\ (min) - t_{11}\ (max) + t_{12}\ (min) \leq 60 - 10 + 0 \leq 50$ ns
Data Setup to End of Write	$t_{DS} \leq t_{10}\ (min) + t_{12}\ (min) - t_{20}\ (max) \leq 60 + 0 - 20 \leq 40$ ns
Data Hold from End of Write	$t_{DH} \leq t_7\ (min) + t_9\ (min) + t_{14}\ (min) - t_{12}\ (max) \leq 24 + 2 + 0 - 10 \leq 16$ ns

SUMMARY

The LH521028 is an ideal solution for many system applications. The density, speed and features serve to reduce system cost while providing increased performance, while significantly reducing the component count and Printed Circuit Board area.

REFERENCES

[1] *Third-Generation TMS320 User's Guide,* Texas Instruments, 1988.

[2] *LH521028 CMOS 64K × 18 Static RAM Data Sheet,* Sharp Corp., 1992.

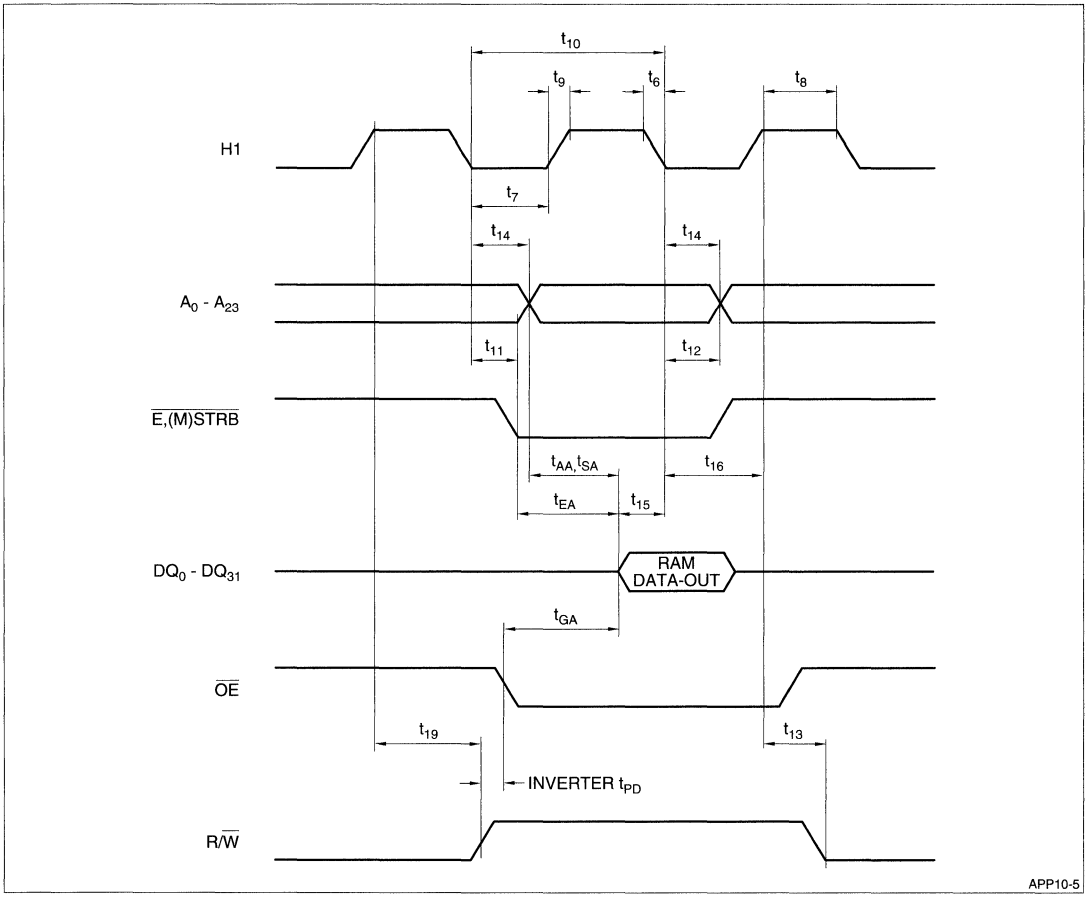


Figure 5. Read Cycle

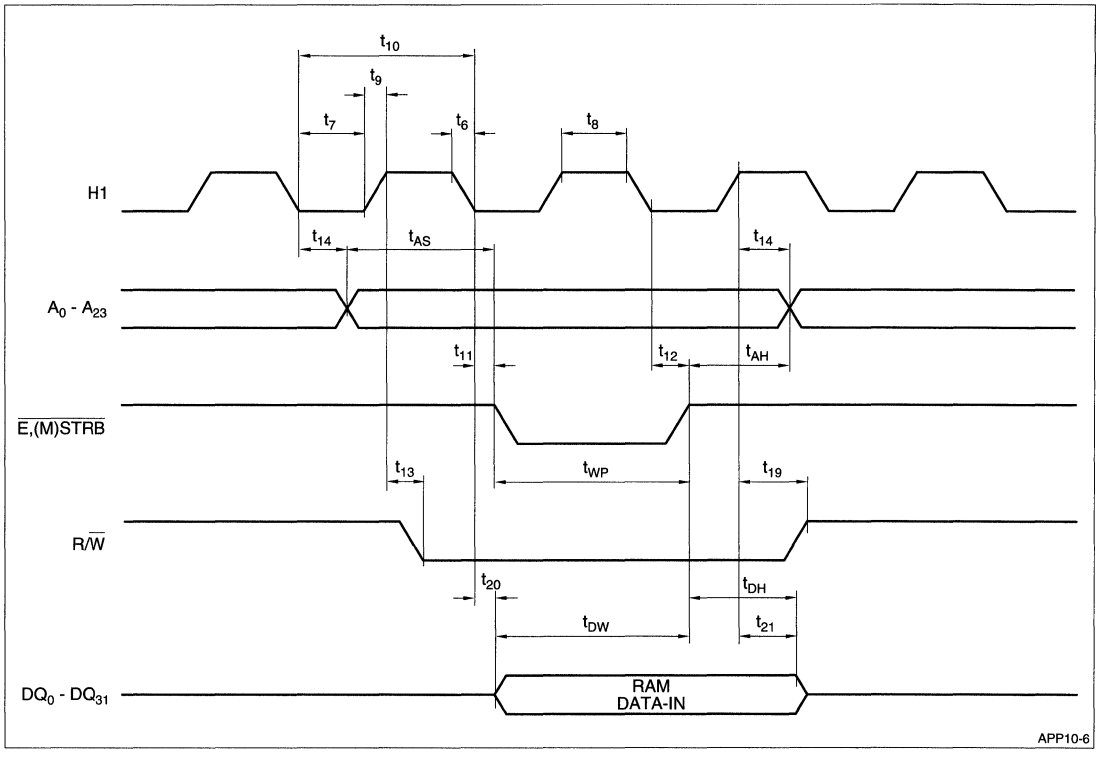


Figure 6. Write Cycle

DC PARAMETERS

INTRODUCTION

CMOS Fast Static RAMs have several parameters that specify DC characteristics. Included are the various Supply Currents, Input and Output Leakage Currents, along with Input and Output Voltage and Current levels. These apparently simple DC specifications are often a source of confusion. This application note attempts to clarify Fast Static RAM DC parameters.

Parameter Categories

CMOS Fast Static RAM DC specifications fall into one of two categories. These categories are DC Operating Conditions and DC Electrical Characteristics. DC Operating Conditions describe those conditions that the system must provide to the static RAM in order for the memory to operate correctly. DC Operating Characteristics describe the behavior of the static RAM.

The DC Operating Conditions include Input Voltage levels (V_{IH} for valid HIGH and V_{IL} for LOW levels) and the Supply Voltage required for proper operation. DC Characteristics include the HIGH and LOW Level Output Voltages, Supply Currents, along with Input and Output Leakage Current values.

The following sections describe the individual specifications.

DC Operating Conditions

DC Supply Voltage

The Supply Voltage is the level required on the V_{CC} terminal(s), reference to V_{SS} (Ground), for proper operation. CMOS Fast Static RAMs typically have Supply Voltage specifications at 4.5 to 5.5 V, typically 5.0 V. This range is often referred to as a 10% supply, where the supply voltage tolerance is $\pm$10%. The supply voltage should never exceed the maximum allowable, otherwise damage can occur to the static RAM. Below 4.5 V, the static RAM may not function in the desired manner. The exception to the minimum Supply Voltage specification is Data Retention mode, which is offered on some static RAMs (this is discussed later).

In the future, the standard Supply Voltage will move to 3.3 V, beginning at 16 Megabits for static RAMs and 64 Megabits for Dynamic RAMs. As transistors shrink in order to fit ever increasing numbers on a chip, they become more sensitive to voltage. Reducing the power supply voltage on future products will increase their reliability, without the need for on-chip voltage regulator circuits. In addition, some static RAMs have separate V_{CC} supply terminals for Output Drivers that drive low impedance loads, i.e. 50 Ω. These terminals often require 3.3 V, in order to reduce power dissipation and provide for better impedance matching. This feature will become more common as static RAMs migrate to the newer configurations with power supply terminals in the center of the package, as opposed to the ends. This new standard provides two sets of V_{CC} and V_{SS} terminals, providing a significant performance advantage.

DC Input Voltage

For CMOS Fast Static RAMs, the DC Input Voltage is tested with the device operating at a very slow cycle time (~10 MHz). The Input Voltage measurements simulate static, or DC operation. This is in contrast to the AC Test Conditions explained in detail in Table 1 and Figure 1. The AC Test Conditions are the conditions where the static RAM is tested at the fastest cycle time. The AC Test Conditions are explained in some detail, in order to provide a better understanding of the DC Specifications.

When operated at their fastest cycle times, CMOS Fast Static RAMs are both specified and tested with wider signal swings, along with specific input signal rise and fall times. The timing reference point is 1.5 V for both input and output signals, which is the same as for standard Bipolar TTL devices. The input signals for most Fast Static RAMs are specified to transition between 0.0 and 3.0 V, with a transition time of 5 ns. Some *very fast* static RAMs, with access times less than 15 ns, are specified to operate with rise and fall times from 2 to 3 ns. Operation with smaller input signal transitions or slower transitions times will result in slower operation of the static RAM. Typical AC Test Conditions are shown in Table 1.

AC Test Conditions

CMOS Fast Static RAMs are tested for speed at the following input levels and transition times, not the DC input levels. Also, input and output timing measurements are made at 1.5 V, not $V_{IL}/V_{IH} - V_{OL}/V_{OH}$, for compatibility with standard Bipolar TTL devices.

Table 1. AC Test Conditions

PARAMETER	RATING
Input Signal Rise/Fall Time	5 ns
Input Signal Transitions	0.0 V to 3.0 V
Input/Output Reference Voltage	1.5 V

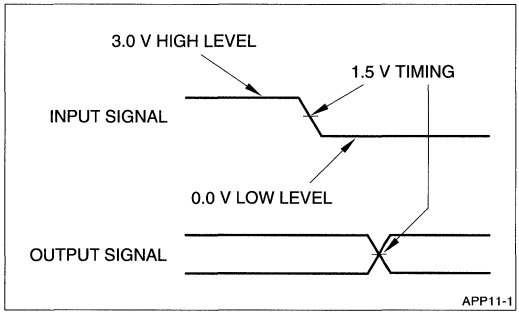


Figure 1. Signal Waveforms

Input HIGH Voltage – V_{IH}

The V_{IH} specification describes the minimum input voltage level that the chip accepts as a valid HIGH level, under static or DC conditions. CMOS Fast Static RAMs require a V_{IH} level of 2.2 V in most cases. Fast Static RAMs are tested operating at the their rated cycle time with the HIGH voltage level at 3.0 V, as seen in the previous section.

CMOS Fast Static RAMs respond faster, inducing more noise, at higher Supply Voltage and colder temperatures. If a Fast Static RAM has an inadequate power supply, the minimum V_{IH} specification is the most sensitive parameter to these conditions.

The maximum V_{IH} level is also important. Depending on the design of the chip, the maximum V_{IH} level is either a diode drop (0.3 or 0.5 V) above V_{CC} or 6.0 V. The reason for the difference is important and related to the Electrostatic Discharge (ESD) protection circuit and Data Retention mode. Systems that use Data Retention mode, or Battery-Backed Operation, in some cases do not provide controlled voltage levels on pins other than the Chip Enable pin. When the system returns to the standard 5 V supply after a period of operation at the reduced battery voltage, the static RAM's other input pins can power-up in the HIGH state. If this happens, the input pin can be driven to a higher voltage than the V_{IH} maximum value, in violation of the specification. This situation results in a current flow from the Input (or Output) pin into the RAM. The Battery usually has a series of switching diodes, used to switch between the Battery and line supply, but this situation can still reduce the static RAM's reliability or disturb the static RAM's operation.

Often, static RAMs intended for operation in Data Retention mode are designed with ESD networks that do not have a reversed diode to V_{CC} on the input pins. The V_{IH} on these input pins can go above the diode drop above V_{CC} and not damage the chip. The Input/Output pins, such as the Data Bus, often have P-channel MOS transistors that have reversed diodes to V_{CC}. N-channel MOS transistors are also used on some chips, usually with a small P-channel transistor, known as a 'Helper,' in parallel. The P-channel usually has enough series resistance to eliminate any problems due to the current flow to V_{CC}. Often, the structure of the Output Buffer eliminates the need for an additional ESD Protection circuit.

Figure 2 shows an example of the circuitry on input pins, with and without the diode in the ESD Protection Circuit connected to V_{CC}. Figure 3 shows an example of the two most common output circuits found in CMOS Fast Static RAMs. The first is a 'Full CMOS' type; the second is a N-channel output, with a P-channel 'Helper' for the pull-up.

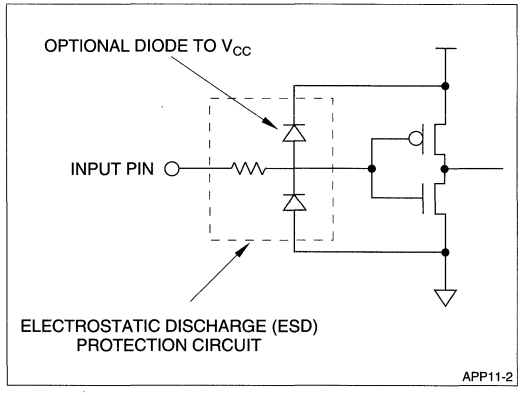


Figure 2. CMOS Input Circuit

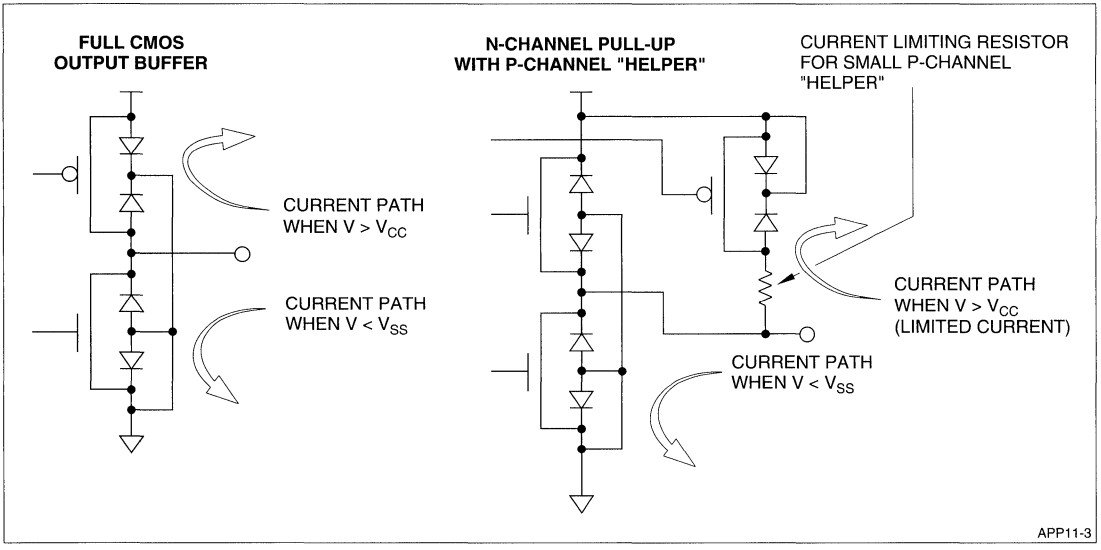


Figure 3. Simplified CMOS Static RAM Output Circuit

Input LOW Voltage – V_{IL}

The V_{IL} specification describes the maximum input voltage level that the chip accepts as a valid LOW level. CMOS Fast Static RAMs require a V_{IL} level of 0.8 V in most cases. Most Fast Static RAMs are tested operating at the fastest cycle time with the LOW voltage level at 0.0 V.

CMOS static RAMs have reversed diodes to V_{SS} on their input pins, as in Figure 2. If the input voltage drops sufficiently below V_{SS}, then current flows from the pin to ground. Systems designed without proper signal termination, especially on the Address and Control pins, often have ringing. The diode structure may clamp the voltage, but current will flow through the input pin to V_{SS}. This additional ground noise may cause on-chip ground bounce, effecting system behavior.

DC Electrical Characteristics

Average Operating Supply Current – I_{CC}

The I_{CC} specification is the average DC current consumed by the static RAM. In the Test Conditions section, I_{CC} is measured at the maximum Address cycle time for the device. The Supply Voltage and Ambient Temperature for the worst case I_{CC} are usually 5.5 V and 0°C. Since most systems typically operate at slower than the fastest cycle time, higher temperature and lower voltages, the actual I_{CC} is somewhat lower than under the worst case conditions.

TTL Standby Current – I_{SB1}

Sharp uses I_{SB1} to describe the TTL Standby Current for CMOS Fast Static RAMs. I_{SB1} describes the supply current consumed by the static RAM when the Chip Enable terminal is HIGH (Standby mode). With most Sharp static RAMs, using Address Power Gating, the only restriction in the Test Conditions section is that the voltage on the Chip Enable terminal is at or above the minimum V_{IH} level, which is Standby mode. For all static RAMs, the other input signals may toggle at any rate up to the fastest cycle time. Static RAMs without Power Gating have a much higher I_{SB1} specification. This is due to the current flowing from V_{CC} to V_{SS} in the Address input buffers as they toggle.

CMOS Standby Current – I_{SB2}

Sharp uses I_{SB2} to describe CMOS Standby Current for Fast Static RAMs. I_{SB2} is very similar to I_{SB1}, with the exception that CMOS input voltage levels are required. CMOS input voltage levels are less than 0.2 V for a LOW level and within 0.2 V of V_{CC} for a HIGH level. On Sharp static RAMs with Address Power Gating, only the Chip Enable input must be held at the valid HIGH level, alleviating the need to control the other inputs. As with I_{SB1}, the Sharp static RAMs with Address Power Gating exhibit a significantly lower I_{SB2} level due to the toggling of the Address bus. Static RAMs without Power Gating have CMOS Standby Current levels only marginally lower than the TTL Standby Current, once again due to the current flowing though the Address buffers as the Address toggles.

Input Leakage Current – I_{LI}

I_{LI} is the Input Leakage Current. CMOS inputs are connected to logic gates, either Inverters or NOR gates. The inputs to these circuits have VERY high impedance, so the current required to drive the inputs at any DC voltage from V_{CC} to Ground consists only of leakage current. The specification values range from 2 to 10 uA. Device terminals damaged by ESD will have Input Leakage Currents in excess of the specification.

Output Leakage Current – I_{LO}

I_{LO} is the Output Leakage Current. Output Leakage Current is similar to Input Leakage Current, except that it describes the leakage current through an output buffer when the buffer is in the high impedance state. I_{OL} is tested with either Chip Enable or Output Enable (or both) at a logic HIGH level.

LOW Level Output Current – I_{OL}

SHARP uses I_{OL} to define LOW level output current. I_{OL} is guaranteed at a given Output Voltage (V_{OL}). This parameter can be tested with DC values, by forcing the current (I_{OL}) and measuring the voltage (V_{OL}), or a dynamic test is performed. The dynamic test is used to drive a known load circuit (a resistor in series to a supply voltage) to the rated voltage level. In order to drive the known load to the specified voltage (in this case V_{OL}), the minimum I_{OL} level is required.

The LOW Level Output Current flows from the output pin to V_{SS} through the N-Channel transistor, as shown in Figure 3.

HIGH Level Output Current

The HIGH Level Output Current is similar to I_{OL}, except that the Voltage referenced is V_{OH}. The current flows from the static RAM's V_{CC} terminal through one of the circuits shown in Figure 3, and then out the Output terminal.

LOW Level Output Voltage – V_{OL}

V_{OL} is the maximum output voltage, when the chip is sinking I_{OL} to V_{SS}. See the discussion on I_{OL}.

HIGH Level Output Voltage – V_{OH}

V_{OH} is the minimum output voltage, when the chip is sourcing I_{OH} from V_{CC}. See the discussion on I_{OH}.

Data Retention Voltage – V_{DR}

Those static RAMs with a Data Retention mode use V_{DR} to describe the lowest supply voltage where valid data is maintained, or Data Retention Voltage. This Supply Voltage is normally obtained from a battery through an isolation diode. The minimum voltage is 2.0 V. In addition to the Supply Voltage, other conditions must be met in order to ensure proper operation during Data Retention mode.

Data Retention Current – I_{DR}

The Supply Current required during Data Retention mode is called I_{DR}, or Data Retention Current, and applies only to those static RAMs with Data Retention mode. The Data Retention Current is the maximum Supply Current required during Data Retention mode. For high-density Fast Static RAMs, the worst case values of I_{DR} range from about 1 to 100 uA. Most Fast Static RAMs use memory cells with four transistors and two load resistors for the cross-coupled inverters. The major factor in a static RAMs' I_{DR} is the current through one of these resistors from V_{CC} to V_{SS}. The resistance varies significantly with temperature, so the typical I_{DR} value at room temperature is about an order of magnitude lower than the worst case.

SUMMARY

This application note describes CMOS Fast SRAM DC parameters. The Test Conditions and Notes for these parameters often provide a useful insight into the characteristics and behavior of the static RAM.

INPUT/OUTPUT LEVEL TESTING

INTRODUCTION

This application note describes the test methodology used to test Input Voltage levels (V_{IL} and V_{IH}) on CMOS Fast Static RAMs. The high speed nature of these devices result in very large transient currents on the power supply terminals. The management of these transient currents is a critical factor, both in the test and system application of CMOS Fast Static RAMs. Included in this application note are a description of CMOS Fast Static RAM input and output circuits, followed by a description of the methodology used in the testing of input and output levels.

CMOS Input Buffer

Figure 1 contains the two most common circuits used by CMOS static RAMs as the first stage input buffer. On the left is a simple Inverter-type input buffer, while a NOR gate is on the right. The first stage input buffer performs level conversion from the TTL levels used externally (LOW ≤ 0.8 V, HIGH ≥ 2.2 V) to the CMOS levels used internally (LOW ≤ $\frac{V_{CC}}{2}$, HIGH ≥ $\frac{V_{CC}}{2}$). The trip point for the first stage input buffer is set near the TTL level midpoint, $\frac{V_{IH} - V_{IL}}{2}$ = 1.5 V, under nominal conditions (T_A = 25°C, V_{CC} = 5.0 V and the semiconductor manufacturer's 'Typical' process). The trip point is the point where the resistance of the pull-up transistor is equal to that through the pull-down transistor in the input buffer, so that the output voltage becomes $\frac{V_{CC}}{2}$.

To achieve a 1.5 V nominal trip point on the Inverter-type input buffer, the P-channel transistor is designed with a much higher On-resistance than that used in a standard CMOS Inverter. Variations in the test conditions will cause the DC trip point to change. Excluding noise, the DC trip point will vary over all conditions by about 20%, from about 1.2 V to about 1.8 V. When noise is added to the equation, primarily originating from the $-L\frac{di}{dt}$ component generated by the switching of output buffers, the noise margin for the DC input levels is reduced.

CMOS Output Buffer

Examples of CMOS output buffers are shown in Figure 2. On the left, there is a Totem Pole output driver constructed from an N-channel pull-down transistor and a P-channel pull-up transistor. The circuit on the right uses both P and N-channel transistors as pull-up devices. The circuit on the right is used in some cases where it is desirable to limit V_{OH}, with a DC load on the output signal.

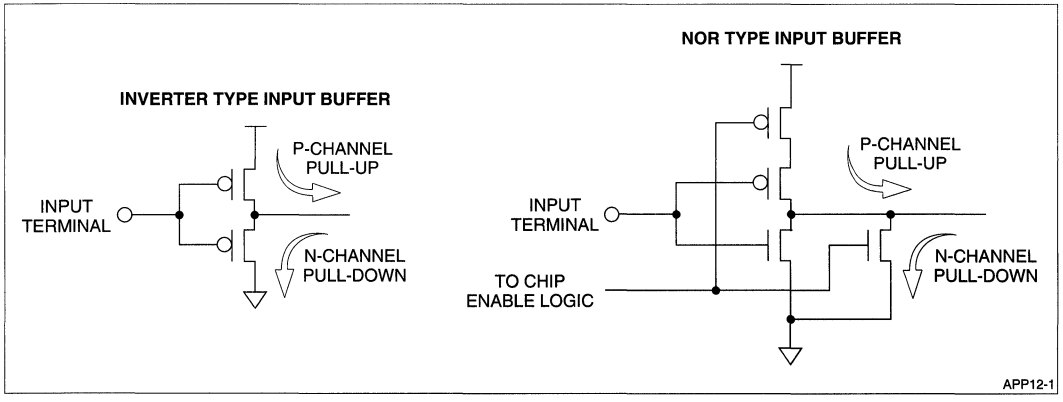


Figure 1. Inverter and NOR Type Input Buffers

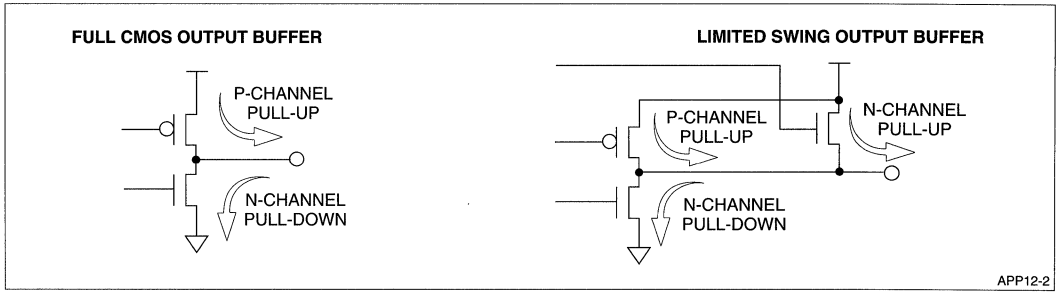


Figure 2. Typical CMOS Static RAM Output Buffers

The output buffer will rapidly switch large amounts of current, so the value of $\frac{di}{dt}$ is large. If the inductance of the power supply path of the test fixture is excessive, then the switching noise ($-L\frac{di}{dt}$) can make a device appear to have difficulty achieving the DC V_{IL} and V_{IH} specifications.

Model of a CMOS Fast Static RAM in a Test Fixture

Figure 3 shows a CMOS Fast Static RAM residing in a test fixture. The static RAM is simplified to show only the input buffer, output buffer and core areas. The plastic package lead inductance is shown, along with the lead inductance added by the test fixture socket used in the test fixture. The package inductance is always present, but the additional inductance added by the socket is eliminated when the chip is mounted on a printed circuit board. Also, the static RAM's input signals in the test environment are open circuits at the end of transmission lines.

Testing AC Input/Output Levels

In order to overcome the issues introduced in the harsh test environment, CMOS Fast Static RAMs are tested for speed with input signals transitioning between levels relaxed from the DC input levels. CMOS Fast Static RAMs on the market today include a table called 'AC Test Conditions,' which describes the conditions used for AC testing.

Example of an AC Test Conditions Table

PARAMETER	TEST CONDITION
Input Pulse Levels	0.0 to 3.0 V
Input and Output Timing Reference Levels	1.5 V
Input Signal Rise and Fall Time	5 ns

The first two entries in the table indicates that input signals swing from 0.0 to 3.0 V during all AC Timing tests. The second condition indicates that AC Timing measurements are based on the time when the signal of interest, either input or output, crosses the 1.5 V level. The third entry is that all input signals have 5 ns rise and fall times. The conditions used to test the DC specifications, input and output levels, differ significantly from those used for AC specifications.

Testing DC Input/Output Levels

A CMOS Fast Static RAM's DC input and output voltage levels are typically tested dynamically, although at slower than minimum cycle times. Figure 4 shows two such timing sets used by Sharp in the testing of DC input and output levels. Note that the inputs transition from the DC V_{IL} and V_{IH} specifications, while outputs are required to meet DC V_{OL} and V_{OH} specifications across the AC test load, thus ensuring adequate I_{OL} and I_{OH} current levels. Figure 5 contains an example of an AC Test Load. The resistor values are selected to ensure that the I_{OH} and I_{OL} specifications are met when the V_{OL} and V_{OH} values are tested.

SUMMARY

CMOS Fast Static RAM DC and AC input levels have precise definitions. When testing CMOS Fast Static RAMs, it is very important to understand the definition of these parameters. Also, the test environment used for testing these parameters must be conducive to high speed testing.

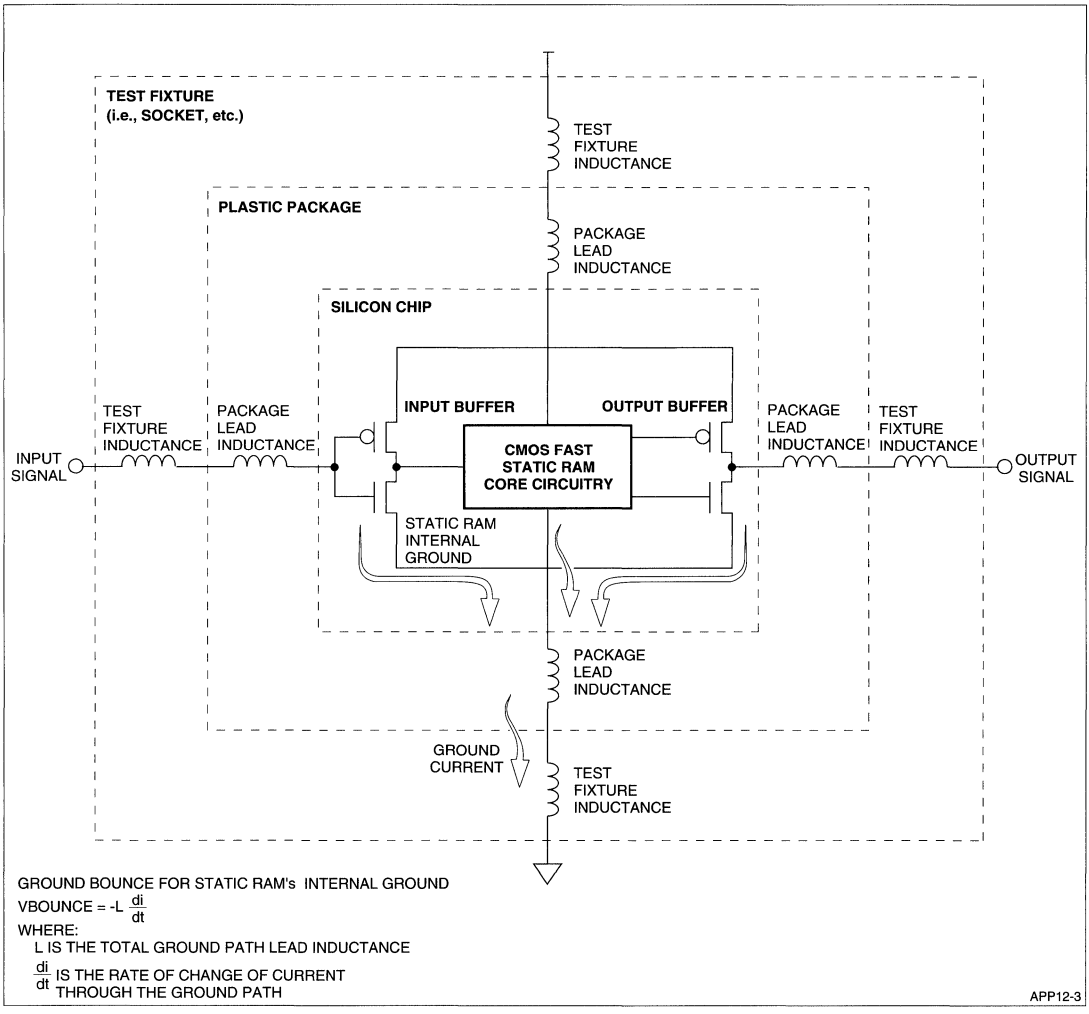


Figure 3. CMOS Fast Static RAM in a Test Fixture

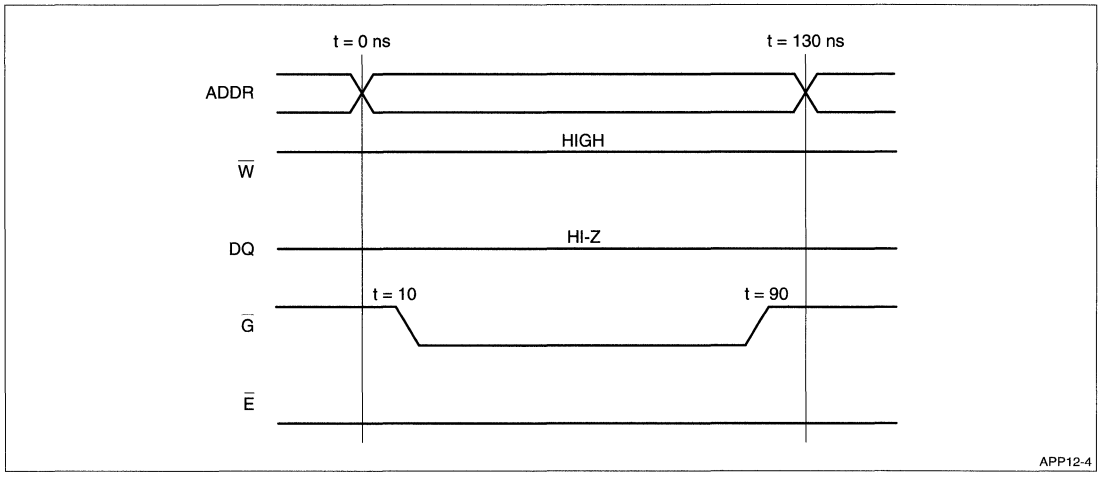


Figure 4. Timing Used for DC Input/Output Level Testing

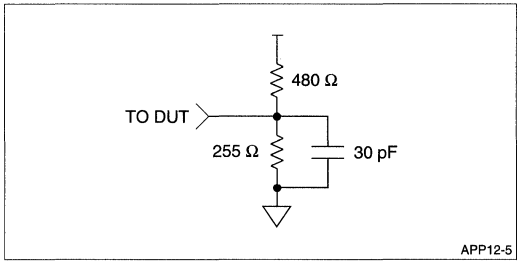


Figure 5. AC Test Load
(Also Used for DC Output Levels)

THE 386SL CPU

INTRODUCTION

There are several new members of the 386/486 microprocessor family, with product announcements by Intel, AMD, Chips and Technology, and probably others in the near future. The potential market for these products is enormous. There are numerous companies manufacturing products based on these chips, with most system manufacturers attempting to differentiate their product lines. Intel has created a controversial new chip, the 386SL and the companion 82360SL ISA I/O peripheral device, to service the high-performance, low-power notebook computer market. the list of features offered by the 386SL chip set includes power management, integrated DRAM controller, serial/parallel interfaces, most of the other peripheral functions required in a PC, along with a cache controller and TAG RAM. With this circuitry, the 386SL comes at a higher price than the standard notebook 32-bit CPU, the 386SX.

The 386SL's optional static RAM cache results in very attractive performance benchmark results, but the results of battery life tests can vary substantially, depending on the test. With constant use, there is little or no improvement in battery life with the 386SL. With intermittent use, the 386SL can provide a substantial savings in battery life. The current 386SL operates with a 5.0 V supply, but a 3.3 V version is rumored. As the debate on the relative merits of the 386SL continues, we will examine the 386SL's cache data RAM interface in this application note.

The 386SL's Cache

The 386SL has an optional port that directly connects with up to 64K bytes of standard asynchronous static RAM, serving as cache data RAM. The cache data RAM can be configured with 0K, 16K, 32K, or 64K bytes, with direct mapped, two- or four-way associativity. The cache configuration is entirely controlled by software, except, of course, for the size of the static RAM.

Given the four available cache sizes, there are several possible cache data RAM configurations, as shown in Table 1.

Table 1. 386SL Cache Configurations

CACHE SIZE	SRAM CONFIGURATION
0K Bytes	N/A
16K Bytes	2 – 8K × 8 SRAMs
32K Bytes	4 – 16K × 4 SRAMs 1 – 16K × 16 SRAM
64K Bytes	2 – 32K × 8 SRAMs 2 – 16K × 16 SRAMs

The 386SL has a very unique cache data RAM interface. The 386SL cache data RAM port connects directly to standard, asynchronous static RAMs without the need for any additional glue logic. The minimum cost system with a 16K byte cache requires two 8K × 8 static RAMs. Maximum performance can be obtained using two slightly more expensive 32K × 8 static RAMs for a 64K byte cache. Four 16K × 4 static RAMs can create a 32K byte cache.

The 386SL's cache data RAM interface has been defined in a way that eliminates the potential for adding value to the cache with improved CPU performance by means other than the cache size. Any value added to the system with static RAM other than the static RAM's size is in the form of reduced system cost, improved packaging, or reduced power demand.

The 386SL Cache Data RAM Signals

The 386SL has a complete cache data RAM interface, requiring only the addition of standard asynchronous static RAMs. The cache data RAM port's signals are defined in Table 2. The interface, shown in Figure 1, is very simple and requires no additional glue logic.

Table 2. 386SL Cache Data RAM Interface Signal Description

386SL SIGNAL	SIGNAL DESCRIPTION
CA[15:1]	Cache Address Bus Connects directly to static RAM's Address inputs. CA2 also connects to the math co-processor.
CCSH#	Cache Chip Select Low Byte Active LOW during Read and Write cycles. Connects directly to static RAM's Chip Enable Input.
CCSL#	Cache Chip Select Low Byte Active LOW during Read and Write cycles. Connects directly to static RAM's Chip Enable Input.
CD[15:0]	Cache Data Bus Connects directly to static RAM's data bus.
CWE#	Cache Write Enable Active LOW pulse during Write cycles. Connects directly to static RAM's Write Enable Input.
COE#	Cache Output Enable Active LOW pulse during Read cycles. Connects directly to static RAM's Output Enable.

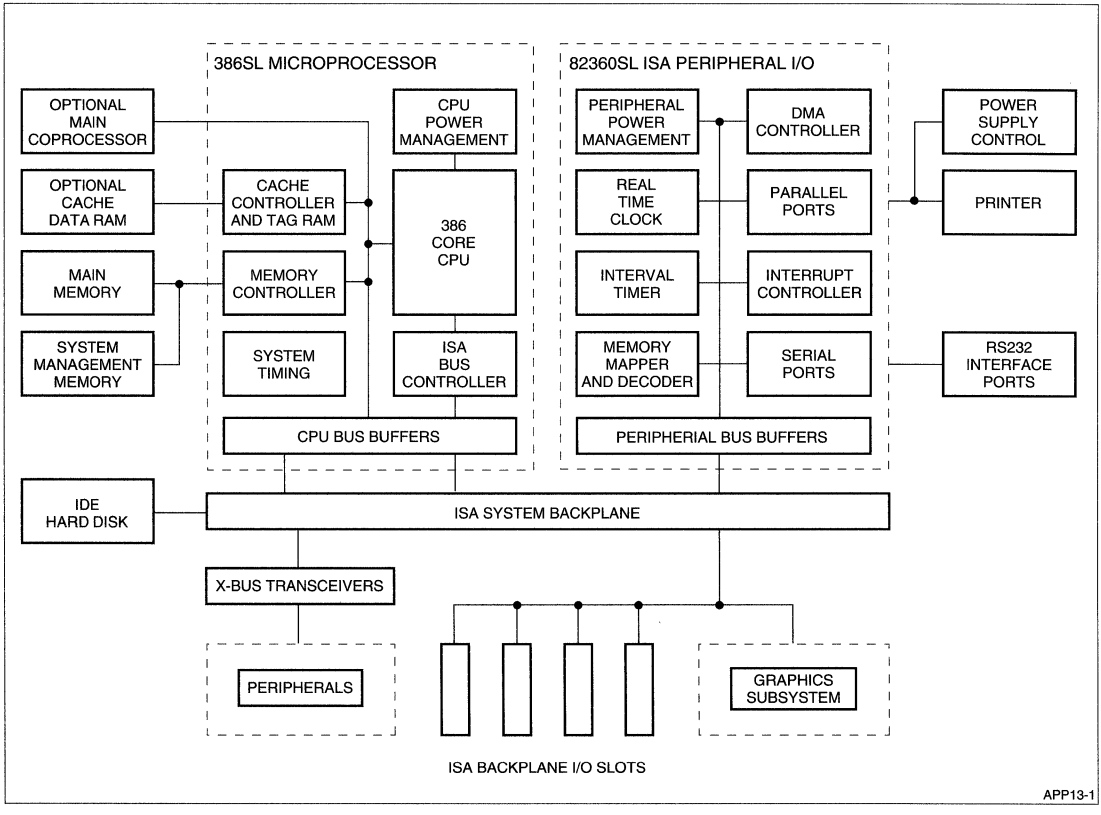


Figure 1. Typical 386SL and Companion 82360SL System

The microprocessor's signals connect directly to the static RAM. Intel uses the '#' symbol as the last character of signal names to indicate that the signal is active LOW. The CA2 Address signal also connects to the math co-processor. The address field ranges from CA1 to CA15, skipping CA0, since the 386SL's memory interface is two bytes wide. Byte-write capability is maintained by asserting only the proper Chip Enable, leaving the other Chip Enable inactive during the Byte-write cycle. Figure 2 shows the timing diagrams as presented by Intel.

The cache data RAM port operates in three distinct modes: Read Hit, Write Hit, and Idle. During the Read Hit cycle, the valid Address is provided to the static RAM on the Address bus (CA[15:1]), the static RAM's appropriate Chip Enable inputs are driven LOW (CCSH# and CCSL#), and the static RAM's Output Enable is driven LOW (COE#). The 386SL asserts the Chip Enable and Address signals just prior to the Output Enable. Given these inputs, the static RAM provides the data in the addressed memory location. At the end of the cycle, the Chip Enable and Output Enable signals are returned to the inactive levels, unless the following cycle is another Read or Write cycle. In this case, the Chip Enable signals remain active.

During Write cycles, the Address signals are driven to the correct levels, the appropriate Chip Enables are asserted, and the static RAM's Write control (CWE#) is pulsed LOW. The static RAM's data bus is then provided with the correct data, which is written into the selected location.

Analysis of the 386SL AC Timing Waveforms

The timing parameters provided with the 386SL are not as complete as one would desire, but with a few rational assumptions (and a lot of imagination), one can determine the required timing for the static RAM. For the Read Hit cycle, the Chip Enables and Address are asserted at least 1 ns prior to the Output Enable (Ct404 and Ct401). For a 20 MHz 386SL, the Output Enable access time must be 36 ns or less (Ct404a), so the Chip Enable and Address access times must be 37 ns or less. As long as the other timing parameters are met, a 35 ns SRAM should match up with 20 MHz 386SL.

The Output Enable is pulsed LOW for a minimum 72 ns (Ct402) and the data from the static RAM must remain valid for 72 ns after the falling edge of Output Enable (Ct404b). These parameters do not correspond to static RAM timing parameters. There are no timing parameters provided for the data bus management (active or inactive), but it is probably safe to assume that bus contention is not a problem when using static RAMs with normal timing.

The Write Hit timing is slightly more detailed, but still incomplete. The Address setup to the start of the Write pulse is not provided, but the Address setup to the end of Write (Ct405) is defined as 36 ns minimum for the 20 MHz 386SL. This value is 1 ns more than the minimum Write pulse width (Ct406) value of 35 ns. The Address hold at the end of the Write pulse (Ct409) is a generous 6 ns. The data setup to the end of the Write pulse (Ct407) is 25 ns, with the hold time (Ct408) of 0 ns minimum.

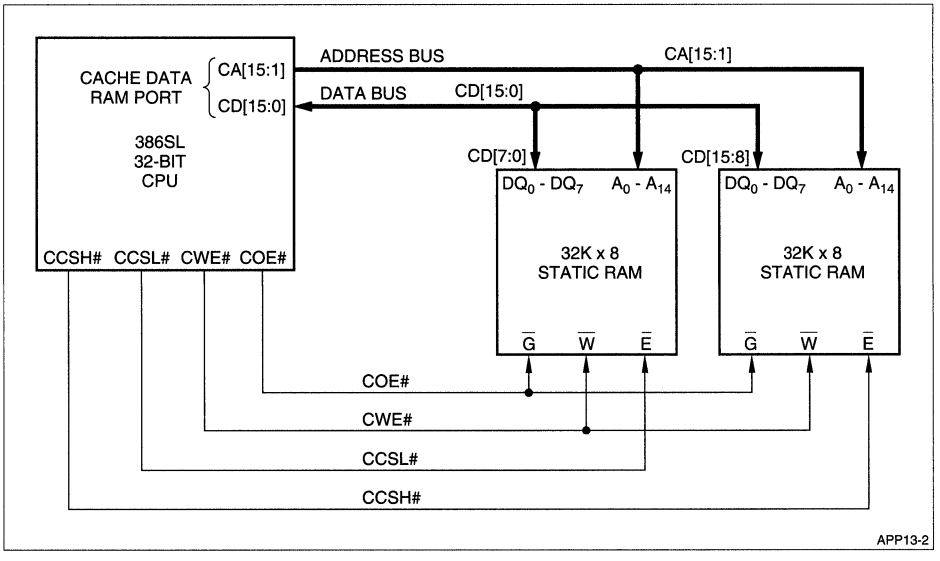


Figure 2. 386SL System With 64K Byte Cache

This timing, though not complete, does suggest that a 35 ns static RAM is suitable for a 20 MHz 386SL. Other than the Address and Chip Enable access times, the only parameter that is tight is the data hold at the end of the Write pulse, which is 0 ns for both a typical static RAM and the 386SL.

The 386SL's AC timing parameters are reproduced in Table 3. The parameters and their descriptions are in the first two columns, followed by a detailed discussion that explains the parameters in detail. Unfortunately, there is not a one-to-one relationship between the AC timing parameters of the 386SL and those of standard static RAMs.

The 386SL's DC Characteristics

The 386SL is intended for use in very low power systems. The output buffers that drive the static RAM are full CMOS in that they swing from rail-to-rail. This is not stated explicitly, but it is implied by the DC specifications table. The V_{OL} maximum is 0.5 V, with a load of 4 mA. The V_{OH} minimum is V_{CC} – 0.5 V, with a –0.18 mA load. With a –2.0 mA load, the V_{OH} minimum value is 2.4 V, for TTL compatibility.

Note that all levels are only guaranteed at 8 MHz, not at full speed. Since the signals swing to within 0.5 V of each rail with a reasonable load, it is safe to assume that the actual swing with no DC load will be less than the 0.2 V from the rails that are specified for static RAM CMOS standby current.

The 386SL documentation describes Idle mode as constant CMOS logic levels on all inputs, with the controls in the inactive state. The Address signals are shown to remain stable until either a Read or Write Hit cycle is performed. This point is critical for the static RAM. The stable CMOS input levels eliminate the potential benefit obtained from power gating the input buffers.

The important static RAM power specification is the Average Supply Operating Current (I_{CC1}). The actual system power supply current differs from the normal static RAM specification in two ways. First, the 386SL provides full CMOS levels for the static RAM inputs. Second, the cycle time is twice as long as the system clock rate, instead of the normal static RAM specification at access time. A 20 MHz 386SL cycles a 35 ns static RAM at 10 MHz, or 100 ns, which is much slower than the usual 35 ns. The net result is that any static RAM may actually consume substantially less current than its rated value during high-speed operation.

The other important issues for the static RAM are the CMOS standby current with stable inputs, which is used during Idle mode, and data retention current, which is used if the supply voltage is reduced to 3.0 V to conserve battery power. Static RAMs that feature power gated inputs have low TTL and CMOS standby current specifications, even when the inputs toggle. Static RAMs that do not have power gated inputs usually have the TTL and CMOS standby current values specified with inputs cycling.

With the stable input signals provided by the 386SL during Idle mode, the value added by power gating the inputs is eliminated. Most static RAMs used with the 386SL will consume only memory array current, or only the amount needed by the chip to maintain data, during the 386SL's Idle mode, whether the static RAM does or does not have power gated inputs.

SUMMARY

The 386SL is a very controversial new flavor of the ubiquitous 386 family. The 386SL is either the new standard CPU for notebook computers, or it will be ravaged in the market by static 386SX chips from a plethora of clone suppliers. If the 386SL is a market success, it will generate a distinct new application for CMOS Fast Static RAMs – notebook computers.

Table 3. 20 MHz 386SL CPU AC Timing Specifications

PARAMETER	DESCRIPTION	MIN	MAX	INTERPRETATION
Ct401	CABUS Setup to COE# Active LOW	–1 ns		During Read Miss cycles, the SRAM's Address becomes valid at least 1 ns prior to Output Enable. Since the Output Enable access time is 36 ns (Ct404a), the SRAM must meet an Address access time of 37 ns.
Ct402	COE# Pulse Width	72 ns		The Output Enable pulse wide is at least 72 ns in duration.
Ct403	CCSH#, CCSH# Active to COE# Active	–1 ns		The HIGH and LOW byte Chip Enables become active at least 1 ns prior to Output Enable. Just as with the Address, the static RAM's Chip Enable access time must meet 37 ns.
Ct404a	CDBUS Setup to COE# Active	36 ns		The static RAM's Output Enable access time should be no more than 36 ns. Also, the Chip Enable and Address access times must meet 37 ns, since they will be valid 1 ns prior to Output Enable (Ct401 and Ct403).
Ct404b	CDBUS Hold from COE# Active	72 ns		The static RAM must maintain valid data during Read cycles until at least 72 ns from the falling edge of Output Enable. Since the Output Enable pulse width is also 72 ns (Ct402), one must assume that the SRAM must simply keep the data valid until Output Enable becomes inactive.
Ct405	CABUS Valid to CWE Inactive High	36 ns		The static RAM's Address setup to End of Write must be no more than 36 ns. Since the Write Pulse width is at least 35 ns (Ct406), one must assume that the Address setup to the Start of the Write Pulse is at least 1 ns. Standard static RAMs typically are specified with 0 ns Address setup to the Start of Write, and an Address setup to End of Write that is equal to the Write pulse width.
Ct406	CWE# Active Width	35 ns		The static RAM's Write Pulse width should be no more than 35 ns. The 386SL uses a Write controlled Write cycle, using the SRAM's Output Enable to prevent bus contention.
Ct407	CDBUS# Setup to CWE# Inactive	25 ns		The static RAM's data setup to End of Write is less than or equal to 25 ns.
Ct408	CDBUS Hold to CWE# Inactive	0 ns		The static RAM's data hold from End of Write should meet 0 ns. Standard static RAMs typically are 0 ns, which should work in most cases, but without any margin to worst case timing.
Ct409	CABUS Hold to CWE# Inactive	6 ns		The static RAM's Address hold from End of Write should meet 6 ns.

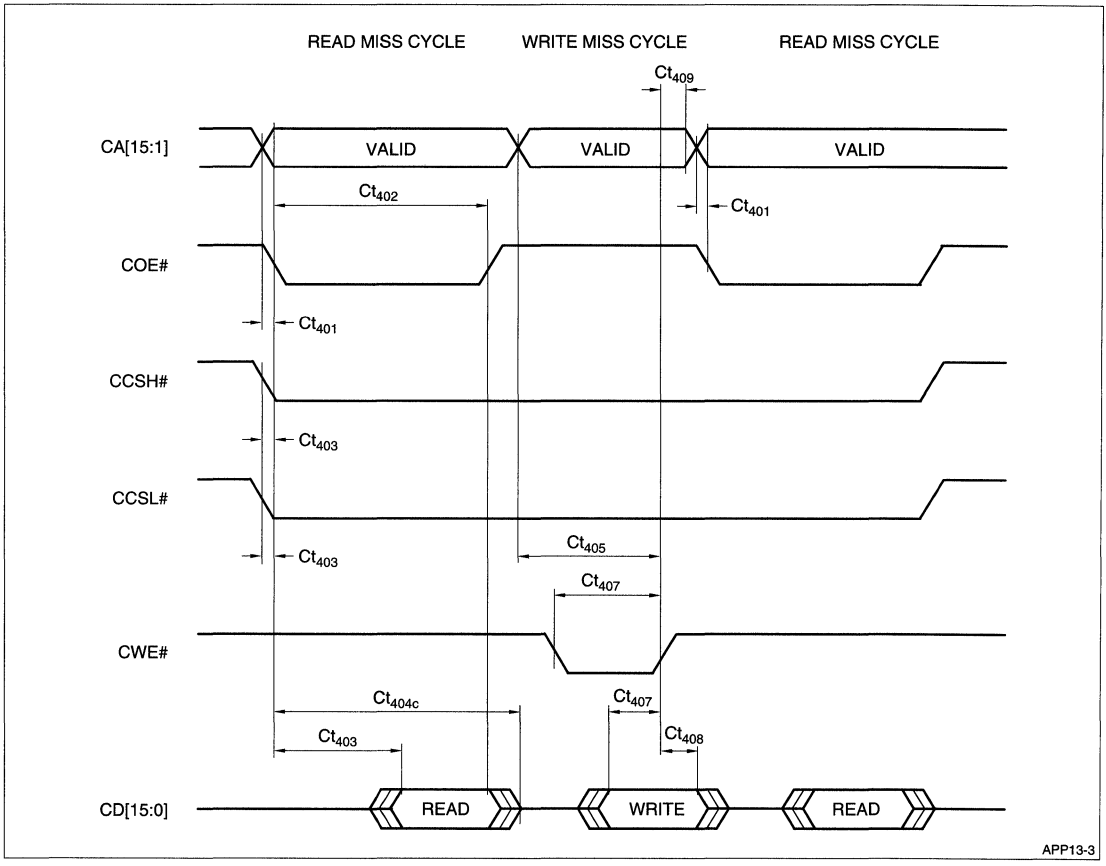


Figure 3. 386SL AC Timing Waveform

AVERAGE SUPPLY CURRENT DATA

Supply Current Characteristics

The SHARP LH521002 is a CMOS Fast Static RAM fabricated in a 0.8 m, 2-layer metal CMOS process. Special care has been taken in the design of the LH521002 to minimize power dissipation under actual system conditions, especially when the system cycle rate is slower than access time.

The standard technique used to specify a Fast SRAM's supply current is with the chip operating at the fastest cycle time allowed, which is usually the same as the address access time (t_{AA}). The vast majority of Fast SRAMs are actually cycled significantly slower than their rated fastest cycle time. The system cycle time is usually on the order of twice the access time. Many Fast SRAMs include I_{CC} specifications at cycle times that are longer than the minimum cycle time, as is the case of the LH521002. Also, data may be extracted from the device for different operating conditions that will provide the system designer and reliability engineer with the information required to understand how the device will perform in the system.

Measurement Conditions

Figures 1 and 2 present actual measured data taken from the LH521002 for Read and Write cycles, graphically illustrating measured supply current versus cycle time. This data has been taken at a variety of ambient operating temperatures, ranging from –10°C to 80°C, with the supply voltage set to 5.0 Volts. The data was obtained under worst case conditions, and the test pattern used toggled all address and data signals during every cycle.

NOTE: During Read cycles, I_{OUT} = 0, which is the same as the normal test specification. The current required to drive the load circuit connected to the output is in addition to the Read cycle current shown.

Understanding the Data

As the data in Figures 1 and 2 shows, the LH521002's power supply current drops significantly at slower cycle times for both Read and Write cycles. The LH521002 was designed to exhibit this characteristic. The sample data is indicative of the supply current that will be seen in systems. Individual devices may exhibit characteristics with slightly higher or lower values, but this data can be used to determine cooling and system power requirements.

During Read cycles, the majority of a Fast SRAM's supply current following access time is simply the DC current required in the TTL input buffers, the bit-line loads and the differential sense amplifiers. The LH521002 uses proprietary circuit design techniques to minimize the supply current required for the sense amplifiers and bit-line loads. The result is that the supply current required during long Read cycles is approximately the Standby Current plus the current required by the TTL input buffers. Additional power savings may be obtained by using CMOS input voltage levels, which eliminate the DC current required by the input buffers.

For Write cycles, a typical Fast SRAM requires DC supply current for long Write cycles in the TTL input buffers and also in the bit-line loads. Again, the LH521002 reduces the DC supply current during long Write cycles by minimizing the current through the bit-line loads.

Example

A 32-bit system uses a 33 MHz CPU with 2 CPU clocks per memory cycle. The memory cycles once every 50 ns, rather than the rough 25 ns access time required for the application. From the graphs, the Read cycle current is about 40 mA and the Write cycle current is about 60 mA. If we assume that 80% of the memory cycles are Read cycles, with 20% Write cycles, and that in worst case the memory bus is in use 100% of the time, then the average supply current is estimated to be 50 mA per device or 400 mA total.

SUMMARY

The supply current required by the LH521002 under real system conditions, especially during longer cycle times, is substantially reduced. Most systems operate at cycle times that are finite and that are substantially longer than the specified access time. The data provided will assist in the estimation of total power actually dissipated by the LH521002 in the system environment.

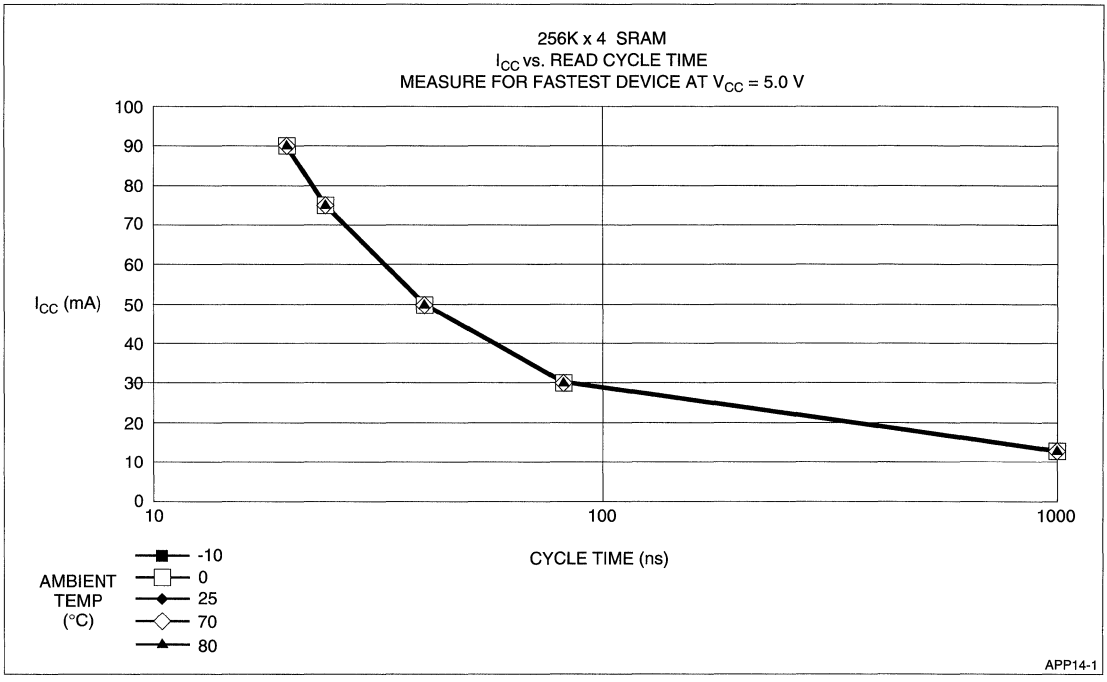


Figure 1. I_{CC} vs. Read Cycle

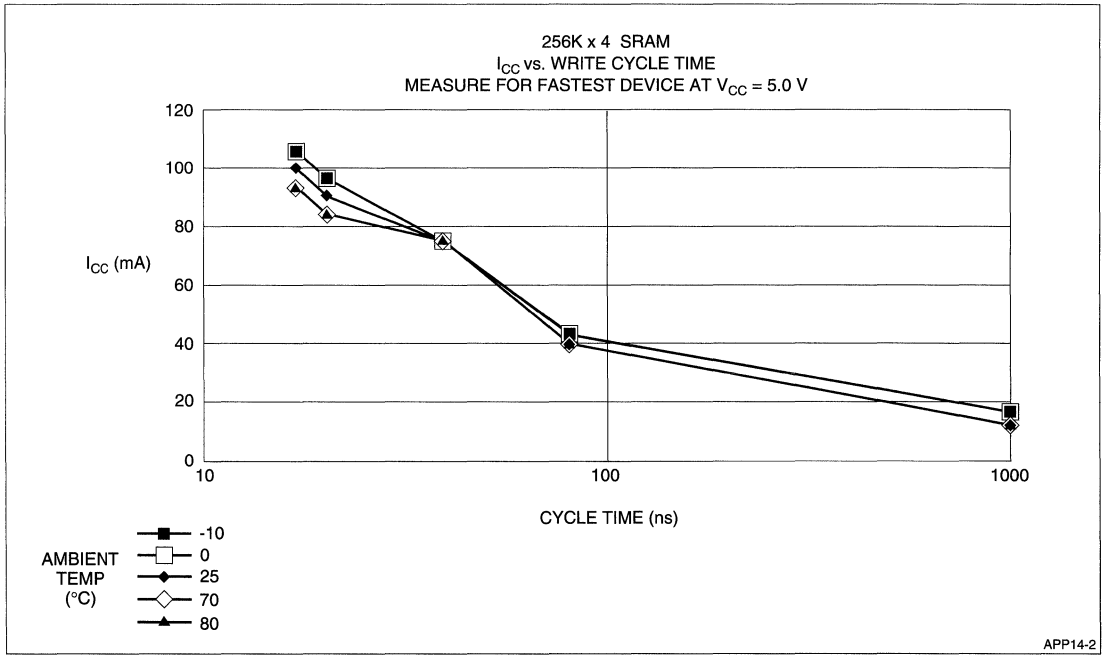


Figure 2. I_{CC} vs. Write Cycle

AVERAGE SUPPLY CURRENT DATA

Supply Current Characteristics

The Sharp LH521008 is a CMOS Fast Static RAM fabricated in a 0.8 m, 2-layer metal CMOS process. Special care has been taken in the design of the product to minimize the device's power dissipation under actual system conditions, especially when the system cycle rate is slower than access time.

The standard technique used to specify a Fast Static RAM's supply current is with the chip operated at the fastest cycle time allowed, which is usually the same as the address access time (t_{AA}). The vast majority of Fast Static RAMs are actually cycled significantly slower than their rated fastest cycle time. The system cycle time is usually on the order of twice the access time. Many Fast Static RAMs include I_{CC} specifications at cycle times that are longer than the minimum cycle time, as is the case of the LH521008. Also, data may be extracted from the device for different operating conditions that will provide the system designer and reliability engineer with the information required to understand how the device will perform in the system.

Measurement Conditions

Figures 1 and 2 present actual measured data taken from the LH521008 for Read and Write cycles respectively, graphically illustrating measured supply current versus cycle time. This data has been taken at a variety of ambient operating temperatures, ranging from $-10^{\circ}C$ to $80^{\circ}C$, with the supply voltage set to 5.0 V.

The data was obtained under worst case conditions and the test pattern toggled all address and data signals during every cycle.

NOTE: During Read cycles, I_{OUT} = 0, which is the same as the normal test specification. The current required to drive the load circuit connected to the output is in addition to the Read cycle current shown.

Understanding the Data

As the data in Figures 1 and 2 show, the LH521008's power supply current drops significantly at slower cycle times for both Read and Write cycles. The LH521008 was designed to exhibit this characteristic. The sample data is indicative of the supply current that will be seen in systems. Individual devices may exhibit characteristics with slightly higher or lower values, but this data can be used to determine cooling and system power requirements.

During Read cycles, the majority of a Fast Static RAM's supply current (following access time) is simply the DC current required in the TTL input buffers, the bit-line loads and the differential sense amplifiers. The LH521008 uses proprietary circuit design techniques to minimize the supply current required for the sense amplifiers and bit-line loads. The result is that the supply current required during long Read cycles is approximately the Standby Current plus the current required by the TTL input buffers. Additional power savings may be obtained by using CMOS input voltage levels, which eliminates the DC current required by the input buffers.

For Write cycles, a typical Fast Static RAM requires DC supply current for long Write cycles in the TTL input buffers and also the bit-line loads. Again, the LH521008 reduces the DC supply current during long Write cycles by minimizing the current through the bit-line loads.

Example

A 32-bit system uses a 33 MHz CPU, with 2 CPU clocks per memory cycle. The memory cycles once every 50 ns, rather than the rough 25 ns access time required for the application. From Figures 1 and 2, the Read cycle current is about 47 mA and the Write cycle Current is about 68 mA. If we assume that 80% of the memory cycles are Read cycles, with 20% Write cycles, and in worst case the meory bus is in use 100% of the time, then the average supply current is estimated to be 50 mA per device or 200 mA total.

SUMMARY

The supply current required by the LH521008 during real system conditions, especially at longer cycle times, is substantially reduced. Most systems operate at cycle times that are finite and that are substantially longer than the specified access time. The data provided in Figures 1 and 2 will assist in the estimation of total power actually dissipated by the LH521008 in the system environment.

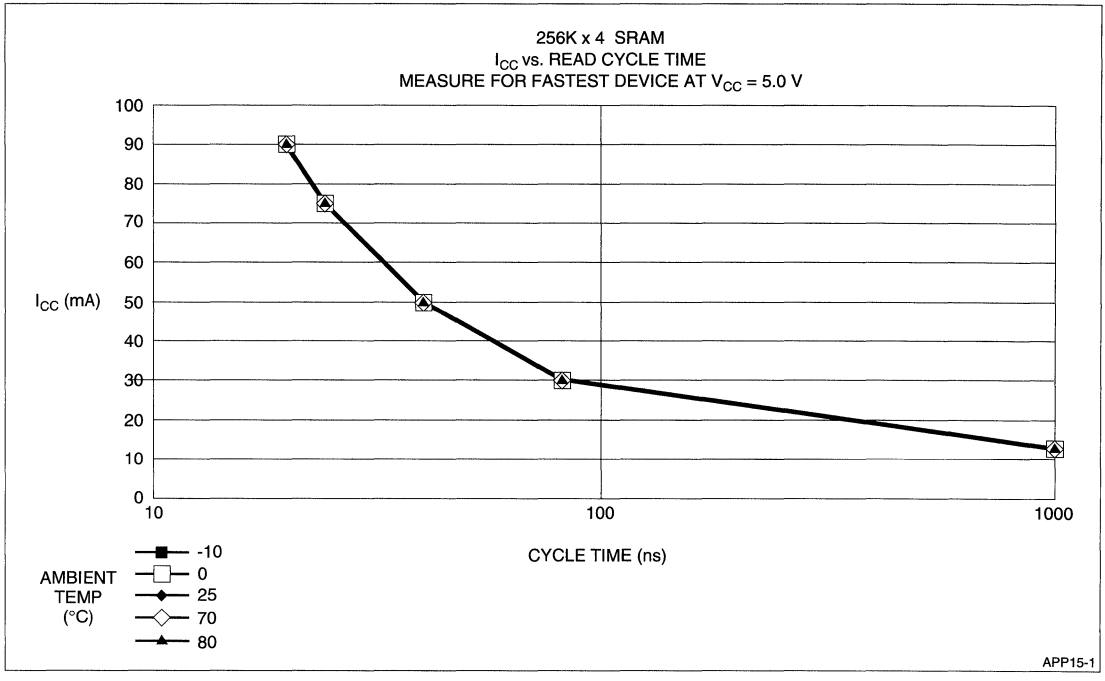


Figure 2. I_{CC} vs. Read Cycle

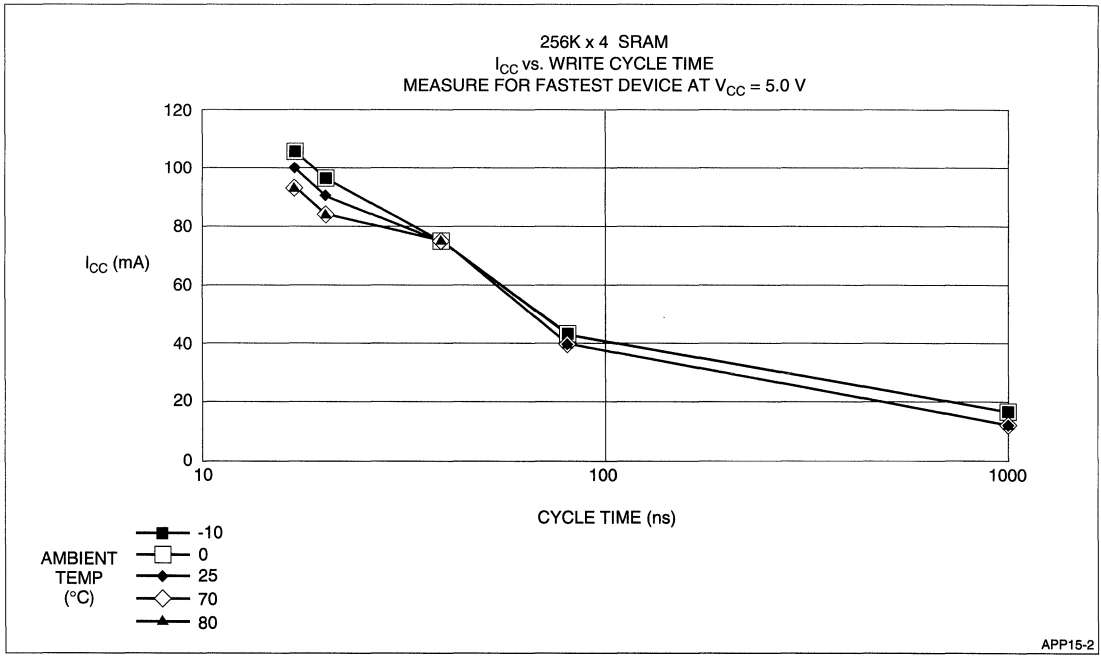


Figure 1. I_{CC} vs. Write Cycle

GROUND BOUNCE MEASUREMENT IN ASYNCHRONOUS STATIC RAMS

INTRODUCTION

Very high speed CMOS Integrated Circuits often have fast output signal edges. These fast edges have resulted in the emergence of a common problem – ground bounce. Ground bounce is the rise of the on-chip ground voltage above the system ground level. Ground bounce is the result of large current transients passing from the on-chip Ground through a package inductance, to the system Ground ($v = -L\ di/dt$). Physically large integrated circuits also may have appreciable on-chip resistance in the supply lines, adding to the problem. The large current transients are often the result of the switching of the chip's output buffers, primarily when a large number of the outputs switch from a HIGH voltage level to a LOW voltage level. In most applications, the outputs drive only PCB traces and CMOS loads, resulting in a capacitive load.

Asynchronous static RAMs pose a special challenge in the measurement of ground bounce, due to their unique operation. In this note, we will examine cause of ground bounce, the application of standard ground bounce test methods to asynchronous static RAMs and the problems with this test methodology, and finally offer some suggestions for evaluating ground bounce in asynchronous static RAMs.

The Cause of Ground Bounce

Figure 1 shows the basic on-chip wiring in a × 8 static RAM. The ground pin is in the lower left hand of the package, DIP or SOJ. There are three input/output pins above the Ground pin and five on the lower portion of the right side of the package. In static RAMs, the Ground reference for all input signals is usually routed separately from that of the outputs. But since these chips have only a single ground pin, the separate ground must meet before reaching the external pin. The distance where both circuits share the same ground path is the area where ground bounce originates.

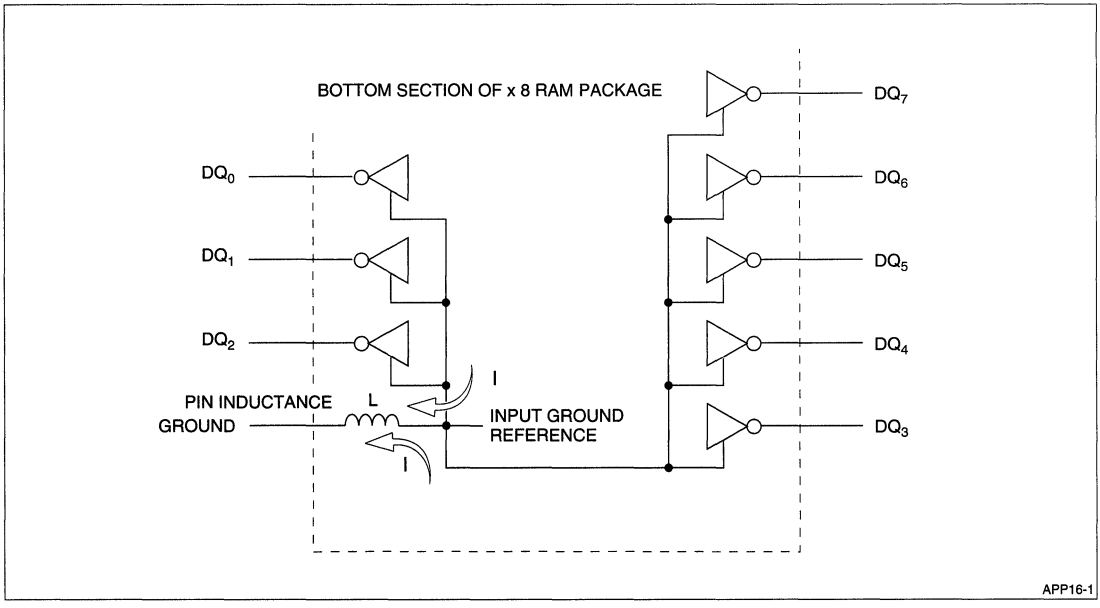


Figure 1. × 8 Static RAM Internal Wiring Diagram

At least some amount of ground bounce will always exists, but the level can be held to a manageable value. When the chip's outputs make a HIGH-to-LOW transition, a large current surge through the pin inductance causes the internal chip ground to 'bounce' above the external reference (v = −Ldi/dt). The large current transients are a by-product of the need to switch the outputs quickly.

Standard Ground Bounce Tests

Ground bounce measurement is a simple test for most devices. Figure 2 is an example of a circuit used to test a 74FCT240 inverting buffer. Seven of the device's inputs are driven from a common source, while the output under test has the input tied HIGH. The output under test should have constant LOW level. Six outputs drive RC loads, as shown. The one remaining output is used as a reference, along with the output under test. Both signals drive a 50 Ω oscilloscope input through a 450 Ω resistor, matching the 500 Ω load on the other signals.

The seven common inputs are driven from a square wave generator, generating the inverted signal on the seven outputs. The test output remains LOW during the test. On every HIGH-to-LOW transition performed by the seven switching outputs, the voltage on the constant LOW output will rise above the static LOW voltage, due to ground bounce. This upward pulse on

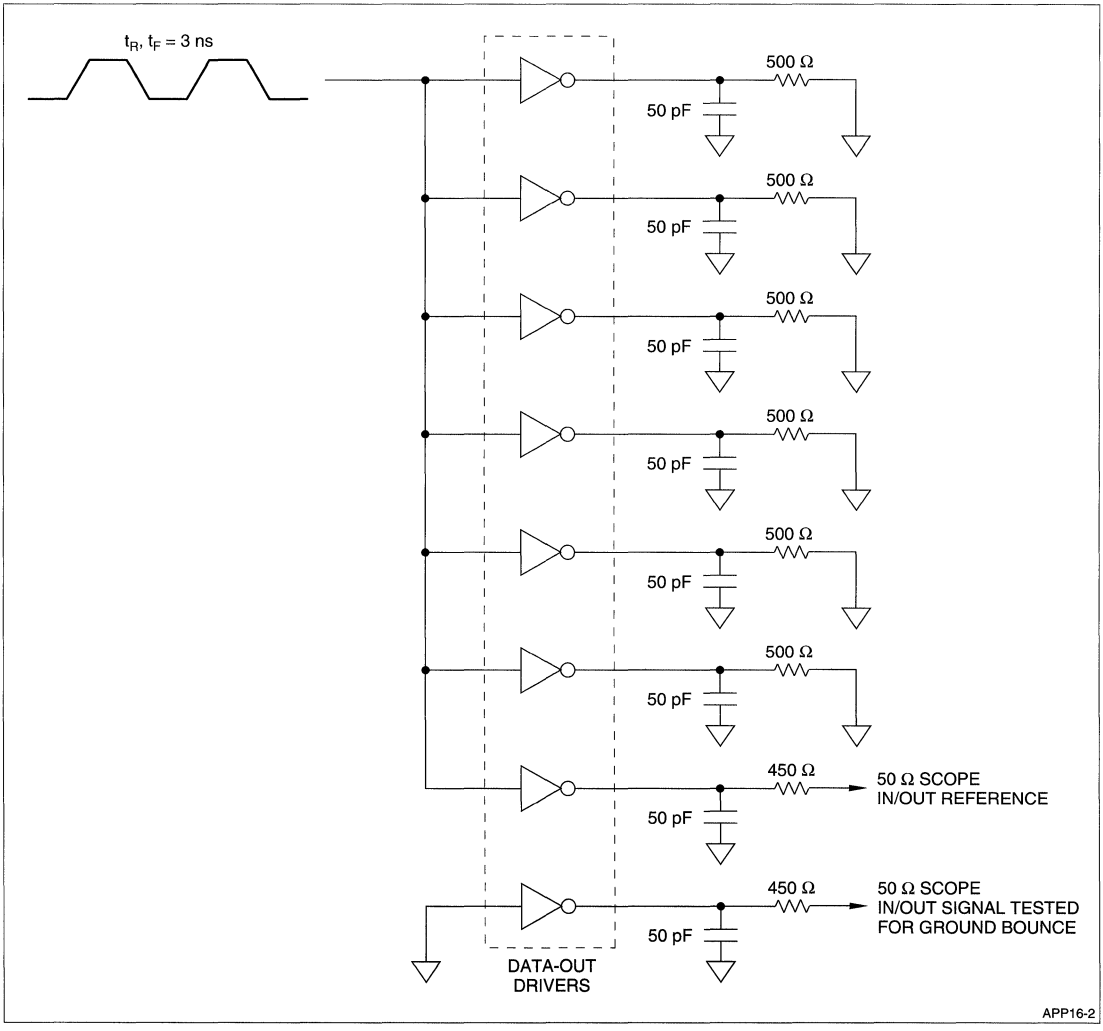


Figure 2. Ground Bounce Test Circuit for 74FCT240 Inverting Buffer

the output signal reflects the ground bounce on the on-chip ground. Each of the outputs is tested, since ground bounce can vary by pin position.

Figure 3 shows a typical output waveform on the reference output and the tested output. Ground bounce in CMOS devices is worst case usually at the lowest temperature and the highest supply voltage, due to relationship between speed and temperature in CMOS technology. Ground bounce can not be entirely eliminated, due to the need to switch the outputs quickly. Ground bounce is considered excessive when the level exceeds the maximum allowable value for the next input stage's LOW level voltage, V_{IL}, usually 0.8 V.

For logic devices that may drive an edge sensitive input such as a register's clock, a ground bounce level exceeding the V_{IL} limit would cause the edge sensitve signal to possibly receive a false trigger. Asynchronous static RAMs can not drive edge sensitive signals, so this is not a problem. The problem does appear in reduced margin of the chip's input signals, due to the shared Ground pin shown earlier.

Application of the Standard Ground Bounce Test to Asynchronous SRAMs

Figure 4 shows an example of the standard ground bounce test applied to an asynchronous static RAM. The results of this test are invalid, as shown by the 'X' in the figure. The results are *invalid* since ALL SRAM outputs are a function of ALL address inputs. When any address input changes state for example, all outputs become invalid as soon as t_{AX} (Data-out Hold from Address Change) and remain invalid until t_{AA} (Address Access) following the address transition. Figure 5 shows the timing that would exist under these conditions, contrasted to the timing desired to conduct this test.

The closest approach to making this test valid is to limit the number of addresses that change to a single address input. In this case, one memory location contains 011 . . . (one LOW, the rest HIGH), and the other location contains 000 . . . (all LOW). Even in this case, the transition of a single address input means that the SRAM's outputs are ALL invalid in the period between t_{AX} and t_{AA}. An asynchronous static RAM does not change exclusivly from one output level to the next. Basic test philosophy eliminates any test that requires the Device Under Test (DUT) to operate in a manner beyond the operational specification. In the case of a static RAM, the chip must meet the two previously stated timing parameters, but is not required to remain in one state or the other during the interval between the two points.

Alternative Tests for Measuring Ground Bounce in Asynchronous Static RAMs

One of the best method for evaluating an asynchronous static RAM's sensitivity to excessive ground bounce is to perform V_{IH} vs. V_{CC} shmoo plots. These are plots that show the results of Pass/Fail tests performed while varying input levels and supply voltages. A device with excessive Ground current will either not pass V_{IH} or V_{IL} tests at the required levels, or will have 'Holes' where the chip will pass at the specified values and fail at others within the required operating range. An example of a Shmoo plot is shown in Figure 6.

For Asynchronous static RAMs, all outputs are function of all Address and Control inputs. A change or transition on any of these inputs may change all output signals. Also, there may be a significant difference in access times between a Fast address and a Slow address. For these reasons and others, static RAM outputs *MUST NEVER* drive edge sensitive inputs. As shown earlier, the constant LOW level on the output under test may not occur.

With the input and output circuits sharing at least part of the same ground pin circuit, excessive ground bounce will appear as input level problems (usually V_{IH}). Input Shmoo plots graphically illustrate the presence of ground bounce, when it is a problem.

SUMMARY

Ground bounce is an important issue among all CMOS devices, including asynchronous CMOS Static RAMs. The unique operation of asynchronous static RAMs precludes the use of standard ground bounce tests. The best tool for understanding the impact of ground bounce induced by output switching in asynchronous static RAMs is V_{IH} and V_{IL} shmoo plots against V_{CC}.

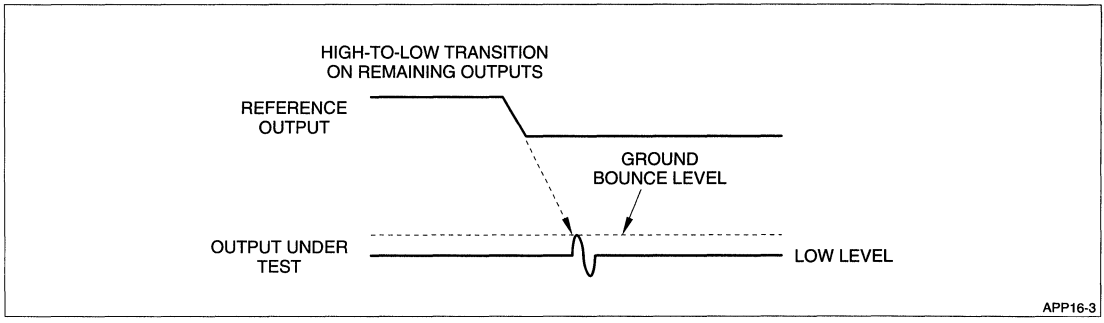


Figure 3. Typical Output Waveforms With Ground Bounce 74FCT240 Inverting Buffer

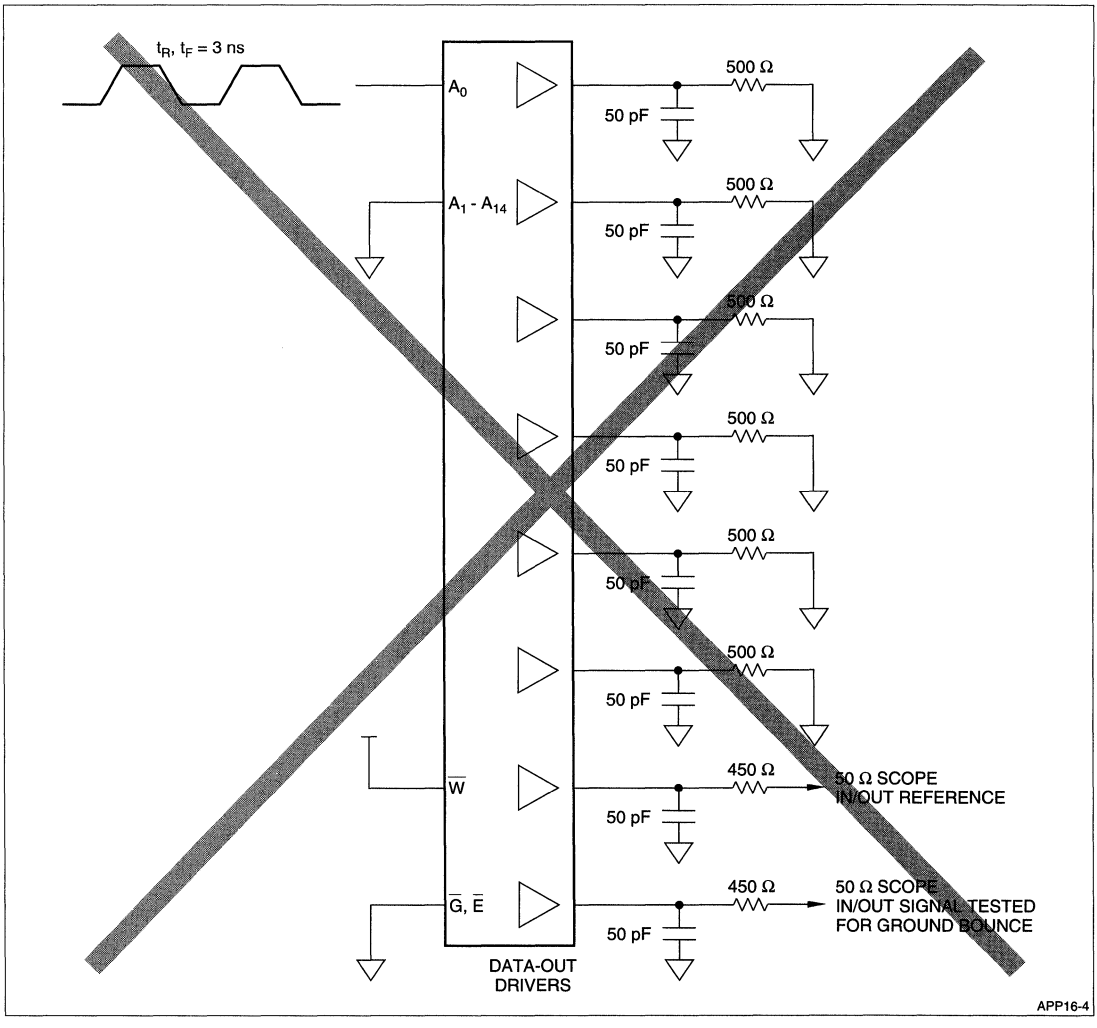


Figure 4. Invalid Test Setup for Measuring Ground Bounce in × 8 Static RAMs

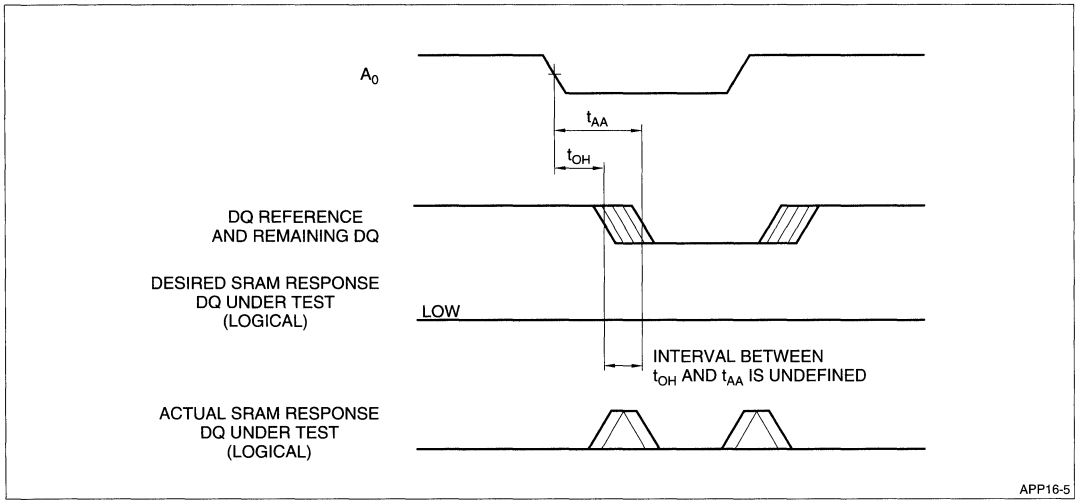


Figure 5. Invalid Timing Set for Measuring Ground Bounce in Asynchronous Static RAMs

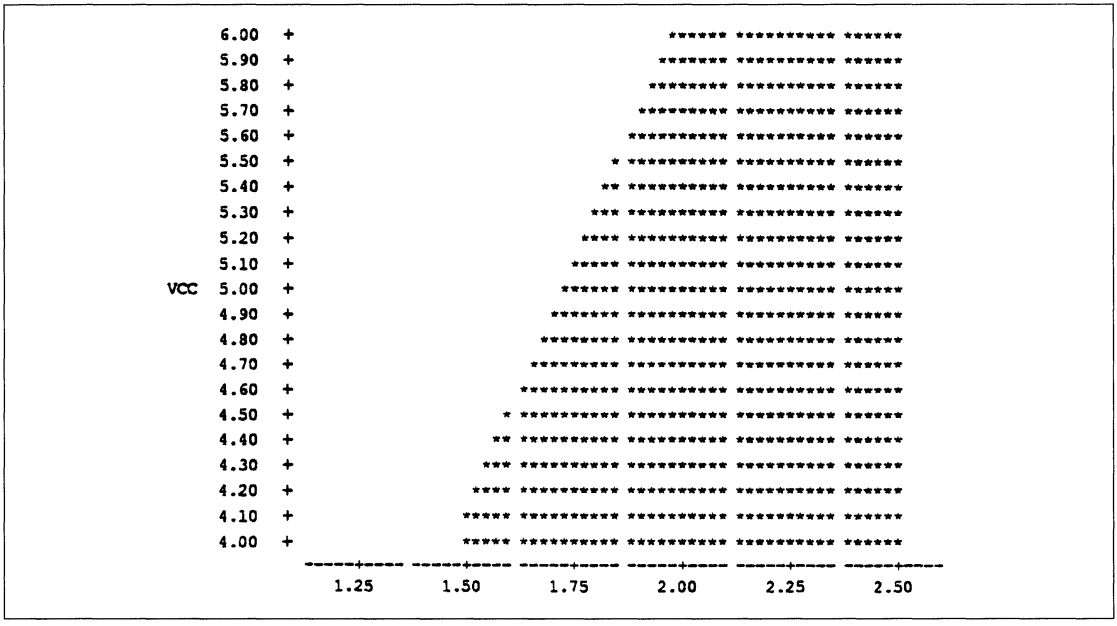


Figure 6. Static RAM V_{IH} vs. V_{CC} Shmoo Plot

USING THE ADDRESS LATCH TO MEET ADDRESS HOLD FROM END OF WRITE (t_{AH})

INTRODUCTION

Write-cycle timing for standard, asynchronous, fast static RAMs is one of the most difficult challenges for system designers, especially during the design of synchronous memory systems. The definition of a synchronous memory system used here is a system where all inputs and outputs are timed from a single clock edge. A typical example of a synchronous system is shown in Figure 1. All input and output signals are timed from the CLK signal. The static RAM Sub-System (shaded region) includes the asynchronous SRAM, along with any circuitry required to interface the static RAM to the synchronous world.

Probably the most difficult timing parameter to successfully achieve in this environment is the Address Hold from the End-of-Write (t_{AH}). This application note describes a technique using the LH521028's Address Latch to solve the Address Hold challenge. First, we need to understand an important definition.

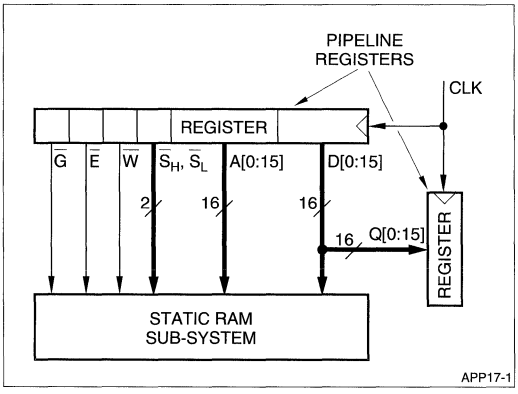


Figure 1. Example of a Synchronous System

The Asynchronous Static RAM Write Pulse

Most synchronous static RAM data sheets are simplified by the use of a pseudo-signal called the Write Pulse, as shown in Figure 2. This signal is sometimes confused with the LOW pulse on the Write Enable control during Write cycles. Actually the Write Pulse is defined as the overlapping period when both Chip Enable and Write Enable are LOW, resulting in a Write operation in the static RAM. This simplification reduces a static RAM Data Sheet to only a few cycle types. A complete specification would require several additional cycle types. Actually, either Chip Enable or Write Enable may initiate or terminate the Write Pulse, so there are four distinct Write cycles.

Most static RAM manufacturers use this technique, which is probably a good example of *Occam's Razor*. The Address Hold time parameter (t_{AH}) is defined as the minimum interval between the first Address transition and the first of the two controls, Chip Enable or Write Enable, to make the LOW-to-HIGH transition, terminating the Write Pulse. The Address Setup is the similar parameter at the start of the Write Pulse. If the signals make the transition at exactly the same point in time, then t_{AH} is relative to that point in time. *If either t_{AH} or t_{AS} are violated, then any or all memory locations may contain the wrong data, not just the desired location.*

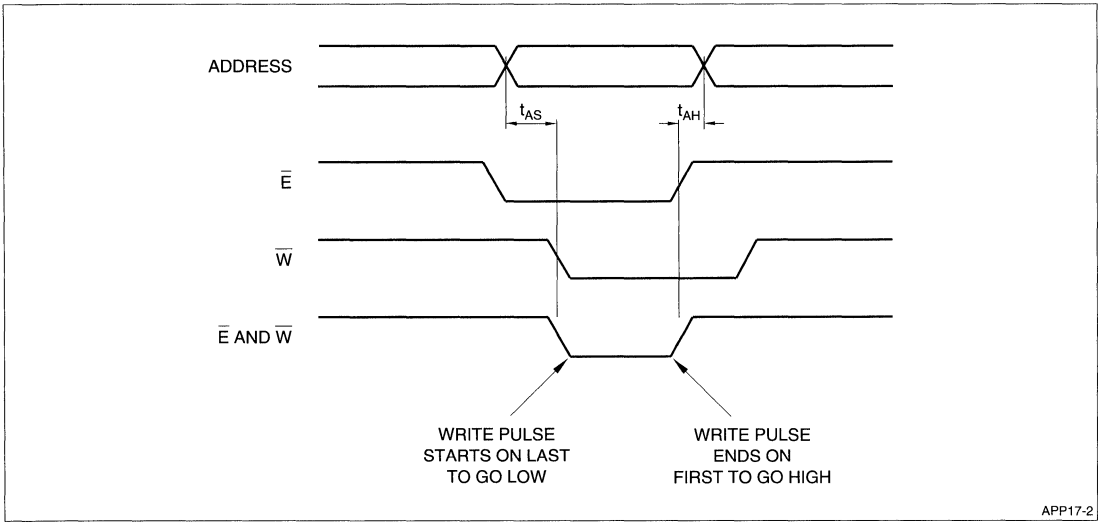


Figure 2. The Standard Asynchronous Static RAM's Write Pulse Specification

Synchronous System Timing

In the system block diagram shown in Figure 1, all inputs to the static RAM Sub-System change on the rising edge of the clock, which is a common occurrence in numerous static RAM applications. Newer microprocessors that use 'burst mode' for on-chip cache fills are one example of a synchronous memory interface, along with many DSP and Embedded Control applications.

Figure 3 shows an example of the system level (not necessarily the static RAM, yet) Read and Write cycle system timing. During the Read cycle, the Address, both Chip Selects, and Output Enable are ASSERTED following the rising edge of the clock; the system databus becomes high-impedance, allowing the static RAM to drive the bus; and the system Write control is DEASSERTED. The timing of the output signals referenced to the clock (CLK) usually is specified as maximum delays, hopefully with minimum delays, but usually never with the differential delay between separate outputs (i.e. skew).

During a Write cycle, the Address, the system Write control, and the appropriate Chip Selects are ASSERTED; Output Enable is DEASSERTED; and the static RAM's Data-Out bus becomes high-impedance (since Output Enable is DEASSERTED), while the system's Databus provides the data for the Write operation. Since the system timing does not guarantee that the Address Setup and Hold times are met during Write cycles, we must derive the signal used as the static RAM's Write Enable control.

NOTE: Chip Enable is always ASSERTED in this example, but could be used for memory depth expansion, Standby mode, or other functions. Chip Enable can also be used to control the Write Pulse, with similar implications.

The timing in this example is straightforward, except for the Write Enable pulse during the Write cycle. The pulse could be generated with the use of delay lines, but they are expensive and often troublesome. Instead, we will use a common technique, simply gating the LOW portion of the clock with the system Write control. The Address Setup Time to the leading edge of the Write pulse is easily achieved, assuming that the Address becomes valid prior to the falling edge of the clock. However, a race condition exists in the Address Hold Time. The maximum delay from the rising edge of the Clock through the gate controlling the Write Pulse must always be less than the minimum hold time of the Address from the rising edge of the Clock, including all loading and circuit board delays. One solution to the problem is somehow to force the Address to remain valid until after the end of the rising edge of the Write Enable signal, which is where the Address latch comes to the rescue.

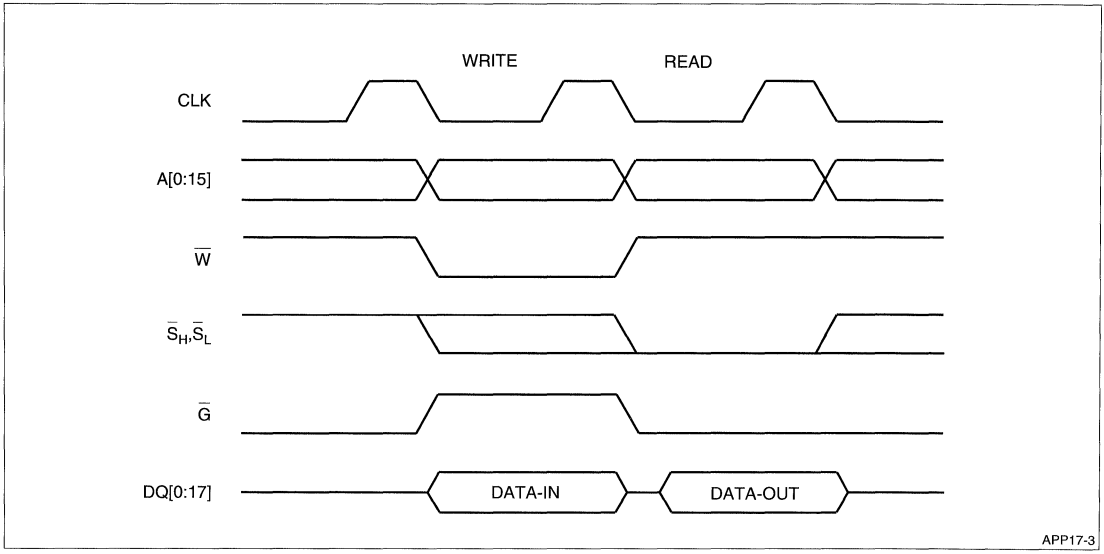


Figure 3. Synchronous System Timing

The LH521028 Provides a Simple Solution

The LH521028 incorporates a 16-bit address latch on-chip, offering a simple solution to this frequently encountered address hold time problem. Connecting the Address Latch Enable (ALE) input on the LH521028 to the static RAM's Write control through a non-inverting buffer eliminates the address hold problem. The purpose of the buffer is to add a time delay, to ensure that the ALE signal follows the Write Enable signal under all conditions. Figure 5 shows the circuit that generates the ALE signal, and Figure 6 illustrates this circuit's timing. This solution assumes that the loading and wiring delays are similar on the Write Enable and Address lines.

When the ALE signal is HIGH, the LH521028's address bus latch is transparent. When LOW, the state of the address bus is latched. The goal is to maintain a valid address to the static RAM array until after the addresses change at the end of the Write. The solution is to connect the Write Enable signal, through a non-inverting gate, to the ALE input. The addresss is latched one gate delay following the beginning of the Write, and remains latched until one gate delay after the end of the Write pulse.

The Address Setup Time is not affected by the latch, since the latch is transparent while ALE is HIGH. Since the address remains latched until the gate propagation delay after the end of the Write pulse, the Address Hold Time is easily achieved. The user *must always* meet the Address Setup Time for the latch (t_{ASL}) and the Address Hold Time (t_{AHL}), as described later.

One note of caution: any changes on the address bus following the end of the Write pulse are ignored until the rising edge of ALE, which follows Write Enable by one gate delay. This is characteristic of the solution to the t_{AH} problem, but can possibly create another problem if the following cycle occurs IMMEDIATELY after the end of the Write cycle. If the next cycle is a Write operation, Write Enable is probably DEASSERTED long enough to allow ALE to become HIGH. If the following cycle is a Read cycle at a different Address, then the Read cycle has a new parameter – Access Time from ALE HIGH (t_{LEA}). Since ALE LOW means that the previous address is latched, ALE must become HIGH to allow the new address to propagate into the static RAM array. The new parameter, t_{LEA}, must be satisfied along with the other access times. Most systems are not paced by t_{LEA}, but this fact must be verified in each design.

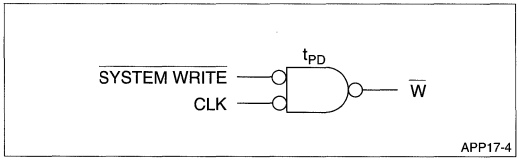


Figure 4. Write Pulse Circuit

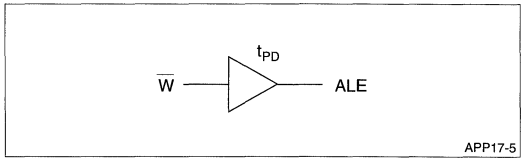


Figure 5. Address Latch Enable Circuit

In systems where *both* Chip Enable and Write Enable control the Write Pulse, a simple OR gate is required to generate the ALE signal, as shown in Figure 7, with timing similar to that shown in Figure 6. In this example, the Addresses are latched when both Chip Enable and Write Enable are ASSERTED. The Address Hold Time is measured from the first rising edge of either of the two signals, which precedes the rising edge of ALE by the gate's propagation delay. If the propagation delay through this gate were 0 nanoseconds, then the output would be the so-called Write pulse.

SUMMARY

The LH521028's address latch can solve several application problems. In this application note, a synchronous memory system's Address Hold Time is achieved, using minimal logic, and taking advantage of the on-chip latches. External latches may be used; but they add propagation delay to the address path, requiring faster, and more expensive, static RAM. A small circuit like a latch is shown to provide substantial benefits, especially when incorporated on-chip.

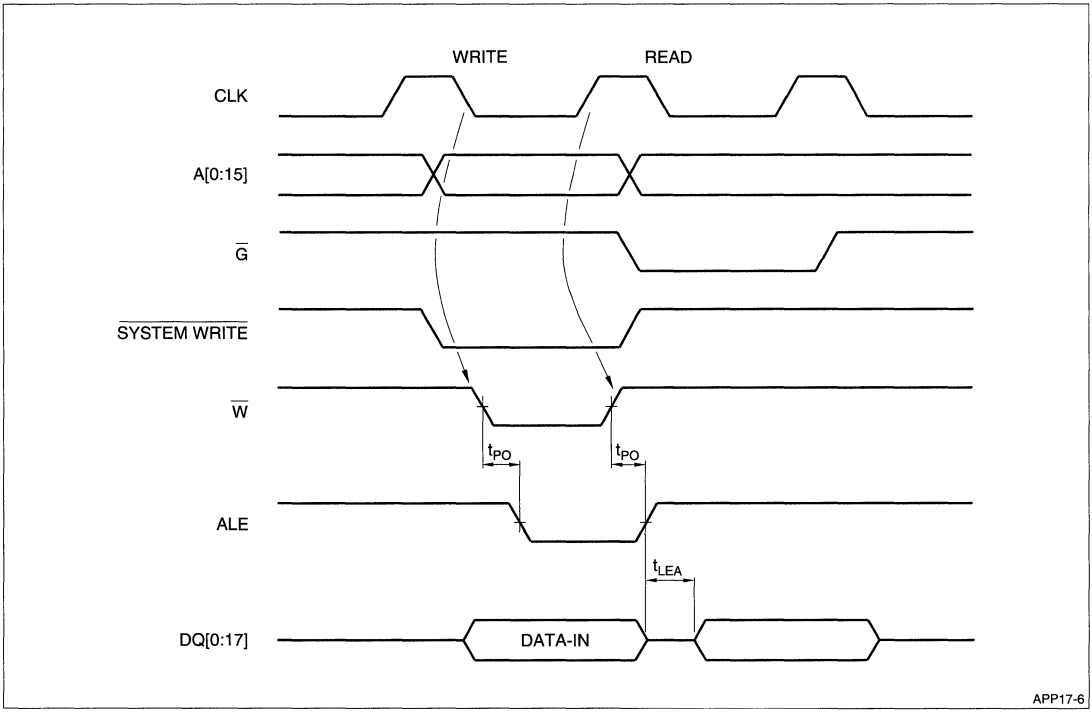


Figure 6. Address Latch Enable Circuit Timing

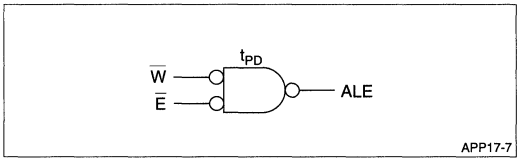


Figure 7. OR Gate (DeMorgon) Used to Generate ALE Signal

LH521028
64K × 18 Static RAM

The Sharp LH521028 is a 64K × 18 CMOS Fast Static RAM, manufactured in 0.8 micron CMOS process. The available speeds range from 20 to 35 nanoseconds. The LH521028 is versatile product, with the capability to solve a wide range of system problems.

The LH521028 is housed in a compact 52-pin Plastic Leaded Chip Carrier (PLCC), measuring 0.750 × 0.750 inches. The pin connections, which are JEDEC standard, is shown in Figure 1. The power supply terminals on the package side are dedicated to the Data bus. The pinout uses a philosophy of one set of power supply terminals to each 4½ Data drivers. This approach minimizes ground bounce for the present day high speed LH521028, as well future versions operating at much higher speeds.

Earlier 32-bit microprocessors, without built-in parity, are very difficult and expensive, to use with parity applied to Fast Static RAM. These microprocessors require an external parity circuit for each byte, with the parity circuit's propagation delay added to the data path. The net result is the requirement for faster, and therefore more expensive, Static RAM.

Several of the latest microprocessor offerings (486, 68040, etc) now incorporate parity into the data path, resulting in a 9-bit data byte and a 36-bit or 72-bit data word. The LH521028, with an 18-bit width, readily accommodates parity in either Cache Data RAM or Main Memory applications. Since there is no external parity circuit with these microprocessors, there is no performance penalty introduced by the addition of parity data protection in systems based on the LH521028.

The LH521028's block diagram is shown in Figure 2. The memory array is divided into two nine-bit bytes, with each array consisting of 64K nine-bit bytes. The primary control signal is Chip Enable ($\overline{E}$). The Chip Enable control functions exactly like the signal with the same name that is found on conventional Static RAMs. This signal controls the state of the chip, including both memory arrays, placing the chip either in Active mode ($\overline{E}$ = LOW) or Standby mode ($\overline{E}$ = HIGH). When in Active mode, Read and Write operations are allowed, as determined by the state of the other control signals. In Standby mode, the power level is reduced to the Standby level, with the state of the other control signals Don't Care. The data busses associated with each of the two arrays are identified as the Low-byte (DQ_0-DQ_8) and the High-bytes (DQ_9-DQ_{17}). Read and Write operations are individually controlled for each byte, with the $\overline{SL}$ and $\overline{SH}$ signals, along with the Write ($\overline{W}$) and Output Enable ($\overline{G}$) signals. The Truth Table for the LH521028 is shown in Table 1.

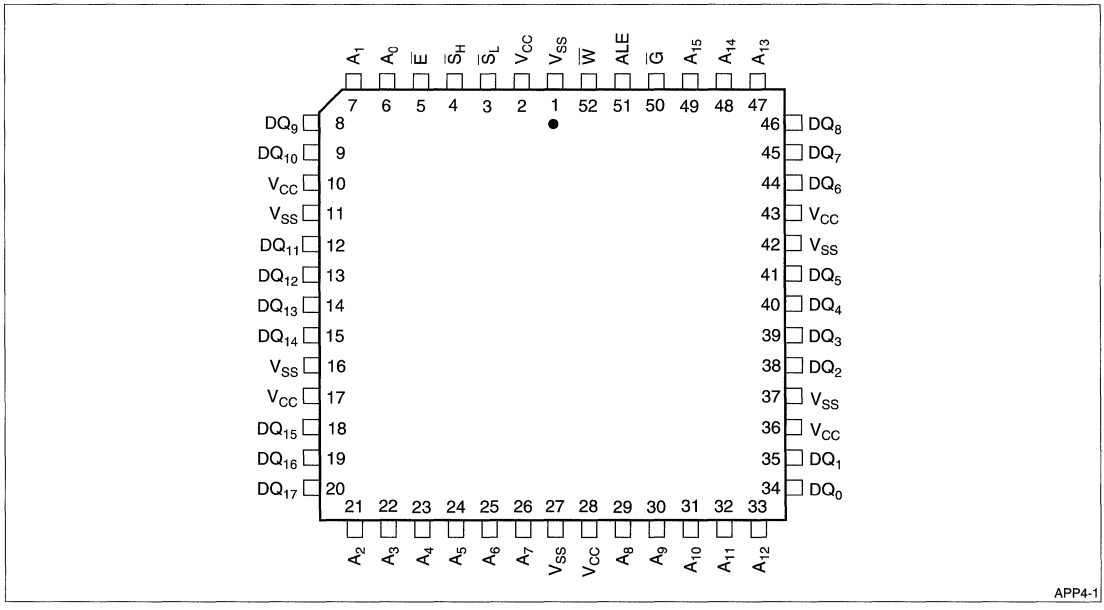


Figure 1. LH521028 Pin Connections

The Byte Select controls are primarily used to implement Byte-write capability. The Byte Select control for the particular byte of interest is enable to allow a Byte-write, while the other control is disabled. At any time the Write Enable signal is active, all data-out signals are high impedance. The result is that during a Byte-write into the Low-byte data bus, the High-byte data bus is high impedance.

The Byte Select controls also allow a Byte-read operation. Byte-read capability is seldom required in the x18 configuration, but does provide a unique capability. The LH521028 can be configured as a 128K × 9 SRAM, by simply connecting the two data busses together and decoding the Byte Select controls. An example of this configuration is shown in Figure 3. Additionally, a different configuration provides a 64K × 9 × 2-way associative Cache Data RAM. Both × 9 configurations also provide the Address latch.

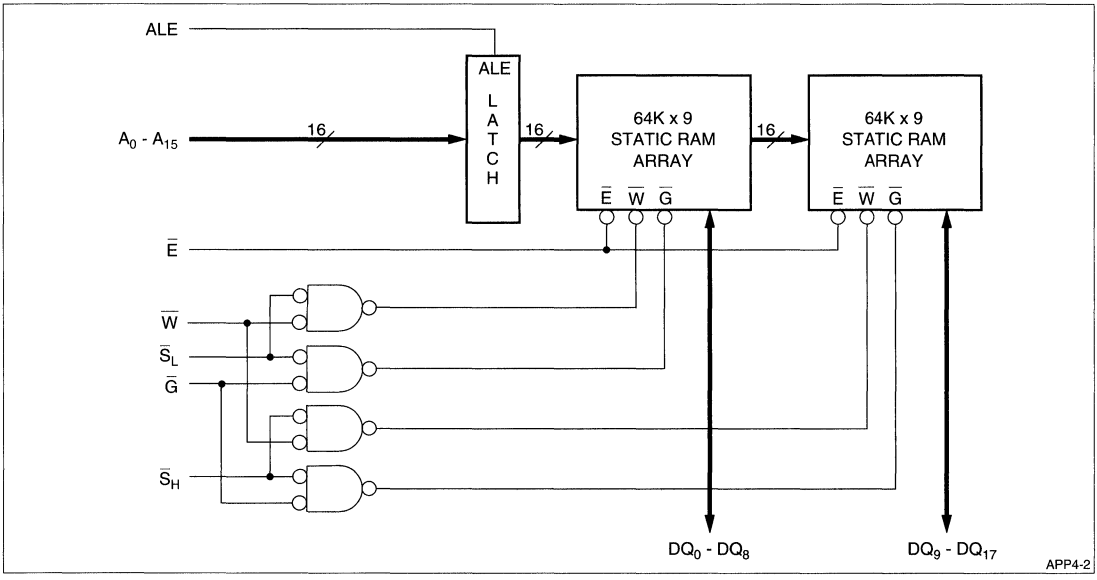


Figure 2. LH521028 Block Diagram

Table 1. Truth Table

OPERATION	$\overline{E}$	$\overline{W}$	$\overline{G}$	$\overline{S}_H$	$\overline{S}_L$	ALE	A_0-A_{15}	DQ_0-DQ_8	DQ_9-DQ_{17}
Standby Mode	1	X	X	X	X	X	X	Hi-Z	Hi-Z
Read High Byte	0	1	0	0	1	X	Address	Hi-Z	Data-out
Read Low Byte	0	1	0	1	0	X	Address	X	Hi-Z
Read Both Bytes	0	1	0	0	0	X	Address	Data-out	Data-out
Read Mode	0	1	X	1	1	X	Address	Hi-Z	Hi-Z
Read Mode	0	1	1	X	X	X	Address	Hi-Z	Hi-Z
Write High Byte	0	0	X	0	1	X	Address	X	Data-in
Write Low Byte	0	0	X	1	0	X	Address	Data-in	X
Write Both Bytes	0	0	X	0	0	X	Address	Data-in	Data-in
Write Mode	0	0	X	1	1	X	Address	X	X
Transparent Address	X	X	X	X	X	1	Address	X	X

The Address latch is a transparent, level sensitive, latch that is controlled by the Address Latch Enable (ALE) control. The latch is transparent when ALE is HIGH and the previous state of the Address bus is latched when ALE is LOW. The Address latch is fully operational at all times, even while the chip is in Standby mode. If desired, one could latch the contents of the Address bus while in Standby mode, then enter Active mode and access the memory. This point is important, since any potential timing relationship between Chip Enable and Address Latch Enable is eliminated.

The Address latch is an important feature in any system that uses Multiplexed Address and Data signals, as in Figure 5, or in systems that require a latch in the address path in order to pipeline the memory access. Multiplexed Address and Data busses are again becoming popular, reducing the pin count (and cost) of embedded RISC microprocessors.

One of the significant advantages of wide word memories is their potential to eliminate the need for Address bus buffers, reducing the chip count and most importantly, eliminating the propagation delay added to the Address path by the Address bus buffers. Systems that require Address latches would otherwise require an external latch while using conventional memories, along with faster, more expensive memories without the Address latch.

Not all systems require or can take advantage of an Address latch. The address latch can be permanently disabled by tying the Address Latch Enable HIGH. The latch is a very simple function when incorporated on-chip, with virtually no impact on the performance or cost of the device.

The LH521028's wide word, 18-bits, minimizes the number of chips required to match the system's word width. In most cases, only one or two chips are required to complete the systems SRAM requirement. This reduced chip count results in a dramatically reduced loading on the Address bus, when compared to a similar system using conventional Static RAMs.

One technique sometimes used in the design of systems using conventional memories that have loading on the Address bus in excess of the microprocessor's rated drive capability is to use a De-rating curve for timing to compensate for the excessive load. This technique can also effectively allow the Address bus to operate without additional buffers, but faster Static RAMs are required. In many cases, the LH521028 will eliminate the need to resort to De-rating curves that are not guaranteed, simply due to the significantly reduced loading on the Address bus. The system designer will have a much higher degree of confidence in the system design, as a result of using published and guaranteed timing parameters, instead of resorting to characteristic tables.

In summary, the LH521028 provides a number of useful features. These include a wide word (18 bits), Byte control, transparent Address latch, along with a power distribution system that eliminates ground bounce both now and in the future generation products.

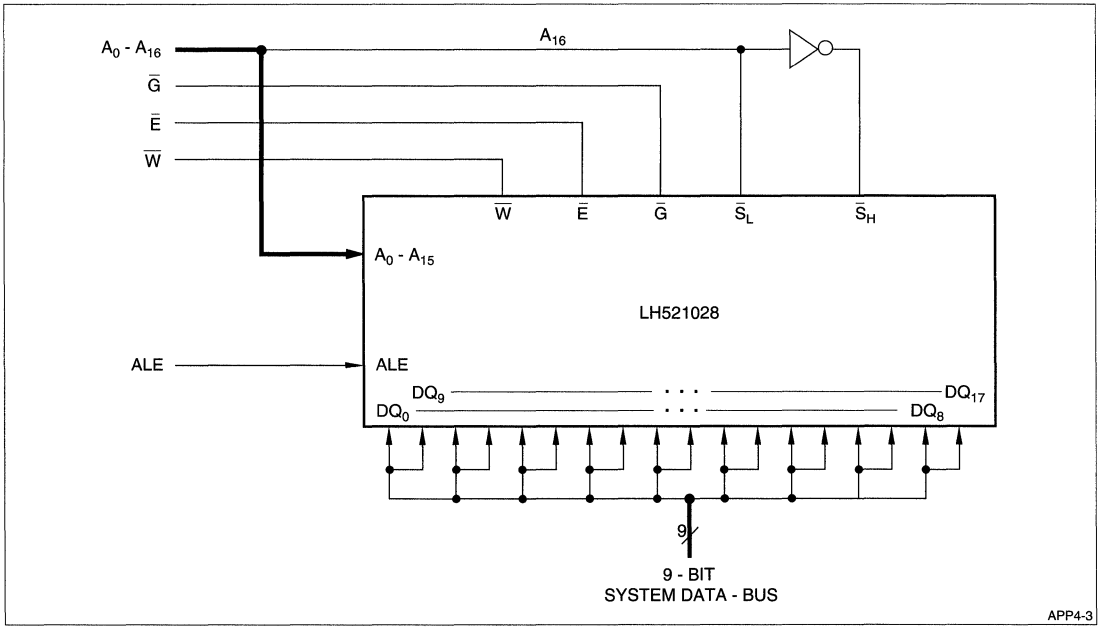


Figure 3. 128K × 9 Configuration With Address Latch

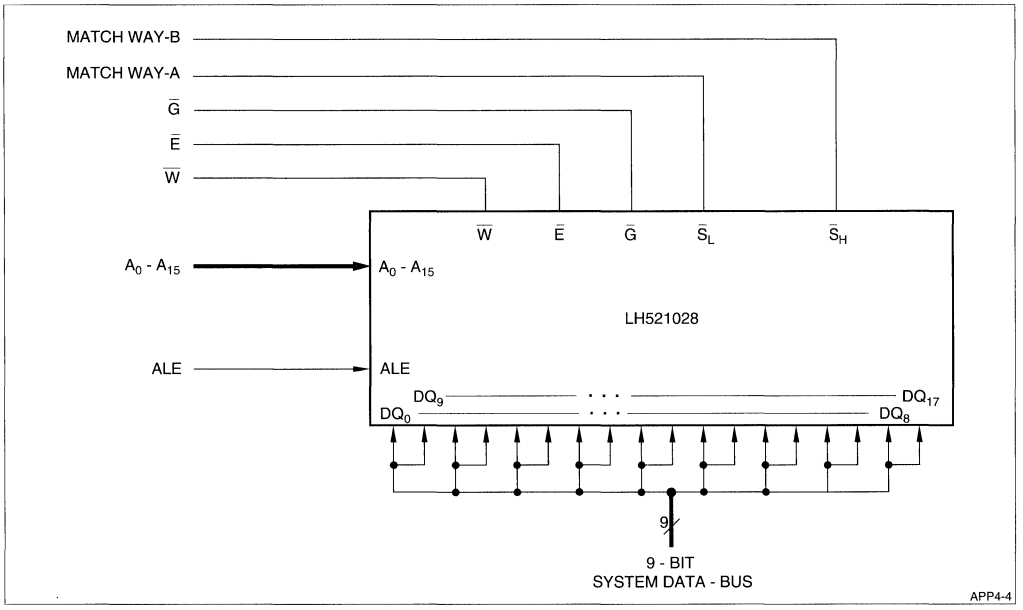


Figure 4. 64K × 9 × 2-Way Cache Data RAM

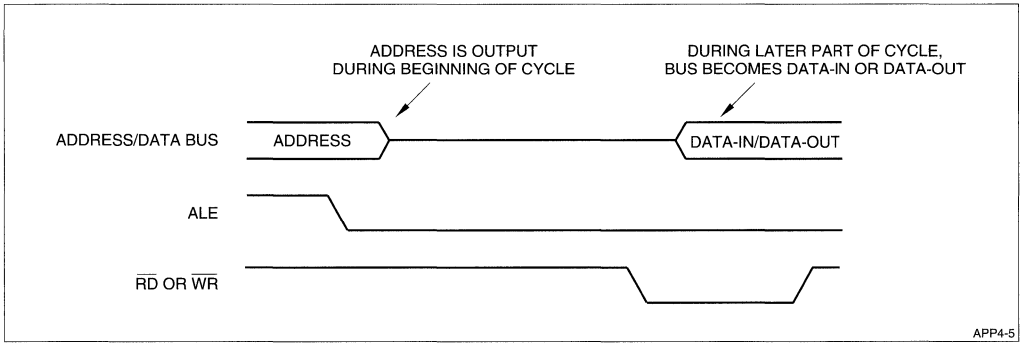


Figure 5. System With Multiplexed Address and Data

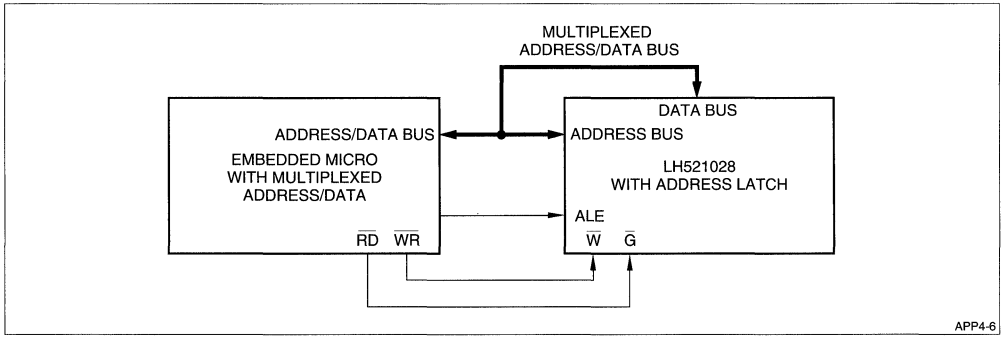


Figure 5 (cont'd). System With Multiplexed Address and Data

USING ADDRESS AND DATA LATCHES TO ELIMINATE WRITE HOLD TIMING ISSUES

INTRODUCTION

The design of high-speed memory systems is often a battle to meet static RAM Write timing parameters. More often than not, the Address and Data hold times at the end of the Write are the most difficult challenges, especially in synchronous memory systems. Systems that use buffers on the Address and Data buses can often use a 'free' solution to the problem, by substituting a transparent latch (i.e., 74XX373) for the buffer.

Write Enable Controlled Static RAM Write Cycle

This application note looks at one type of Write cycle – Write Enable controlled. For simplicity, assume that the Chip Enable control is tied LOW, so that only the Write Enable control is used to create the Write pulse. Also, assume the Output Enable, if present on the static RAM, is tied LOW. The same general principles apply to more complex systems.

Figure 1 contains the timing waveforms for a Write controlled Write cycle. All timing parameters, except for the Write cycle time, reference the falling or the rising edge of Write Enable. The Address is valid before the falling edge of Write Enable, meeting the t_{AS} Address setup specification. For most static RAM chips, t_{AS} is 0 nanoseconds. The Data bus is valid at the t_{DW} Data setup before the end of the Write pulse. Both the Address and Data buses are valid until the end of the Write pulse at the rising edge of Write Enable. The Address Hold time is t_{AH} and the Data Hold time is t_{DH}; both are typically 0 nanoseconds. In addition, the minimum Write pulse width is t_{DW} and the minimum Write cycle time is t_{WC}.

System Timing

Synchronous systems often generate the Write signal, the Data, and the Address simultaneously, usually from the same active clock edge. In this case, the skew between the Clock-to-Valid delays of the Address and Data bus, to the Write control, allow a violation of the Address setup, Address Hold, or Data Hold times. Usually, the Data bus is valid well ahead of the required point in time.

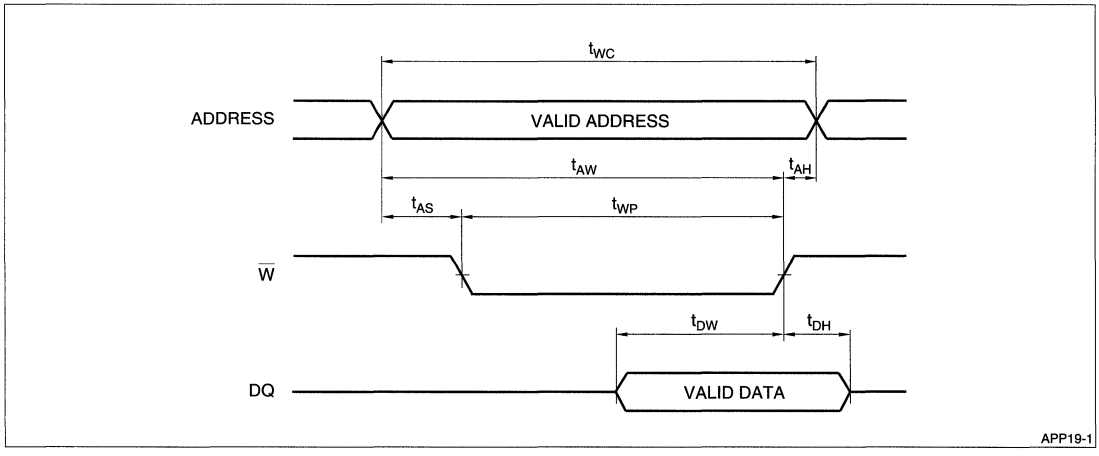


Figure 1. Write Controlled Static RAM Write Cycle

In order to meet the Address setup parameter, an AND function generates the static RAM's Write pulse from the synchronous Write signal and the second half of the Clock. Figure 2 shows a typical circuit and Figure 3 contains the resulting timing waveform. This circuit provides a generous Address setup time to the leading edge of the Write pulse, assuming that the clock HIGH time is longer than the Clock-to-Address valid delay.

Even if no skew exists between signals generated from the system clock, the AND function adds at least a few nanoseconds delay to the time the Write signal is active LOW. This delay, along with the skew between the signals, results in a problem with the Address and Data Hold times. The Address and Data Hold times are missed by the amount of skew between the signals plus the AND function's propagation delay. Since the Address and Data Hold times are 0 nanoseconds from the trailing edge of the Write pulse, all that is required is to keep the buses valid until after ($t_{AH} > 0$) the rising edge of Write enable.

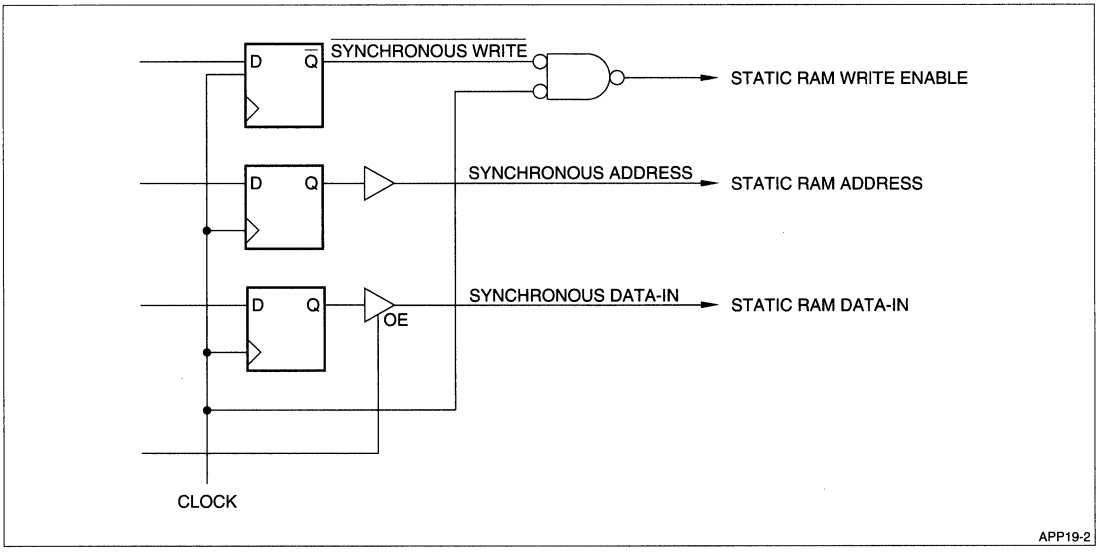


Figure 2. Circuit Generating Write Pulse in Synchronous System

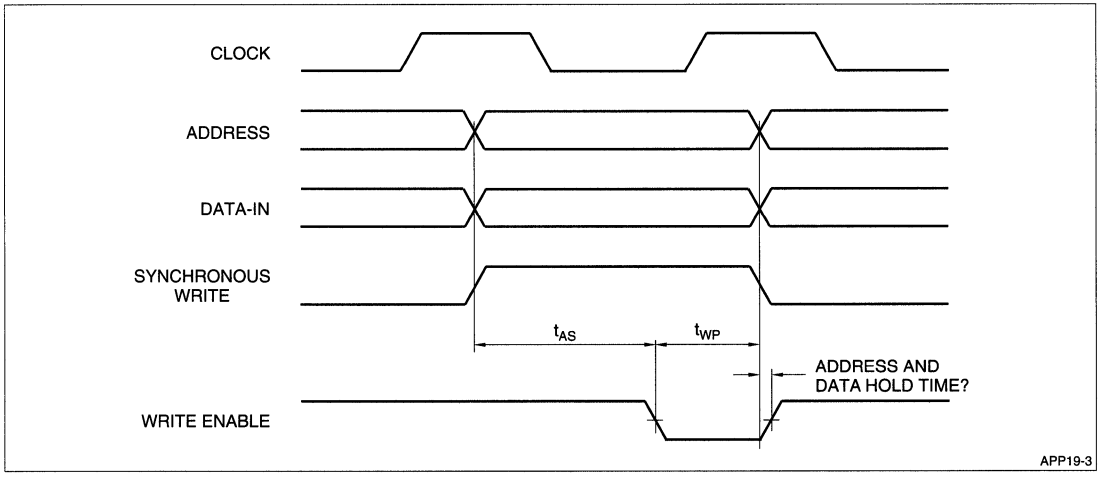


Figure 3. Timing With Clock Generated Write Pulse

Use Latched Buffers on Address and Data Buses

Using latched buffers on the Address and Data buses allows the extension of their valid interval by latching the contents of the buses and releasing them after the static RAM's Write Enable control has made the LOW-to-HIGH transition. This function is created by simply using the static RAM's Write control to control the transparent latches, as shown in Figure 4. The Address and Data Hold times become no less than the minimum propagation delay through the latches, which is always greater than 0 nanoseconds.

The delay through the latches should exceed the delays on the PCB, but an analysis should verify this fact. A latched buffer with a slower minimum propagation delay is also a possibility.

Although simple, this solution comes with a cost:

1. Usually the only true cost is incurred in a Read cycle immediately following a Write cycle. The Address hold time comes at the expense of a delayed Address bus in the following cycle, which reduces the acceptable Address access time if the following cycle is a Read. Many synchronous systems require a dummy clock cycle between Write and Read cycles to prevent bus contention on the Data bus. Even if there is no dummy cycle allowing for Data bus management, and the delay on the Address bus penalizes a following Read cycle, a faster static RAM solves the problem. The Address and Data Hold times are constant with all static RAM speeds, so the user may have no choice other than faster static RAMs in order to meet Write cycle timing requirements.
2. Increased component count, unless the latches replace simple buffers.
3. Reduced Data setup time, unless the latches replace buffers on the Data bus. If the Data becomes valid early in the cycle, which is common in synchronous systems, there would be more than sufficient margin in the Data setup time even with the latch.
4. Reduced Address setup time at the beginning of the Write, which is usually not a problem since the Write pulse is delayed by 1/2 clock period.

SUMMARY

Write cycle timing for fast static RAMs often create a difficult challenge. A well-placed latch is often a simple solution, especially when a latched buffer replaces a buffer.

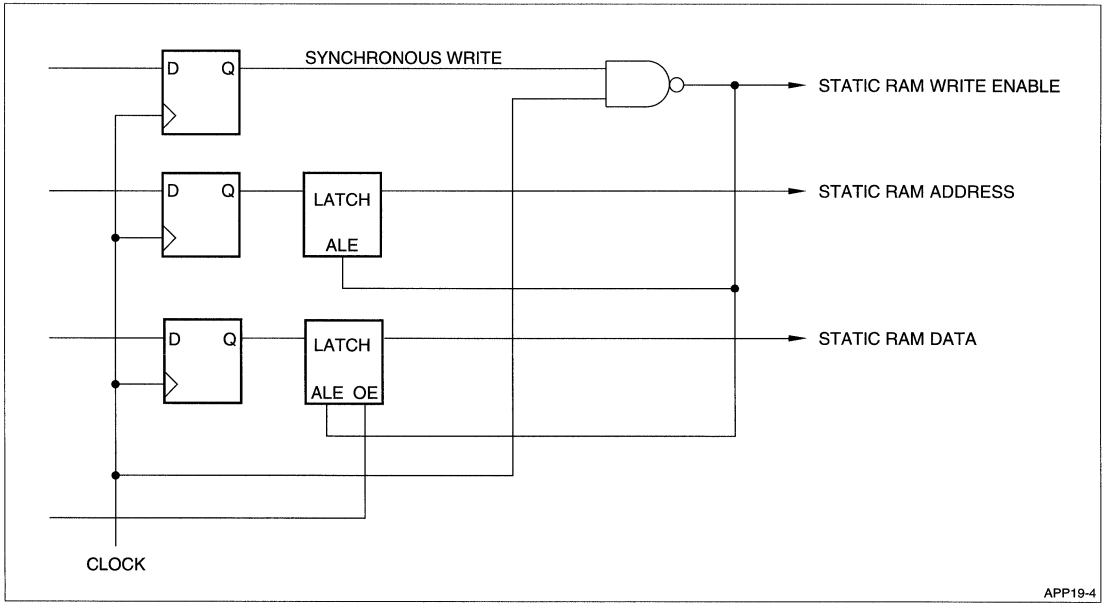


Figure 4. Improved Circuit Meeting Hold Time Requirements

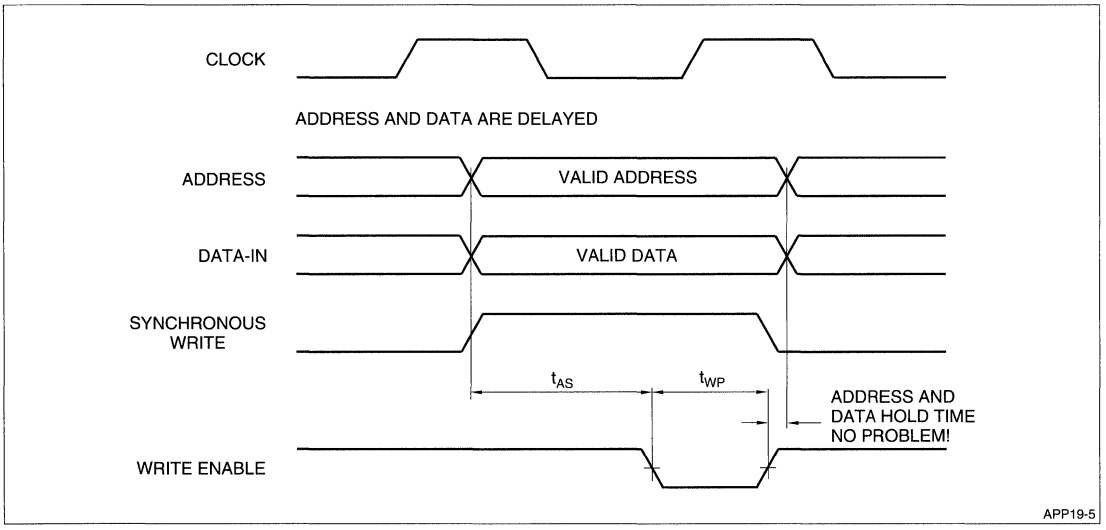


Figure 5. Timing for Improved Circuit

SHARP

Application Note

Pseudo-Static RAMs

256K (32K × 8),1M (128K × 8), 4M (512K × 8)

Bob Laird, Field Applications Engineer

INTRODUCTION

The Sharp LH5P8512 is a 4M bit Pseudo-Static RAM (PSRAM) manufactured on a 0.8 μm CMOS process in one of Sharp's state-of-the-art wafer fabs. The LH5P8512 is the newest member of SHARP's family of PSRAMs that began with the 256K bit LH5P832.

Sharp is committed to expanding its presence in the PSRAM market, which has been growing rapidly over the last few years. Initially developed as a low-cost alternative to a Static RAM (SRAM), the PSRAM has also successfully replaced Dynamic RAMs (DRAM) in certain applications. This application note will discuss the advantages and disadvantages of using a PSRAM over an SRAM or DRAM.

DEFINITIONS

A **DRAM** is a volatile, dynamic memory IC. Volatile means that if it loses power, it will also lose its memory. Dynamic means that it must constantly be refreshed since the memory cell itself utilizes a small storage capacitor that must be constantly recharged. Refresh is accomplished when the memory contents of a row of cells are read by the sense amplifiers, and the logic states that were read are amplified and written back to the cells.

An **SRAM** is a volatile, static memory IC. Static means that it will maintain its memory indefinitely with no external clocking, as long as power is supplied. The memory cell is a simple latch that is either set or reset depending on the data that was written to it.

A **PSRAM** is a volatile, dynamic memory IC. It uses the same memory cell structure as a DRAM. Like a DRAM, a PSRAM must be refreshed, but it incorporates the refresh timing and control logic on-chip to simplify and minimize external logic. Having the refresh logic circuitry on-chip also provides more flexibility to the designer to reduce the power consumption of the PSRAM. The benefit of this is discussed later.

The Pseudo SRAM got its name because it was designed to be offered in the same package and have the same pinout as an SRAM. A PSRAM, while pin-for-pin compatible with an SRAM, would not be a direct drop-in replacement because it must be refreshed.

Schematic diagrams of the memory cell structures for the DRAM/PSRAM and SRAM are shown in Figure 1. For the SRAM, the cell structure shown is the most commonly used by SRAM vendors. There are SRAMs available with 6-transistor cell structures for ultra-low power applications, but they will not be discussed in this application note.

Pinout diagrams for the 4M bit versions of the PSRAM, SRAM, and DRAM are shown in Figure 2.

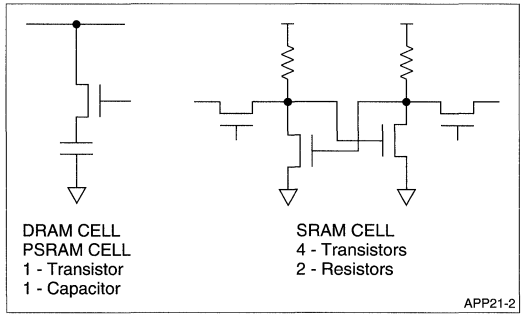


Figure 1. Memory Cell Structures for DRAM/PSRAM and SRAM

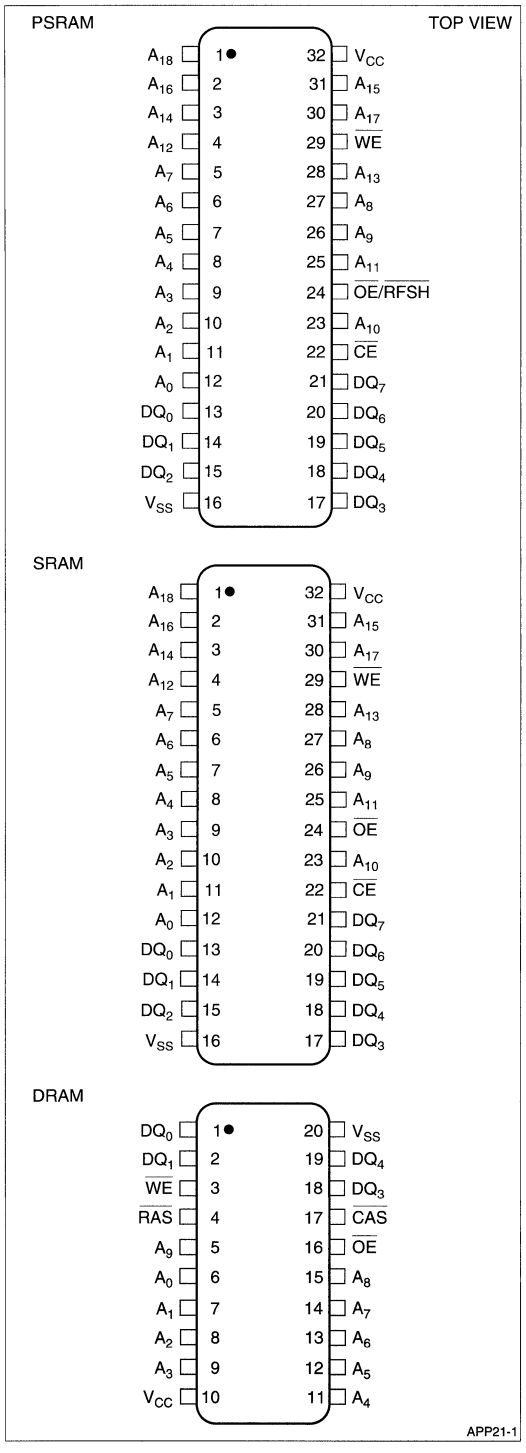


Figure 2. Pin Connections for PSRAM, SRAM and DRAM Packages

ADVANTAGES AND DISADVANTAGES

The major considerations when deciding which memory IC to use in your design are as follows:

1) Cost
2) Density and speed
3) Design ease
4) Power consumption

1) Cost

The PSRAM carries a small-cost premium over DRAM because of its added circuitry. The premium has been 1.2× to 1.4×, but it is expected to decrease as the volume usage of PSRAMs increases.

The DRAM has traditionally been the high-volume memory of choice based on its low cost per bit, high density, and small package. It is ideal for large systems that use many Megabytes of memory.

The SRAM is the most expensive of the three technologies. The memory cell area of an SRAM is much greater than that of the DRAM/PSRAM, resulting in a much larger die size. On a cost-per-bit basis, an SRAM is typically 3× to 4× that of a DRAM. The differences in cell complexity can be seen in Figure 1.

2) Density and Speed

Design engineers selecting the density of their memory ICs must consider what is currently available on the market versus next generation devices. The PSRAM utilizes a DRAM for its base design, so it tends to lag behind the DRAM design introduction by a year or more. Since the DRAM, with its small cell size, is easier to manufacture than an SRAM of the same density, DRAMs and PSRAMs lead SRAMs to market by at least one generation. Today, leading DRAM vendors are just starting small volume production of 16M bit DRAMs while the 4M bit PSRAM is readily available on the market. Leading SRAM manufacturers are just starting to sample 4M bit SRAMs. New DRAMs have historically gone to market with a very high initial cost. The PSRAM tends to go to market with a cost comparable to its corresponding DRAM, which has come down the cost curve, making it attractive for new designs. Compared to an SRAM then, the PSRAM will be one generation ahead in density, and very cost competitive from its introduction.

PSRAMs (and DRAMs) compete against SRAMs in what would be considered the slow-speed classification. The 4M bit PSRAM is offered with access times ranging from 60 ns to 80 ns and cycle times from 110 ns to 130 ns, respectively. Slow 1M bit SRAMs are available with access times of 70 ns or 100 ns. For an SRAM, the access time and cycle time are the same simplifying system timing requirements. DRAMs have speeds comparable to or slightly faster than PSRAMs.

Fast SRAMs, with sub 45 ns access times, fulfill a speed requirement for high-end systems that DRAMs and PSRAMs cannot fill. Therefore, PSRAMs and DRAMs do not compete against fast SRAMs.

3) Design Ease

A DRAM is the most complex device to design with because of its multiplexed addressing and refresh circuitry requirements. The address provided by the microprocessor while it is accessing memory does not multiplex the address. Therefore, decode logic is required between the DRAM and microprocessor to perform the multiplexing function. No such logic is required for PSRAMs or SRAMs.

The refresh specification for the DRAM and PSRAM defines the maximum time allowed between refreshes. This means that both the DRAM and PSRAM require a timer circuit to monitor the refresh time, and also control logic to provide the refresh clocking. Both the DRAM and PSRAM have internal counters that can be used to provide the refresh address. On DRAMs, it is called a $\overline{CAS}$ before $\overline{RAS}$ refresh. On PSRAMs, it is called Auto refresh. In either case, external refresh clocking is required. An example of the refresh timing is shown in Figure 3. This timing represents the refresh of one row internal to the memory. The 4M bit PSRAM requires that 2048 rows be refreshed every 32 ms.

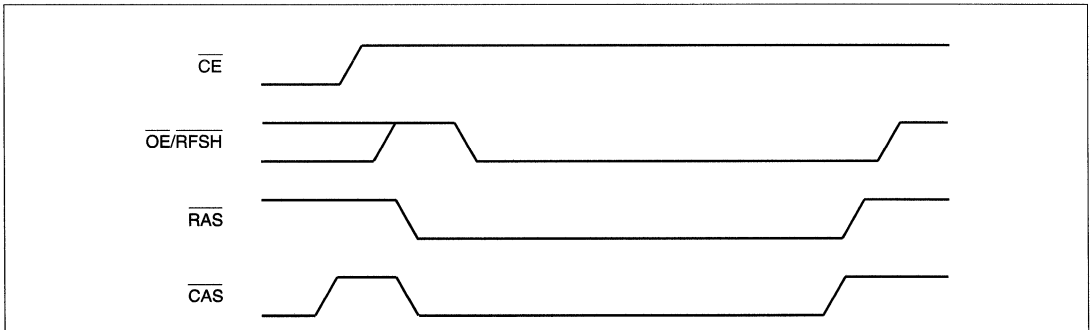
One large advantage that a PSRAM has over a DRAM is a Self Refresh mode. The PSRAM can be put into Self Refresh mode in which an internal clock provides the refresh timing, and no external timing is required. It is very useful during standby mode when memory is not being accessed. The timing for Self Refresh looks the same as that for Auto Refresh except that the OE/RFSH pin must be held LOW for longer than 8 µs. After 8 µs, the internal Self refresh clock begins to automatically refresh the memory array. It will continue to refresh the array as long as OE/RFSH is held LOW with CE held HIGH.

The standard through hole package for the 4M bit PSRAM (and SRAM) is a 32-pin, 600-mil PDIP. For the 4M bit DRAM (1M × 4), it is a 20-pin, 300-mil DIP. The DRAM package is the smaller package, which is preferred because it helps reduce board size. One reason that it has a smaller package is that the DRAM has fewer outputs than the others. The DRAM is organized as 1M x 4 while the PSRAM and SRAM are organized as 512K x 8. The wide width of the PSRAM and SRAM is advantageous for small memory systems. For example, assume that a system required no more than 512Kb of density. If this system used one of the common microprocessors on the market that has a 32-bit data bus, then you would need four 512K × 8 memories to meet the required data width. The same system would need eight 1M × 4 DRAMs to meet the same 32-bit width. Recently, some DRAM vendors announced plans to offer devices with ×8 configurations, targeting small and medium size systems.

A second reason for the smaller package is that the DRAM has almost half the number of addresses as the PSRAM or SRAM. This is because it uses multiplexed addressing. Multiplexing allows you to clock in two different 10-bit addresses on the same address pins to achieve the 20 address bits required to select one of the 1M different memory locations. On the PSRAM and SRAM, you must provide all of the required 19 address bits at the same time to select one of the 512K memory locations.

4) Power Consumption

There is not a significant difference in operating power consumption for a PSRAM, SRAM, and DRAM. The operating power is composed primarily of switching currents from the CMOS peripheral circuits sur-



NOTES:
1. PSRAM auto refresh (cycle $\overline{RFSH}$ pin while $\overline{CE}$ inactive HIGH).
2. DRAM $\overline{CAS}$ before $\overline{RAS}$ (cycle $\overline{CAS}$ active LOW before $\overline{RAS}$).

APP21-3

Figure 3. Refresh Timing

rounding the memory arrays, and those circuits are reasonably similar for all three memories. Differentiation is seen in standby power. The term standby means that no Reads or Writes are taking place. A sample of standby current (I_{SB}) specifications is given below. Note that a DRAM must have its supply voltage maintained within the operating range (4.5 V ~ 5.5 V) during standby, but with SRAMs and 4M bit PSRAMs the supply voltage can be dropped to a data retention voltage (typically 3 V) during standby, significantly reducing power consumption.

	4M Bit PSRAM	4M Bit DRAM	1M Bit SRAM
I_{SB} max @ V_{CC} = 5 V	100 µA	1 mA	100 µA
I_{SB} max @ V_{CC} = 3 V	50 µA	–	50 µA

The standby current given for the PSRAM at V_{CC} = 3 V is actually a self refresh current, so it includes the additional current required to perform the internal refresh. The standby current given for the DRAM does not include refresh current. The typical standby current seen with an SRAM is actually much lower than that for a PSRAM, making SRAMs the best choice for low standby power consumption.

Low-standby power consumption is a critical parameter for battery operated systems such as Notebook computers. A decrease in power consumption directly translates to an increase in battery life. PSRAMs have proven to be an ideal choice for main memory in the Notebooks where low-power consumption is required at a reasonable cost.

For systems requiring a small amount of memory to be battery backed up (usually a small 3 V lithium battery), an SRAM is still the memory of choice, due to its low I_{SB} and static operation.

SUMMARY

Pseudo-Static RAMs have advantages and disadvantages when compared to either SRAMs or DRAMs. The market has shown that the PSRAM can win designs against either, depending on the application. A summary of the advantages and disadvantages is given below.

ADVANTAGE	DISADVANTAGE
Versus SRAM	
Much lower cost	More complex design
Higher density	Higher standby power
Versus DRAM	
Lower standby power	Higher cost
Ease of design	Larger Package
	Lag in time-to-market

INTEGRATED CIRCUITS

MICROCONTROLLER

SHARP'S SM5K5, 4-BIT MICROCONTROLLER

'A CHIP DOES IT ALL' POWER MANAGEMENT, BATTERY CHARGING, AND MONITORING FUNCTIONS

Vijay Desai, Field Applications Engineer

INTRODUCTION

This application note describes various features/benefits of SHARP's 4-bit microcontroller – 'A chip does it all' power management, battery monitoring, and charging functions. It is an ideal chip for a handheld, compact battery operated computers, instruments, and other consumer devices.

The SM5K5 is a CMOS 4-bit low voltage operating (2.2/5.5 V) single-chip microcontroller which integrates 4-bit parallel processing functions 2 kbit ROM, 128 × 4 RAM, 10-bit A/D, timer/ counter, and provides three types of interrupts and four levels of subroutine stack. Refer to Table 1.

SYSTEM CONFIGURATION

Please refer to the Block Diagram in Figure 1.

Arithmetic Logic Unit (ALU) and Carry Flag (C)

By instruction, the 4-bit parallel processing performs binary addition at the ALU when linked together with the RAM, C flag, and accumulator (A) register.

A-Register and X-Register

The A-register (accumulator) is a 4-bit general purpose register that is linked together with the ALU, C flag, and RAM and is used primarily to transfer numeric values and data. Data transfer with the various I/O terminals are primarily performed by the A-register.

The X-register (4-bit) can be used as a temporary register. By instruction, it can load itself with the contents of the A-register or exchange its contents with that of the A-register.

Table 1.
SM5K5 Features and Benefits

FEATURES	BENEFITS
On-chip A/D converter Resolution: 10 bit Channels: 4 Error: ±2.5 LSB	Provides a super fine voltage control to prevent an over charging of secondary batteries.
On-chip 8-bit Timer/Counter with 15-bit divider	Provides cost + space savings.
On-chip ROM (2 K × 8) and RAM (128 × 4)	Provides for programming flexibility.
50 instruction sets, 3 interrupts, and 4 levels of subroutines nesting with 2 ms/inst. cycle @ 3.0 V	Provides tremendous programming flexibility and easy to implement real time applications.
Operating Speed: 1 to 61 µs (2 MHz to 32.768 kHz)	Ideal for power efficient applications.
24 I/O ports	Gives tremendous flexibility and cost savings.
Low voltage operation: 2.2 V to 5.5 V	Provides for longer battery life.
Extended operating temperature: –20 to 70°C	Provides a rugged environment.
Compact Packages: 32-pin SOP and 36-pin QFP	Provides optional space savings.

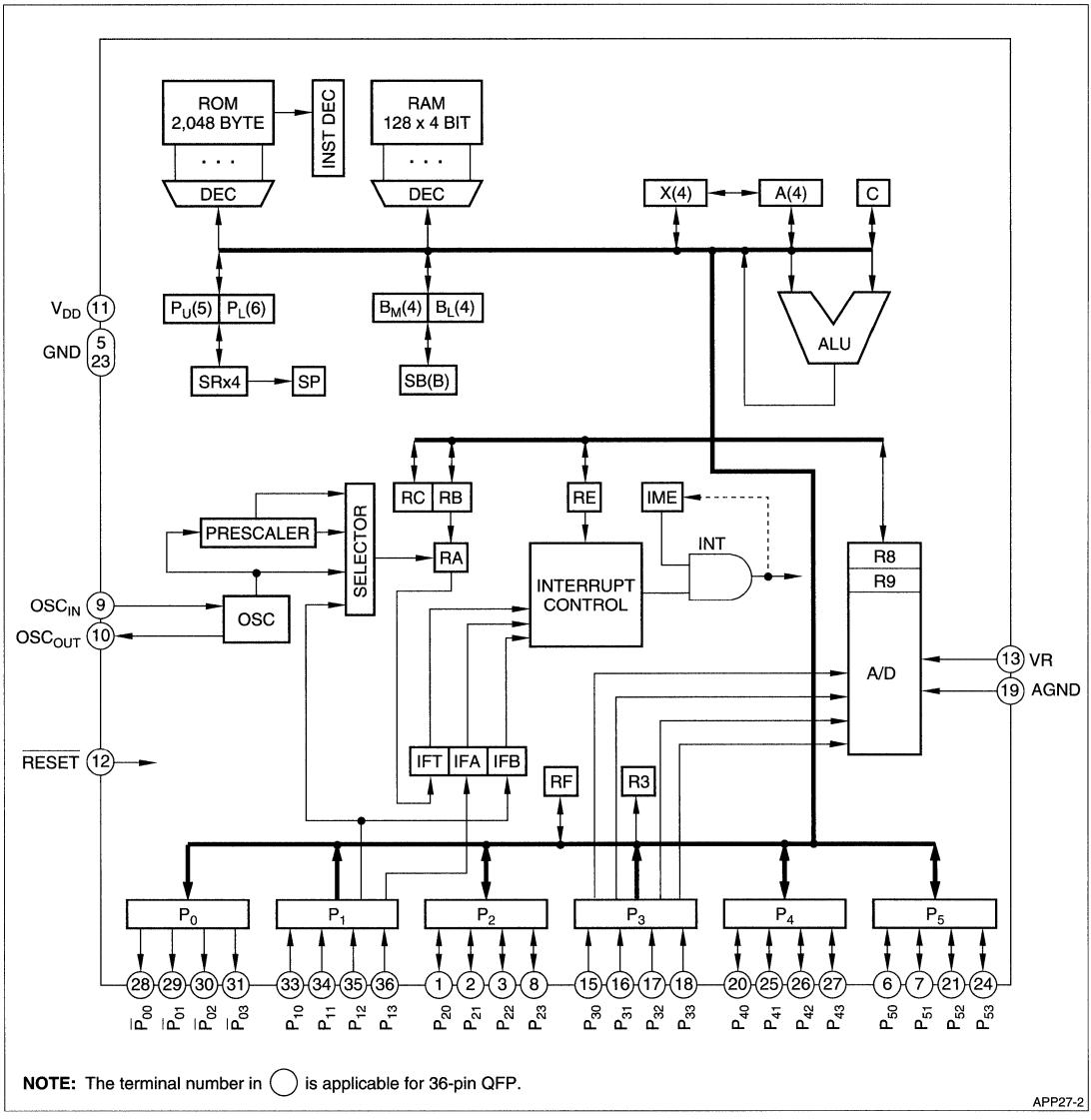


Figure 1. SM5K5 Block Diagram

Program Counter (PC) and Stack Register (SR)

A ROM address is specified by the PC. The PC is comprised of 12 bits, where 6 bits (Pu) are used to specify the step. Pu is a register and PL is a binary counter.

The Stack Register (SR) comprises four levels to allow a maximum subroutine nesting of four levels. During the subroutine jump, the contents of the PC are pushed into SR, and the stack pointer (SP) is automatically incremented. When returning from a subroutine, the contents of SR are popped to the PC and the SP is automatically decremented.

Program Memory (ROM)

The ROM is use for program storage. The ROM capacity is 20 pages × 64 steps = 1,280 steps. The ROM is organized into 16 pages, where one page is organized into 64 steps.

Data Memory (RAM)

The RAM is used for data storage. The RAM capacity consists of 16 × 4 × 4 = 256 bits.

Mode Registers

These registers – commonly referred to as 'mode registers' – control the functions of the SM5K5 and also serve as a counter/timer. Both R8 and RB are 8-bit mode registers; R3, RC, RE, and RF are 4-bit mode registers.

System-Clock Generator and Divider

The fundamental frequency which is input through OSCin and OSCout is divided into two in order to generate the system-clock frequency. System-clock period is the same as the instruction execution cycle. However, the instruction's execution cycle of two-word instructions is twice that of a one-word instruction.

FUNCTIONAL DESCRIPTION

Reset Function

Reset function initializes the SM5K5 system. When the input on the RESET pin goes LOW, the system enters reset condition after two command cycles. Reset causes the following changes:

1. I/O terminals are set to input terminals.
2. All mode registers are reset.
3. Output latch register PO is reset, causing its four terminals to go HIGH.
4. Interrupt-request flags (IFA, IFB, IFT) and the interrupt-master-enable flag (IME) are reset, disabling all interrupts.

Standby Feature

The Standby feature saves power and is categorized into 'Stop mode' and 'Halt mode,' one of which is selected at a time according to the operating condition.

When the STOP command is applied, the chip stops the master clock and forces the entire system into Stop mode. In Halt mode, since the divider is operating, the timer can run during the Halt mode. This mode is used when it is necessary to start the system quickly upon removal of the Standby mode.

Interrupt Feature

The SM5K5 allows three types of masking interrupts. An interrupt is triggered by the timer or external input, such as from the keyboard to process the program periodically. Control of interrupt enable and disable is made through an operation of the mode register RE (serves as mask flag) and the master-enable flag IME. The interrupt processing starts when IME is set to 1 and a bit in the mode register RE is set, and the corresponding interrupt event has occurred.

Timer/Counter

The SM5K5 has a pair of built in timers/counters. They are used to handle interrupts, counting, and as an interval timer. The timer/counter consists of an 8-bit count register RA, a modulo register RB, a 15-bit divider, and a 4-bit mode register RC.

A/D Conversion

There is a built-in 10-bit A/D converter having four channel multiplexer analog inputs. The A/D converter operates in A/D conversion mode and comparison mode. In the A/D conversion mode, the analog input from the P3 terminal is converted to a digital value. In the comparison mode, it compares the input analog amplitude with that of a reference voltage that is set inside of the SM5K5.

The P30-P33 terminals can be used as analog voltage inputs. One or more of these four inputs can be set to assume A/D terminal by the bit operation of the mode register R3. The A/D converter is controlled by the bits set in the mode register R8.

An Application Example

The SM5K5 is a very low-cost, low-power (2.2 V to 5.5 V) single-chip microcomputer that could be used in a variety of applications, including handheld PDA/PCAs, battery-operated medical instrumentations, computers, and many consumer products such as smart toys, coffee makers, etc.

A typical system configuration using the SM5K5 for battery charging, monitoring, and power management in computer application is shown in the Figure 2.

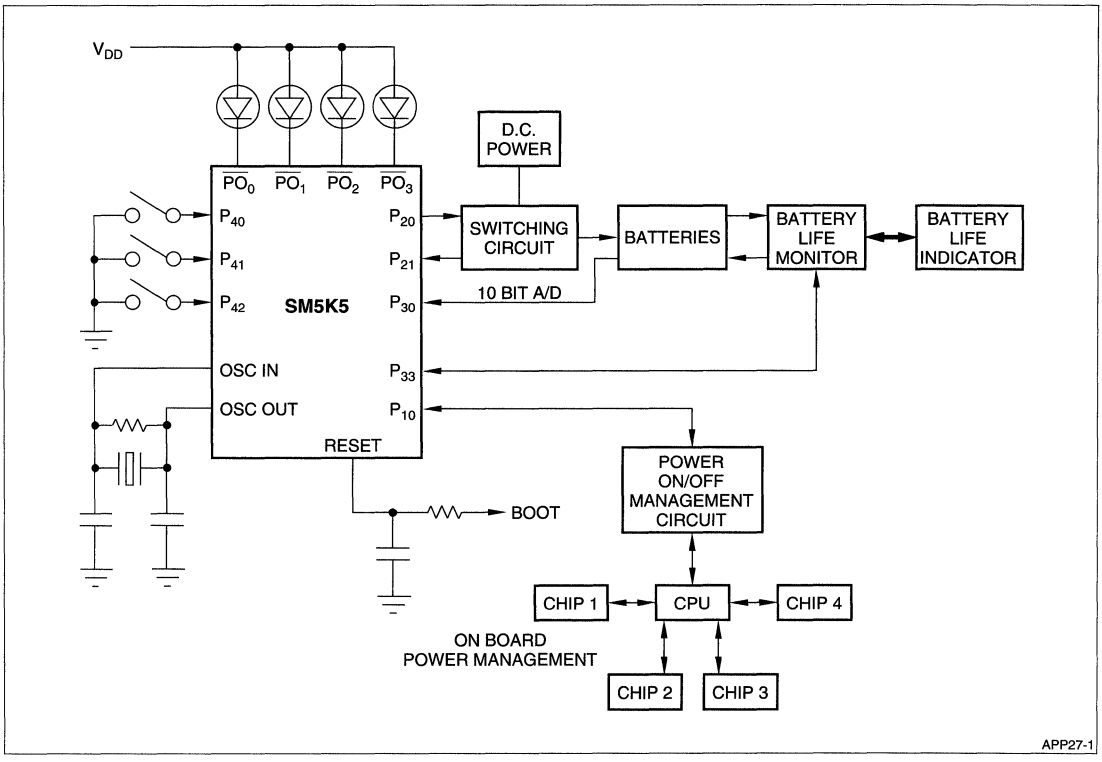


Figure 2. Battery Charger, Battery-Life Indicator

LIQUID CRYSTAL DISPLAYS

POWER-ON SEQUENCING FOR LIQUID CRYSTAL DISPLAYS; WHY, WHEN, AND HOW

Charles Guthrie, Field Applications Engineer

ABSTRACT

Liquid Crystal Displays require special considerations upon power-up. The following explains why power-up sequencing is necessary, the required timing of the power-up sequence, and circuitry necessary to accomplish proper power-up sequencing. Each LCD display technology is discussed individually.

INTRODUCTION

Liquid crystal displays are being used more often as the display of choice in computer applications, most notably in laptop, notebook, and pen based computers. These displays are popular because they are reliable, power efficient, compact, light weight, and easily installed in hardware. Currently, one of the most overlooked design considerations is the required power on sequencing. Liquid crystal displays vary from simple one line monochrome units, up to full page graphic displays having both monochrome and color formats. This paper will discuss the reason power-up sequencing is necessary, ways to accomplish the required sequencing in both passive and active matrix displays, and provide examples of circuitry to power up and down the display properly.

POWER-ON SEQUENCING IS NECESSARY

Power-on sequencing is required mainly to protect the liquid crystal from exposure to any DC voltage. What needs to be accomplished upon powering up the display is that the 5 V must be started first. This allows the onboard logic to become active and starts an internal clock (the M clock) which sets up an AC wave form on the display electrodes. Even with very short intervals of exposure to VEE, without the "M" clock started first, the liquid crystal will begin to break down and change state. This change of state manifests itself as a change in color of the liquid crystal and eventually, the formation of gas bubbles. When this happens, the damage is permanent and the display will eventually be rendered useless. This AC wave form, the "M" clock, is controlled using a phase lock loop circuit. When VCC is started, this clock begins oscillating and resonates between VCC and VEE. As the voltage for VEE is increased, the display gains contrast. The contrast on the display is optimized by adjusting VEE.

After VCC has stabilized, the external clock and data signals should be introduced to the display module. After the clock and data signals are stable, VEE can be turned on. One word of caution, if you are using a display panel with display enable, there is a time limit as to how long the display can be function without VEE. For intervals longer than 20 ms, the display logic may latch up thus requiring a full reset. In order to fully understand the reasons for a specific power-up sequence, each display type will be discussed and the sequencing requirements for that display will be fully explained.

CHARACTER MODULE

Sharp manufactures a line of Dot Matrix Character Modules including 16 x 1; 16 x 2; 20 x 2; and 40 x 2 displays. Because these displays operate from a single power supply, power-up sequencing is not as critical as in other units. However, care must be taken to insure that the power supply is up to at least 4.5 V within 10 ms of turn on. If the display has not reached this level prior to latching the internal reset, the reset circuit will not operate normally and may latch up the display. Also, if power is interrupted, the display power must be held off for at least 1 ms prior to reapplying power or the display may once again latch up due to failure to trigger the internal reset circuit (Figure 1).

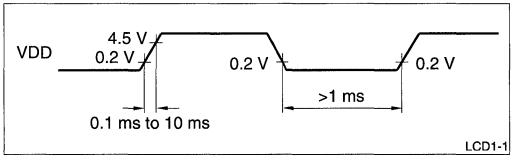


Fig. 1. Power-up Sequencing for Character Modules

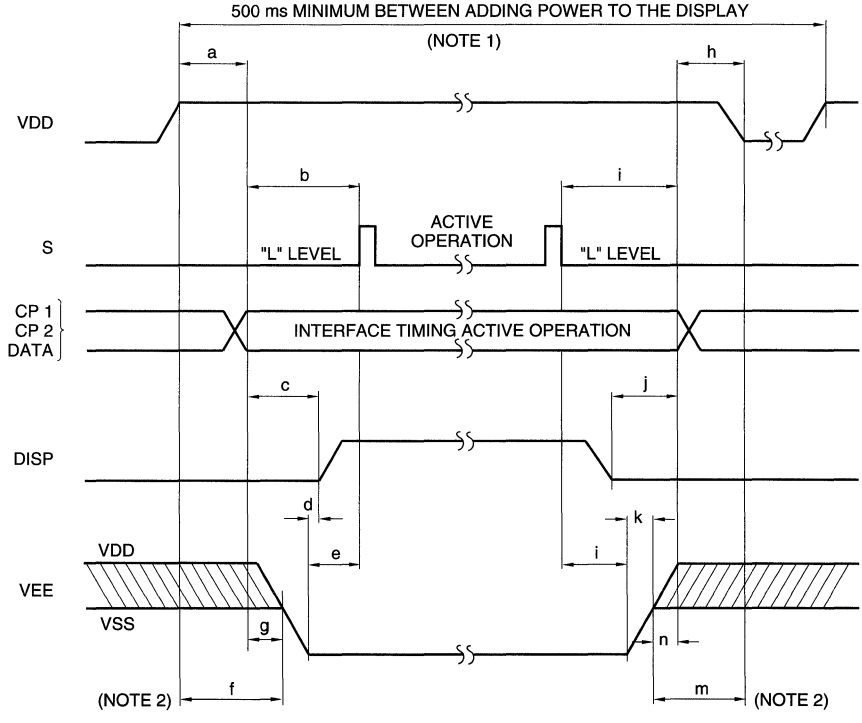
Upon power-up, you must reset the display in software. Instructions as to how to accomplish this are published in the Sharp "Display Unit Users Manual for Dot Matrix LCD Units with Built in Controllers."

PASSIVE DISPLAYS

Large area passive displays, both Monochrome and Color, require care in implementing the power-up sequencing of the display. Because the operating voltage of the display is higher than the DC breakdown voltage of the fluid, care must be taken to provide protection for the liquid crystal fluid. The objective of the power-up sequencing procedure is to start an internal clock, the "M" clock, and to insure stable operation of the CMOS circuitry by applying VDD prior to introduction of the logic signals. By following the recommended start up procedure, we can prevent the device inputs from being subjected to voltages greater than +0.2 VDC above VDD. If the logic is allowed to be exposed to voltages exceeding VDD +0.2 V, noise will disperse through the drivers and possibly cause lockup of the logic.

The "M" clock sets up an AC wave form which prevents DC current from flowing through the Liquid Crystal. According to the fluid manufacturer, DC voltages as low as 25 mV for even short periods of time will cause a breakdown of the liquid crystal material and eventually render the display useless. The recommendations outlined in the attached timing diagram will assure the user that the display will be fully protected from damage and premature failure.

Display Types which use the above power sequencing are: LM64P70 Series; LM64P80 Series; and the LM64C031 as well as the older Passive Matrix Displays. (Figure 2)



NOTES:

1. If during the power up sequence, the power is interrupted to the display, power must remain off for at least 500 ms, measured from the original initialization of VDD, or damage may occur to the display drivers. This time interval is necessary to allow the CMOS logic to return to a ground potential and stabilize prior to restarting the logic.
2. VEE can be at either VDD or VSS prior to turn on. If display enable is not used, then VEE must be at VDD until activation. This allows the pull up attached to display enable to hold the panel reset until activation of VEE. This will prevent the display from latching up.

LCD1-2

Fig. 2. Power-up Sequencing for Passive Displays

Table 1. Power-up Sequencing for Passive Displays

POWER ON			POWER OFF		
SYMBOL	WITH DISPLAY ENABLE	WITHOUT DISPLAY ENABLE	SYMBOL	WITH DISPLAY ENABLE	WITHOUT DISPLAY ENABLE
a	0 ms MIN	0 ms MIN. 20 ms MAX.	h	0 ms MIN.	0 ms MIN. 20 ms MAX.
b	0 ms MIN	20 ms MIN.	i	0 ms MIN.	20 ms MIN.
c	20 ms MIN	—	j	20 ms MIN.	—
d	0 ms MIN	—	k	0 ms MIN.	—
e	—	0 ms MIN.	l	—	0 ms MIN.
f	0 ms MIN	—	m	0 ms MIN.	—
g	—	0 ms MIN. 100 ms MAX.	n	—	100 ms MIN.

ACTIVE MATRIX DISPLAYS

Sharp Electronics is presently involved in manufacturing Thin Film Transistor (TFT) Active Matrix LCDs for use in Office Automation Applications and Audio Visual Applications. Each of these technologies require specific power-up sequencing to insure reliable operation of the displays. The following summarizes those requirements.

Small TFT Audio Visual Displays

Small TFT audio visual displays include the LQ4xxxx and LQ6xxxx displays. These displays, like the Passive units are susceptible to the introduction of the negative voltage (VSL) prior to initializing the logic voltage (VSH). The following diagram graphically represents the necessary timing considerations for these modules. (Figure 3)

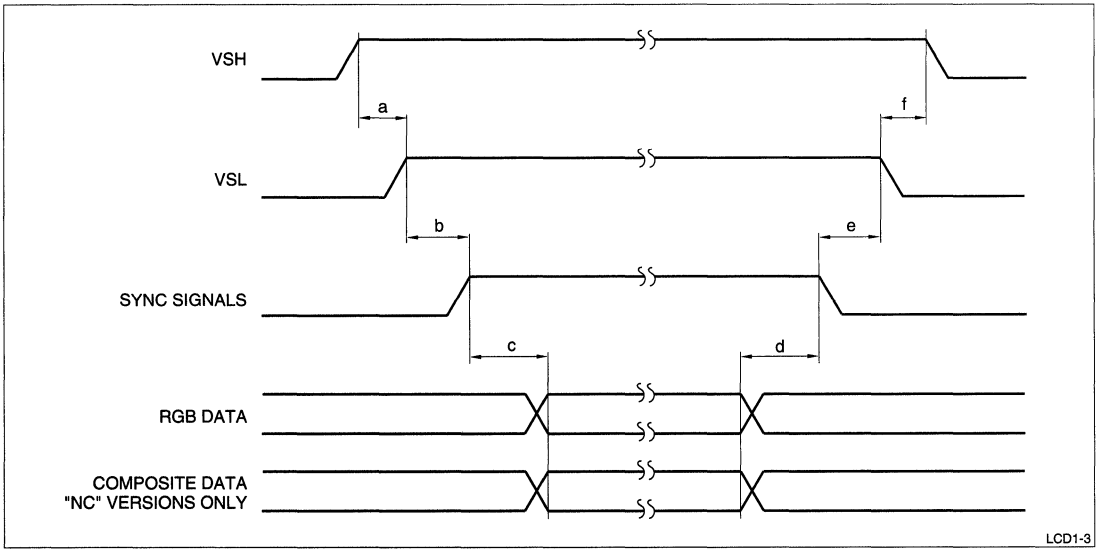


Fig. 3. Power-up Sequencing for Audio Visual TFT Displays

Table 2. Power-up Sequencing for Audio Visual TFT Displays

POWER ON			POWER OFF		
SYMBOL	MINIMUM	MAXIMUM	SYMBOL	MINIMUM	MAXIMUM
a	0 ms	100 ms	d	0 ms	20 ms
b	2 ms	20 ms	e	0 ms	20 ms
c	0 ms	20 ms	f	0 ms	100 ms

TFT Displays Without Controllers

In addition to the above mentioned audio visual displays, Sharp has a few "Special Application" displays such as the LQ323Y11 which do not come with a display controller. What makes these displays unique is that separate voltages are required for the source and gate drivers. In all instances, the "High" level voltages should be started prior to the "Low" level voltages. This means that VSH (+5 V typ.) and VGH (+13 V typ.) should be started prior to VSL (-8 V typ.) and VGL (-20 V typ.). The minimum and maximum timing requirements specified above can be used for the TFT displays without controllers by applying the standard for VSH in the table to VSH and VGH on the display and the standard for VSL in the table to VSL and VGL on the display. All other timing for the display will also be as represented in the table.

Large TFT Office Automation Displays

The original offering in office automation products were the LQ10 and 011 products. These displays require multiple power supply inputs of +5 VDC at 0.45 watts and +12 VDC at 4.68 watts. As with other displays, it is necessary to power up the displays in the proper sequence to prevent latch up of the display logic. Latch up of the display causes a failure of the "M" clock to start and subjects the liquid crystal to DC Voltages. If latch up occurs, the liquid crystal itself will be electrochemically decomposed thus permanently damaging the display. The timing diagram (Figure 4) must be implemented in order to assure that the display will not be damaged.

The newest offering in large format TFT displays is the LQ9D011. With this display, there is a requirement for only one power supply. Resident circuitry divides the single supply into the various voltages required by the panel. The only requirement for this display is that the power must be present prior to, or simultaneously with the input signals. If not, the display may latch up. The circuit used for the LQ10 TFTs should be implemented on the LQ9 display, to supply power when the input signals are present, and to remove power if the HSYNC or CLK are interrupted. Because of the display configuration it will only be necessary to connect VDD in Figure 5 to the +5 V input on the display. The +12 V is not necessary on the LQ9 displays.

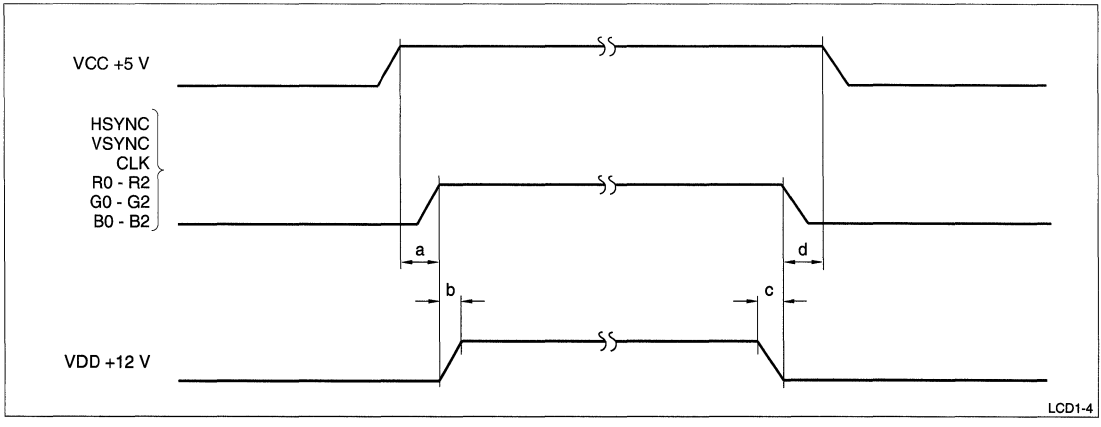


Fig. 4. Power-up Sequencing for LQ10xxxx Office Automation Displays

Table 3. Power-up Sequencing for LQ10xxxx Office Automation Displays

POWER ON			POWER OFF		
SYMBOL	MINIMUM	MAXIMUM	SYMBOL	MINIMUM	MAXIMUM
a	0 ms	100 ms	c	0 ms	20 ms
b	2 ms	20 ms	d	0 ms	100 ms

Power-on Sequencing Circuitry

The following circuit (Figure 5) is set up to sequence the power supply on the Sharp LQ10D011 display. The circuit senses clock and horizontal sync, energizes the power supplies whenever these signals are present. If for any reason the HSYNC or CLK signals are interrupted, the relay will open thus interrupting VDD and VEE to the display. This will prevent any possibility of damaging the display. The circuit is set up to open the power supply line if the signals are lost for more than 28 ms. The power will be restored if the two signals return.

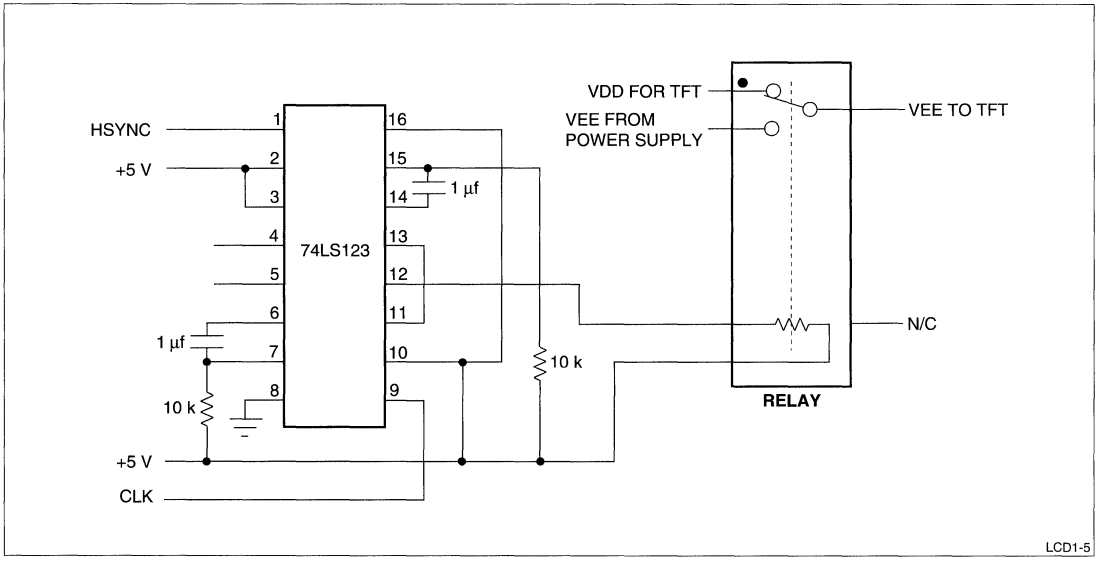


Fig. 5. Power Supply Sequencing Circuit for LQ10xxxx TFT Displays

Power Supply Sequencing for Passive Displays

Figure 6 is set up to sequence the power supply on the Sharp passive displays including monochrome displays and the LM64C031 passive color display. The circuit senses upper and lower data clocks and energizes the power supplies whenever these signals are present. If for any reason the clock signals are interrupted, the relay will open thus interrupting VDD and VEE to the display. This will prevent any possibility of damaging the display through the introduction of Direct Current to the liquid crystal material. The circuit is set up to open the power supply line if the signals are lost for more than 28 ms. The power will be restored if the two signals return.

NEW GENERATION CONTROLLERS

Chips and Technologies Controllers

The new generation controllers include signals specifically to control the sequencing of the display power. In their new Vampire board with the 65520/530 chip set, Chips & Technologies supplies ENAVDD and ENAVEE signals to control the VDD and VEE voltages respectively. The signals are intended to be control lines which drive external power devices sized to handle the required current. Figure 7 is a diagram of the required circuitry and the timing provided by the controller. A contrast adjustment circuit is illustrated in the Chips and Technology literature and may be added to this circuit diagram as required. The contrast circuit has been removed from this example for clarity.

In the 65520, the timing for VDD and VEE on and off is fixed, however, in the 65530, the timing is programmable using the POWER SEQUENCING DELAY REGISTER (XR5B). The power-on delay time in the 65520 is set at 32 ms for each interval, a, b, c, & d. In the 65530, the interval default is 32 ms, however, the delay may be programmed up to 60 ms in 4 ms intervals on the power-up cycle, and up to 480 ms in 32 ms intervals on the power-down mode. (Figure 8)

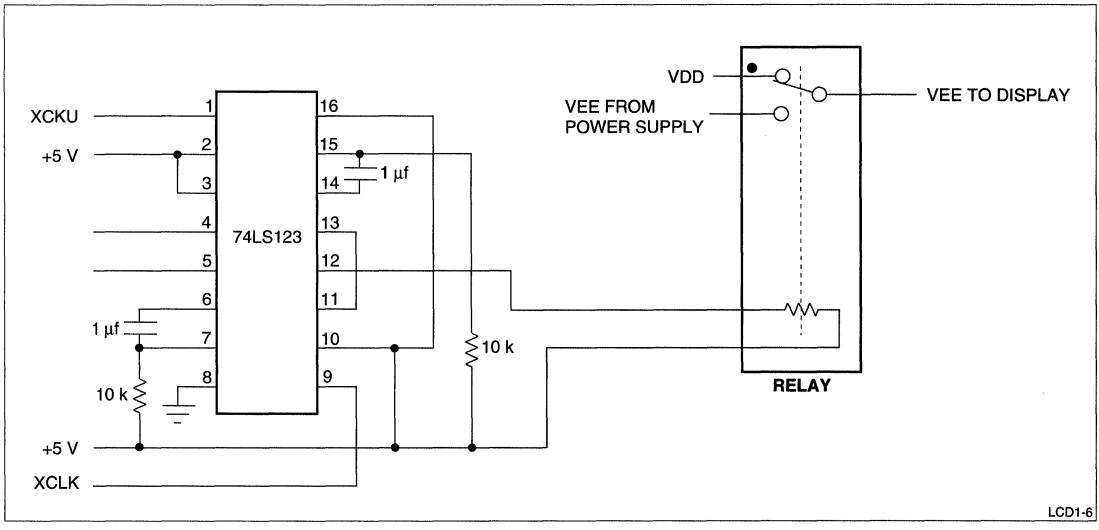


Fig. 6. Power Supply Sequencing Circuit for Passive Displays

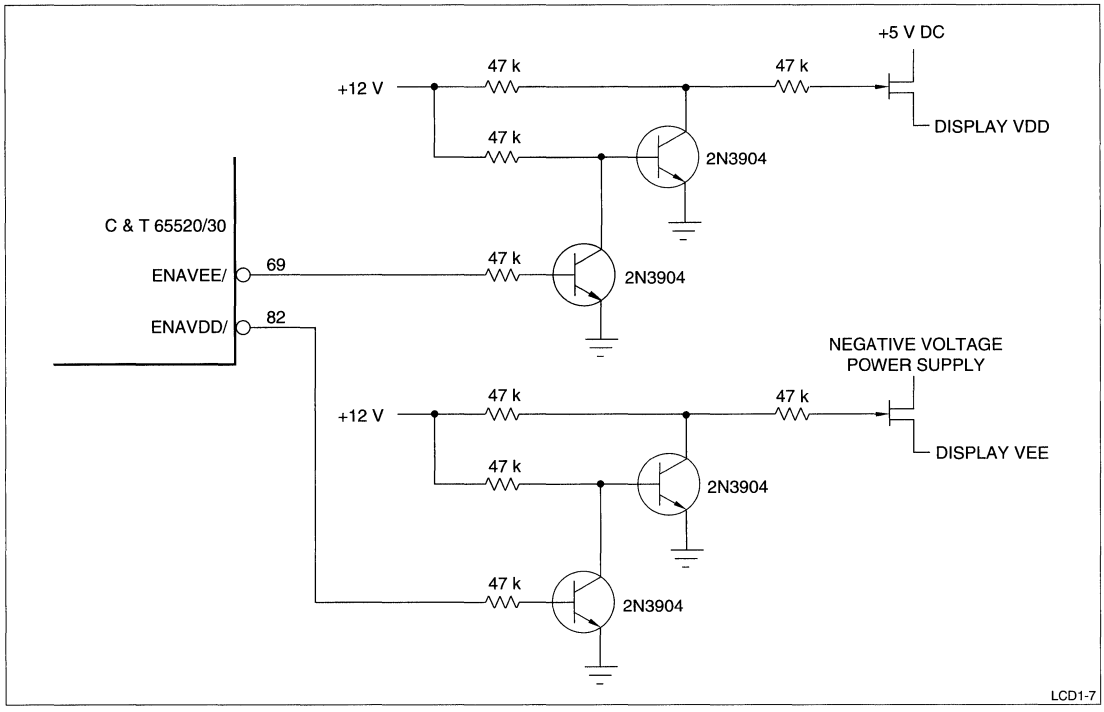


Fig. 7. Chips and Technology Power-up Sequencing Circuit 65520/30 Demo Board

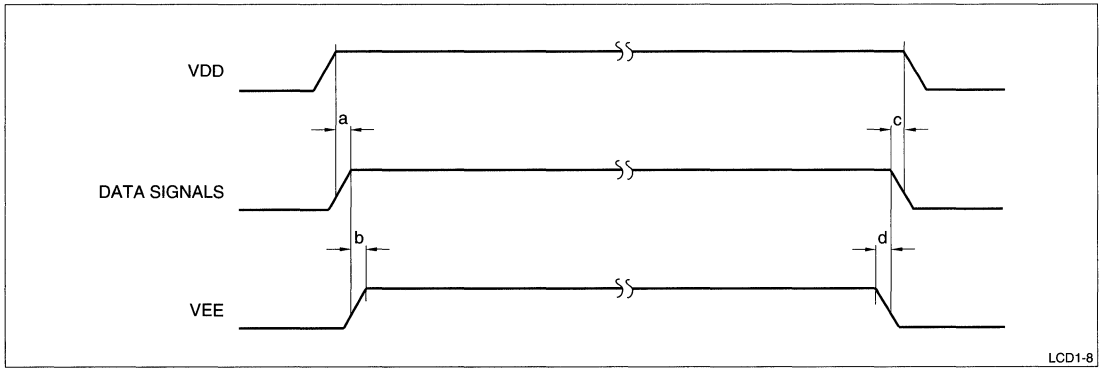


Fig. 8. Chips and Technology Power-up Sequencing Circuit 65520/30 Demo Board

Cirrus Logic Controllers

Cirrus Logic also supplies FPVCC and FPVEE signals from their GD6340 controller. The controller manages the power up/down sequencing completely. The only additional components necessary are the actual power devices as illustrated in Figure 9, and are sized to handle the required current.

The Cirrus Logic GD6340 has logic to correctly sequence the power-on and off in LCD applications. The FPVCC and FPVEE pins are simply CMOS outputs which supply control signals intended to drive external power management components. Sample circuitry used to accomplish the proper power management is illustrated in Figure 9. A contrast control circuit may be added to the above circuitry. Please refer to the Cirrus Logic documentation for recommended schematics for this circuit. The timing (Figure 10) for intervals a, b, c, and d, are all set at 37 ms in the Cirrus Logic controller.

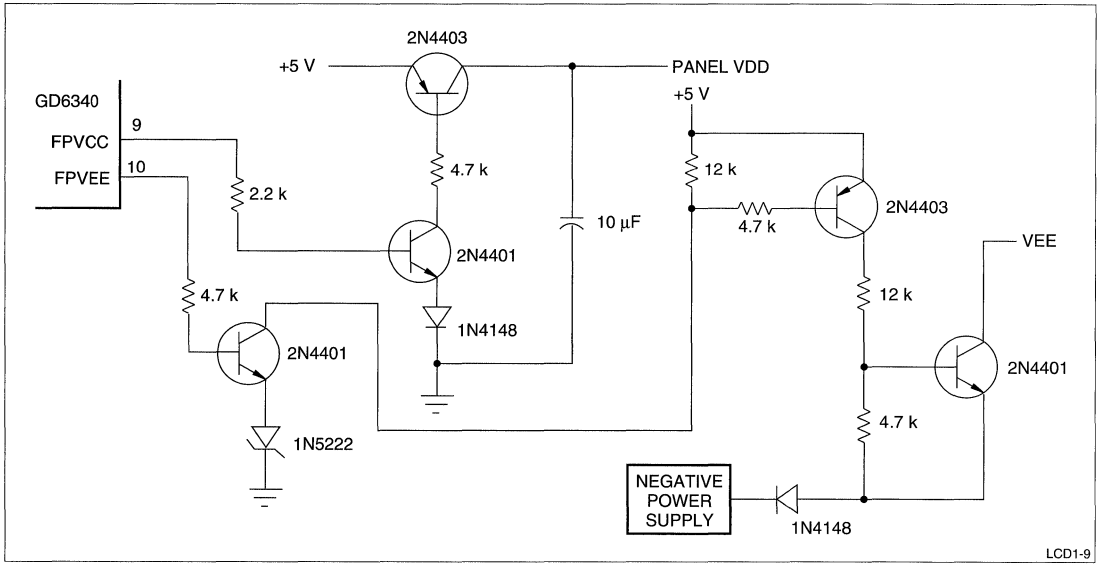


Fig. 9. Cirrus Logic Power-up Sequencing Circuit GD6340 Demo Board

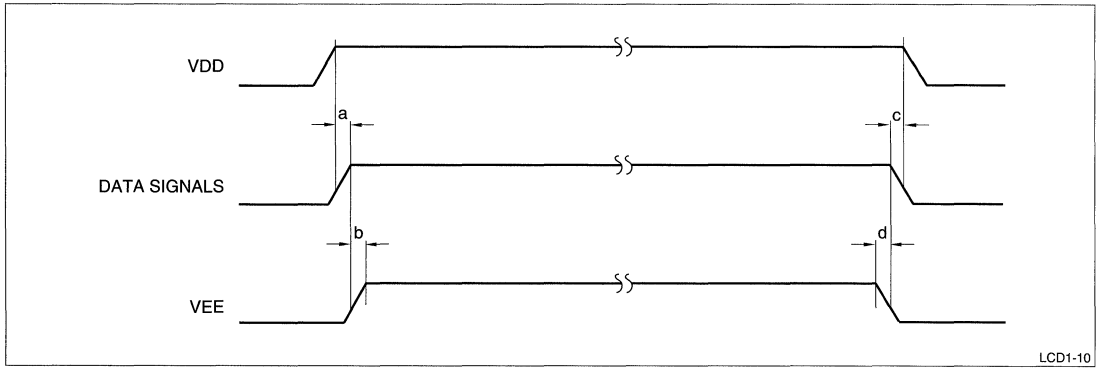


Fig. 10. Timing for Cirrus Logic GD6340 Power-up Sequencing

Western Digital Controllers

Within the Western Digital chip set, there is a signal (LCDN) dedicated to controlling the Power-on/off sequencing for the LCD. In their demonstration board, the signal is routed through the circuit (Figure 11) to enable control of power initialization and to delay the removal of power until the logic signals have been turned off. The intervals a, b, c, and d, are set to be approximately 30 ms (Figure 12). Contrast adjustment can be added to the circuit.

CONCLUSIONS

Power-up sequencing of Liquid Crystal Displays is critical. This application note is intended to express the importance of this function and to make suggestions as to the required implementation. All of the major controller chip manufacturers have recognized the importance of this feature and have implemented logic to assist the designer in incorporating this feature into his hardware.

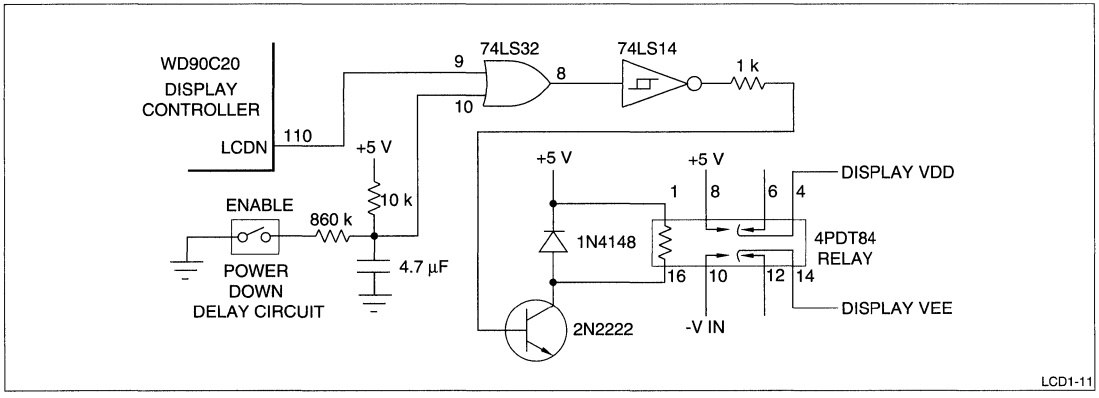


Fig. 11. Western Digital Power-up Sequencing Circuit WD90C20 Demo Board

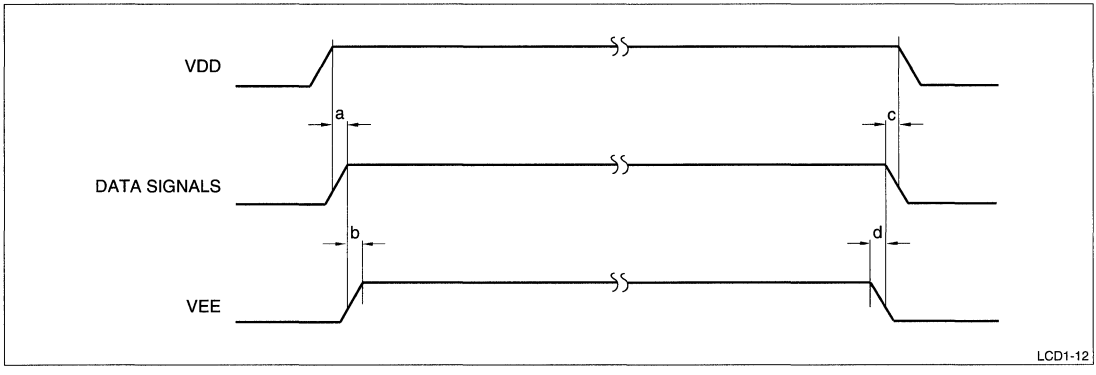


Fig. 12. Timing for Western Digital WD90C2X Power-up Sequencing

AN EXAMINATION OF ACTIVE MATRIX TECHNOLOGIES AND COMPONENTS

Harry Miller, Field Applications Engineer

OBJECTIVE

The purpose of this paper is to examine the different active matrix technologies and review the theory behind the different technologies because breakthroughs in the different technologies are happening every day. This paper will review the technologies, discuss some of the problems, and propose some solutions brought about through various research.

INTRODUCTION

Flat panel displays required a technology breakrough, not only to compete with CRTs but more importantly to provide applications that did not and do not exist with CRTs (i.e., color laptop computers, portable T.V.s, and helmet mounted flight simulation devices to name a few). The design criteria was to achieve a display comparable to a CRT, utilizing matrix addressing, while overcoming crosstalk and offering cost effective solutions. The technology that allowed this breakthrough is active matrix technology.

BACKGROUND

Before an examination of the different technologies is possible, an understanding of addressing and crosstalk is required.

Addressing

Direct addressing is the hard wiring of each pixel to driver amplifier. In a 640 x 480 resolution display it would require (640 x 480) 307, 200 driver amplifiers. This method is very expensive because of the number of drivers and physically, the driver circuitry would be very large. Alternative methods of driving the displays had to be found. The method used for driving most flat panel displays is matrix addressing. In matrix addressing each pixel is electronically connected between one row lead and one column lead, therefore the number of driver amplifiers is equal to the number of rows plus the number of columns. In a 640 x 480 resolution display the number of drivers can be reduced from 307, 200 to 1,120 a significant reduction in space and expense. In order to use matrix addressing the display media must posses a strong nonlinear characteristic to prevent partial selection of non-addressed pixels (crosstalk). A nonlinear component can be added; usually a transistor or diode, an active nonlinear device. [1]

Crosstalk

There are several causes of crosstalk; including leakage current, capacitive coupling between address lines, and line delay.

Leakage current

While matrix addressing reduced the number of drivers that are needed it also means that a single driver will drive more than one pixel. A single driver may drive one or more rows or columns. This means that even nonselected pixels will have a voltage on a row or column. This voltage is less than the voltage required to turn the pixel on. However if the active device (transistor or diode) has leakage current, the unaddressed pixel may turn on, causing crosstalk. The leakage current is often expressed as the ratio of Ion/Ioff.

Capacitive coupling

There are two causes of capacitive coupling between address lines. One is the parasitic capacitive coupling between data lines. The second source of capacitive coupling is the high speed of the driving wave form used in matrix addressing. The higher the speed, the greater the probability of capacitive coupling. [2]

Line delay

There are two major causes of line delay. One is the RC line delay between gate terminals of adjacent pixels. The other is the resistance of the LCD cell itself. Figure 1 is a representation of the resistance in the LCD cell.

Rc = represents the resistance across the LCD cell

Rc^1 = represents the resistance through the ITO layer of a LCD cell

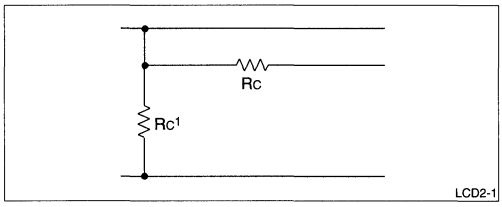


Fig. 1. LCD Cell Resistance

Measurement Method

A method of measuring crosstalk is illustrated below. A brightness measuring device is used to measure the brightness Ya with all the pixels on (white). A pattern (usually a box) is then displayed on the screen and the same spot is measured Yb (Figure 2). The cross modulation ratio can then be measured using the following formula:

$$\text{Cross modulation ratio} = \frac{|Yb/Ya|}{Ya} \times 100$$

Ya = brightness of area without pattern (cd/m)

Yb = brightness of area with pattern (cd/m)

Ya and Yb must be measured at the same spot.

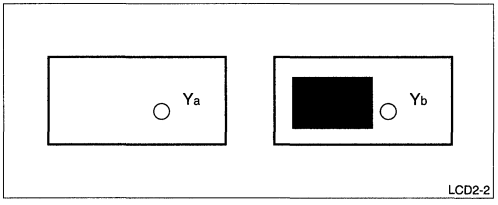


Fig. 2. Method for Measuring Crosstalk

EVALUATION

As we evaluate the different technologies, keep in mind the design criteria: to achieve a display comparable to a CRT, utilizing matrix addressing and good grayscales, while overcoming crosstalk and maintaining cost.

Diode Devices

1. Two Terminal Devices:

A. a-Si Back-to-Back Diodes (Figure 3)

The advantages are:

- Ion/off 10^8 (lower leakage current, improved crosstalk)
- Simple Process (Cost)

The difficulties are:

- The breakdown voltage is neither reproducible or reliable, making large screen uniformity and grayscales very difficult.

B. a-Si Back-to-Back Ring Diodes (Figure 4)

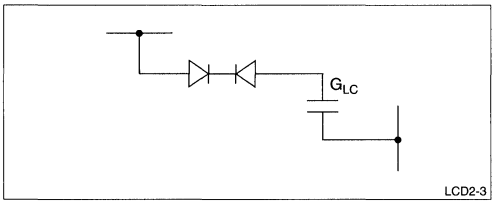


Fig. 3. a-Si Back-to-Back Diodes Equivalent Circuit

The advantages are:

- Same as a-Si Back-to-Back Diodes

The difficulties are:

- One diode, low-voltage threshold voltage
- More diodes, more complex
- Voltage not optimized to liquid crystal threshold

C. Metal Insulator Metal (MIM) (Figure 5)

The advantages are:

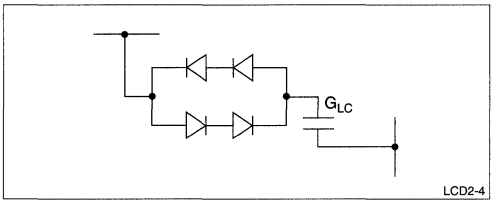


Fig. 4. a-Si Back-to-Back Ring Diodes Equivalent Circuit

- Fewer Masks (2-3) (Cost)
- Light insensitive (TaO MIM)
- High Resolution
- No Crossovers

The difficulties are:

- Crosstalk
- Tight Tolerances: Voltage and temperature (can influence image sticking), insulator thickness (can influence panel uniformity), device area and material issues.
- Complex Cover Plate

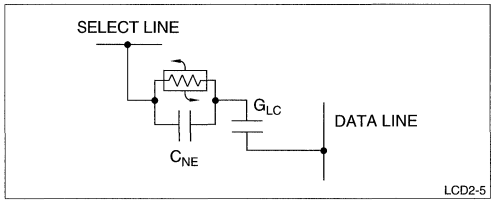


Fig. 5. MIM Equivalent Circuit

MIM LCDs were the first active matrix displays. The MIM is electrically equivalent to a bidirectional zener diode. The bidirectional aspects allows the display to be driven by symmetrical square waves with no DC ie., component.

In addition there is lateral MIM technology. Lateral MIM technology uses several lateral MIM devices per pixel. The results are high density per pixel (better resolution), high contrast ratio over a wide temperature range, a redundancy in technology for pixel defects and symmetrical I/V characteristics. [5]

3. Three Terminal Diodes: D^2R (Figure 6)

Advantages:

- Charged storage
- Diodes are small
- Redundancy, series/parallel
- ON/OFF approximately $1O^8$

Disadvantages:

- 2X row interconnect lines (Increase number of drivers are necessary)
- Poor storage capacitors
- Complicated color plate back plane
- Column lines and ITO on top of filters

4. Three Terminal TFTs:

A. a-Si TFT (Figure 7)

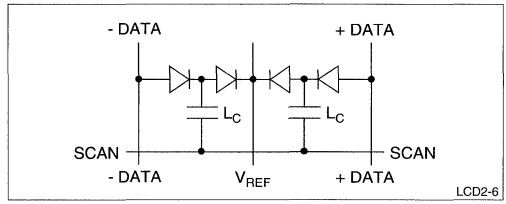


Fig. 6. D^2R Equivalent Circuit

a-Si TFTs are low-temperature polysilicon TFTs. During the manufacturing process the silicon is deposited as an amorphous film then recrystallized it to polysilicon by annealing at temperatures between 300° C and 600° C. This has advantages compared to conventional processing of polysilicon TFT fabrication that requires temperatures as high as 1000° C. The lower temperature process allows the use of low cost glass substrate instead of quartz or high temperature glass. [9]

Advantages:

- Larger temperature process, glass instead of quartz, (cost)
- Lower area capability
- Good uniformity over a large area

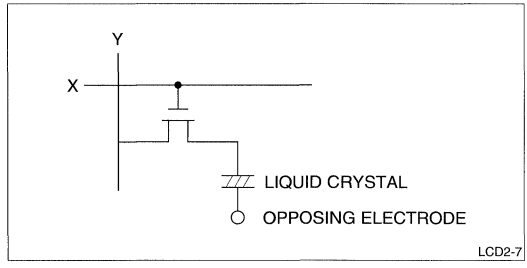


Fig. 7. a-Si TFT Equivalent Circuit

Disadvantages:

- Low mobility and high photo conductivity [8], (results is higher leakage current)
- Complex process (Compared with two terminal devices)

Work is being done to improve the leakage current characteristics of polysilicon TFTs at lower temperature process. Including composite fabrication of below 450° C by using laser induced crystallization of PE-CVDed amorphous silicon.

B. poly - Si TFT

Conventional high temperature polysilicon TFTs.

Advantage:

- High mobility (high-drive currents, high-speed circuitry)

Disadvantage:

- Practical applications reduced to devices formed on small area quartz substrates.
- Complex process
- High temperature process

TFT technology is by far the leading technology for active matrix displays because of the high contrast and high picture quality, ability to display good grayscales and superior response time. While MIM technology is attractive because of the cost advantages (fewer masks), the manufacturing of TFT displays has far out numbered the MIM displays because of the superior display performance.

Of the different TFT technologies a-Si TFTs are considered the leading candidates for large area active matrix displays since they can be fabricated at low temperatures on inexpensive glass substrates.

ADVANCES IN COMPONENT TECHNOLOGY

Drivers

1. 3-bit Digital source driver driven by a single 5 V supply for TFT LCD. [6]

   A 3-bit digital source driver has been developed by Sharp Corporation for computer application (CA) displays. The advantages of a digital driver over an analog driver are as follows:

   - Increased sampling speed. The maximum sampling speed of this new driver is 20 MHz, more than three times as fast as the traditional analog sampling driver.
   - Data can be sampled more precisely.
   - Greater computer compatibility because the computer is a digital device
   - +5 V operation eliminates the need for a +12 V supply

2. 16 level grayscale driver for TFT LCD [7]

   A 16 level analog grayscale driver has been developed by Fujitsu which has the following advantages:

   - 16 level grayscale will provide 4096 colors, compared to 512 colors of the standard 8 level grayscale drivers.
   - If frame rate control (FRC) is used it is possible to obtain 64 grayscale levels capable of displaying 262,144 colors. This is useful for multimedia applications.

3. 6-bit Digital source driver driven by a single 5 V supply TFT LCD. [10]

   Sharp has developed a 6-bit digital source that can generate 64 gray levels. For audio visual (AV) and computer applications (CA). The driver is driven by a single +5 V source and is packaged by a TAB (tape automatic bonding) process the drivers when used with the Inter-frame modulation method can produce both AV and CA display, thereby combining CA and AV technology in a single display.

Color Filters

Red, green and blue (RGB) color filters are a very important component of a display. Table 1 summarizes the different color filter application methods.

Color filters are arranged in different formations depending on the design criteria. There are four methods to arrange the color filters in a display (Figure 8).

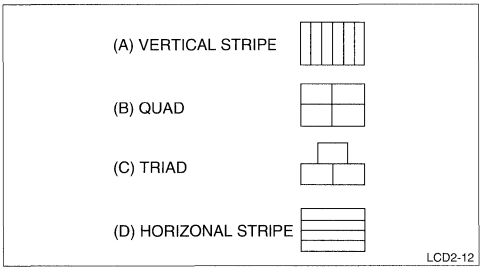


Fig. 8. Methods of Arranging Color Filters in a Display

A. **Vertical Stripe:** Most popular arrangement for CA (computer application) type of displays.

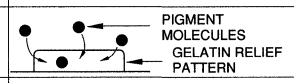
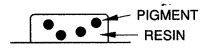
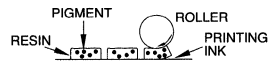
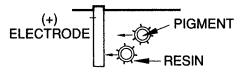
   Good arrangement for good character display capabilities. Requires three vertical data line drivers and one horizontal gate line driver for each pixel.

B. **Quad:** Requires two vertical drivers and two horizontal gate line drivers for each full color pixel.

C. **Triad:** Used for AV application because of good appearance, such as TV applications.

D. **Horizontal Stripe:** Requires one vertical data line driver and three horizontal gate line drivers. Because a 640 x 480 panel has more vertical lines than horizontal lines, the horizontal stripe configuration is the most economical. The major disadvantage is that line-buffer memory is required except in applications that have frame memory. [12]

Table 1.
Color Filter Formation Methods

METHOD NAME	DESCRIPTION
Gelatin Dyeing: Patterned resin is dyed.	
Pigment Impregnation: Resin containing pigment is made into a pattern.	
Printing: Color ink containing pigment is printed on.	
Electroplating: A resin coat is electroplated on the pigment surface.	

Backlighting Technology

Since LCDs do not emit light, an external light source is needed. The most common light sources are: HCFTs (hot cathode fluorescent tubes) or CCFTs (cold cothode fluorescent tubes). The HCFTs are usually brighter, but their life is not as long (10,000 hours). The CCFTs are not as bright, but they have a longer life (20,000 hours).

The tubes can be placed behind the LCD panel for maximum brightness but results in an increased panel thickness; or on the edge of the panel for less brightness but better brightness uniformity and a thinner display.

PUTTING IT ALL TOGETHER

Sharp has put all the technologies together to offer some of the most innovative active matrix displays available in the market. An example of such a display is the LQ9D011 for computer applications. Let's take a look at the LQ9D011 and see how Sharp has combined the various technologies for a state of the art panel. LQ9D011 (640 x 480 x 3 TFT 512 color VGA panel for laptop applications):

1. TFT technology providing high-contrast and high-picture quality, 8 grayscales, and 80 ms typical response time.
2. CCFT Edge Light: Low power (laptop application), thin (12mm), good brightness uniformity and still providing 70 nt typical brightness.
3. 3-bit, 5 V drivers: Providing 512 colors with only 5 V requirement. It is packaged using a Tape Automated Bonding (TAB) process and measures approximately 25 mm x 20 mm with 120 outputs per chip.
4. It utilizes a vertical stripe color filter pattern very good character display.
   s

REFERENCES

[1] Tannas, Lawrence, Jr., *Flat-Panel Displays and CRT's*

[2] Perez, Richard., *Electronic Display Devices*, pp. 230-241

[3] Refioglu, H., *Electronic Displays*, pp. 114-115

[4] Kamagami, S., pp. 119-202, 1991, *International Display Research Conference*

[5] Kotoyoshi, Takahashi, pp. 247-250, 1991, *International Display Research Conference*

[6] Okada, Hisao, pp. 111-114, 1991, *International Display Research Conference*

[7] Takahara, Kazuhiro, pp 115-118, 1991, *International Display Research Conference*

[8] Endo, Y., pp. 203-206, *International Display Research Conference*

[9] Mitra, U., pp. 207-210, *International Display Research Conference*

[10] Okada, Hisao, pp. 251-254, *International Display Research Conference*

[11] LCD Display, *The Leading Edge in Flat Panel Display*, Sharp Corporation

[12] Matino, Haruhiro, pp. 400-403, *SID 90 Digest*, 1990

INTERFACING SHARP'S LJ64ZU50 TO ATLANTIC DIGITAL'S PC4558

Dan Gutierrez, Field Applications Engineer

INTRODUCTION

Electroluminescent displays (EL) have a strong hold on the industrial computer market in areas of flat-panel technology. LCD is getting better visually but it is still lacking in operating temperature and shock/vibration specifications. Also, the TFT technology is used in some applications and it is increasing in popularity but the cost is still to high. EL can replace monochrome CRT displays because of its short depth (less than an inch) and equal if not better visual performance.

Depth is a crucial dimension because it saves factory floor space. The amber/orange color of EL is the one draw back which takes time to get accustomed to. The red color of plasma has not been accepted in ergonomics as a "comfortably" pleasing to the eye display. EL is the perfect choice as long as battery operation is not required. EL draws about 22 watts as compared to 2.2 watts of a monochrome LCD.

The purpose here is to offer those who have never used an EL display in a DOS environment a way to assemble a development tool quickly. For others this may be the solution for their application. This application note is intended to take one through the simplest path to a working EL VGA resolution display.

GENERAL INFORMATION

The LJ64ZU50 is Sharp's grayscale solution for EL. (This model is the same as the LJ64ZU52 but we will continue to refer to it as the ZU50 for this application note.) The resolution is 640 x 480 VGA. This resolution is now common in industrial workstations and their corresponding application software.

The grayscale capability of the display has been improved. Because of the shadowing characteristic of grayscale other vendors rely on dithering and hatching to simulate grayscale. Sharp has continued to design with pulse width modulation to achieve true grayscale, and ultimately has succeeded in reducing the shadowing effects. Table 1 indicates some key characteristics for this display.

Table 1.
LJ64ZU50 Characteristics

Overall Dimensions	267 x 197 x 20 (36° C)
Active Area	211.1 x 158.3
Dot Pitch	0.33 x 0.33
Power	22 W
Luminance	20fL (minimum)
Weight	800 grams

I chose the PC4558 Atlantic Digital controller because of its simple cabling capability with the ZU50. Atlantic has other high-end display controllers which require a little more effort in preparing an interface cable, but offers a few more features. Some of these are indicated in Table 2.

Table 2.
Atlantic's Display Controller Features

PC455H	Half-card Size	Can Run All of Sharp's Monochrome LCDs
PC456HA	Half-card Size	Same as 455H With 64 Grayscales
PC520	Oversized Half Card	Can Run 10.4" and 8.6"TFT
PC530	Same as 520	Change One IC Chip for Duty Color and TFT Dual Purpose

The card is 100 percent register level compatible with IBM VGA so one can use off-the-shelf compatible software like Windows and others without the need of any special drivers. A specific video BIOS ROM is recommended here for the Sharp ZU display. These are available from Atlantic Digital. Refer to page three for detailed board specifications.

Atlantic Digital® is a trademark of Atlantic Digital Corporation.

INTERFACE CABLE

As pointed out in the beginning of this discussion the cabling is quite simple. Besides the parts described above, the only other requirement is a 26-pin ribbon cable with appropriate headers. The nomenclature and pin designators are different, but the ribbon cable interconnects are pin-for-pin between the display and the card. The following shows the pin out for both systems (Tables 3 and 4).

Table 3.
Sharp's LJ64ZU50 Pin-out

TERMINAL NO.	SIGNAL NAME OR SUPPLY VOLTAGE	TERMINAL NO.	SIGNAL NAME OR SUPPLY VOLTAGE
A-1	NC	B-1	NC
A-2	D-11 (Data Signal)	B-2	D 10 (Data signal)
A-3	D 13 (Data signal)	B-3	D 12 (Data signal)
A-4	D 01 (Data signal)	B-4	D 00 (Data signal)
A-5	D 03 (Data signal)	B-5	D 02 (Data signal)
A-6	NC	B-6	NC
A-7	CK D (Data Xfer clock)	B-7	GND
A-8	HD (Horz sync signal)	B-8	GND
A-9	VD (Vert sync signal)	B-9	GND
A-10	GND	B-10	GND
A-11	connect to A13 & B13 *	B-11	connect to A13 & B13 *
A-12	V L (+5 V)	B-12	V L (+5 V)
A-13	VD *(+12 V)	B-13	V D (+12 V)

*These will route the "VMID" from the card that originally for the ZU49. Note: Be sure to use 12 VDC in place of 24 VDC.

Table 4.
PC4558 Interface Card Pin-out

TERMINAL NO.	SIGNAL NAME OR SUPPLY VOLTAGE	TERMINAL NO.	SIGNAL NAME OR SUPPLY VOLTAGE
2	NC	1	NC
4	OVID 1	3	OVID 0
6	OVID 3	5	OVID 2
8	OVID 5	7	OVID 4
10	OVID 7	9	OVID 6
12	NC	11	NC
14	PLASCLK	13	GND
16	OWGTCLK	15	GND
18	OVSYNC	17	GND
20	GND	19	GND
22	+VMID	21	+VMID
24	+5 FUSED	23	+5 FUSED
26	NC	25	NC

Note: Pin 26 and 25 are interconnected to 22 and 21 at the display end.

As you can see, laying one chart on top of the other verifies the pin compatibility. One should take note of the connector polarity mark given in the display specification. It faces the correct side, but it should be drawn at the opposite end near pin B-1 and A-1, which are pin 1 and 2 respectively.

INTERFACE SIGNALS AND TIMING OVERVIEW

The interface signals are very simple to understand. These are four bits of grayscale information for each pixel, and there is odd and even data. Please refer to the timing diagram (Figure 1). The odd data is bits D 00 through D 03 with D 00 as the LSB. The even data is D 10 through D 13 with D 10 as the LSB. All Bits "0" yields dark or full off, while all bits "1" will yield bright or fully on. Consequentially, 8 bits of data are latched in parallel. All the grayscale interpretations are achieved internally by the PC4558 as any VGA compatible board. The display accepts the TTL data that is clocked in via the data transfer clock CK_D.

The last CK_D pulse latches columns 639 and 640. This is followed by a HD or horizontal pulse, which has a dual function. First it enables the output of the display drivers for the row of data which has been loaded. Second, it starts the latching of the next row of data. There is a short 1.3 μs delay between these two events. The VD or vertical pulse synchronizes everything back to the first row. A minimum of four horizontal timings is needed between row 480 and row one.

POWER SUPPLY CAUTIONS

The PC4558 has built in DC to DC converters for other types of displays, but for our purpose one simply needs to route the computer system 12 VDC through the 2 amp fuse (F1). The PC4558 provides this, as shown in the schematic on page 5 (lower left hand corner of the schematic). One must make certain that only F1 and F3 are installed on the PC4558. This will automatically set VMID to +12 VDC for the Sharp ZU50. Fuse F3 routes VL of +5 VDC for the display logic circuits.

CONCLUSION

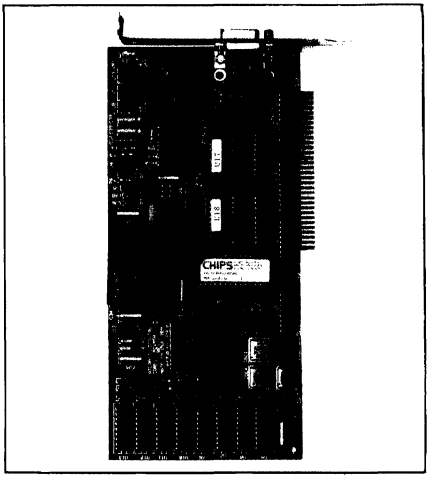
We would like to encourage the use of EL in more diverse areas of the display industry. Hopefully the simplicity of preparing a development tool such as this will help engineers evaluate EL more often for more applications. There is certainly an abundance of display driver support in the USA. Atlantic Digital is not by any means the exclusive recommendation from Sharp. This is only an example of which an engineer can start software development and make initial evaluations of display performance quickly.

Atlantic Digital Corporation

PC4558
PC/AT VGA COMPATIBLE FLAT PANEL CONTROLLER

PRODUCT OVERVIEW

The PC4558 is a complete VGA-Compatible flat panel controller for the PC/AT bus. The PC4558 supports a wide variety of LCD, PLASMA, and ELECTROLUMINESCENT flat panel displays as well as standard TTL CRTs.



FEATURES

- 100 percent register level compatible with IBM VGA.
- Standard DB9 TTL video output connector for CRT.
- On board invertor for CCFT LCD backlight (optional)
- On board adjustable negative bias supply for LCD contrast, control.
- Contrast control for LCDs accessible at rear panel or remotely with addition of external potentiometer.
- On board connectors for HITACHI and SHARP LCDs, OKI PLASMA, and SHARP ELECTRO-LUMINESCENT flat panels.
- Auxiliary output connector provided for support of flat panel displays from other vendors.
- On board BIOS boots in CRT or Flat Panel mode.
- Provides backward compatibility to EGA, CGA, Hercules, and MDA on flat panel displays.
- Supports flat panel displays of up to 720 × 480 resolution.
- Supports gray scale and monochrome flat panels.
- Provides 16 levels of gray on monochrome LCD and PLASMA panels by using Frame Rate Control or PWM.
- Provides 8 or 64 colors on color flat panels.
- Supports Single Panel/Single Drive, Dual Panel/Single Drive, Dual Panel/Dual Drive LCDs and Single Panel/Single Drive PLASMA and EL panels.
- Provides Horizontal and Vertical compensation to map an image in one resolution to a flat panel with a different resolution.
- Flat Panel Clock signals routed through PAL for support of non-standard displays.
- Connectors provided for external power input to PLASMA and EL displays as well as + 12V for external DC/DC convertors.

SPECIFICATIONS

Bus Interface	8 BIT PC/AT BUS
Form Factor	3.9H × 8.3L
Frame Buffer	256K
Display Resolution (Flat Panels)	640 × 200 640 × 350 640 × 400 640 × 480 720 × 480 User Programmable
Display Resolution (CRTs)	640 × 200 CGA 640 × 350 EGA 640 × 400 640 × 480 VGA (TTL) 720 × 348 Hercules User Programmable
Color Resolution (CRT)	16 From Pallete of 64 (CRT)
	8 or 64 Colors (color LCD)
Vertical Refresh	60-70 Hz Non-interlaced (CRT)
Horizontal Frequency	15.5 to 35 KHz (Varies with Resolution)
Power Requirements	+ 5V @ 1A + 12V @ 300 MA
Operating Temperature	0 to 40 Degrees Celsius
Relative Humidity	0-95% Non-Condensing

Atlantic Digital® is a trademark of Atlantic Digital Corporation.

OUTPUT CONNECTORS

CRT TTL Video (J2)	DB9 Subminiature	
Inverse Video Select (J3) (Flat Panels)	1 × 3 Header 1-2 Normal Video 2-3 Inverse Video	
External Power (J4) (Flat Panels)	3 Pin Conn 1 +VHI 2 +VMID 3 GND	
External LCD Contrast (J5)	1 × 3 Header 1-2, 3 2OK Potentiometer 2 × 13 Dual Row Header 2 × 10 Dual Row Header	
Sharp EL (J9)	2 × 8 Dual Row Header	
Sharp LM64048Z LCD (J10)	2 × 7 Dual Row Header	
Sharp LM64148Z LCD (J11)	2 × 8 Dual Row Header	
Auxiliary Flat Panel (J15)	2 × 13 Dual Row Header	
Pin Description	1,2	+VMID
	3,4	+VHI
	5, 6, 8	GND
	7	WGTCLK
	11, 13, 15, 17	VIDO-7
	19, 21, 23, 25	
	10, 12	+5V Fused
	14, 16	+12 Fused
	18	PLASCLK
	20	LCDCLK1
	22	HSYNC
	24	VSYNC
	26	BLANK
CCFT Backlight AC (J13)	2 Pin Conn	

ORDERING INFORMATION

Part Number	Description
PC4558-I	With CCFT Invertor & LCD Contrast Control
PC4558-NI	With LCD Contrast Supply Only
PC4558-O	Without Backlight Invertor or LCD Contrast Supply

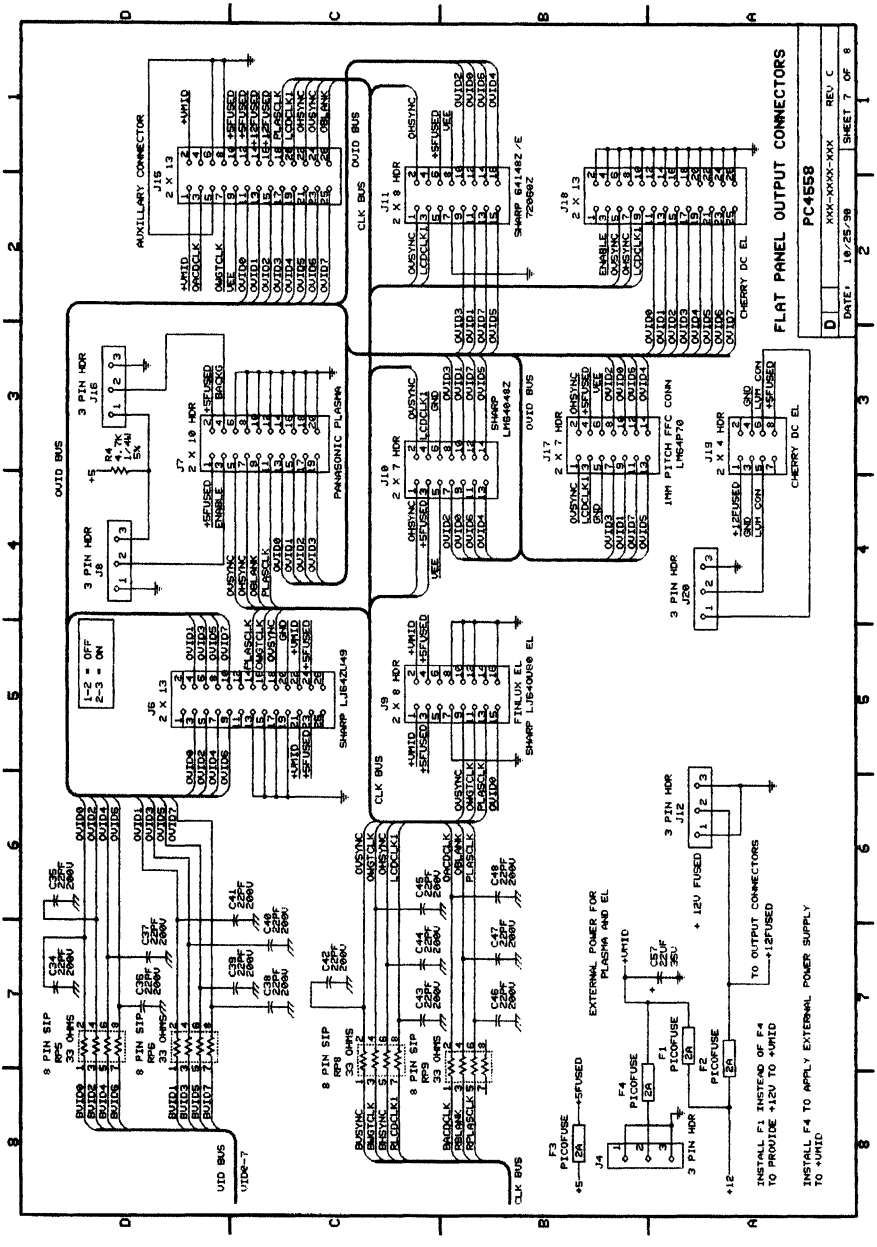
Contact Factory for custom requirements or other DC/DC convertor options.

AT, VGA, CGA, MDA are trademarks of International Business Machines.

851 West State Road 436, Suite 1051 Altamonte Springs, Florida 32714 407/788-4200 FAX 407/788-8692

Copyright 1989

Atlantic Digital® is a trademark of Atlantic Digital Corporation.



Atlantic Digital® is a trademark of Atlantic Digital Corporation.

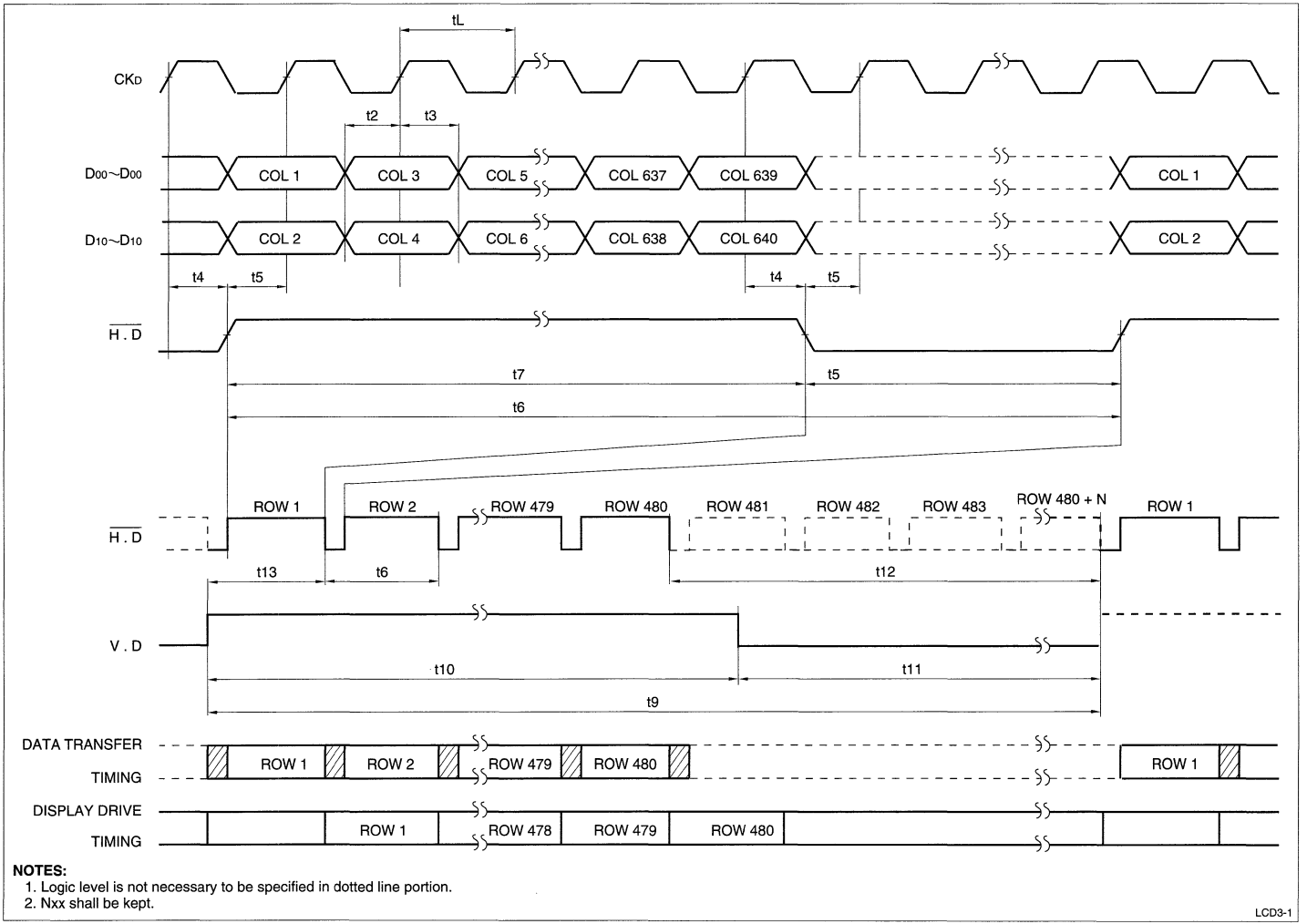


Fig. 1. Interface and Display Drive Timing

NOTES:
1. Logic level is not necessary to be specified in dotted line portion.
2. Nxx shall be kept.

IMAGE QUALITY: MEASUREMENTS AND DEFINITION

Alan Dragon, Field Applications Engineer

INTRODUCTION

This application note will explain image quality of displays by defining commonly used measurement terminology and by describing the visual capabilities of the observer. In understanding image quality, it is important that the limits of the display and that of the human-eye system are comprehended. For a display to "look acceptable" there are many intangibles that come into play such as aesthetic taste, cultural conditioning, intelligence, attentiveness, and patience. In order to equally assess different displays and display technologies, image quality must be carefully defined in tangible and measurable terms. When characterizing a display, sensory qualities such as luminance, contrast, color, and spatial frequency depend and interact with each other. Although the focus will be on flat panel displays (LCDs and ELs), all of these terms and measuring techniques can be applied to any display technology.

It is well known that many terms used in our everyday technical discussion are often misused, such as brightness vs. luminance and resolution vs. display format. Although it is often hard to break habits, this application note will define all of these terms so there is at least an understanding of the technically correct definition.

HUMAN VISUAL RESPONSE

It is important to understand how we process and perceive visual stimulus to understand how image quality is defined. The human eye is composed of two photoreceptors; rods and cones. These photoreceptors are distributed in different parts of the retina and have different sensitivities to light energy. The cones are concentrated in the center of the eye and the rods are concentrated 18-20° off center, linearly decreasing their density out to approximately 80°. Rods and cones have the following characteristics:

- Rods have greater sensitivity to light than do cones.
- Cones are differentially sensitive to the wavelengths of light (i.e., can "see" colors) whereas rods can only "see" shades of gray.
- Cones have greater resolving capabilities than do rods, thus provide greater discrimination of details.
- Rods are more sensitive to temporal changes in luminance levels and are more likely to see flicker. This is why we can see flicker more readily in our peripheral vision.

Based on these attributes, the human eye processes visual information with the following characteristics:

- When the intensity of a multi-color display or image is lowered, the colors will drop out sequentially into gray levels. The first colors to shift to gray are blues and reds with the last being greens. This shifting of colors to gray is known as the Purkinje shift.
- The ability to detect resolution depends on the size of the object, luminance of the object, luminance of the background and ambient, and the color of the object.
- Lower luminance levels on a display require greater contrast for equal discriminability (i.e., a high-contrast low-light output LCD can look as good as a lower-contrast high-light output LCD).
- The human eye is sensitive to the color spectrum between the wavelengths of 380 and 720 nm. This sensitivity shifts based on ambient lighting conditions. There is approximately a 50 nm shift of the sensitivity curve, lower in wavelength, from daytime to nighttime vision.
- As the human eye-brain system is differential input, the eye is very sensitive to luminance and color differences. However, as the average luminance level increases on a display, the eye is less sensitive to absolute luminious differences.

The two primary measurements used in display image quality characteristics are luminance and spectral radiance. Resolution is also important in defining the interaction between the information density on the display and the resolving capability of the human visual system.

LUMINANCE

Luminance is the psycho-physical measure of perceived radiant power under carefully controlled and defined conditions. Luminance can also be defined as the quantitative measure of brightness and is measured in English units as footlamberts (fL) and in SI (International System) units as candela/m^2 (nit).

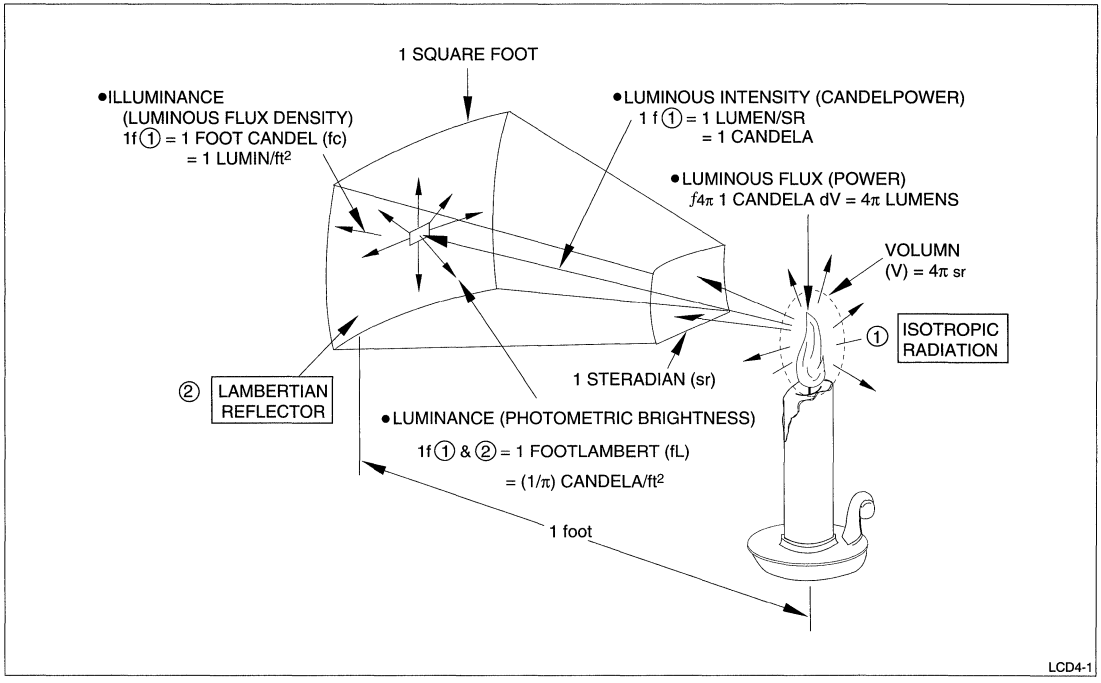


Fig. 1. Footlambert

A footlambert is the luminance reflected off a 1 square foot Lambertian (perfectly diffusing) reflector illuminated by 1 foot candle (fc) (Figure 1). Figure 2 shows the same setup in SI units. Table 1 gives the conversion factors between the two systems along with a chart (Figure 3) to show the electrical equivalent term.

BRIGHTNESS

Brightness is a term that is commonly misused when specifying light output levels. Brightness is purely a psycho-physiological attribute; the subjective response to electromagnetic energy that occurs when the human eye system has adapted to a particular radiation level. To put it in simple terms, a radiance and intrinsic luminance. BRIGHTNESS IS NOT A MEASURABLE VISUAL PARAMETER.

RADIANCE

Radiance is a measure of the rate of energy flow from an electromagnetic source usually measured in power units; watts or joules/second. All colormetric analysis requires the use of radiance measurements as a function of wavelength (spectral radiance) and subsequent conversion to luminance or chrominance values. Chrominance is the physical combination of the dominant wavelength (hue) and purity (saturation).

Hue is the dominant wavelength of the color as subjectively perceived by the human eye system. Saturation is the degree to which the hue of a color subjectively appears to be undiluted by its complimentary color to form white. If there is no trace of apparent white in the color, it is said to be fully saturated.

When explaining about color it is important to understand how the human eye processes color. Figure 4 shows the human eye response to radiant power. This clearly indicates the eye does not have the same sensitivity to all colors. Our eyes are much more sensitive to green wavelengths of light. In order to get the same perceived luminance (sensory response) for all colors, the radiance power levels (lumens/watt) are different. It is important to note that lumen of red, green, or blue light has the same sensory response, only the amount of radiant power to get to that level is different.

When describing white light, it is EQUAL LUMINANCE levels across the visible spectrum that gives white its color. EQUAL RADIANCE levels across the visible spectrum however, will give a magenta color. A sometimes confusion concept, but as long as radiance and luminance levels of wavelengths are kept separate, there shouldn't be any comprehension problems.

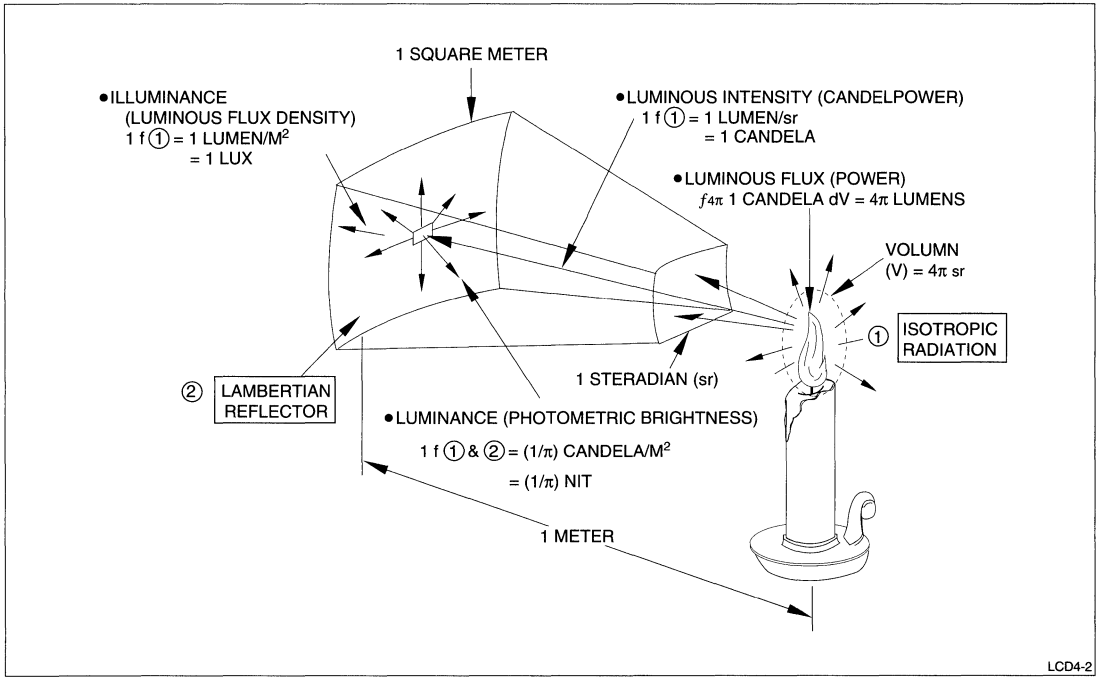


Fig. 2. Footlambert in SI Units

Table 1.
Conversion Factors and Electrical Equivalents

NO. OF MULTIPLIED BY EQUALS NO. OF	FOOTLAMBERT	CANDELA/m^2	MILLILAMBERT	CANDELA/in^2	CANDELA/ft2	STILB
Footlambert	1	0.2919	0.929	452	3.142	2,919
Candela/m^2 (nit)	3.426	1	3.183	1,550	10.76	10,000
Millilambert	1.076	0.3142	1	487	3.382	3,142
Candela/in^2	0.00221	0.000645	0.00205	1	0.00694	6.45
Candela/ft^2	0.3183	0.0929	0.2957	144	1	929
Stilb	0.00034	0.0001	0.00032	0.155	0.00108	1

ELECTRICAL	LUMINOSITY TERM	LUMINOSITY UNIT
Power (rate of energy flow)	Luminous Flux	Lumen = 1/680 watt/Luminosity Function
Power-source Output	Intensity (Power-Source)	Candela = Lumen/Steradian
Delivered Power	Luminance (Surface)	Nit = Lumen/Steradian/$meter^2$ = Candela/m^2
Power-transfer Efficiency	Transmittance	Transmittance Factor = 0.0 to 1.0
	Reflectance	Reflectance Factor = 0.0 to 1.0

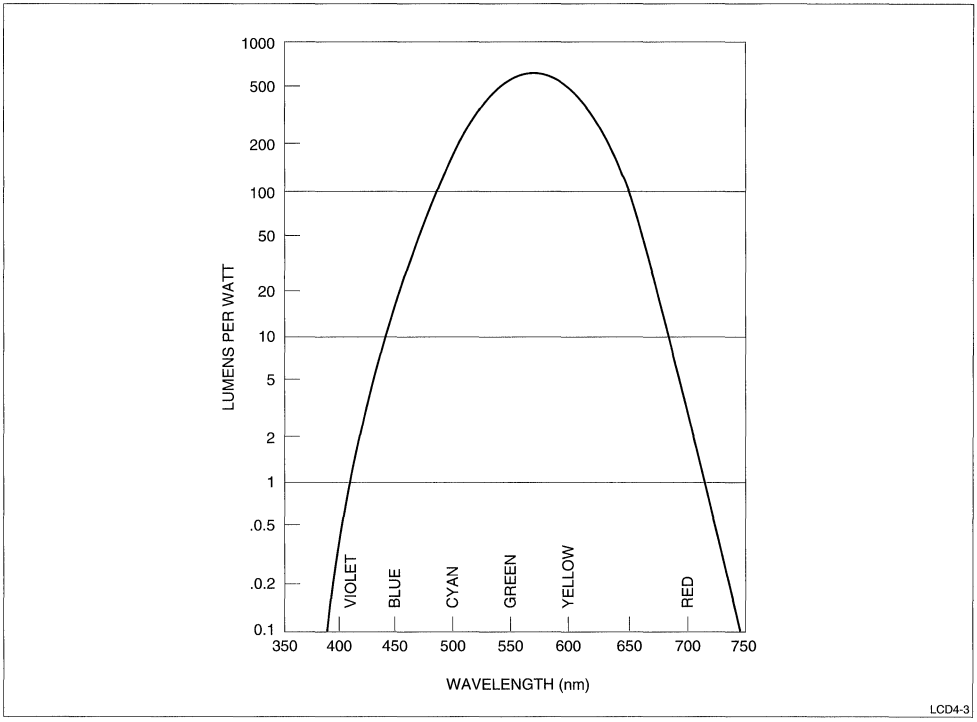


Fig. 3. Electrical Equivalent Term

COLOR

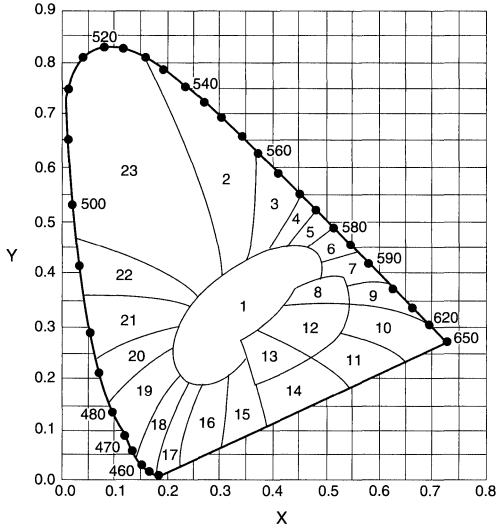
One of the best ways to describe the three dimensional color characteristics (luminance, hue, and saturation) in a two dimensional chart is by using the CIE (Commission Internationale de L'Eclairage) System. (Figures 4 and 5) The original 1931 chart revised in 1955 by the NBS (Nation Bureau of Standards) and adopted by the EIA (Electronic Industries Association) is also called the Kelly Chart and can be used today for any type of color characterization. The chart is based on the supersaturated primaries of red, blue, and green. Any color can be characterized by its x and y coordinates. The chart assumes all colors are generated by equal radiance sources, which puts white at the center of the chart. Color temperature can also be readily specified by this chart. Color temperature is often used to specify white balance. The color temperature of a source is the temperature in Kelvins at which an ideal black-body radiator would emit the same spectral power distribution. Any set of primaries can be plotted and aligned to any reference white. The compliment of every primary color is a combination of the other two primaries on the chart. Two colors are considered compliments if they connect thru the center white point on the chart.

SHADES OF GRAY

Specifying shades of gray is often a game of specmanship. The industry standard has been to assume the detectable increment in luminance between two levels is the square root of 2 or 1.41. When calculating gray level, the original luminance level, size and shape of the object, number of objects, and adapting luminance should all be taken into account. As this can be very difficult, it is sometimes better to define dynamic range (maximum-minimum luninance) to determine performance.

UNIFORMITY

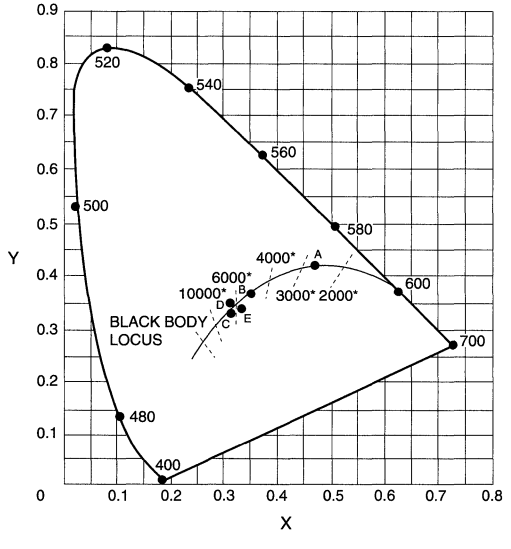
Uniformity, or the absence of, called nonuniformity, is the gradual change of luminance and chrominance from one display area to the next. It can be broken down into large-area and small-area nonuniformity. Large-area nonuniformity is usually a luminance change over the entire display (i.e., edge-to-edge, edge-to-middle). Because these changes are very gradual, a 50% change from edge to edge sometime can not be noticed. At low display luminance levels though, even these gradual changes may be seen. Small-area nonuniformity is often referred to as pixel to pixel changes in luminance or chrominance.



1. ILLUMINANT AREA
2. YELLOWISH GREEN
3. YELLOW-GREEN
4. GREENISH YELLOW
5. YELLOW
6. YELLOWISH ORANGE
7. ORANGE
8. ORANGE PINK
9. REDDISH ORANGE
10. RED
11. PURPLISH RED
12. PINK
13. PURPLISH PINK
14. RED PURPLE
15. REDDISH PURPLE
16. PURPLE
17. BLUISH PURPLE
18. PURPLISH BLUE
19. BLUE
20. GREENISH BLUE
21. BLUE-GREEN
22. BLUISH GREEN
23. GREEN

LCD4-4

Fig. 4. Three-Dimensional Color Characteristics



LCD4-5

Fig. 5. Three-Dimensional Color Characteristics

These changes are usually very visible due to the sensitivity of the human eye to changes in luminance in close proximity.

PIXEL

The smallest resolvable spatial-information element on any display is called a pixel. The pixel can be subdivided further to achieve color. Each red, green, and blue element is referred to as a subpixel (Figure 6). The spatial dimension of pixel can be defined by the pixel size and pixel pitch (Figure 7). Fill factor (Figure 8) is another parameter used when image quality measurements are taken over an area (more than one pixel).

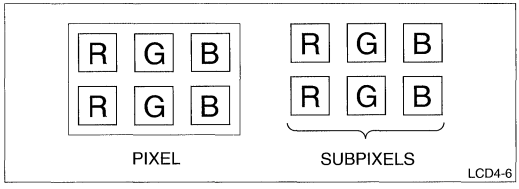


Fig. 6. Subpixel

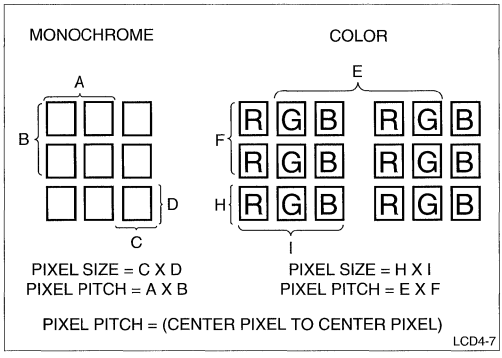


Fig. 7. Spatial Dimension of a Pixel

RESOLUTION

Resolution is probably the most misused term when describing image quality. For example, when describing a LCD, 640 x 480 is not the resolution. This should be referred to as the display format. Resolution is a very complicated measurement which deals with both the display and the human eye. For discretely addressed displays such as LCDs, resolution is usually measured in resolvable elements per unit measurement (i.e., pixels (dots)/inch). For analog addressable displays such as CRTs, resolution can be defined as above or more accurately as the spatial frequency (lines/inch) at which an observer can no longer discriminate the light and dark bars of a square wave pattern. There are numerous methods to measure resolution, so it is important that whatever method is used, it is consistent when comparing two display to each other.

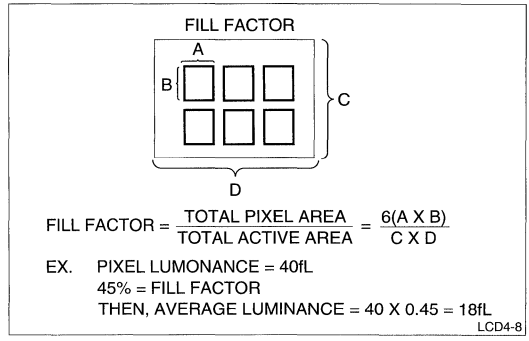


Fig. 8. Fill Factor of a Pixel

CONTRAST RATIO

$$\text{Contrast Ratio} = \frac{\text{Maximum Luminance}}{\text{Minimum Luminance}}$$

Although this is a simple formula, many factors must be taken into consideration. All ambient and display parameters must be defined in order to calculate a meaningful ratio. As is shown in Figure 9, if ambient conditions are not taken into account, the contrast ratio can vary significantly. Another factor is the area for the measurement. It can be on-pixel ("white") to off-pixel ("black") or on-area ("white") to off-area ("black"). In LCDs the angle of measurement to the display surface should be defined as the contrast ratio varies over the angle. Again, as long as all conditions are controlled and noted, an accurate comparison of displays can be made.

VIEWING ANGLE

Viewing angle is an important parameter to define in non-emissive display such as a LCD. Because there are limitations inherent to the technology, the viewing angle can help characterize the display more accurately. LCDs usually have their viewing angle defined as minimum contrast ratio over a certain angle. This angle is specified by its x and y direction on the face of the display.

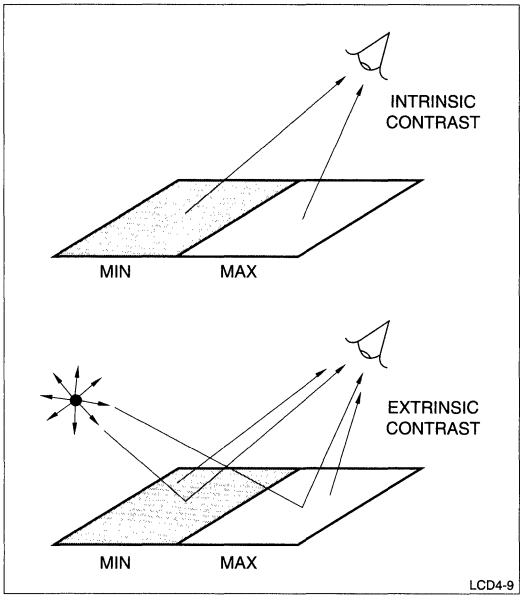


Fig. 9. Contrast Ratio

RESPONSE TIME

Response time is the time it takes a pixel to change state from on to off ("black" to "white"). This time includes all electrical and physical delays. It is defined as the transition time from the 10% level to the 90% level of luminance output (Figure 10). In LCDs, because the rise and decay times are usually unequal, both are specified; $\tau\gamma$ (10%-90%), τd (90%-10%). Rise and decay times may be combined to give a total response time.

CONCLUSION

As has been explained, image quality is characterized by a close interaction between the display and human eye system. Both must be understood to fully comprehend how to perform accurate and meaningful measurements. In order to compare displays of the same technology or even different technology, carefully controlled conditions and techniques must be implemented to get comparable results. Unfortunately, sometimes in the end, it may be a subjective visual response that determines the choice of the display, contrary to measured data. With the understanding of the terminology used in image quality analysis, it is hoped subjective decisions can be kept to a minimum.

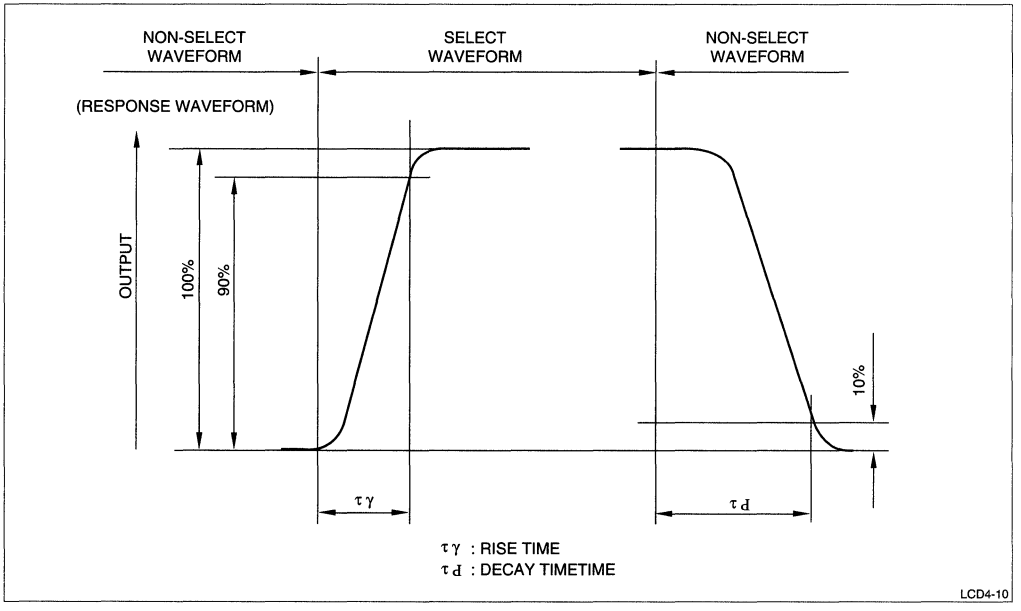


Fig. 10. Response Time

MECHANICAL DESIGN CONSIDERATIONS FOR INTEGRATING LIQUID CRYSTAL DISPLAYS INTO NOTEBOOK & PEN BASED COMPUTERS

Charles Guthrie, Field Applications Engineer

ABSTRACT

In portable computers, there is a need to reduce overall size and weight of the units. The following guidelines focus on several problems facing the designer. Discussions include the maintenance of flatness and rigidity, thermal management, placement of the inverter, heat sinking the CCFT, and protecting the face of the display.

INTRODUCTION

As more companies begin the manufacturing of their next generation of computers, there is a need to reduce the overall size and weight of the units to improve their portability. This has sparked the need for more compact designs where the various components are placed in closer proximity, thus making them more susceptible to interaction from signal noise and heat dissipation. The following is a series of guidelines for the placement of the display components, and a summary of suggestions for overcoming difficult design constraints associated with component placement.

In notebook computers, the thickness of the display housing is important. The design usually requires the display to be in a pivotal structure so that the package may be folded down over the keyboard for transportation. Also the outline dimensions must be minimal so that the overall package will remain as compact as possible. These two constraints drive the display housing design and placement of the various components. The following is the result of actual experiences with design problems, and promotes practical solutions for the designer.

The problems facing the pen based computer designer will be similar to those realized in designing notebooks. In addition, however, pen based designs will require the implementation of protection for the face of the display. In pen based applications, as the pen moves across the surface of the display, the pen could scratch the soft plastic that the front polarizer is made of. For this reason, the front of the display must be protected. We will discuss various methods for protecting the face of the display while minimizing the effects of the cover material on the optical properties of the display image.

Within this application note, we will also discuss the need to specify the flatness of the bezel. Suggestions will be presented concerning acceptable construction techniques to accomplish a sound design. Further, we will identify the display components likely to cause problems due to heat buildup, while discussing the methods used to minimize the effects of the heat on the display.

Even though this paper is intended to assist the designer in considering the problems associated with the development of a display system for notebook and pen based computers, the ideas expressed here are not to be construed as being the only solutions to the various problems, and have not been assessed as to whether they may infringe on any patents issued or applied for.

FLATNESS AND RIGIDITY OF THE BEZEL

In the notebook computer, the bezel has several distinct functions. It houses the display, the inverter for the backlight, and in some instances, the controls for contrast and brightness of the display. The bezel is usually designed to tilt in order to compensate for the optimum viewing angle of the display and its relationship to the person operating the computer.

It is important to understand that the bezel must provide a mechanism to keep the display flat, particularly at the mounting holes, in that even subtle changes in flatness will place uneven stress on the glass which can cause variations in contrast across the display. Slight changes in pressure may cause significant variation in the apparent contrast of the display. Also, at the extremes, significantly uneven pressures can cause the display glass to fail.

Because the bezel must be functional in maintaining the flatness of the display, consideration must be made for the strength of the bezel. Care must be taken to provide structural members, while minimizing the weight of the unit.

This may be executed using a parallel grid, (Figure 1a) normal to the edges of the bezel, or angled about 45° off of the edges of the bezel (Figure 1b). The angled structure may be more desirable in that it will

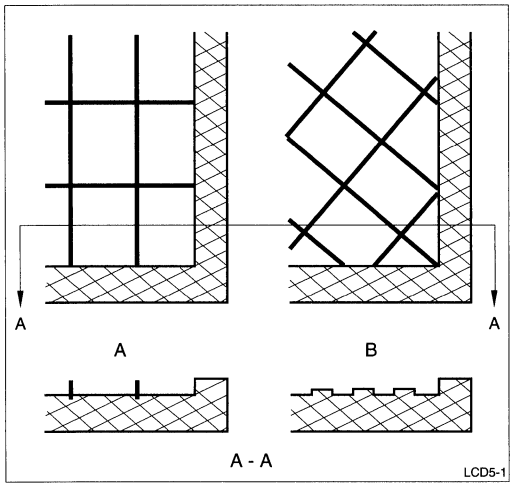


Fig. 1a,b. Parallel and Angled Bezel Grids

provide resistance to torquing the unit while lifting the cover with one hand. Again, the display is sensitive to the stresses due to uneven pressure on the display housing.

Another structure that will provide excellent rigidity, but adds more weight to the computer would be a "honeycomb" structure (Figure 2). This "honeycomb" structure resists torquing from all directions and tends to provide the best protection for the display.

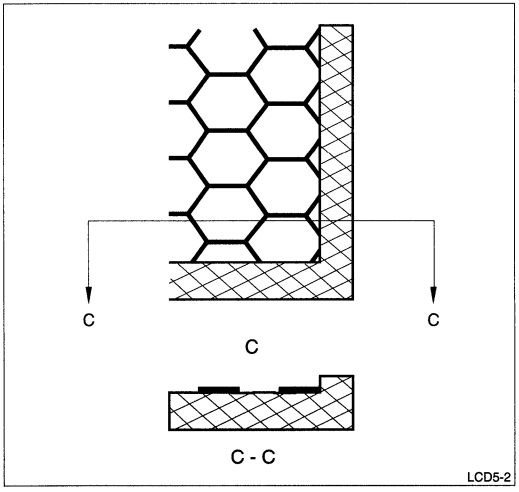


Fig. 2. Honeycomb Bezel Grid

With each of these structures, it is easy to provide mounting assemblies for the display. Threaded inserts (blind nuts) can be molded into the housing. The mounting may be done to either the front or rear of the bezel, however, attachment to the rear may provide better rigidity for the display and easier placement of the mounting hardware.

One last caution is worth noting in the development of a bezel. The bezel should be engineered to absorb most of the shock and vibration experienced in a portable computer. Even though the display has been carefully designed, the notebook computer can present extraordinary problems, related to use and abuse, for the design engineer.

AVOIDING HEAT BUILDUP IN THE DISPLAY

Several of the display components are sources for heat problems. If thermal management is not taken into account in the design of the display bezel, the display may be adversely affected evidenced by a loss of contrast uniformity. The Cold Cathode Fluorescent Tube (CCFT) itself gives off a small amount of heat relative to the amount of current dissipated in its glow discharge. Likewise, even though the inverters are designed to be extremely efficient, there is some heat generated. The build up of heat in this component will be aggravated by the typically "tight" designs currently being introduced (Figure 3). There is little ventilation designed into most display bezels. To compound the problem, the plastics used are poor thermal conductors, thus causing the heat to build up which may affect the display.

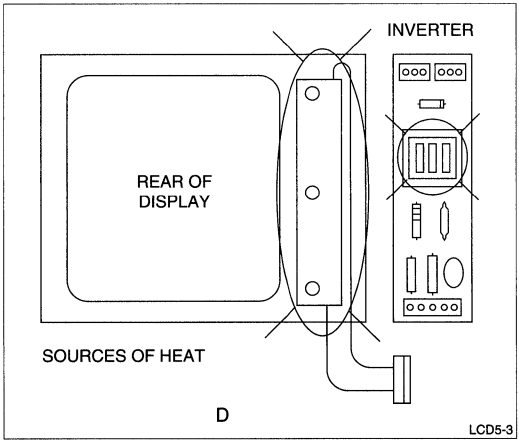


Fig. 3. Typical "Tight" Design

Some current designs suffer from poor placement of the inverter and/or poor thermal management techniques. These designs can be improved, even where redesign of the display housing, with improved thermal management, is impractical.

One of the most common mistakes in current designs is that there has been no consideration for the build up of heat from the CCFT. Typically, the displays for notebook applications have only one CCFT to minimize the power requirements for the display. This lamp is usually placed along the right edge of the display. Since the lamp is placed very close to the display glass, it can cause a temperature rise in the liquid crystal. It is important to note that variations in temperature of as little as five degrees centigrade can cause an apparent non-uniformity in the contrast of the display. Variations caused by slightly higher temperature variations will cause objectionable variations in the contrast and the appearance of the display.

To further aggravate the situation, some designs have the inverter placed in the bottom of the bezel. This has a tendency to cause the same variations in contrast, particularly when the housing does not have any heat sinking for the inverter. This problem will manifest itself as a "blooming" of the display, just above the inverter. This "blooming" looks like a washed out area where in the worst case, the characters on the display fade completely, thus making the display unreadable in that area.

The following section will discuss the recommended methods for overcoming these design problems.

PLACEMENT OF DISPLAY COMPONENTS

A. The Inverter

One of the things that can be done is to design the inverter into the base of the computer with the motherboard. In some applications, however, this is impractical in that this design would require the high voltage leads to be mounted within the hinges connecting the display bezel to the main body of the computer. This causes a problem with strain relief of the high voltage leads, and thus with U.L. Certification.

One mistake, made most often, is placing the inverter at the bottom of the bezel next to the lower edge of the display. It is a fact that heat rises, yet this is one of the most overlooked problems in new notebook designs. Even though the inverters are very efficient, some energy is lost in the inverter in the form of heat. Because of the insulating properties of the plastic materials used in the bezel construction, heat will build up and affect the display contrast as previously explained.

Designs with the inverter at the bottom can be improved in one of three ways. The inverter can be relocated away from the display; heat sinking materials can be placed between the display and the inverter (Figure 5); or ventilation can be provided to remove the heat.

In mature designs, it may be impractical to do what is obvious and move the inverter up to the side of the display towards the top of the housing. In these cases, the inverter may be insulated from the display with a "heat dam". One method of accomplishing this would be to use a piece of mica insulator dye cut to fit tightly between the inverter and the display. This heat dam would divert the heat around the end of the display bezel to rise harmlessly to the top of the housing. Mica is recommended in this application because of its thermal and electrical insulating properties.

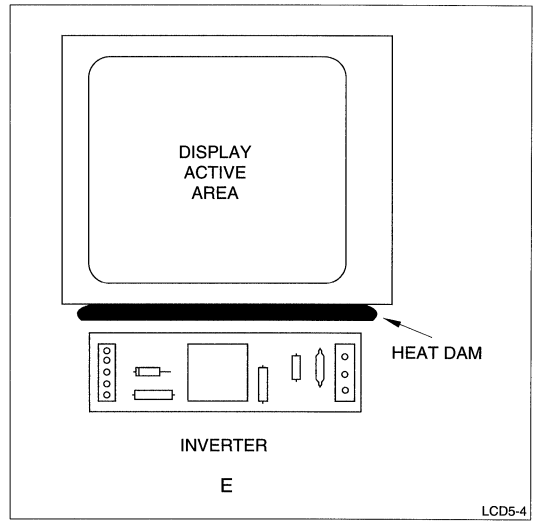


Fig. 4. Heat Dam

The last suggestion as to how to remove the heat would be to provide some ventilation to the inverter area. This would have to be done very carefully to prevent exposing the high voltage. Ventilation may not be a practical solution in that resistance to liquids and dust would be compromised.

The best solution for the designer of new hardware would be to only consider the placement of the inverter to the side of the display and at the top of the bezel. In existing designs, the effects of heat from the inverter, even in tight housings, has been minimal or nonexistent.

B. Heat Sinks For The CCFT

One problem that is aggravated by the placement of the inverter at the bottom of the bezel would be the heat dissipated by the CCFT. In designs where the inverter is placed up and to the side of the display, fading of the display contrast due to the heat from the CCFT is not a problem. However, when the inverter is placed at the bottom of the bezel, some designs have experienced a loss of contrast aggravated by the heat from the CCFT combined with the heat from the inverter.

In cases where the inverter must be left at the bottom, and the CCFT is causing a loss of contrast, the problem can be minimized by using an aluminum foil heat sink, not to remove the heat from the display (Figure 5), but to dissipate it over the entire display area, thus normalizing the display contrast. The aluminum foil is easy to install and in some present designs has successfully improved the display contrast.

Remember that the objection to the contrast variation stems more from nonuniformity than from a total loss of contrast. (Note: Contrast can be adjusted with VDE). The aluminum foil represented in Figure 3 represents a configuration that has worked well.

PROTECTING THE FACE OF THE DISPLAY

One of the last considerations in the design of notebook and more importantly, pen based computers is the protection of the face of the display. The front polarizer is made of a mylar base and thus is susceptible to scratching. The front protection for the display, along with providing scratch protection, may also provide an antiglare surface.

There are several ways that scratch resistance and antiglare surfaces can be incorporated. A glass or plastic cover may be placed over the display, thus providing protection. The material should be placed as close to the display as possible to minimize possible parallax problems associated with reflections off of the cover material. With antiglare materials, the further the material is from the front of the display, the more "frosting" of the antiglare material will take place causing distortion of the display image.

In pen applications, the front anti-scratch material is best placed in contact with the front glass of the display. The cover glass material normally needs to be slightly thicker to protect the display from distortion when pressure is being exerted on the front.

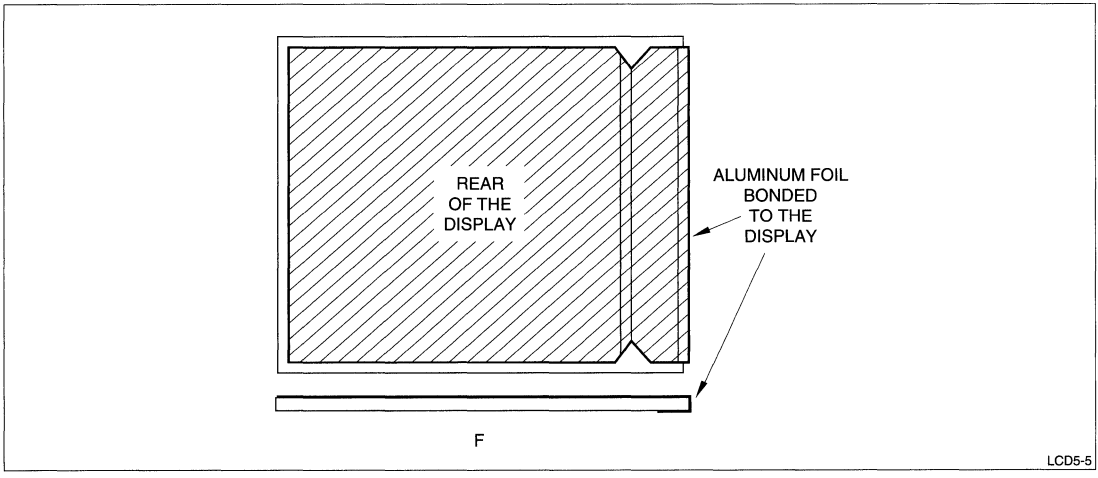


Fig. 5. Aluminum Foil Heat Sink

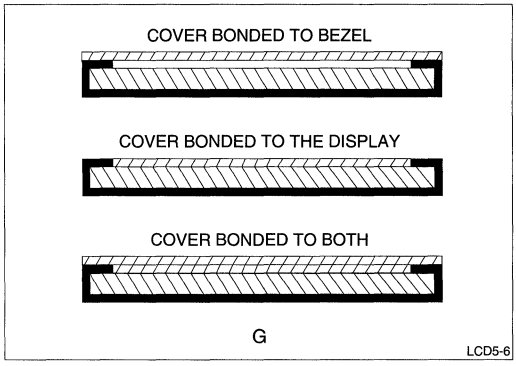


Fig. 6. Aluminum Foil Heat Sink

There are several methods for making the pen input devices. Some use the front surface of the cover glass to provide input data, and some use a field effect to a printed wiring board on the back of the display. When the pen input is on the front of the display, the input device is usually on a glass surface.

To limit specular reflection in this application, the front cover glass should be bonded to the display. Care must be taken to insure that the coefficient of thermal expansion is matched for all of the materials used in this system.

Because of the difficulties encountered with the bonding of the cover glass, and the potential to destroy the display through improper workmanship, we strongly recommend that you consult an engineer, experienced in the bonding of a cover glass to a liquid crystal display module.

CONCLUSIONS

There are several problems that are easily overcome with careful planning and design. This application note is intended to assist the designer and to minimize problems already experienced by others.

FLAT PANEL DISPLAY CONTROLLERS FOR PC APPLICATION

David Blass, Product Marketing Engineer

The popularity and demand for flat panel displays replacing CRTs grows annually. With the increasing popularity of these display devices, there is an ever increasing challenge to have the flat displays maintain the same advanced performance characteristics as the CRT with little to no concession. The fact remains, however, that the flat panel display is widely different from the CRT. In addition, all of the flat display technologies, LCD, EL and Plasma, differ greatly from each other. Since all of the displays have different operational and cost differences, there is no obvious display for all applications. A state-of-the-art controller must be able to run all of the displays, maximizing their particular strengths while minimizing their weaknesses. This has been the ultimate challenge for the leading chip manufacturers.

The key in maximizing the operation of the various displays is the understanding of the differences between the CRT and the various flat display technologies. A CRT has a large active area that can display different resolutions, given it's timing parameters. The CRT gun can place pixels and rows of data anywhere it is directed by the control signals. In this way, low resolution displays are placed on high resolution CRTs and the "blank" areas of the displays can be "filled" or compensated. The flat display is different because they all have a fixed matrix of rows and columns of pixels. This means that lower resolution video cannot give the panel the necessary signals to drive the display. Apart from the physical differences between the CRT and flat displays, another major obstacle is the "translation" of color CRT information to monochrome displays. Color to grayscale representation must be transparent to the software and the video controller, while at the same time, be accurate for the viewer. That is to say that the color representation in black and white must have the same tone quality and intensity so the video information is not compromised. Finally each flat technology exhibits different performance variances that prevent the display from appearing identically like a CRT. Response timing, contrast, and phosphors properties require unique handling to either minimize a performance liability or as a method of enhancing a displays characteristics.

In the 1986 time frame, the chip industry tackled the driving of the flat display with the early introduction of the CGA compatible laptop computers. Due to poor operational characteristics, LCD displays had only limited demand prior to that point. Given comparable resolutions, the timing requirements for EL and Plasma technologies were only slightly different than the CRT. As the demand for LCD CGA controllers grew, early chip manufacturers succeeded in lowering the component count of the CGA function and were able to add the required control lines to operate the LCD display. It had to have a unique controller that could drive the display while being totally transparent to the host PC. From this beginning, new families of controller chips for the PC compatible, was born. Later, controllers grew to meet the requirements of the EGA specification. In this case, the LCD models from Japan were all 640 x 400. The EGA specification was 640 x 350, so the first tack of the new controllers had to compensate for this resolution/timing difference. The flat display EGA controller was designed to map out the 350 lines of data onto the 400 line display by double writing every seventh line. This filled the 400 display with 400 lines of data but the PC worked with the standard 350 of data from its standpoint.

Resently, the demands on video controllers have increased. VGA is now the ubiquitous video standard. Because of this, it is all but impossible to buy a monochrome LCD in any other resolution but 640 x 480. The other technologies; TFT color LCD, passive duty-drive LCD color displays, EL and Plasma need to be driven in the VGA environment. To keep up with these new display technologies while not "playing favorites" by not supporting one display type, the modern chip manufacturers have to introduce entirely new families of display controllers. These new controllers have the latest architecture that allows them to keep up with panels that have not yet been released.

The trends in display controllers dictate that the display controller must be able to operate all of the current displays and be demanded by the market to drive soon-to-be-released models and technologies. Secondly, the manufacturers must respond to the con-

stant pressure from the market to be competitively priced while increasing performance. Other concerns for today's controllers manufacturers are the migration to 3.3 V operation, supporting higher resolutions with the XGA-compatible specification, simultaneously operating a CRT with a flat display, supporting very large color palettes and the ability to be bus compatible for the newest trends in PC architecture. All the while, the performance and compatibility of the hardware and the software cannot be jeopardized.

To these ends, the modern chip manufacturer is introducing families of controllers rather than just one device. Each family has a core design that is code and register compatible with all the other members. This core design was developed to have all of the design features that are required in today's market, while maintaining a competitive price. For scaled back, limited operation applications a chip is offered with reduced specifications, primarily targeted towards black and white LCD operation in a notebook operation. This device typically offers less function but at a reduced price. In some cases, a subset of this device is offered in extremely low power/chip count operation for limited LCD only VGA function. This is the number one market demand by the notebook industry. On the other extreme, there is a superset of the basic device that maximizes colors, speed and upward compatibility with the next generation of video standards. These devices will cost more than the base device but will out perform them as well.

The purpose of this application note is to outline the display controllers currently released to the market by the top video controller manufacturers. Each manufacturer, presented in alphabetical order, has prepared by themselves, a one page overview of their leading devices. Following this section is a matrix of specifications. This will allow the engineer to better evaluate each device from the manufacturer as well as comparing them to other manufacturers in an "apple to apple" comparison. This document has been reviewed by its participants for fairness and accuracy. As any specification document, various parameters may change over time. Though every effort was taken to be correct and thorough, Sharp Electronics cannot guarantee the accuracy of the information presented.

CHIPS AND TECHNOLOGIES
CHIPS 65520/530 MONOCHROME/COLOR FLAT PANEL/CRT CONTROLLER

OVERVIEW

The 65520 or pin-compatible 65530 supports up to 1280 x 1024 16 grayscales or 800 x 600 64 grayscales on Sharp's monochrome LCD and EL panels. The 65520 increases the color palette of Sharp's 512-color TFT LCDs to 24,389 colors. The pin-compatible 65530 supports Sharp's single or dual-drive color STN LCDs with a 226,981 color palette and 512-color TFT LCDs with a 185,193 color palette. The 65520 or 65530 provides a variety of fully programmable options to enhance display quality, such as:

- Gray scaling algorithm to reduce flicker on fast-response,"mouse quick" LCDs
- Vertical compensation techniques to completely fill the panel with lower resolution software via line replication, blank line insertion or TallFonts™
- SMARTMAP™ color to grayscale conversion to optimize the foreground/background contrast in text modes
- The 65520 or 65530 provides simultaneous CRT display with single or dual-drive LCDs (3 MHz or faster).

The 65520 or 65530 requires only three external components - CHIPS' 82C404 programmable clock synthesizer and two memories - for a complete VGA sub-system on the motherboard. Memory options include two or four 256K x 4 DRAMs or 256K x 4 VRAMs or two 512K x 8 DRAMs. VRAMs provide significantly higher performance and lower power consumption than DRAMs for video memory.

The 65520 and 65530 provide high performance for Graphics User Interface (GUI) application software. The controllers employ a FIFO and write buffer architecture providing zero wait-state operation with the ISA bus. The 65520 supports the 386SL's "PI" local bus, and the 65530 supports the 386SL's "PI" and 386DX/SX's local CPU buses.

The 65520 and 65530 are both optimized for minimum power consumption. The 65530 provides "mixed" 3.3 V and 5.0 V operation, so that the VGA, video memory and bus interface operate 3.3 V and the panel interface operates at 5.0 V. The 65520 operates at either 5.0 V or 3.3 V. Both the 65520 and 65530 supports self-refresh 256K x 4 DRAMs and 512K x 8 DRAMs during suspend/resume power down.

Table 1.
CHIPS 65520/530 Monochrome/Color Flat Panel/CRT Controller

FEATURES	BENEFITS
Drives Monochrome & Color LCD, EL & Plasma Panels	Flexible Panel Support
1024 x 768 16 Colors/800 x 600 256 Colors On CRTs	Super VGA Resolution Support
Simultaneous LCD/CRT Display	Facilitates Making Presentations
Programmable Gray Scaling/Color Generation Algorithm	Supports "Mouse Quick" LCDs
Four (4) Chip VGA Sub-System	Minimum Chip Count/Board Space
Supports Various Video Memory Configuration	Range of Price/Performance Options
Local Bus Support; Write Buffer/FIFO; 0-Wait State Pin	High Graphics Performance
3.3 V or 5.0 V Operation	Extends Battery-Base Operation
Full VGA and Backwards Compatibility	Runs All Application Software

CHIPS AND TECHNOLOGIES
CHIPS 82C9001A PC VIDEO WINDOWING CONTROLLER

OVERVIEW

CHIPS PC Video controller provides real-time video data acquisition, scan rate conversion and display windowing control for displaying live video with VGA graphics on flat panel displays and CRT monitors. PC Video supports the industry-standard video formats (NTSC, PAL, SECAM, S-VHS & RGB). The size of the video window is controlled by PC Video's input cropping and scaling features. The position of the video window is controlled by independent X-Y coordinates and by color keying. PC Video's ability to save and restore acquired data to/from system disk or memory enables a PC Video sub-system to serve as a low-cost frame grabber. PC Video and the 82C457 provide a cost-effective method for displaying video from a tape, CD-ROM, camera of digital tuner in a VGA graphics window on a CRT monitor and Sharp's color TFT LCD.

Table 2.
CHIPS 82C9001A PC Video Windowing Controller

FEATURES	BENEFITS
Merges Video Input With VGA Graphics	Provides Cost-Effective Multi-Media Solution
Very Cost Effective	Reduces End-User Price From > $2,000 to < $500
Still Frame Capture & Display	Serves As Frame Grabber
Programmable Window Size & Position	Provides End-Users With Full Control
Direct Interface With CHIPS 82C457	Supports Simultaneous Display on LCDs & CRTs

CHIPS AND TECHNOLOGIES
CHIPS 82C457 Full-Color VGA Flat Panel/CRT Controller

OVERVIEW

The 82C457 along with CHIPS' 82C9001A Video Windowing Controller displays a live video image with VGA graphics simultaneously on a CRT monitor and Sharp's color TFT LCD. The 82C457 increases the color palette of Sharp's 512-color TFT LCDs to 24,389 colors. The 82C457's programmable vertical compensation techniques stretch lower resolution software to completely fill 640 x 480 resolution LCDs.

Table 3.
CHIPS 82C457 Full-color VGA Flat Panel/CRT Controller

FEATURES	BENEFITS
Direct Interface To PC Video CHIPSet	Full Motion Video Capability on LCD/CRTs
24,389 Color On TFT LCDs	Direct Interface To Sharp's Color TFT LCDs
800 x 600 x 16 Colors On CRT Monitors	Super VGA Resolution Support
Full VGA and Backwards Compatibility	Runs All Application Software
Five (5) Chip VGA Sub-System	Low Chip Count/Board Space

CHIPS AND TECHNOLOGIES
CHIPS F8680 SINGLE CHIP PC

OVERVIEW

The F8680 integrates into a single chip all of the functionality required to implement a PC compatible computer except memory. The chip integrates a high-performance 8086 compatible CPU, CGA Graphics Controller, PC Systems Logic, 16C450-Compatible UART, and Power Management into a single 160 PQFP.

The CGA controller supports 640 x 400, 640 x 200 and smaller resolution single drive LCDs. The controller features programmable Frame Rate Control and "Visual Map". Visual map provides excellent visual contrast on any LCD Panel. All 256 possible foreground/background contrast combinations are programmable. Only one 32K x 8 120 ns SRAM is required for the video memory.

Table 4.
CHIPS F8680 Single Chip PC

FEATURES	BENEFITS
16 Programmable grayscales using Visual Map	Excellent Visual Contrast
Programmable Frame Rate Control	Reduced Flicker
Sleep/Suspend Mode Support	Increased Battery Life
Small Panel Support	Non-Standard LCD Display
High Integration	Reduced Power and Board Size

CHIPS AND TECHNOLOGIES
CHIPS 82C426 CGA LCD/CRT CONTROLLER

The 82C426 provides eight flicker-free grayscales or colors on LCD panels with up to 640 x 400 resolutions on CRT monitors. The 82C426's SMARTMAP™ feature improves the contrast of text on monochrome LCDs. The 82C426 employs a single 32K x 8 SRAM for video memory, enabling a highly integrated CGA sub-system.

Table 5.
CHIPS 82C426 CGA LCD/CRT Controller

FEATURES	BENEFITS
Up to 640 x 400 Resolution LCDs	Direct Interface To Small, Low-Cost LCDs
Up to 8 grayscales or Colors	Flicker-Free Display Quality
SMARTMAP™	Improvides Contrast of Text on LCDs
8 x 8 or 8 x 16 Font	VGA Quality Text In CGA Sub-System
SLEEP Mode	Extends Battery Llfe

CIRRUS LOGIC
CL-GD6410 LCD VGA CONTROLLER
FOR NOTEBOOK COMPUTERS

OVERVIEW

The CL-GD6410 is a single-chip VGA controller optimized for use in notebook computers, where reduced form factor and low power consumption are critical design objectives. With the CL-GD6410, a complete motherboard VGA controller requires only four or five ICs, and can fit within four square inches (excluding power sources and connectors). A true two-DRAM video memory (256K x 4), on chip RAMDAC, direct-connect ISA (PC AT) bus interface, and direct-connect LCD interface all help to minimize the form factor.

By using Cirrus Logic's Frame Accelerator technique, the CL-GD6410 is able to provide a high vertical refresh rate for dual-scan LCD panels while operating at approximately one-half the clock speed of other LCD controller solutions; this provides a significant reduction in full-active power consumption and extends battery life. In addition, standby and suspend modes are supported in the hardware of the CL-GD6410 to enable multiple levels of system power management.

The CL-GD6410 provides 64 shades of gray on monochrome LCD panels. Duty-cycle modulation, combined with dynamic pattern-management algorithms, provide 640 x 480-resolution grayscales with no apparent flicker. Pixel-doubling and stripping techniques provide increased grayscale in the VGA high-color Mode 13. In all cases, the Cirrus Logic grayscale provides consistent linear-step functions, making smooth transitions from black, through the grayscale, to white. With a direct connection to 512-color TFT LCD panels, the CL-GD6410 provides a single-controller solution for monochrome and TFT color portable computers. The CL-GD6410 also provides a direct interface to the CL-GD6340, Cirrus Logic's color LCD interface controller, for STN color LCD panels.

The CL-GD6410 supports SimulSCAN™, a technique introduced by Cirrus Logic for achieving simultaneous CRT and LCD operation. SimulSCAN™ supports both single and dual-scan LCDs, and both fixed and multifrequency analog CRTs. Monochrome LCDs may be operated in reverse video (pagewhite) simultaneously with normal CRT operation.

FEATURES

- Single-chip VGA controller
- IBM-VGA-hardware-compatible
- Simultaneous CRT and LCD (SimulSCAN™) operation
- Two 265K x 4 DRAM video memory for small form factor
- Integral RAMDAC
- Integral LCD panel interface
  - Control and data buffering
  - Power sequencing logic
- Direct connection to ISA (PC AT) bus
- Frame-Accelerator for low active power
- Standby and suspend modes to save power
- Expanded operational range: 5 V ±0%
- 64-shade grayscale on monochrome STN LCD
  - NTSC sum-to-gray color mapping
  - Multiple sum-to-gray weighting options
- Direct connection to 512-color TFT LCD panels
  - Single-controller design for STN monochrome and TFT color LCD's
- Graphics and text expansion of VGA modes on LCD
- 800 x 600 x 16 color on analog CRT
- 8- or 16-bit CPU interface
- Packaged in 160-pin (EIAJ-standard) QFP package
  - Pin-out optimized for efficient board layout

CIRRUS LOGIC
CL-GD6412 LCD VGA CONTROLLER
FOR MIXED-VOLTAGE NOTEBOOK COMPUTERS

OVERVIEW

The CL-GD6412 is a single-chip VGA controller optimized for use in systems with a mix of 3.3 V and 5.0 V components, where quick implementation of a notebook computer with reduced power consumption is the critical design objective. In a design using the CL-GD6412, the internal logic can use a 3.3 V power supply for reduced power consumption. The video memory, host bus interface, panel interface, and clock interface may each be implemented at either 3.3 V or 5 V. The voltages of these interfaces may be mixed in virtually any combination.

Multiple levels of system-power management are supported in the hardware of the CL-GD6412. Standby mode can be driven by software, or an internal counter. Suspend mode can be driven by software or hardware by means of a suspend pin. The hardware suspend pin allows a nearly complete shutdown of the device, with no bus decoding for very low power consumption. The CL-GD6412 is based on the proven CL-GD6410 architecture, one of the industry's most popular LCD VGA controllers. Basic design and programming models are unchanged. This allows for the highest confidence in quick development schedules.

With the CL-GD6412, a complete motherboard VGA controller requires only four or five ICs, and three external pull-up/pull-down configuration resistors. It can be designed in less than four square inches (excluding power sources and connectors). A two-DRAM video memory interface (256K x 4), on-chip RAMDAC, direct-connect ISA (PC AT) bus interface, and direct-connect LCD interface, all help to minimize the form-factor.

The CL-GD6412 has enhanced algorithms for dynamic pattern-management within the Frame Rate Duty Cycle, providing a grayscale with minimum apparent flicker, even on 4 MHz or quick response 'mouse-quick) panels. In all cases, the Cirrus Logic grayscale provides smooth transitions from black, through the grayscale, to white, for maximum display quality of realistic images.

The CL-GD6412 supports SimulSCAN™, a technique introduced by Cirrus Logic for achieving simultaneous CRT and LCD operation. Reverse video is controlled separately for the LCD display data path, so that page-white LCD operation may occur simultaneously with normal CRT operation.

FEATURES

- Single-chip VGA controller
- Mixed voltage: 3.3 V or 5.0 V on any major interface
  - Supports JEDEC Number 8 low voltage CMOS standard 3.3 V ±0.3 V
  - Mix 3.3 V or 5.0 V system components with no external-level converters
- IBM VGA hardware-compatible
- Two 265K x 4 DRAM video memory for small form-factor
- Integrates RAMDAC
- Integrates LCD panel interface
  - Control and data buffering
  - Power sequencing logic
- Direct connection to ISA (PC AT) bus
- Frame-Accelerator for low-active power
- 64 shade grayscale on monochrome STN LCD
  - NTSC sum-to-gray color mapping
  - Multiple sum to gray weighting options
- Enhanced flicker-reduction algorithms for 4 MHz and quick response LCDs
- Direct connection to 512-color TFT (Thin Film Transistor) LCD
  - Single-controller design for STN momochrome and TFT color LCDs
- Graphics expansion and compression maps CRT modes to fixed-resolution LCD
- Packaged in 160-pin (EIAJ-standard) QFP package

CIRRUS LOGIC
CL-GD6420 LCD VGA CONTROLLER
FOR HIGH-RESOLUTION NOTEBOOK COMPUTERS

OVERVIEW

The CL-GD6420 is a single-chip VGA controller optimized for use in high-end notebook computers, where high resolution CRT capabilities and high preformance are critical design objectives. The CL-GD6420 is based on the proven architecture of the CL-GD6410. The CL-GD6420 adds a scaleable video memory capability, supporting up to eight 265K x 4 DRAMs.

By using the Cirrus Logic Frame-Accelerator technique, the CL-GD6420 is able to provide a high vertical refresh rate for dual-scan LCD panels while operating at approximately one-half the clock speed of non-accelerated LCD controller solutions. This provides a significant reduction in full-active power consumption, extending the battery life of notebook computers. Standby and Suspend Modes are supported in the hardware of the CL-GD6420, to enable mutiple levels of system power management. Standby mode can be initiated by software, by a programmable on-chip timer, or a separate standby pin.

The CL-GD6420 provides 64 shades of gray monochrome LCD panels. Duty cycle modulation, combined with improved dynamic pattern management algorithms, provide 16 shades of gray with minimum perceivable flicker, even on 4 MHz and fast response (mouse-quick) LCD panels. Grayscale enhancement provides 64 apparent shades of gray on the LCD for 640 x 480 x 265 color extended mode operation. Pixel-doubling and stippling techniques provide increased grayscale in the VGA Mode 13H.

With a direct connection to 512-color TFT LCD panels, the CL-GD6420 provides a single-controller solution for 64 grayscale monochrome and 256-simultaneous-color portable computers. Extended color mode support allows 640 x 480 resolution with 256 colors TFT LCD panels. For color STN LCD panels, the CL-GD6420 provides a direct interface to the CL-GD6340 color LCD interface controller. The CL-GD6420 supports SimulSCAN™ operation, a technique introduced by Cirrus Logic for achieving simultaneous CRT and LCD operation. SimulSCAN™ supports both single and dual-scan LCDs, and both fixed multifrequency analog CRTs. Monochrome LCDs may be operated in reverse video (page-white) simultaneously with normal CRT operation.

FEATURES

- Single-chip IBM VGA hardware-compatable controller
- Up to 1 MByte (2, 4, or 8) 256K x 4 DRAM Video Memory
- Extended resolution up to 1024 x 768 with 265 colors on CRT
- Simultaneous display on LCD panel and CRT
- Integral RAMDAC
- Integral LCD panel interface
  - Control and data buffering
  - Power sequencing logic
- Direct connection to ISA (PC AT) bus
- Standby and Suspend Modes to save power
  - Internal standby counter or hardware standby pin
- Expanded operational range: 5 V ±10%
- 64-shade grayscale on monochrome STN LCD
  - NTSC sum-to-gray color mapping
  - Multiple sum-to-gray weighting options
- Enhanced flicker-reduction algorithms for 4 MHz and quick response LCDs
- Direct connection to 512-color TFT LCD panels
- Graphics and text expansion on LCD
- Packaged in 160-pin (EIAJ standard) QFP package

CIRRUS LOGIC
CL-GD6340 COLOR LCD INTERFACE CONTROLLER

OVERVIEW

The CL-GD6340 is a flat-panel interface chip for use with the CL-GD64xx family of LCD VGA graphics chips to provide full-color displays on color LCD panels. By using multiple techniques for color shading, the CL-GD6340 is able to increase the number of colors an LCD can display. This interface controller can also produce up to 64 shades of gray on monochrome LCD panels. The CL-GD6340 supports active matrix panels with up to 1, 2, 3, or 5 bits per pixel (8, 512, 4K, or 32K-color panel capability). It supports color passive (STN) LCD with single and dual-scan architectures.

The CL-GD6340 handles subtle graduations of shade and hue, expanding an LCD panel's total color (or grayscale) capability, and providing a display quality that compares with CRTs. With a complete Cirrus Logic color LCD VGA solution, even eight-color panels can support all VGA modes (including mode 13), displaying 256 simultaneous from a palette of thousands.

The CL-GD6340 features a highly-programmable panel interface, with timing and power-sequencing logic. The CL-GD6340 can provide SimulSCAN™ operation, driving analog CRT and digital LCD in parallel. SimulSCAN™ supports both single and dual-scan LCDs, and both fixed and multifrequency analog CRTs. Monochrome LCDs may be operated in reverse video (page-white) simultaneously with normal CRT operation.

FEATURES

- On-chip support for active-matrix LCD panels
  - 8, 512, 4K, or 32K-color panels
  - Monochrome panels with or without grayscaling capability
- On-chip support for passive-matrix LCD panels
  - Dual scan
  - Single-chip
- On-chip buffers and logic to support monochrome- and color-passive LCD panels (single- and dual-panel displays)
- Provides full-color VGA on color panels
  - 256 simultaneously displayed colors in VGA mode 13
  - 16 simultaneously displayed colors in VGA mode 12
  - Selected from palette of thousands of colors
- Supports all VGA modes on monochrome LCD panels
  - 64 shade grayscale on passive monochrome panels
  - 16 shade grayscale on active-matrix panels without grayscaling capability
- Simultaneous displays on analog CRT and LCD
- Easy interface with Cirrus Logic VGA controller chips
- On-chip RAMDAC and RAMDAC extension registers
- 32 MHz operation
- Packaged in EIAJ-standard 100-pin plastic Quad Flat Pack (QFP) package

Table 6.
Western Digital

FEATURES			90C26	90C26A	90C24	COMMENTS
Host Interface	Local Bus 16-bit I/O 16-bit Memory 32-bit Memory Zero Wait State Write Buffer Virtual Memory ISA Bus Micro Channel 386SL PI Bus Intel PCI Bus		NO YES YES NO YES YES YES YES YES NO NO	NO YES YES NO YES YES YES YES YES YES NO	INTEGRATED YES YES YES YES YES YES YES YES YES NO	True 32 bit local bus 90C26/C26A - 2 cycles, 90C24 - programmable to 4 cycle depth
Display Memory	Memory Achitecture Maximum Memory data Interface Minimum DRAM required Maximum MCLK frequency	 CRT Flat Display	FAST PAGE 16 BIT 2-256X4 50 MHz 50 MHz	FAST PAGE 16 BIT 2-256X4 50 MHz 50 MHz	FAST PAGE 32 BIT 1-256X16 85 MHz 85 MHz	 C24 can support simultaneous display w/o additional buffer
Package			144 pin QFP	144 pin QFP	208 pin QFP	
Built-In RAMDAC			YES	YES	YES	90C24 built-in RAMDAC designed for 3.3V and 5.0V operation
Built-In Clock Generator			NO	NO	YES	Built-in clock is designed for both 3.3V and 5.0V operation, clock frequencies are programmable
Power Supply Voltages	5 VDC 3.3 VDC Mixed Voltages		YES YES NO	YES YES YES	YES YES YES	 3.3V (-10%) to 5.0V (+10%), Integrated level shifters
Power Down Modes	Panel Off- VGA active Standby- VGA Powered Down Memory maintained		YES YES	YES YES	YES YES	 WD offers "Deep Sleep" mode: 50 micro-amps power consumption
Panel Power Sequence Control			NO	NO	NO	Panel power sequencing managed in conjunction with core logic
CRT Display	640 times 400 640 times 480 800 times 600 1024 times 768	Simultaneous colors/Color Palette	256/256K 256/256K 256/256K 16/256K	256/256K 256/256K 256/256K 16/256K	64K/256K 64K/256K 256/256K 256/256K	90C24 supports 16 bit True Color in modes up to 640 X 480 90C24 supports 16 bit True Color in modes up to 640 X 480 1024 X 768 non-interlaced
Flat Panel Display Support	1/240 TSTN B/W LCD Monochrome EL Gray Scale EL Color TFT 1/480 STN Duty Color 1/240 STN Duty Color		YES YES YES YES YES NO	YES YES YES YES YES NO	YES YES YES YES YES YES	 Directly drives color TFT panels, no additional logic required 90C26/C26A interface to 90C55 STN interface device, 90C24 directly drives color STN panels with no additional logic

Table 6 (cont'd). Western Digital

FEATURES			90C26	90C26A	90C24	COMMENTS
Panel Interface	8-bit 9-bit 12-bit 16-bit		YES YES YES YES	YES YES YES YES	YES YES YES YES	
Monochrome Panels	640 x 400 640 x 480 1024 x 768	# of Gray Shades * *	64 64 N/A	64 64 N/A	64 64 64	WD TrueShade grayscales utilize 64 distinct SGF pixel dithering patterns
STN Duty Color 640 x 480	VGA Mode 13 Super VGA	Simultaneous colors/Color Palette *	256/256K 255/256K	256/256K 256/256K	64K/256K 256/256K	90C24 supports 16 bit true color True color supported in 640 times 480 SVGA mode only
TFT Color 640 x 480	VGA Mode 13 Supter VGA	Simultaneous colors/Color Palette *	256/185K 256/27K	256/185K 256/27K	64K/256K 256/256K	90C24 supports 16 bit true color True color supported in 640 x 480 SVGA mode only
Simultaneous Display	CRT and Mono LCD CRT and EL CRT and 1/480 Duty Color CRT and 1/240 Duty Color CRT and TFT Color		YES YES YES NO YES	YES YES YES NO YES	YES YES YES YES YES	90C26/C26A & 90C24 support simultaneous display with full support for full screen vertical expansion of displayed images. The optional frame buffer for simultaneous display may be powered off or used as video memory when not in simultaneous display mode. The 90C24 does not require a separate frame buffer for simultaneous display.
Display Enhancement Features	Greyscale/Color Generation Algorithm Color to Grayscale Reductions Programmable mapping RAM Vertical Centering Horizontal Centering Vertical Stretching Tall Fonts Blank Line Insertion White Text Enhancement	 NTSC Equal Green Only Text Mode Text Mode	YES YES YES YES YES NO YES YES YES NO YES	YES YES YES YES YES NO YES YES YES NO YES	YES YES YES YES YES NO YES YES YES NO YES	 Provides full prog. of grayscales. Select best 16/64 in std. VGA All devices support both hardware and software vertical expansion techniques. Proprietary line duplication algorithm provides superior display quality vs. competitive Nth line techniques.
Window Acceleration	Bitblt Hardware Cursor Line Draw		NO NO NO	NO NO NO	YES YES YES	BitBlt w/packed pixel mode for improved performance Programmable cursor size, 32 x 32 or 64 x 64 MS Line Assist & Bressenham
Release To Market/Available			March/June	June/Sept	Aug/Nov	

Table 7.
Cirrus Logic

FEATURES			GD6410	GD6412	GD6420	GD6340	Comments
Host Interface	Local Bus 16-bit I/O 16-bit Memory 32-bit Memory Zero Wait State Write Buffer Virtual Memory ISA Bus Micro Channel 386SL PI Bus Intel PCI Bus		YES/ with PAL YES YES NO YES YES YES YES NO YES/ with PAL NO	YES/with PAL YES NO NO YES YES YES YES NO YES/with PAL NO	YES/with PAL YES YES NO YES YES YES YES NO YES/with PAL NO	YES/with PAL YES YES YES YES YES YES YES YES YES NO	Yes/ with PAL
Display Memory	Memory Architecture Maximum Memory data Interface Minimum DRAM required Maximum MCLK frequency	 CRT Flat Display	DRAM 256K BYTES (2) 256K x 4 44 MHz 32 MHz	DRAM 256K BYTES (2) 256K x 4 40 MHz	DRAM 1 MBYTE (2) 256K x 4 44 MHz	DRAM 1 MBYTE (2) 256k x 4 44 MHz	
Package			160 QFR (EIAJ)	160 QFP (EIAJ)	160 QFP (EIAJ)	160 QFP (EIAJ)	
Built-In RAMDAC			YES	YES	YES	YES	
Built-in Clock Generator			NO	NO	NO	NO	
Power Supply Voltages	5 VDC 3.3 VDC Mix Voltages		YES NO NO	YES YES YES	YES NO NO	YES YES YES	
Power Down Modes	Panel Off - VGA active Standby - VGA Powered Down memory maintained		YES YES	YES YES	YES YES	YES YES	
Panel Power Sequence Control			YES	YES	YES	YES	
CRT Display	640 x 400 640 x 480 800 x 600 1024 x 768	Simultaneous colors/Color Palette * * *	256/256K 16/256K 16/256K n/a	256/256K 16/256K 16/256K n/a	256/256K 16 or 256/256K 16 or 256/256K 16 or 256/256K	256/256K 256/256K 256/256K 16/256K	
Flat Panel Display Support	1/240 TSTN B/W LCD Monochrome EL Gray Scale EL Color TFT 1/480 STN Duty Color 1/240 STN Duty Color		YES YES YES YES NO NO	YES YES YES YES NO NO	YES YES YES YES NO NO	YES YES YES YES YES YES	 GD6340 is color companion chip for GD6410/12/20

Table 7 (cont'd).
Cirrus Logic

FEATURES			GD6410	GD6412	GD6420	GD6340	COMMENTS
Panel Interface	8-bit 9-bit 12-bit 16-bit		YES YES NO NO	YES YES NO NO	YES YES NO NO	YES YES YES YES	
Monochrome Panels	640 x 400 640 x 480 1024 x 768	# of Gray Shades * *	16 or 64 64 16	16 or 64 16 or 64 16	16 or 64 16 or 64 16 or 64	16 or 64 16 or 64 16 or 64	
STN Duty Color 640 x 480	VGA Mode 13 Super VGA	Simultaneous colors/Color Palette *	NO NO	NO NO	NO NO	256/226K 256/4K	
TFT Color 640 x 480	VGA Mode 13 Super VGA	Simultaneous colors/Color Palette *	256/185K 16/3K	256/185K 16/3K	256/185K 256/185K	256/185K 256/185K	
Simultaneous Display	CRT and Mono LCD CRT and EL CRT and 1/480 Duty Color CRT and 1/240 Duty Color CRT and TFT Color		YES YES NO NO YES	YES YES NO NO YES	YES YES NO NO YES	YES YES YES YES YES	
Display Enhancement Features	Grayscale/Color Generation Algorithm Color to Grayscale Reductions Programmable mapping RAM Vertical Centering Horizontal Centering Vertical Stretching Tall Fonts Blank Line Insertion White Text Enhancement	 NTSC Equal Green Only Text Mode Text Mode	YES YES 6 OPTIONS YES YES YES YES YES YES YES YES	YES YES 6 OPTIONS YES YES YES YES YES YES YES YES	YES YES 6 OPTIONS YES YES YES YES YES YES YES YES	YES YES 87 OPTIONS YES YES YES YES YES YES YES YES	 Maximizes test foreground/background
Window Acceleration	Bitblt Hardware Cursor Line Draw		YES YES YES	YES YES YES	YES YES YES	YES YES YES	Enhances contrast of test foreground. Enhances contrast of test background Enhances test foreground and background
Release To Market/Available			Production	Sample 4/92	Sample Now	Sample 6/92	

Table 8.
Chips and Technology

FEATURES			CHIPS 65530	CHIPS 65520	COMMENTS
Host Interface	Local Bus 16-bit I/O 16-bit Memory 32-bit Memory Zero Wait State Write Buffer Virtual Memory ISA Bus Micro Channel 386SL PI Bus Intel PCI Bus		YES YES YES YES YES YES NO YES/ with PAL YES YES NO	YES YES YES NO YES YES NO YES YES YES NO	65530 - direct; 65520 requires external PALs Direct interface Direct interface Increases performance Increases performance Provides linear address space Direct interface Direct interface Direct interface
Display Memory	Memory Architecture Maximum Memory data Interface Minimum DRAM required Maximum MCLK frequency	 CRT Flat Display	DRAM/VRAM 1 MBYTES (2) 256K x 4 65 MHz 65 MHz	DRAM/VRAM 1 MBYTES (2) 256K x 4 56 MHz 56 MHz	Performance/power/cost flexibility
Package			160 QFP	160 QFP (EIAJ)	
Built-In RAMDAC			YES	YES	
Built-in Clock Generator			NO	NO	Use external 82C404 clock synthesizer
Power Supply Voltages	5 VDC 3.3 VDC Mix Voltages		YES YES YES	YES YES YES	
Power Down Modes	Panel Off - VGA active Standby - VGA Powered Down memory maintained		YES YES	YES YES	
Panel Power Sequence Control			YES	YES	
CRT Display	640 x 400 640 x 480 800 x 600 1024 x 768	Simultaneous colors/Color Palette * * *	256/256K 256/256K 256/256K 16/256K	256/256K 256/256K 256/256K 16/256K	 65530 supports Non-interlaced & interlaced
Flat Panel Display Support	1/240 TSTN B/W LCD Monochrome EL Gray Scale EL Color TFT 1/480 STN Duty Color 1/240 STN Duty Color		YES YES YES YES YES YES	YES YES YES YES NO NO	Direct interface (no frame buffer required) Direct interface Direct interface Direct interface Direct interface Direct interface

Table 8 (cont'd).
Chips and Technology

FEATURES			CHIPS 65530	CHIPS 65520	COMMENTS
Panel Interface	8-bit 9-bit 12-bit 16-bit		YES YES YES YES	YES YES YES YES	Direct interface Direct interface Direct interface Requires external 374 latch
Monochrome Panels	640 x 400 640 x 480 1024 x 768	# of Gray Shades * *	64 64 16	64 64 16	At 640 x 400 and 320 x 200 resolutions At 640 x 400 and 320 x 200 resolutions
STN Duty Color 640 x 480	VGA Mode 13 Super VGA	Simultaneous colors/Color Palette *	256/226K 256/226K	NO NO	At 320 x 200 resolution At 640 x 480 resolution
TFT Color 640 x 480	VGA Mode 13 Super VGA	Simultaneous colors/Color Palette *	256/185K 256/185K	256/24K 256/24K	At 320 x 200 resolution At 640 x 480 resolution
Simultaneous Display	CRT and Mono LCD CRT and EL CRT and 1/480 Duty Color CRT and 1/240 Duty Color CRT and TFT Color		YES YES YES NO YES	YES YES NO NO YES	Requires 100 ns 64K x 4 VRAM frame buffer Direct Direct Direct
Display Enhancement Features	Greyscale/Color Generation Algorithm Color to Grayscale Reductions Vertical Centering Horizontal Centering Vertical Stretching Tall Fonts Blank Line Insertion White Text Enhancement Color Text Enhancement	 NTSC Equal Green Only Text Mode Text Mode	Programmable YES YES YES YES YES YES YES YES YES YES	Programmable YES YES YES YES YES YES YES YES YES YES	Reduces flicker at low refresh rates Automatic & programmable Automatic & programmable Programmable line replication Loads larger font Programmable Swaps bright white and dim white Programmable color to grayscale mapping
Window Acceleration	Bitblt Hardware Cursor Line Draw		NO NO NO	NO NO NO	
Release To Market/Available			Now (From 11/91)	Now (From 4/92)	

INTERFACE & INTERCONNECTION FOR 4-INCH & 6-INCH TFT / LCDs

Alan Dragon, Field Applications Engineer

INTRODUCTION

Sharp Electronics Corporation has introduced a series of small TFT (Thin Film Transistor) LCD modules to be used in a wide range of video applications. These displays are well suited for use in Portable TV/VCR Entertainment Systems, Test/Control Equipment, Control and Entertainment displays for Aviation, along with Automotive, Navigational and Imaging Applicatons. The current product line-up includes diagonal sizes of 3", 4" and 5.7" (Table 2, page 2). With three formats available (234V x 32.5H, 234V x 720H, 240V x 720H), NTSC and PAL Video Standards can be supported by various models within the product line (Note 1). All 4" and 5.7" models are available with a 6 o'clock or a 12 o'clock viewing direction for optimum performance with any mounting orientation.

Sharp TFT LCDs use the normally white mode of operation for an excellent contrast ratio and superior color reproducibility. This is characterized by a contrast ratio of 30:1 with a light output of 120 nits (35 foot-lamberts). With these specifications, the TFT LCD modules can be used in various lighting environments.

The possibility of battery operation is enhanced by the low power consumption of the TFT display. Total power is typically 2.6 watts, with 1.7 watts of that total being consumed by the backlight.

VIDEO SIGNAL STANDARDS

NTSC/PAL

It will be useful to understand the video standards of NTSC and PAL before explaining the actual interface of the TFT displays. NTSC (National Television System Committee) and PAL (Phase Alternation Line) are two different color encoding methods for broadcasting or sending color video information. Most countries around the world have adopted one of these two standards. The remaining countries have adopted SECAM, which is based on the PAL Standard (Table 3). This application note will concentrate on the timing characteristics of the NTSC and PAL Standards and leave the explanation of color encoding and decoding methods up to video textbooks.

Note 1: The current product line reflects improvements made to earlier models (Table 1).

Table 1.
Transition from Prototype Models to Current Production Models

CURRENT MODEL	PROTOTYPE MODEL
LQ4RE01 NTSC/PAL 6 o'clock	LQ424Y02 [NTSC 6 o'clock] LQ424P01 [PAL 6 o'clock]
LQ4RA01 NTSC/PAL 6 o'clock	LQ424A01 [NTSC 6 o'clock] LQ6MA01 [PAL 6 o'clock]
LQ4RE02 NTSC/PAL 12 o'clock	LQ4NA02 [NTSC 12 o'clock] LQ4MA02 [PAL 12 o'clock]
LQ6RA01 NTSC/PAL 6 o'clock	LQ6NA01 [NTSC 6 o'clock] LQMA01 [PAL 6 o'clock]
LQ6RA02 NTSC/PAL 12 o'clock	LQ6NA01 [NTSC 12 o'clock] LQ6MA02 [PAL 12 o'clock]
LQ6RA02 NTSC/PAL 12 o'clock	LQ6NA02 [NTSC 12 o'clock] LQ6MA02 [PAL 12 o'clock]

The basic difference between NTSC and PAL is the number of lines per frame. NTSC uses 525 lines per frame at a 60 Hz field rate and PAL uses 625 lines per frame at a 50 Hz field rate (Table 4).

When an image is broken up into more horizontal lines, the resolution and image quality improve accordingly. Both standards consist of two interlaced fields. These two fields (designated odd and even) make up one full frame. The alternating odd and even fields make the actual frame rate 1/2 of the field rate (Figure 1). Although the two fields alternate, the human eye will superimpose and blend the odd and even fields to give the appearance of one continuous and flicker-free image.

Table 2.
Sharp's Small TFT Color LCD Modules

MODEL	DIAGONAL SCREEN SIZE (INCHES)	PIXEL FORMAT (V x H)	INPUT CAPABILITY	BACKLIGHT	VIEWING DIRECTION	EFFECTIVE VIEWING AREA (W x H) (mm)	DOT PITCH (W x H) (mm)	OUTLINE DIMENSIONS (W x H x L) (mm)	WEIGHT (GRAMS)	INTERFACE
LQ4RE01	4	234 x 479	NTSC/PAL	N/A	6 o'clock	81.9 x 61.8	0.171 x 0.264	122 x 100 x 6.6	135	N/A
LQ4RE02	4	234 x 479	NTSC/PAL	N/A	12 o'clock	81.9 x 61.8	0.171 x 0.264	122 x 100 x 6.6	135	N/A
LQ4RA01	4	234 x 479	NTSC/PAL	HCFT/Built-in	6 o'clock	81.9 x 61.8	0.171 x 0.264	110.2 x 85.8 x 20.7	170	Analog RGB
LQ4RA02	4	234 x 479	NTSC/PAL	HCFT/Built-in	12 o'clock	81.9 x 61.8	0.171 x 0.264	110.2 x 85.8 x 20.7	170	Analog RGB
LQ4NC01	4	234 x 479	NTSC	HCFT/Built-in	6 o'clock	81.9 x 61.8	0.171 x 0.264	110.2 x 85.8 x 20.7	180	Composite/Analog RGB
LQ4NC02	4	234 x 479	NTSC	HCFT/Built-in	12 o'clock	81.9 x 61.8	0.171 x 0.264	110.2 x 85.8 x 20.7	180	Composite/Analog RGB
LQ6RA01	5.7	240 x 720	NTSC/PAL	CCFT/Built-in	6 o'clock	113.8 x 87.6	0.158 x 0.365	149.4 x 117 x 23	310	Analog RGB
LQ6RA02	5.7	240 x 720	NTSC/PAL	CCFT/Built-in	12 o'clock	113.8 x 87.6	0.158 x 0.365	149.4 x 117 x 23	310	Analog RGB
LQ6NC01	5.7	240 x 720	NTSC	CCFT/Built-in	6 o'clock	113.8 x 87.6	0.158 x 0.365	149.4 x 117 x 23	320	Composite/Analog RGB
LQ6NC02	5.7	240 x 720	NTSC	CCFT/Built-in	12 o'clock	113.8 x 87.6	0.158 x 0.365	149.4 x 117 x 23	320	Composite/Analog RGB
LQ6MC01	5.7	240 x 720	PAL	CCFT/Built-in	6 o'clock	113.8 x 87.6	0.158 x 0.365	149.4 x 117 x 23	320	Composite/Analog RGB
LQ6MC02	5.7	240 x 720	PAL	CCFT/Built-in	12 o'clock	113.8 x 87.6	0.158 x 0.365	149.4 x 117 x 23	320	Composite/Analog RGB
LQ323Y11	3	234 x 382.5	NISC	N/A	6 o'clock	61.7 x 44.5	0.161 x 0.190	94.2 x 78.5 x 61	80	N/A
LQ323P07	3	234 x 382.5	PAL	N/A	6 o'clock	61.7 x 44.5	0.161 x 0.190	94.2 x 78.5 x 61	80	N/A

Note: All specifications are subject to change.

Table 3. International Television Standards

LINES PER FRAME: 525 FIELD RATE: 60 Hz COLOR CODING: NTSC	LINES PER FRAME: 625 FIELD RATE: 50 Hz COLOR CODING: PAL	LINES PER FRAME: 625 FIELD RATE: 50 Hz COLOR CODING: SECAM
Antiqua, West Indies	Algeria	Afars and Issas
Bahamas	Australia	Arab Republic of Egypt
Barbados	Austria	Bulgaria
British Virgin Islands	Bahrain	Czechoslovakia
Canada	Bangladesh	East Germany
Chile	Brunei	France
Costa Rica	Brazil (525/60)	Greece
Cuba	Denmark	Haiti
Dominican Republic	Federal Republic of Germany	Hungary
Ecuadoron Republic	Finland	Iran
El Salvador	Hong Kong	Ivory Coast
Guatemala	Iceland	Iraq
Japan	Ireland	Lebanon
Mexico	Italy	Luxembourg
Netherlands Antiles, West Indies	Jordan	Mauritius
Nicaragua	Kuwait	Monaco
Panama	Malaysia	Morocco
Peru	Netherlands	Poland
Phillipines	New Zealand	Reunion
St. Kitts, West Indies	Nigeria	Saudi Arabia
Samoa (U.S.)	Norway	Tunisia
Surinam	Oman	USSR
Province of Taiwan	Pakistan	Zaire
Trinidad, West Indies	Oatar	
Trust Territory of Pacific	Singapore	
United States of America	South Africa	
	Spain	
	Sweden	
	Switzerland	
	Tanzania	
	Thailand	
	Turkey	
	United Arab Emirates	
	United Kingdom	
	Yugoslavia	
	Zambia	

Table 4. NTSC and PAL Timing Standards

NTSC	PAL
Lines / Frame: 525	Lines / Frame: 625
Lines / Field: 262.5	Lines / Field: 312.5
Field Rate: 60 Hz	Field Rate: 50 Hz
Frame Rate: 30 Hz	Frame Rate: 25 Hz
Display Period: 24 OH	Display Period: 28 OH
Horizontal Interval: 63.5 μs (1H)	Horizontal Interval: 64.0 μs (1H)
Vertical Interval: 16.7 ms (262.5H)	Vertical Interval: 20.0 ms (312.5H)
Vertical Blanking: 1.42 ms (22.5H)	Vertical Blanking: 2.08 ms (32.5H)
Vertical Sync Pulse: 254 μs (4H)	Vertical Sync Pulse: 256 μs (4H)

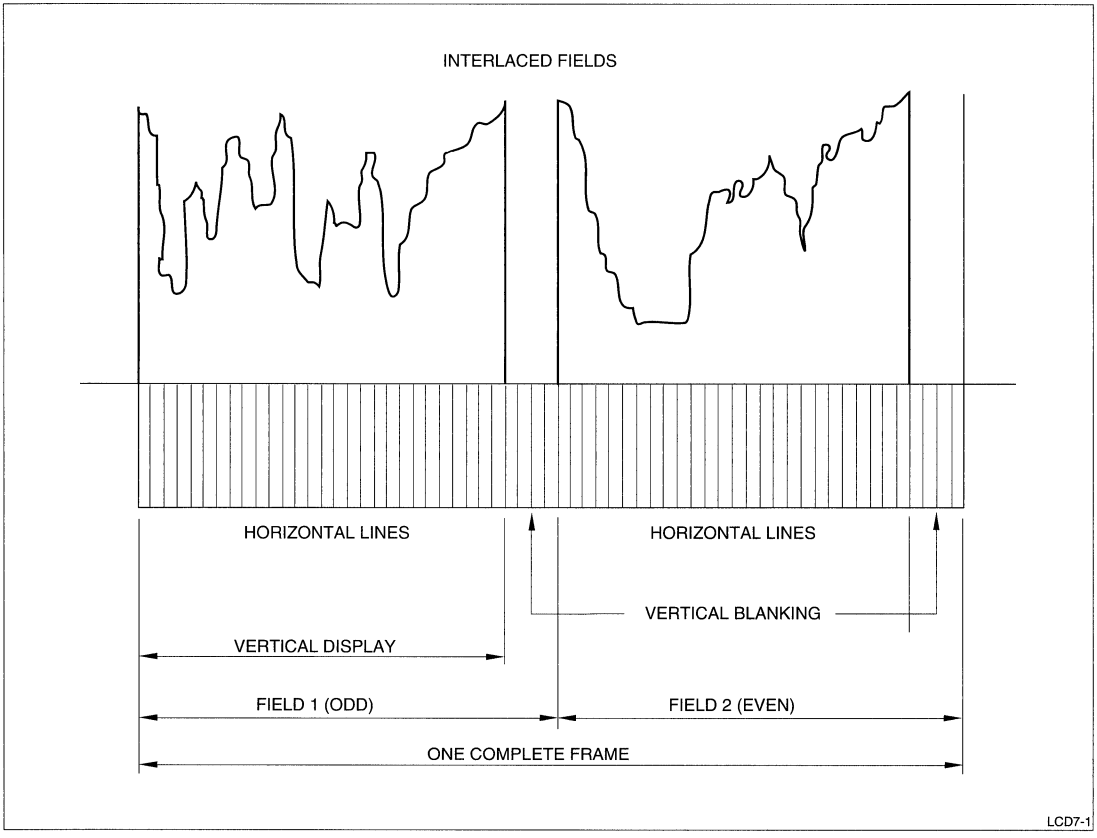


Fig. 1. Interlaced Fields

Composite/Non-Composite Video Signals

Video signals can be transmitted in a variety of formats. Sharp TFT LCD modules will support the following video configuration:

- Composite Video - Video information is combined with horizontal and vertical sync and color burst information into one signal (Figure 2).
- Analog RGB (non-composite) - Separate red, green and blue video signals used in conjuction with composite sync or separate horizontal and vertical sync (Figure 3).
- Composite Sync - Horizontal and vertical sync are combined into one signal (Figure 4).

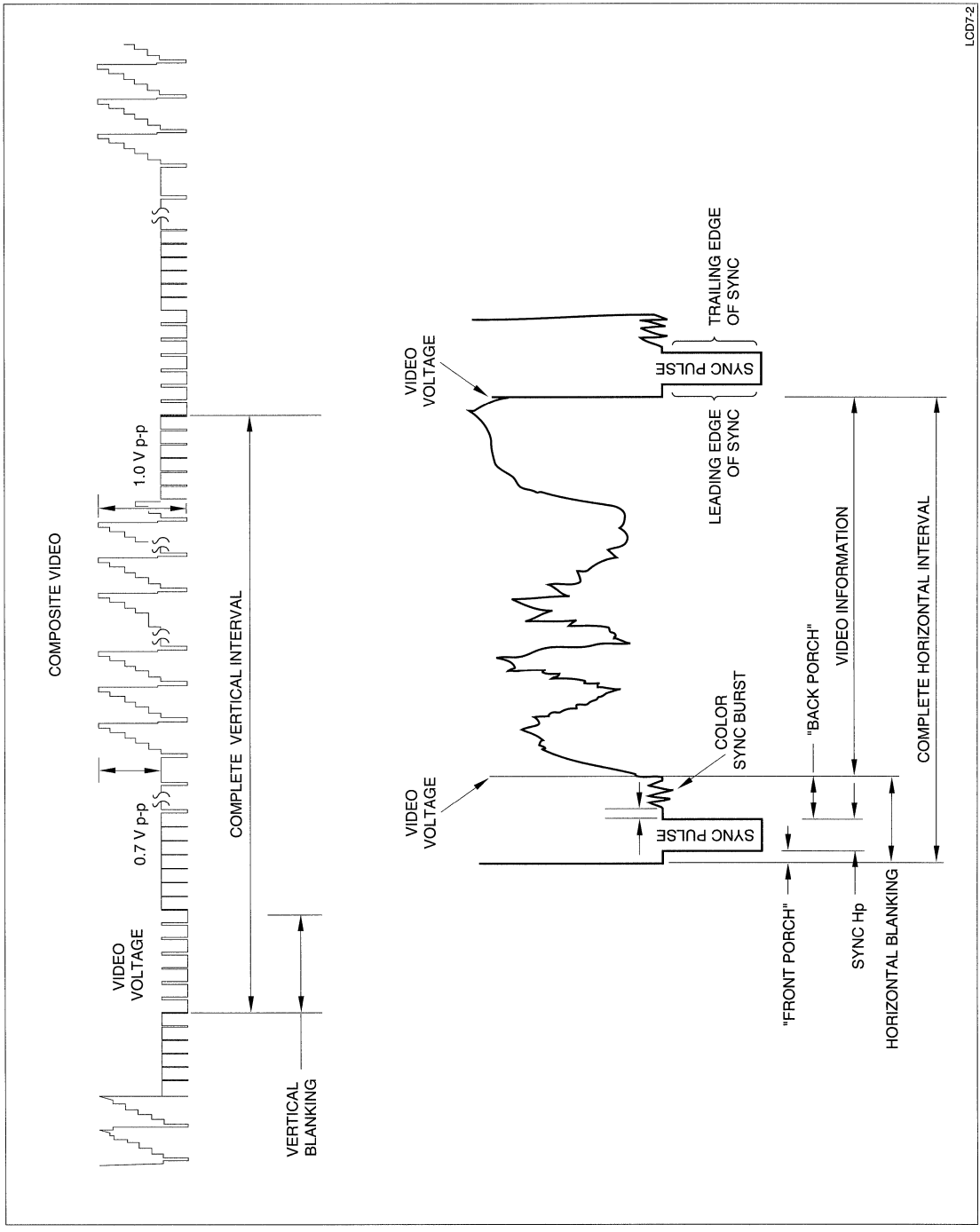


Fig. 2. Composite Video

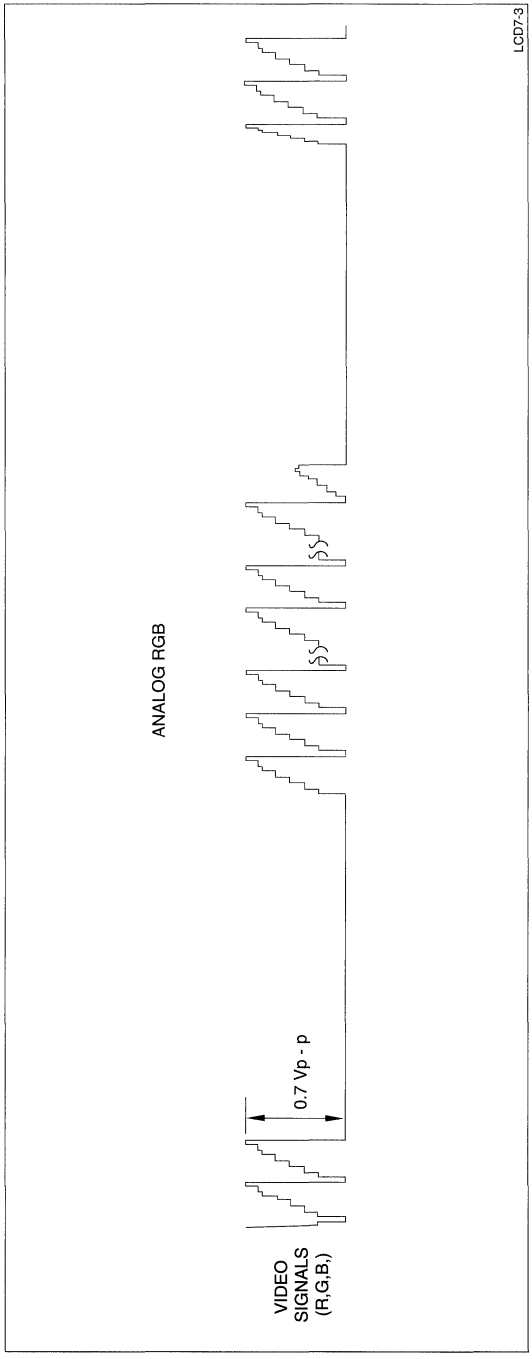


Fig. 3. Analog RGB (non-composite)

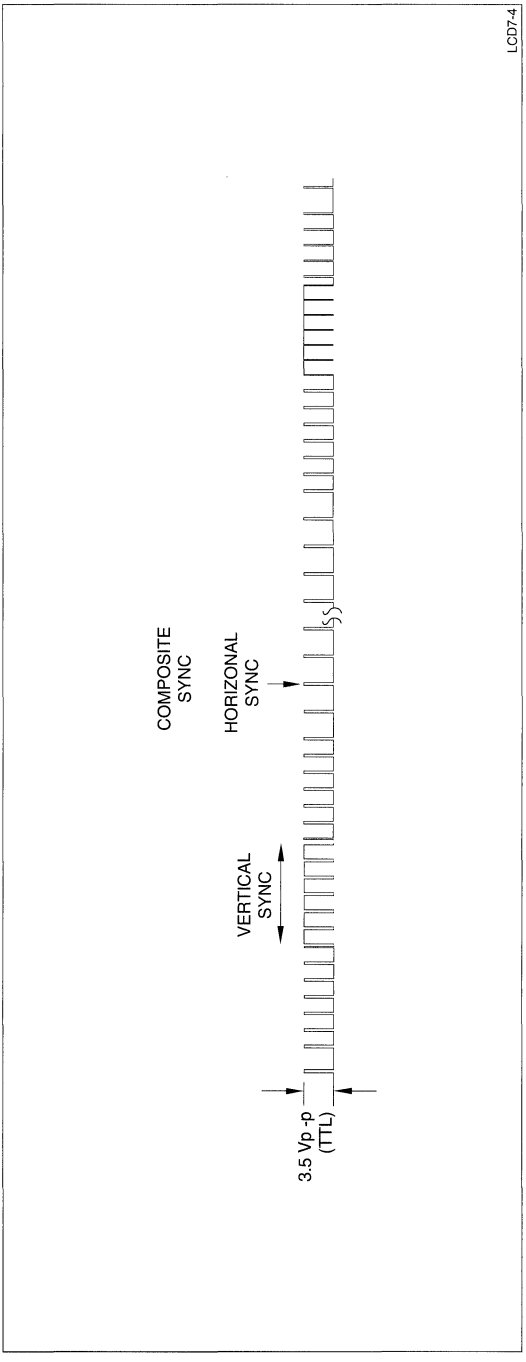


Fig. 4. Composite Sync

GENERAL INFORMATION

Viewing Direction

All 4" and 5.7" color TFT LCD modules are offered with a choice of viewing direction. The last digit of the part number will indicate the viewing direction. A "1" will indicate a 6 o'clock viewing direction while a "2" indicates a 12 o'clock viewing direction. The relationship of the viewing direction optimizes the contrast ratio for +10° to -30° in the vertical axis and the 12 o'clock viewing direction optimizes the contrast ratio for +30° to -10° in the vertical axis (Figure 5). The horizontal viewing angle remains even and consistent for all TFT modules.

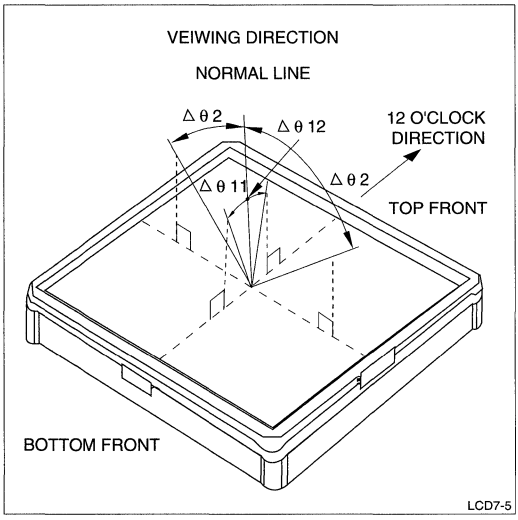


Fig. 5. Viewing Direction

Interlaced Operation

As described earlier, NTSC and PAL video signals consist of two interlace fields. When these fields are viewed on a monitor by the human eye, they appear as one continuous image. In a CRT monitor, these two fields are interwoven on the screen (Figure 6). In a Sharp TFT LCD there are either 240 or 234 lines printed on the glass depending on the model (Note 2). Since the odd and even fields consist of 240 lines each, the fields are scanned on the same lines of the LCD one after another (Figure 7). As with a CRT, the eye integrates these two fields into one complete frame and the resulting image appears to be a full 480 lines on the LCD.

Displaying the video information on the PAL version of the LCD is more involved. The odd and even fields

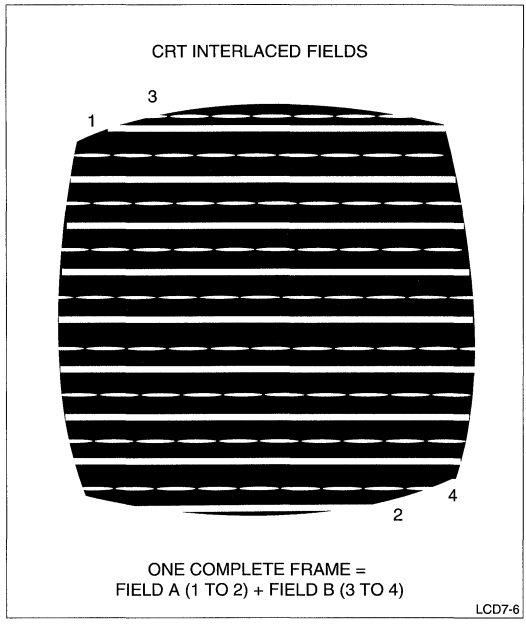


Fig. 6. Interlaced Operation

are made up of 280 lines each. In order to display the information on the 240 line LCD as done in the NTSC mode, the video signal must be reduced to 240 lines/field. This is done by removing 40 lines/field through a method called MBK-PAL (MaBiki "Thinning in Japanese). In this method every seventh line is removed by using the following algorithm (Figure 8):

(14n + 11)H, (12n + 19)H even field

(n=1, 2, 3, ... 20)

(14n + 16)H, (14n + 22)H odd field

Pixel Configuration

Each pixel of Sharp's small color TFT LCDs consists of 3 sub-pixels. These red, green, and blue sub-pixels can operate at any output level to give full color capability to each pixel. A 234 V x 382.5 H, 234 V x 479 H or 240 V x 720 H pixel format is availiable depending of the particular TFT LCD model. The pixel format is defined as the number of sub-pixels (RGB). The actual number of pixels is as follows: 234 V x 128 H (128 x 3 RGB approx. 382.5), 234 V x 160 H (160 x 3 RGB approx. 479), 240 V x 240 H (240 x 3 RGB approx. 720). The LCDs use a delta pixel configuration with each addressable pixel residing on the same line (Figure 9). This configuration makes the LCD displays well suited for video information.

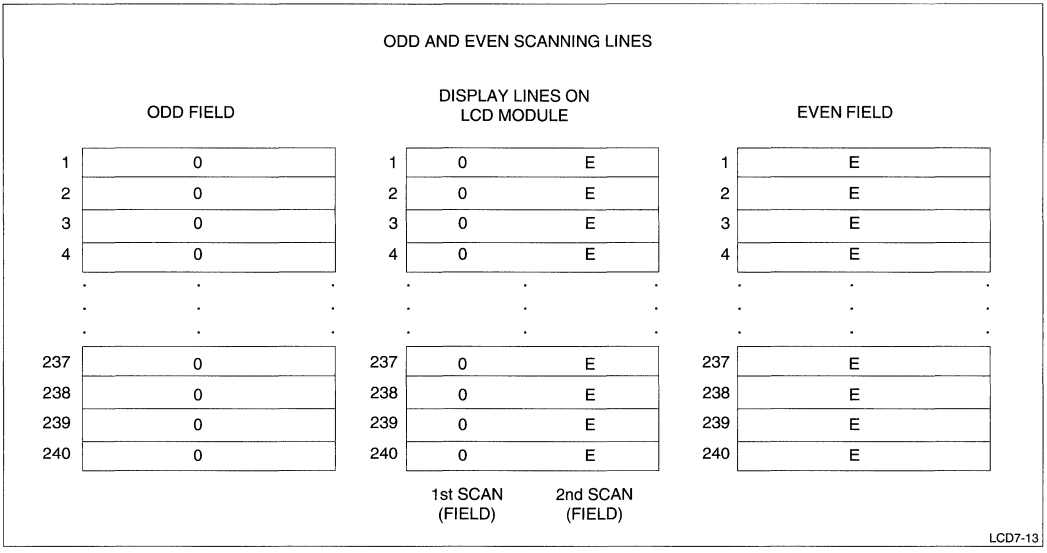


Fig. 7. Odd and Even Scanning Lines

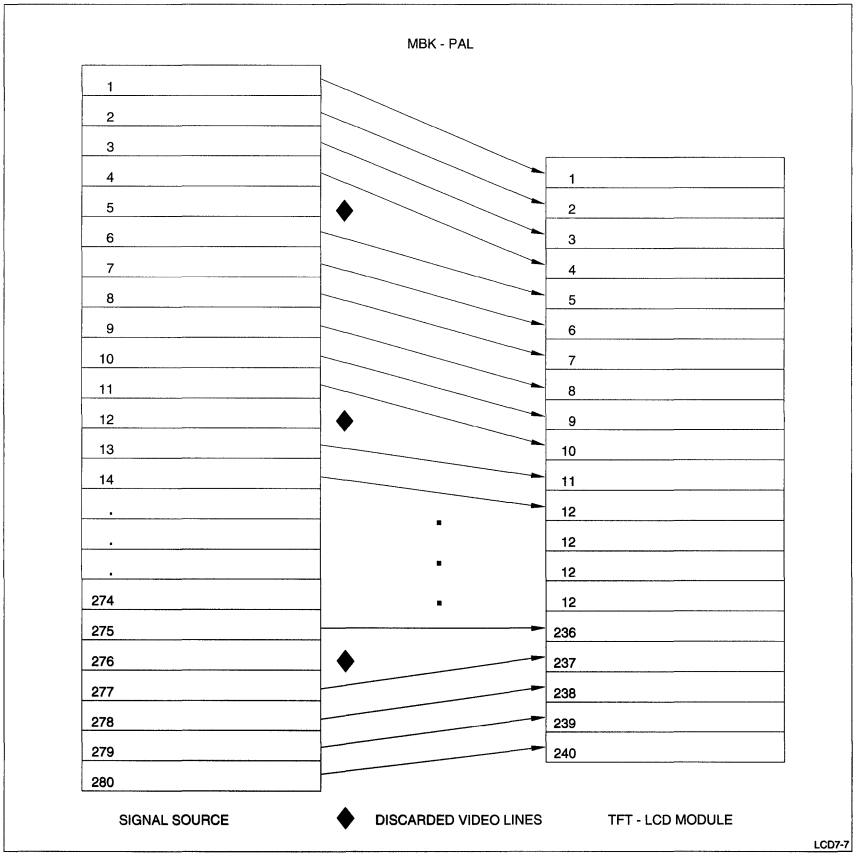


Fig. 8. MBK-PAL Method

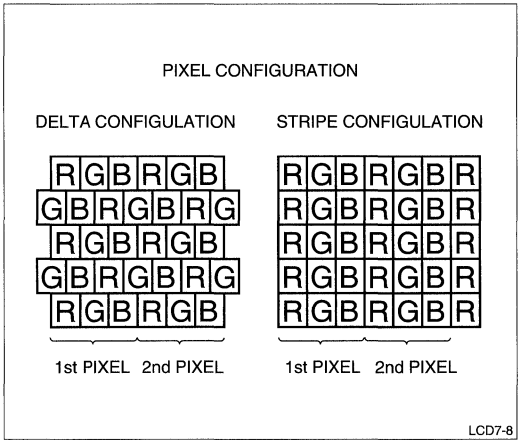


Fig. 9. Delta Pixel Configuration

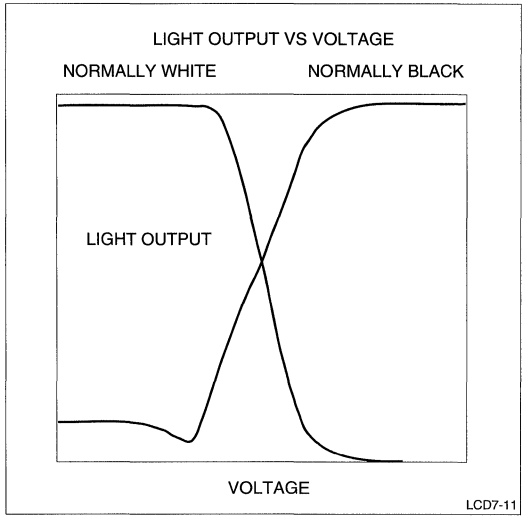


Fig. 10. Light Output vs. Voltage

Normally White Mode

Sharp Color TFT LCD modules operate in the normally white mode. This gives Sharp TFT displays a higher contrast ratio than those using the normally black mode. The main difference between normally white mode and normally black mode is the orientation of the linear polarizers. Normally white polarizers are perpendicular to each other (Figures 11a). When the power is removed in a normally black mode (Figure 11b), the liquid crystal material does a 90° twist. There is still some light at low voltages that will pass through in this configuration (Figure 10). The normally white mode will not let this light through in the off state due to the perpendicular arrangement of the polarizers.

Note 2: The LCD models which have 234 lines drop the last 6 lines in each field to accept the 240 line NTSC input.

Gamma Correction

Gamma correction refers to the exponential variation in the signal input to provide a linear transfer function from the signal source to the display-luminance output. The luminous output of the LCD is not directly proportional to the display-signal input under normal conditions. This is due to the characteristics of the LC material. A similar condition occurs in a CRT monitor where the luminous output of the phosphors are not directly proportional to the signal input. In order to compensate for this non-linear response, a gamma correction is applied to the monitor (LCD or CRT) to give a linear luminous output (figure 16). Although values vary among monitors, the gamma correction factor for Sharp LCDs is 2.2. This is the same as the accepted value for most CRT monitors.

Gamma is defined as the power exponent used to approximate the curve of display-luminance output versus signal input amplitude. When using a factor of 2.2, for example, a 50% - of-full-scale input voltage will result in a 22%-of-full-scale luminous output (0.5/2.2 = 0.22).

All Sharp small color TFT modules have the gamma correction built into the module with the exception of LQ323Y11/P07 and LQ4RE01/2. Sharp's basic design policy is all color TFT LCDs without built-in interfaces will have to be supplied with an external gamma correction, whereas all color TFT LCDs with a built-in interface do not need an external gamma correction.

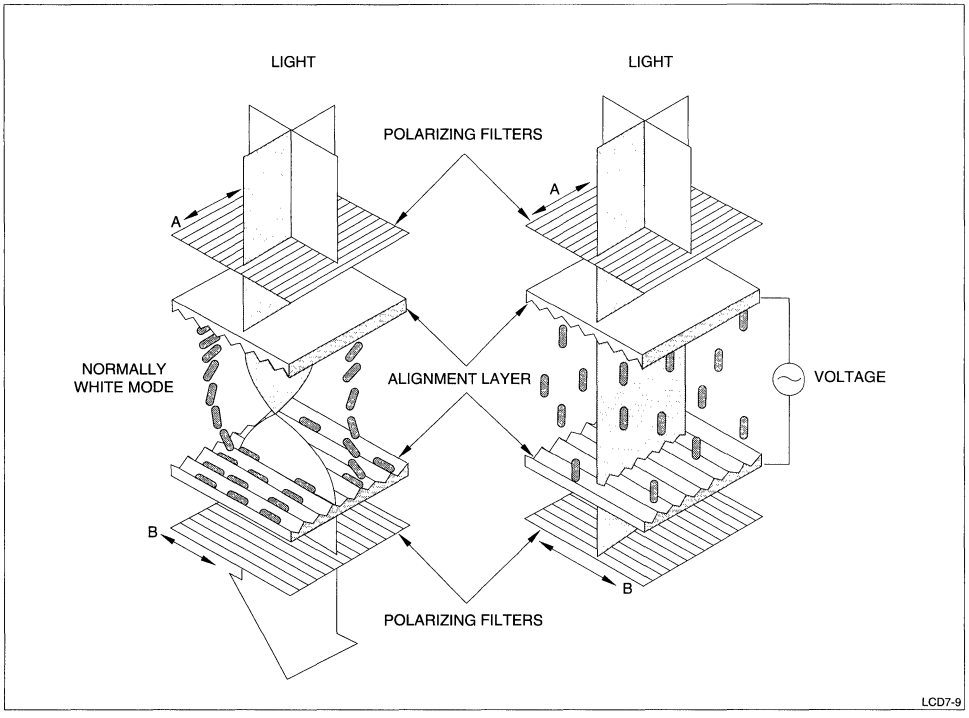


Fig. 11a. Normally White Mode

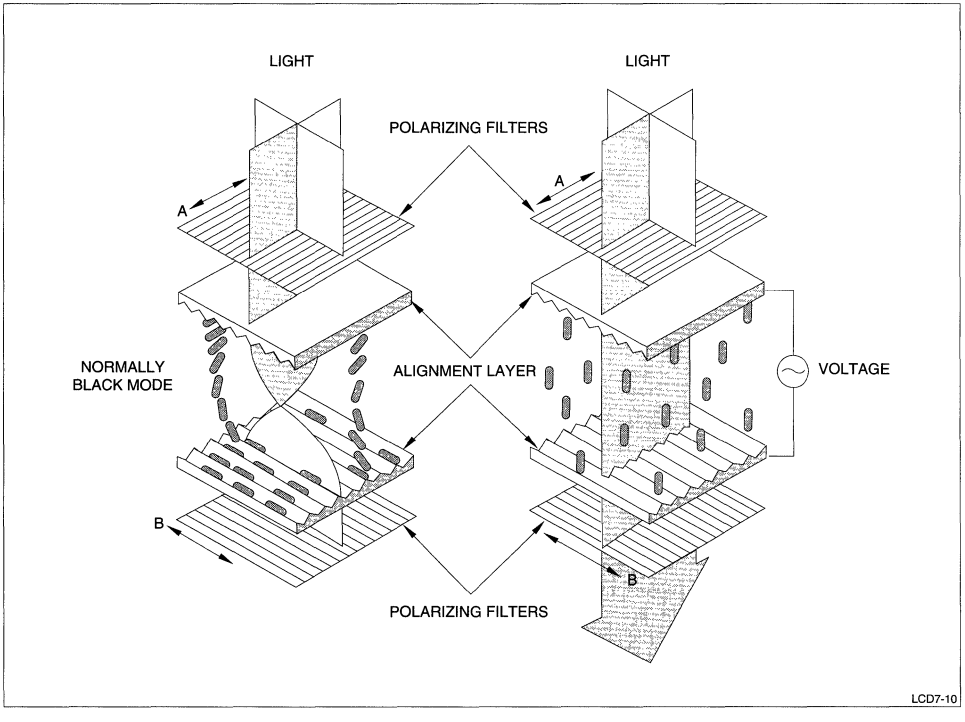


Fig. 11b. Normally Black Mode

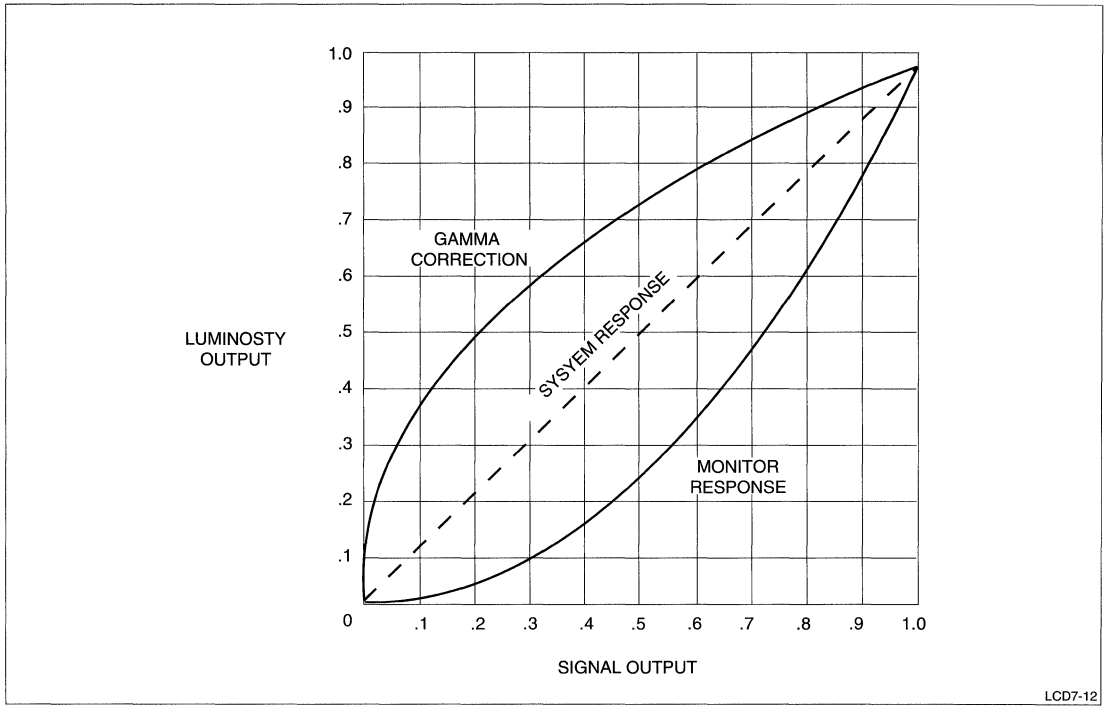


Fig. 12. Linear Luminous Output

Connectors

All the TFT modules use Flat Printed Circuit (FPC) cables and connectors for signal I/O connection. The various modules come with either 16 or 20 conductor cables/connectors with a 1.0 mm or 1.25 mm pitch. Refer to Table 5 for the configuration with each model. The 1.0 mm and 1.25 mm Pitch FPC cables/connectors can be obtained from varous companies including the following:

- ELCO (Laguna Hills, CA, 714-830-8383)
  1.00 mm/1.25 mm FPC connectors
- Molex (Lisle, IL, 708-969-4550)
  1.0 mm/1.25 mm FPC connectors
- Parlex (Salem, NH 603-893-0040)
  1.0 mm/1.25 mm FPC cables, custom cables with connectors
- Sumitomo (Freehold, NJ, 908-409-3990)
  1.00 mm/1.25 mm FPC cables/connectors.

All the LCD module connectors/cables, along with their mating connectors/cables, are referenced in their individual specifications.

Table 5.
COLOR TFT LCD PRODUCTS THAT REQUIRE MATING FLEX CIRCUIT CABLES

DISPLAY MODEL	CONNECTOR/CABLE ON LCD	MATING CONNECTOR/CABLE
LQ323Y11	16 Conductor FPC Cable; 1.0 mm pitch	16 Conductor FPC Connector; 1.0 mm pitch
LQ323P07	16 Conductor FPC Cable; 1.0 mm pitch	16 Conductor FPC Connector; 1.0 mm pitch
LQ4RE01/2	16 Conductor FPC Cable; 1.0 mm pitch	16 Conductor FPC Connector; 1.0 mm pitch
LQ4RA01/2	20 Conductor FPC Connector; 1.25 mm pitch	20 Conductor FPC able; 1.25 mm pitch
LQ4NC01/2	20 Conductor FPC Connector; 1.0 mm pitch	20 Conductor FPC Cable; 1.0 mm pitch
LQ4RB03	16 Conductor FPC Connector; 1.0 mm pitch	16 Conductor FPC Cable; 1.0 mm pitch
LQ4RB11	16 Conductor FPC Connector; 1.0 mm pitch	16 Conductor FPC Cable; 1.0 mm pitch
LQ6RA01/2	20 Conductor FPC Connector; 1.25 mm pitch	20 Conductor FPC able; 1.25 mm pitch
LQ6NC01/2	20 Conductor FPC Connector; 1.0 mm pitch	20 Conductor FPC Cable; 1.0 mm pitch

Backlight

The backlight on the 4" and 5.7" TFT LCD modules is replaceable. The replacement backlight modules can be obtained from Sharp as needed. The backlight on the 4" modules uses HCFTs (hot cathode fluorescent tubes) and on the 6" modules the backlight uses CCFTs (cold cathode fluorescent tubes). The main differences between the two technologies are: 1) HCFTs have to pre-heat the filament, 2) HCFTs tend to have a higher light output, and 3) HCFTs will have a shorter life.

The Sharp part number for the 4" replacement backlight module is LQ0B01 and the part number for the 6" replacement module is LQ0B04.

The input for backlight modules varies from 110 Vrms @20-50 kHz for the HCFT modules to 330 Vrms @20-50 kHZ for the CCFT modules. These modules also require an ignition voltage of 600 - 800 Vrms for initial discharge of the backlight tubes.

Inverters to drive these voltages are presently available from Sharp and also from Endicott Research Group, Endicott, NY (607-754-9187). These inverters usually run off an input of either +5 V, +12 V or +24 V DC. The Sharp part numbers for the inverters are LQ0J06 for the 4" modules and LQ0J04 for the 6" modules. The LQ0J06 comes with an input voltage connector and two output cable assembles to connect directly to the backlight module. The LQ0J04 also comes with an input voltage connector, but can connect directly to the backlight module without cables.

Handling/Cleaning

The face of the Color TFT LCD modules consists of an outer polarizer. The polarizer is made from a soft material and care should be taken not to scratch this surface. Any scratch to the surface will adversely affect the operation and quality of the display. All mounting configurations should include some type of clear protective covering to prevent any damage to the polarizer and module. The protective laminate, shipped with the module, should only be removed immediately before the protective covering is placed in front of the face of the module.

If some dust happens to get on the surface of the display, it should be blown off with an ionized air gun to prevent any static build-up of dust. Wiping the unit with a cloth will scratch and damage the polarizer. If, in the worst case, dirt or a smudge appears on the glass, Sharp recommends using a soft, dry cloth with a very small amount of petroleum benzene.

As with any LCD display, do not expose the display to prolonged direct sunlight/UV rays as the filters and LC material will degrade over time. A clear protective antiglare filter that also filters UV may provide the best all-around solution to protect the LCD from the environment.

INTERFACING TO THE DISPLAY MODULE

Interfacing to the various modules varies from model to model. Each model will be examined and their interfacing requirements defined.

LQ323Y11/P07

These two display modules support either NTSC (LQ323Y11) or PAL (LQ323P07) video standards. They require an external interface to control the input data and sync signals, along with an external backlight. DC input voltages of +5.0 V, +13.0 V, -8.0 V and -20.0 V also must be provided. To prevent damage to the LCD, the power voltages to the source and gate drivers must be applied in the following power on/off sequence:

On: +5.0 V-->+13.0 V-->-20.0 V-->-8.0 V

Off: -8.0 V-->-20.0 V-->+13.0 V-->+5.0 V

Separate RGB video signals, alternating in polarity, must be used to input the data. The FRP signal output (TTL level clock signal = 1/2 horizontal sync frequency) should be combined with the RGB video signals to give the alternating video inputs. The alternating video signals prevent electrochemical degradation of the LC material which can be caused by applying either all negative or all positive voltages.

Composite horizontal and vertical sync is also required as an input to these units. A separate horizontal and vertical sync output from the module can be used to sync video overlays on the screen.

Vcom (common electrode driving signal) is a required clock signal, in phase with the alternating video signal, which controls the amplitude of the video signal. Controlling the amplitude of this signal will control the contrast of the LCD module.

These displays can be easily controlled by using one of Sharp's interface ICs for LCDs. These interface ICs have onboard gamma correction and polarity inversion among other useful features. The IR3P89 is a general purpose color LCD interface IC which processes RGB video into RGB signals suitable for input into the LCD module. This part only performs basic functions and does not have any chroma processing capabilities. The IR3P90B (NTSC) and IR3P96 (PAL) can accept either a composite video input or separate RGB input and has the capability of performing extensive video/chroma processing. Depending on the application, one of these ICs will be able to efficiently fulfill the 3" TFT LCD interface requirements.

There are only two adjustments available to modify the visual performance of the modules. Vcom, as described earlier, is an input signal that will change the display contrast. H-POS is an adjustment on the module that can readjust the image if it is not centered horizontally. All other controls are preset during production and should not be adjusted by the user to maintain optimum performance of the module.

LQ4RE01/2

These modules accept either NTSC or PAL video inputs by changing the TTL level of the mode change terminal. An external backlight along with a series of input voltages and processing signals are required to drive these units. The input DC supply voltages required are +5.0 V, +13.0 V, -8.0 V and -20.0 V. These voltages supply the power for the source and gate drivers and should be applied in the following power on/off sequence:

On +5.0 V-->+13.0 V-->-20.0 V-->-8.0 V

Off -8.0 V-->-20.0 V-->+13.0 V-->+5.0 V

The video input signals required are composite sync along with separate RGB signals alternating in polarity by using the FRP output signal. The FRP signal is a TTL level clock signal operating at 1/2 the horizontal sync frequency. The FRP signal when combined with the analog video signal should give a combined alternating signal of 5 Vp-p. The polarity of the video signal must be alternated to prevent the LC material from electro chemical degradation.

As with the 3" modules, Sharp interface chips IR3P89, IR3P90B and IR3P96 can be used on the LQ4RE01/2 models to perform the required processing of the interface signals.

Horizontal and vertical sync are also available as outputs to synchronize video overlays on the screen such as time/channel indication for a TV application.

There are a series of adjustments that can be used to adjust the visual characteristics of the LCD. Vcdc (common electrode driving signal) is a DC bias input that can be used to optimize the contrast of the display. This adjustment is optimized during manufacturing and should not be readjusted by the user under normal operating conditions. If operating conditions change, such as a reduction in the input power voltage level, Vcdc can be readjusted to optimize the performance of the display module by using the adjustment method described in the TFT-LCD module specification. The H-Pos (horizontal position) control on the module can adjust the position of the picture on the display. All other controls are preset at the factory and should not be touched in order to assure optimum performance.

LQ4RA01/2, LQ6RA01/2

These 4" and 5.7" modules are compatible with both NTSC and PAL Video standards. The replaceable backlight is already incorporated into these units and the user need only supply a DC-AC inverter to power the backlight. The video section requires DC power supply voltages of +5.0 V and -8.0 V applied in the following power on/off sequence:

On +5.0 V-->-8.0 V

Off -8.0 V-->+5.0 V

The video drive signals consist of composite sync and separate RGB analog video signals. The sync accepts a standard composite sync signal or a standard composite video input (1.0 Vp-p). The video inputs consist of separate (3) RGB analog signals (0.7 Vp-p). There are two sets of video RGB inputs that can be multiplexed by a TTL control signal (VSW) for displaying two separate video sources. The horizontal and vertical sync outputs can be used to synchronize overlays on the display an might be done in a television application.

As with all of the color TFT LCD modules, there are adjustments to control the display visual characteristics. Brightness (BRT) and contrast (Vcdc) can be adjusted by changing the DC voltage to these two inputs. For optimum performance Sharp recommends not readjusting Vcdc under normal operating conditions. There are also a number of adjustments on the back of the module. Most of these have been adjusted at the factory for optimum performance and display quality. The adjustments that are user accessible are contrast, brightness and horizontal position. The contrast (Vcdc) and brightness adjustments as indicated, can be adjusted either externally or on the module itself.

LQ4NC01/2, LQ6NC01/2, LQ6MC01/2

This series of 4" and 5.7" modules operates from a composite video input. The "N" series is compatible with a NTSC input and the "M" series is compatible with a PAL input. The backlight is in a modular form and can be easily replaced.

The video section requires a DC power supply of +5.0 V and -6.0 V to be powered on/off simultaneously. As with all of the modules, the video signals should be applied after the supply voltages. The video input signals consist of a composite video input along with separate (3) RGB analog video inputs. The separate RGB video inputs synchronized with the horizontal and vertical sync outputs can be multiplexed with the composite video input to give onscreen overlapping images.

The demodulation of the color burst is performed internal to the module and the color gain (COL) and tint (TIN) can be adjusted either on the back of the module or external to the module by changing the DC input voltage. Contrast and brightness may also be adjusted on the module or externally. The horizontal position control on the back of the module can be used to center the image, but all other adjustments on the back of the module are performed at the factory and should be left alone for optimum operation of the module.

CONCLUSION

Sharp will also be introducing a series of new small TFT color LCD modules. The LQ4RB03 will have a similar interface to the LQ4RE01, but will include the backlight module and come in a slightly more compact package. The LQ4RB11 and LQ6RB11 are 4" and 5.7" LCDs that will also be similar to the LQ4RE01 interface, but will have an optional backlight module and a lower resolution. These units will be targeted toward more cost sensitive applications.

Sharp is also working on increasing the temperature range of the TFT color modules. Sharp's initial goal is to introduce units with a storage temperature of -30° C to +80° C and an operating temperature of -10° C to +70° C. By 1993 Sharp plans to have units available with the capability of a storage temperature of -30° C to +90° C and an operating temperature of -30° C to +85° C. Sharp's efforts to fulfill these goals will concentrate on all aspects of the LCD module: driver/interface ICs, interconnect technology, backlighting technology, polarizer technology, and the LC glass structure itself. Sharp is very confident it can meet the demand for modules which are subject to severe environmental conditions.

With a variety of models available, Sharp's small TFT color LCD modules can meet the design requirements needed for compact monitor applications. The variety of available models also offers a range of cost/performance options. The realm of applications is only limited by the imagination of the designer.

REFERENCES

Raster Graphics Handbook. Second Edition. Conrac Division, Conrac Corporation. New York: Van Nostrand Reinhold Company. 1985.

TV/Video Sync: Primer and Product Note Opt. 005-1. Hewlett/Packard. November, 1981.

GUIDELINES FOR CONTROLLING ELECTROSTATIC DISCHARGE IN A FIELD SERVICE ENVIRONMENT

Harry Miller, Field Applications Engineer

OBJECTIVE

CMOS and MOS devices used in Sharp LCD's are very susceptible to electrostatic discharge failure. The objective of this paper is to provide electrostatic control guidelines for Sharp LCD field service engineers and salesmen. The subjects will include a brief introduction, sources of electrostatic discharge, setup of static controlled work areas, and handling procedures. The guidelines may be used for setting up work stations in an office, manufacturing facility, booths at trade shows or any place that will require ESD protection. The paper will not cover packing, customer facility evaluation or notification to the factory of static problems by customers.

INTRODUCTION

Electrostatic discharge control is often talked about but not put into practice by field engineers and salesmen. Part of the problem is that the effects of electrostatic discharge are not always immediately apparent. Among the catastrophic failures caused by ESD are EOS, electrical overstress failures. EOS failures may start out with an ESD from a charged person resulting in a dielectric layer or small p-n junction melt within a MOS or CMOS device. The failure is usually categorized as a shorted or open device not as a defect caused by ESD. Appendix B illustrates an EOS failure on a Sharp color TFT display. CNDs (cannot duplicate failures) are failures often caused by transient pulses due to static. Stephen Halperin, in a paper on Static Control Management, has stated that 60 percent of service calls resulting in CND reports have been attributed to static in the user's environment.

SOURCES OF ELECTROSTATIC DISCHARGE

When setting up a static controlled area environmental factors must be evaluated (Appendix A). The environmental factors include floors, walls, ceilings, fixtures, work surfaces and major equipment.

Floors

The floor is the single greatest contributor to static generation on people, rolling equipment and sliding material because of their movement across it's surface. There are two ways to correct flooring problems. The first is by using antistat material on existing floors. The second method is by installing high performance tiles or carpets. ESD protective floors should have the ability to conduct charges to the floor. Walls, ceilings, light fixtures and vent grills are of major consideration where dust control and air flow are important. Dust control and airflow are related because air movement is a static generator which causes dust to be attracted to surfaces.

Work Surfaces

Work surface materials which dissipate a maximum charge to ground in approximately 0.2 seconds under ambient conditions are preferred. The rate of charge dissipation will allow the charge to disperse without arcing. Each work surface should have its own ground path. Serially connecting work stations will cause resistance buildup and inhibit proper discharge.

Major Equipment

Conveyors, power tools, and hand tools can all contribute to static discharge. If the work area is properly grounded, but the tools used for debug or repair are not, static discharge between the tool and device will result. Power tools and test equipment should be third-wire grounded. Leads from test equipment can carry substantial static charges. These leads should be grounded before touching them to ESD sensitive devices. Hand tools should be treated with antistat or be made of ESD protective items.

ESD WORD STATION

A basic ESD work station is illustrated in Figure 1. It consists of an ESD protective top, wrist strap containing a current limiting resistor and a conductive floor or floor mat.

A static dissipating or antistatic material is preferred for use as a work station for the following reasons:

1. It provides series resistance in the discharge path if an ungrounded person lays down an ESD sensitive item.
2. It reduces the chance of sparking from charged persons touching and discharging themselves to the table top.

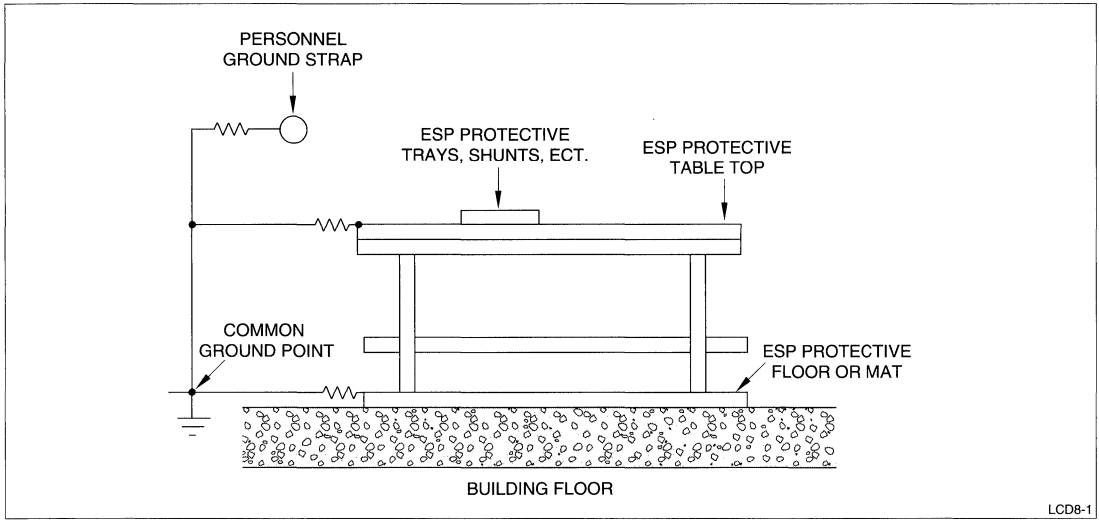


Fig. 1. Basic ESD Work Station

3. It prevents a short circuit between electrical connections if an assembly is energized while on the table top.
4. It reduces the risk of hard grounding the table top if grounded test equipment is placed on it.

OPTIONAL EQUIPMENT FOR ESD WORK STATIONS

Conductive Floors, Shoes, Heel Straps

Used in high traffic areas. Heel straps may be used in place of a wrist strap for convenience of having both hands free. Two heel straps should be used one on each foot. Every time a person picks up their foot the potential of not being properly grounded is present if two heel straps are not used.

Ionizers

For dissipation of charges from insulators or fluid flow process. In manufacturing environments ionizers can be used when static causes foreign matter to become attached to equipment.

ESD Protective Clothing

For use in clean room or other areas where clothing can be a prime source of static.

Controlled Relative Humidity

For dry climates where relative humidity is typically 25% or less.

HANDLING PROCEDURES

1. When ESD work station is available:
   - Open packages containing ESD material at ESD workstations.
   - Ensure the ground strap is properly connected to the workstation or to the equipment chassis.
   - Ensure the wrist or ankle strap is touching skin.
   - Do not touch ESD sensitive parts when handling ESD sensitive material (i.e. When handling LCDs, hold display by frame and do not handle CMOS devices)
   - If ionizers are used turn them on for a least ten minutes.
   - Keep ESD sensitive material in their protective packaging until needed.
   - Prior to transferring ESD sensitive material from one container to another connect the two carriers together and to ground.
   - After placing a carrier containing ESD sensitive material on a work bench allow two to three minutes for discharge.
2. When ESD work stations are not available and repair is required:
   - Store ESD sensitive material in ESD protective packing until ready for use.
   - Shut off power to equipment prior to removal or replacement.
   - Touch the grounded equipment chassis with your hand to dissipate any charges.

- If testing or probing is necessary, ground the test leads prior to touching ESD sensitive material.
- Remove failed ESD sensitive material and place in ESD protective packaging.
- Touch the package of the replacement device to the equipment chassis for a minimum of 10 seconds prior to opening.
- Remove replacement device and install it. Handle the replacement device by the frame only.

3. Displaying demonstration units at customer sites:
   - Field service kits are available to set up an ESD work station.
   - The portable work station may be grounded using the screw in a wall fixture that is connected to third wire ground.
   - When a customer wants to handle the product, they should be connected through a wrist strap to the ESD work station.

PRIME SOURCES OF STATIC ELECTRICITY IN AN ELECTRONICS WORK AREA

Source: Contributors to High Static Generation

1. **People:** Walking, rocking in a chair, getting up from a chair, putting on or removing work smocks, any brisk or repetitive motion.
2. **Clothing:** Clean room smocks, synthetic shirts, sweaters, pants etc., Shoes with synthetic (such a neoprene, foam rubber) soles, and heels.
3. **Chairs, work stools:** Vinyl covered, fiberglass, finished (such as, varnished, shellacked, polyurethane coated), waxed.
4. **Workbench tops:** Plastic, vinyl, covered, finished, waxed.
5. **Floors:** Vinyl, finished wood, sealed concrete, waxed floors.
6. **Parts bins, trays, tote boxes:** Plastic, finished wood or metal.
7. **Packaging material:** Plastic foam, bubble pack, cushioning, polyethylene bags, vials, containers, Styrofoam pellets (shells, peanuts).
8. **Tools:** Plastic solder suckers, hand tools (coated handles), soldering irons (with ungrounded tips), heat guns (used for heat shrink tubing, drying conformal coating, etc:), solvent brushes (synthetic bristles).
9. **Manufacturing Equipment and Processes:** Drying ovens, degreasers, sandblasters, temperature chambers, solder flow machines, integrated circuit handlers or loaders, spraying conformal coating, spray painting and cleaning, fault isolation, cryogenic sprays.

HISTORY

Missing colors were observed on the Sharp TFT color display. IC13 (LZ95F26), an ASIC device, was removed and submitted for analysis.

Purpose

To determine the failure mechanism and investigate its cause.

Methods

Curve tracer analysis, decapsulated using red fuming nitric acid, and optical microscopy.

Results

The curve tracer analysis revealed that pins 9, 11, and 12 were open. The device was decapsulated using red fuming nitric acid. Optical microscopy at the passivation level revealed severely damaged metallization near the bond pads of the failing pins.

CONCLUSIONS

The device exhibited opens due to severely damaged metal leading from the bond pads of the failing pins. The nature of the observed damage indicated that it was caused by EOS (Electrical Overstress).

REFERENCES

[1] McAteer, O.J., *An overview of the ESD problem*, Inst. Phys. Ser. No 85: Section 2

[2] Halperin, Stephen, *Guidelines for Static Control Management*

[3] Barr, T.N., Electrostatic Discharge Control

[4] Frank, Donald, E., *ESD Considerations for Electronic Manufacturing*, Westec Conference, March 21-24, 1983

ELIMINATION OF CROSSTALK IN TFT DISPLAYS

Harry Miller, Field Applications Engineer

OBJECTIVE

This paper examines the cause of crosstalk and describes a method of measuring crosstalk.

INTRODUCTION

It is common knowledge that TFT displays reduce crosstalk compared to conventional passive LCD displays. What is not always understood are the causes of crosstalk and how it is eliminated. The importance of discussing crosstalk is that Sharp has done proprietary work to nearly eliminate crosstalk from active TFT displays providing a decisive edge over the competition.

Crosstalk

There are two major causes of crosstalk. One is the resistance between gate terminals of adjacent pixels and the other is the resistance of the LCD cell itself. Figure 1 is a representation of the resistance in the LCD cell.

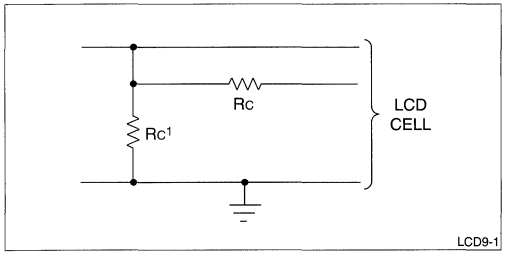


Fig. 1. Resistance in LCD Cell

$R_C{}^1$ = the resistance across the LCD cell

R_C = the resistance through the ITO layer of a LCD cell

Figure 2 (see page 2) is an equivalent circuit with R_S and R_G representing the resistance between the source and gate terminals of adjacent pixels.

Figure 3 (see page 2) is a simplified equivalent circuit where R_S and R_G are the total bus resistance and R_{LC} is the sum of R_C and RC^1.

$R_S = R_S \times V$

$R_G = R_G \times H$

$R_{LC} = R_C + RC^1$

Measurement Method

A method of measuring crosstalk is illustrated in Figure 4. A brightness measuring device is used to measure the brightness Ya with all the pixels on (white). A pattern (usually a box) is then displayed on the screen and the same spot is measure Yb. The cross modulation ratio can then be measured using the following formula:

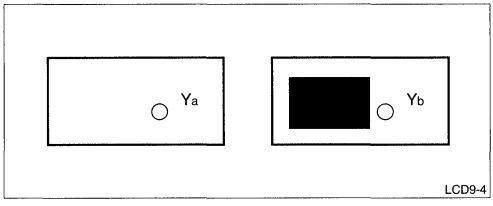


Fig. 4. Method for Measuring Crosstalk

$$\text{Cross modulation ratio} = \frac{|Yb - Ya|}{Ya} \times 100$$

Ya = brightness of area without pattern (cd/m^2)

Yb = brightness of area with pattern (cd/m^2)

Ya and Yb must be measured at the same spot.

SUMMARY

Sharp has done a great deal of work to eliminate crosstalk in TFT displays. We are using a proprietary feedback circuit to reduce crosstalk to the point it is nearly eliminated.

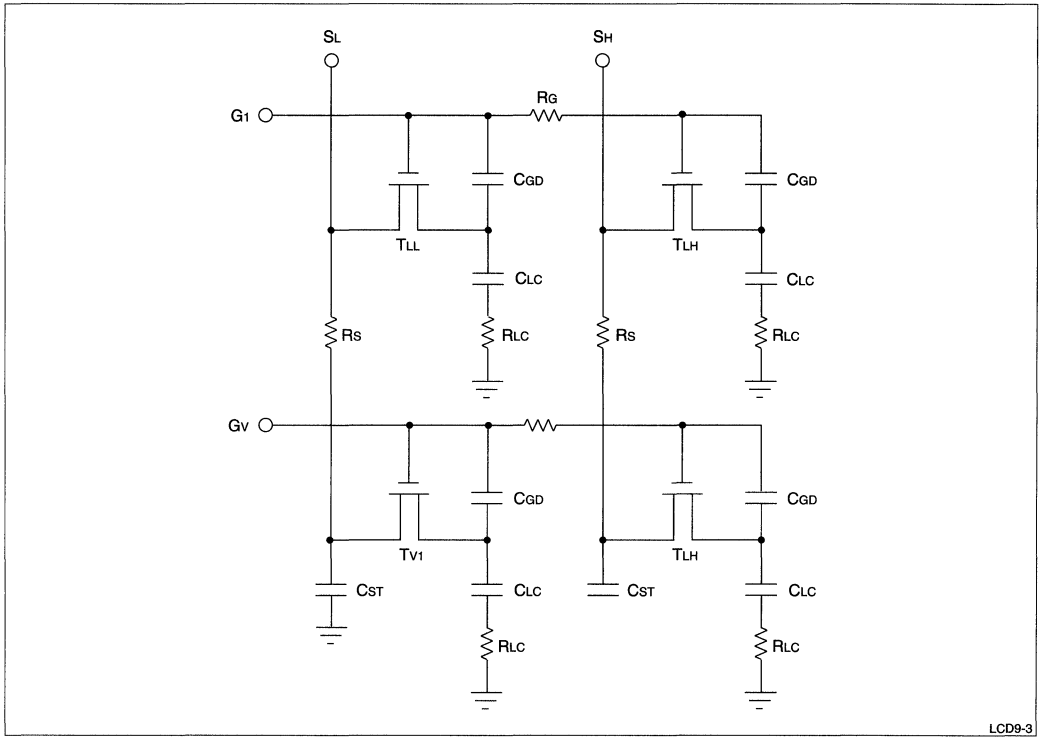


Fig. 2. Equivalent Circuit

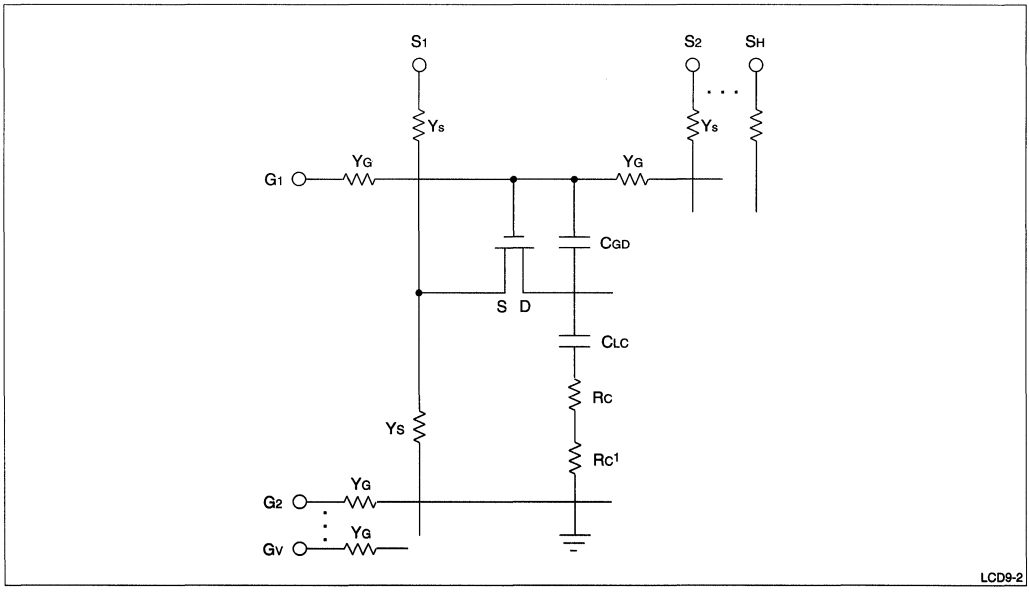


Fig. 3. Simplified Equivalent Circuit

Designing Small Modular Power Supplies for the Operation of Passive Liquid Crystal Displays

Charles Guthrie, Field Applications Engineer

ABSTRACT

Due to the nature of Passive Liquid Crystal Displays, several different voltages are required for their correct operation. This application note discusses the power requirements for the various display products, and provides suggested circuitry to generate the required voltages from a single 5 V or 12 V power source.

INTRODUCTION

Liquid crystal displays have complex power requirements. Because the Liquid Crystal material responds to a true RMS voltage, and the multiplex ratio in full-page displays is high, a large AC voltage is required to properly drive the fluid. These conditions create a requirement for a pure AC waveform with a maximum DC offset of only 25 mV. In order to accomplish this 'net zero' DC voltage, both positive and negative power supplies are required. The positive and negative voltages are asymmetrical around ground to compensate for voltage levels to drive the system logic. System power supplies have traditionally been +5 VDC. The driver logic is also set at +5 V in order to make the logic in the display drivers compatible.

At first glance, one would assume that the net panel voltage would be DC in that the voltages are not symmetrical around ground. Rules concerning the differential between VDD and the input signal levels require that the maximum power supply voltage be maintained at VDD ±0.2 V. Since the liquid crystal requires an amplitude of at least 28 V, the drive voltage must be asymmetrical around ground. Because of this, the display controller must adjust the duty cycle of the positive and negative going pulses to provide a net 'zero offset' DC voltage.

Sharp LCDs have an internal clock, the 'M' clock, which is used to develop a pure AC waveform with a net 'Zero' volt DC offset. In most applications, the 'M' clock is operated at a constant frequency, thus providing the offset by modulating the pulse width of the wave form. In some applications it is desirable to control the frequency of the 'M' clock to reduce flicker. Because the frequency of the 'M' clock must be carefully controlled, and the amplitude of the DC voltages must be stable, care must be taken to ensure that the power supplies are stable and devoid of any ripple.

The following is a compilation of circuits which may be adapted into computer designs to provide the various voltages required to operate the display. The circuits presented are adapted from literature provided by the various manufacturers. Some of the circuits have been breadboarded to verify their suitability in the applications; however, if you are considering using any of these designs in your application, it is strongly recommend that you contact the respective manufacturer and work with their application engineers to fine tune the circuits to your individual needs. This report is intended to explain the requirements for, as well as the principals of operation, for the various circuits.

PASSIVE DISPLAYS

Passive Displays by nature are very complex in their power requirements. The logic driving the display operates off of the system +5 V power supply (VDD). An additional negative power supply is required to provide sufficient voltage to reliably twist the liquid crystal. In addition, in backlit units, the backlight requires an elevated voltage to fire the backlight tube. If the backlight is an EL lamp, the voltage required is between 150 and 250 V. In the case of Cold Cathode Fluorescent Tubes, an initial firing voltage of between 800 to 1,500 V is required to start the discharge, and 280 to 400 V is required to sustain the lamp. Hot Cathode and incandescent lamps are not used in Passive displays and will not be discussed in this section.

MONOCHROME DISPLAYS

Monochrome full page displays typically require a total of about 28 VAC for efficient operation. The following circuits are designed to take either a 12 V or 5 V input and generate the necessary negative offset voltage required to properly twist the liquid crystal. Because the output must be stable yet variable, the following circuits include the necessary elements to provide variable negative voltages. The circuits described are designed to take the +5 V or +12 V input from the main power supply and convert them to the required levels. These circuits are also designed to provide stable, efficient, and economical solutions to the unique requirements for proper display operation.

MAXIM NEGATIVE VOLTAGE LCD BIAS POWER SUPPLY

The Maxim 759 is a CMOS inverting switch mode regulator with an internal power MOSFET. The circuit is designed to deliver up to 30 ma at –24 V with 80% efficiency. The circuit comes in a 16-pin DIP package or a 16-pin SMT package. With the SMT package, the whole circuit can be built in an area of approximately 1.5 inches square using surface mount components. The circuit has an on/off pin to hold the supply off until the proper power on sequencing has been accomplished. This pin can be controlled with the output enable incorporated in most state-of-the-art controllers. The 100 K potentiometer is used to adjust the contrast on the display.[1] Refer to Figure 1.

The Max759 incorporates a high-performance current mode pulse width modulation control scheme coupled with a simple buck-boost switching circuit to provide tight output voltage regulation and low noise. The fixed frequency oscillator is factory trimmed to 165 kHz, allowing for easy noise filtering. The devices are production tested in an actual application circuit, and output accuracy is guaranteed at ±5% over all specified conditions of line, load, and temperature.[2]

Features of the Maxim 759 include:

- Conversion of positive to negative voltages
- 1.25 W guaranteed output
- 83% typical efficiency
- 1.7 mA quiescent current
- 0.7 µA typical shutdown mode
- 165 kHz fixed frequency oscillator
- current mode PWM
- under voltage lockout and soft start
- simple application circuitry

LINEAR TECHNOLOGY NEGATIVE VOLTAGE BIAS POWER SUPPLY

Figure 2 shows the circuit diagram for the Linear Technology variable voltage negative power supply. The circuit is half switcher and half charge pump technologies. The charge pump is driven by the flying mode at Vsw. This circuit uses the LT1172 inverting voltage regulator. This supply generates an output between –10 and –30 VDC. The output current is up to 70 mA at better that 80% efficiency.[3] In low voltage operation where the input voltage is less than 6 V, Vin and Vbatt can be tied together. Refer to Linear Technology application note #49 for further information about this power supply. The circuit has an enable to allow the power supply to be turned on through proper power up sequencing. The unit operates with an input voltage of between 3 and 20 V. R1 is used to adjust the contrast on the display.

The entire circuit can be built in less than 1.5 square inches of board space with surface mount components. The following circuit is available as a demo kit from Linear Technology.[4] The kit is evaluation kit #004. The evaluation module provides –10 to –30 V output at currents from 1 mA to 70 mA. Efficiency of the circuit is greater than 80%. Features of the circuit are similar to those for the Maxim circuit diagrammed above; however, the output current is specified to be a little higher.

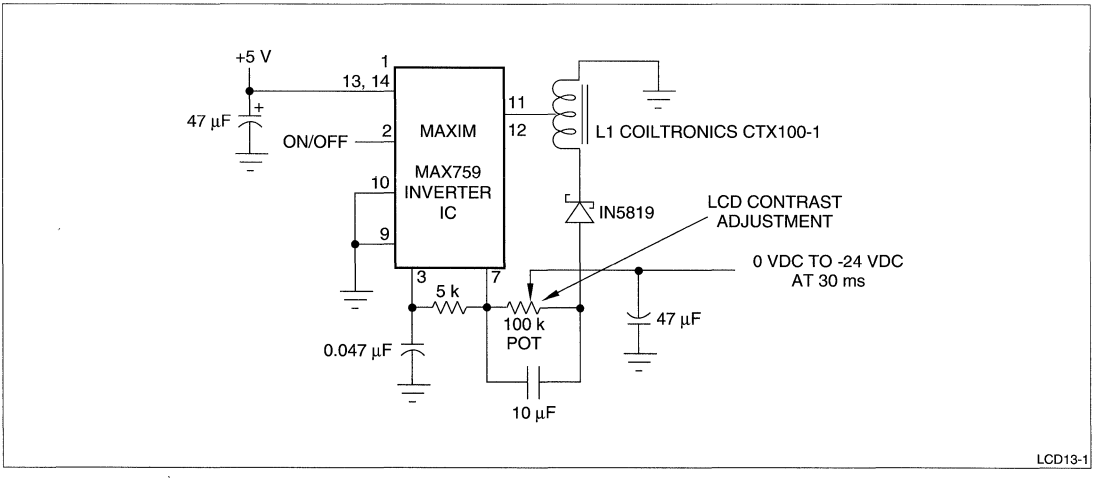


Figure 1. Maxim –24 V Variable Power Supply

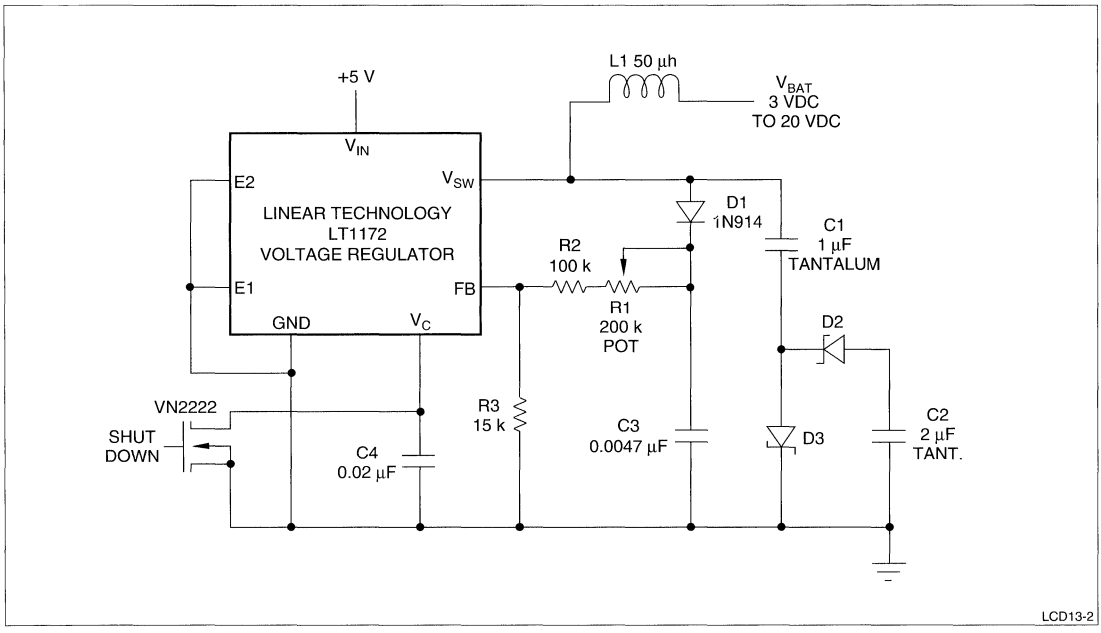


Figure 2. Linear Technology -30 V Variable Power Supply

CCFT INVERTER POWER SUPPLY

Linear Technology makes an integrated circuit suitable for developing a custom inverter for use in driving the backlight used in most passive monochrome Liquid Crystal Displays. According to the Linear Technology Demo Manual DC011, the inverter circuit will work with an input voltage of 4.75 to 25 VDC and will provide CCFT output voltages at up to 7.5 mA. Regulation is to within 1% and efficiency is better than 78% over the input voltage range. Parts for the unit, in quantity, costs less than $5.00.

In designing a CCFT power supply, one must consider the lamp characteristics. The lamps are a difficult load to drive, particularly with a switching regulator.[5] The lamps have a 'negative resistance' characteristic; the starting voltage is significantly higher than the operating voltage. Typically, the starting voltage is between 1,000 and 1,500 V. The sustaining voltage is between 250 and 400 V depending on the internal gas mixture. The bulbs require a pure AC waveform with little or no DC component. Any DC component in the power supply will eventually damage the tube and shorten its life.

The 'negative resistance' of the bulb and induced snap back, combined with the frequency compensation problems associated with switching regulators, can cause severe loop instabilities particularly on startup. Once the lamp is in its operating region, it assumes a linear load characteristic, easing stability criteria.[6]

Bulb operating frequencies are typically between 20 kHz and 100 kHz with a sine waveform. The sine wave low-harmonic conditions minimizes R.F. emissions.[7]

Care must be taken in laying out the inverter circuit. The coupling capacitor must be carefully located to minimize leakage paths on the circuit board. Please refer to the design guidelines in the demo manual from Linear Technology before doing layout of the following circuit. Refer to Figure 3.

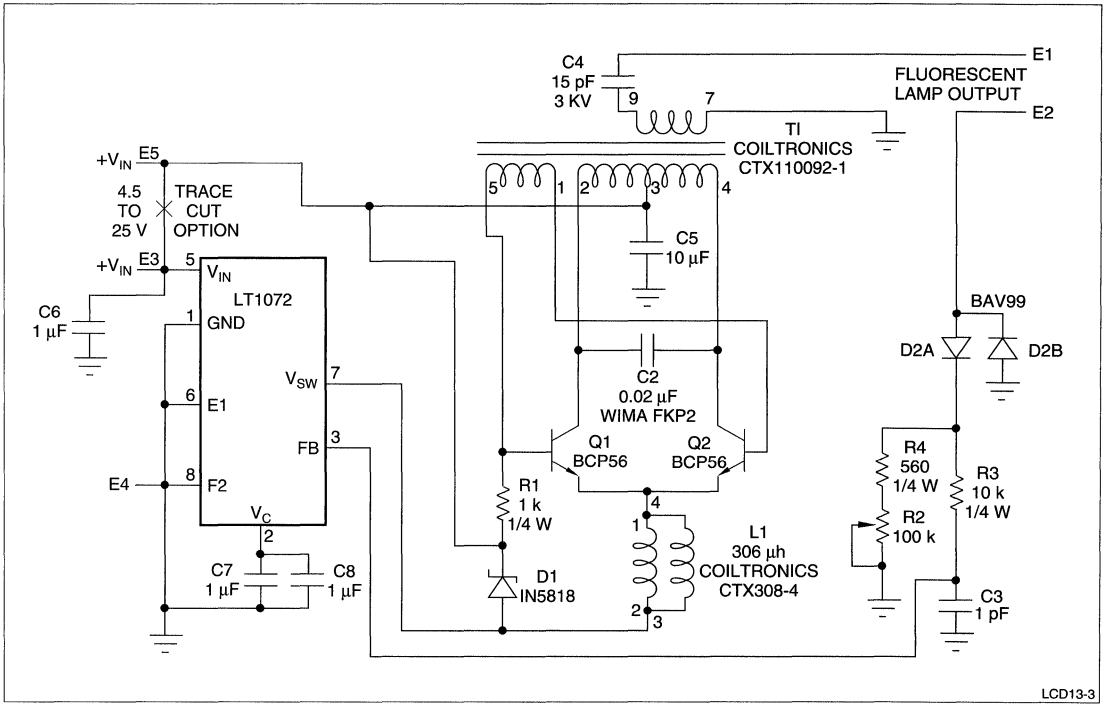


Figure 3. CCFT Inverter Power Supply

COLOR DISPLAYS

Passive color displays require a positive VEE power supply. The supply has to be capable of producing +20 to +50 VDC with at least 50 mA current. The design for these power supplies must take into account that the current fluctuates depending on the number of pixels which are activated. As the image shifts, the demand for power will fluctuate. This can cause flickering or modulation if the power supply is not stable and able to adjust to the fluctuations created by the display.

One of the most important components to be considered in all of the designs illustrated below is the inductor. The core of the inductor must be selected carefully in order to provide proper stability in the power supply circuit. Cores which are too small will modulate the brightness and make the display look as if it is fading on and off. In both designs, it is recommended that you start with a 100 µH choke that is rated at a peak current of at least 3 amps. This is required by the high switching currents needed to boost the voltage from 5 V to as much as 50 V.

The power supplies shown in Figures 4 and 5 have been adjusted to provide the required stability. Both supplies use a switching regulator and a charge pump design to provide the boost voltage. The advantage of this circuit is that it takes up a relatively small space on the mother board and the design is extremely efficient.

CONCLUSIONS

One of the things that becomes evident as you examine the previous data is that small efficient power supplies can be designed to meet the needs of the Liquid Crystal Displays used in most modern notebook computers. The advantage these supplies offer the designer manifests itself in the fact that the main power supply may be made less complex. Typically, when the required voltages are designed into the main power supply, the net design is much larger than required by the display and its related components. These supplies offer an inexpensive alternative. If you are interested in more information on the power supplies described in this paper, please contact:

Maxim Integrated Products
120 San Gabriel Drive
Sunnyvale, CA 94086
Phone: (408) 737-7600
Fax: (408) 737-7194

Linear Technology Corporation
1630 Mc Carthy Blvd.
Milpitas CA 95035-7487
Phone: (408) 432-1900
Fax: (408) 434-0507
Telex: (408) 499-3977

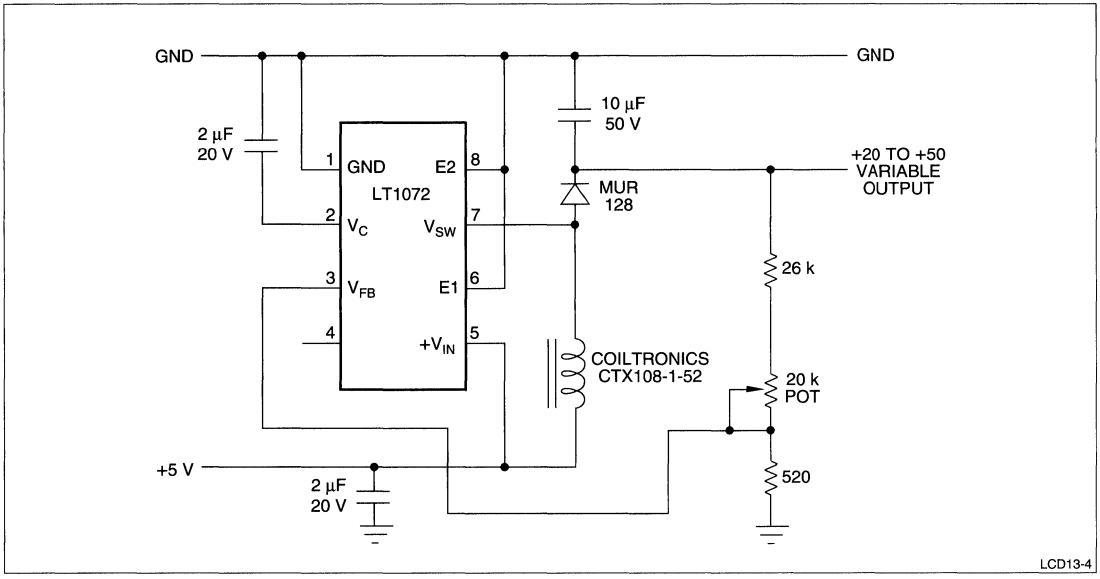


Figure 4. Linear Technology Positive Voltage Variable Power Supply

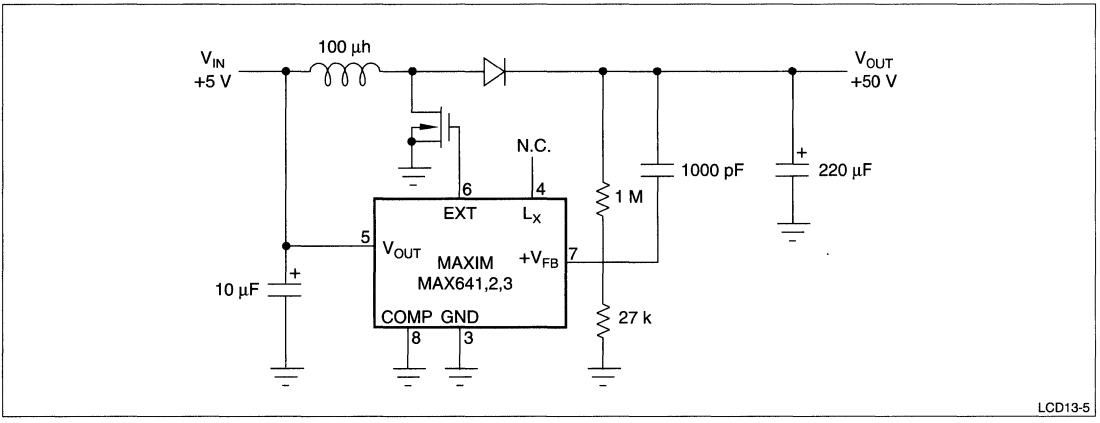


Figure 5. Maxim Positive Voltage Variable Power Supply

REFERENCES

[1] *Power Supplies Analog Design Guide*, August 1991

[2] *Product Specification*, Max759, Maxim, 1991

[3] *Illumination Circuitry for Liquid Crystal Displays*, Jim Williams, August 1992

[4] *Demo Manual DC011*, Lienary Technology Corporation, August 1992

[5] *Illumination Circuitry for Liquid Crystal Displays*, Jim Williams, August, 1992

[6] *Illumination Circuitry for Liquid Crystal Displays*, Jim Williams, August, 1992

[7] *Illumination Circuitry for Liquid Crystal Displays*, Jim Williams, August 1992

Touch Screens for Flat Panel Applications

David Blass, Senior Application Engineer

INTRODUCTION

Input devices for computers systems traditionally have been keyboards or keypads. These devices are simple and very low cost. As systems become more complex with advance hardware, keyboards have grown in size, adding many single-stroke key functions to aid in high-speed use of the computer. Keyboards have grown to include well over 100 key switches, which makes them very large devices. Reducing the overall size and complexity of the keyboard while increasing its functionality is gaining popularity.

Non-keyboard systems fall into two main categories: those that are independent from the video output (i.e., tabletop digitizing tablets or a mouse), and those that work in conjunction with the video output (i.e., over screen input systems). Over screen systems have two main types of input devices: pen input and touch screens. Pen input systems use the pen as the input device. These systems are used when hand-writing recognition or capture is important or if there are very small targets on the screen. The touch screen devices are activated by the human finger. This is important for quick operation and response where a stylus (pen) would be awkward. Finger operated touch switches, however, are not well-suited for small targets or when hand-writing is required.

In both cases, the actual switching mechanism sits directly on the video screen. In this way, the actual key switch can have its legend, appearance or location appear directly on the scene. Further, when the computer operation no longer requires the key switch function, the system simply erases the switch in order to display other key switches. Using this feature, a computer system can have a limitless number of key switches and functions.

THEORY OF OPERATION

Pen input and touch screens are similar in overall systems design and interaction. Each device places an X,Y grid across the face of the device. Touching the device with either a pen or the fingertip, sends the corresponding coordinate pair to the computer. The computer is interrupted while it searches its program to see what function is attached to that particular coordinate pair. An X/Y coordinate can refer to a key-code or an entire command string or complex macro. In either case, the touch switch coordinates and their programmed functions are interrelated to the video screen, but are separate in function. The video screen is not directly tied to the touch switch. Software programming of the computer is the link between the video screen and its touch buttons and the touch targets on the touch screen placed over it. The communication link to the computer can be a serial or a parallel ASCII link via a communication port or to the keyboard port. The touch device can also communicate directly to the computer at the bus level using binary or Hex. Communications at the bus level is much faster and more direct than the other approach. In a PC environment, the bus-level approach is the most preferred method because of the performance enhancement and the hardware and software are readily available. Non-IBM PC systems will probably find ASCII communication far easier to deal with since no new hardware will need to be developed.

All pen input digitizers operate in a high-resolution mode with a large number of targets per inch. Touch switch overlays come two ways. They can be full matrix switches, similar to the digitizer, with very high resolution or with a fixed matrix, offering lower resolution. The full matrix types use an analog input of either a resistance or capacitance value. These values are then translated into a digital coordinate. The fixed matrix type are wide X/Y traces, whose intersecting point are quite large, typically one half inch square. The resolution is quite coarse and the number of touch targets is very low. The only benefit to fixed-matrix-touch switches is the low price of its controller. A low-cost keyboard decoder is all that is required rather than the more complicated and costly A/D converter based converter board. Resistive and capacitive technologies are offered in both matrix arrays while acoustical wave is offered only as full matrix. Infrared is listed as a full-matrix type but in reality is truly a high-resolution fix-matrix type of touch switch.

TOUCH-INPUT SYSTEMS

Considerations For Software

Pen-input systems need to recognize the limitations of handwriting recognition software. Needless to say handwriting capture is fail-proof. Handwriting is simply captured but not recognized by the computer. Handwriting

recognition is another story. Today, handwriting recognition drivers are still prone to enough errors to greatly limit the operation of many applications. It is best to limit application software to predefined character target areas. In addition, the software should force the user to print in capital letters in these targets. This should greatly the efficiency and accuracy of the application software.

With touch applications, always remember, **keep it simple**!!! Displaying too many key switches or making the display too complicated can completely defeat the primary purpose of a touch screen. The primary function of touch switches is to eliminate and replace the keyboard. For this reason, it is advised that the touch switch replication of the keyboard should do more than match a keyboard key by key. That is to say, to simply display key caps that represent, for instance, an entire qwerty keyboard is far less functional than higher-level function switches. There are two compelling reasons for this. First, the a touch switch version of a keyboard, qwerty, 10 key or otherwise, is far less convenient to use than a keyboard. Considering the angle of the screen, the size and layout of the screen, and the lack of tackle feedback will greatly affect the user's ability to quickly and efficiently use the 'touch screen' keyboard. Most typists prefer a 10 to 13 degree angle of the keyboard to maintain a proper typing angle. In normal operation mode, it is rare to find a video display at this angle. Most full-size keyboards are 15 inches wide. Even keyboards in notebook computers are 11 inches wide. The most common widths of flat displays are less than 10 inches wide. This fact makes it nearly impossible to accurately reproduce a full size keyboard on a flat display. Lastly, and perhaps most important, typing on a flat screen offers the operator zero tactile feedback. Without feeling the switch action of the keys, typing is very difficult.

More important than the mere keyboard replacement, consider a higher role of the touch screen. With advanced computer power, coupled with the increased use of the computer by novice users, the visual aid of the video output computers now plays an active role in aiding the user. Touch screens facilitate the computer to 'ask' the user particular questions required for an operation. Rather than requiring the user to think about the operation and then laboriously typing in the answer character by character, the computer can now present specific questions in graphical form to the user who will merely 'touch' the answer. Once the user touches the correct button/answer, the screen can instantly reconfigure itself with a new set of questions or steps in an operation. With the use of these high-level 'buttons,' the computer will guide the user through the entire operation. Human factor studies have found that operations increased 30% by using touch switches over the manual keystroke operations.

When laying out a screen for use with a touch screen keep in mind the following guidelines:

- Keep available key switches to a minimum. Reading and reacting to key switches goes down proportionally to the number on the screen. Too many forces the operator to read and then think about the choices. This is very time consuming.
- Keep the key switches large. Not only should there not be many but if they are too small or too close together there is a large possibility for error or operating delay. Human factor studies have determined that touch targets should be a minimum of 3/8 inch square and have a 1/8 inch gap between the switches. This 1/2 center key switch is the minimum. Again, they should be as large as possible.
- Use graphics to enhance the function of the switch. Make the switch appear to be 3-dimensional and reverse video or XOR the video when the switch is activated. Using audio feedback can further enhance this feature.

Finger-driven touch switches can, in limited use, be used to replicate and replace a mouse. The pen input type touch device, as mentioned, is best suited for this application. It can travel quickly, has the ability to respond with mouse 'clicks' and has the fine resolution necessary for mouse replacement. In general, finger-driven touch switches cannot adequately replace the mouse function. The touch resolution and the size of the finger does a poor job in this respect. Finger driven touch switches have mouse drivers developed for PC applications but is advised to consider them only in custom applications where the software understands that the 'mouse' is a human finger.

Considerations For Hardware

Choosing the best touch device for an application depends largely on the cost performance requirements. As with most electronic hardware, there is a variance between the different technologies and their features. Typically, low-cost touch technologies have limited performance or environmental characteristics, whereas the more expensive types have superior performance. The following is a listing of the various technologies, outlining their particular strengths and weaknesses.

Resistive

Resistive touch screens consist of two transparent layers: the bottom layer can be made of glass or rigid plastic, while the top layer is flexible Mylar. Each of the layers are separated by insulating spacers. The touch controller puts a voltage gradient across the conductive layers oriented in the X and Y directions. Pressure from a finger or stylus causes the outer layer to come in contact with the inner layer.

When contact is made, the conductive layers electrically short to each other and the resistance value is sent to the controller. The controller then puts the R values through an analog-to-digital converter where they result in a digital coordinate.

Features:

- Highest resolution – Resistive touch switches are capable of the highest resolution of any overlay touch technology. They are capable of reaching 4096 × 4096 touch points across the screen.
- Low cost – Resistive touch is the lowest cost touch technology. In the fixed matrix format, resistive touch including the decoder circuitry can cost as little as $50 or less for the system. A full matrix, analog resistive touch can cost less than $100 complete. The decoder circuitry to 'read' the analog values is neither complicated nor costly to develop.
- Adaptable for different sizes – A resistive touch switch can be made to accommodate various sizes for a reasonably low cost for the NRE. If initial quantities are sufficient, some manufacturers will waive the NRE charge completely. Resistive technology can be developed in different sizes with very short turn-around times.

Concerns:

- Calibration – The resistance value at a coordinate location can drift over time and temperature. This requires the screen to be calibrated at regular intervals to ensure the touch target on the display corresponds to the resistance values assigned to it.
- Delicate – The top resistive layer is Mylar. This layer can be scratched or torn if a sharp or metallic item is used for a touch stylus. It can also be burnished to a dull, hard-to-see-through surface with constant use.
- Heat can also damage the resistive touch switch. Touching the top layer when very warm, +40°C, can actually stretch the Mylar, which ruins the switch.
- Transparency – Resistive touch switches require a large amount of Indium Tin Oxide (ITO). This resistive layer can cut the transmissivity of the touch screen by up to 27%. This can greatly affect the viewability of the flat screens, especially with TSTN LCD and EL displays.

Infrared Red

Infrared switches use emitting and detecting LED arrays, in an X and Y arrangement over the face of the display. When touched, paired LED beams are broken on both axis. The broken beam coordinate is interrupted much as a simple keyboard. The LEDs are arranged on 1/4 inch physical centers, but with the aid of coordinate averaging, I.R. touch switches can interpret a virtual beam between two physical beams. This give I.R. switches an effective resolution of 1/8 inch centers.

Features:

- Simplicity – I.R. touch switches are very simple X/Y arrays. They can be decoded with a common keyboard decoder circuit.
- Solid state – I.R. switches use all solid state devices. There is no mechanical mechanism to wear out or break.
- Transparent – Since the switch mechanism is light, there is no overlay to block the light transmission of the flat panel. 100% of the light of the screen is transmitted, with no loss or distortion from the touch switch.
- Sealable – I.R. touch switches are completely enclosed in their bezel, easily allowing them to be sealed from the environment.

Concerns:

- Coarse resolution – I.R. switches can only be built on 1/4 inch centers. Even with software averaging, the closest resolution support by I.R switches is 1/8 inch centers. On standard 9.5 diagonal displays, this equates to 63 × 32 touch targets.
- Complex to build/develop – I.R. switches are printed circuit boards with the LED array of emitters and detectors. The layout and design of this board is reasonably complex and costly to develop. Since the LED array fits on the front of the display, the IR switch must be made for specific display models. It is common not to find IR switches in production for various displays.
- Parallax – Since the LED array sits over the face of the screen, the beam path can sit as much as 1/8 inch over the display face. This means that a user could touch or activate the switch without actually touching the screen.
- False triggers – Since it only requires breaking the light beam to activate the switch, anything in the path of the LEDs can trigger the device.

Capacitive

Capacitive overlay technology uses a touch sensor that is a glass overlay with a conductive coating bonded to its surface. A low current flows across the capacitive panel and establishes the frequencies of four oscillator circuits at the panel's corners. When the screen is touched by a conductive stylus, the impedance alters the frequency of the four oscillators. The touch coordinate is calculated from the differential frequency change of the four oscillators. This, in turn, determines the X and Y coordinate of the touch activation.

Features:

- Durability – Capacitive touch switches are unaffected by dirt, water or grease.
- Fast response – It responds faster to the touch than resistive or acoustical wave technologies but cannot respond 'before' the touch like Infrared touch.
- Reliability – The user touches a sheet of glass with only a coating on it. There is no mechanical switch mechanism to wear out or fail.

Concerns:

- Non-capacitive touches – Capacitive touch switches uses the conductivity of the stylus (human finger, metal stylus, etc.). Non-conductive items cannot be used with conductive touch switches. This is especially important in medical applications when rubber gloves are used.
- Circuit Sensitivity – The touch switch can be affected by the emissions of the display. Additionally, electrically noisy circuits can also disrupt the operation of the switch.

Acoustical Wave

Acoustical wave touch sensors utilizes the ability of inaudible, high frequency acoustical waves, traveling over the surface of the sensor at very precise speeds in very straight lines. Acoustical signals generated by an X or Y transmitting transducer travels to a reflective array, then across the surface of the screen, to a second reflective array and finally, to a receiving transducer.

Features:

- Z Axis – Acoustical wave switches have the ability to measure the Z axis on the switch. The harder the user presses on the switch the dampening of the sound waves can produce a Z axis reading.
- Transmissivity – Acoustical wave switches are essentially clear glass offering very high transmissivity for an overlay type switch.

Concerns:

- Damage – The transducer of the switch sits either on the front or edge of the glass. It is exposed and very prone to damage.
- Contamination – Surface contamination, water or heavy grease, can absorb some of the acoustical frequency and register a touch.

Pen Input

Pen-based digitizer touch switches use an embedded sensor board under the flat display. The embedded sensor generally radiates electromagnetic energy. The pen stylus picks up this frequency and returns it to the touch controller. Some pen-based stylus' are directly wired into the system while others are cordless.

Features:

- Embedded sensor in display – Digitizers have the unique ability to be built into a display.
- Handwriting recognition – Because of the high speed/resolution of the sensor and the pen stylus, a digitizer pen-based touch can be used for handwriting recognition applications.

Concerns:

- Pen – All pen-based systems need the pen. The tethered based pens are inconvenient and awkward. Non-tethered pens are costly and easily misplaced.
- Display technology – Due to emissions, metallic frames and radiation from the display, not all display technologies are suitable for pen-based applications. Currently, TSTN LCD technology is the only display technology utilizing pen-based touch systems.

VENDORS

The following is a listing of some manufacturers of the touch switch technologies mentioned in this report. The list covers all of the technologies; it is not necessarily a complete list. Although every step was taken to make the list complete and comprehensive, it is impossible to included all manufacturers.

Resistive

Elographics
105 Randolph Road
Oak Ridge, TN 37830
(615) 482-4100

Touch Technology
5524 Bee Caves Road
Austin, TX 78746
(512) 328-9284

W. R. Brady
Thin Film Prod. Div.
8225 West Parkland Court
Milwaukee, WI 53223
(414) 355-8300

Transparent Devices
717 Lakefield Road
Bldg. D
Westlake Village, CA 91361
(805) 497-8500

Infrared

Carroll Touch
P.O. Box 1309
Round Rock, TX 78680
(512) 244-3500

Dale Electronics
1122 23rd Street
Columbus, NE 68601
(402) 563-6505

Capacitive

Microtouch
10 State Street
Woburn, MA 01801
(617) 935-0080

Touch Technology
5524 Bee Caves Road
Austin, TX 78746
(512) 328-9284

Acoustical Wave

Elographics
105 Randolph Road
Oak Ridge, TN 37830
(615) 482-4100

EXZEC
1840 Oak Ave.
Evanston, IL 60201
(708) 866-1869

Pen-Based Digitizers

Wacom
501 S.E. Columbia Shores Blvd.
Suite 300
Vancouver, WA 98661
(206) 750-8882

Kurta
3007 East Chambers
Phoenix, AZ 85040
(602) 276-5533

Scriptel
4145 Arlingate Plaza
Columbus, OH 43228
(614) 276-8402

Logitech
6505 Kaiser Drive
Fremont, CA 94555
(510) 795-8500

CalComp
2099 Gateway Place, Suite 550
San Jose, CA 95110
(408) 441-4126

THE DEFINITION OF TERMS AND AN UNDERSTANDING OF THE TECHNIQUES USED FOR THE INCOMING INSPECTION STANDARDS OF LIQUID CRYSTAL DISPLAYS

Charles Guthrie, Field Applications Engineer

ABSTRACT

This Application Note deals with the parameters associated with the incoming inspection of liquid crystal displays. This document explains the terms used in describing defects and provides the criteria from which display lots are accepted or rejected. Even though this document is based on the Sharp Electronics incoming inspection standards, the terms used and inspection procedures may be applied to any manufacturer's product.

INTRODUCTION

Agreement upon the standards for the incoming inspection of Liquid Crystal Displays (LCDs) has been difficult for both display manufacturers and their customers, O.E.M. computer manufacturers. These problems have resulted from a misunderstanding of the capabilities of the manufacturers to produce cost-effective 'Perfect' products and their customer's need for totally defect-free, visually-perfect displays. In the past, problems have resulted from the impossible requirement for 100% defect-free displays. The following is a compilation of the standards which have been developed to assess the quality of displays and the acceptability of defects, classified as major and minor, and the quantification of these defects as to the acceptance or rejection of any manufacturing lot through the customer's incoming inspection.

Items to be discussed include: a recommended test station for the incoming inspection of displays, recommended test patterns and their objectives in testing the display quality, a glossary of terms used in the inspection of displays, and specific standards for the various types of displays manufactured.

SHARP INCOMING INSPECTION STANDARDS

The following is a list of the inspection standards used to create this document. Each item discussed may apply to one or more of the documents listed. Please refer to the applicable standard for the allowable defect levels.

S-U-006	Incoming Inspection Standards for Passive LCDs
S-U-008	
S-U-012-01	
S-U-014-01	
LQ6xxxx	Incoming Inspection Standards for TFT - LCDs
LQ4xxxx	
LQ9D011	
LQ10D011	

TEST STATION SETUP

The test station for the incoming inspection of liquid crystal displays shall consist of an anti-static workstation that is properly connected to earth ground. Grounding of the test station is accomplished by installing a copper rod driven into the earth until it is three feet below the low ebb of the water table (usually around 10 feet). In addition, fluorescent lighting shall be provided from a single 20 W fluorescent tube that is approximately 50 cm from the work surface (approximately 20 inches).

A computer test station is recommended, preferably using hardware similar to that used in the final assembly. Timing to the display should be identical to that used in the final hardware and fall within the parameters of the manufacturer's specification, in order to assess the display as it appears in final products. Timing changes can cause visual anomalies that may make the display appear to be defective when, in fact, it performs to specification when installed in the final system.

Another component that is recommended at the test station is a display stand with 'quick connect' hardware to facilitate testing of the display. The following items *should not* be part of the display test station: an illuminated magnifier, a microscope, or a high-intensity light source. These items are unacceptable since their only purpose would be to 'Magnify' defects and to amplify problems which will never be seen by customers.

Figure 1 depicts a typical test setup including the lab bench, test station, and lighting. Please note that the lighting should be above and directly over the center of the workstation. In addition, there should be sufficient space around the test station to allow moving the boxes of displays in and out. For full-page monochrome displays, the boxes are packaged with only 20 units per carton. In the case of a notebook computer manufacturer, this translates to an area the size of a pallet each for incoming and outgoing materials.

TEST PATTERNS

The following test patterns are recommended since they are simple to implement and shows all possible defects in the display.

Alternate Lines – Vertical

This test identifies any column drivers that are shorted. Shorted drivers appear as 'fat lines,' which either extend up from the bottom to the center of the display (bottom driver shorted), or down from the top to the center (top driver shorted) (Figure 2).

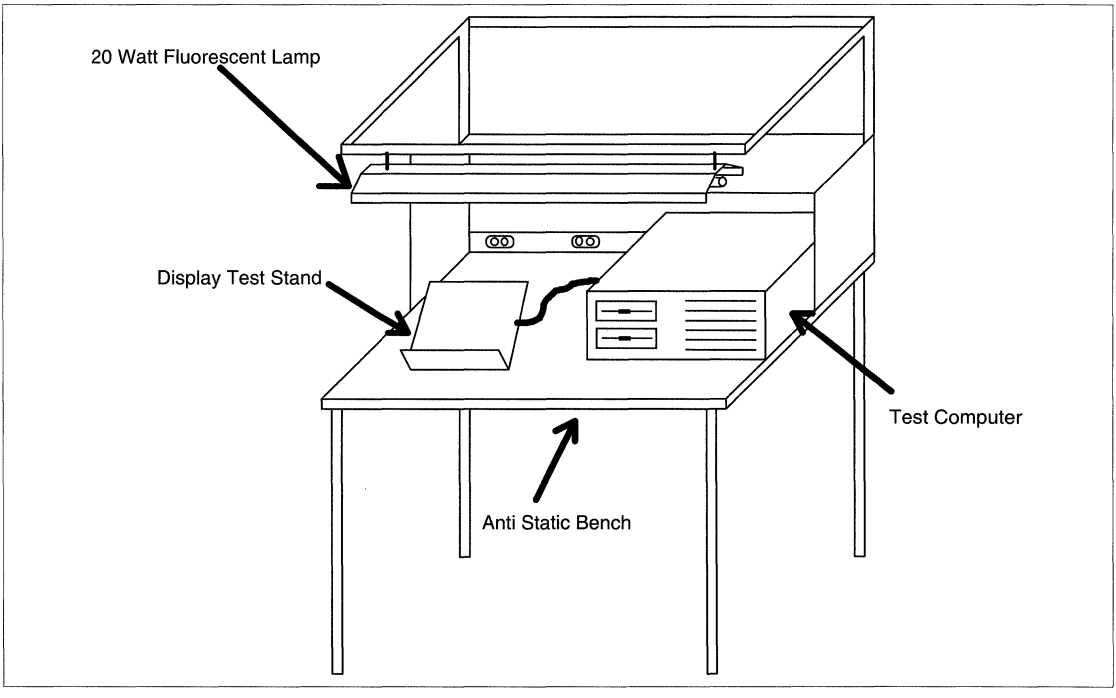


Figure 1. Test Station Setup

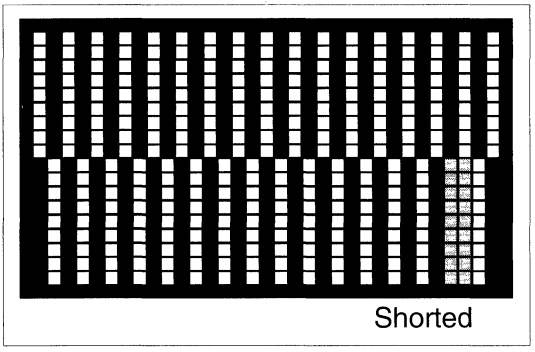


Figure 2. Alternate Lines – Vertical

Alternate Lines – Horizontal

This test identifies any row drivers that are shorted. Shorted drivers appear as 'fat lines,' which extend across the display (Figure 3).

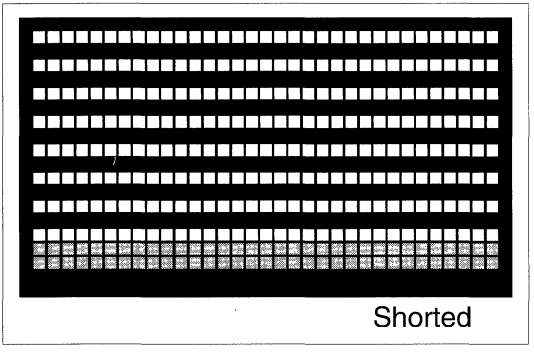


Figure 3. Alternate Lines – Horizontal

All Pixels Off

During this test, the technician assesses the display for white spots, white lines, and the uniformity of the retardation film (Figure 4).

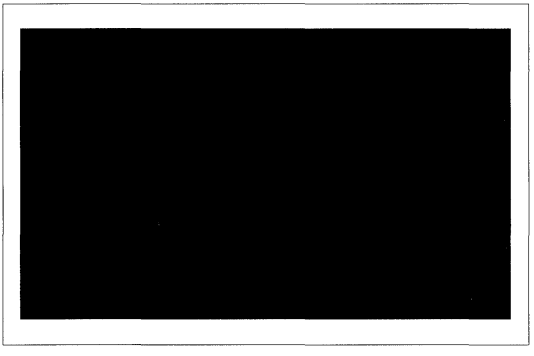


Figure 4. All Pixels Off

All Pixels On

During this test, the technician assesses the display for black spots, black lines, dot width uniformity, and the center gap line (Figure 5).

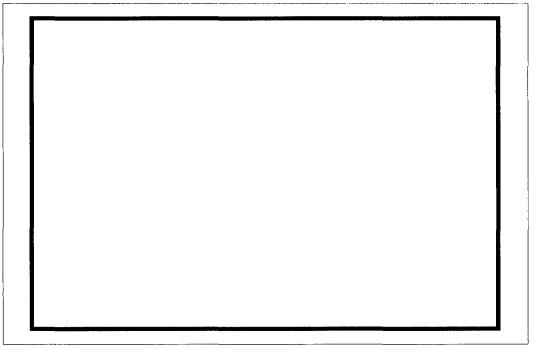


Figure 5. All Pixels On

Full Page Text

This test assesses whether black or white lines and spots interfere with proper character generation (Figure 6).

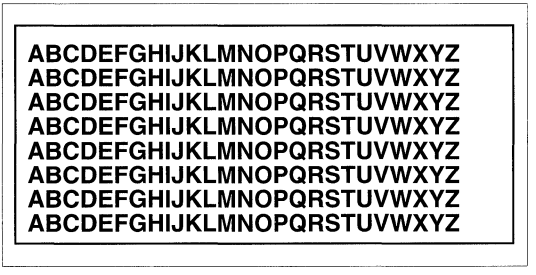


Figure 6. Full Page Test

Visual Inspection

Even though this test does not require a specific test pattern, the technician should inspect the display for the following before accepting or rejecting the display: defects in the bezel plating, scratches on the front polarizer, bubbles under the polarizer, stains on the display surface, and a bent or deformed bezel.

GLOSSARY OF TERMS

In order to understand the terms used for the inspection of Liquid Crystal Displays, the following glossary of terms has been compiled. A comprehensive explanation of the meaning of the term is included and, where applicable, examples to help the inspector better understand the meaning of the term.

The Buyer's Rights

The buyer has the right to conduct an incoming inspection of the units received at his own cost, at the destination specified in the relevant Bill of Lading, and on the basis of the unit specification. Terms of the inspection procedure are identified in the incoming inspection standard provided by the display manufacturer. Any required deviations to the incoming inspection standard must be agreed to *prior to the acceptance of the Purchase Order.* All claims resulting from the inspection judgment must be made *in writing* within 90 days after the Bill of Lading for ocean freight. If the buyer fails to notify the seller within the above specified time, the lot shall be regarded as accepted by the buyer.

The buyer has the right to accept the lot based upon the incoming inspection standard, and that the quality of the lot is regarded as accepted by the buyer, or to reject the lot based on the same. If any one lot of the products is rejected *due to defects attributable to the seller,* and if the seller agrees with this fact clearly, the *seller* shall select one of the following three options:

1. The buyer shall return the defective lot to the location specified by the seller, and the seller shall inspect all products within the lot and repair or replace the defective products.
2. The seller shall inspect all products in the buyer's facilities within a reasonable period, and repair or replace the defective products.
3. The buyer shall inspect the lot in question at the seller's cost agreed upon separately, return all defective products to the seller's specified location, and the seller shall repair or replace the defective products.

The disposition of the rejected lot shall be determined upon mutual agreement and the seller shall confirm the agreement. Final selection of the option to be used will be at the discretion of the seller, and the seller shall notify the buyer within a reasonable amount of time after being notified that the lot has been rejected by the buyer.

Major Defect

A major defect may substantially reduce the usability of the product in product applications. Major defects substantially degrade the image quality of the display such that the defect distracts from, or distorts, the information content of the display. Major defects also affect the reliability of the display and may significantly shorten the published M.T.B.F. of the product. Examples of major defects are shorted lines, spots which cover more than one pixel, or lines which are out.

Minor Defect

A minor defect does not substantially degrade the information content of the display. Minor defects deviate from the existing standards; however, they are not related to the effective use of the product or its operation. Examples of minor defects include scratches on the bezel, very small black or white spots in the display, or non-uniformity of the retardation film.

Incoming Inspection Method

Unless otherwise agreed upon by both parties, the incoming inspection method shall be per Mil-Std-105D, Inspection Level II, normal inspection, single sampling.

AQL

The AQL (Acceptable Quality Level) shall be based on the sampling table found in Mil-Std-105D and shall be based on the delivered lot size. The AQL level for major defects is in accordance with the respective inspection document, or based on agreed levels prior to placement of the purchase order. These levels for all of the above mentioned documents are 1.0 for major defects and 2.5 for minor defects. Please refer to the table in Mil-Std-105D to determine the number of defects allowable for a given lot size.

Sampling Inspection Method

The inspection level is level II as specified in Mil-Std-105D. It is important that the buyer keep records on the display serial numbers and the results of the inspection, so that the defects reported to the seller may be verified. Failure to keep accurate records will only delay the disposition of suspected non-compliant material.

Environmental Conditions

The inspection shall be accomplished at room temperature, 20 to 25°C, and at 65% relative humidity ± 5%.

Lot Size

The sampling inspection method shall be a normal inspection with single sampling. This means that only random samples of the quantity specified in Mil-Std-105D shall be used and that if there are no failures in the random sample, the lot shall be deemed to be acceptable.

Inspection Conditions

The inspector shall be at least 30 cm (12 inches) from the front surface of the display. Lighting shall be at approximately 1600 Lux (the output of one 20 W fluorescent lamp 50 cm from the display surface). Inspection criteria that falls outside of these simple rules are not allowed, and determinations based on non-compliant standards will not be allowed.

Display On Inspection

The supply voltage shall be normal for the display being tested, and the viewing angle from the technician to the display shall be as per the product specification. The display must be adjusted to optimum contrast. Judgement at other settings are not considered.

Judgment for Accepted Units and Subsequent Disposal

Acceptable Lot

If the number of defects found in the units sampled from the lot is *equal to or less than* the applicable AQL, the lot in question shall be regarded as acceptable.

Rejected Lot

If the number of defects exceeds the applicable AQL, the lot shall be rejected. At this point the buyer shall inform the seller of the detailed results of the inspection within the period specified in the terms of the purchase order.

Final rejection of the lot shall be on the basis of the incoming inspection standard and due to defects attributable to the seller, and the seller admits this fact clearly. Disposal of the rejected lot shall be as per the three criteria set up in 'The Buyer's Rights.'

Black Spots – Type I

When the unit operates, Spots appear dark in the display patterns and remain unvaried in terms of size and shade regardless of the adjustment of VEE (Figure 7).

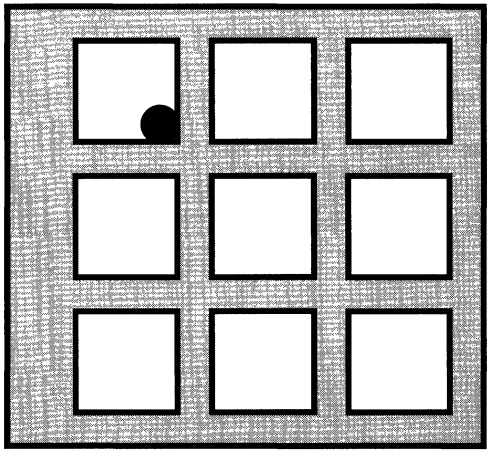


Figure 7. Black Spots – Type I

Black Spots – Type II

When the unit operates, Spots appear dark in the display patterns and vary in terms of size and shade with the adjustment of VEE (Figure 8).

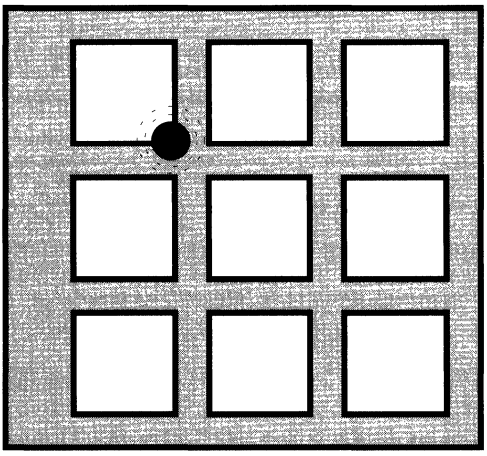


Figure 8. Black Spots – Type II

Black Lines – Type I

When the unit operates, lines appear dark in the display patterns and remain unvaried in terms of size and shade regardless of the adjustment of VEE (Figure 9).

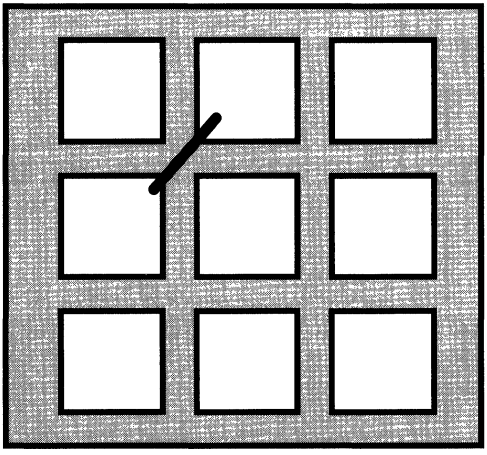


Figure 9. Black Lines – Type I

Black Lines – Type II

When the unit operates, lines appear dark in the display patterns and vary in terms of size and shade with the adjustment of VEE (Figure 10).

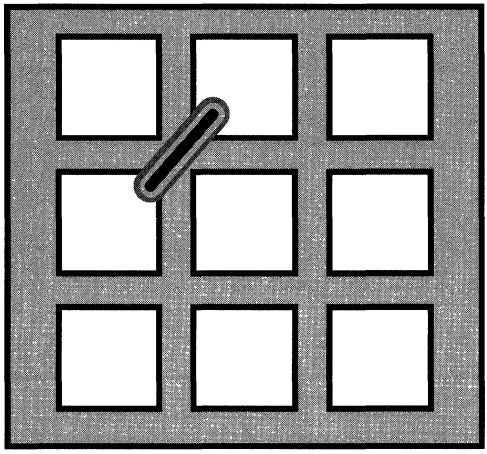


Figure 10. Black Lines – Type II

White Spots – Type I

When the unit operates, Spots appear light in the display patterns and remain unvaried in terms of size and shade regardless of the adjustment of VEE (Figure 11).

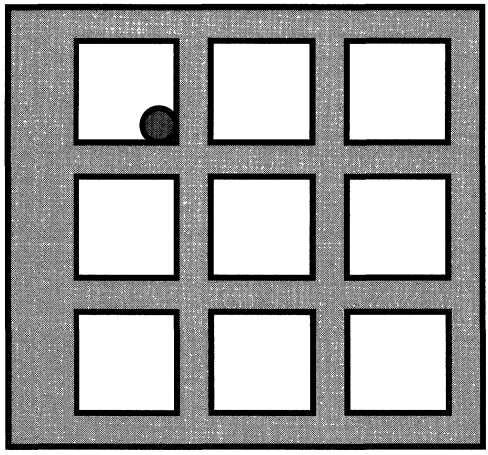


Figure 11. White Spots – Type I

White Spots – Type II

When the unit operates, Spots appear light in the display patterns and vary in terms of size and shade with the adjustment of VEE (Figure 12).

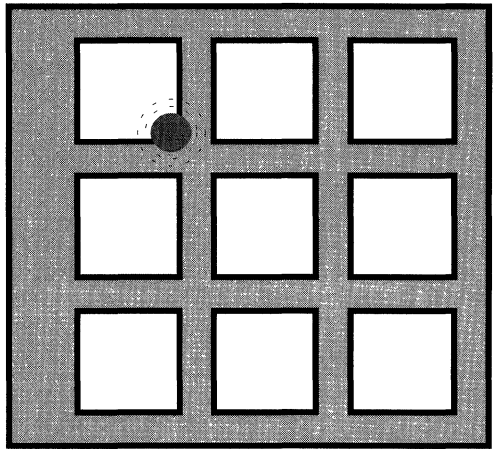


Figure 12. White Spots – Type II

White Lines – Type I

When the unit operates, lines appear light in the display patterns and remain unvaried in terms of size and shade regardless of the adjustment of VEE (Figure 13).

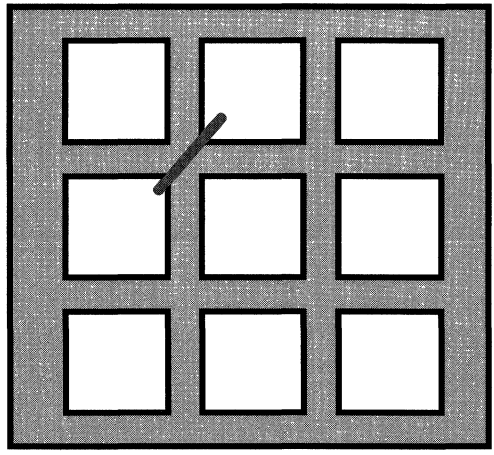


Figure 13. White Lines – Type I

White Lines – Type II

When the unit operates, faint lines appear in the display patterns and vary in terms of size and shade with the adjustment of VEE (Figure 14).

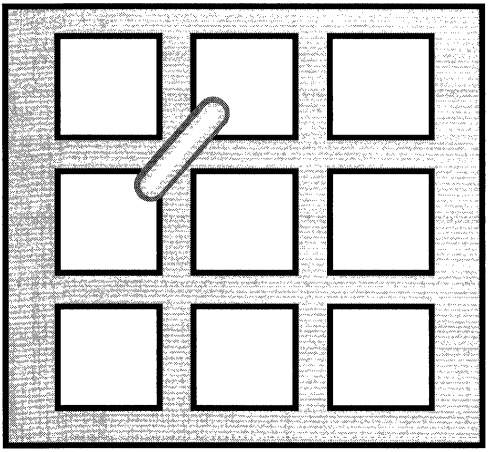


Figure 14. White Lines – Type II

Red, Green or Blue Spots (Color Displays)

These are spots which may be seen clearly on or outside of the color filters. These spots are caused by defects in the color filter panel where the wrong color is displayed or deposited in a color filter area.

Uniformity of Retardation Film

When the unit operates, irregular color appears due to non-uniform retardation film.

Vertical Line Off

When the unit operates, one or more vertical lines do not appear in the display patterns (Figure 15).

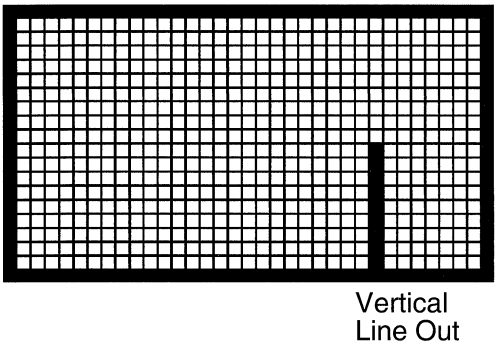


Figure 15. Vertical Line Off

Horizontal Line Off

When the unit operates, one or more horizontal lines do not appear in the display patterns (Figure 16).

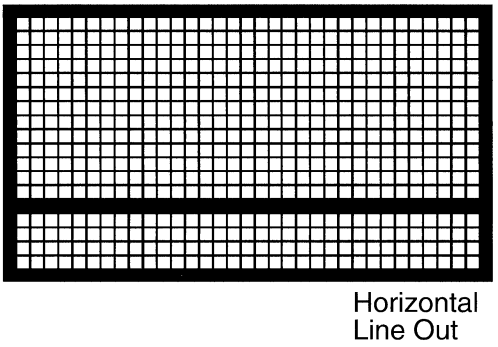


Figure 16. Horizontal Line Off

Cross Lines Off

When the unit operates, one or more vertical and horizontal lines do not appear in the display patterns (Figure 17).

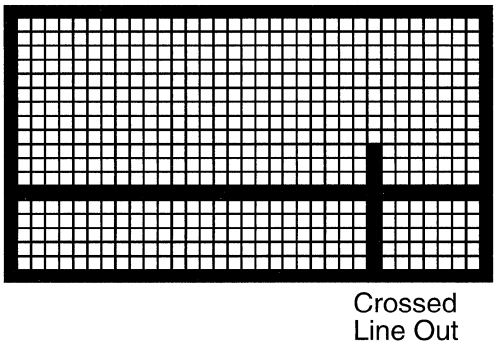


Figure 17. Cross Lines Off

Dot Width

Variations in the individual dot widths occur due to variations in the processing if the photographic image on the display substrates (Figure 18).

Center Line Gap

An excessive center line gap occurs between the upper and lower display sections due to irregularities in the display image processing.

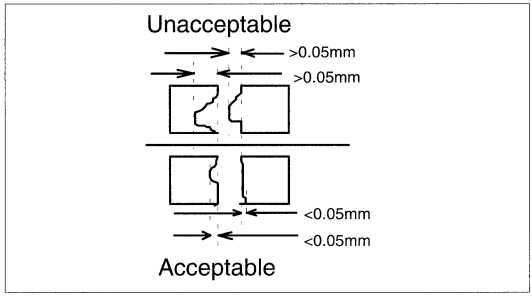


Figure 18. Dot Width

Polarizer Scratches and Dents

Polarizer scratches and dents should be ignored except in cases where they cause black or white lines or spots as per the individual criteria listed below. The scratches and dents *shall not* be considered when they can only be seen under high-intensity lighting reflected off of the polarizer causing magnification.

Polarizer Inclination Bubbles

Extremely small bubbles may be found under the polarizer. These should be ignored unless they exceed the criteria specified for the type of part being inspected as outlined below.

Stains on the LCD Panel Surface

Stains on the polarizer may include very light stains from packaging materials to water marks. Almost all stains are considered minor defects since they do not interfere with the image quality of the finished display.

Rainbow Color

Rainbow colors within the display are typically caused by variations in the cell gap between the front and rear glass. There shall be no noticeable rainbow colors within the active area of any display when it is being driven at its optimum contrast.

Bezel Appearance

Minor scratches in the bezel finish are to be ignored. If there are deep scratches, or areas where the plating is missing and the base metal is exposed, the bezel may be rejected. This is a minor defect. Irregular plating or coatings are to be ignored.

MONOCHROME/COLOR PASSIVE DISPLAY STANDARDS

The following is a summary of the reject criteria specifically related to the determination of defects in passive displays. This criteria is clearly defined as it appears in specifications S-U-012-01 (monochrome) and S-U-014-01 (passive color) concerning major and minor defects.

Vertical Line Out

If any vertical line is out on the display, this is considered a major defect and the display is to be rejected.

Horizontal Line Out

If any horizontal line is out on the display, this is considered a major defect and the display is to be rejected.

Black/White Spots – Type I

If there is a spot within the active area whose diameter is >0.6 mm, it is to be considered a major defect and the panel is to be rejected.

If there are spots within the active area whose diameter is ≥0.35 mm and ≤0.6 mm, this is a minor defect. Reject the panel if there is more than one spot.

If there are spots within the active area whose diameter is ≥0.20 mm and ≤0.35 mm, reject the panel if there are three or more spots.

If there are spots within the active area whose diameter is smaller than 0.20 mm, these are not rejectable (Figure 19).

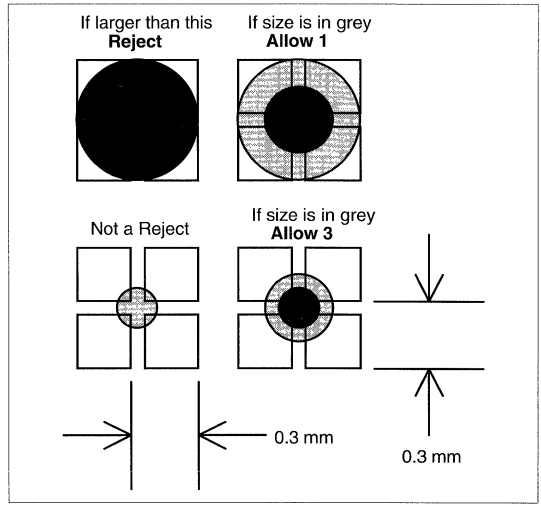


Figure 19. Black/White Spots – Type I

Black/White Lines – Type I

If there is a line within the active area whose length is >1 mm and whose width is >0.05 mm, it is considered to be a minor defect toward the AQL sample; however, the panel is to be rejected.

If there is a line within the active area whose length is <5.0 mm and whose width is >0.02 and ≤0.05 mm, it is considered to be a minor defect toward the AQL sample; however, the panel is to be rejected (Figure 20).

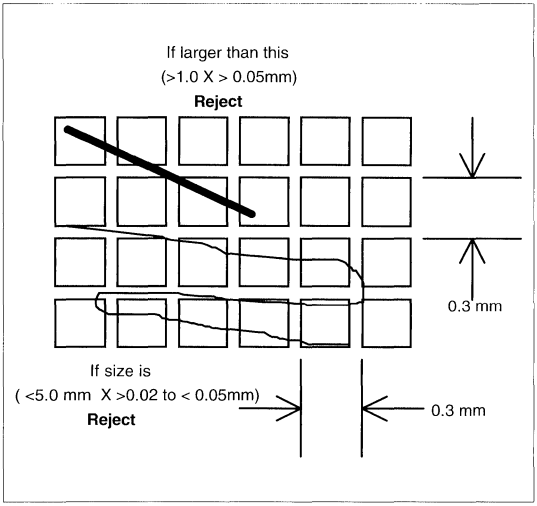


Figure 20. Black/White Lines – Type I

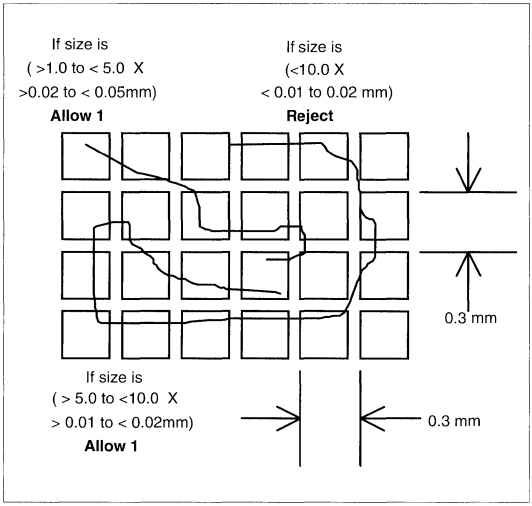


Figure 21. Black/White Lines – Type I

If there is a line within the active area whose length is >1.0 and ≤5.0 mm and whose width is >0.02 and ≤0.05 mm, this is considered to be a minor defect toward the AQL sample; however, one defect is allowed in any panel.

If there is a line within the active area whose length is >5.0 mm and <10.0 mm, and whose width is between >0.01 and ≥0.02 mm, it considered to be a minor defect toward the AQL sample; however, the panel is rejected.

If there is a line within the active area whose length is >5.0 and ≤10.0 mm and whose width is >0.01 and ≤0.02 mm, this is considered to be a minor defect toward the AQL sample; however, one defect is allowed in any panel (Figure 21).

Black Spots – Type II

In panels under 13K mm^2, the whole screen will be assessed for defects with the criteria for a quadrant in larger displays. For larger displays, each quadrant of the screen will be assessed for black spots. If there is a spot >1.2 mm in diameter, it will be considered to be a minor defect in the AQL sample; however, the panel will be rejected.

If there is a spot >0.85 mm in diameter but ≤1.2 mm in diameter, allow up to two per quadrant.

If there are spots <0.85 mm in diameter, they are not considered to be defects (Figure 22).

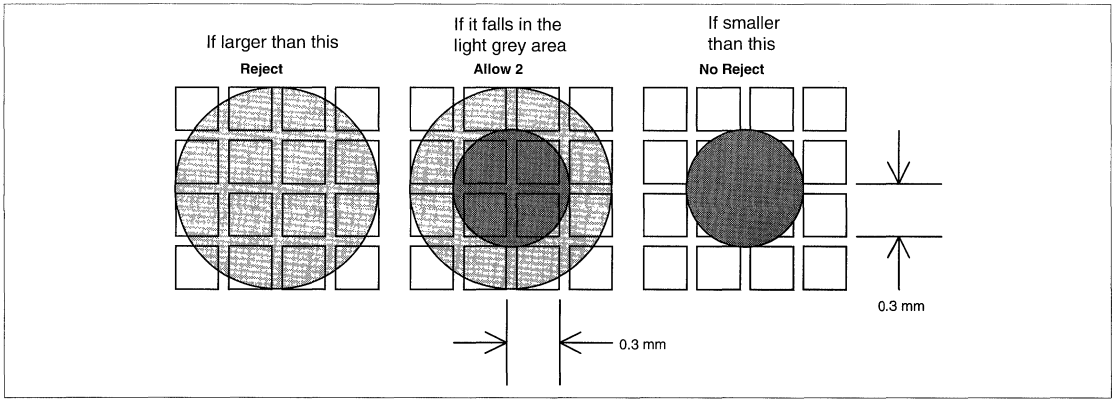


Figure 22. Black Spots – Type II

Black Lines – Type II

No noticeable black lines shall be found within the active area. Black lines are considered to be a minor defect in the AQL sample.

White Spots – Type II

In panels under 13K mm^2, the whole screen will be assessed for defects with the criteria for a quadrant in larger displays. For larger displays, each quadrant of the screen will be assessed for white spots. If there is a spot >1.2 mm in diameter, it will be considered to be a minor defect in the AQL sample; however, the panel will be rejected (Figure 23).

If there is a spot >0.85 mm in diameter but ≤1.2 mm in diameter, allow up to two per quadrant.

If there are spots <0.85 millimeters in diameter, they are not considered to be defects.

White Lines – Type II

No noticeable white lines shall be found within the active area. White lines are considered to be a minor defect in the AQL sample.

Retardation Film Uniformity

Irregular color due to slight variations in the retardation film should be ignored. Objectionable variations that may reduce the display quality are considered to be minor defects in the AQL sample and the panel can be rejected.

Dot Width

Variations in the width of the dots on the display may be up to 0.05 mm. Widths which exceed this criteria are considered to be minor defects in the AQL sample, and may cause the panel to be rejected (Figure 24).

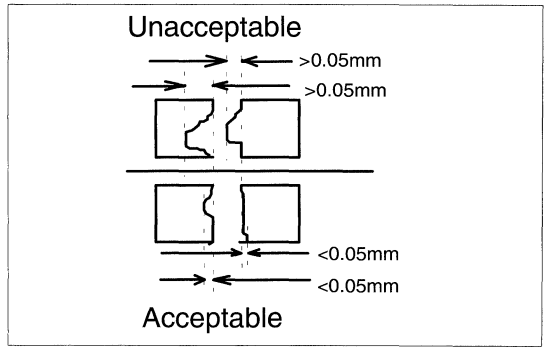


Figure 24. Dot Width

Center Gap Line

The center line gap between the upper and lower halves of the screen may vary up to 0.07 millimeters. Excessive gaps in excess of 0.07 millimeters would be considered a minor defect in the AQL sample and would be cause for rejection of the display. This condition applies to monochrome displays only.

Scratches/Dents in the Polarizer

Scratches and dents in the polarizer shall be assessed in accordance with the rules for black and white spots (Type I) and lines (Type I) as previously outlined. They shall not be deemed a defect if they do not affect the image quality as viewed normal to the display. Assessment of scratches and dents through 'Magnification' as previously defined *is not allowed.*

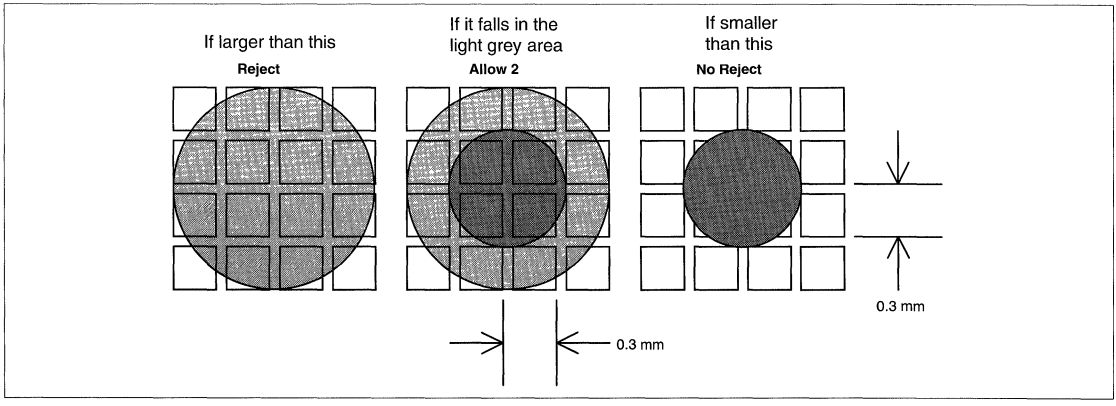


Figure 23. White Spots – Type II

Polarizer Bubbles

Bubbles under the polarizer shall be assessed as follows:

- If the diameter is >1.0 mm, it will be considered to be a minor defect in the AQL sample; however, the panel will be rejected.
- If the diameter is >0.5 and ≤1.0 mm, allow one in the display active area. It will be considered to be a minor defect in the AQL sample.
- If the diameter is <0.5 mm, ignore them since they are not a defect.

Stains

Stains which can be cleaned off with a soft cloth are considered acceptable. All others will be considered to be a minor defect in the AQL sample. The panel is rejectable.

Backlight Failure

If the backlight is inoperable due to failure of the fluorescent tube or wire harness, the display is to be rejected as a major defect.

COLOR TFT DISPLAY STANDARDS

The guidelines for testing TFT displays are exactly the same as for Passive LCDs. The test station and test equipment required are the same except that Audio Visual displays will require an analog input device to properly drive the display electronics. The differences in display inspection criteria will be explained in the following section.

Audio Visual Display Standards

Special Ambient Conditions for Inspection

- Viewing Distance: 350 mm ±50 mm
- Ambient illumination: 100 to 150 Lux
- Panel Temperature: 35°C maximum
- Panel Surface Luminance: 100 Nits or more
- Viewing Angle: Normal to the display surface (90° to the surface)

Black Spots

There may be up to two black spots in the active area, ≤0.5 mm in diameter. If there are more than two black spots, or the spots are >0.5 mm in diameter, the display should be rejected.

Nap

Nap refers to fibrous material such as lint within the display cavity. Up to two fibers may appear in the display ≤3 mm long.

Scratches

Up to three scratches that are ≤10 mm long and 0.04 mm wide may be on the front of the display.

Dents

Up to three dents ≤0.5 mm in diameter may be on the front of the display.

Active Area Defects

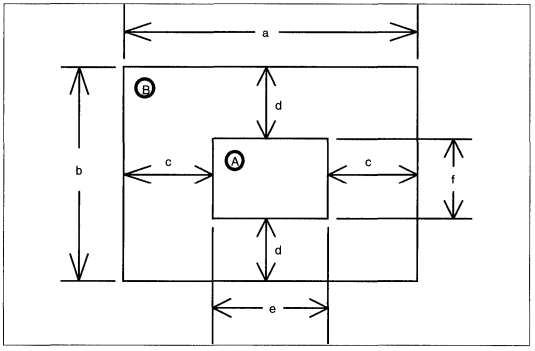


Figure 25. Defect Zones

PART NUMBER	a	b	c	d	e	f
LQ4XXXX	81.9	61.8	19.9	14.5	42.1	32.8
LQ6XXXX	113.5	87.6	27.7	21.1	58.4	45.4

Acceptance Standards

Refer to Table 3.

Table 3. Acceptance Standards

		LQ4XXXX			LQ6XXXX		
ZONE		A	B	TOTAL	A	B	TOTAL
Red Green Blue	High Level	1	3	4	1	3	4
	Low Level	2	5	7	3	6	9
Total		3	8	9 *	4	9	9 *

* There can be no more than 9 defects per zone.

Dark Spots

Up to two clusters of two non-functioning adjacent elements are allowed on the display. Up to 10 defects are allowed on a display; however, there must be 5 mm between each failure. This excludes the two allowable clusters previously mentioned.

ZONE	A	B	TOTAL
Black Spots	5	8	9 Max.

Pixel Failures

In order to assess whether a pixel failure is a high-level or low-level failure, the display must be observed through various neutral density transmission filters. It is recommended that the test technician be supplied with a set of filters constructed as per the guidelines supplied in Appendix 1, *Making Incoming Inspection Filters* later in this application note.

High-level failures occur in pixels which are still apparent when observed through a 2% transmission filter.

Low-level failures are apparent when viewed with a 5% filter, yet are not apparent through a 2% filter.

A non-operating pixel which is not visible through a 5% transmission filter is not considered to be a defect.

Scratches in the Color Filter

The color filters used in generation of the color spectrum for the display are located internally in the display glass envelope. Even though care is taken to prevent scratching of the color filter material, scratches will occasionally occur. Scratches are considered to be high level if they are visible over one-half of the pixel. Scratches are not considered defective if they cover less than one-half of the pixel (Figure 26).

Figure 26. Scratches in the Color Filter

Scratches on the Chrome Mask

Within the display is a Chrome mask, which is deposited between the color filters, that is used for contrast enhancement and as a light shield to prevent activation of the TFT by stray light. Scratches are high-level defects if the scratch is over 50 micrometers in diameter (about 25% of a pixel). Scratches are low-level defects if the scratch is around 50 micrometers in diameter. Finally, scratches are not a defect if the scratch is below 50 micrometers.

Defect Proximity

Limits have been developed to determine how close any two defects may be in the display active area. High-level defects must be separated by 15 mm (0.6 inches) minimum. Low-level defects must be at least 5 mm (0.2 inches) apart.

Display Non-Uniformity

In assessing display uniformity, the display needs to be powered up with each of the three primary colors as well as white and black. The display is observed through a 1% transmission filter using the ambient conditions as previously described. There should not be any variation in the display uniformity as the display is viewed with each color respectively.

Display Persistence

Display persistence refers to the potential problem of a latent image appearing for a period of time after the display is blanked. For this test, a static image should be placed on the screen for a minimum of five seconds. As the display is blanked, the image must disappear completely within 10 seconds. If the image does not disappear within the specified time, this is considered to be a major defect and the display should be rejected.

Other Defects

Other defects found in incoming inspection of small Audio Visual TFT displays include the following:

- Line and Spot defects as with passive displays – Major Defects
- Pixel Failures – Minor defects as per the above tables
- Vertical Stripes – Minor Defects
- Contrast Ratio – Minor Defect
- Power Dissipation – Minor Defect

Office Automation Display Standards

In assessing the quality of Office Automation displays, the same policies and procedures previously outlined apply. Inspection criteria specifically applicable to Office automation displays are discussed in the following section. Office Automation displays include 9-inch, 10-inch, 12-inch, and 14-inch striped format displays with a digital

RGB interface. These displays are specifically designed for use in computer applications such as Notebook Computer displays, Overhead Projection Displays, and in Computer Workstations.

Special Ambient Conditions for Inspection

- Viewing Distance: 350 mm ±50 mm
- Ambient illumination: 300 to 700 Lux (Normally 500 Lux)
- Panel Temperature: 20-25°C
- Panel Surface Luminance: 100 Nits or more
- Viewing Angle: Normal to the display surface (90° to the surface)

Black Spots

There may be up to four black spots in the active area that are up to 0.5 mm in diameter. If there are more than four black spots, or the spots are larger than 0.5 mm in diameter, the display should be rejected.

Nap

Nap refers to fibrous material, such as lint, within the display cavity. Up to four fibers may appear in the display up to 3 mm long.

Scratches

Up to six scratches that are up to 10 mm long and 0.04 wide may be on the front of the display.

Dents

Up to six dents that are less than 0.5 mm in diameter may be on the front of the display.

Active Area Defects

Defect Zones

On active matrix color displays for office automation, the defect zone includes the entire active area.

Acceptance Standards

From time to time, displays may be shipped with one or more picture elements that are non-functional. These elements may appear as lighted elements on a black background (all pixels activated), or as darkened elements on a lighted background (all pixels deactivated). If elements remain lighted on a black background, the following table applies as to the allowable level of defects. High- and low-level defects are evaluated using the 1%, 2%, and 5% transmission filters as explained below.

Off Pixels

Because of the visual effect resulting from pixels being non-functional in the 'ON' state, the following rules apply as to the proximity of multiple failures:

- Any two Green, high level failures *cannot* occur within four spots.
- If two non-functioning elements are joined together, both cannot be high-level failures.
- Only two joined failures are allowed, with one being a high-level and one being a low-level failure.
- Only five joined failures are allowed if both elements are low-level failures.

		ZONE TOTAL
Red Green Blue	High Level	12 max.
	High/Low Level	25 max.

On Pixels

The following table applies to dark picture elements on a display with all elements turned on.

Up to two clusters of two non-functioning adjacent elements are allowed on the display.

Up to 10 defects are allowed on a display; however, there must be 5 millimeters between each failure. This excludes the two allowable clusters mentioned above.

	ZONE TOTAL
Black Spots	12 max.

Pixel Failures

In order to assess whether a pixel failure is a high-level or low-level failure, the display must be observed through various neutral density transmission filters. It is recommended that the test technician be supplied with a set of filters constructed per the guidelines supplied in Appendix 1, *Making Incoming Inspection Filters* later in this paper.

High-level failures occur in pixels which are still apparent when observed through a 2% transmission filter.

Low-level failures are apparent when viewed with a 5% filter, yet are not apparent through a 2% filter.

A non-operating pixel, which is not visible through a 5% transmission filter, is not considered to be a defect.

Scratches in the Color Filter

The color filters used in generation of the color spectrum for the display are located internally in the display glass envelope. Even though care is taken to prevent scratching of the color filter material, scratches will occasionally occur. Scratches are considered to be high level if they are visible over one-half of the pixel. Scratches are not considered a defect if they cover less than one-half of the pixel (Figure 27).

Figure 27. Scratches in the Color Filter

Scratches on the Chrome Mask

Within the display is a Chrome mask, which is deposited between the color filters, that is used for contrast enhancement and as a light shield to prevent activation of the TFT by stray light. Scratches are high-level defects if the scratch is over 50 micrometers in diameter (about 25% of a pixel). Scratches are low-level defects if the scratch is around 50 micrometers in diameter. Finally, scratches are not considered a defect if the scratch is below 50 micrometers.

Defect Proximity

Limits have been developed to determine how close any two defects, other than joining defects as described above, may be in the display active area. High-level defects must be separated by 15 mm (0.6 inches) minimum. Low-level defects must be at least 5 mm (0.2 inches) apart.

Display Non-Uniformity

In assessing display uniformity, the display needs to be powered up with each of the three primary colors as well as white and black. The display is observed through a 1% transmission filter using the ambient conditions described previously. There should not be any variation in the display uniformity as the display is viewed with each color respectively.

Display Persistence

Display persistence refers to the potential problem of there being a latent image appearing for a period of time after the display is blanked. For this test, a static image should be placed on the screen for a minimum of 30 minutes. As the display is blanked, the image must disappear completely within 10 seconds. If the image does not disappear within the specified time, this is considered to be a major defect and the display should be rejected.

Other Defects

Other defects found in incoming inspection of small Audio Visual TFT displays include the following:

- Line and Spot defects as with passive displays - Major Defects
- Pixel Failures - Minor defects as per the above tables
- Vertical Stripes - Minor Defects
- Contrast Ratio - Minor Defect
- Power Dissipation - Minor Defect

CONCLUSIONS

In producing this document, it has been my intention to clearly define the various terms and procedures used for the incoming inspection of Liquid Crystal Displays. The goal of this document is to become an aid by which both manufacturers and users alike may understand the requirements of inspection, and the terms used in discussing various defects found in the finished displays. The incoming inspection standards document still remains the document of record in the event of any conflicting information.

APPENDIX 1 – MAKING INCOMING INSPECTION FILTERS

Inspection filters are used to determine if the display defects are high level or low level. The following describes how to make go/no go gauges to determine high-level or low-level defects.

Several gauges are required in order to completely assess whether a defect is a high-level or low-level defect. The gauges needed are as follows:

- 1 each 1% transparency filter with a cross-hair target (optional)
- 1 each 2% transparency filter with a cross-hair target (optional)
- 1 each 5% transparency filter with a cross-hair target (optional)
- 1 each clear filter with a cross-hair target (optional)
- The clear filter should be made on a one-inch square opaque substrate with a 10 mm hole in the middle where the filter is.
- The 1% filter should be made on a one-inch square opaque substrate with a 10 mm hole in the middle where the filter is.
- The 2% filter should be made on a one-inch square opaque substrate with a 30 mm hole in the middle where the filter is.
- The 5% filter should be made on a one-inch square opaque substrate with a 10 mm hole in the middle where the filter is.

The Fuji neutral density filters shown in Table 4 have been previously used to make the various filters. The filters described here are neutral density optical filters that are Tri Acetyl Cellulose base. U.S. Filter suppliers may be able to make transmission filters. U.S. sources may exist for these filters, however, we have not evaluated any to date.

Table 4. Fuji Neutral Density Filters

FUJI #	TRANSMISSIVITY	% TRANSPARENCY
0.3	0.5012	50.12%
0.5	0.3160	31.60%
0.6	0.2510	25.10%
0.7	0.1995	19.95%
0.8	0.1580	15.80%
0.9	0.1260	12.60%
1.0	0.1000	10.00%
1.2	0.0630	6.30%
2.0	0.0100	1.00%

Appropriate filters may be made by combining the Fuji filters as shown in Table 5.

Any of the combinations shown in Table 5 will provide satisfactory results. For each type of filter, a 1%, 2%, or 5%, only one combination is required. Each filter stack should be bonded together using a suitable optical adhesive in order to prevent further losses due to the intermediate interfacial layers. If optical adhesives are not available, then the filter should be tightly placed together without air gaps between filters.

Table 5. Fuji Filter Combinations

	FUJI #	TRANSITIVITY	% TRANSMISSION	
For a 1% Filter	2.0	0.0100	1.00%	or
	1.2 + 0.8	0.0630 X 0.1580	0.99%	or
	1.0 + 1.0	0.1000 X 0.1000	1.00%	
For a 2% Filter	0.9 + 0.8	0.1260 X 0.1580	1.99%	or
	0.7+ 0.1	0.1995 X 0.1000	1.99%	or
	0.5 + 1.2	0.3160 X 0.0630	1.99%	
For a 5% Filter	0.6 + 0.7	0.2510 X 0.1995	5.00%	or
	0.5 + 0.8	0.3160 X 0.1580	4.99%	or
	0.3 + 1.0	0.5012 X 0.1000	5.01%	

APPENDIX 2 – AQL SAMPLING TABLES (MIL-STD-105D)

Sample Size Code Letters

LOT SIZE	SPECIAL INSPECTION LEVELS				GENERAL INSPECTIONS LEVELS		
	S - 1	S - 2	S - 3	S - 4	I	II	III
2 - 8	A	A	A	A	A	A	B
9 - 15	A	A	A	A	A	B	C
16 - 25	A	A	B	B	B	C	D
26 - 50	A	B	B	C	C	D	E
51 - 90	B	B	C	C	C	E	F
91 - 150	B	B	C	D	D	F	G
151 - 280	B	C	D	E	E	G	H
281 - 500	B	C	D	E	F	H	J
501 - 1200	C	C	E	F	G	J	K
1201 - 3200	C	D	E	G	H	K	L
3201 - 10000	C	D	F	G	J	L	M
10001 - 35000	C	D	F	H	K	M	N
35001 - 150000	D	E	G	J	L	N	P
150001 - 500000	D	E	G	J	M	P	Q
500001 and over	D	E	H	K	N	Q	R

MIL-STD-105 D

Single Sampling Plans for Normal Inspection (Master Table)

Sample Size Code Letter	Sample Size	Acceptable Quality Levels (AQL) (Normal Inspection)																										
		0.010	0.015	0.025	0.040	0.065	0.10	0.15	0.25	0.40	0.65	1.0	1.5	2.5	4.0	6.5	10	15	25	40	65	100	150	250	400	650	1000
		Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re
A	2	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↓	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31
B	3	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31	44 45
C	5	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31	44 45	↑
D	8	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31	44 45	↑	
E	13	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31	44 45	↑		
F	20	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑	↑	↑	↑
G	32	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑						
H	50	↓	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑							
J	80	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑								
K	125	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑									
L	200	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑										
M	315	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑											
N	500	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑												
P	800	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑													
Q	1250	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑														
R	2000	↑		1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑															

↓ = Use first sampling plan below arrow. If sample size equals or exceeds lot size, do 100% inspection.

MIL-STD-105 D

↑ = Use first sampling plan above arrow.

Ac = Acceptance number

Re = Rejection number

Single Sampling Plans for Tightened Inspection (Master Table)

Sample Size Code Letter	Sample Size	Acceptable Quality Levels (AQL) (Tightened Inspection)																												
		0.010	0.015	0.025	0.040	0.065	0.10	0.15	0.25	0.40	0.65	1.0	1.5	2.5	4.0	6.5	10	15	25	40	65	100	150	250	400	650	1000
		Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re
A	2																			1 2	2 3	3 4	5 6	8 9	12 13	18 19	27 28
B	3															0 1			1 2	2 3	3 4	5 6	8 9	12 13	18 19	27 28	41 42
C	5														0 1			1 2	2 3	3 4	5 6	8 9	12 13	18 19	27 28	41 42	
D	8													0 1			1 2	2 3	3 4	5 6	8 9	12 13	18 19	27 28	41 42		
E	13												0 1			1 2	2 3	3 4	5 6	8 9	12 13	18 19	27 28	41 42			
F	20											0 1			1 2	2 3	3 4	5 6	8 9	12 13	18 19						
G	32										0 1			1 2	2 3	3 4	5 6	8 9	12 13	18 19							
H	50									0 1			1 2	2 3	3 4	5 6	8 9	12 13	18 19								
J	80								0 1			1 2	2 3	3 4	5 6	8 9	12 13	18 19									
K	125							0 1			1 2	2 3	3 4	5 6	8 9	12 13	18 19										
L	200						0 1			1 2	2 3	3 4	5 6	8 9	12 13	18 19											
M	315					0 1			1 2	2 3	3 4	5 6	8 9	12 13	18 19												
N	500				0 1			1 2	2 3	3 4	5 6	8 9	12 13	18 19													
P	800			0 1			1 2	2 3	3 4	5 6	8 9	12 13	18 19														
Q	1250		0 1			1 2	2 3	3 4	5 6	8 9	12 13	18 19															
R	2000	0 1			1 2	2 3	3 4	5 6	8 9	12 13	18 19																
S	3150			1 2																							

↓ = Use first sampling plan below arrow. If sample size equals or exceeds lot size, do 100% inspection.

MIL-STD-105 D

↑ = Use first sampling above arrow.

Ac = Acceptance number

Re = Rejection number

Single Sampling Plans for Reduced Inspection (Master Table)

Sample Size Code Letter	Sample Size	Acceptable Quality Levels (AQL) (Reduced Inspection)*																												
		0.010	0.015	0.025	0.040	0.065	0.10	0.15	0.25	0.40	0.65	1.0	1.5	2.5	4.0	6.5	10	15	25	40	65	100	150	250	400	650	1000
		Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re
A	2	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↓	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31
B	2														0 1	↑		0 2	1 3	2 4	3 5	5 6	7 8	10 11	14 15	21 22	30 31
C	2													0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	14 17	21 24	↑
D	3												0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	14 17	21 24	↑		
E	5											0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	14 17	21 24	↑			
F	8										0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑	↑	↑	↑			
G	13									0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑							
H	20								0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑								
J	32							0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑									
K	50						0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑										
L	80					0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑											
M	125				0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑												
N	200			0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑													
P	315		0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑														
Q	500	0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑															
R	800	↑		0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑																

↓ = Use first sampling plan below arrow. If sample size equals or exceeds lot size, do 100% inspection.

MIL-STD-105 D

↑ = Use first sampling plan above arrow.

Ac = Acceptance number

Re = Rejection number

* = If the acceptance number has been exceeded, but the rejection number has not been reached, accept the lot, but resume normal inspection.

PC Video Board Interfaces for Sharp Video LCD Displays

Alan Dragon, Senior Field Application Engineer

INTRODUCTION

This application note will concentrate on PC video boards that will interface directly with Sharp's color LCD video displays. The Sharp LCD displays that can interface with these video boards are: the LQ4RA/NC Series, the LQ6RA/NC Series, and the LQ9RA03. All of these displays are NTSC-compatible, but each series has a different input configuration. The 'NC' series uses a composite video input, and the 'RA' series uses an R,G,B, composite sync (4-line) video input. (The Sharp color video displays, 'LQ3' series and 'LQ4RE' series, cannot interface directly with standard video signals without additional logic, and thus will not be covered.) The 'RA' series is PAL-compatible and video boards with this option will be indicated. Some of the video boards are also compatible to Apple Macintosh, and this will be noted.

VIDEO BOARDS

One of the main reasons to use a PC video board is to display standard PC software (programs/images) on a video display rather than a standard VGA-type monitor. This might be done for presentation purposes (large TV monitor or projection TV) or to record the information on a VCR. All of the video boards that are mentioned have VGA to NTSC (or PAL) conversion capabilities. This means that to use one of these video boards, the PC being used must have the hardware/software to handle a VGA format. Although the focus of this application note is on video boards, there are also many stand-alone boxes that perform a similar function. These boxes can take a standard VGA monitor connector and provide an NTSC/PAL output to the appropriate video monitor. These boxes tend to be significantly more expensive than video PC boards and require a standard VGA card to be already installed in a PC. Some of the video board manufacturers mentioned also provide these remote box solutions. They can be contacted directly for further information.

It is important to understand that some of the VGA modes have display formats of 640 × 480 pixels and that the Sharp video displays have NTSC formats of 160H × 234V pixels up to 320H × 456V pixels. When scan conversion is done from VGA to NTSC, resolution will be lost. This should be taken into account when displaying small text or fine detail on the display. VGA application software may need to be optimized for use on a video compatible display from Sharp.

Most of the video boards listed in Table 1 are basic VGA to NTSC/PAL conversion interfaces. (This list is not a complete list of manufacturers with this type of product, but rather a thorough and useful list of up-to-date products.) Many of the manufacturers listed have a number of video boards with varying degrees of functionality. These boards can have such features as frame grabbing, overlaying and manipulation of various video and graphics sources, and general multimedia processing. The cost of these boards will also vary directly with the amount of functionality. The lower cost boards will have less input and output video sources and may require special driver software to control the conversion. The higher end boards may not even need an existing VGA board as they have VGA functionality on-board and can output VGA simultaneously (along) with video. Some video boards have VGA controllers on-board and can operate by themselves. Other video boards need the existing VGA controller card already in the PC to connect to the VGA feature connector.

Most of these video boards provide genlock circuitry which is required for high-quality videotaping of the converted signal. Horizontal and vertical positioning are also available on many models along with chroma keying which allows overlaying of live video sources on computer generated graphics. Image quality will vary depending on the functionality and, ultimately, the price of the video board. The application of the video output will determine what product and what level of quality should be acceptable.

One manufacturer, Infonix, has designed a board that was made specifically for the 6″ and 4″ 'RA' series displays. This board mounts directly on the back of the LCD and will connect to a second Infonix card in a PC bus slot. This system is mechanically and electrically tuned to the Sharp 'RA' series displays.

INTERFACE CABLING

Each video interface board has its own set of connector/cabling requirements. Some have D-sub input and output connectors and others have RCA phone plugs and DIN S-Video connectors for video outputs. All of these boards must have some type of interface connector board between the display and the video board. The input connectors on the displays are not standard video connectors and other signals, such as DC supply voltages and control signals, must also be input to the display. (Interfacing and cabling to the Sharp small 6″ and 4″ TFT color displays can be found in Sharp's application note, *'Interfacing and Interconnection for 4-Inch and 6-Inch TFT LCDs.'*) Cables from the video boards (i.e., VGA feature connector to video board, D-sub to S-video, etc.) can either be ordered as accessories from the manufacturers or from other video accessory catalogues.

INTERFACE SOFTWARE

Some of the video boards require software drivers to control the conversion features. This software is provided with the video board when necessary. One of the special routines is called TSR (Terminate and Stay Resident) and modifies the PC's output along with controlling the conversion. Many manufacturers include multimedia software packages with their video boards, and these may include software drivers for DOS or Microsoft Windows.

CONCLUSION

If there is a requirement to operate a Sharp color video LCD from a PC, there are two basic choices. The first is to drive the LCD from a VGA to NTSC/PAL video board and the second is to drive the LCD from a separate scan conversion remote box. The video board can be obtained with many different levels of functionality and depending on the application required. The information given in Table 1 should provide users with enough detail to make a decision on what product will best suit their needs. Sharp does not endorse any one vendor, but suggests the user choose a vendor based on the user's individual requirements.

ACKNOWLEDGEMENT

Apple Macintosh is a registered trademark of Apple Computer, Inc.

Table 1.
Video Board Manufacturers

MANUFACTURER	MODEL	PC COMPATIBILITY	VIDEO INPUTS [1]	VIDEO OUTPUTS [1]	NTSC/PAL COMPATIBILITY	FEATURES	COMMENTS
Cardinal Technologies, Inc. 1827 Freedom Road Lancaster, PA 17601 Phone: (717) 293-3000 (800) 722-0094 FAX: (717) 293-3055	SNAPplus	PC/AT, 386 PS/2 Model 35	RGB, S-Video, NTSC	RGB, S-Video, NTSC, VGA Out	NTSC/PAL Input; NTSC Output	• Genlock • Real Time Video Special Effects	• All software drivers included • Does not require existing VGA Card
Infonix Aerospace Systems 639 Davis Road League City, TX 77573 Phone: (713) 554-4240 FAX: (713) 554-6768	MVC-6000	PC/AT, must plug into existing VGA Card Feature connector	RGB	RGB, S-Video, NTSC	NTSC Input/ Output	• Initialization software requires no driver or TSR	• Boards are electrically and mechanically made specifically for the 'LQ6RA' series LCD • Must be combined with MDI-6000A or MDI-6000B to operate with LCD • MDI board provides interface functionality, MVC board provides scan conversion
Magni Systems, Inc. 9500 SW Gemini Drive Beaverton, OR 97005 Phone: (503) 626-8400 FAX: (503) 626-6225	VGA Producer Pro	PC/AT, must plug into existing VGA Card Feature connector	S-Video, NTSC	S-Video, NTSC	NTSC Input/ Output, PAL Input/Output (2 versions)	• No special drivers required • Flicker stabilizer circuitry • Genlock	• Remote control for video special effects • Comes with 3 video software production packages • PS/2 version available for MCA bus
Omnicomp Graphics Corporation 1734 W. Sam Houstan Parkway Houston, TX 77043 Phone: (713) 464-2990 FAX: (713) 827-7540	ViVA Basic	PC/AT, must plug into existing VGA Card Feature connector		RGB, S-Video, NTSC	NTSC Output	• Flicker filter	• Software utilities provided, including TSR • Omnicomp offers many other multi-media boards with varying degrees of functionality
Redlake Corporation 718 University Avenue Suite 100 Las Gatos, CA 95030 Phone: (408) 399-5000 FAX: (408) 354-7428	PC2TV	PC/XT, must plug into existing VGA card rear 15-pin high density connector			NTSC/PAL Output (2 versions)	• Anti-flicker circuitry • No utility software required	

Table 1 (cont'd).
Video Board Manufacturers

MANUFACTURER	MODEL	PC COMPATIBILITY	VIDEO INPUTS [1]	VIDEO OUTPUTS [1]	NTSC/PAL COMPATIBILITY	FEATURES	COMMENTS
Truevision 7340 Shade Land Station Indianapolis, IN 46256 Phone: (317) 841-0332 FAX: (317) 576-7700	Video VGA	PC/AT	S-Video, NTSC	S-Video, NTSC, VGA Out	NTSC Input/ Output	• Flicker reduction module available • Genlock	• VGA/Multi-media software drivers included • Truevision offers many multi-media graphic engines for wide range of video board products
	NuVista +	Macintosh II	RGB, S-Video, NTSC	RGB, S-Video, NTSC	NTSC Input/ Output, PAL Input/Output (2 versions)	• Genlock • Digital Linear keying and chroma keying	• Multi-media software drivers included
Visionetics International 21311 Hawthorne Boulevard Suite 235 Torrance, CA 90503 Phone: (310) 316-7940 FAX: (310) 316-7457	VGA Link	PC/AT, 386; must plug into existing VGA Card Feature connector	S-Video, NTSC	S-Video, NTSC	NTSC Input/ Output	• Genlock • Chroma keying	• Ribbon cable to VGA feature connector provided
	VGA/TV Mate	PC, XT/AT; must plug into existing VGA Card Feature connector		S-Video, NTSC	NTSC/PAL Output	• Anti-flicker circuitry	• Requires TSR driver (included)
	VIGA VGA+	PC/AT	S-Video, NTSC	S-Video, NTSC, VGA Out	NTSC/PAL Output	• Includes VGA drivers and utilities	
Willow Peripherals 190 Willow Avenue Bronx, NY 10454 Phone: (212) 402-9500 FAX: (212)402-9603	VGA-TV 4000	PC/AT		S-Video, NTSC, VGA Out	NTSC/PAL Output (2 versions)	• Horizontal and vertical positioning • 8-bit and 16-bit auto switchable	• Supports all VGA modes • VGA software drivers included
Worldwide Video, Inc. 107 Main Street Newark, NJ 07105 Phone: (201) 491-5147 FAX: (201) 491-5149 (Distributes USVideo Products)	TVGA	PC XT/AT, PS/2 Model 30			NTSC	• 8-bit and 16-bit auto switchable	• Genlock overlay moduleand digitizer module available • VGA software drivers included

NOTE:

1. RGB designates R,G,B Composite Sync
   NTSC designates composite NTSC Video

BACKLIGHTING

Alan Dragon, Senior Field Application Engineer

INTRODUCTION

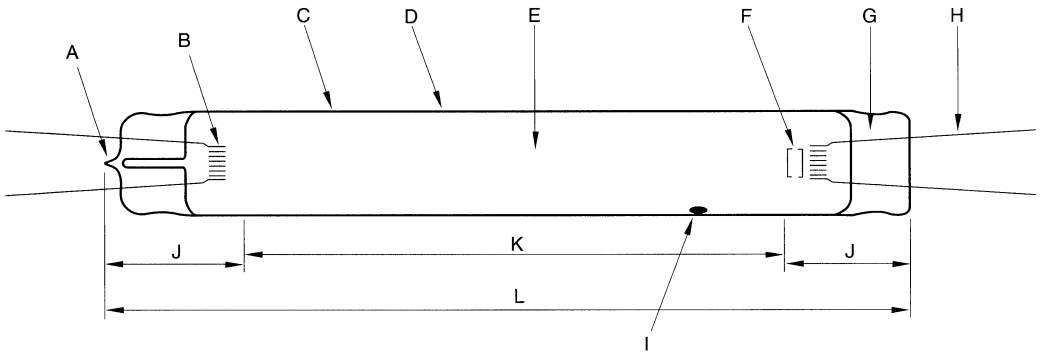
The passive, non-emissive structure of a liquid crystal display makes the backlight an integral part of this display sub-system. Although some LCDs (reflective) can operate without an integrated backlight, they still need an illuminated ambient environment for useful operation. The majority of large area, multi-purpose LCDs require an integrated backlight to operate in a wide range of ambient conditions. The backlighting technologies most often used for LCDs are hot and cold cathode fluorescence (HCFT and CCFT), electroluminescence, cathodeluminescence, and incandescence/metal halide (using a light pipe to bring the light to the display from the light source). The advantages and disadvantages of each technology are shown in Table 1. Although each technology will be described, the emphasis will be on HCFTs and CCFTs.

CATHODE FLUORESCENCE OPERATING THEORY

Hot cathode and cold cathode fluorescent tubes have similar operating mechanisms and differ only in the way the cathode emits electrons. The hot cathode emits electrons by thermionic emission and the cold cathode emits electrons by secondary emission. The flow of electrons emitted by the cathode creates a gas plasma arc. The electrodes in an AC tube configuration alternately function as cathode and anode (Figure 1). When the tube conducts, the electrons are accelerated, changing the electrical energy into electron kinetic energy. These electrons ionize mercury atoms to carry the current, and in addition, collide with orbital electrons in mercury and buffer-gas atoms. This interaction results in the excitation of mercury atoms to higher energy levels. When the excited atoms relax to a lower energy state, ultraviolet (UV) electromagnetic radiation is an emitted characteristic of mercury. The phosphor deposited on the inside of the tube transforms this radiation into visible light. The electromagnetic radiation emitted by excited mercury atoms is at its resonance wavelength of 254 nm. Each phosphor has its own excitation spectra – a wavelength range of electromagnetic radiation that can be absorbed and converted to visible light. Phosphors for fluorescent tubes are optimized to transform the 254 nm radiation to visible light with high efficiency.

Table 1.
Backlight Technologies

TYPE	ADVANTAGES	DISADVANTAGES
Hot and Cold Cathode Fluorescence	• High Efficiency • Low Cost	• Temperature Sensitive • Moderate Dimmability
Electroluminescence	• Thin Profile • Uniform Distribution	• Low Luminance • Limited Color
Cathodoluminescence	• High Luminance • Wide Dimming Range • Wide Temperature Range	• Thick Profile
Incandescence/Metal Halide (Fiber Optic Light Pipe)	• High Luminance • Wide Dimming Range	• High Infrared • Large Area Light Source



A Exhaust Tubulation tip-off
B Emitter or Cathode
C Phosphor coating on the inside of the glass envelope
D Glass Envelope
E Buffer Gas, usually argon or a mixture of inert gases
F Region of Cathode - fall voltage
G Glass Seal area
H Lead Wires
I Mercury to produce mercury vapor
J Lamp area that emits no usable light
K Length of plasma ARC
L Overall lamp length

LCD10-1

Figure 1. Fluorescent Tube Structure

HOT CATHODE FLUORESCENT TUBES

In an HCFT, the cathode is usually a tungsten coil (sometimes called a filament) coated with an alkaline oxide material (Figures 2 and 3). Electron emission occurs when the filament is heated to approximately 1,000°C. This is sometimes referred to as electrons 'boiling' off the emitter. When the coated coil is heated by the operating current flowing into the tube, thermionic emission occurs from a 'hot-spot' on the coil. A high current density can be produced from just a small hot-spot on the emitter. During operation of the tube, this hot-spot will move along the surface of the coil boiling off electrons.

When the operating temperature on the coil cannot be reached from arc-current heating alone, continuous supplemental resistive heating will be required. Although less efficient than using only arc-current heating, it does offer a wide range of operational currents with good life.

Arc-current is only used at relatively high tube currents. At lower tube currents, using only arc-current heating, the hot spot becomes too cold, the cathode sputters away, and the tube life is substantially reduced. This is caused by a cathode structure suited for hot-cathode operation undergoing secondary emissions. Under proper operating conditions, the life of a thermionic emitter should be proportional to the amount of emitter coating held by the cathode coil.

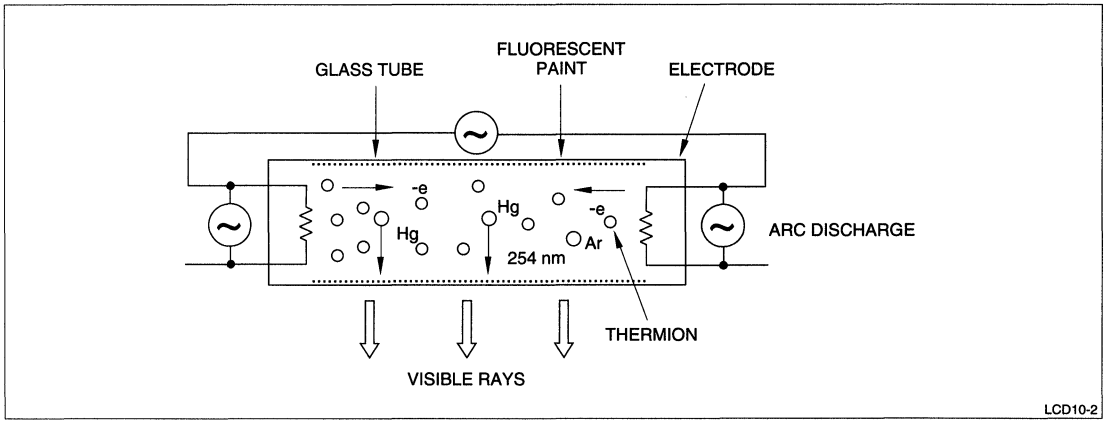


Figure 2. Hot Cathode Fluorescent Tube (HCFT)

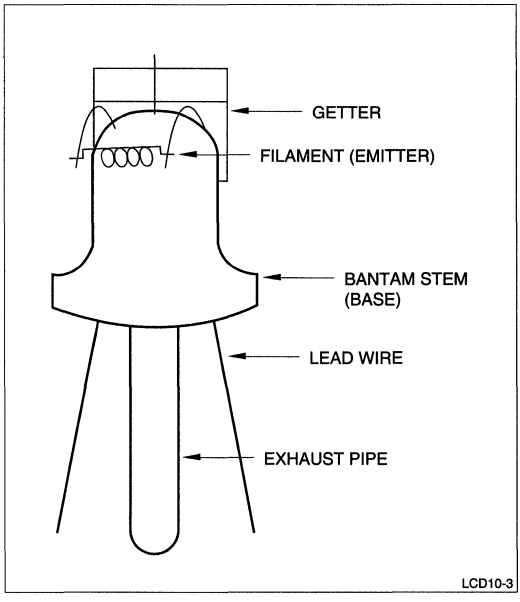


Figure 3. Electrode Part of Hot Cathode Fluorescent Tube

COLD CATHODE FLUORESCENT TUBES

The secondary emissions of a cold cathode have a significantly lower cathode current density than thermionic emissions. During this 'glow-discharge' mode the positive mercury ions are accelerated towards the negative cathode emitter. When these ions bombard the emitter, electrons are released into the arc stream. The cathode is usually constructed as a metal cylinder with an alkaline oxide emitter coating on its inside surface with the closed end of the cylinder facing away from the arc (Figures 4 and 5).

Secondary emissions require a larger emitting surface compared to thermionic emissions, with the actual surface area increasing proportionally with the tube operating current. This increase in area may result in larger 'dead' space at the ends of the tube. The structure of CCFTs make them ideal for long life and tolerant of short duty cycles.

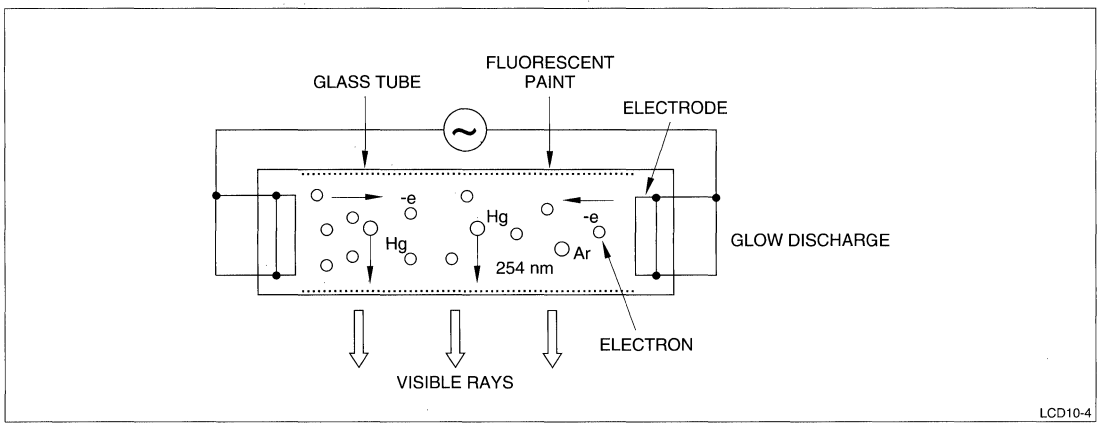


Figure 4. Cold Cathode Fluorescent Tube (CCFT)

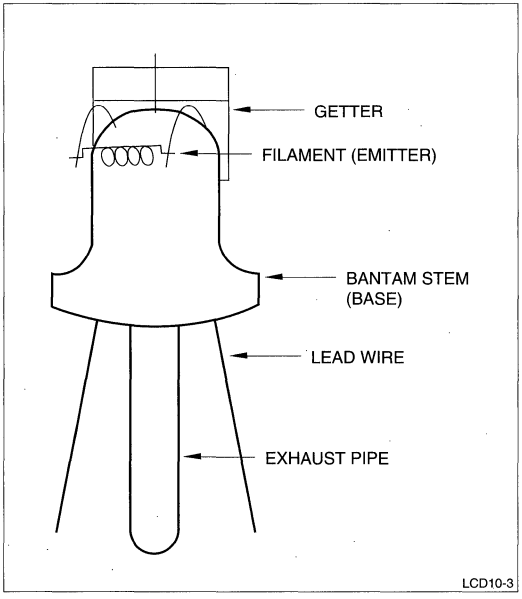


Figure 5. Electrode Part of Cold Cathode Fluorescent Tube

A comparison of HCFTs and CCFTs is shown in Table 2.

Table 2.
Hot Cathode (HCFT) vs. Cold Cathode (CCFT)

PARAMETER	HCFT	CCFT
Emission Type	Thermionic	Secondary
Starting Voltage (Vrms)	360	1,000
Filament Voltage (Vrms)	7.5 [1]	N/A
Filament Current (mArms)	105	N/A
Lamp Voltage (Vrms)	61	455
Lamp Current (mArms)	100	7.0
Frequency (kHz)	40	40
Power (W)	6.1	3.2
Life (Hours)	5,000	15,000

NOTES:
All electrical parameters are relative typical values.
1. Filament preheating time 1.5 seconds.

LUMINANCE

The light output of a fluorescent tube depends on the current density. As current density increases, it is less efficient in producing UV radiation and thus the light output does not increase linearly with operating current. Because light output is related to current density and not just operating current, the tube diameter is an important parameter in determining the maximum luminance output of the tube. Luminance is also affected by temperature (Figure 6). As the temperature decreases, the light output also decreases. This is a function of the mercury vapor pressure which is controlled by the coolest point on the tube wall. The temperature on the tube wall is directly affected by changes in the ambient temperature.

TEMPERATURE

As mentioned, light output of a fluorescent tube is affected by temperature. Discharge voltage is also related to temperature (Figure 7). Colder temperatures require a much higher voltage level to start cathode emissions. The life of a fluorescent tube is not directly affected by the ambient temperature during steady state operation.

LIFE

The life of a fluorescent tube is usually defined as the minimum time it takes the luminance of the tube to reduce to 50% of the original level. For HCFTs the average life is 3,000-5,000 hours and for CCFTs the average life is 10,000-20,000 hours. Life is also affected by tube diameter and length. Larger cathode areas for electron emission will take longer to deplete their coated material. The absolute end of life (where the tube will fail to ignite) occurs when the emission material on either electrode depletes completely. The depletion of emission material occurs during the normal starting and operating cycle of the tube.

DESIGN CONSIDERATIONS

When interfacing the display system with fluorescent tubes, certain parameters must be considered. The most important parameters for CCFTs are: 1) minimum discharge voltage, 2) operating voltage, 3) frequency, and 4) tube current. The minimum discharge voltage is the minimum voltage required to ignite the tube at the tube's specified end-of-life and minimum operating temperature. The operating voltage is the voltage required to drive the tube once the tube has ignited and is operating at a steady

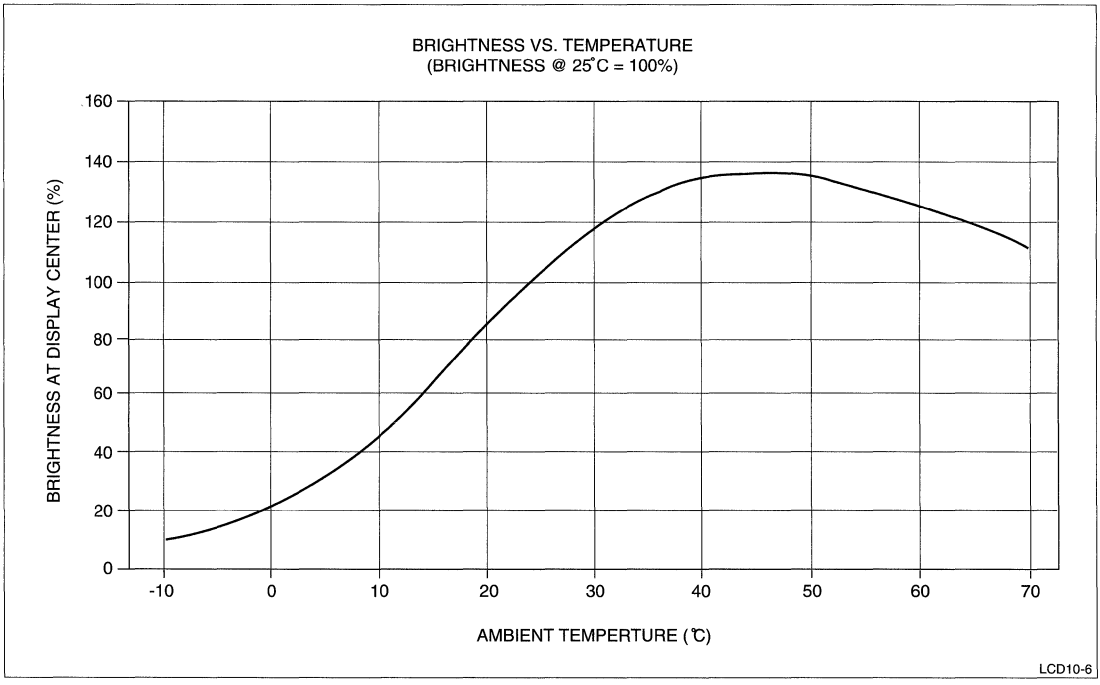


Figure 6. Brightness vs. Temperature

state condition. This voltage is reached after a series 'ballast' capacitor reduces the input impedance, and thus the starting voltage, after the tube is ignited. The frequency of this AC voltage is usually between 20 to 60 kHz and is specified as such to not interfere with the LCD electronics. The tube current is the RMS current passing through the tube. The current directly affects life, luminance, and power consumption of the CCFT. These parameters must be designed to efficiently drive the tubes.

The display structure is important when defining the backlight. Lead cable lengths and metal bezels of an LCD can interact directly with the tube and produce capacitive voltage losses. This needs to be taken into account when designing drive circuitry for CCFTs.

The mercury content of fluorescent tubes may need to be addressed when disposing tubes or operating in a closed environment. Although the mercury content is very negligible, some applications may warrant close scrutiny of this amount.

OTHER BACKLIGHT TECHNOLOGIES

An electroluminescent backlight is composed of EL material sandwiched between a front and rear electrode. When an electric field is applied across the two electrodes, the EL material is excited and visible light is emitted as electrons return to lower energy levels. The advantage of EL backlights is their thinnest, but short life and low luminance output may outweigh this advantage.

The cathodoluminescent backlight operates under a theory of operation similar to a CRT. This type of backlight has extremely high luminance output (up to 13,000 fL) along with an extremely high power requirement.

An incandescent or metal halide light source is combined with a fiber optic link to make another type of backlight. This method has the advantage of having a very thin backlight, without any heat concerns and still produces a very high luminance output. The light source can be placed in a convenient location and the light energy piped over to the display on this fiber optic link.

CONCLUSION

When using different backlight technologies, the operating parameters must be considered for a successful display design. Power supply, operating temperature, and size specifications will all have an effect on the backlight. When all parameters are examined, the wide use of HCFTs and CCFTs in LCDs can easily be understood.

REFERENCES

Information Display, Vol. 5, No. 11, November 1990

Information Display, Vol. 8, No. 2, February 1992

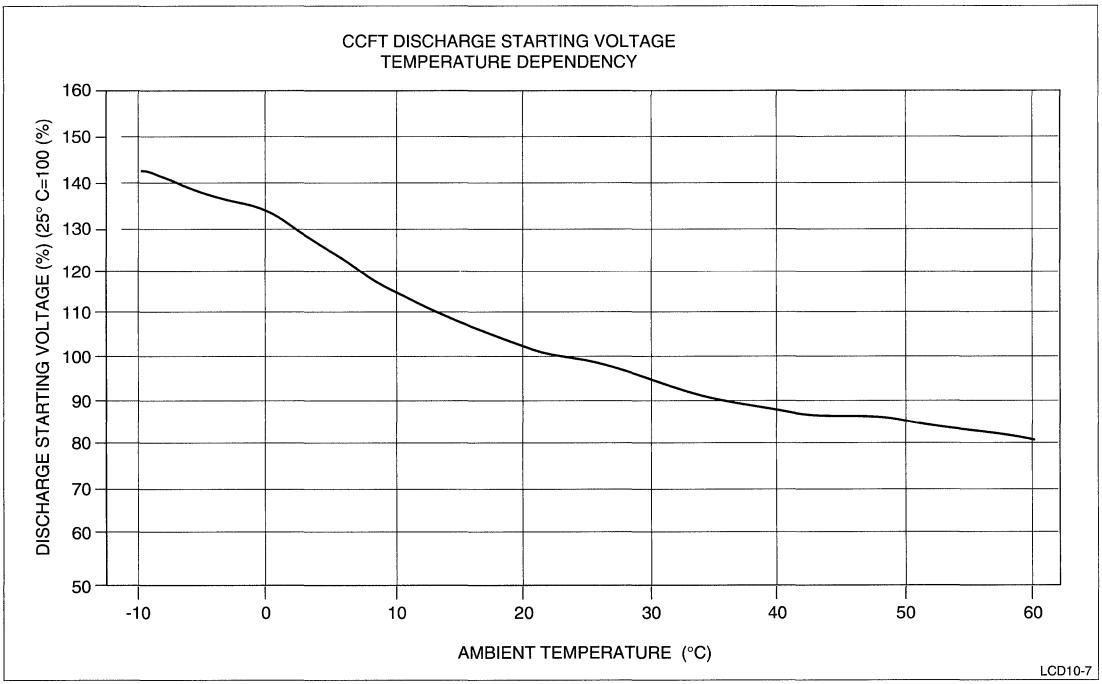


Figure 7. CCFT Discharge Starting Voltage

BACKLIGHTING MANUFACTURERS

Fluorescent Backlights

Flat Candle Company
4165 N. Sinton
Colorado Springs, CO 80907
Phone: 719-260-8088
Fax: 719-260-8089

LCD Lighting, Inc.
11 Cascade Blvd
Milford, CT 06460-9998
Phone: 203-876-1520
Fax: 203-877-7212

Voltarc Technologies, Inc.
186 Linwood Avenue
Fairfield, CT 96430
Phone: 203-255-2633
Fax: 203-259-1194

Electroluminescent Backlighting

BKL, Inc.
421 Feheley Drive
King of Prussia, PA 19406
Phone: 215-277-2910
Fax: 215-277-2956

Durel Corporation
645 W. 24 Street
Tempe, AZ 85282
Phone: 602-731-6200
Fax: 602-731-6210

Eltech
181 Gibralter Road
Horsham, PA 19044
East Sonoma, CA 95467
Phone: 215-441-0404
Fax: 215-441-8299

World Products, Inc.
19654 8th Street
East Sonoma, CA 95467
Phone: 707-996-5201
Fax: 707-996-3380

Other Backlighting Technologies (Metal-Halide, Xenon, etc.)

ILC Technology, Inc.
399 Java Drive
13343 Paxton Street
Sunnyvale, CA 94089
Phone: 408-745-7900
Fax: 408-744-0829

JKL Components Corp.
13343 Paxton Street
Pacoima, CA 91331
Phone: 818-896-0019
Fax: 818-897-3056

Lumitex,Inc.
11941-H Abbey Road
North Royalton, Ohio 44133-9908
Phone: 216-237-5483
Fax: 216-237-5743

FCC CONSIDERATIONS

Dan G. Gutierrez, Field Applications Engineer

ABSTRACT

Flat panel technology in computing devices is our focus here but this paper will apply in most cases to system-level applications. This is due to the fact that one cannot foresee every varying configuration of systems for a particular flat-panel display. The following explains a portion of the emissions concern with FCC, why it is so important, and what is expected of the designer in the beginning stages. The material in this paper is not intended to replace any authorized publications in regards to regulatory emissions and FCC.

INTRODUCTION

Sharp's flat-panel displays are basically considered components. For this reason one cannot say that we can get specific approvals at this level. FCC applies to emissions at the system and peripherals level. Our products are designed with compliance in mind but one cannot expect every kind of measure to be taken at the component level. Sharp's manufacturing is reactive to customers' needs and justifiably most of them can achieve compliant system designs.

STATIC INTERFERENCE IN THE HOME AND OFFICE

FCC has two classifications: Class A or Class B. Class B applies to home or personal use and Class A applies to all others. Radio-frequency interference is the main problem with systems using high-frequency clocks. These signals interfere with neighboring receivers. The bulk of these receivers are found in the home in the form of AM/FM radios, televisions and some two-way and HAM radios. Wireless remote and cellular phones are abundant and can also be susceptible. Both classifications are used to distinguish acceptable signal levels for each environment. The acceptable levels in the home are more stringent. Interestingly enough with the advent of multimedia, the office will become more sensitized to the interference levels. Static interference comes in through the receivers antennae, referred to as radiated interference, or through the power lines as conducted interference.

The source of interference is typically the same for all environments. Nature is certainly a contributor to noise but typically thunderstorms are the most noticeable. Primarily the man-made aspect of interference is experienced and for that reason it is necessary for some regulation through the FCC. Any time electrical energy is present one can expect this interference in the form of electro-magnetic interference or EMI. Automobiles, power tools, appliances, transmitters, office equipment, test equipment, medical equipment, and computers to name a few. If you paint a picture of the modern version of these in your mind you can realize that almost all of these have some form of visual display for the user interaction.

DEFINITIONS

The Federal Communications Committee was created by Congress to regulate emissions and license broadcasting. There are a number of bureaus within the FCC. The Office of Engineering and Technology (OET) deals with authorizations and grants, while the Field Operations Bureau (FOB) enforces regulation. According to a 1987 document noted FCC/OET MP-4 the OET office is at:

Federal Communications Commission
Authorization and Evaluation Division
7435 Oakland Mills Road
Columbia, MD 21046

MP-4 happens to be on RF measurement procedures. Copies of those materials on general rules and regulation can be requested there. The regulation which applies to interference is Title 47 code of federal regulations (CFR) Parts 2 and 15, although there are 42 parts or sections. Part 2 deals with general rules. Part 15 is on RF devices. Each part is divided into subparts. Subpart J applies to all digital devices using 10 KHz or greater clocks which are not regulated within other rules.

Subpart A refers to:

- Class A and B distinctions
- Personal and business computer distinctions

Class A is 'A computing device that is marketed for use in a commercial, industrial or business environment; exclusive of a device which is marketed for use by the general public, or which is intended to be used in the

home.' which requires verification. Verification involves the manufacturer insurance of compliance without any need to submit data or samples. They are required to have test data though on file at the manufacturers side. The limits are:

Class A Radiation Limits

FREQUENCY (MHz)	DISTANCE (m)	FIELD STRENGTH (μV/m)
30-88	30	30
88-216	30	50
216-1000	30	70

NOTE: Class A measurements can be made at any distance between 3 and 30 m. The test results can be scaled inversely with the distance. For example at 3 m the allowable field strength in the frequency range 30 to 88 Mhz is 300 μV/m.

Class A Conducted (Power Line) Limits

FREQUENCY (MHz)	MAXIMUM VOLTAGE (μV)
0.45-1.6	1000
1.6-30	3000

Class B is 'A computing device that is marketed for use in a residential environment notwithstanding use in commercial, business and industrial environment.' These are devices which require certification. Peripherals which fit the description for a Class B computing device also require certification. Certification involves submitting test configurations and data to the FCC by the applicant. This is a much more difficult process than Class A because of this and its more stringent signal limitations:

Class B Radiation Limits

FREQUENCY (MHz)	DISTANCE (m)	FIELD STRENGTH (μV/m)
30-88	3	100
88-216	3	150
216-1000	3	200

Class B Conducted (Power Line) Limits

From 0.45 to 30 MHz the maximum voltage fed back to the line at any frequency must be less than 250 μV.

There are conditions which help determine if the application is business or personal (Class A or B). For Class B all three of the following should apply:

1. If it's available through retail or mail order catalog to the general public (not niche catalogs that cater to specific commercial applications).
2. Advertisements directed at the general public via magazines or marketing brochures rather than specific commercial users.
3. Operates on battery power or house current.

One must demonstrate to the commission that because of price or performance the system could not be applicable in the home or for personal usage.

APPLICATIONS AND COMPLIANCE

As one can see, both radiated and conducted types of EMI are potentially disruptive to broadcast signals at the receiver. Compliance to FCC emissions can also benefit the end product by allowing immunity at internal signal lines to external noise. One now has to consider on how to comply to FCC limits. The limits described above are maximum limits. The designer and manufacturer must allow for some margin of manufacturing tolerances. Typically most will allow about 6 dB minimum below the given limits for both classifications. This is usually enough margin for most but one must take caution with ones own manufacturing tolerance capability. The equipment needed to do the testing is quite expensive. These will be described in more detail later. Most manufacturers contract out to what are called EMI test sites and labs. The availability of these test facilities depends on the location. They are not generally hard to find. It's best to use an FCC authorized site to allow for transferability of test data to the commission.

Flat panel displays affect system design at the video level when attempting to pass FCC. Typically the video controller generates clocks which range between 50 Hz and 20 MHz for monochrome LCD and EL displays. Generally, the video clocks encompass the higher part of that range (10 to 20 MHz). In color LCD displays, the frequency of video clocks is about 28 MHz at the high end. These are the areas where one must concentrate to alleviate 'hot spots' in the FCC range of frequencies. Most of the applications in monochrome are easily passed. It is in color displays where the designer is challenged in resolving for a compliant system.

Incompatible shielding practices can cause problems. If a system's main enclosure is shielded well and the display bezel is left unshielded, conducted emissions from the system can be radiated through the display housing. Refer to Figure 1.

Some typical measures one would take for attenuating EMI would be:

- Board Layout Practices – Here the video graphics control board is laid out with compliance in mind. Close attention to ground paths and plains is critical. A long ground path can vary in impedance at different frequencies. Additionally output impedance's of the video clock lines may require some damping by way of ferrite beads or capacitors to smooth out any spikes. Very sharp rise and fall times can cause amplified EMI 'hot spots.'

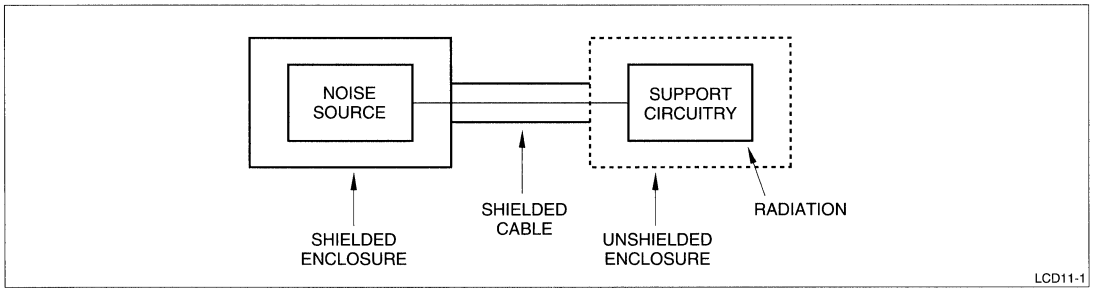


Figure 1. Incompatible Shielding

- Cabling – The shortest path to the display is recommended. Instances where this is not possible one must enhance shielding or experiment with different grounding points. Add ferrite cores to filter out noise. These can prove to be very effective counter measures.
- Metal Enclosures – Full metal enclosures or conformal coating of plastic enclosures are recommended for stubborn emissions.
- Separate Interconnect Boards – This facilitates flexibility in adding measures to reduce emissions without having to do new board layouts or bulky cabling.

TEST FACILITIES AND EQUIPMENT

There are two type of testing facilities. An open field test site is preferred for final determination of level compliance. These require an open field away from any power lines or metal structures. A baseline background noise is established to make any corrections to the data. In some cases the equipment under test (EUT) is laid out in the open on some wooden platform with appropriate power connections and monitoring cables. These are routed to a lab environment at some distance so as not to cause any skew in the testing. The platform is set 30 meters from dipole antenna for class A, 3 meters for class B.

There are five main items which are maximized for worst case readings.

1. The EUT must be rotated as in a turntable to maximize the signals received back at the lab area. Usually these tables are remotely controlled for rotation to allow expedience.
2. Cable placement is optimized for worst case.
3. Configuration of the system is optimized by attaching cables and or peripherals on all I/O connectors.
4. Worst case software is run to peak the overall speed of all the system clocks. At the display side a full screen of scrolling H's is a good test.
5. Receiving antenna orientation horizontal, vertical, and height is adjusted for worst case.

An enclosed shielded lab environment can be used to do EMI testing but it's only used for determining 'hot spots.' One should realize that frequency profiles in an enclosed lab environment can hide some unexpected problem areas. It's advisable not to make any manufacturing decisions based on enclosed lab findings. The final decisions should always come from an open test site. One could be fighting an apparent problem area in an enclosed lab test and not know that in actuality the problem is non-existent in the real test. This could cost valuable engineering time for both the manufacturer and the components vendors.

Conducted tests are done on both the hot and neutral lines of the system power cord. Readings are taken separately for both. The center of the test equipment is a line impedance stabilizer network (LISN). This 'box' provides uniform impedance to the equipment under test (EUT). Signals are tapped off of this box on to a spectrum analyzer to measure the range of 450 KHz to 30 MHz. Class A measurements break it down in two ranges. For 450 to 1600 KHz the readings should not exceed 1000 μV. From 1600 KHz to 30 MHz it should not exceed 3000 μV. Much lower voltage levels are allowed for line interference in the home or Class B. Throughout the entire range the reading should not exceed 250 μV.

MANUFACTURING CHANGE EFFECTS

Whenever changes are made in the manufacturing of the system or its components several rules will apply.

For certified equipment such as for Class B one can refer to FCC article 2.1043. Simply, a change which has lower emissions requires little or no action by the manufacturer. On the other hand if one changes the product and it reduces the amount of margin (even if still under the FCC limits) this will require retesting and submittal of new test data. For this reason one must submit the worst case manufactured capability of the product at FCC initial

submittal. This can help the compliance engineer when doing audits during the life of the product.

For verified equipment or Class A one can refer to the 'Public Notice of April 7,1982' document from FCC. In this case typically the manufacturer can simply retest and file his findings internally. Again no application is needed with the FCC.

LABELING FOR COMPLIANT SYSTEMS

Labeling is defined explicitly by the FCC. Specific wording must appear on the compliant system. FCC Class B labeling should appear such as:

> – Grantee Name Here –
>
> FCC ID: XXXYYY1234
>
> Certified to comply with Class B limits, Part 15 of FCC Rules. See instructions if interference to radio reception is suspected.
>
> Made in . . .

FCC Class A labeling requirements should appear such as:

> This equipment complies with the requirements in Part 15 of FCC Rules for a Class A computing device. Operation of this equipment in a residential area may cause unacceptable interference to radio and TV reception requiring the operator to take whatever steps are necessary to correct the inter-

In addition to labels one must have an insert to the systems instruction manual which has specific actions to be taken by the user to remedy the interference. Also precautions are given when changing out system components such as power cords or peripherals. One can obtain this wording directly from the FCC Part 15, subpart J Rules.

CONCLUSIONS

FCC rules are extremely detailed and can suffer from interpretation. One would expect to see more claims with FCC in regards to testing and misuse of the rules. Surprisingly most of the claims are related to labeling of the products. In the history of emissions claims only during the hey day of CBs did the amount of claims rise. Notebook computers have been abundant and no news of increased claims have been apparent. This just shows that manufacturers have not had difficulty in complying specially with the use of flat panels rather than the obtrusive CRT.

REFERENCES

[1] *EMI for EDP, Telecom, and Medical Devices*, A Seminar, by Dash, Straus and Goodhue, Inc., Compliance Engineering

[2] *FCC Procedure for Measuring RF Emissions from Computing Devices*, FCC/OET MP-4 (1987)

[3] FCC, Code of Federal Regulations, Article 47, Parts 2 and 15 subpart J.

CCD FOR AUTOMOTIVE USE WITH COLOR LCD DISPLAY

Chuck Bohac, Field Applications Engineer

INTRODUCTION

With the decreasing costs of color Liquid Crystal Displays (LCDs), increased operating temperature range, improved ruggedness, and obvious small size and weight, the color LCD is finding its way into automotive applications. One application for its use is in rear view imaging. The integration of an LCD with a Charge Coupled Device (CCD) camera can be used to view a very large area behind (and directly below) the rear of a vehicle. The popularity of mini vans and utility vehicles is on the rise. These vehicles have more 'blind spots' than other conventional automobiles. With the newer automotive quality LCDs, additional rearward viewing can improve the safety of these vehicles and reduce the number of accidents due to limited viewing before and during rear drive.

This application note describes one technique used to design an interface between the SHARP LQ6NC02 Liquid Crystal Display and a color CCD camera with an automotive electrical system.

SYSTEM REQUIREMENTS

1. SHARP LQ6NC02. (Chosen for its 12 o'clock viewing angle because the viewer will be looking down on the display.)
   NOTE: The 'NC' version requires an NTSC standard interface.
2. A wide-angle color CCD camera that operates from +9 VDC and produces an NTSC-compatible output.
3. Operation from a standard +12 V automotive electrical system.
4. Input voltage protected for voltages up to +20 V, in the event two auto batteries are connected in a series during jump starting.
5. Limiting the distance between the CCD camera and the display at 25 feet or less.

DESIGN REQUIREMENTS (SUMMARY)

1. Convert the input voltage (+12 V to +18 V) into the required voltages to operate both the display and the camera.
2. Provide sufficient input protection. If the input voltage exceeds +20 V, the unit is protected.
3. Provide for adequate power on/off sequencing to prevent the LCD from degrading.
4. Provide Contrast, Brightness, Color, and Tint adjustments.
5. Ideally, the printed circuit board is mounted on the back of the LCD.

DESIGN (DETAILED)

1. Convert the input voltage (+12 V to +18 V) into the required voltages to operate both the display and the camera.

   The maximum voltage and current requirements for both the LCD and the CCD during normal operation over temperature are shown in Table 1.

   Voltage Regulators: SHARP's PQ05RF1, PQ09RF11, and PQ12RF11 voltage regulators are used for two reasons. First, they have a low drop-out voltage (0.2 V max.). Secondly, they have TTL controlled inputs. This is important because the LCD has a power-up sequence requirement (see Design Requirement #3). TTL control of the +5, +12, and –8 voltages is essential. Also, by using a TTL controlled device, power-up sequencing becomes very easy.

   DC-to-DC Converter: A Computer Products DC-to-DC converter is used to invert the +12 V input to +12 V and -12 V output. The –12 V output is needed to input to the –8 V regulator. An NTE932E negative voltage regulator is used. (See Figure 1.)

Table 1. Voltage and Current Requirements

PARAMETER	VOLTAGE		CURRENT
	MIN.	MAX.	MAX.
LCD Positive Supply	+4.8	+5.2	155
LCD Negative Supply	–7.6	–8.4	120
BackLight	+11.0	+13.0	560
CCD Camera	+8.8	+9.2	500

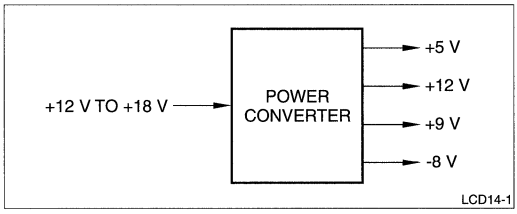


Figure 1. Block Diagram of Power Converter

2. Provide sufficient input protection. If the input voltage exceeds +20 V, the unit is protected.

   The system inputs are protected in the event that two batteries are accidentally connected in series (i.e., jumping a car). Therefore, Zener Diode D1 is selected for 20 V and placed at the power input of the board. If the input voltage exceeds +20 V, the Zener breaks down, shorts the input to ground and blows the inline fuse. Therefore, the board, display, and camera is protected (Figure 2).

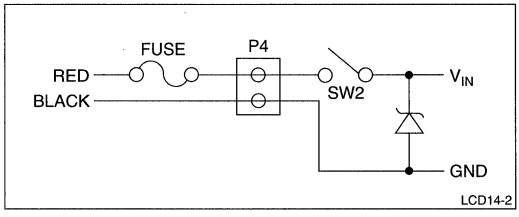


Figure 2. Schematic

3. Provide for adequate power on/off sequencing to prevent the LCD from degrading.

   See Figure 2 for a circuit that provides the proper power-up sequencing for the LCD.

4. Provide Contrast, Brightness, Color, and Tint adjustments.

   Contrast, brightness, color, and tint potentiometers have been added to enhance the adjustability of each respective parameter. The 4-position dip switch (S1) is used to enable or disable the effects of the potentiometers.

5. Ideally, the printed circuit board is mounted on the back of the LCD.

   The printed circuit board (PCB) was laid out to bolt directly to the back of the LCD using standoffs (although there is no requirement that the PCB be mounted directly behind the LCD). The PCB could quite possibly be mounted in some other location in the vehicle with the flat ribbon cable connecting the two. It should be noted that if a long (+12") flat ribbon cable is used, it should be shielded for EMI (both emissions and susceptibility).

CONNECTORS

P1 The connector on the LCD is an XHP-3 produced by the Japan Solderless Terminal Co. (JST). A similar connector is required for the PC board. An FPC cable is used to interconnect between the display and the circuit board. Additional suppliers for the cable as well as the connector include Molex and Raintree.

P2 This two pin connector spacing is the standard 0.1". Any standard Berg type connector can be used to mate to the inverter cable.

P3 This four pin connector spacing is the standard 0.1". Any standard Berg type connector can be used to mate to the CCD camera.

P4 This connector mates the fuseable link (Red, Vin (+13.8V)) to the PCB. The Black wire is connected to automotive ground.

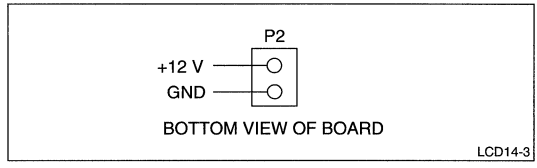


Figure 3. P2 Connector

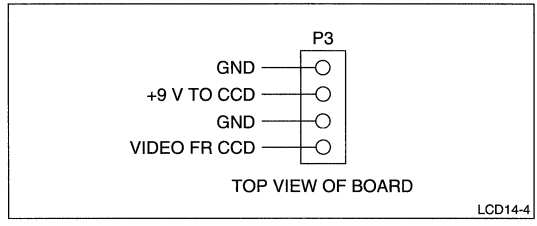


Figure 4. P3 Connector

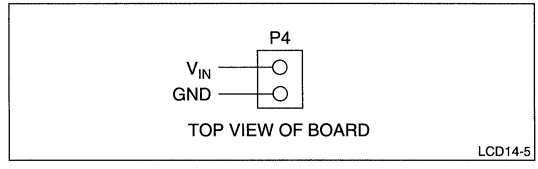


Figure 5. P4 Connector

ELECTROMAGNETIC COMPATIBILITY

Brian Graham, Field Applications Engineer

INTRODUCTION

Flat panel displays are becoming more prevalent in today's digital electronics. The displays are operating at higher frequencies due to larger data formats and increased color capabilities. With higher operating frequencies, electromagnetic compatibility (EMC) becomes increasingly important. EMC should be approached at the system level in the design phase in order to ensure the most economical and time efficient design. Unexpected noise problems can lead to expensive time consuming solutions which are not optimized for the design. This application note is a discussion of EMC as it relates to flat panel technology.

FLAT PANEL DISPLAYS AND EMC

The clock signal is generally the highest frequency signal in a digital system and the cause of most EMC problems. Flat panel displays have three clocks associated with them. The names of the clock signals vary from specification to specification but they are similar in operation. The three clocks are: frame clock, latch clock and data clock. The data clock is the highest frequency of the three clocks. The data clock is used to load data into shift registers in the display. The data clock for color displays is more than three times faster than for monochrome displays. Therefore color displays tend to be noisier than monochrome displays. The latch clock is used to transfer the data to latch registers that energize the display pixels. This is the next highest speed clock. The latch clock is at a significantly lower frequency than data clock but should still be considered a noise source due to the harmonics it presents to the system. The frame clock is the slowest of all three clocks and generally is in the frequency range between 40 Hz and 80 Hz. This is the least likely of the display clocks to cause a problem.

The data lines for a display can also cause EMC emissions since they are being manipulated at the same frequency as the data clock. Some displayed screen patterns emit more radiated noise than others due to the increased activity on the data lines.

Drive frequencies for several of Sharp's displays are listed in Table 1. The data clock frequencies in Table 1 are close to the frequencies where FCC regulations become applicable. This is the reason special attention should be given to the display interface.

Table 1. Clock Frequencies

DISPLAY	F_{DATA}	F_{LATCH}	FFRAME	RISE TIME MAX
LQ9D011	25 MHz	31.5 kHz	60 Hz	13 ns
LM64C08	3.4 MHz	14 kHz	60 Hz	13 ns
LM64P839	2.3 MHz	14 kHz	60 Hz	50 ns

EMC

There are two types of compatibility that must be taken into account when discussing EMC: noise susceptibility and noise generation. Noise susceptibility refers to the product interfering with itself or being susceptible to other noise generators in its vicinity. Noise generation refers to the capability of the product to interfere with other electronic devices located in its vicinity. Noise susceptibility is generally taken care of in the design phase due to its ability to keep the design from working. Noise generation, however, generally gets ignored until testing begins. The proper use of system design will avoid problems with both susceptibility and generation.

Radiated noise can be classified into two categories: differential and common mode. The following is a discussion of differential and common mode noise.

Differential Mode Noise

Differential mode noise is magnetic in nature and can be represented by the circuit diagram in Figure 1. It can be modeled as an inductive element that is transformer coupled to a low impedance noise receiver. Differential mode noise occurs where current is flowing around a loop formed within an electronic circuit. The loop is formed by the signal and return path of the current. The magnetic field is formed by the fluctuating current flowing in the loop. The field has the shape of a toroid and is maximized around the edges of the loop.

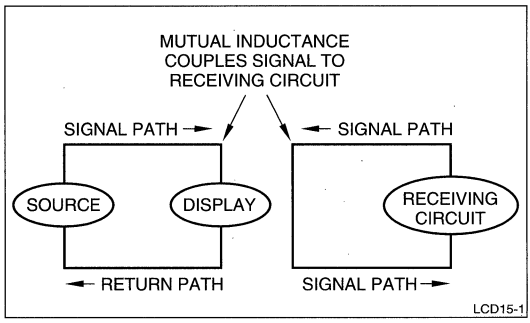


Figure 1. Differential Mode Noise

Current carrying loops that have perimeters less than 1/4 wavelength are very effective radiators. This is because the current within the loop is in phase. larger loops that have perimeter greater than 1/4 wave length are less effective radiators because the current is out of phase with itself. In the far field the electric field due to a current carrying loop can be defined as:

$$E = 263 \times 10^{-16}(f^2AI)(1/r)$$

E is in volts/meter, f is frequency, A is the area of the loop, I is the current in the loop, and r is the distance from the loop. This equation has been corrected for ground reflection. This equation is for a small loops with a perimeters of less than 1/4 wavelength but can be used to estimate field strength for larger loops.

The designer has control over three of the parameters in this equation. The loop area can be minimized, the current can be minimized or the frequency content of the signal can be minimized. Note that the frequency component is squared. If the harmonic content of the signals can be reduced, the largest reduction in differential noise can be accomplished.

The key to controlling differential mode noise is by using proper circuit layout techniques. Differential mode noise is a function of current loop area, as seen in equation 1. Minimizing loop area is the starting point for reducing differential noise. The total system should be analyzed to determine signal and return paths. This will give a clear understanding where current loops exist.

The critical lines where frequency is the highest in the LCD display are the clock lines. The data lines can also cause problems due to the high speed data transfer taking place. Within the display these loops are minimized as much as possible through the use of ground planes. The designer has control over the interface cable extending between the display and the system board. This is where a reduction in loop area can easily be accomplished using proper cable construction. Cable construction is listed below in the order of preference:

- Coax or simulated coax is best. Simulated coax is a flex cable with a ground plane. The shield for the coax or the ground plane for the simulated coax should be grounded at both ends.
- Tri-Lead wire. This is wire that surrounds the clock lead on either side by a ground wire.
- Twisted pair. Each signal wire is twisted with a return wire.
- Flat cable or ribbon cable where each signal wire is next to a ground wire.

The cabling should be secured in place to ensure identical placement of the cable in production. Securing the cable also eliminates the possibility of low frequency fields occurring during mechanical vibration due to the movement of the current carrying conductors.

The frequency and magnitude of the current around the loops should be minimized where possible in the system. These parameters, however, are only minimally controllable within the drive and timing specification of the display.

Common Mode Noise

Common mode noise is capacitive in nature and can be represented by the circuit diagram in Figure 2. A circuit board or cable forms one plate of the capacitance and the second plate is ground. As the circuit card varies in potential with respect to ground, a current flows in the capacitor which in turn generates an electric field. The circuit or cable acts as a monopole antenna. The following equation describes the electric field emitted by a common mode source in the far field:

$$E = 12.6 \times 10^{-7}(fL\bar{I})(1/r)$$

E is in volts/meter, f is frequency, L is the length of the antenna, $\bar{I}$ is the common mode current, and r is the distance from the antenna. The designer has control over frequency, the length of the antenna, and the common mode current flow.

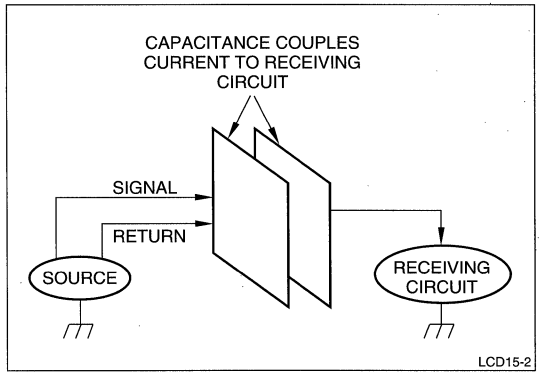


Figure 2. Common Mode Noise

Common mode noise is generally harder to predict and contain than differential mode noise. This is because common mode currents are not part of the circuit operation and are not easily defined like differential mode currents. Common mode currents are also more effective radiators than differential mode currents. A common mode current of a few micro-amps can radiate as much as a differential mode current of several milliamps.

Investigating common mode noise should begin with the cables that exit the unit. Cables connecting peripherals and even the display will act as good common mode antennas. In order to control noise emission from a common mode noise source, the common mode current must be minimized. The initial step in minimizing common mode current is determining the current path and its magnitude. Determining the path and magnitude can be accomplished using a high frequency current probe. The high frequency current probe is placed around the signal and return lines in question and If a current is detected it is the common mode current.

When the current path has been identified, there are several methods that can be used to control the common mode current. These methods are:

- Raise the impedance seen by the common mode current.
- Minimize the common mode voltage

Raising the impedance seen by the common mode current is probably the most effective method for controlling common mode noise. This is accomplished by adding a common mode choke in series with the signal and return lines in question. A common mode choke is made of ferrous material shaped into a toriod. Most transformer core manufacturers produce common mode chokes.

The most effective application of a common mode choke occurs when all current paths are encompassed by the choke. This indicates that if shielding is used, then the ground for the shield should also run through the common mode choke. In order for a common mode chock to be effective, the impedance must be greater than 100 Ω at the frequency of interest. Most common mode chocks will supply less than 20 db of impedance so, in some cases, it may be necessary to implement more than one fix. The designer should plan room for a common mode choke around the display drive signal cable so that it may be incorporated if necessary. Some connector companies have available connectors that incorporate a common mode chokes.

Minimizing the common mode voltage can be accomplished by minimizing voltage transitions, using grounding techniques and shielding techniques. Minimizing common mode voltage can be accomplished by using 3 V logic in place of 5 V logic. Some of Sharp's newer panels incorporate 3 V logic. Implementing a ground plane for differential mode noise can also decreases common mode voltage. The ground connection from interface cables should also be kept as close to the external ground, (chassis ground), as possible to avoid potential differences.

SHIELDING

Shielding is another way to reduce common mode and differential mode noise. Noise radiation has different characteristics as the distance from the noise source increases. Near a noise source, the transmitted field can be characterized as either low impedance consisting of mostly current, or high impedance consisting mostly of a voltage wave form. This is near field noise. As the distance from the noise source increases the impedance converges to the characteristic impedance of the medium, for air approximately 377 Ω. This is far field noise. The dividing point occurs at approximately 1/6 wavelength. Shielding techniques for far field differ from near field shielding. There are two methods by which shielding eliminates radiation: absorption and reflection. In the near field, magnetic fields are primarily effected by absorption and electrical fields are effected by reflection. In the far field, absorption is the primary method of loss for both magnetic and electric fields. FCC testing is considered far field high frequency. An effective shield for far field high frequency purposes can be made form copper, steel, or aluminum. Conductive coatings can also be used if their resistivities are low enough – a few ohms per square. Some ESD coatings are not effective as shields due to their high resistivity. Thickness of the shield material also effects the ability of the shield to absorb radiated noise.

Shielding effectiveness also depends on shield integrity – how many apertures there are. The aperture size can have more effect on shield effectiveness than the shield material. The area of the opening is less important than the maximum linear dimension. Therefore the linear dimensions of openings in the shield should be kept as small as possible.

The shield does not have to be grounded to be effective. However, to give the RF current a path to ground and keep the shields potential from varying, grounding the shield is a good design practice.

CONTRIBUTING FACTORS

Other factors that may contribute to the radiating frequencies are the rise time of the clock signals and the reflection characteristics of the drive line. Drive line reflection is caused by impedance mismatches between the driving and the receiving circuitry. Drive line reflections can cause high frequency ringing on either the voltage or current wave form of a signal line. Drive line reflections are a function of the source impedance, the receiving impedance, and the line impedance. The parameters under control of the designer are the driver and line impudences. The designer should plan to compensate at

the driver because the cable length is generally not known until late in the design.

Signal rise time can also contribute to noise problems. The rise time of the clock signal is a determining factor in the spectrum of noise radiated. Figure 3 illustrates the noise spectrum of a clock signal. The spectrum of the noise increases from the fundamental of the clock frequency at 20 db/dec until a frequency determined by the rise time where the noise flattens. The break frequency can be determined by the following formula:

$F = 1/\Pi x t_{(RISE)}$

Therefore the rise time of signals should be maximized where possible to ensure the lowest breakpoint in the spectrum.

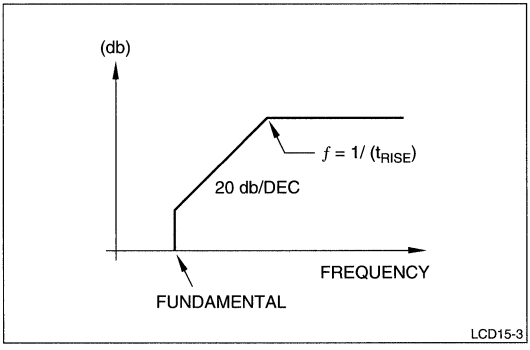


Figure 3. Noise Spectrum of a Clock Signal

THE FCC

The governing organization in the United States for EMC is the FCC. Digital devices are covered under FCC regulations, part 15, subpart J. Subpart J governs both emissions and power line conducted noise. There are two classifications for equipment: class A and class B. Class A defines the acceptable level for radiated noise in equipment intended for use in the industrial environment. Class B defines the acceptable level for radiated noise in equipment intended for use in the residential environment. Class B is much more stringent than class A. Most computer systems must conform to class B requirements. Table 2 describes the acceptable noise levels for class A and B equipment. Prior to testing, you should contact the FCC to obtain a copy of part 15, subpart J so that you have detailed test information.

Table 3. FCC Conducted Emission

FREQUENCY (MHz)	CLASS A µV	CLASS B µV
0.45-1.6	1000	250
1.6-30	3000	250

CONCLUSION

LCD displays are becoming a prevalent technology. The clock speeds have increased with the demand for higher color content and increased data formats. The system designer is facing a bigger challenge to meet FCC regulations. System design must take into account EMC from the initial phase of development in order to meet cost and schedule. EMC requirements can only be met if the designer understands the techniques to control differential mode and common mode noise. These techniques are:

- Limit differential mode noise by minimizing current loop area within the system. This can be accomplished by using proper cabling techniques and ground planes.
- Limit common mode noise by using common mode chokes on cables exiting the unit and reduce common mode voltages by using 3 V logic and ground planes.
- Minimize reflection on display drive lines.
- Maximize clock signal rise time.
- Incorporating shielding as necessary.

REFERENCES

Noise Reduction Techniques in Electronic Systems, Second Edition, by Henry W, Ott. John Wyley and Sons, New York, 1988.

Digital Design with Standard MSI and LSI, Second Edition, by Thomas Blakeslee. John Wyley and Sons, New York, 1979.

Controlling Radiated Emission by Design, by Michael Mardiguian. Van Nostrand Reinhold, New York, 1992.

Electromagnetics, Second Edition, by John Kruas and Keith Carver. McGraw-Hill Inc., New York, 1973.

Table 2. FCC Radiated Emission

FREQUENCY (MHz)	CLASS A		CLASS B	
	MEASURING DISTANCE	FIELD STRENGTH	MEASURING DISTANCE	FIELD STRENGTH
30-88	30 m	30 µV/m	3 m	100 µV/m
88-216	30 m	50 µV/m	3 m	150 µV/m
216-1000	30 m	70 µV/m	3 m	200 µV/m

RF COMPONENTS

TELEVISION RECEIVE-ONLY APPLICATIONS

Robert Stuart, RF Applications Engineer

INTRODUCTION

Television Receive-Only (TVRO) refers to the reception of signals from a satellite, without capability or intention of transmitting signals back to the satellite. Typical applications are the satellite dish in a neighbor's backyard, and those used by Cable Television companies to receive program material. Very Small Aperture Terminal (VSAT) applications differ in that they have the capability to transmit back to the satellite, as well as receive information. Their typical application is for insurance sales offices, major automobile dealerships, chain stores, and others who want to stay in contact by way of an immediate medium.

TVRO encompasses the satellite television receiver industry as well as commercial applications of satellite technology for data purposes. Many industrial companies use satellites as a convenient medium for transmitting data to a large number of receive sites at the same time. Due to their position above the earth, most communication satellites are constantly in view to a large portion of the earth's surface. Some versions of satellite receive dishes continuously observe a particular satellite. Other versions have the ability to adjust their position to view different satellites. This provides for continuous reception capability with a large number of possible signal sources from which to choose.

The Cable Television industry has actively used satellite transmission capability since the mid-1970's to provide television broadcast signals to a wide area of coverage. Home Box Office® (HBO) network was the first movie channel to uplink to satellite for distribution of its signals. At the same time, the young satellite receiver industry got a boost as satellite hobbyists were able to begin offering systems to others, because there were now more signals to capture.

OVERVIEW

TVRO covers a broad range of equipment and system applications. Specific equipment is manufactured for TVRO applications, such as Low-Noise Block (LNB) downconverters, tuners, receive antennas (dishes), and other specific items. This discussion will outline the basic TVRO environment, introduce some of the equipment used, and explain a few of the particular factors governing satellite television reception. For more in-depth reading, refer to the references.

SATELLITES

The term "satellites" in this context refers to the man-made communication devices placed in orbit above the earth. Most of these satellites are solar-powered, and have fixed life expectancies. Functional lifetime is usually dependent on how long the satellite can be expected to withstand the harsh environment of space. Satellites are continually bombarded by solar and cosmic radiation, micrometeorites, and other particles. The materials that are used in the manufacture of the satellite will eventually breakdown, failing to operate, or failing to protect a vital component, leading to breakdown of the entire system.

Companies that operate satellites plan on replacing them at an interval slightly shorter than the expected lifetime of the unit. As an additional provision, they will often launch a spare satellite and place it in a parking orbit, where it will be held until needed. All satellites are provided with some degree of maneuvering capability, so that they maintain precise locations above the earth, or moved to new locations. Satellites are becoming more plentiful as new capabilities are added, and their expanding capabilities attract more users.

Geostationary

Geostationary satellites are placed in orbits above the earth's equator, often referred to as equatorial orbits. This satellite orbit region is also referred to as the Clarke Orbit in honor of science fiction author Arthur C. Clarke, who suggested placing satellites in this orbit in an article that appeared in Wireless World Magazine in 1945. [1] The period of orbit around the earth is timed to exactly match one earth-day. In this position, they appear to be stationary above a fixed point on the surface of the earth, thus called geostationary. A satellite receive antenna that is aimed at a geostationary satellite can be installed and left in place, once properly aligned. The satellites actually wander a bit, but are kept in their assigned positions by ground controllers who continuously monitor their position. The satellites have small maneuvering jets and can be directed to adjust their relative position to the surface of the earth.

Geostationary satellites are being placed in orbits with 2-degree spacing, as measured from the center of the earth. At this spacing, satellites must maintain their orbital positions within ±40 kilometers, or ±0.1 degrees of longitude of their assigned point. [2]

Inclined Orbit

Inclined Orbit satellites do not occupy a true equatorial orbit, but wander North and South of the equator by a small amount. These satellites are at a different distance from the earth than those in equatorial orbits. The total length of the path that they fly also equates to one earth-day, providing a longitudinal fixed reference. The elevation (how high it appears in the sky, relative to the equator) of the satellite as viewed from the earth will increase and decrease during the earth-day. Satellite antennas for use with inclined orbit satellites must have a mechanism to allow for the change in position of the elevation component of the satellite. In addition, some sort of tracking device is required to ensure that the antenna accurately follows the target satellite.

Inclined orbit satellites are coming into greater usage as older satellites are replaced with newer versions, and the old units still have viable capability left. The older satellites are then positioned in inclined orbits and new signals sent to them. This new transmission capacity is not the primary commercial signal medium and thus commands a lower market price. As such, customers that may have not been able to afford transmission costs on the premium satellites, may be able to afford these new services, which are in lower demand. The cost of the earth station to receive the signals is greater due to the requirement to track the movement of the satellite, but the aggregate transmission costs are lower.

Polar Orbit

Polar orbit satellites cross over the polar regions of the earth during their rotational period. This is the only effective manner in which to adequately cover areas of the earth's surface beyond 80 degrees North or South latitude. Polar satellites are often at much lower altitudes than equatorial satellites and may complete one orbit in eighteen hours or less. Receiving antennas for polar systems must be capable of continuously tracking the satellite through the period of time that it is within view.

BROADCAST BANDS

Several frequency bands are used with satellites for the transmission of television broadcast signals. The first television signals were sent on C-band frequencies. Later advances in technology allowed for transition to Ku-band frequencies, bringing smaller components and lower overall costs.

C-Band

Early satellite receiving systems operated at C-band frequencies as technology available at that time used 3.4 GHz to 4.8 GHz frequencies. C-band satellites are still in use and carry a large portion of the total communications and television traffic. The components for C-band reception are physically large and expensive to manufacture. Receive antennas for commercial purposes are often 7 meters or 12 meters in diameter. Component size and power efficiency place a limit on the operational capability packed into one satellite. C-band satellites can experience interference from terrestrial microwave transmissions, depending on the location of the transmit or receive facility.

KU-Band

With improvements in technological capabilities, new satellites are being launched that incorporate Ku-band 11 GHz to 13 GHz frequencies and higher power levels. The advantages of the Ku-band systems are smaller satellite components, smaller earth station components, and earth stations became more portable. Additionally, the Ku-band systems do not experience interference from terrestrial microwave systems as do C-band systems. Ku-band satellites generally carry more capability than a C-band satellite of comparable size.

The penalties for operation at Ku-band frequencies are the atmospheric and environmental effects experienced. Ku-band frequencies are more susceptible to signal fading and reception noise due to snow and rain scatter degradation. C-band wavelengths (corresponding to frequency) range from 3.46 inches at 3.4 GHz to 2.44 inches at 4.8 GHz. Ku-band wavelengths range from 1.06 inches at 11 GHz to 0.905 inches at 13 GHz. Due to the shorter wavelengths of Ku-band frequencies, signals are more susceptible to physical interference by rain and snow flakes. Snowflakes can be as wide as the wavelength of the signal. C-band signals are usually two or three times the physical dimensions of rain or snow accumulations in the atmosphere, and experience less interference by these mediums.

All electromagnetic transmissions experience some fading due to atmospheric absorption, caused by the magnetic dipole moment of the oxygen and water vapor molecules. The difference in attenuation, in dB per kilometer is greater than a factor of 10 from C-band to Ku-band [3], with Ku-band suffering the greater attenuation.

ANTENNAS

Satellite receiving antennas may be one of several types: Flat-plate, Cassegrain, Off-set fed, or prime focus. Prime-focus antennas are the popular parabolic reflectors seen in most applications. They are made of metal or metalized fiberglass material. These are called prime-focus antennas as the LNB, the receiving element, is located at the prime focus of the parabolic reflector. The prime focus is the point where all of the received electro-magnetic energy converges in front of the reflector at a common point. As an analogy, a flashlight bulb is located at the prime focus of its reflector. The flashlight bulb is the source of reflected energy, rather than the receptor (Figure 1).

Off-set fed antennas are a variation of a parabolic antenna with the prime focus off-set from the center of the reflector material. These antennas are more elongated in shape so that the incoming signal energy is not directly blocked by the LNB positioned in front of the reflector. The LNB will be positioned away from the center of the reflector, but the antenna is shaped to focus the energy into the LNB in this position.

Cassegrain antennas mount the LNB at the center of the reflector, and place a smaller hyperbolic reflector at the prime focus of the main reflector, to focus the signal energy from the main reflector into the LNB (Figure 1). Flat-plate antennas are generally either a Fresnel lens variety or a phased-array. These are not as well known and are more expensive in general. With advances in manufacturing technology, phased array antennas may become more common as new, higher-power satellites enter service, as these antennas are generally smaller due to cost. With the higher output power of the new satellites, smaller antennas become efficient and economically viable.

LNBs

LNB is an acronym for Low-Noise Block down-converter. Low-noise refers to the amount of electrical signal noise added by the LNB to a received signal. With lower added noise, the overall signal recovery of the system is improved. Sharp LNBs are manufactured for the Ku satellite broadcast band, in the 10.95 GHz to 12.75 GHz range. An LNB is used to convert the 12 GHz satellite signal to lower frequencies in the 950 - 2,000 MHz range. This conversion reduces the cost of the processing electronics by the elimination of complex and expensive microwave components. These lower frequencies are more easily handled by commercially available components used in other electronic applications.

Feedhorn

LNBs are usually paired with a feedhorn which allows entry of the signal energy in an efficient manner, while keeping the environment out. The feedhorn protects the LNB from the effects of weather and other intruders such as microorganisms, insects, and small animals. The feedhorn material is transparent to the received signal energy.

Polarization

North American broadcast satellites linearly polarize their signal in either the vertical or horizontal plane. Signals in one polarization plane do not interfere with signals in the other plane. Thus, two sets of channels may be transmitted on the same frequency band without causing mutual interference. An LNBF or feedhorn must be oriented so that it can receive one of the two orthogonal signals. An LNBF is a LNB with an integrated feedhorn assembly. These are usually purchased as a single unit. The new family of Direct Broadcast Satellite (DBS) systems being developed for introduction into North America, Europe, and Asia use circular polarized signals. This provides two additional sets of non-interfering channels that may occupy a given frequency band. Left-polarized and Right-polarized signals require an appropriate LNB or LNBF to receive them.

Polarizers

Mechanical polarizers have been added to some LNBs so that a single installation may receive both polarizations of signals, choosing Horizontal/Vertical or Left/Right circular polarization. The LNB monitors the supply line voltage to change the position of the polarizer. These LNBs do not offer dual bands, as the voltage monitoring is used for selecting polarity. The mechanical movement of the polarizer allows the LNB to select the other orientation of signals from the satellite. Some newer models of LNBs use two signal pick-ups and electrically select which input signal orientation will be received and converted.

Dual Polarity Feeds

In some installations, the ability to receive both polarizations at the same time is desired. Commercial applications often have use for this capability. Probes are inserted into the LNB to pick up each signal orientation, and two IF output channels are provided to the

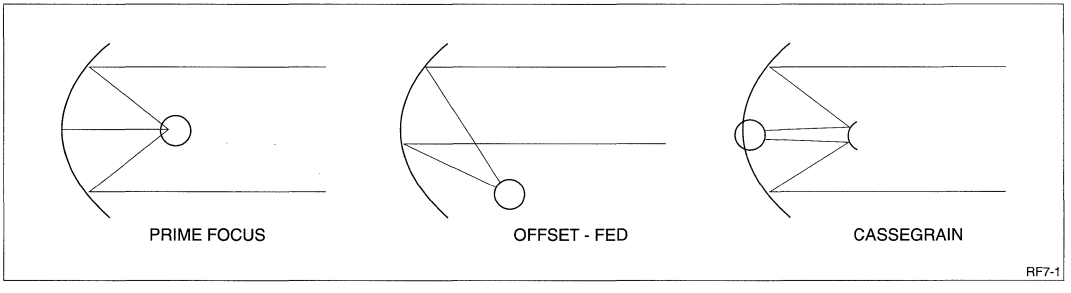


Figure 1. Prime Focus, Offset-fed, and Cassegrain Antennas

receive equipment. Some LNBs provide the capability to select which signal orientation is provided to either of the two LNB outputs.

TUNERS

The Tuner is the frequency selective component in the TVRO receiver. The LNB converts the entire receive band from the satellite broadcast frequencies (C-band or Ku-band) down to the tuner input frequencies. The full range of received channels is available at the tuner input. Through the tuning process, the desired channel will be selected and the unit will provide either an Intermediate Frequency (IF) output for further processing, or demodulated video and audio signals for viewing and listening.

Tuning Process

The tuning process is performed by mixing a Local Oscillator (LO) signal with the input signals in a mixer. The mixer is the device used to combine the received channel signals with the LO signal, to produce an IF output with the desired channel signal. The LO is operated at a frequency higher than the received channel spectrum. A general rule of thumb for mixers is that the signals present at the IF output are the sum, the difference, and the two original frequencies, or groups of frequencies. In tuning applications, the frequency of the LO is controlled by voltage, becoming a voltage controlled LO. A Voltage Controlled Oscillator (VCO) allows for flexible tuning control over the desired operation range (Figure 2).

To tune a channel that may be at 1,000 MHz, the VCO must be positioned at a frequency higher than the desired channel. The difference in frequency must be exact to place the mixer IF output at the desired frequency. In this case, for an IF of 479.5 MHz, to tune a channel frequency of 1,000 MHz, the VCO would be positioned to 1,000 MHz + 479.5 MHz = 1,479.5 MHz. In the mixer, the combination of the 1,479.5 MHz VCO and the 1,000 MHz input signal will produce a 479.5 MHz output IF difference signal (1,479.5 MHz - 1,000 MHz = 479.5 MHz). Thus, 1,000 MHz has been translated to 479.5 MHz. An input at 1,700 MHz could be translated to 479.5 MHz the same way (2,179.5 MHz - 1,700 MHz = 479.5 MHz).

The other frequency components of the process are also present, but are undesired in the final output. For the first example, they are the original frequency, 1,000 MHz, the Local Oscillator Frequency, 1,479.5 MHz, and the undesired sum frequency, 2,479.5 MHz A filter on the IF output of the mixer will be used to bandpass the desired 479.5 MHz difference signal, and reject the undesired components (Figure 3).

IF SYSTEM

The Intermediate Frequency (IF) system amplifies and filters the channel signal of interest. Unwanted channel signals will be suppressed to eliminate sources of interference. The first IF amplifier is controlled by the AGC circuits to adjust the signal level after the LC filter and prior to the Surface-Acoustic-

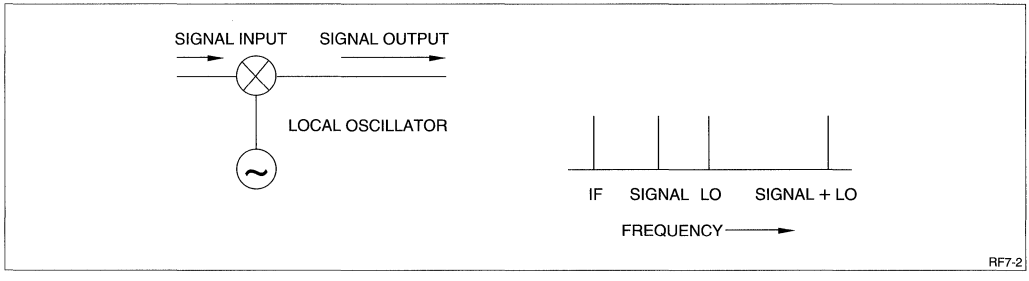


Figure 2. Tuner Mixer Block Diagram

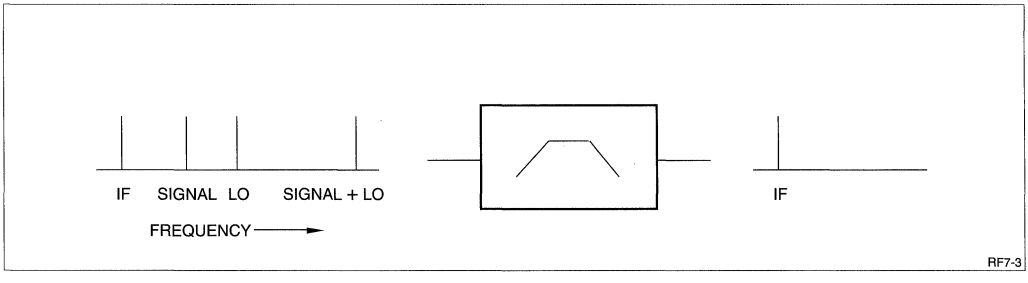


Figure 3. Bandpass SAW Filter and Frequency Spectrum

Wave (SAW) filter. The SAW filter output will then be amplified to provide the correct level into the demodulator assembly (Figure 5).

Filter

The primary bandwidth filter is a Surface-Acoustic-Wave (SAW) filter that has a well defined frequency response. The out-of-band rejection by the SAW filter eliminates most of the signal energy not associated with the selected channel. The well defined characteristics and stability of SAW filters over changes in operating environment are two of the reasons for their use. The bandpass effect is shown in Figure 3.

Multiple Filters

Tuners are available with the capability to select from multiple IF SAW filters. Selection capability is provided by means of an RF switch that connects one of the filters into the IF signal path. An external pin(s) is provided on the tuner assembly to allow for receiver system control of the filter selection. Currently, modules with up to three different bandwidth filters are available in some tuner types.

AGC

Automatic Gain Control (AGC) is incorporated into the tuner to maintain optimum internal signal levels over a defined signal range at the RF Input. The AGC system monitors the IF signal applied to the demodulator and develops a DC voltage proportional to the signal level. This AGC voltage is then amplified and sent to the IF AGC amplifier and the RF Attenuator. As the input signal level increases, the RF Attenuator will decrease the signal level to the mixer and the IF amplifier may decrease its amount of amplification, depending on the AGC level. The AGC system works in the opposite manner for decreasing input signal levels. If tuner components change over time and operating conditions, the AGC system will help to compensate (Figure 5).

DEMODULATOR

Most satellite broadcasts for video transmission use Frequency Modulation (FM) as opposed to terrestrial video broadcast using Amplitude Modulation. The demodulator also uses a Phase-Locked-Loop (PLL) to recover the video signal. In preparation for satellite uplink, the amplitude-varying video signal is converted to a frequency-varying modulated signal. The demodulator PLL in the receiver tries to track the frequency-varying modulated signal, producing an error signal that is proportional to the amount of frequency change. This error signal is then processed as the recovered baseband video signal.

Another signal is added at the transmission facility to help prevent damage to the satellite amplifier components. A 30 Hz triangle wave is superimposed on the FM video signal to keep the signal moving, with respect to frequency, in the satellite's electronic components. This prevents the satellite's power amplifiers from remaining stationary at a single frequency for any period of time, which could damage those components. This is a protection feature to ensure the operational capability of the satellite. This signal is removed at the receiver and will not interfere with the transmission content.

OUTPUTS

VIDEO

Baseband (demodulated) video is provided as one of the outputs from the demodulator section of the tuner. Baseband video is different from a modulated carrier such as broadcast by a local television station. This is the composite video signal that includes all of the information required by a television to present an image. Baseband video corresponds to the video signal that is provided from video recorders and players directly to a monitor. The composite video signal is used as the modulating signal in preparation for satellite uplink or terrestrial broadcast.

For satellite signal reception, the video output occupies a 4.5 MHz bandwidth when the program source material has been deviated with 17 MHz p-p FM modulation at the transmit site (Figure 4). The 4.5 MHz bandwidth includes all of the signals in the baseband video to provide a color television picture for North American NTSC-M format systems. For this example, 17 MHz represents a fully deviated signal from the uplink facility. The use of a wide bandwidth allows for greater signal recovery at the receiver site.

VIDEO MEASUREMENTS

The recovered video signal includes all of the signal information that is required to produce a television picture, and is known as composite video. There is no guarantee that all components of the signal are correct in all respects. After a tuner is installed in a receiver system, the engineer must make certain measurements to insure that the recovered video is not being affected by either the tuner or the supporting video circuits. Care must be taken by the engineer so that the process of making the measurements does not skew the test results. Poorly performed measurements will give the impression that the tuner or video circuits have a problem, when in fact the measurement practice is at fault. Insuring that the receiver system properly recovers the video and audio signals is essential.

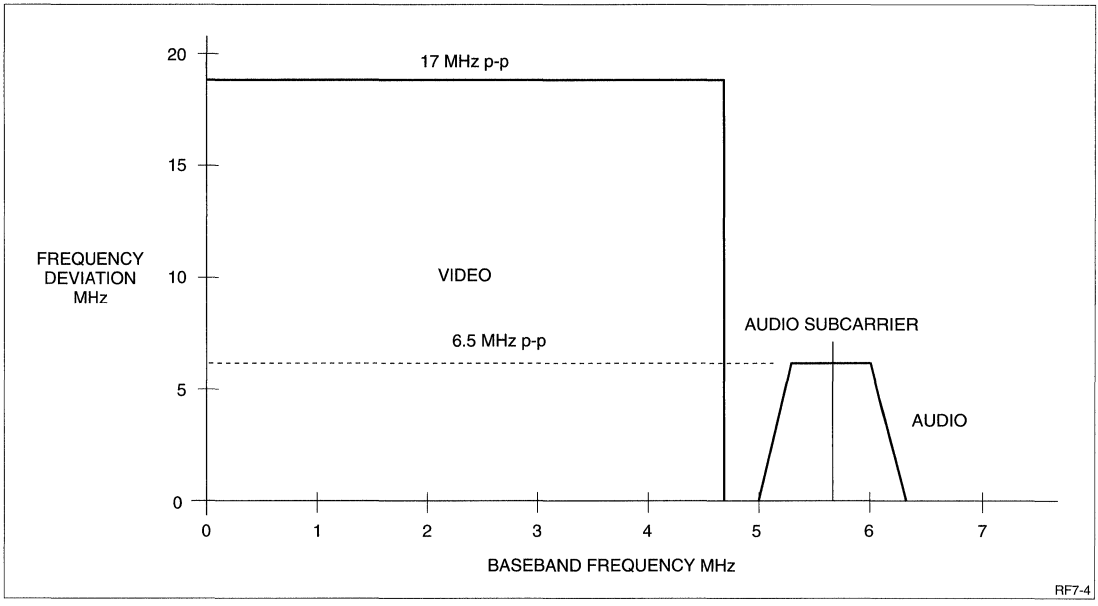


Figure 4. Combined Video and Audio Outputs

AUDIO

The recovered Audio signal is provided on the same output pin as the video information. The audio is carried on a 5.727 MHz subcarrier that positions it above the recovered video signal (Figure 4). When the audio signal at the transmit site is deviated ±3.25 MHz, a ±727 kHz signal will be recovered from the IF demodulator. An audio demodulator integrated circuit will then be used to remove the audio subcarrier and provide an audio output.

The Audio output should also be tested for audio frequency response and amplitude distortions. These distortions may be created at the broadcast facility or at the Direct Broadcast Satellite (DBS) receiver. Recovery of audio signals, including stereo processing, must be performed external to all Sharp tuners. Most manufacturers of Integrated Receiver/Descrambler (IRD) units want control of audio processing in order to offer audio quality commensurate with the cost of the receiver. Stereo, or other extra audio information will be provided on the audio subcarrier along with the regular audio information.

BLOCK DIAGRAM

The block diagram (Figure 5) is representative of a standard DBS tuner. The RF input range is 950 ~ 1,750 MHz. The LNB power is provided through the RF input connector and the LNB Power connection point on the tuner. The input signal is AC coupled into the RF Amplifier. The signal flow path is from the RF amplifier, through the RF Attenuator, the Image Filter, an additional RF Amplifier, and into the Mixer. The Mixer is driven by the signal from the VCO and its Buffer Amplifier. The combined mixer output goes to the LC Filter and the difference signal is passed into the AGC section of the tuner.

The IF AGC amplifiers pass the signal to the SAW bandpass filter, an additional IF Amplifier, and into the PLL Demodulator. The demodulator provides the baseband video and audio on the subcarrier (Figure 4). The AFT output is provided for monitoring the tuning accuracy of the unit. The AGC detector monitors the final IF signal and drives the RF AGC to control the RF Attenuator, and provides an AGC monitor point. TUN is the tuning voltage supply point. PSC out is the prescaled Voltage Controlled Oscillator (VCO) output for return to the Phase-Locked-Loop tuning control system.

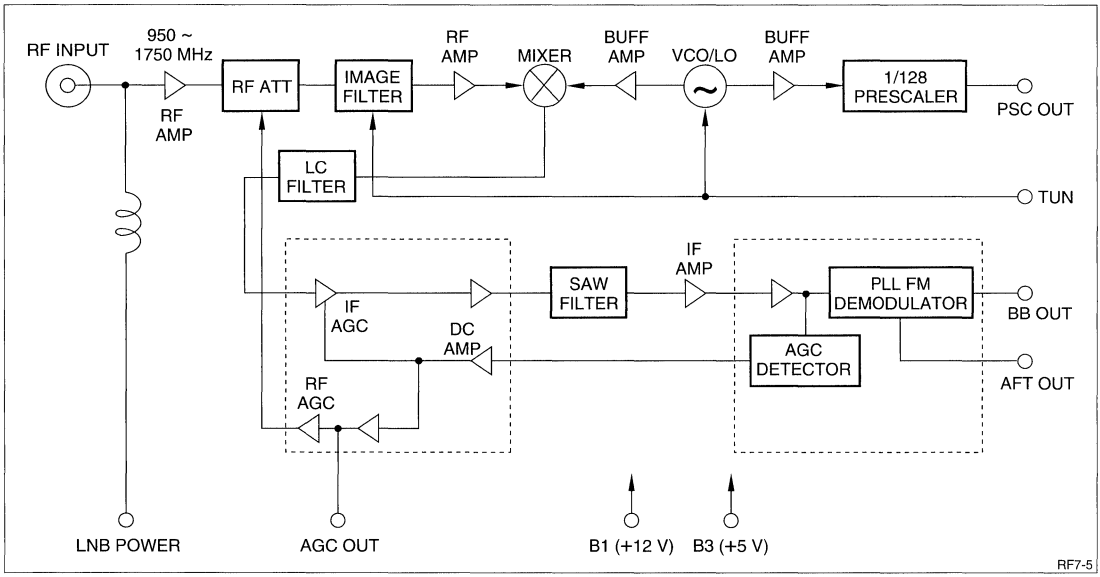


Figure 5. Standard DBS Tuner Block Diagram

GLOSSARY

AGC Automatic Gain Control. Used to automatically compensate for changes in signal strength in the tuner IF system.

DBS Direct Broadcast Satellite. General term used in conjunction with equipment and services associated with the satellite reception process and industry.

Demodulator Converts a modulated signal into the original source signal components.

GHz Acronym for giga-Hertz, one billion cycles-per-second. See MHz.

LNB Low-Noise Block downconverter. Used to convert signal energy at satellite transmission frequencies to lower, more easily processed frequencies for tuning and channel selection.

MHz Acronym for mega-Hertz. One Hertz is defined as one cycle-per-second, for periodic signals. One MHz is equal to one million cycles-per-second.

PLL Phase-Locked-Loop. Device used to accurately control frequency in tuner devices and other applications.

RF Radio Frequency. General reference to electromagnetic energy that will propagate through free space, including a vacuum.

REFERENCES:

[1] *The Ku-Band Satellite Handbook*, Mark Long, 1987, p.21

[2] Ibid, Long, p.25

[3] *Reference Data for Radio Engineers*, Howard Sams & Co., 6 ed., 1975, p.28-18,28-19

ADDITIONAL REFERENCES

Single Conversion DBS Tuners, Training Topic DBS01, Sharp Electronics Corporation, 28 August 1992.

Sharp LNBs, Training Topic LNB01, Sharp Electronics Corporation, 28 August 1992.

The Ku-Band Satellite Handbook, Mark Long, Howard W. Sams & Co., 1987, ISBN:0-672-22522-0.

Reference Data for Radio Engineers, Howard Sams & Company, 6th ed., 1975, ISBN: 0-672-21218-8.

Illustrated Encyclopedic Dictionary of Electronics, John Douglas-Young, Parker Publishing, 1981, ISBN: 0-13-450791-6.

For Sharp Products: BSF-BSFF Series Tuners

Robert Stuart, RF Applications Engineer

OVERVIEW

This training topic is intended to provide background material to anyone who is not familiar with Sharp's Single-Conversion Direct-Broadcast Satellite (DBS) Tuners. The concept and purpose for the tuner, and each of the major sections of the tuner block diagram (Figure 1) will be explained. Topics will start with basic concepts, advancing to more complex information where appropriate.

INTRODUCTION

The Direct Broadcast Satellite (DBS) Tuner is the first section of the satellite receiver system that receives the signal from the antenna Low-Noise-Block (LNB) converter. The LNB for this example, is mounted in the focal point of the parabolic receive antenna. The antenna focuses the received signal energy onto the LNB feedhorn, which then passes it into the LNB. The LNB will then convert the 11.7 ~ 12.5 GHz broadcast signal to a 950 ~ 1750 MHz Intermediate Frequency (IF), an 800 MHz bandwidth in this example. (LNBs are available in different types, with conversion bands of 950 ~ 1,450 MHz, 950 ~ 1,750 MHz, 950 ~ 2,000 MHz, and 950 ~ 2,050 MHz. The LNB for this example will convert the received energy to the 950 ~ 1,750 MHz band.) For more in-depth information on LNBs, refer to the *BSCA-BSCH, BSCW Series Application Note.* The 950 ~ 1,450 MHz bandwidth is used primarily in the United States. The 950 ~ 1,750 MHz through 950 ~ 2,050 MHz bandwidths are primarily used in European and Asian satellite systems.

After the broadcast signal has been converted by the LNB, the signal is sent to the tuner. The 800 MHz-wide broadcast signal now occupies a frequency region that is easier to process than the 12 GHz signal. The entire 800 MHz-wide band is provided to the tuner input. The tuner will then be instructed by the receiver system to select an individual channel from the 800 MHz band, for conversion and demodulation, to provide video and audio information for viewing. The tuner is the conversion module that allows user control and selection of the program material.

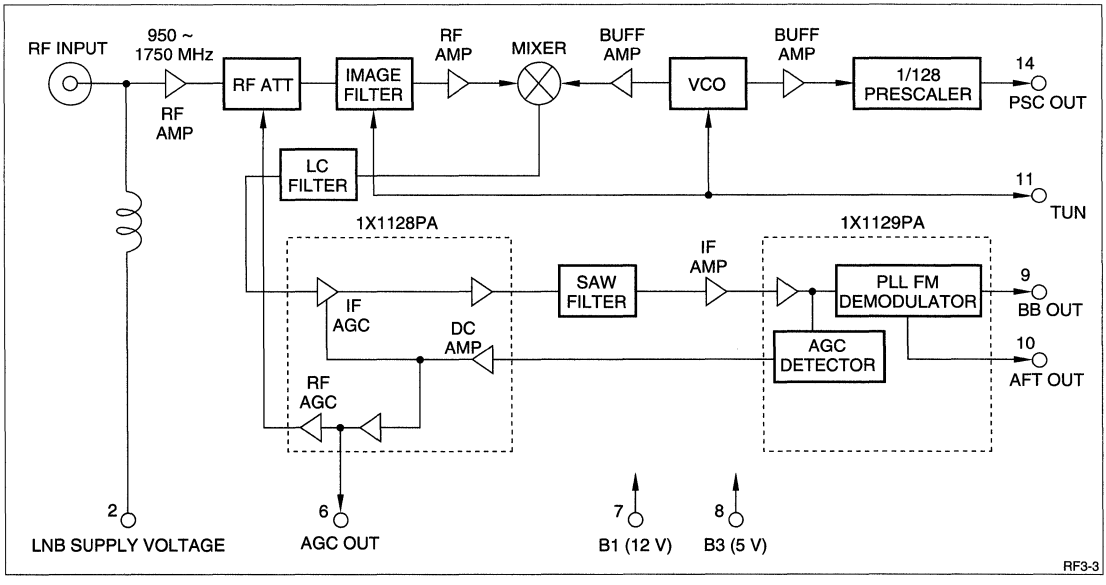


Figure 1. DBS Tuner Block Diagram

RF INPUT

The RF Input is the point of application for the received signal from the LNB. Refer to the tuner block diagram in Figure 1. The RF Input must be capable of handling the full bandwidth of the received signal energy without degradation. A 950 ~ 1,750 MHz signal would not be effectively handled by a 950 ~ 1,450 MHz tuner. This lower bandwidth tuner cannot tune the full band range, and the input is not configured to properly handle the wider input frequency range. This is explained more fully in the next topic, Input VSWR.

Input VSWR

The Input Voltage Standing-Wave Ratio (VSWR) is a specification that expresses how well the RF Input is electrically matched. The input electrical condition is stated to be a 75 Ω (Ohm, unit of resistance or complex impedance) environment. If the input is exactly 75 Ω in all respects, then the system is considered properly terminated, or matched, and all of the signal will be received by the tuner with no loss. In the real world, the electrical characteristics of the tuner input are not perfect and will change slightly. Some signal will always be lost, but the intent is to reduce the amount of loss whenever possible.

VSWR is defined as a "ratio of standing-wave maximum amplitude to standing-wave minimum amplitude". [1] The standing-wave minimum amplitude is generated by energy that is reflected from non-ideal characteristics in the RF Input, often expressed as a mis-match. A VSWR = 1 is ideal. As the input match becomes less ideal, the VSWR figure will increase. A VSWR = 2 is poorer than a VSWR = 1.5. In the tuner specification, a maximum value of VSWR = 2 indicates that the RF Input mis-match will be no worse than that given value. From this VSWR figure, an engineer can determine how much signal may be lost under worst-case conditions. For more information, refer to the Glossary/Appendix.

LNB Power

Electrical power for the LNB is provided by way of the coaxial cable that connects the LNB to the tuner RF Input. The RF Input is AC-coupled to the tuner so that the signal energy is passed into the tuner assembly. The power for the LNB is a DC voltage, and will not be conducted to the tuner components. Any strong noise signals that are present on the DC power may be passed into the tuner input, or effect the operation of the LNB itself. For this reason, the LNB power source must be well-regulated and relatively noise-free to ensure the most efficient system operation.

RF AMPLIFIER

The RF Amplifier is used to raise the level of the incoming signal. Specific levels are required for proper operation of some components, such as the mixer. The RF amplifier at the input helps compensate for losses in the coaxial line from the antenna and LNB. The RF amplifier just prior to the mixer compensates for the loss associated with the Image Filter, ensuring that the mixer-drive level is correct for proper low-noise operation. A properly operated mixer will have a minimum number of spurious products in its output, and will perform in a predictable manner.

RF ATTENUATOR

The RF Attenuator is used to control the level of the signal entering the conversion chain. The input signal may change due to atmospheric conditions, or changes in the actual broadcast signal. The RF attenuator will be controlled by the AGC system to vary the input-signal level to maintain the proper operating conditions. It is part of a closed-loop system that will automatically track changes in the signal level so that the final output does not demonstrate any noticeable signs of the change.

IMAGE FILTER

The Image Filter is a tracking bandpass filter that is used to isolate the mixer section from the RF Input. Even though the block diagram shows signal flow from left to right, some of the energy from the VCO will migrate through the mixer and try to go out the RF Input port. Many regulatory agencies specify the level of signal energy that can be present at the RF Input due to the operation of the tuner itself. A well designed filter will provide a good impedance match between stages and attenuate reflected energy inside the tuner. Signal energy that tries to go from the mixer and through the filter will be reflected by the output of the input RF amplifier. This reflected signal will also be converted, generating spurious signals that will be delayed in time, and interfere with the original received signal. A good filter will minimize the amount of signal reflection and attenuate signals outside of the desired signal band.

The tracking aspect of the filter is seen when the operating frequency of the tuner is changed. As the voltage on the TUN pin is adjusted to change the tune frequency of the VCO, the filter will also change its center frequency of operation. This change in bandpass frequency helps limit the number of individual channel frequencies that are fed to the mixer at one time. The mixer is not an ideal device and generates signal distortions based on the total number of signals and the level of each signal present. The amount of distortions indicate the degree by which the mixer is imperfect. Changing the filter center frequency during tuning limits the number of signals that are presented to and converted in the mixer, improving the intermodulation performance of the mixer and the overall conversion process.

MIXER

The mixer is the device used to combine the received channel signals with the VCO signal, to produce an IF output with the desired channel signal. The VCO is operated at a frequency higher than the received channel spectrum. A general rule-of-thumb for mixers is that the signals present at the IF output are the sum, the difference, and the two original frequencies, or groups of frequencies. The mathematical expression for this operation is:

$$e(t) = m(t) \cos \omega_2 t + m(t) \cos (2\omega_1 \pm \omega_2)t$$

The input signal is represented by m(t) cos ω1(t) and the Local Oscillator (LO) is represented by cos ω2. The input signal has been translated to a new frequency, which is the LO plus or minus the input frequency. To work through the operation in general mathematics, the low end of the input range minus the LO is 11.7 GHz - 10.75 GHz = 0.95 GHz, which is 950 MHz. The upper end of the input range minus the LO is 12.5 GHz - 10.75 GHz = 1.75 GHz, or 1,750 MHz. In this manner, 11.7 to 12.5 GHz has been translated to 950 ~ 1,750 MHz. The sum of each of these signals is also present, though not desired or used in this type of application.

To tune a channel that may be at 1,000 MHz, the VCO must be positioned at a frequency higher than the desired channel. The difference in frequency must be exact to place the mixer IF output at the desired frequency. In this case, for an IF of 479.5 MHz, to tune a channel frequency of 1,000 MHz, the VCO would be positioned to 1,000 MHz + 479.5 MHz = 1,479.5 MHz. In the mixer, the combination of the 1,479.4 MHz VCO and the 1,000 MHz input signal will be combined to produce a 479.5 MHz output IF difference signal. In this manner, 1,000 MHz has been translated to 479.5 MHz. An input at 1,700 MHz could be translated to 479.5 MHz the same way. The other frequency components of the process are also present, but are undesired in the final output. The LC Filter block on the IF output of the mixer will be used to bandpass the desired signal, and reject the undesired components.

VOLTAGE CONTROLLED OSCILLATOR (LOCAL OSCILLATOR)

The Voltage Controlled Oscillator (VCO) is the signal reference used to select the desired channel signal for conversion. The VCO is tuned to a frequency above the channel of interest. The VCO must be tunable, yet stable at the selected signal frequency to ensure a quality conversion process. If the VCO is noisy in amplitude, or unstable in frequency, then the IF output will be poor. Amplitude variations will effect the quality of the recovered signal by causing variations in the amplitude of the IF signal. These variations directly influence the video content of the recovered signal. Frequency variations of the VCO, or jitter, are known as "Phase Noise." The VCO is essentially tuned to a given frequency, but is experiencing small frequency variations. Poor system design will contribute to phase noise errors in the IF signal due to inserted error signals. Phase noise degrades the recovered signal because the IF will contain additional signal energy that was not part of the original signal. Frequency Modulation is used to transmit video information, so frequency errors directly become signal errors. Larger instability of the VCO will cause the IF to be pulled to a greater degree and degrade the overall reception process.

The VCO is tuned by the application of an external tuning voltage. The Phase-Locked-Loop (PLL) frequency synthesizer block will provide a DC voltage proportional to the frequency to be tuned. This tuning voltage must be noise-free and well controlled so as not to inject frequency related problems into the VCO. The tuning voltage is an easy entry point for frequency errors as the voltage directly effects the output of the VCO.

Buffer Amplifier

The VCO output is passed through two Buffer Amplifiers. These amplifiers isolate the VCO from the mixer and prescaler, respectively. Isolation prevents either the mixer or prescaler from affecting the VCO frequency or amplitude output, due to component variation, or changes in the environmental operating conditions. For example, if the input to the mixer is not a perfect impedance match, as described above, it will tend to load the VCO output by a different amount at different frequencies. The buffer amplifier can drive the changes in the mixer without affecting the output of the VCO.

Prescaler

The prescaler divides the VCO frequency, or prescales, for the frequency tuning system. Usually the frequency control system is a PLL device. The PLL loop will compare the scaled VCO signal with the PLL reference frequency to compensate for tuning errors. The prescaler, in this example, divides the VCO output frequency by the constant, 128. Some prescalers may be externally controlled to divide by different constants, as commanded by the PLL control system.

PLL SYSTEM

A Phase-Locked-Loop (PLL) system is generally used as the frequency control for DBS tuners. As the name implies, phase information is actually used to track the operation of the VCO. The PLL device is clocked by a high-stability fixed-reference frequency. The PLL is provided with tuning data by the system microprocessor for each frequency that the system is tunes. Using this information, the PLL looks at the phase difference between the reference and the VCO input from the prescaler. Phase errors correspond to

frequency errors in the loop. Phase differences between the reference and the prescaled VCO output indicate a difference in frequency, as frequency is the rate of change of phase. When the phase difference between the reference and the VCO sample indicates that phase errors are at a minimum, or zero, the VCO is at the correct frequency.

This comparison is performed in the PLL device. The data loaded by the microprocessor is usually a count value that is proportional to the difference between the reference frequency and the desired prescaled VCO input. The prescaled VCO frequency input will be higher in frequency than the reference frequency. The count value provided by the microprocessor will be loaded into counter-registers internal to the PLL. As each cycle of the prescaled VCO input reaches the PLL chip, the counter, or counters if more than one, will be decremented in a predetermined fashion. The output signal from the counter-registers will be compared to the reference frequency. The PLL chip will then make a phase comparison and provide output signals that may be used to correct the VCO tune frequency to place the tuner on channel.

For each channel that the tuner is programmed to receive, there will be a unique count value for the PLL to use as comparison data. This ability allows the PLL to tune and lock to a wide range of channel frequencies, while still maintaining close control over the frequency of the VCO. The flexibility offered by PLL circuits has led to a new generation of lower cost, high quality tuners and other devices across the RF industry.

External PLL

Some DBS tuners, such as the Sharp BSFA series, generally do not have an internal PLL. A prescaler or VCO output is provided for connection with a PLL system. This allows the receiver system engineer to design the PLL system, choosing integrated circuits, filters, and the VCO driver. Some applications may have special requirements that would not be adequately addressed by a pre-designed PLL system. Also, by not including the PLL circuits, the tuner is generally lower in cost to the customer.

Internal PLL

Other DBS tuners are available with an internal PLL system. The tuner module is interfaced with the receiver system microprocessor for the loading of tuning data. Internal PLL systems are designed by the factory engineer to provide accurate and stable frequency control for the tuner. In many cases, the receiver system designer has no need for special tuning requirements, and the modular design fits all requirements. Internal PLL tuners may have a somewhat higher price than non-PLL tuners, depending on production volume.

IF SYSTEM

The Intermediate Frequency (IF) system amplifies and filters the channel signal of interest. The unwanted channel signals will be suppressed to eliminate any source of interfering signals. The first IF amplifier is controlled by the AGC circuits to adjust the signal level after the LC filter and prior to the SAW filter. The SAW filter output will then be amplified to provide the correct level into the demodulator assembly.

Filter

The primary bandwidth filter is a Surface-Acoustic-Wave (SAW) filter that has a well defined frequency response. The out-of-band rejection by the SAW filter eliminates most of the signal energy not associated with the selected channel. SAW filters are used for their well defined characteristics, and stability over operating environment changes.

Multiple Filters

Capability for incorporation of multiple IF SAW filters is designed into many tuner products. Filter selection is provided by an RF switch that connects one filter or the other into the IF signal path. An external pin(s) is provided on the tuner assembly to allow for receiver system control over the bandwidth selection. Currently, up to three different bandwidth filters are available in some types of tuner modules.

AGC

Automatic Gain Control (AGC) is incorporated into the tuner to maintain optimum internal signal levels over a defined range of signal levels at the RF Input. The AGC system monitors the IF signal applied to the demodulator and develops a DC voltage proportional to the signal level. This AGC voltage is then amplified and sent to the IF AGC amplifier and the RF Attenuator. As the input signal level increases, the RF Attenuator will decrease the signal level to the filter and the IF amplifier may decrease its amount of amplification, depending on the AGC level. The AGC system works in the opposite manner for decreasing input signal levels. If tuner components change over time and operating conditions, the AGC system will help to compensate.

IF Ports

In some tuners, an internal IF filter may not be incorporated. In this case, two IF port connectors are installed to support an external IF loop. A filter is inserted in the external IF loop to provide the desired bandwidth. This may be a custom requirement by a particular customer, or a basic feature of the tuner, depending on the application. Data recovery applications often require unusual bandwidths as compared to standard video and audio recovery. IF ports allow for design flexibility.

DEMODULATOR

Most satellite broadcasts for video transmission use Frequency Modulation (FM) as opposed to terrestrial broadcast using Amplitude Modulation (AM). The demodulator also uses a PLL to recover the video signal. In preparation for satellite uplink, the amplitude-varying video signal is converted to a frequency-varying modulated signal. The demodulator PLL in the receiver tries to track the frequency-varying modulated signal, producing an error signal that is proportional to the amount of frequency change. This error signal is then processed as the recovered baseband video signal.

Another signal is added at the transmission facility to help prevent damage to the satellite components. A 30 Hz triangle wave is superimposed on the FM video signal to keep the signal moving, with respect to frequency, in the satellite's electronic components. This prevents the satellite's power amplifiers from remaining stationary at a single frequency for any period of time, which could damage those components. This is a protection feature insuring the operational capability of the satellite.

AGC Detector

The Automatic Gain Control (AGC) detector monitors the level of the IF signal at the input to the demodulator. The AGC detector output will be a DC voltage that is proportional to the IF signal level. The detector input is monitored, as the detector needs a steady input level for consistent recovery of the video signal. The AGC output will be amplified and used to control the RF Attenuator and the IF AGC amplifier, maintaining the level at the detector relatively constant at all times.

OUTPUTS

Video

Baseband (demodulated) video is provided as one of the outputs from the demodulator section of the tuner. Baseband video is different from a modulated carrier such as broadcast by a local television station. This is the composite video signal that includes all of the information required by a television to present an image. Baseband video corresponds to the video signal that is provided from video recorders and players directly to a monitor.

The video output occupies a 4.5 MHz bandwidth when the program source material has been deviated with 17 MHz p-p FM modulation at the transmit site, as shown in Figure 2. The 4.5 MHz bandwidth includes all of the signals in the baseband video to provide a color television picture for North American NTSC-M format systems. For this example, 17 MHz represents a fully deviated signal from the uplink facility. The use of a wide bandwidth allows for greater signal recovery at the receiver site.

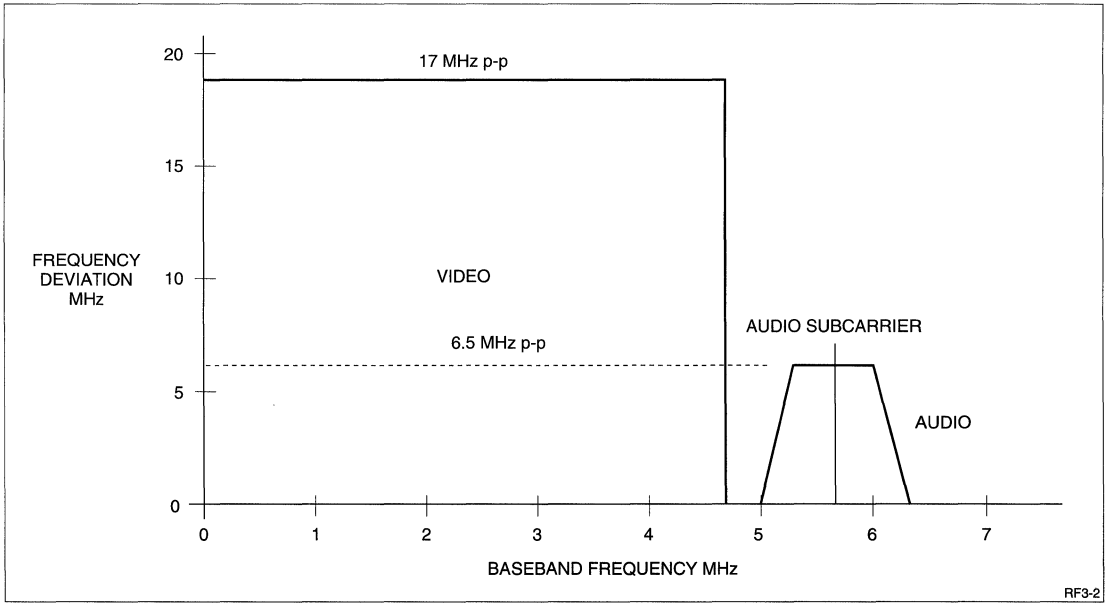


Figure 2. Combined Video and Audio Outputs

Coupling

The video output on many tuners is AC coupled with a capacitor to the video circuits. This means that there is an AC signal path from the tuner output to the video circuits, but no DC path. In applications where a tuner module is being used to replace an earlier model, this point needs to be considered. Some earlier version tuners had DC coupled outputs. If a tuner requiring AC coupling is inserted in a circuit designed for a DC coupled tuner, the output signal will be very low, essentially shorted out. Addition of an appropriate coupling capacitor between the new tuner and the original video circuits will generally solve this problem (Figure 3).

Video AFT

The Video Automatic Fine Tuning (AFT) output provides a method for the demodulator to indicate how precisely the tuner is on channel. The PLL in the demodulator expects the IF input to be at a particular frequency. If for some reason the tuner is tuned below the desired frequency, the AFT output will provide a DC level that is shifted away from its normal level. The polarity of the DC change in voltage indicates whether the tuner is above or below the correct frequency. Usually the AFT output will be summed in with the VCO control voltage to automatically correct for small tuning errors. The tuning errors may be due to the satellite broadcast system, or to slight changes in the components of the receiver system. AFT generally compensates for small errors. Large tuning errors would be corrected by the microprocessor communicating with the main frequency control PLL.

Video Measurements

The recovered video signal includes all of the information that is required to produce a television picture, and is known as composite video. There is no guarantee that all components of the signal are correct in all respects. After a tuner is installed in a receiver system, the engineer must make certain measurements to insure that the video signal is not being affected by either the tuner or the supporting video circuits. Care must be taken by the engineer so that the process of making the measurements does not skew the test results. Poorly performed measurements will give the impression that the tuner or video circuits have a problem, when in fact the measurement practice is at fault. A few of the more common measurements will be touched on here.

Frequency Response "Frequency-response measurements evaluate a system's ability to uniformly transfer signal components of different frequencies without affecting their amplitudes. This parameter, which is also known as gain/frequency distortion or amplitude versus frequency response, evaluates the system's amplitude response over the entire video spectrum." [2] Frequency response problems can cause a wide variety of picture problems. The picture quality may be poor and the specific cause difficult to ascertain.

Differential Gain Differential Gain is present if chrominance gain is affected by the luminance level. [3] Luminance is the portion of the video signal that carries the information describing how bright an individual pixel in the television image will be. Chrominance carries the information about which color and how much color is in each pixel. The combination of luminance and chrominance are used in a television receiver to properly reconstruct the picture image with the correct colors and intensity. [4]

Differential Phase Differential Phase distortion is present if a signal's chrominance phase is affected by luminance level. [5] This type of distortion can cause changes in picture hue when the picture brightness changes, or colors may not be properly reproduced in high-brightness areas. [6] This is because the chrominance signal carries the image color information, and if this signal changes, so will the colors in the displayed image. The chrominance phase is measured relative the color subcarrier signal, which is also part of the composite video signal.

There are a wide variety of other video tests that should be performed on a video recovery system to ensure correct video delivery. The three tests outlined above only provide a little insight into the complexity of the video signal and the need to ensure that all

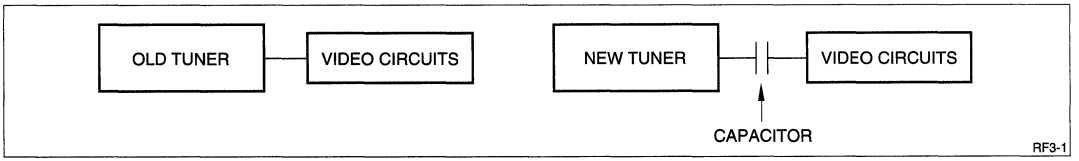


Figure 3. Inserting New Tuner in Old Tuner Circuit

portions of it are correct for a television picture that provides comfortable viewing. Refer to the Appendix for additional information on television measurements and informational references.

Audio

The recovered Audio signal is provided on the same output pin as the video information. The audio is carried on a 5.727 MHz subcarrier that positions it above the recovered video signal (Figure 2). When the audio signal at the transmit site is deviated ±3.25 MHz, a ±727 kHz signal will be recovered from the IF demodulator. An audio demodulator integrated circuit will then be used to remove the audio subcarrier and provide an audio output. The Audio output should also be tested for audio frequency response and amplitude distortions. These distortions may be created at the broadcast facility or at the DBS receiver. Recovery of audio signals, including stereo processing, must be done external to all Sharp tuners. Most manufacturers of Integrated Receiver/Descrambler (IRD) units want control of audio processing in order to be able to offer audio quality commensurate with the cost of the receiver. Stereo, or other extra audio information will be provided on the audio subcarrier along with the regular audio information.

GLOSSARY/APPENDIX

Ohm (Ω) Unit of electrical resistance. In complex circuits such as those seen in RF transmission lines and circuits, the unit of Ohms is used to define complex impedance in addition to electrical resistance. In the impedance plane, both resistive and reactive elements exist. If the reactive element is zero, then the real component (resistance) remains. Reactive components of impedance are due to inductive and capacitive components as seen by the circuit.

Video Measurements The performance of correct video measurements is a specialty and an industry all to itself. These measurements should be approached only after reviewing the correct procedures and methodologies for each test. Incorrect procedures and methods will create erroneous readings and lead to poor conclusions about the operating capability of the equipment under test.

Books and application notes covering television measurements are available from several vendors. Source material for this document was drawn from publications obtained from Tektronix, Inc., a manufacturer of television measurement systems. Use of these source materials does not imply endorsement of any kind by Sharp Electronics Corporation. Tektronix contact point: P. O. Box 1700, Beaverton, Oregon, 800-835-9433, Fax: 503-641-7245.

Voltage Standing Wave Ratio (VSWR), defined as:

$$VSWR = \frac{|V(x)|max.}{|V(x)|min.} = \frac{1+|\Gamma_0|}{1-|\Gamma_0|}$$

where Γ0 is the reflection coefficient of the load, or terminating input in the case of this discussion. [7] The VSWR is a ratio expressing the incident energy compared to the reflected energy at a fixed reference point or plane. In this discussion, the reference point (or plane) is the RF Input of the tuner. An ideal match with VSWR = 1 implies that all of the incident, or incoming energy is perfectly coupled into the RF Input with no loss. If the RF Input has any imperfections (real world) then some amount of energy will be reflected. VSWR is an indicator of how well the RF Input of this tuner matches its expected characteristics. This argument holds for any reference point or plane in an RF transmission line system. Bear in mind that transmission line theory operates in the complex impedance plane and has many attributes that are not discussed here. Any good text on transmission line theory will explain this topic in greater depth.

REFERENCES

[1] Illustrated encyclopedic Dictionary of Electronics, John Douglas-Young, 1981, Parker Publishing Co.

[2] Television Measurements, NTSC Systems, Margaret Craig, p. 37, Tektronix, Inc., Beaverton, OR

[3] ibid, Craig, p. 52

[4] Solving the Component Puzzle, John Horn, p.2, Tektronix, Inc., Beaverton, OR

[5] ibid, Craig, p.48

[6] ibid, Craig, p.48

[7] Microwave Transistor Amplifiers, Guillermo Gonzalez, 1984, p. 4 - 8, Prentice Hall

ADDITIONAL REFERENCES

RF/IF Designer's Handbook, 1992, Mini-Circuits Division of Scientific Components.

Principles of Communications, Ziemer/Tranter, 1985, Houghton Mifflin Company.

Telecommunication Transmission Handbook, Roger L. Freeman, 1991, Wiley & Sons.

BSCA-BSCH, BSCW Series
Low-Noise Blocks

For Sharp Products: BSCA-BSCH, BSCW Series

Robert Stuart, RF Applications Engineer

OVERVIEW

Sharp Corporation manufactures a large variety of Ku band Low-Noise Blocks (LNBs) in different shapes, sizes, and electrical performance characteristics. Though some of these products are older and may no longer be in production, all LNBs are based on similar concepts. Newer products offer a variety of features that were not available in earlier models, such as dual polarization reception capabilities, and smaller size.

After completion of this material, you should be familiar with the basic concepts of LNBs and how and why they are used. At that time you should be able to review a LNB specification and understand most of the key information that is used as the basis for selecting a LNB for a particular application.

INTRODUCTION

LNB is an acronym for Low-Noise Block down-converter. Low-noise refers to the amount of electrical signal noise added by the LNB to a received signal. With lower added noise, the overall signal recovery of the system is improved. Sharp LNBs are manufactured for the Ku satellite broadcast band, in the 10.95 GHz to 12.7 GHz range. An LNB is used to convert the 12 GHz satellite signal to more usable frequencies in the 950 - 2000 MHz range. This conversion lowers the cost of the processing electronics by the elimination of complex and expensive microwave components. These lower frequencies are more easily handled by commercially available components used in other electronic applications.

BACKGROUND

Early satellite receiving systems operated at C-band frequencies as technology available at that time used 3.4 GHz to 4.8 GHz frequencies. With improvements in technological capabilities, new satellites were launched that incorporated Ku-band technology operating at higher frequencies and higher power levels. The advantages of the Ku-band systems are smaller satellite components, smaller earth station components, and earth stations become more portable. Additionally, Ku-band systems experience minimal interference from terrestrial microwave systems, as do C-band systems. The trade-off for operation at Ku-band frequencies are the atmospheric and environmental effects experienced. Ku-band frequencies are more susceptible to signal fading and reception noise due to snow and rain scatter degradation.

SATELLITES

Most satellites associated with Ku-band broadcast of signals for commercial and household use occupy allocated spaces above the equator. These satellites have been placed in equatorial orbits with rotational periods equal to the rotation of the earth. This creates the appearance of the satellite being stationary over a fixed point on the earth. The satellites actually wander a bit, due to gravitational effects of the moon and other minor orbital deviations. Ground control stations monitor the position of each satellite and send commands to the satellite, firing control jets to make orbit corrections to maintain position. Earth station receiving antennas may then point at the same location in the sky and always receive the desired signal.

Inclined-orbit satellites occupy positions other than directly above the equator. Their orbits may be polar or slight deviations above and below the equator. Their daily rotational periods may differ from one earth day. Therefore, receiving antennas must be equipped to track the movement of the satellite.

ANTENNAS

Satellite receiving antennas may be one of several types: Flat-plate, Cassegrain, Off-set fed, or prime focus. Prime-focus antennas are the popular parabolic reflectors seen in most applications. They are made of metal or metalized fiberglass material. These are called prime-focus antennas as the LNB, the receiving element, is located at the prime focus of the parabolic reflector. The prime focus is the location where all of the received electro-magnetic energy converges in front of the reflector at a common point. As an analogy, a flashlight bulb is located at the prime focus of its reflector.

Off-set fed antennas are a variation of a parabolic antenna with the prime focus off-set from the center of the reflector material. Cassegrain antennas mount the LNB at the center of the reflector and place a smaller hyperbolic reflector at the prime focus of the main reflector. Flat-plate antennas are either a Fresnel lens variety or a phased-array. These are less common and are generally more expensive. With advances in manufacturing technology, phased array antennas

may become more common as new, higher-power satellites enter service. These antennas are generally smaller due to cost. With the higher output power of the new satellites, smaller antennas become efficient and economical.

BLOCK DIAGRAM

A block diagram for an LNB is shown in Figure 1. The signal input is at the WR-75 waveguide adapter on the left. In many applications, a feedhorn will be attached to the LNB input. The LNA is a Low-Noise Amplifier used to boost the incoming signal. The term BPF stands for Band-Pass Filter, used to reject unwanted signals. The Mixer combines the input signal with the LO (Local Oscillator) signal to produce the desired output signal. The IF Amp amplifies the Intermediate Frequency produced at the output of the mixer. The IF output will be presented to the input of the tuner module in the receiver system. The Power Supply supplies the required operating voltages for the LNB.

FEEDHORN

The feedhorn protects the LNB from the effects of weather and other intruders such as microorganisms, insects, and small animals. The feedhorn is transparent to the electro-magnetic energy that is focused by the antenna, passing it to the signal input of the LNB.

LOW NOISE AMPLIFIER

The low-noise amplifier increases the strength of the received signal. The low-noise requirement is essential for efficient operation of the LNB so the recovered signal is not degraded by injected noise from the LNB. Noise indexes are given in terms of deci-Bels (dB) or noise temperature (Kelvin scale). Good LNBs have noise figures that are low. The noise figure is an indication of the amount of noise added by the LNB during reception of a signal. A noise figure of 1.5 dB or lower is generally considered good for most applications, however in some circumstances where designs are being pushed to the limit, noise figures of 1.2 are barely adequate. LNB noise figure measurements are performed across the entire operating range of the device at time of manufacture. The overall LNB noise figure rating will be determined by the highest level measured across the operating range.

BAND-PASS FILTER

The bandpass filter is used to reject unwanted signals received by the LNB. If the input signal is allowed to include frequencies beyond the desired signal band, those undesired signals will also appear at the input of the mixer. In addition, the total noise contribution of the LNB to the received signal is a function of bandwidth. A wider bandwidth would allow a greater amount of noise to be carried through with the desired signals. Restricting the signal bandwidth will hold LNB noise to a minimum, yet allow the recovered signals to pass unrestricted.

MIXER

The signal mixer is used to combine two signals to produce a different result. In this case, the input signals at 11.7 to 12.2 GHz will be converted to a 950 to 1,450 MHz Intermediate Frequency. The mixer output is a function of the input signal combined with the Local Oscillator signal to provide an Intermediate Frequency output signal. A general rule-of-thumb for mixers is that the signals present at the output are the sum, the difference, and the two original frequencies, or groups of frequencies.

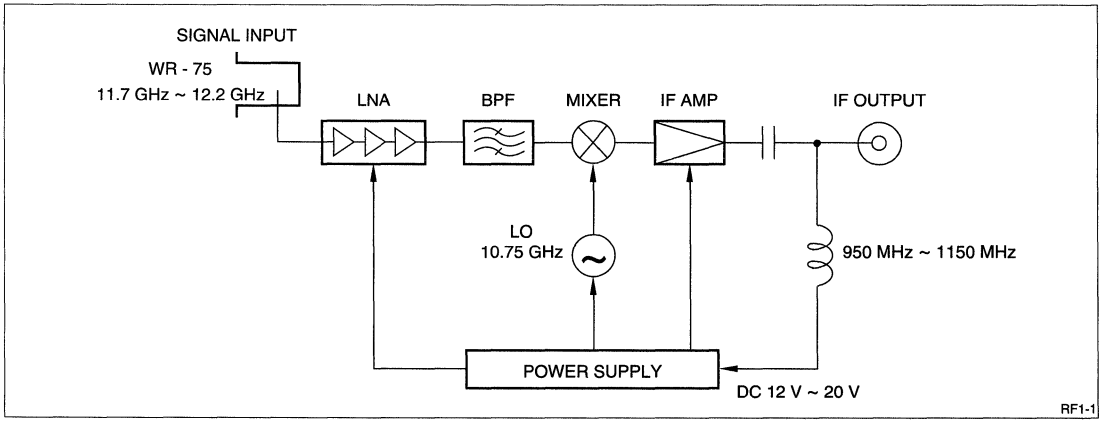


Figure 1. LNB Block Diagram

The Local Oscillator is operating at 10.75 GHz for the LNB in the block diagram (Figure 1). The mathematical expression for this mixing operation is:

$$e(t) = m(t) \cos \omega_2 t + m(t) \cos(2\omega_1 \pm \omega_2)t$$

The input signal is represented by m(t) and the Local Oscillator (LO) is represented by ω_2. The input signal has been translated to a new frequency, which is the LO plus or minus the input frequency. To work through the operation in general mathematics, the low end of the input range minus the LO is 11.7 GHz - 10.75 GHz = 0.95 GHz, which is 950 MHz. The upper end of the input range minus the LO is 12.2 GHz - 10.75 GHz = 1.45 GHz, or 1,450 MHz. In this manner, 11.7 to 12.2 GHz has been translated to 950 ~ 1,450 MHz. The other high-frequency components of the process are also present, but not desired. In general, the tuner module will filter them from its operating band. For a more thorough explanation of this process, refer to the references listed on page 4.

LOCAL OSCILLATOR

The Local Oscillator is a fixed-frequency oscillator used as part of the frequency-conversion process. This signal is used as the conversion reference. The purity of the LO signal will affect the quality of the recovered signal. If the LO is noisy, that noise will also appear on the output signal. If the LO shifts in frequency, then the output of the mixer will shift as well. The LO must be stable over changes in temperature and humidity to ensure a predictable, high quality output from the LNB. Stability information is usually provided in an LNB specification.

IF AMPLIFIER

The IF amplifier boosts the strength of the mixer output. If designed as such, an IF amplifier may also be more sensitive to the band of frequencies to be amplified, thus helping reject the undesired mixer output signals. The IF amplifier also maintains the 75 Ω impedance environment for the coaxial cable to the tuner module. A well designed IF amplifier presents a nearly constant impedance to the receive system over the entire band of amplified frequencies.

POWER SUPPLY

The electrical power for the LNB is supplied via the coaxial cable from the tuner module. The power is supplied as a DC voltage. The IF amplifier output is AC coupled to the coaxial cable. As seen in the block diagram, the DC voltage is applied to the Power Supply unit. The Power Supply will convert the DC supply voltage to the required operating voltages for the LNB. If the DC supply voltage is noisy, then the LNB will exhibit some of this noise in the recovered output signals. The noise may be on the supply voltage as provided from the receiving unit, or induced into the LNB or coaxial cable if located near a high noise source. Additionally, if the coaxial cable is damaged, noise may enter through apertures in the shielding braid that encloses the cable.

CONVERSION GAIN

All LNBs have a minimum amount of signal power gain associated with them. The gain is an increase in the signal level as compared to the received signal. Gain figures are usually given in terms of dBs. A power gain of 40 dB is an amplification factor of 10,000. Conversion gain is listed as the minimum gain measured across the total receive band of the LNB. LNBs with a high conversion gain may not have the best noise figures. For this reason, noise figure and conversion gain are considered when designing a receiver system.

FOCAL LENGTH

The LNB is normally positioned at a focus point in front of the reflector dish. The exact position is determined by the shape of the reflector. In deep dish designs, the LNB feedhorn is located close to the reflector. In shallow designs the LNB is further away. This leads to the f/D rating; focal length to Diameter of the reflector dish. Antennas with a f/D rating of 0.25 to 0.33 are considered deep dishes, where reflectors with a rating of 0.35 to 0.45 are considered shallow. An LNB will work most effectively when matched with a feedhorn that has a f/D rating appropriate for the antenna on which it will be installed, thus seeing the area of the dish most efficiently. The f/D rating applies to discrete feedhorns when purchased separately from the LNB. When the feedhorns and the LNB are obtained as a single unit, they are known as a LNBF; LNB with Feedhorn. In this case, the LNBF would have an associated f/D rating.

MULTIPLE BAND

Some LNBs are constructed to receive more than one broadcast band. In the previous example, the receive frequency range was 11.7 to 12.2 GHz. Other LNBs are able to receive 12.2 to 12.7 GHz, 12.25 to 12.75 GHz, and other frequency bands. Some new LNBs are capable of receiving two or three different bands (Figure 2). The LNB is commanded to change receive band by changing the supply voltage fed to the LNB. In general, the supply voltage may be +12 V. To change to the second receive band, the supply voltage is raised to more than +18 V.

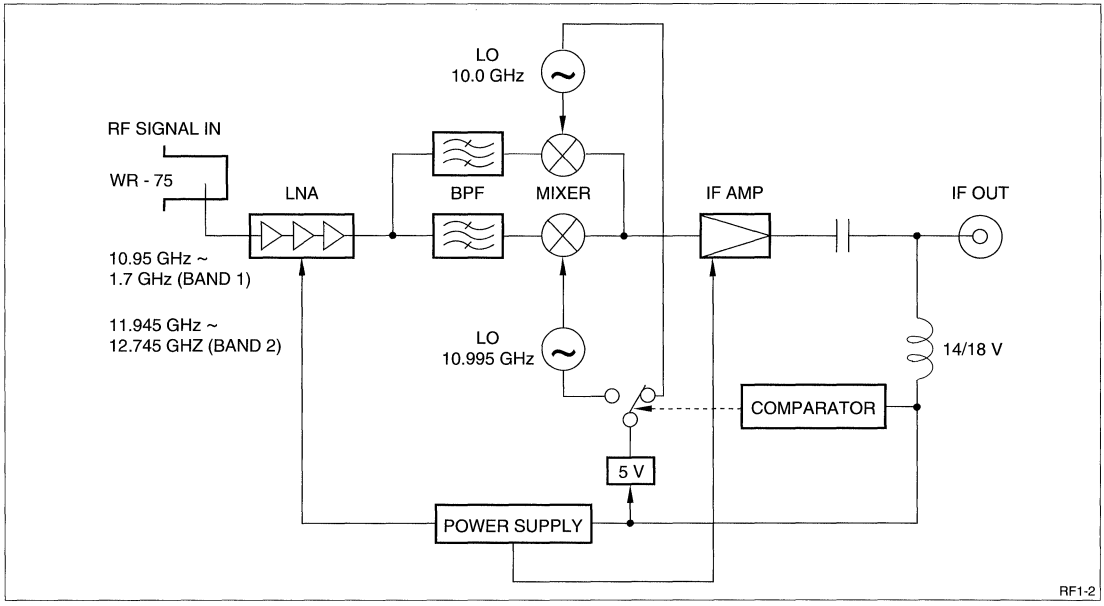


Figure 2. LNB Block Diagram

A comparator in the LNB monitors the supply voltage and actuates a switch to select an alternate LO signal in the LNB. The output IF signal will be the product of the new LO signal and the new input band. In some cases, the IF output may be different than for the original receive band. The Power Supply module in the LNB continues to provide the correct operating voltages to the LNB circuits while the higher input voltages are applied. As can be expected, multiple-band LNBs are more complex and more expensive than single-band units.

POLARIZATION

North American broadcast satellites linearly polarize their signal in either the vertical or horizontal plane. Signals in one polarization plane do not interfere with signals in the other plane. Thus, two sets of channels may be transmitted on the same frequency band without causing mutual interference. An LNBF or feedhorn must be oriented so that it can receive one of the two orthogonal signals.

The new family of Direct Broadcast Satellite (DBS) systems being developed for introduction into North America, Europe, and Asia use circularity polarized signals. This provides two additional sets of channels that may occupy a given frequency band. Left-polarized and right-polarized signals require an appropriate LNB or LNBF to receive them.

POLARIZERS

Mechanical polarizers have been added to some LNBs so that a single installation may receive both polarizations of signals, choosing Horizontal/Vertical or Left/Right circular polarization. As in the dual receive-band LNBs, the LNB monitors the supply line for a change in voltage to change the position of the polarizer. These LNBs do not offer dual bands, as the voltage monitoring is used for selecting polarity. The mechanical movement of the polarizer allows the LNB to select the other orientation of signals from the satellite. Newer LNBs perform the switching function electronically and do not use mechanical polarizers (Figures 3 and 4).

DUAL POLARITY FEEDS

In some installations, the ability to receive both polarizations at the same time is desired. Commercial applications often have use for this capability. Probes are inserted into the LNB to pick up each signal orientation, and two IF output channels are provided to the receive equipment. Some LNBs provide the capability to select which signal orientation is provided to either IF output. The selection is accomplished in the same manner as band selection. Refer to Figure 5.

REFERENCES

The Ku-Band Satellite Handbook, Mark Long, Howard W. Sams & Co., 1987.

Principles of Communications, Second Edition, Ziemer/Tranter, Houghton Mifflin Co., 1985

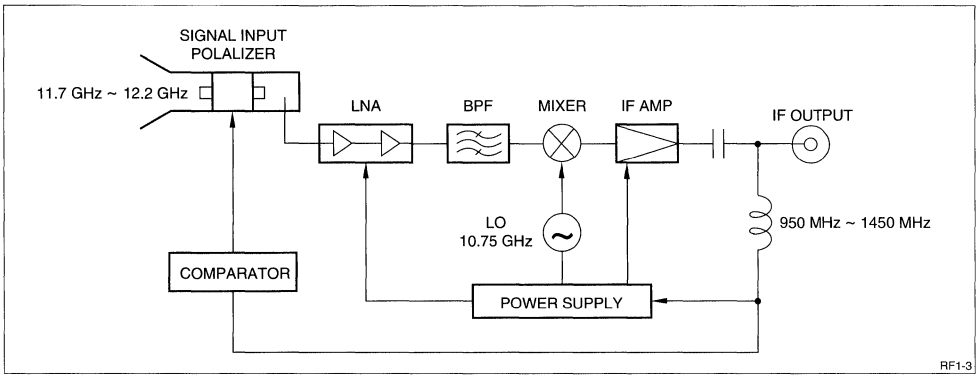


Figure 3. H/Y LNB Block Diagram

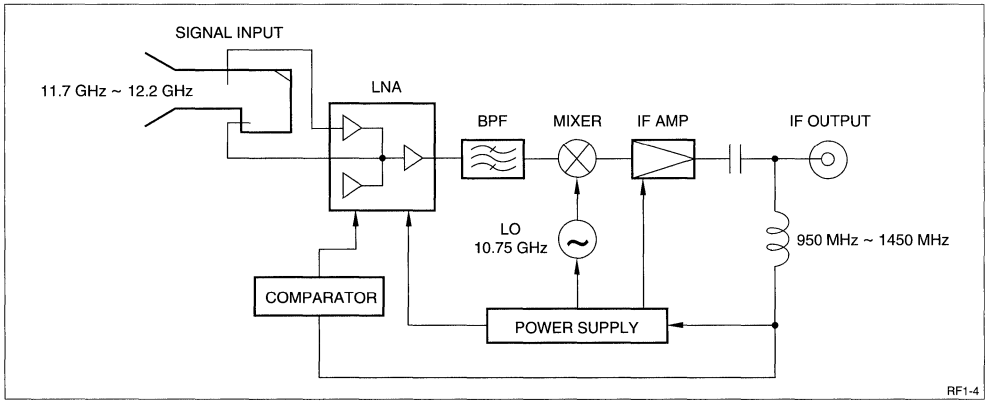


Figure 4. H/Y LNB Block Diagram

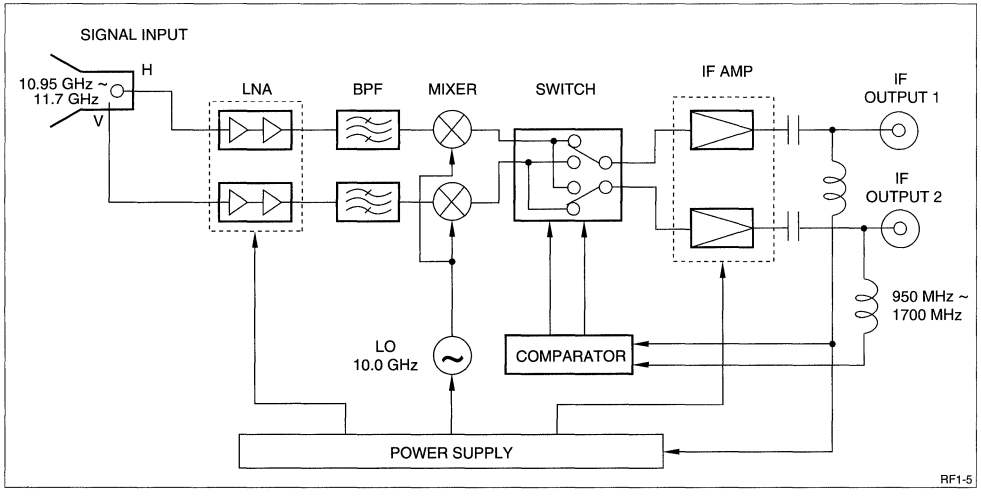


Figure 5. H/Y LND Block Diagram

PROGRAMMING THE MC145158-2 PLL For Sharp Products: BSFB73H02/3/4

Robert Stuart, RF Applications Engineer

The Motorola® MC145158-2 Serial-Input PLL Frequency Synthesizer is the primary tuning control in SHARP BSFB73H02/3/4 DBS tuners. These tuners provide an Intermediate Frequency (IF) output that may be used for a wide variety of applications. The IF output frequency for each tuner is different, however the tuning mechanism is identical. The second local oscillator (2nd LO) frequency is unique for each different tuner. Comparison data is provided in Table 1.

Table 1.
Local Oscillator and IF Output Comparison

P/N	INPUT RANGE MHz	1st IF MHz	2nd LO MHz	OUTPUT IF MHz
BSFB73H02	950 - 1,450	612	655.5	43.5
BSFB73H03	950 - 1,450	612	656.0	44.0
BSFB73H04	950 - 1,450	612	682.0	70.0

Automatic Fine Tune (AFT) control is provided for the 2nd LO. The AFT input's voltage range of +1 V to +11 V (+6 ± 5 V) with a scaling factor of 900 kHz/V typical. Unpredictable operation may result if the AFT control voltage is taken below +1.0 V.

The MC145158-2 is clocked from a 10 MHz crystal. The Reference Divider, R, should be set to a value of 40 to match the 250 kHz PLL Reference Frequency and 1 MHz PLL Step Size indicated in the General Specifications list for the tuner. The Reference Divider should be set to 40 when the MC145158-2 is initialized; 10 MHz/250 kHz = 40.

Figure 1 shows the overall Voltage-Controlled-Oscillator (VCO) loop. The VCO output is immediately divided by four and fed to the modulus-two prescaler. The prescaler divide ratio is controlled by the PLL Modulus Control output. The combination of the divide-by-four and the 250 kHz Reference Frequency provide the 1 MHz PLL Step Size.

The MC145158-2 specification gives the relationship of the programmable-divide ratio as $N_T = (N \times P + A)$, where N is the number programmed into the internal 10-bit N counter, P is the external ÷32/÷33 prescaler, and A is the number programmed into the internal 7-bit +A counter. To close the loop in Figure 1,

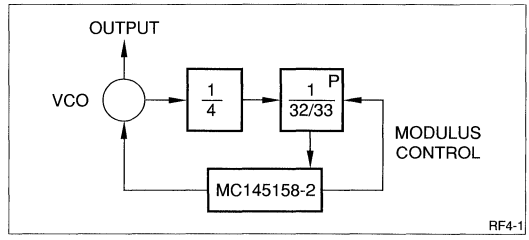


Figure 1. VCO Loop Diagram

the VCO output must match the internal reference frequency of the MC145158-2. The 1/4 prescaler will designated N_P in the relation $N_T = N_P \times (N \times P + A)$. The Reference Frequency will be recovered from $F_{VCO} \div N_T = 250$ kHz. The phase difference between the internal Reference Frequency and the scaled VCO output will be compared to provide the phase detector output signals.

Example: Tuning to 950 MHz

If the desired frequency is 950 MHz, then the VCO will be tuned above the selected frequency by a value equal to the 612 MHz IF frequency. Therefore:

$F_{VCO} = F_{TUNE} + 612$ MHz.

The working relationship is:

$$\frac{F_{VCO}}{\text{Ref. Freq.}} = N_P \times (N \times P + A) \text{ changing to:}$$

$$\frac{F_{VCO}\ \text{MHz}}{0.25\ \text{MHz}} = 4 \times (N \times 32 + A)$$

where: $F_{VCO} = 950$ MHz, Ref. Freq. = 0.25 MHz (250 kHz), $N_P = 4$, and P = 32. Cancel the MHz units and multiply by 0.25 on both sides of the equation to provide the final relationship $F_{VCO} = (N \times 32 + A)$.

Therefore: $F_{VCO} = 950$ MHz + 612 MHz = 1,562 MHz, and work with the number 1,562.

$1{,}562 = N \times 32 + A$, divide both sides by 32.

1,562/32 = N + A/32. Working with the left side of the equation, 1,562/32 = 48.8125.

Interpreting N + A/32 as an integer plus a fraction and 48.8125 MHz as an integer plus a decimal fraction, set N = 48.

Then set the decimal remainder 0.8125 = A/32.

$A = 32 \times 0.8125$. $A = 26$.

In summary: N= 48, A = 26, P = 32, for the desired tune frequency of 950 MHz.

Repeating the example for 1,125 MHz, in a condensed form:

F_{VCO} = 1,125 MHz + 612 MHz = 1,737 MHz, use 1,737. $1,737 \div 32 = 54.28125$.

Therefore: $N = 54$, $A = 32 \times 0.28125 = 9$

Summary: N = 54, A = 9.

Table 2 shows various frequencies of the input tuning range and their associated N and A values.

Table 2.
Input Frequencies with N and A Values

FREQUENCY MHz	N	A	FREQUENCY MHz	N	A
950	48	26	1,250	58	6
1,000	50	12	1,300	59	24
1,050	51	30	1,350	61	10
1,100	53	16	1,400	62	28
1,150	55	2	1,450	64	14
1,200	56	20			

For further information on programming the Motorola MC145158-2, refer to the Motorola Semiconductor Technical Data sheets for the MC145158-2 series, and Motorola product application notes.

REFERENCES

Mr. Don Voss, Motorola Corporation.

Motorola Semiconductor Technical Data, MC145158-2 Series.

RECOMMENDED READING

AN535, *Phase-Locked Loop Design Fundamentals*, Motorola Semiconductor Products, Inc., 1970.

AR254, *Phase-Locked Loop Design Articles*, Motorola Semiconductor Products, Inc., Reprinted with permission from Electronic Design, 1987.

BR504/D, *Electronic Tuning Address Systems*, Motorola Semiconductor Products, Inc., 1986.

DESIGNING WITH THE LC7215
For Products Employing the Sanyo® LC7215 PLL Synthesizer

Robert Stuart, RF Applications Engineer

The Sanyo LC7215 Phase-Locked-Loop Synthesizer is used in several different Sharp tuner modules. The designer may choose the synthesizer internal reference frequency from four programmable options. Selection of the reference frequency is dependent on the prescaler installed in the tuner module in use, and the system application. The 11.16 MHz crystal clock frequency as referenced in the LC7215 specification is used in all applications of the LC7215 in Sharp tuners.

In this particular example, the tuner module uses an internal divide-by-256 prescaler on the Voltage-Controlled-Oscillator (VCO) output. The programmable Divider N should be loaded with a value that is the ratio of the prescaled VCO input frequency divided by the selected F_{REF} (Figure 1).

Relevant equations for this synthesizer are:

$$F_{REF} = \frac{F_{VCO}}{256 \times N} \qquad N = \frac{F_{VCO}}{256 \times F_{REF}} \qquad F_{REF} = \frac{F_{CRYSTAL}}{No}$$

"No" is the reference divider as selected by serial data bits Q and R. Four different PLL Reference Frequencies (No) may be chosen: 1 kHz, 5 kHz, 9 kHz, and 10 kHz. These choices allow for selection of different VCO step increments in the tuner. The step value is the prescaler value times F_{REF}: $256 \times F_{REF}$ (Table 1).

The step increment is used to determine the appropriate value for N in the expression:

$$N = F_{VCO} / [256 \times F_{REF}]$$

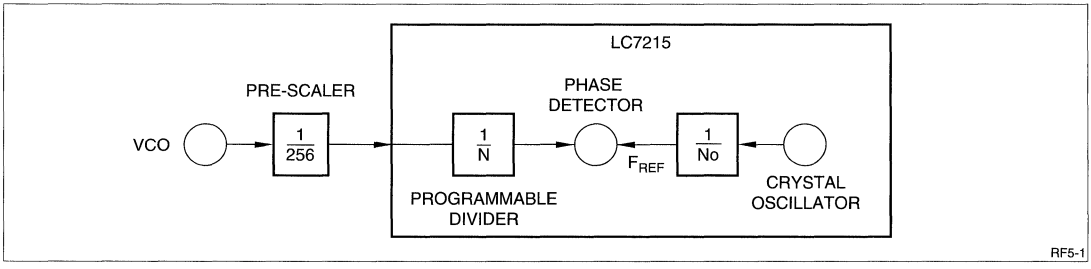


Figure 1. LC7215 Tuner Module

Table 1.
Prescaler Step Values

REFERENCE FREQUENCY No	F_{REF} kHz	PRESCALER	STEP INCREMENT kHz	N 950 MHz INPUT	N 1,750 MHz INPUT
1 kHz	11.16	256	2,856.96	500	780
5 kHz	2.232	256	571.392	2501	3902
9 kHz	1.24	256	317.440	4503	7023
10 kHz	1.116	256	285.696	5003	7803

The Programmable Divider value N must be determined to correctly divide the prescaled VCO signal so that it may be compared to F_{REF} in the phase comparator. Selection of a higher Reference Frequency allows for smaller step values when tuning, if desired.

The Programmable Divider will accept a 14-bit value, in the range from 0 to 16,383. All of the N values listed in Table 1 are within this range. Values for N are listed for tuning frequencies of 950 MHz and 1,750 MHz. To tune to 950 MHz, with a selected No of 9 kHz, use the equation for N listed.

$$N = F_{VCO} / [256 \times F_{REF}]$$

The resultant step increment is 317.44 kHz. The value of F_{VCO} for 950 MHz is 950 MHz + 479.5 MHz = 1,429.5 MHz. (The standard IF frequency for these tuners is 479.5 MHz).

Therefore:

$$N = 1{,}429.5 \text{ MHz} / 317.44 \text{ kHz} = 4503.$$

This value for N would be appropriate for the selected conditions.

Bit S is used to select the programmable divider's input frequency range; 0.5 to 2.5 MHz, or 2.3 to 13 MHz, (or 20 MHz depending on device suffix). For Sharp tuners using the LC7215, this bit value is always set to "1", to select the 2.3 to 13 MHz range. A 11.16 MHz crystal is used as the reference for all applications with the LC7215.

For further specific information, please refer to Sanyo specification number 3584A, or your local Sanyo representative.

For Sharp Products: VTSS6USZF/1, VTSS7USZF/1, VTSS6USFF, RFIVU3, DUNTR0104, E6095, E7853, RFSP7US1, RFSO7US3, RFSN7US2, CAJA35U01 Series

Robert Stuart, RF Applications Engineer

OVERVIEW

The Cable Television (CATV) environment encompasses a wide variety of possible applications from headend equipment to in-the-home Set-Top Converters (STC). A selection of tuners, demodulators, and modulators for use in CATV applications are available in a variety of combinations. These devices are described, and basic specification sheets are provided. The progression of development from single tuner, demodulator, and modulator packages into combined units is outlined.

INTRODUCTION

In a large number of CATV applications, tuners are mixed with microprocessor-controlled systems for remotely controlled selection of the tuned channel. This may be at the headend or in the subscriber's home. The tuning range required for these applications covers a minimum of 50 MHz to 550 MHz, and in some systems as high as 1074 MHz. Most CATV systems do not exceed 550 MHz at the present time. The tuners described here cover the 50 MHz to 550 MHz range, or 50 MHz to 801 MHz range.

All of the devices described here are usable in some CATV applications. The internal-Phase-Locked-Loop (PLL) tuners employ a three-wire serial interface for programming the PLL device. The addition of an external PLL device of the designer's choice is required for the final tuner described.

Six of the tuner products covered are single-conversion tuners, and the seventh device a double-conversion tuner. Various combinations of tuner, demodulator, and modulator combinations are available, in different package styles. All are intended for consumer or related commercial applications and are not specifically designed for instrumentation quality products or life-support medical applications.

VTSS6USZF TUNER

The VTSS-series tuners are the heart of the internal-PLL single-conversion tuner family that is currently available. These tuners are designed for use in systems that comply with the NTSC-M specification (United States and North America) for broadcast television. The Toshiba® TD6359P PLL Integrated Circuit (IC) is used as the frequency control device. The TD6359P automatically takes care of band-switching requirements to enable the tuner to cover the entire 55.25 MHz to 801.25 MHz tune range. A three-wire serial interface is provided for loading control data into the PLL. The interface is a three-wire serial interface that is easily connected to a microprocessor with lines for data, data clock, and load enable.

The tuner dimensions are 69.4 mm × 45.6 mm × 13 mm (2.73 in. × 1.79 in. × 0.51 in.). The RF input is a RCA connector, mounted on the end of the unit. Power requirements are +9 V for tuner operation, +5 V for the PLL chip, and +33 V for the internal Voltage Controlled Oscillator (VCO). An AGC input is provided. The output is a standard composite Intermediate Frequency (IF) with an inverted spectrum. The IF is referenced to the video carrier at 45.75 MHz. The audio carrier is at 41.25 MHz. Standard demodulators are available to recover the video and audio signals.

EIA/IS-31

The VTSS6USZF is compliant to Electronic Industries Association interim standard EIA/IS-31. This standard covers recommended design guidelines for rejection of educational FM interference to broadcast Channel 6 television reception. This tuner incorporates a narrow band-reject filter to prevent interference to the Channel 6 audio carrier. This compliance applies to products for off-air reception of broadcast television signals, such as television receivers and VCRs. Cable television receivers and decoders are an exception item to the specification, as long as off-air reception is not a customer capability. Contact Global Engineering Documents, as noted in the appendix, for a copy of the specification.

VTSS6USZF1 Tuner

This tuner is the same as the VTSS6USZF with the exception of EIA/IS-31 compliance. The band-reject filter is not incorporated in this device. The VTSS6USZF1 is used in applications where EIA/IS-31 compliance is not required.

VTSS6USFF Tuner

This tuner is the same device as VTSS6USZF, with a different input connector. A standard "F"-series threaded connector is used in place of the RCA connector on the original model.

VTSS7USZF/1 Tuner

This model variation is the same functional tuner as VTSS6USZF, with the tuner supply voltage changed from +9 V to +12 V. This series of tuners is used in applications where a +12 V supply is either available, or preferred. The two models provide EIA/IS-31 compliance (VTSS7USZF) and no filter (VTSS7USZF1).

RFIVU3 DEMODULATOR

The RFIVU3 is an integrated demodulator assembly for use with the VTSS-series tuners. The RFIVU3 replaces the earlier model number RIFU0508GEZZ. In all respects it has the same functionality.

The input to the demodulator is the combined IF signal from VTSS-series tuners. The outputs are demodulated video and audio signals. An AFT output is provided for monitoring the frequency tuning point of the tuner, with a 7 V AFT range provided. An AFT Mute function is incorporated for control of the output video and audio signals. The AGC output is provided for direct control of the AGC input in the selected tuner. The demodulator operates from a +9 V source. In application, it is necessary to align the tuner and demodulator as a set for optimum operation. This demodulator may also be used with any tuner that provides a standard 45.75 MHz IF output.

DUNTR0104 MATCHED PAIR

As an obvious next step, a tuner and demodulator can be paired to provide a set of modules. The DUNTR0104 is a matched set of VTSS6USZF1 and RFIVU3 functional blocks. Both units are aligned together at the factory and are intended to be used as a matched set of drop-in modules where the tuner-demodulator function is required.

The tuner-demodulator combination shown in Figure 1 demonstrates VTSS6USZF and RFIVU3 connected together. The DUNTR0104 pair would be connected in like manner. The IF signal line comes out of the tuner and into the demodulator. The demodulator monitors the IF signal level out of the tuner and provides feedback by way of the AGC return line. If the input to the tuner is operated across the correct range, the tuner and demodulator will work together. The AFT output may be monitored for tuning accuracy on the currently tuned channel. The AFT Voltage will provide an indication of whether the tuner is on channel, above the desired frequency, or below the desired frequency. The AFT Error Voltage is determined as a difference Voltage from the nominal 4.5 V AFT center point. The AFT voltage is scaled to provide an estimate of how far off-channel the tuner is positioned.

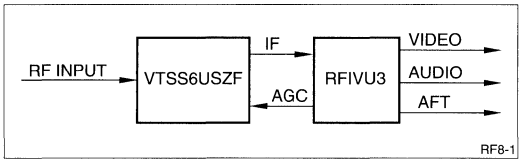


Figure 1. VTSS6USZF and RFIVU3

E6095 MODULATOR

Many RF applications require that baseband video and audio signals be modulated for transmission to television receivers. The E6095 modulator provides a standard NTSC-M modulated output on either broadcast Channel 2 or 3. The output channel is electrically controlled by an external interface pin, asserting 0 V or 5 V. The modulator operates from a +5 V supply.

The four input connections are +5 V, video, audio, and the switch control for selecting the desired channel. The modulated RF output signal is provided through an "F"-series threaded connector.

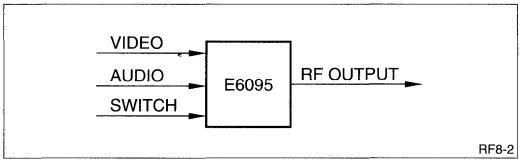


Figure 2. E6095 Modulator

E7853 MODULATOR

The E7853 provides the same general functional capabilities as the E6095, with the addition of an "Antenna In" connection. This allows for the connection of the modulator into a system where the incoming antenna or cable signal may be passed to the output. Control of the RF pass-through capability is provided electrically by a pin on the modulator.

The RF pass-through has a frequency passband of 55 MHz to 890 MHz, providing capability for most applications. The modulator is rated for a 6.5 dB maximum insertion loss. Isolation of 60 dB minimum between the antenna input and the output port is provided.

The RF output from the E7853 is selected between standard broadcast NTSC-M channels 3 and 4. The selection is made by a mechanical switch on the modulator housing. The E7853 also incorporates a "Tuner Out" port. This allows for the E7853 to be the complete system interface with the outside world. A

sample of the RF Input signal is coupled to the Tuner Out port at all times. In addition, if used in a cable television "set-top-converter", the input signals are provided on the RF Output, returning all television or VCR capabilities to the customer. The RF pass-through feature passes the input signals to the RF Output until the modulator function is selected.

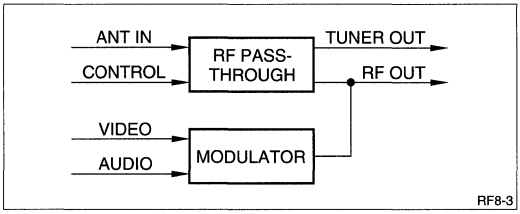


Figure 3. E7853 Modulator

RFSP7US1 COMBINATION TUNER/DEMOD

This unit is a combination tuner and demodulator in one package. It incorporates the functions of VTSS6USZF and RFIVU3, or DUNTR0104. The PLL is the TD6380P, similar to the device used in VTSS6USZF. An IF sample port is provided for monitoring the signal level or frequency characteristics of the recovered IF signal. Audio, video, and AFT outputs are also provided. The AFT Mute function is available as well.

The AGC control line from the demodulator to the tuner is not accessible to the system designer. If AGC or complete IF access is required, refer to the RFSN7US2 description below. Audio, video, and AFT interface circuits for test purposes are described in the full RFSP7US1 specification, and the Application Note covering internal PLL tuners.

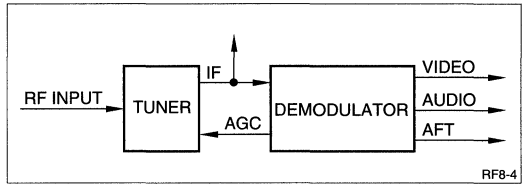


Figure 4. RFSP7US1 Combination Tuner and Demodulator

RFSO7US3 COMBINATION TUNER/DEMOD/MOD

A modulator similar to E7853 has been added to the RFSP7US1 package. The RF pass-through function is retained, so that maximum system design potential is available. The two RF connectors are mounted on the end of the unit to allow for a mounting position that requires minimum panel area.

The modulator section has separate inputs so that reception and transmission features may operate independently. The modulator output is electrically controlled via an external interface pin. Output channel 3 or 4 is selected electrically through an interface pin. The internal tuner sample point is always active, allowing the tuner and demodulator functions to be used at all times, regardless of the state of the RF pass-through and the modulator output.

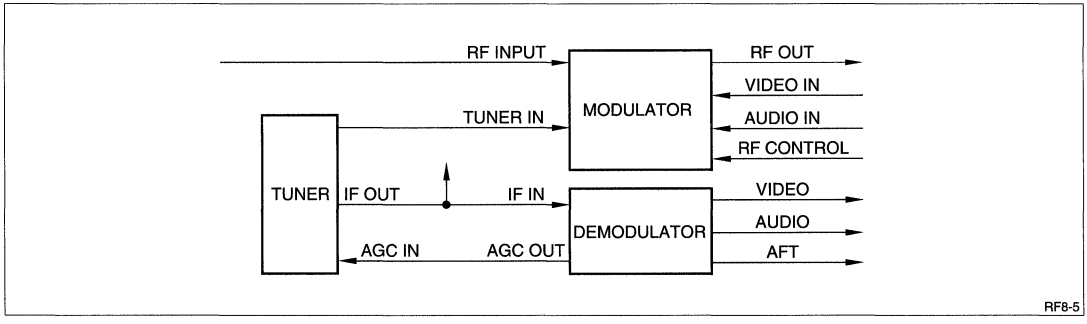


Figure 5. RFSO7US3 Combination Tuner, Demodulator, and Modulator

RFSN7US2 COMBINATION TUNER/DEMOD/MOD

The RFSN7US2 is similar to the RFSO7US3 in general concept. It differs in that all three functional blocks are completely separate at the electrical interface, except for the tuner RF input. There are no internal connections between the tuner and demodulator blocks. This configuration is important when the AGC or IF signals must be manipulated for a particular purpose, such as descrambling a signal, or special compensation circuits. The RF connectors are mounted on the side of the housing instead of the end of the unit. This allows for cable connections internal to the set-top converter with directional couplers and other devices, when there is no direct requirement for the tuner unit to be panel mounted.

The modulator section has separate inputs so that reception and transmission features may operate independently. The modulator output is electrically controlled via an external interface pin. Output channel 3 or 4 is selected by a mechanical switch on the housing. The internal tuner sample point is always active, allowing the tuner and demodulator functions to be used at all times, regardless of the condition of the RF pass-through and the modulator output.

CAJA35U01 DOUBLE CONVERSION TUNER

The CAJA35U01 is currently in a preliminary status. It was designed for standard NTSC broadcast CATV applications. The internal bandpass filter has a nominal 6 MHz 3 dB bandwidth with low ripple and phase error. A tuning range of 50 MHz to 550 MHz is accomplished with the addition of an external PLL device. A standard 45.75 MHz IF output is produced. An internal divide-by-128 prescaler provides a fixed divider value for the First Local Oscillator. The Second Local Oscillator provides access for AFT control with approximate scaling of 250 kHz/V. The block diagram for the CAJA35U01 is shown in Figure 7.

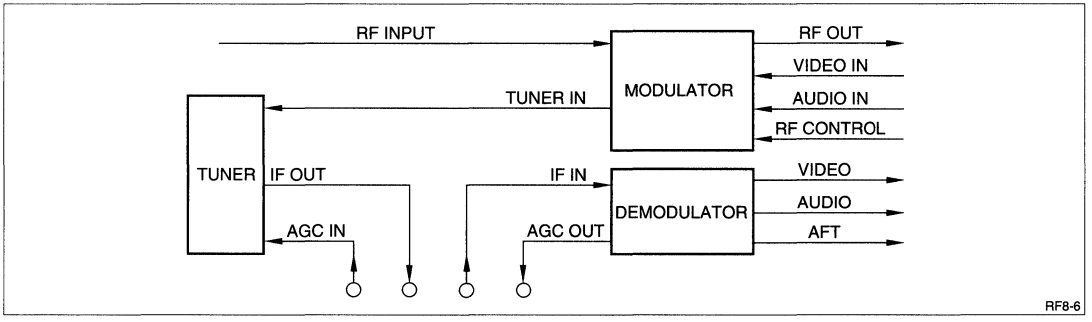


Figure 6. RFSN7US2 Combination Tuner, Demodulator, and Modulator

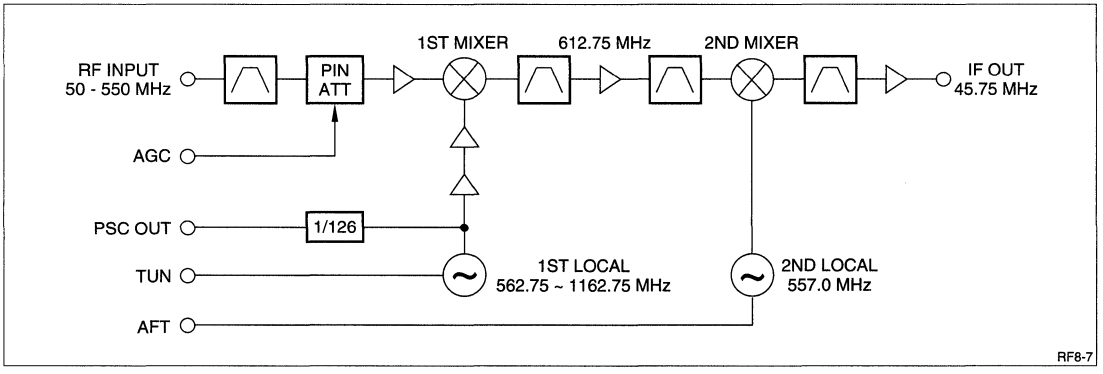


Figure 7. CAJA35U01 Block Diagram

VTSS6USZF-SERIES SPECIFICATIONS

ITEM	SPECIFICATION
Receiving System	USA standard NTSC-M
Frequency Coverage	55.25 MHz to 801.25 MHz, referenced to video carrier
Input Impedance	75 Ω, unbalanced
IF Output Load Impedance	75 Ω
Intermediate Frequency	Video IF: 45.75 MHz, Audio IF: 41.25 MHz
Nominal Supply Voltages	B: +9 V, 80 mA; BT: +33 V, 3 mA; AGC: +6.3 V, 0.5 μA; BP: +5 V, 85 mA

VTSS6USZF-SERIES ELECTRICAL CHARACTERISTICS

ITEM	TYPICAL	LIMITS
Antenna Input VSWR	1.5	4.0 max.
Power Gain		
AIR	35 dB	25 ~ 44 dB
CATV	32 dB	22 ~ 44 dB
UHF	32 dB	25 ~ 44 dB
Gain Deviation Between Channels		
VHF	8 dB	13 dB
UHF	6 dB	11 dB
AGC Dynamic Range		
VHF	50 dB	40 dB min.
UHF	45 dB	30 dB min.
Image Rejection, Nominal Gain		
AIR	70 dB	55 dB min.
VHF	60 dB	50 dB min.
UHF	60 dB	45 dB min.
Image Rejection, AGC Attenuation = –30 dB		
AIR	60 dB	50 dB min.
VHF	50 dB	40 dB min.
UHF	50 dB	35 dB min.
IF Rejection, Nominal Gain		
VHF Low	80 dB	55 dB min.
VHF High	90 dB	60 dB min.
UHF	85 dB	60 dB min.
IF Rejection, AGC Attenuation = –30 dB		
VHF Low	60 dB	50 dB min.
VHF High	70 dB	50 dB min.
UHF	65 dB	50 dB min.
Local Oscillator Frequency Drift, 25°C ± 25°		
VHF	±15 kHz	±50 kHz
UHF	±15 kHz	±50 kHz
Local Oscillator Frequency Drift, 1 Hour Warmup		
VHF	±15 kHz	±50 kHz
UHF	±15 kHz	±50 kHz
PLL Stability	±15 kHz	±50 kHz
Tuning Voltage Ripple	5 mVp-p	10 mVp-p

VTSS7USZF-SERIES SPECIFICATIONS

ITEM	SPECIFICATION
Receiving System	USA standard NTSC-M
Frequency Coverage	55.25 MHz to 801.25 MHz, referenced to video carrier
Input Impedance	75 Ω, unbalanced
IF Output load impedance	75 Ω
Intermediate Frequency	Video IF: 45.75 MHz, Audio IF: 41.25 MHz
Nominal Supply Voltages	B: +12 V, 80 mA; BT: +33 V, 3 mA; AGC: +6.3 V, 0.5 μA; BP: +5 V, 85 mA

VTSS7USZF-SERIES ELECTRICAL CHARACTERISTICS

ITEM	TYPICAL	LIMITS
Antenna Input VSWR	1.5	4.0 max.
Power Gain		
AIR	35 dB	25 ~ 44 dB
CATV	32 dB	22 ~ 44 dB
UHF	32 dB	25 ~ 44 dB
Gain Deviation Between Channels		
VHF	8 dB	13 dB
UHF	6 dB	11 dB
AGC Dynamic Range		
VHF	50 dB	40 dB min.
UHF	45 dB	30 dB min.
Image Rejection, Nominal Gain		
AIR	70 dB	55 dB min.
VHF	60 dB	50 dB min.
UHF	60 dB	45 dB min.
Image Rejection, AGC Attenuation = –30 dB		
AIR	60 dB	50 dB min.
VHF	50 dB	40 dB min.
UHF	50 dB	35 dB min.
IF Rejection, Nominal Gain		
VHF Low	80 dB	55 dB min.
VHF High	90 dB	60 dB min.
UHF	85 dB	60 dB min.
IF Rejection, AGC Attenuation = –30 dB		
VHF Low	60 dB	50 dB min.
VHF High	70 dB	50 dB min.
UHF	65 dB	50 dB min.
Local Oscillator Frequency Drift, 25°C ± 25°		
VHF	±15 kHz	±50 kHz
UHF	±15 kHz	±50 kHz
Local Oscillator Frequency Drift, 1 Hour Warmup		
VHF	±15 kHz	±50 kHz
UHF	±15 kHz	±50 kHz
PLL Stability	±15 kHz	±50 kHz
Tuning Voltage Ripple	5 mVp-p	10 mVp-p

RFIVU3 SPECIFICATIONS

ITEM	SPECIFICATION
Receiving Standard	USA Standard System NTSC-M, 55.25 MHz to 801.25 MHz
Receiving System	Intercarrier Sound, 4.5 MHz
Intermediate Frequency	Picture IF: 45.75 MHz Sound IF: 41.25 MHz
Nominal Input Impedance	IF Input Terminal: 75 Ω unbalanced. (Shall be connected to IF Output Terminal of tuner.)
Supply Voltage	DC 9 V AFT Mute On 0 V, Off Open
Terminal Output Voltages	RF AGC Out 8.0 V (8.0 ~ 0 V) AFT Out 4.5 V (1.0 ~ 8.0 V)
Operating Voltages	B Voltage DC 9 V ± 10% AFT Mute On 0.3 V max.
Consumed Current	65 mA max.
Weight	24 ± 10 Grams
Temperature	Storage: –25°C ~ +75°C Operating: –10°C ~ +60°C

RFIVU3 ELECTRICAL CHARACTERISTICS

ITEM	UNITS OR RANGE	TYPICAL	LIMITS	REMARKS
Selectivity Characteristics				IF AGC 20 dB, CH 10
Picture Carrier Level	fp	6 dB	6 ±3 dB max.	See complete specification.
Sub-Carrier Level	fc	6 dB	6 ±3 dB max.	
Adjacent Sound Carrier Level	Adj.fs	50 dB	40 dB min.	
Adjacent Picture Carrier Level	Adj.fp	45 dB	40 dB min.	
Sound Carrier level	fs	21 dB	21 ±4 dB max.	
Picture Output Level				Standard Color Bars
Brightness Signal Level		1 Vp-p	1 ±0.2 Vp-p max.	87.5% Modulation. See complete specification.
V/S Ratio		7.15/2.85	6.9/3.1 ~ 7.4/2.6 max.	
Differential Gain		±8%	±12% max.	Stair-step, CH 10, 80% Modulation, 90 dBμV
Differential Phase		±6°	±10° max.	Stair-step, CH 10, 80% Modulation, 90 dBμ
Picture Output S/N Ratio		42 dB	41 dB min.	Weight filter OFF, White Level 50%, 90 dBμ
AFT Output Voltage		4.5 V	1.0 ~ 8.0 V	
Local Oscillator Frequency Error		±25 kHz	±50 kHz max.	At AFT Voltage 4.5 V
Picture Frequency Characteristics	100 kHz ~ 3 MHz	±1.5 dB	–1 ±3 dB max.	Multi-burst 100 kHz standard 80% Modulation, 90 dBμ
	3.58 MHz	–3.0 dB	–3 ±4 dB max.	
Sound S/N Ratio		50 dB	42 dB min.	Standard Color Bars, 87.5% Mod., 20 kHz, ±25 kHz dev., 400 Hz standard.

DUNTR0104A SPECIFICATIONS

This specification is a combination of VTSS6USZF1 and RFIVU3. Please refer to those specifications for preliminary details, or request complete specifications for full information.

E6095 SPECIFICATIONS

ITEM	SPECIFICATION
Transmission System	USA Standard NTSC-M
Output Channel	Channel 2, 3
Output Impedance	75 Ω, unbalanced
Power Requirement	+5 V

E6095 ELECTRICAL CHARACTERISTICS

ITEM	SPECIFIED VALUE	REMARKS
Video Input Impedance	1 kΩ, ±30%	Measure at 0 ~ 4.2 MHz
Input Signal Level	1 Vp-p	82 Ω load, negative sync signal
Modulation Factor	78 ±7%	82 Ω load
V/S Ratio	7, +0.25, −0 3, +0, −0.25	Input: Staircase 1Vp-p V/S = 7/3, negative sync
Amplitude/Frequency Characteristics	+2, −2 dB max.	See Specification
Differential Gain	8% or less	See Specification
Differential Phase	8° or less	See Specification
Modulation Variation With Respect to APL	Within ±3%	Measure modulation variation over a range of 10 ~ 90% APL with respect to 50% APL
S/N Ratio	45 dB (p-p/rms)	Measure with respect to standard demodulator output
Audio Input Impedance	50 kΩ or more, unbalanced	Measure at 0.1 ~ 10 kHz
Input Signal Level	−8.7 dBs	0.8 Vp-p sinewave
Amplitude Frequency Response	+2, −2 dB (100 Hz ~ 10 kHz) +3, −3 dB (11 kHz ~ 100 kHz)	Measure deviation from 100 Hz to 100 kHz with 1 kHz as reference
Modulation	100 ±20%	100% ±25 kHz
Distortion	2% or less	
S/N	45 dB or more	Including buzz
Video Carrier Frequency	Within 100 kHz	Test at 25°C and 65% RH
Audio Carrier Frequency	Within 4,500 Hz, ± 5 kHz	Test at 25°C and 65% RH
Audio Carrier Difference Level	Within 16 ±3 dB	Difference between video and audio output levels, audio unmodulated
Spurious Response Within Bandwidth	60 dB or less	Between video and audio carriers
920 kHz Beat	60 dB or less	See Specification
Output VSWR	Within 3 dB	
Current Consumption @ +5 V	20 mA or less	

E7853 SPECIFICATIONS

ITEM	SPECIFICATION
Transmission System	USA Standard NTSC-M
Output Channel	Channel 3, 4
Output Impedance	75 Ω, unbalanced
Power Requirement	+5 V

E7853 ELECTRICAL CHARACTERISTICS

ITEM	SPECIFIED VALUE	REMARKS
Video Input Impedance	1 kΩ ±30% unbalanced	Measure at 0 ~ 4.2 MHz
Input Signal Level	1 Vp-p	82 Ω load, negative sync signal
Modulation Factor	77% ±7%	82 Ω load
V/S Ratio	7, +0.25, −0 3, +0, −0.25	Input: Staircase 1 Vp-p V/S = 7/3, negative sync
Amplitude/Frequency Characteristics	+2, −3 dB max.	See Specification
Differential Gain	10% max.	See Specification
Differential Phase	10° max.	See Specification
Modulation Variation With Respect to APL	±5% max.	Measure modulation variation over a range of 10 ~ 90% APL with respect to 50% APL
S/N ratio	45 dB (p-p/rms)min.	Measure with standard demodulator output.
Audio Input Impedance	30 kΩ min., unbalanced	Measure at 0.1 ~ 10 kHz
Input Signal Level	1.24 Vp-p	
Amplitude/Frequency Characteristics	+2, −3 dB	Using 1 kHz as a standard in the 100 Hz ~ 10 kHz range, measure deviation between reference.
Modulation Factor	100 ±20%	100% = ±25 kHz
Distortion Rate	3% max.	
S/N ratio, Including Buzz	45 dB min.	
RF Switch Frequency Band	55 MHz to 890 MHz	
RF Switch Insertion Loss, ANT IN to TV OUT	6.5 dB max.	Measure 55 MHz to 890 MHz
Distribution Loss, ANT IN to TUN OUT	6 dB max.	Measure 55 MHz to 890 MHz
RF Switch Isolation	60 dB min.	

RFSP7US1 SPECIFICATIONS

ITEM	SPECIFICATION
Receive System	USA standard NTSC-M
Frequency Coverage	181 Channels, 55.25 MHz to 801.25 MHz
Input Impedance	75 Ω, unbalanced
Output Impedance	Video out: 1 kΩ, Audio out: 4.7 kΩ
Intermediate Frequencies	Video carrier: 45.75 MHz, Audio carrier: 41.25 MHz
Nominal Supply Voltages	BT: +31 V, 5 mA max.; BP: +5 V, 86 mA max.; B: +12 V, 190 mA max.

RFSP7US1 ELECTRICAL CHARACTERISTICS

ITEM	TYPICAL	LIMITS	REMARKS
Noise Figure	8 dB	13 dB	
Image Rejection, −47 dBm Input			
VHF AIR	70 dB	60 dB min.	
VHF CATV	60 dB	50 dB min.	
UHF	60 dB	45 dB min.	
Image Rejection, −17 dBm Input			
VHF AIR	60 dB	50 dB min.	
VHF CATV	50 dB	40 dB min.	
UHF	50 dB	40 dB min.	
IF Rejection, −47 dBm Input			
VHF Low	80 dB	55 dB min.	
VHF High	90 dB	60 dB min.	
UHF	85 dB	60 dB min.	
IF Rejection, −17 dBm Input			
VHF Low	60 dB	45 dB min.	
VHF High	70 dB	50 dB min.	
UHF	65 dB	40 dB min.	
Citizen's Band Rejection, S/I	50 dB	40 dB min.	See Specification
Picture Output Level	2 Vp-p	2 ±0.4 Vp-p	Channel 10, 1 kΩ termination, fp 70 dBμV, 100% White
Differential Gain	3%	16% max.	Channel 10, fp 90 dBμV, 80 IRE stairstep
Differential Phase	3°	16° max.	Channel 10, fp 90 dBμV, 80 IRE stairstep
S/N Ratio	48 dB	43 dB min.	fp 70 dBμV, White 100%, 100 kHz ~ 4.2 MHz filter
Frequency Characteristics			Channel 10, fp 70 dBμV, Mulitburst
1.0 MHz	−0.5 dB	−3 ~ +2 dB	
2.0 MHz	−0.5 dB	−3 ~ +2 dB	
3.0 MHz	−1.0 dB	−6 ~ +0.5 dB	
3.58 MHz	−2.3 dB	−6 ~ +0.5 dB	
Synchronization Ratio	28.5%	23.6 ~ 33.6%	Channel 10, fp 70 dBμV, SMPTE Colorbar
Sound Output Level	250 mVrms	180 ~ 320 mVrms	See Specification
AFT Output			See Specification
Voltage	0.5 ~ 11.5 V	1.0 ~ 11.0 V	
Frequency Accuracy	25 kHz	100 kHz max.	

RFSO7US3 SPECIFICATIONS

ITEM	SPECIFICATION
Receiving Channels	USA Standard System NTSC-M, 55.25 MHz to 801.25 MHz
Output Channels	USA NTSC-M Channels 3 and 4, 61.25 MHz, 61.25 MHz
Intermediate Frequency	Picture IF: 45.75 MHz, Audio IF: 41.25 MHz
Nominal Input and Output Impedance	75 Ω
Tuning System	Phase-Locked-Loop, Toshiba® TD6382N or equivalent.
Supply Voltages	+5 V, +12 V, +33 V
Consumed Current	320 mA total, see specification
Temperature	Storage: –20°C ~ +75°C Operating: –10°C ~ +60°C

RFSO7US3 ELECTRICAL CHARACTERISTICS

ITEM	TYPICAL	LIMITS	REMARKS
RF Output, Modulated			Input 1Vpp with 1KΩ load.
Modulation Factor	80%	80 ±5%	
Differential Gain	2%	7% Max.	
Differential Phase	1.5°	8° max.	
S/N Ratio	55 dB	48 dB min.	
Frequency Characteristics	0 dB	0 ±3 dB	
Change in Modulation Factor to APL	1%	3% max.	
RF Output, Audio System			Input 1 kHz 1.24 Vp-p sine-curve.
Modulation Factor	±25.0 kHz	±25.0 kHz, ±5.5 kHz	
Distortion Rate	0.3%	2% max.	
S/N Ratio	56 dB	48 dB min.	
Frequency Characteristics	0 dB	0 ±3 dB	
RF Output			Measured Sync level at white 100% signal, input 1Vpp.
Video Carrier Center Frequency Accuracy	±10 kHz	±100 kHz max.	
Audio Carrier Center Frequency Accuracy	4.5 MHz	4.5 MHz ±7 kHz	
Video Carrier Output Level	66 dBμV	66 ±3 dBμV	66 dBμV = 6 dBmV
Video/Audio Ratio	16 dB	16 ±3 dB	
RF Output Spurious			
Specific Frequency	70 dB	60 dB min.	fp ~ fp + 4.5 MHz
Other Frequencies	50 dB	30 dB min.	0 ~ 1 GHz, except fp ±4.6 MHz
RF Switch			
Insertion Loss	3.0 dB	6.5 dB max.	55 ~ 806 MHz
Isolation	70 dB	60 dB min.	61 ~ 72 MHz
Return Loss	8.0 dB	4 dB min.	61 ~ 72 MHz
Image Rejection			at –47 dBm input (+1.75 dBmV)
VHF AIR	70 dB	60 dB min.	
VHF CATV	60 dB	50 dB min.	
UHF	60 dB	45 dB min.	

RFSN7US2 SPECIFICATIONS

ITEM	SPECIFICATION
Receiving Channels	USA Standard System NTSC-M, 55.25 MHz to 801.25 MHz
Output Channels	USA NTSC-M Channels 3 and 4, 61.25 MHz, 61.25 MHz
Intermediate Frequency	Picture IF: 45.75 MHz, Audio IF: 41.25 MHz
Nominal Input and Output Impedance	75 Ω
Tuning System	Phase-Locked-Loop, Toshiba® TD6382N or equivalent
Supply Voltages	+5 V, +12 V, +33 V
Consumed Current	320 mA total
Temperature	Storage: –20°C ~ +75°C Operating: –10°C ~ +60°C

RFSN7US2 ELECTRICAL CHARACTERISTICS

ITEM	TYPICAL	LIMITS	REMARKS
RF Output, Modulated			Input 1Vpp with 1 kΩ load.
Modulation Factor	77.5%	77.5 ±5%	
Differential Gain	2%	7% Max.,	
Differential Phase	2°	8° max.	
S/N Ratio	55 dB	48 dB min.	
Frequency Characteristics	0 dB	0 ±3 dB	
Change in Modulation Factor to APL	1%	3% max.	
RF Output, Audio System			Input 1 kHz 1.24 Vp-p sine-curve.
Modulation Factor	±25.0 kHz	±25.0 kHz, ±5.5 kHz	
Distortion Rate	0.3%	2% max.	
S/N Ratio	56 dB	48 dB min.	
Frequency Characteristics	0 dB	0 ±3 dB	
RF Output			Measured Sync level at white 100% signal, input 1Vpp.
Video Carrier Center Frequency Accuracy	±10 kHz	±100 kHz max.	
Audio Carrier Center Frequency Accuracy	4.5 MHz	4.5 MHz ±7 kHz	
Video Carrier Output Level	66 dBμV	66 ±3 dBμV	66 dBμV = 6 dBmV
Video/Audio Ratio	16 dB	16 ±3 dB	
RF Output Spurious			
Specific Frequency	70 dB	60 dB min.	fp ~ fp + 4.5 MHz
Other Frequencies	50 dB	30 dB min.	0 ~ 1 GHz, except fp ±4.6 MHz
RF Switch			
Insertion Loss	3.0 dB	6.5 dB max.	55 ~ 806 MHz
Isolation	70 dB	60 dB min.	61 ~ 72 MHz
Return Loss	8.0 dB	4 dB min.	61 ~ 72 MHz
Image Rejection			at –47 dBm input (+1.75 dBmV)
VHF AIR	70 dB	60 dB min.	
VHF CATV	60 dB	50 dB min.	
UHF	60 dB	45 dB min.	

CAJA35U01 SPECIFICATIONS

ITEM	SPECIFICATION
Receiving Systems	NTSC-M, Standard, HRC, IRC
Receiving Frequency Range	50 MHz to 550 MHz
Input Level	–55 dBm to –30 dBm; –6.25 dBmV to +18.75 dBmV
Nominal Input/Output Impedance	75 Ω
Output Frequencies	Picture Carrier: 45.75 MHz Sound Carrier: 41.25 MHz
First Local Oscillator Frequency	662.75 MHz to 1162.75 MHz
Second Local Oscillator Frequency	567 MHz

CAJA35U01 ELECTRICAL CHARACTERISTICS

ITEM	MIN.	TYPICAL	MAX.	UNITS	CONDITION
Input Return Loss	5	8		dB	On channel selected
Output Return Loss	8	12		dB	Between P and S carrier
Conversion Gain	24	28		dB	RF Input to IF Output
Noise Figure		8	11.5	dB	
Intermodulation i) 2nd Order/3rd Order Beats Standard 82 Channels, –42 dBm Input ii) Composite Triple Beat Standard 82 Channels, –42 dBm Input			 –57 –57	 dB dB	
Cross Modulation, Standard 82 Channels, –42 dBm Input			–57	dB	
Local Oscillator Leakage at Input Terminal			–50	dBm	1st Local Oscillator fundamental
Local Oscillator Leakage at Output Terminal			–40	dBm	2nd Local Oscillator fundamental
Spurious Signals at Input Terminal			–75	dBm	50 MHz to 550 MHz
Spurious Signals at Output Terminal			–75	dBm	41.25 MHz to 45.75 MHz
Output Bandpass Response i) P and S Carrier Level From Peak: P Carrier S Carrier ii) Bandwidth at 3 dB Down	 4.5	 6.0	 –2 –2 9.0	 dB dB MHz	The Second Local Oscillator is adjusted so that the Output P carrier is within 50 kHz of nominal frequency.
Current Consumption		80 55 20 1	120 80 30 3	mA mA mA mA	B1 (+9 V) B2 (+9 V) Bp (+5 V) TUN
AGC Voltage	0		8	V	
AFT Voltage	1.5	4.5	7.5	V	
Prescaler Output Voltage	0.6	1.0		Vp-p	AC coupled

APPENDIX

Global Engineering Documents
2805 McGaw Avenue
Irvine, CA 92714
Phone: 714-261-1455
Fax: 714-261-7892
(800) 854-7179

For complete component specifications, please contact your nearest Sharp office and request the following specifications by model number:

- VTSS6USZF
- VTSS6USZF1
- VTSS6USFF
- VTSS7USZF
- VTSS7USZF1
- RFIVU3
- DUNTR0104
- E6095
- E7853
- RFSP7US1
- RFSO7US3
- RFSN7US1
- CAJA35U01

VTSS, RFSO/SP Series
Internal-PLL Electronic Tuners

For Sharp Products: VTSS, RFSO/SP Series

Robert Stuart, RF Applications Engineer

OVERVIEW

This Application Note covers several different Sharp product series that utilize Toshiba® TD-Series Phase-Locked-Loop (PLL) Integrated Circuits (ICs), and the Sanyo® LA7530 Demodulator IC. All of the Sharp products covered here are useful for Television, Video Cassette Recorders, Multimedia, or other similar applications.

INTRODUCTION

Three Sharp product series use different members of the Toshiba TD-series of PLL ICs. These tuners operate in similar manner, with variations dependent on the PLL used. All the PLL devices are similar as well, with the internal frequency step dependent on the version of IC used. Sharp RFSO and RFSP-series tuners use the Sanyo LA7530 Demodulator. Background information for this device is included here, clarifying the material provided in the specification sheets.

VTSS-Series

The VTSS-series of tuners are PLL-tuned and cover the VHF and UHF frequency bands. A total of 181 standard North American NTSC-M channels are within the tune range of the VTSS6USZFx and VTSS7USZFx model lines. Channels in the Standard, HRC, and IRC frequency plans may be tuned with these products. Other non-PLL tuners are available, though are not addressed here. The Sanyo LA7530N is not used in VTSS-series products as they are tuner modules only.

RFSP/RFSO-Series

These products include the RFSP and RFSO series of tuner products. RFSP-series products are combination tuner-demodulator units that cover the NTSC-M North American 181 channel broadcast range. The RFSO-series of products are tuner-demodulator-modulator products that include all RFSP-series capabilities and include a modulator for signal output capability. In addition, all Standard, HRC, and IRC channels may be tuned.

TOSHIBA® TD-SERIES PLL ICs

A variety of Toshiba TD-series PLL ICs are used in Sharp electronic tuner products. The TD6359P, TD6380P, and TD6382P are similar in function and programming, yet have slightly different characteristics. Table 1 provides information about the three ICs, and where each is used.

TD6359P/TD6380P

The TD6359P and TD6380P use an 18-bit serial data string for configuring the PLL. The Toshiba PLL specifications provide an expression for computing the oscillator lock frequency:

$$(1) \quad f_{OSC} = f_R \times 8\ ((32 \times M) + S)$$

The (1) at the beginning indicates that it is equation 1 in this document. The Voltage Controlled Oscillator (VCO) output frequency term f_{OSC} is determined by the output from the PLL IC. The internal Reference Frequency, f_R, is the reference crystal frequency, which in this case is 4 MHz. The PLL specification provides the expression f_R = crystal osc./512. However, $f_R \times 8$ is equal to the step frequency of the tuner, for TD6359P and TD6380P. By substituting step frequency in place of $f_R \times 8$, the expression simplifies to:

$$(2) \quad f_{OSC} = [\text{step freq}] \times ((32 \times M) + S)$$

M is the Main Counter value loaded into the PLL, and S is the Swallow-count value. As an example of the above expression, we will tune the VCO in a tuner using the TD6359P, to 101 MHz. The required VCO frequency to tune NTSC-M Channel 2, which is at 55.25 MHz, referenced to the video carrier, is 101 MHz. The Intermediate Frequency (IF) output for these tuners is 45.75 MHz. As part of the conversion process, the VCO is the conversion Local Oscillator (LO), and is tuned above the desired frequency by an

Table 1.
PLL Specifications

IC	CRYSTAL REFERENCE FREQUENCY	STEP	VTSS	RFSP	RFSO
TD6359P	4 MHz	62.5 kHz	X		
TD6380P	4 MHz	62.5 kHz		X	X

amount equal to the IF. Thus 101 MHz – 55.25 MHz = 45.75 MHz produces the desired IF output frequency. For a given channel frequency, add 45.75 MHz, and the required VCO frequency will result.

The maximum M count possible is 511, and the maximum S count is 31. One of the most straightforward ways to derive the PLL constants M and S, is to refer to equation 2, repeated below:

$$f_{OSC} = [\text{step freq}] \times ((32 \times M) + S)$$

The range for S is 0 to 31, and has the least impact on the overall count value. Therefore, set S = 16, which is a midrange value. By multiplying [step freq] through the parentheses, we can rearrange and solve for M:

$$(3) \quad \frac{f_{OSC} - [\text{stepfreq}] \times S}{[\text{step freq}] \times 32} = M$$

From Table 1, the value of [step freq] for the TD6359P and TD6380P is 62.5 kHz. Using equation 3 will provide a value for M = 50.

$$3a) \quad \frac{101\ \text{MHz} - [62.5\ \text{kHz}] \times 16}{[62.5\ \text{kHz}] \times 32} = 50$$

Putting this M-value back into equation 2 and using S = 16 will produce: f_{OSC} = 101 MHz, as desired.

For the TD6359P and TD6380P PLL ICs, the use of S = 16 is a good start. The M-count value is the only count that needs to be modified to change the tuned channel frequency. Fine adjustment of the tuned frequency is effected by changing the S-count value. Refer to the section on AFT for further information.

To send this information to the PLL IC, it is combined with 4 bits of Band Switch data. The PLL IC automatically controls the three different tuning bands VHF Low, VHF High, and UHF by responding to the Band Switch bits in the serial data string. The Band bits are provided in the separate tuning tables for each tuned channel.

To convert the M and S values to the serial data string, the expression 32 M + S is used. For channel 2 at 55.25 MHz, M = 50 and S = 16. In the serial data string, the Swallow counter S is allocated 5 bits, bits 0 through 4, at the end. The S value of 16 decimal converts to 10 Hex, which is 10000 binary. The last 0 occupies the Least Significant Bit (LSB) of the data string.

The M value of 50 is allocated 9 bits, in positions 5 through 13. The value of 50 must be multiplied by 32, providing the result 1,600. The Hex value is 640, which must be ORed with the S value to create the final value binary value: 00011001010000.

The Band Switch bit configuration for channel 2 is 0001. This value is added to the Most Significant Bit (MSB) end of the data string. The process is shown in Table 2.

Table 2.
Combined Values

DECIMAL	HEX	BINARY
S 16	10	10000
Mx32 = 1,600	640	011001000000
Combined Value	650	011001010000
Add 2 Bits	0	00011001010000
Band Bits (4)	1	000100011001010000

As only twelve bits are shown as a result of this combined value and 14 bits are allowed, bits 12 and 13 must be filled in with zeroes. In the case where the M and S values generate a larger number, these bits will have values and the leading zeroes not added. It is essential to make sure that the entire fourteen bits are captured correctly.

CHECKING YOUR NUMBERS

Had we not chosen to pick S = 16, we would need to take a different approach. From equation 2, dividing both sides by [step freq]:

$$(4) \quad \frac{f_{OSC}}{[\text{step freq}]} = ((32 \times M) + S)$$

Following through,

$$(4a) \quad \frac{101\ \text{MHz}}{62.5\ \text{kHz}} = ((32 \times M) + S) = 1{,}616$$

The expression ((32 × M) + S) is equated to 1,616. At this point, some value must be assumed for either S or M. Again, selecting S = 16 will provide the previous result. Due to M being multiplied by 32, one more than the maximum value of the S-counter, increasing or decreasing M by 1 will be equivalent to changing S by 32. Therefore, the approach to getting the correct count value is more dependent on M, than S. Dividing 1,616 by 32 provides M and the decimal fraction of S/32, as shown

$$(5) \quad 32M + S = 1{,}616 = M + \frac{S}{32} = 50.5$$

The expression M + S/32 can be viewed as an integer plus a decimal fraction, referenced to 32. Thus M = 50, and S/32 = 0.5. Multiplying both sides of S/32 = 0.5 by 32 provides S = 16. This is the same result as determined from equation 3, where S was chosen to be 16. Other values may be chosen, however they may prove to be awkward in algorithms to determine the required tune frequency values.

Tables of programming information for the TD6359P and TD6380P are available, listing the M and S counts for most channel frequencies, and the corresponding VCO frequency.

TD6359P/TD6380P Algorithm Shortcut

The TD6359P and TD6380P both incorporate a 32-bit divider as part of the internal prescaler. In this case, the value of 32 directly corresponds to the 32 in the expression:

$f_{OSC} = [\text{step freq}] \times ((32 \times M) + S)$

For direct serial programming of the PLL IC where the separate values of M and S are not required for any particular purpose, the complete value of 32M + S = 1,616 may be used.

For example, when calculating the required values for channel 2 at 55.25 MHz, the result of 32M + S = 1,616 was broken into M and S values, when M = 50 and S = 16. So 32(M) + 16 = 1,616, which is the preliminary result of expression (3a). Therefore it is not necessary to actually calculate the discrete values of M and S unless required. The process was described above for clarity of the subject. If the internal divider was of a different value, then the complete calculation would be required.

Once the value of 1,616 is determined, then can be broken onto the serial data string consisting of 4 Band Switch bits, and 14 Data bits, as shown in Table 3.

Table 3.
Serial Data String

BAND BITS	DATA BITS
Band 1	1,616 (Hex 650)
0001	00011001010000

for the combination data string of:
000100011001010000.

The key step in this process is to convert the decimal number to a Hexadecimal value, and then convert it to binary. In this case it is important to make sure that the full 14 data bits are captured and included in the serial data string. In the case of tuning to 505.25 MHz, the Hex value is 2270, which will naturally generate sixteen binary bits of data. Only the lower fourteen bits will be used. At the maximum end of the tune range, the generated Hex value for the oscillator will only occupy fourteen bits of data.

Programming Note

Care should be taken to ensure that extra clock pulses are not presented on the clock line prior to disabling the Load Enable line on the PLL. Extraneous clock pulses will create the appearance of the tuner not operating correctly, as it will tune erratically. This is due to the data stream being shifted by one or more additional bits, providing a division of all register values, and incorrect band switching control bits being asserted.

Programming errors will create a wide variety of problems that generate symptoms that are misleading. If the circuit design appears to be correct in all respects, and difficulties are experienced, examine the timing of the clock, data, and load enable lines carefully. In addition, ensure that the correct data bits are being sent to the PLL for your desired channel.

Band-Switching

The TD-series of PLL ICs provide four pins for control of band-switching in electronic tuners. Different sections of the tuner are used for the VHF and UHF reception bands. In this manner, one section of the tuner is enabled and another section disabled for the specified tuning range. This function is automatically controlled from the PLL IC directly, as decoded by the PLL from the serial tuning data string.

This function is transparent to the operator of the tuner module. The first four bits of the serial data string indicate which band is to be selected for operation. Each tuner specification will indicate which band selection is required for each channel in its tuning range. The band switch range and associated bits are shown in Table 4. Band bit 3 activates the EIA IS-31 filter and does not affect the tuning performance.

Table 4.
Band Switch Range and Associated Bits

CHANNELS	FREQUENCIES (MHz)	BIT PATTERN
2 – G	55.25 – 157.25	0001
H – W + 11	163.25 – 361.262	0010
W + 12 – W + 83	367.262 – 793.25	1000

SANYO® LA7530N DEMODULATOR

The RFSP7US1 and RFSO7US1 incorporate the Sanyo LA7530 Demodulator IC. This demodulator provides baseband Video and Audio outputs from an IF signal input. The demodulator includes an Audio Mute function and an AFT (S-Curve) output.

Audio and Video Mute (Blanking)

The Audio and Video Mute function provides external control of the Audio and Video outputs. When the Mute pin is taken below 0.7 V, both the Audio and Video outputs are disabled. The LA7530N specification indicates 0.7 V typical, and 0.5 V minimum, as the Mute threshold voltage. The LA7530 specification shows the Mute pin shorted to ground through a switch when Mute is asserted. A switching transistor is recommended as an alternative to using a relay or switch. Mute control is asserted through one of the IF AGC filter pins on the IC. The video and audio outputs may require several seconds to recover once Mute is no longer asserted.

AFT (S-Curve) Output

An Automatic Frequency Tuning (AFT) or Automatic Fine Tuning output is provided for monitoring how precisely the tuner is on frequency (in some specifications, the AFT output will be indicated as S-Curve). The LA7530N specification lists AFT sensitivity as shown in Table 5.

Table 5.
7AFT Sensitivity

MINIMUM	TYPICAL	MAXIMUM
80 mV/kHz	110 mV/kHz	150 mV/kHz

The specification lists the AFT center point as +6 V. The AFT output has a positive sense, meaning that if the tuner is above the desired frequency, the AFT output will be more positive than +6 V. Likewise, if the tuner is below frequency, the AFT output will be below +6 V.

AFT Application

All tuners covered herein are PLL tuned. The PLL is controlled by means of a serial digital data string. Due to this tuning method, it is not possible to simply sum the AFT output into an operational amplifier to automatically correct the tuning frequency, as may be done with analog tuning. An Analog-to-Digital Converter (ADC) can be used to monitor the AFT output, allowing the system microprocessor to provide corrections to the PLL serial data. Correction algorithms can be developed from the tuning equations listed above, depending on which PLL is in use.

Check the appropriate specification for the test circuit for use with the AFT output. This circuit normally consists of two 120 kΩ resistors configured as a simple voltage divider from the positive voltage supply. The junction of the two resistors is connected to the AFT pin. This circuit will allow the AFT output to be monitored for correct voltages. The circuit connection is to pin 16 for RFSO7US1 and RFSP7US1 (Figure 1).

AFT, AUDIO, AND VIDEO OUTPUTS

TD6359P/TD6380P AFT Correction

As discussed in the section on the TD-series PLLs, the S-value used in the data stream for regular NTSC-M channels will be 16. The step frequency used in these PLLs is 62.5 kHz. Each increment or decrement of the S-count will change the tuned frequency by the frequency step value. Based on the frequency step value, an AFT deviation of 62.5 kHz will provide an AFT output of +12.875 V. The AFT specification for the RFSP-series tuners indicates the AFT range is +6 V ±5.5 V. Obviously the nominal voltage swing associated with 62.5 kHz of deviation exceeds this range.

The consideration that applies here is that the frequency accuracy for the AFT output is 25 kHz nominal, 100 kHz maximum error. An option is to use a voltage comparator instead of an ADC. If the AFT voltage exceeds nominal thresholds, such as +6 V ±4.5 V, then a tuning correction is required. Using the nominal 25 kHz AFT accuracy specification, the 62.5 kHz of tuning error could occur with an AFT output of ±4.125 V.

Tuners using the TD6359P and TD6380P PLLs cannot make corrections for frequency errors of less than 62.5 kHz. Therefore, precise measurement of the AFT output is not required.

Audio Output

The audio output from the RFSO7US1 and RFSP7US1 modules is shown in Figure 1. This circuit is recommended, not mandatory, and is suggested for testing the unit as a separate module. It may be advantageous to incorporate these audio circuits in your final design to provide effective coupling.

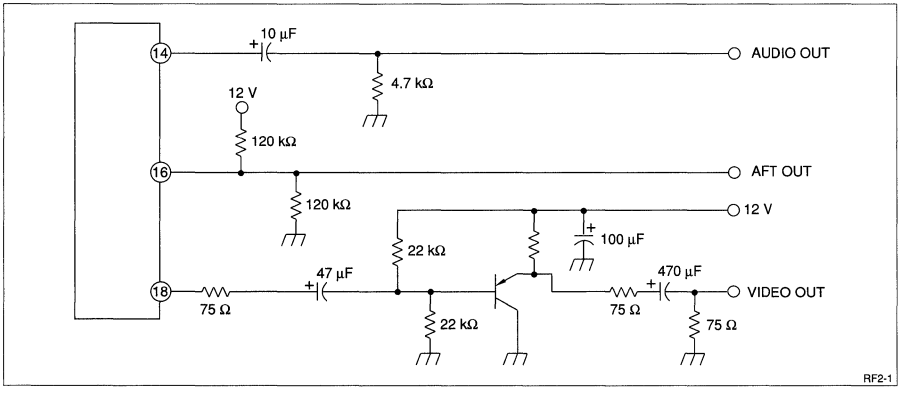


Figure 1. RFSP7US1 and RFSO7US3 Audio, AFT, and Video Outputs

Video Output

The video output shown in Figure 1 provides proper buffering and biasing for the video output of RFSP7US1 and RFSO7US1. If the video signal is taken directly from the pin on the unit, the picture quality will be significantly degraded as the DC level is not correctly set. The buffer stage helps restore the correct DC level to the video signal. Selection of the transistor is left to the designer. The transistor should have a minimum 7 MHz frequency bandwidth when operated in this capacity. A unique buffer stage of your own design will also work if system isolation and video DC signal level is considered.

The DC level of the output waveform should be referenced to the blanking level shown in Figure 2. The blanking level is usually set to 0 V. Signal amplitudes above the blanking level are referred to as going in the direction of "white" (as displayed on the television receiver), and signals going below the blanking level are in the "black" region. This is a very general description and not intended to be exhaustive. Refer to other materials on video waveform specifications and measurements for complete details.

The video coupling circuit for RFSN7US2 is shown in Figure 2. It is significantly simpler than for the other two products. This is a result of the LA7577 demodulator chip used in this application. All of the above circuits may be replaced by circuits of the designer's choosing. These circuits are recommended for testing and may be used in the final design.

TUNING TABLES

Tables with tuning values for M and S based on nominal VCO frequencies are available on request. Tables are provided with the TD6359P and TD6380P specifications, and apply only to the TD6359P and TD6380P.

REFERENCES

Television Measurements, NTSC Systems, Margaret Craig, Tektronix, Inc.,1990.

RFSO7US1 specification, 16 June, 1992, Sharp Electronics Corporation.

RFSP7US1 specification, 16 June, 1992, Sharp Electronics Corporation.

TD6359P specification, 26 July, 1989, Toshiba Corporation.

TD6380P, TD6382P, Toshiba Corporation.

LA7530, Sanyo Corporation.

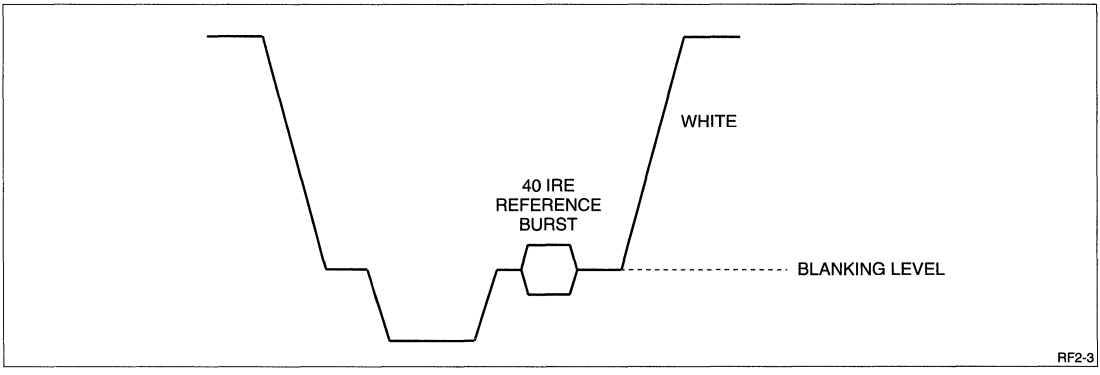


Figure 2. Sample Video Waveform

RY5AR01/AT01/BD01/AR021
Infra-Red Data Communication

For Sharp Products: RY5AR01/AT01/BD01/AR021

Sharp Electronics Corporation, Microelectronics Group, RF Products Marketing Group, 206-834-2500

OVERVIEW

Infra-Red Data Transmission (IRDT) and reception capabilities offer great potential to eliminate or reduce the number of wires and interconnecting cables required for data transfer. Applications such as a wireless mouse reduce the clutter associated with a desktop computer. Palmtop and laptop computers benefit from their small size and light weight. In this environment, bulky connectors add to the size and weight of the unit.

IRDT also has potential for many other applications in industry where data must be transferred: palmtop-palmtop data transfer, remote unit data down-load at the end of the day, computer tablets, barcode readers, wireless LANs, data logging equipment and more. Any data transfer application where a mechanical connector interface is impractical is a potential application for IRDT.

RY5 SERIES

The SHARP RY5 series Infra-Red Data Communications products provide small size, low-power serial data transmission. Current technology provides a data transfer rate of 38,400 bits-per-second (bps) at a distance of 1 meter. All RY5 series devices are TTL/CMOS compatible, operate from a 5 V source, and are readily interfaced with existing digital circuitry. A 500 kHz Amplitude Shift Keying (ASK) modulation process combined with an envelope detector insures reliable recovery of transmitted data and significantly reduces the possibility of interference from other IR sources. IR devices inherently demonstrate a high resistance to EMI noise.

Four products are currently available in the RY5 series: RY5AR01 IR Data Receiver, RY5AT01 IR Data Transmitter, RY5BD01 Bidirectional Data Transceiver, and the RY5AR021 enhanced range IR Data Receiver.

General Specifications

RY5 series units have the characteristics shown in Table 1.

Table 1. Operating Characteristics

Transmission Wavelength	900 to 1,050 nanometers
Modulation Frequency	500 kHz ± 50 kHz
Modulation Process	Amplitude Shift Keying
Range	1 meter (3 meters for RY5AR021)
Data Rate	38,400 bits-per-second
Supply Voltage	5 VDC ± 0.3 V

RY5AR01 IR Data Receiver

The RY5AR01 IR Data Receiver is designed for reception of Amplitude-Shift-Keying (ASK) format serial data. The unit is composed of a photodiode, amplifier chain, bandpass filter, detector, and hysteresis comparator.

The photodiode is the detector for the IR signal. The recovered signal is amplified through a chain of five amplifiers. An internal AGC and limit stage is used to prevent saturation of the amplifier chain if a strong signal is present. A bandpass filter is incorporated to select the desired signal source as the primary input recovered by the receiver system. The transmitter module provides a 500 kHz modulation source that receives gating information from the transmitting data signal. The receiver will respond to this modulation source and reject signals without the 500 kHz signal. The detector responds to the presence of a correctly modulated IR source and drives the hysteresis comparator to provide an inverted TTL/CMOS compatible output.

The receiver external interface is composed of three pins in a shrouded connector block. The pins are power, ground, and an inverted sense output signal.

RY5AT01 IR Data Transmitter

The RY5AT01 IR Data Transmitter is designed for transmission of 500 kHz ASK-modulated serial data. The unit is composed of an input buffer, a 500 kHz modulator, and the output IR LED. The application of a positive logic '1' (+5 V) to the input enables the 500 kHz modulation at the IR output. A logic '0' (0 V) disables the output. As serial data is applied to the input, the output is gated to correspond to the data stream. Observe correct data protocols for the transmission and reception of serial data to insure correct recovery of the transmitted data. The transmitter external interface is composed of three pins extending flat in the plane of the circuit board. Power, ground, and a signal input are provided.

RY5BD01 IR Bidirectional Data Unit

The RY5BD01 IR Bidirectional Data Unit is designed for transmission and reception of 500 kHz ASK-modulated serial data. This unit incorporates all of the features of the RY5AR01 IR Data Receiver. The transmitting LEDs are the only portion of the RY5AT01 transmitter in this package. The 500 kHz modulation source and input buffer must be incorporated external to the unit. All other data transmission characteristics are the same.

RY5AR021 Three-Meter IR Data Receiver

The RY5AR021 is an enhanced-range version of the RY5AR01 IR Data Receiver. It has identical electrical characteristics with the exception of the enhanced range. The mechanical package has been changed to incorporate two photodetectors. This unit is used in conjunction with the standard RY5AT01 one-meter transmitter to provide a total three-meter effective operating range. If additional transmission distance is required, a discrete transmitter unit should be developed and used with the RY5AR021. A discrete transmitter implementation will allow for additional transmitting diodes to increase the total radiated signal.

CIRCUIT APPLICATION

RY5AR01

The RY5AR01 IR Data Receiver is a compact, lightweight unit designed for reception of Amplitude-Shift-Keying (ASK) format serial data. The unit is composed of a photodiode, amplifier chain, bandpass filter, detector, and hysteresis comparator (Figure 1).

The photodiode is the detector for the IR signal. Even though the received IR energy is in the 900-1,050 nanometer range, consideration should be given to reducing the amount of light received from ambient or local light sources. Light sources such as the sun, halogen, and tungsten lamps have a fairly high IR output over a broad frequency range. Reception of broadband light sources by the photodiode detector will tend to raise the internal noise level and reduce the overall sensitivity of the receiver circuits. One easy way to test for the amount of background light interference is with an IR transparent window. Refer to the *IR Window* section. If high ambient IR levels are expected, consideration should be given to the incorporation of an IR filter in the system design.

The recovered signal is amplified through a chain of five amplifiers. An internal AGC and limit stage is used to prevent saturation of the amplifier chain if a strong signal is present. A signal that is at the low end of the receiver sensitivity will be amplified to a usable level. A bandpass filter is incorporated to select the desired signal source as the primary input recovered by the receiver system. The bandpass characteristic suppresses signals in the 38 kHz modulation band used by many in-home consumer products such as television receivers and VCRs. In addition, many new inverter-supply fluorescent fixtures for commercial and in-home application use power sources driven in the 20 kHz to 38 kHz region. The light sources can generate background noise that in some circumstances will raise the overall noise and interference levels in a room.

The receiver will respond to the 500 kHz transmitter signal source and reject signals without the 500 kHz modulation. For example, an IR signal source in the 900-1,050 nanometer band with 38 kHz modulation will be rejected, and the receive unit will not generate an output.

The detector stage is an envelope detector which responds to the presence of a correctly modulated signal. The detector output feeds the hysteresis comparator which provides a TTL/CMOS compatible output. A positive-going signal that exceeds the upper comparator threshold will set the comparator, and a negative-going signal that exceeds the lower threshold will reset the comparator. Hysteresis minimizes the probability of receiving incorrect information. [1] The

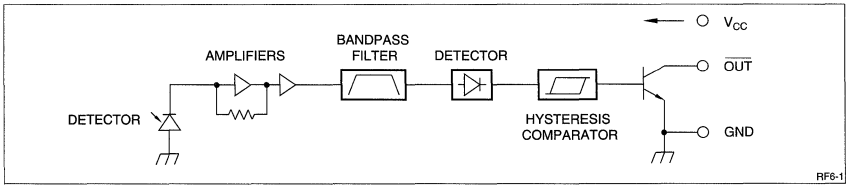


Figure 1. RY5AR01 Block Diagram

comparator output is the logical complement of the received signal. For a logical '1' at the transmitter, a logic '0' is presented at the output of the receiver. The data receiver has an open-collector output that can be used with signals that do not exceed +10 V. A 33 kΩ pull-up resistor from V_{CC} is recommended in a normal +5 V system architecture. The nominal design current for the output is 150 μA. Do not exceed 200 μA. The receiver external interface is composed of three pins in a shrouded connector block; power, ground, and the inverted sense output signal. (See Figure 2.)

Information about where to obtain matching connectors is included in the *References* section.

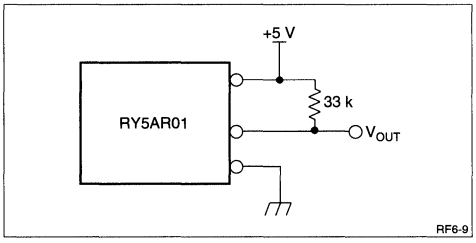


Figure 2. RY5AR01 Output Configuration

IR Window

A convenient source of IR transparent material is to use a piece of color film. Obtain a roll of color print film with an ASA rating of 100. Pull the film out of the can until the end is reached and expose the film to sunlight for several seconds. Roll the film back into the can and have it processed. Explain to your photo shop what you are doing so that they will correctly process the film and not try to make prints. The processed film material will provide a good IR test filter. [2] Double layers of the processed film material will be difficult to see through, yet will remain transparent to most IR wavelengths.

RY5AT01

The RY5AT01 IR Data Transmitter is a compact, lightweight unit designed for transmission of 500 kHz ASK-modulated serial data. The unit is composed of an input buffer, a 500 kHz modulator, and the output IR LED (Figure 3).

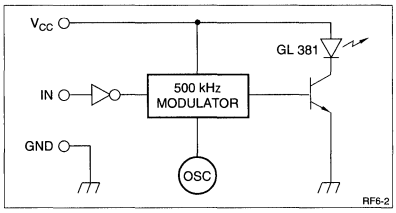


Figure 3. RY5AT01 Block Diagram

The application of a positive logic '1' (+5 V) to the input inverting buffer enables the 500 kHz modulation at the IR output. A logic '0' (0 V) disables the output. Serial data at the input creates a corresponding modulated output. Observing correct data protocols for transmission and reception of serial data will insure correct recovery of the transmitted data. Half-duplex protocols will provide high reliability data transmission. Full-duplex operation is difficult in the IR environment due to local echo at each end of the transmission system, and from reflective objects in short-range line-of-sight. Half-duplex operation can achieve high transfer rates and low error rates when packet transfer and error checking are employed.

The transmitter external interface is composed of three pins extending flat in the plane of the circuit board. Power, ground, and a signal input are provided.

Amplitude Shift Keying

The 500 kHz modulator runs continuously and is enabled by the input signal. The modulation process is Amplitude-Shift-Keying (ASK) (Figure 4). In this application, ASK essentially refers to the presence or absence of a modulated carrier. [3] In transferring data files, 'mark' and 'space' correspond to logic '1' and '0' respectively. [4] With a logic '1' (mark) applied at the input, a modulated signal will be gated to the output IR LED. With a logic '0' (space) at the input, no signal will be sent. In this condition, the IR carrier will not be transmitted. The time period of a logic '1' data bit is filled in with the 500 kHz modulation frequency. The envelope detector at the receiver will respond to the presence of the modulation envelope. The rising and falling edges of the envelope will drive the hysteresis comparator to generate the recovered output. Probability data for the correct transmission and recovery

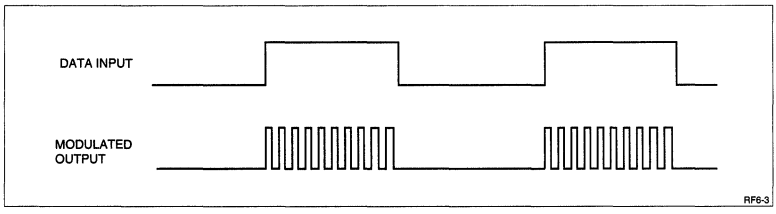


Figure 4. ASK Modulation Diagram

of ASK data is available in reference 1. With correct application of these modules, error rates of $10e^{-07}$ can be demonstrated.

RY5BD01

The RY5BD01 IR Bidirectional Data Unit is a compact unit designed for transmission and reception of 500 kHz ASK-modulated serial data. This unit incorporates all of the features of the RY5AR01 IR Data Receiver. The transmitting LEDs are the only portion of the RY5AT01 transmitter in this package. The 500 kHz modulation source and input buffer are added in external supporting hardware for the unit. (See Figures 5 and 7).

The RY5BD01 unit is designed for mounting on the surface of a circuit board. Consider a reflector as an easy method for conducting the IR energy to and from the unit. This will allow for direct mounting of the module without a separate circuit board mounted perpendicular to a normal backplane or motherboard (Figure 6). The mirror surface should be mounted as close as possible to the LEDs and photodiode.

Reflector Material

A recommended material for use as the reflector is 'PET' (Polyethylene Terephthalate) with 'Al' layer by evaporative deposition. This material is essentially a Polyethylene substrate with a reflective layer and should be available as an off-the-shelf item. Or, a similar material with a highly reflective surface should suffice.

Modulator Circuit

The RY5BD01 does not include a 500 kHz modulator circuit in order to keep the package size small. A modulator circuit of some type is required to provide the modulation signal and drive the IR LEDs. One possible circuit is shown in Figure 7. The ceramic resonator and CMOS inverters comprise the 500 kHz oscillator. The transistor arrangement is the LED driver network. The limit resistor controls the LED forward current at different duty cycle rates. The ceramic resonator in the oscillator has been chosen for the good frequency stability offered.

Try to maintain the 500 kHz modulation frequency within ±25 kHz of the fundamental. This will guarantee dependable signal recovery over temperature and other operating environment changes. The diagram in Figure 7 shows a low-cost, reliable, low gate-count design, but not the lowest possible gate-count. Other modulator schemes can be devised. The primary considerations should be frequency stability, and for a logic '1' (+5 V) at the Signal Input, the LEDs are enabled to provide an IR output. If a 500 kHz sub-multiple of a microprocessor clock is available, as shown in Figure 8, then it may be used as the modulation signal source. In that case, it would be applied to the second NAND gate input in place of the clock signal.

A stable RC oscillator comprised of CMOS gates is described in National Semiconductor® Application Note 118[5]. This circuit can be used to provide a 500 kHz source when no other methods are desired. (See Figure 9.)

The frequency of operation is described by the expression $f = 1/[2C(0.405\ R_{EQ} + 0.693\ R1)]$. R_{EQ} is the parallel combination of R1 and R2.

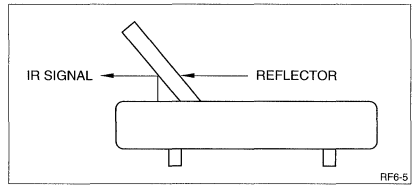


Figure 6. RY5BD01 With Reflector

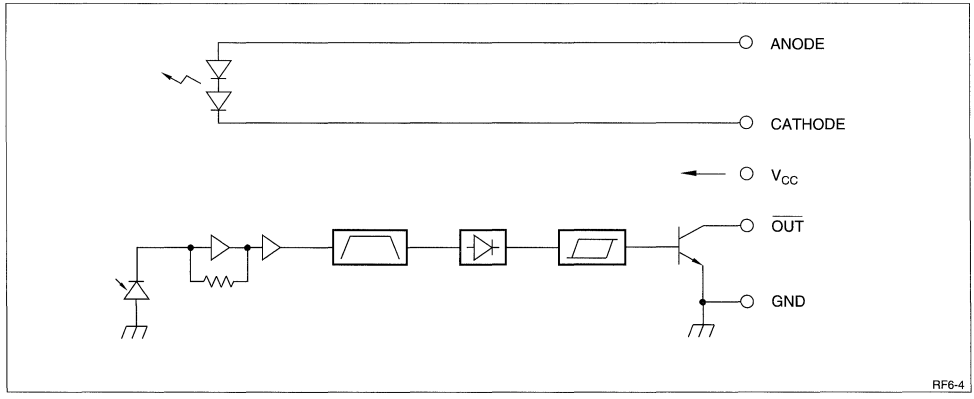


Figure 5. RY5BD01 Block Diagram

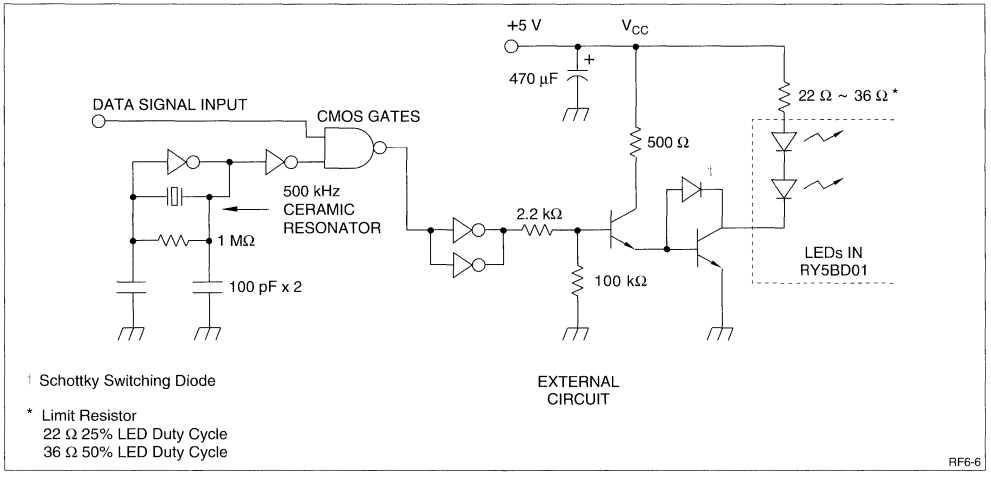


Figure 7. Sample 500 kHz Modulator Circuit

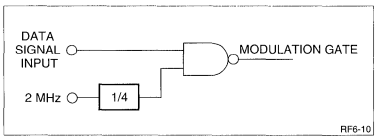


Figure 8. 500 kHz Modulation Source

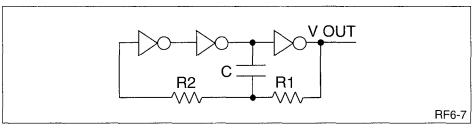


Figure 9. Three-Gate Oscillator

SAMPLE SYSTEM

A sample RS-232 connection using two RY5AT01 IR Transmitters and two RY5AR01 IR Receivers is shown in Figure 10. Standard RS-232 pin assignments are assumed here. This example employs the use of MAXIM[6] MAX233 RS-232 Drivers/Receivers, which may be operated from +5 VDC without any additional external components. This configuration duplicates a standard RS-232 connection for two-way data transfer.

Insure that the correct line-driver connections are made from the MAX233 device to the RS-232 port and to the TTL/CMOS interfaces. Short connection lengths from the MAX223 chips and the RY5AT01 and RY5AR01 modules are possible without additional power supply decoupling. If the run lengths are more than 3 inches, consider additional power supply decoupling if your system environment is noisy.

For full-duplex operation, isolate the co-located transmitter and receiver pairs from direct viewing by each other. A simple divider plate or other physical separation will suffice. When fully isolated, full-duplex operation may be achieved. Without separation, a local echo will be generated at each end of the transmitting system. In this normal operating environment, half-duplex operation is standard, and depending on the setup, the system software must strip the echo characters from the received data.

Currently, the RY5BD01 bidirectional data unit will support half-duplex operation only, due to the transmitter and receiver devices being co-located on the module. High data transfer rates may be accomplished by the use of direction control routines in the supervisory software in your final system. System test with a modem software package may appear to demonstrate full-duplex operation when echo canceling is chosen as a set-up parameter. Half-duplex operation will be more reliable and provide an overall lower error rate. Data transfer protocols that use a packet format that are not dependent on simultaneous bidirectional handshakes will operate most effectively.

Most commercially available communications software packages for Personal Computers (PCs) may be used to test the IR link. Data rates up to 38,400 bits-per-second work with minimal data errors, if recommended system parameters are not exceeded. Check your computer system data transfer specifications to determine if higher data rates can be supported without errors. Direct communication between the two modules is a first step in demonstrating the operation of the IR link.

Another method is to prepare a test file for evaluating the serial connection. Prepare the file prior to the

test. After configuring the port setup in the software, select 'file transfer.' Use a file transfer protocol without error correction to demonstrate a one-way data file transfer, with or without handshake. Check for file transfer errors. If present, check for system parameters that have been violated, such as range. Bidirectional capabilities may be tested by selecting a protocol that employs echo canceling, half-duplex handshake, or packet transfer methods. The file may be tested for data errors by using comparison routines that are part of the PC system programs.

Manufacturing Considerations

Several considerations should be kept in mind for best overall system results: the ground plane of the circuit board on which the RY5AR01 or RY5BD01 is mounted should be extended under the area of the photodiode. Including ground plane under as much of the units as possible will be of additional help. The receiver circuits and photodiode are susceptible to Electro-Magnetic Interference, and additional ground plane helps reduce the unit sensitivity. In the same context, the mounting tabs of the metal enclosures should also be soldered to plated-through holes that are tied to the ground plane.

Efficient mounting of the RY5BD01 requires the relief of several areas on the circuit board. Figure 11 provides information on the areas that need to be removed.

All of the RY5-series parts should be hand-mounted and soldered. These devices are not suitable for IR process systems used in Surface-Mount Technology production lines. The high IR temperatures beyond 100°C will soften the material in the IR photodiode and LEDs, distorting the optical characteristics. Under circumstances where due caution is exercised, the RY5AR01, RY5AR021, and RY5AT01 may be processed on a flow-solder system as long as the components do not heat soak to a 100°C level and solder will not wick to the underside of the modules. Short temperature transitions on the leads for less than five seconds should not harm the devices if the above stated considerations are kept in mind.

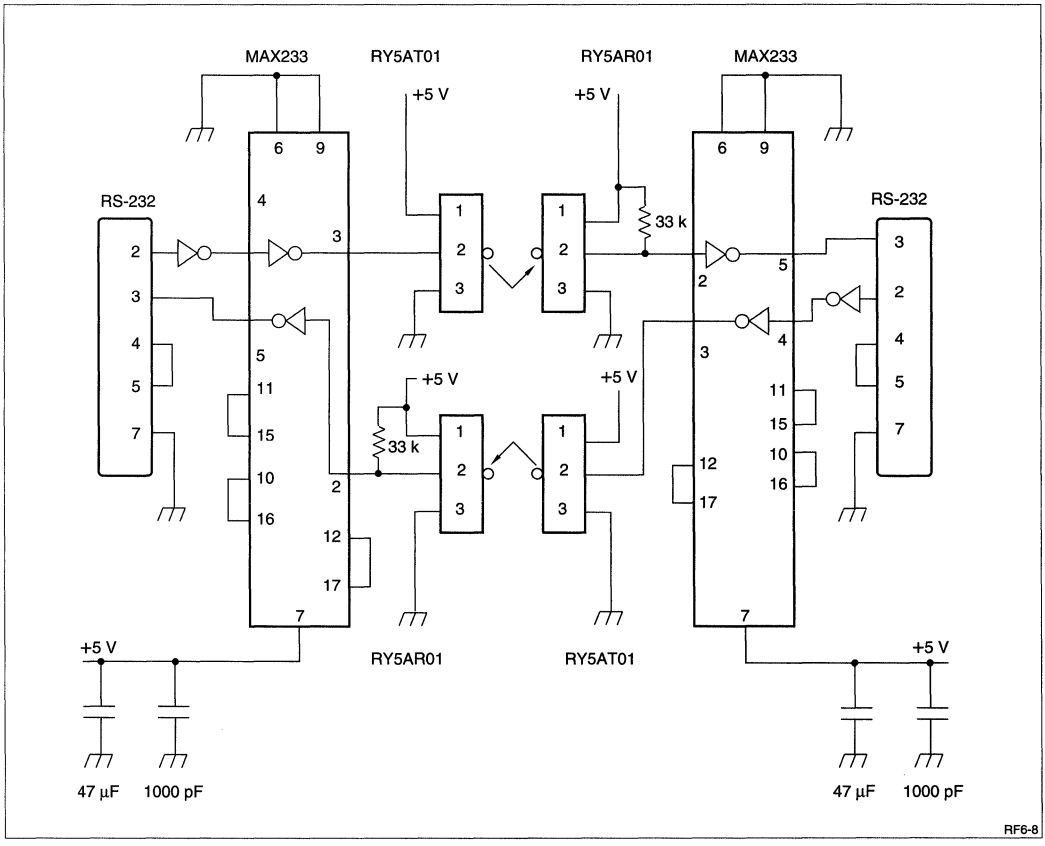


Figure 10. Bidirectional System Diagram

REFERENCES

[1] *Reference Data For Radio Engineers*, Howard Sams & Co., 1975, Ch. 23.

[2] Lance Kempler, Sharp Electronics, Opto-Electronics Field Applications Engineer, Camas, WA

[3] *Reference Data for Radio Engineers*, Howard Sams & Co., 1975, Ch. 23.

[4] *RS-232 Simplified*, Byron W. Putnam, Prentice-Hall, 1987.

[5] *National Semiconductor, CMOS Logic Databook*, AN-118, National Semiconductor Corporation, 2900 Semiconductor Drive, P.O. Box 58090, Santa Clara, California 95052-8090, 408-721-5000.

[6] *MAXIM Integrated Products*, 120 San Gabriel Drive, Sunnyvale, California 94086, 408-737-7600.

[7] Matching Connectors for RY5AR01: J.S.T. Corporation, 1200 Business Center Drive, Suite #400, Mt. Prospect, Illinois 60056, 708-803-3300, facsimile 708-803-4918. Crimp style connector: Part number PHR-3; matching contact Part number SPH-002T-P05S. IDC style connector Part Number 03KR-8M for 28 gauge wire; 03KR-6S for 26 gauge wire.

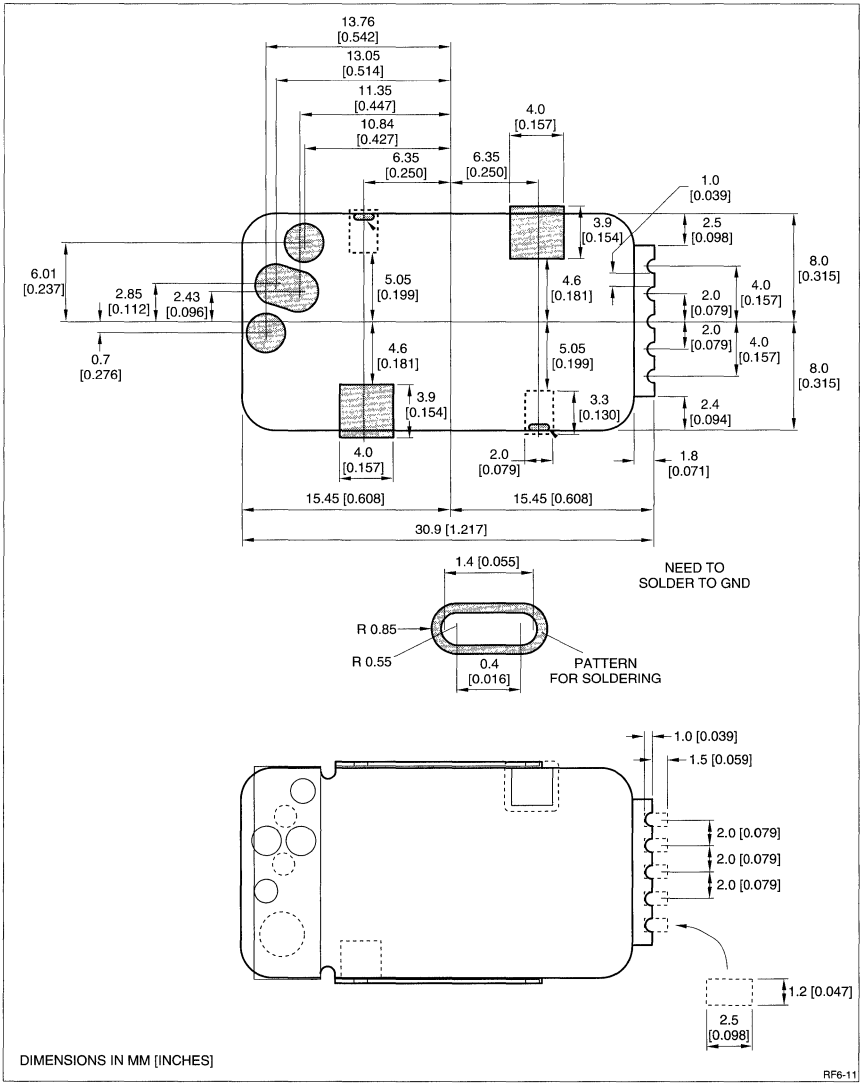


Figure 11. RY5BD01 Mounting and Soldering Details

OPTOELECTRONICS

'INFRARED' LIGHT-EMITTING DIODE APPLICATION CIRCUITS

Serial Connection And Parallel Connection

Figure 1 shows the most basic and commonly used circuits for driving light-emitting diodes.

In Figure 1(A), a constant voltage source (V_{CC}) is connected through a current limiting resistor (R) to an LED so that it is supplied with forward current (I_F). The I_F current flowing through the LED is expressed as $I_F = (V_{CC} - V_F)/R$, providing a radiant flux proportional to the I_F. The forward voltage (V_F) of the LED is dependent on the value of I_F, but it is approximated by a constant voltage when setting R.

Figures 1(B) and 1(C) show the circuits for driving LEDs in serial connection and parallel connection, respectively. In arrangement (B), the current flowing through the LED is expressed as $I_F = (V_{CC} - V_F \times N)/R$, while in arrangement (C), the current flowing through each LED is expressed as $I_F = (V_{CC} - V_F)/R$ and the total supply current is $N \times I_F$, where N is the number of LEDs.

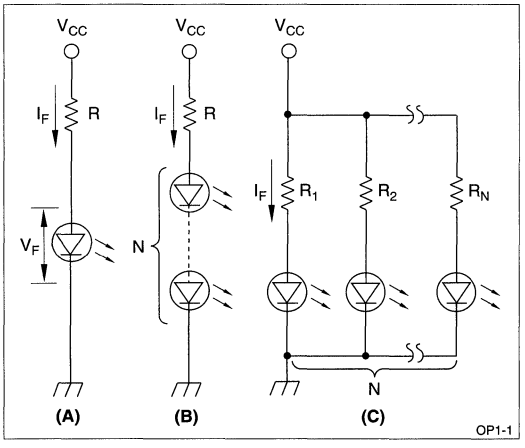


Figure 1. Driving Circuit of Light-Emitting Diode (LED)

The V_F of an LED has a temperature dependency of approximately -1.9 mV/°C. The operating point for the load R varies in response to the ambient temperature as shown in Figure 2.

Constant Current Drive

To stabilize the radiant flux of the LED, the forward current (I_F) must be stabilized by using a constant current source. Figure 3 shows a circuit for constantly driving several LEDs using a transistor. The transistor (Tr_1) is biased by a constant voltage supplied by a zener diode (ZD) so that the voltage across the emitter follower loaded by resistor R_E is constant, thereby making the collector current ($I_C = I_F$) constant. The I_C is given as $I_C = I_E = (V_Z = V_{BE})/R_E$. If too many LEDs are connected, the transistor enters the saturation region and does not operate as a constant current circuit. The number of LEDs (N) which can be connected in series is calculated by the following equations.

$$V_{CC} - N \times V_F - V_E > V_{CE}\ (sat)$$

$$V_E = V_Z - V_{BE}$$

These equations give:

$$N < (V_{CC} - V_Z + V_{BE} - V_{CE}(sat))/V_F$$

Figures 4 and 5 show other constant current driving circuits that use diodes or transistors, instead of zener diodes.

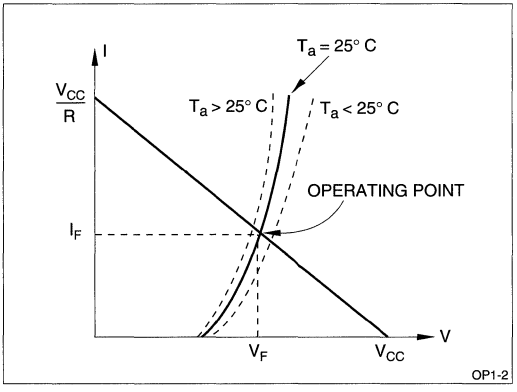


Figure 2. Current vs. Voltage of Light-Emitting Diode (LED)

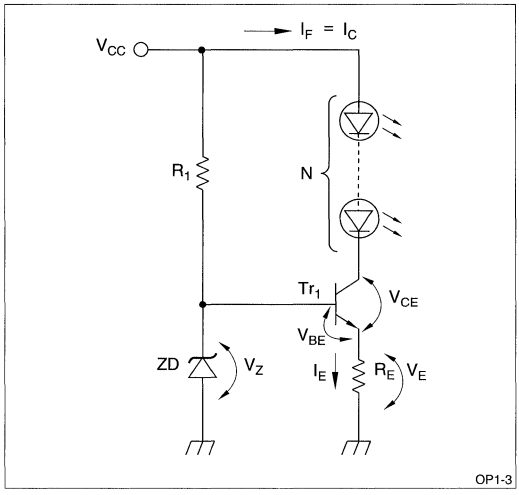


Figure 3. Constant Current Driving Circuit (1)

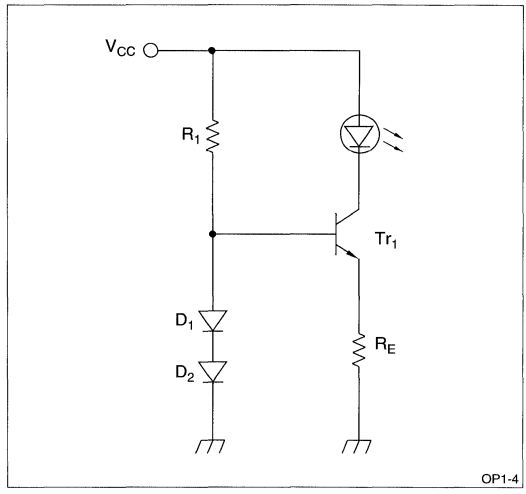


Figure 4. Constant Current Driving Circuit (2)

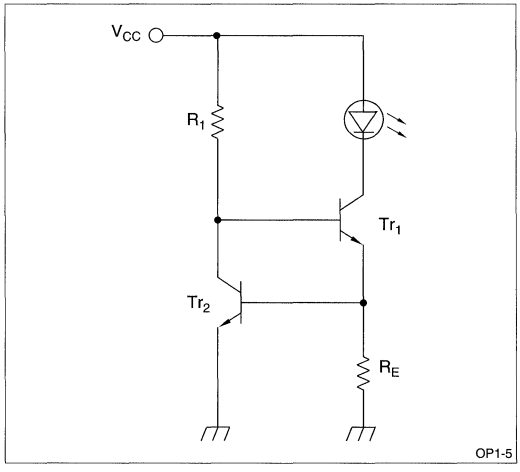


Figure 5. Constant Current Driving Circuit (3)

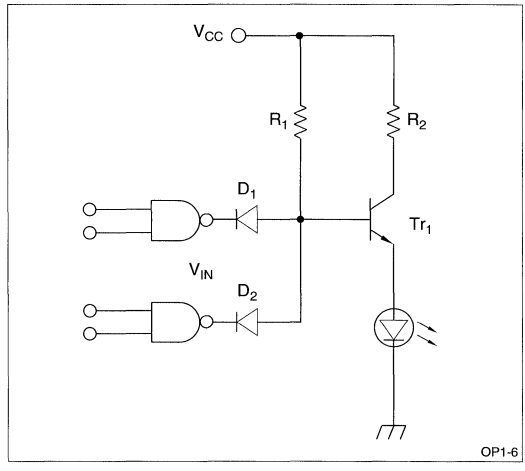


Figure 6. Connection with the TTL Logic Circuit (1)

Driving Circuit Activated By A Logic IC

Figures 6 and 7 show LED driving circuits that operate in response to digital signals provided by TTL or CMOS circuits.

Figure 8 shows a driving circuit connected with a high level logic circuit.

In Figure 6, a high input signal V_{IN} from a TTL circuit makes the NPN transistor (Tr_1) conductive so that the forward current (I_F) flows through the LED. Accordingly, this circuit operates in the positive logic mode, in which a high input activates the LED.

In Figure 7, a low input signal V_{IN} from a TTL circuit makes the PNP transistor (Tr_1) conductive so that the forward current flows through the LED. This circuit operates in the negative logic mode, in which a low input activates the LED.

In Figure 8, the circuit operates in the positive logic mode, and current I_F is stabilized by constant current driving so that the radiant flux of LED is stabilized against variations in the supply voltage (V_{CC}).

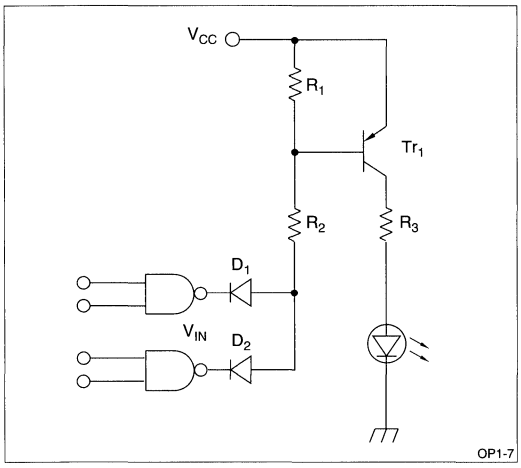


Figure 7. Connection with the TTL Logic Circuit (2)

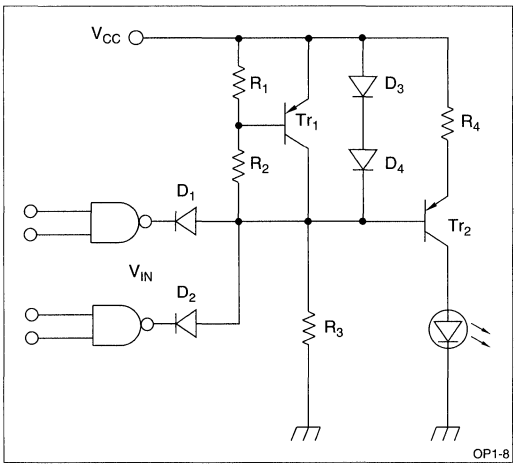


Figure 8. Connection with the TTL Logic Circuit (3)

Driving Circuit With An AC Signal

Figure 9 (A) shows a circuit in which an AC power source supplies the forward current (I_{F1}) to an LED. A diode (D_1) in inverse parallel connection with the LED protects the LED against reverse voltage, suppressing the reverse voltage applied to the LED lower than V_{F2} by using a reverse voltage protection diode of an LED. The LED provides a radiant flux proportional to the applied AC current, (emitting only in half wave).

Figure 9 (B) shows the driving waveform of the AC power source.

Figure 10 (A) shows a driving circuit which modulates the radiant flux of LED in response to a sine wave or modulation signal. Figure 10 (B) shows modulation operation.

If an LED and light detector are used together in an environment of high intensity disturbing light, it is difficult for the light detector to detect the optical signal. In this case, modulating the LED drive signal alleviates the influence of disturbing light and facilitates signal detection.

To drive an LED with a continuous modulation signal, it is necessary to operate the LED in the linear region of the light-emitting characteristics. In the arrangement of Figure 10, a fixed bias (I_{F1}) is applied to the LED using R_1 and R_2 so that the maximum amplitude of the modulation signal voltage (V_{IN}) lies within the linear portion of the LED characteristics. Moreover, to stabilize the radiant flux of the LED, it is driven by a constant current by the constant current driving circuit shown in Figure 3. The capacitor (C) used in Figure 10 (A) is a DC signal blocking capacitor.

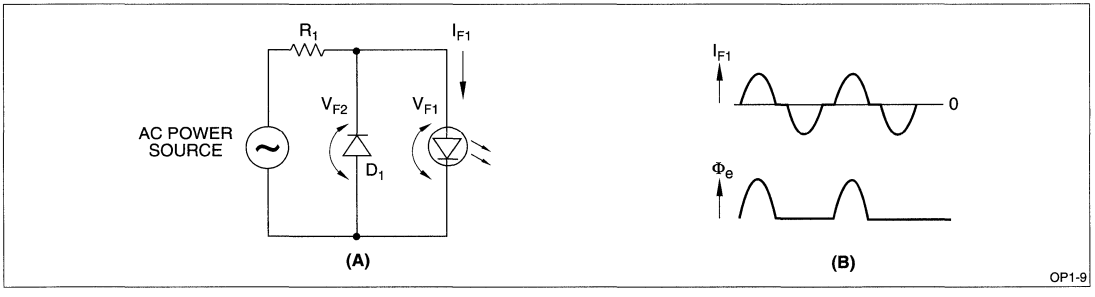


Figure 9. (A) Driving Circuit with AC Power Source (B) Driving Waveform

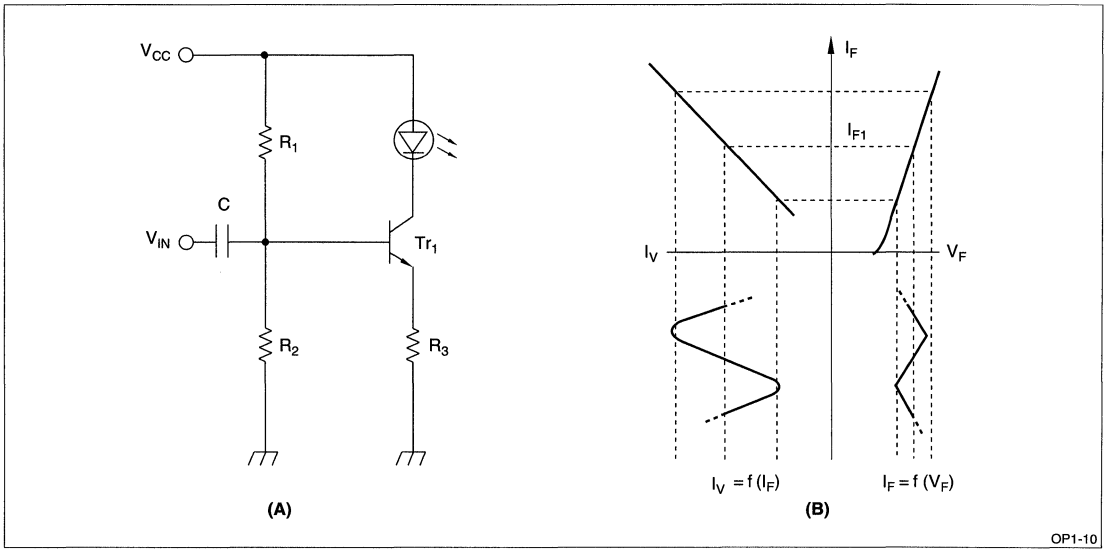


Figure 10. (A) Modulation Driving Circuit (B) Modulation Operation

Pulse Driving

LED driving systems fall into three categories: DC driving system, AC driving system (including modulation systems), and pulse driving system.

Features of the pulse driving system:

1. Large radiant flux
2. Less influence of disturbing light
3. Information transmission

1. The radiant flux of the LED is proportional to its forward current (I_F), but in reality a large I_F heats up the LED by itself, causing the light-emitting efficiency to fall and thus saturating the radiant flux. In this circumstance, a relatively large I_F can be used with no risk of heating through the pulse drive of the LED. Consequently, a large radiant flux can be obtained.

2. When an LED is used in the outdoors where disturbing light is intense, the DC driving system or AC driving system which superimposes an AC signal on a fixed bias current provides low radiant flux, making it difficult to distinguish the signal (irradiation of LED) from disturbing light. In other words, the S/N ratio is small enough to reliably detect the signal. The pulse driving system provides high radiant flux and allows the detection of signal variations at the rising and falling edges of pulses, thereby enabling the use of LED-light detector where disturbing light is intense.

3. Transmission of information is possible by variations in pulse width or counting of the number of pulse used to encode the LED emission.

Figures 11 through 14 show typical pulse driving circuits. Figure 15 shows the pulse driving circuit used in the optical remote control. The circuit shown in Figure 11 uses an N-gate thyristor with voltage between the anode and cathode oscillated at a certain interval determined by the time constant of $C \times R$ so that the LED emits light pulse. To turn off the N-gate thyristor, resistor R_3 must be used so that the anode current is smaller than the holding current (I_H), i.e., $I_H > V_{CC}/R_3$. Therefore, R_3 has a large value, resulting in a large time constant ($\tau \pm C \times R_3$) and the circuit operates for a relatively long period to provide short pulse widths. The circuit shown in Figure 12 uses a type 555 timer IC to form an astable multivibrator to produce light pulses on the LED. The off-period (t_1) and the on-period (t_2) of the LED are calculated by the following equations.

$$t_1 = 1n2 \times (R_1 + R_2) \times C_1$$

$$t_2 = 1n2 \times R_2 \times C_1$$

The value of R_1 is determined so that the rating of I_{IN} of a 555 timer IC is not exceeded, i.e. $S_1 > V_{CC}/I_{IN}$.

This pulse driving circuit uses a 555 timer IC to provide wide variable range in the oscillation period and light-on time. It is used extensively.

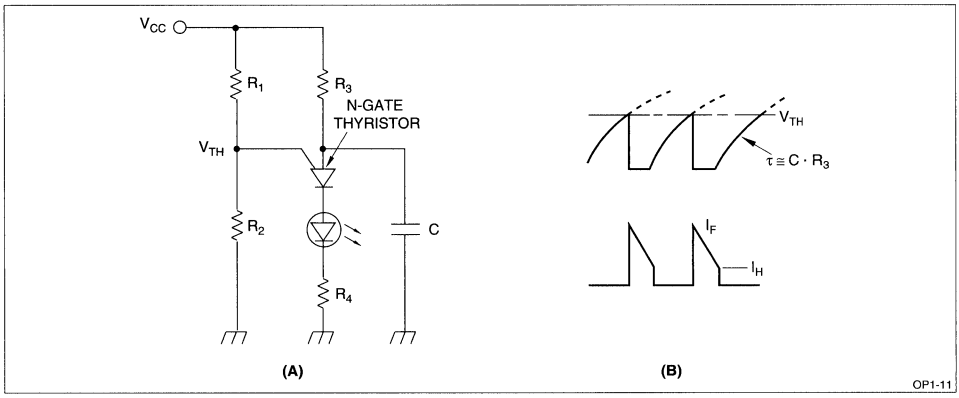


Figure 11. (A) Pulse Driving Circuit using N-Gate Thyristor (B) Operating Waveform

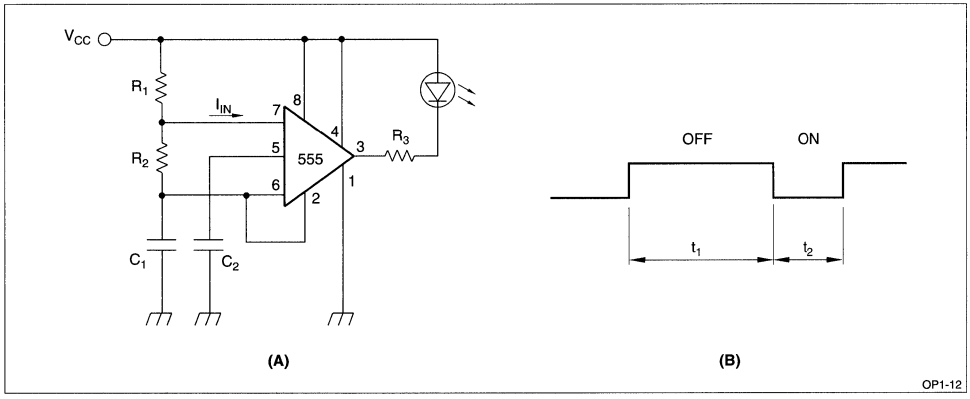


Figure 12. (A) Pulse Driving using a 555 Timer IC (B) Output Waveform

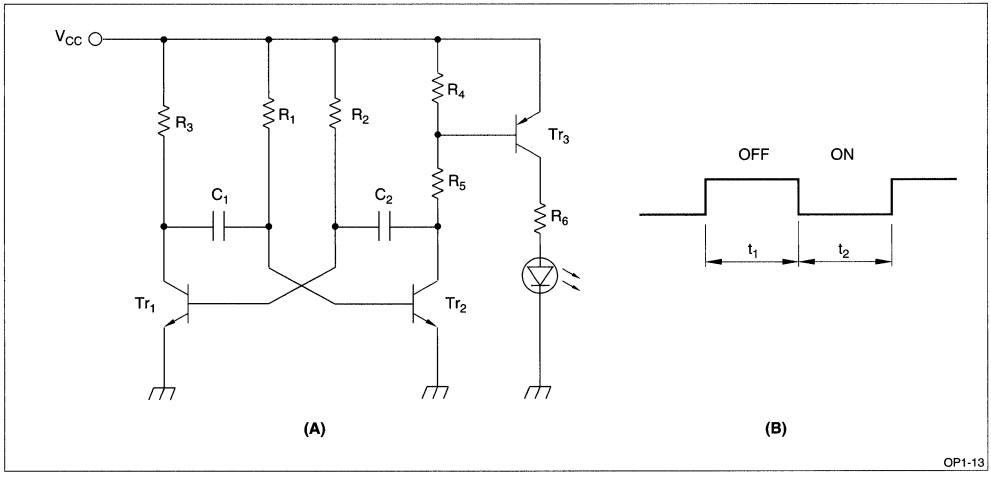


Figure 13. (A) Pulse Driving Circuit using Astable Multivibrator (B) Output Waveform

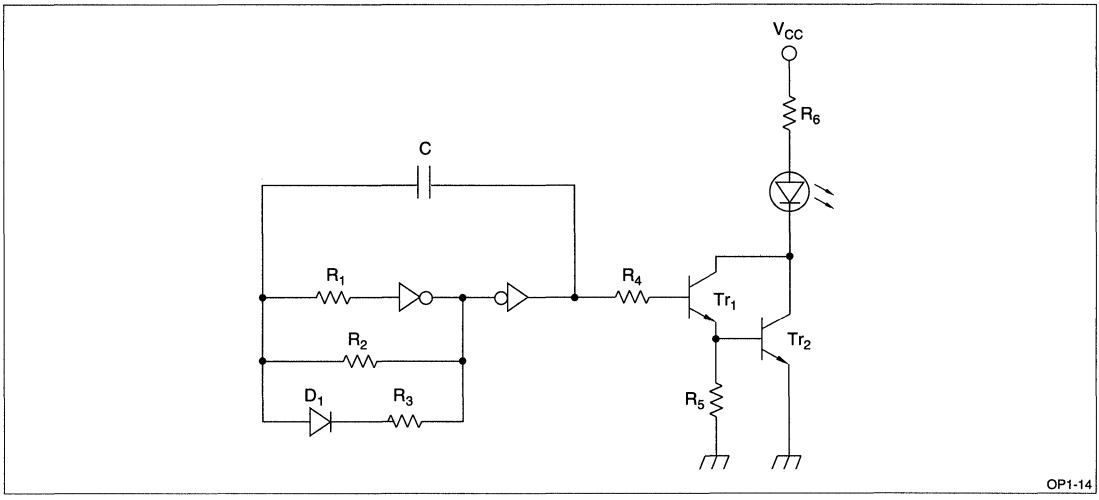


Figure 14. Pulse Driving Circuit using CMOS Logic IC

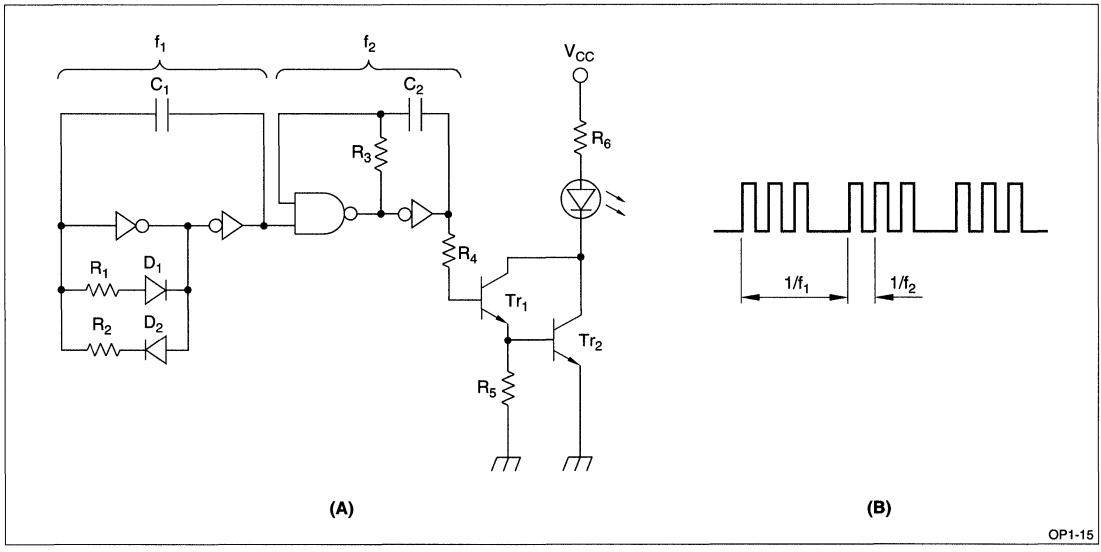


Figure 15. (A) Pulse Driving Circuit
(B) Output Waveform

The circuit shown in Figure 13 uses transistors to form an astable multivibrator for pulse driving an LED. The off-period (t_1) of the LED is given by $C_1 \times R_1$, while its on-period (t_2) is given by $C_2 \times R_2$. For oscillation of this circuit, resistors must be chosen so that the R_1/R_3 and R_2/R_5 ratios are large.

The circuit shown in Figure 14 uses a CMOS logic IC (inverter) to form an oscillation circuit for pulse driving an LED. The pulse driving circuit using a logic IC provides a relatively short oscillation period with a 50% duty cycle.

Figure 15 (A) shows an LED pulse driving circuit used for the light projector of the optical remote control and optoelectronic switch. The circuit is arranged by combining two different oscillation circuits i.e., a long period oscillation (f_1) superimposed with a short period oscillation (f_2) as shown in Figure 15 (B). Frequencies f_1 and f_2 can be set independently.

PHOTODIODE/PHOTOTRANSISTOR APPLICATION CIRCUITS

Fundamental Photodiode Circuits

Figures 16 and 17 show the fundamental photodiode circuits.

The circuit show in Figure 16 transforms a photocurrent produced by a photodiode without bias into a voltage. The output voltage (V_{OUT}) is given as $V_{OUT} = 1_P \times R_L$. It is more or less proportional to the amount of incident light when $V_{OUT} < V_{OC}$. It can also be compressed logarithmically relative to the amount of incident light when V_{OUT} is near V_{OC}. (V_{OC} is the open-terminal voltage of a photodiode).

Figure 16 (B) shows the operating point for a load resistor (R_L) without application of bias to the photodiode.

Figure 17 shows a circuit in which the photodiode is reverse-biased by V_{CC} and a photocurrent (I_P) is transformed into an output voltage. Also in this arrangement, the V_{OUT} is given as $V_{OUT} = I_P \times R_L$. An output voltage proportional to the amount of incident light is obtained. The proportional region is expanded by the amount of V_{CC} {proportional region: $V_{OUT} < (V_{OC} + V_{CC})$} . On the other hand, application of reverse bias to the photodiode causes the dark current (I_d) to increase, leaving a voltage of $I_d \times R_L$ when the light is interrupted, and this point should be noted in designing the circuit.

Figure 17 (B) shows the operating point for a load resistor R_L with reverse bias applied to the photodiode.

Features of a circuit used with a reverse-biased photodiode are:

1. High-speed response
2. Wide-proportional-range of output

Therefore, this circuit is generally used.

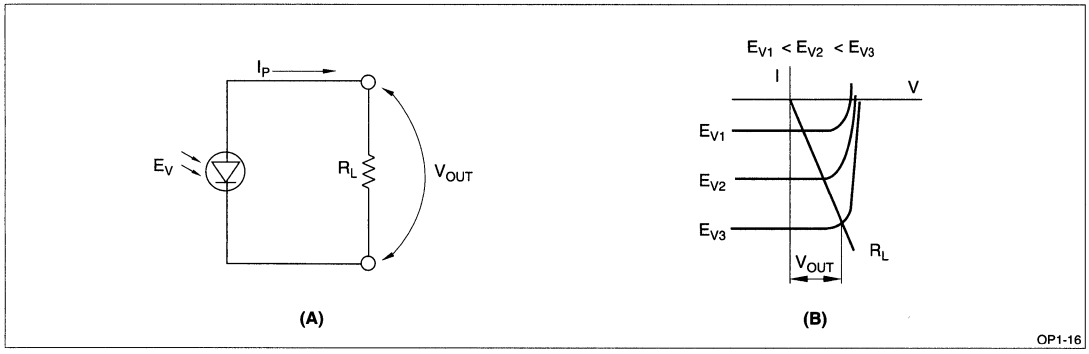


Figure 16. (A) Fundamental Circuit of Photodiode (without bias)

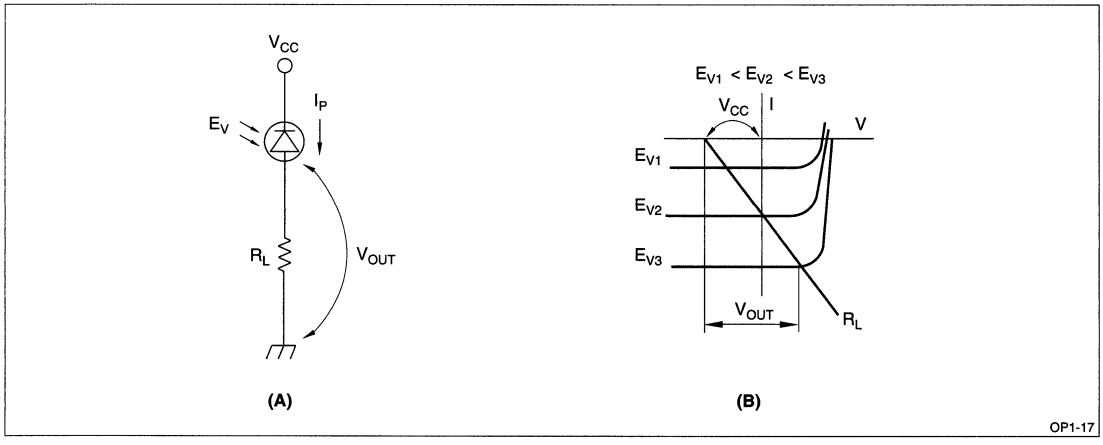


Figure 17. Fundamental Circuit of Photodiode (with bias)

The response time is inversely proportional to the reverse bias voltage and is expressed as follows:

$r = C_j \times R_L$

$C_j = A(V_D - V_R) - \frac{1}{n}$

C_j: junction capacitance of the photodiode

R_L: load resistor

V_D: diffusion potential (0.5 V ~ 0.9 V)

V_R: Reverse bias voltage (negative value)

n: 2 ~ 3

Photocurrent Amplifier Circuit Using The Transistor Of Photodiode

Figures 18 and 19 show photocurrent amplifiers using transistors.

The circuit shown in Figure 18 are most basic combinations of a photodiode and an amplifying transistor. In the arrangement of Figure 18 (A), the photocurrent produced by the photodiode causes the transistor (Tr_1) to decrease its output (V_{OUT}) from high to low. In the arrangement of Figure 18 (B), the photocurrent causes the V_{OUT} to increase from low to high. Resistor R_{BE} in the circuit is effective for suppressing the influence of dard current (I_d) and is chosen to meet the following conditions:

$R_{BE} < V_{BD}/I_d$

$R_{BE} > V_{BE} / \{I_P - V_{CC}/(R_L \times h_{FE})\}$

Figure 19 shows simple amplifiers utilizing negative feedback.

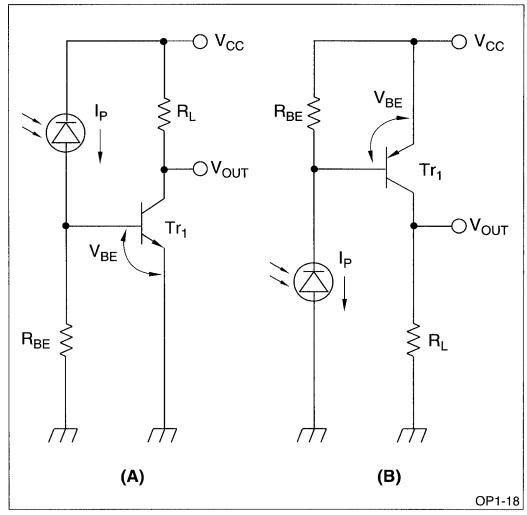


Figure 18. Photocurrent Amplifier Circuit using Transistor

In the circuit of Figure 19 (A), the output (V_{OUT}) is given as:

$V_{OUT} = I_P \times R_1 + I_B \times R_1 + V_{BE}$

This arrangement provides a large output and relatively fast response.

The circuit of Figure 19 (B) has an additional transistor (Tr_2) to provide a larger output current.

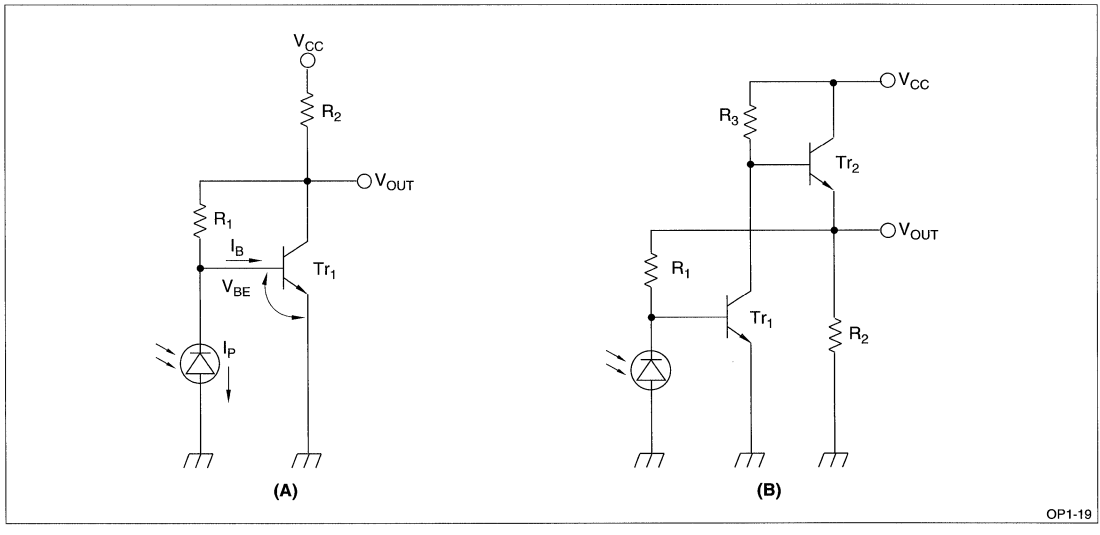


Figure 19. Photocurrent Amplifier Circuit with Negative Feedback

Amplifier Circuit Using Operational Amplifier

Figure 20 shows a photocurrent-voltage conversion circuit using an operational amplifier. The output voltage (V_{OUT}) is given as $V_{OUT} = I_F \times R_1$ ($I_P \cong I_{SC}$). The arrangement utilizes the characteristics of an operational amplifier with two input terminals at about zero voltage to operate the photodiode without bias. The circuit provides an ideal short-circuit current (I_{SC}) in a wide operating range.

Figure 20 (B) shows the output voltage vs. radiant intensity characteristics. An arrangement with no bias and high impedance loading to the photodiode provides the following features:

1. Less influence by dark current

2. Wide linear range of the photocurrent relative to the radiant intensity.

Figure 21 shows a logarithmic photocurrent amplifier using an operating amplifier. The circuit uses a logarithmic diode for the logarithmic conversion of photocurrent into an output voltage. In dealing with a very wide irradiation intensity range, linear amplification results in a saturation of output because of the limited linear region of the operational amplifier, whereas logarithmic compression of the photocurrent prevents the saturation of output. With its wide measurement range, the logarithmic photocurrent amplifier is used for the exposure meter of cameras.

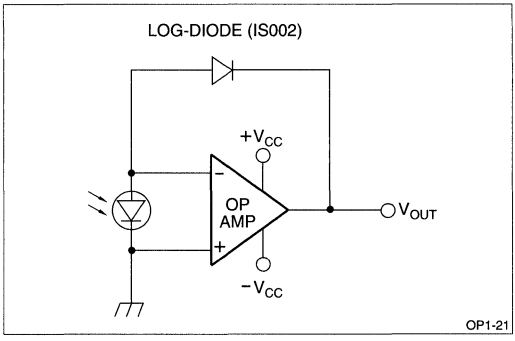


Figure 21. Logarithmic Photocurrent Amplifier using an Operational Amplifier

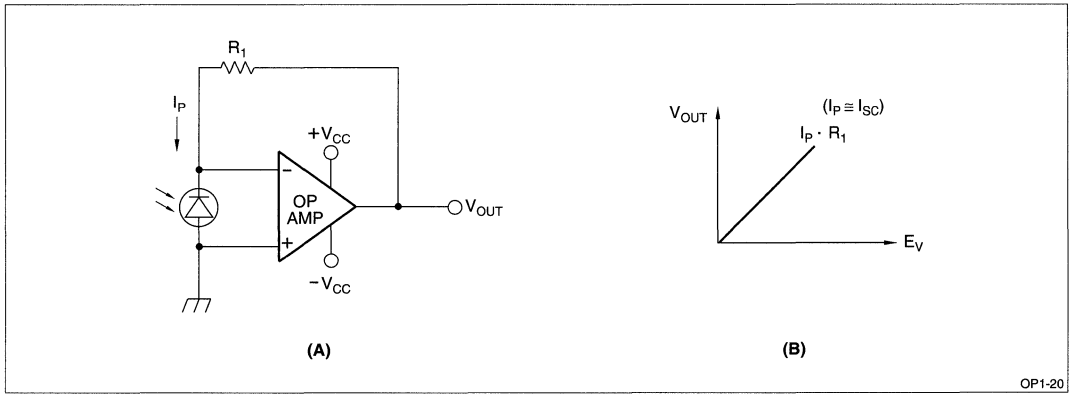


Figure 20. Photocurrent Amplifier using an Operational Amplifier (without bias)

Light Detecting Circuit For Modulated Light Input

Figure 22 shows a light detecting circuit which uses an optical remote control to operate a television set, air conditioner, or other devices. Usually, the optical remote control is used in the sunlight or the illumination of a fluorescent lamp. To alleviate the influence of such a disturbing light, the circuit deals with pulse-modulation signals.

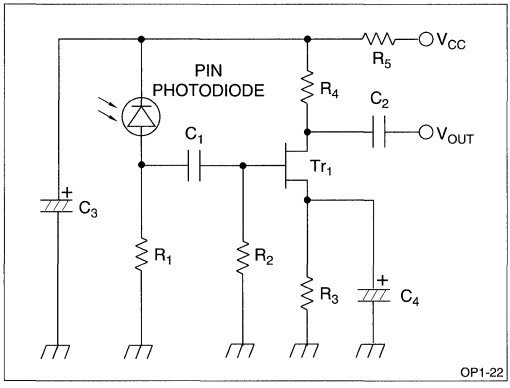


Figure 22. Light Detecting Circuit for Modulated Light Input PIN Photodiode

The circuit shown in Figure 22 detects the light input by differentiating the rising and falling edges of a pulse signal. To amplify a very small input signal, an FET proving a high input impedance is used.

Color Sensor Amplifier Circuit

Figure 23 shows a color sensor amplifier using a semiconductor color sensor. Two short circuit currents (I_{SC1}, I_{SC2}) conducted by two photodiodes having different spectral sensitivities are compressed logarithmically and applied to a subtraction circuit which produces a differential output (V_{OUT}). The output voltage (V_{OUT}) is formulated as follows:

$$V_{OUT} = \frac{kT}{q} \times \log\left(\frac{I_{SC2}}{I_{SC1}}\right) \times A$$

Where A is the gain of the differential amplifier. The gain becomes $A = R_2/R_1$ when $R_1 = R_3$ and $R_2 = R_4$, then:

$$V_{OUT} = \frac{kT}{q} \times \log\left(\frac{I_{SC2}}{I_{ISC1}}\right) \times \frac{R_2}{R_1}$$

The output signal of the semiconductor color sensor is extremely low level. Therefore, great care must be taken in dealing with the signal. For example, low-biased, low-drift operational amplifiers must be used, and possible current leaks of the surface of P.W.B. must be taken into account.

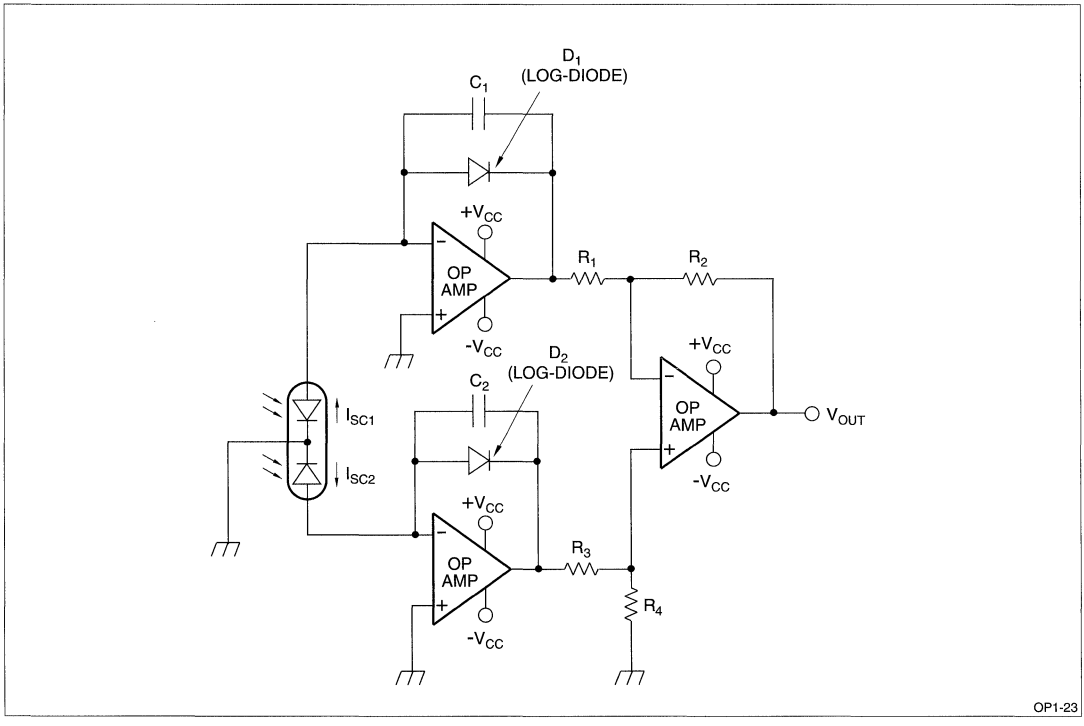


Figure 23. Color Sensor Amplifier Circuit

Fundament Phototransistor Circuits

Figures 24 and 25 show the fundamental phototransistor circuits. The circuit shown in Figure 24 (A) is a common-emitter amplifier. Light input at the base causes the output (V_{OUT}) to decrease from high to low. The circuit shown in Figure 24 (B) is a common-collector amplifier with an output (V_{OUT}) increasing from low to high in response to light input. For the circuits in Figures 24 (A) and 24 (B) to operate in the switching mode, the load resistor (R_L) should be set in relation with the collector current (I_C) as $V_{CC} < R_L \times I_C$.

The circuit shown Figure 25 (A) uses a phototransistor with a base terminal. A R_{BE} resistor connected between the base and emitter alleviates the influence of a dark current when operating at a high temperature. The circuit shown in Figure 25 (B) features a cascade connection of the grounded-base transistor (Tr_1) so that the phototransistor is virtually less loaded, thereby improving the response.

Amplifier Circuit Using Transistor

Figures 26 (A) and 26 (B) show the transistor amplifiers used to amplify the collector current of the phototransistor using a transistor (Tr_1). The circuit in figure 26 (A) increases the output from high to low in response to a light input. The value of resistor R_1 depends on the input light intensity, ambient temperature, response speed, etc., to meet the following conditions:

$R_1 < V_{BE}/I_{CEO}, R_1 > V_{BE}/I_C$

Where I_{CBO} is the dark current of phototransistor and I_C is the collector current.

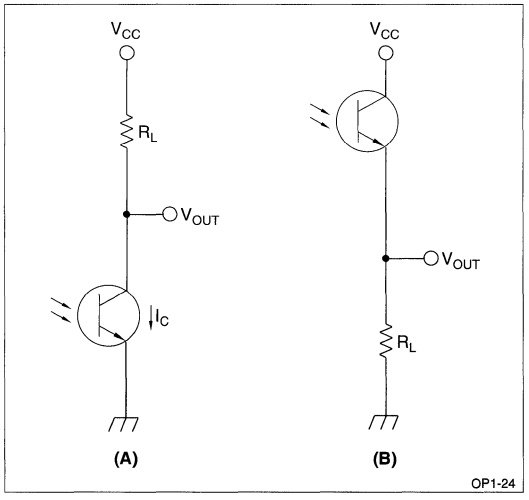


Figure 24. Fundamental Phototransistor Circuit (I)

Modulated Signal Detection Circuit

Figures 27 (A) and 27 (B) show the circuits used to detect a modulated signal such as an AC or pulse signal. The phototransistor has a base terminal with a fixed bias through resistors R_1 and R_2. An R_4 emitter resistor maintains the DC output voltage constant. A modulated signal provides a base current through bypass capacitor C causing current amplification so that the signal greatly amplified.

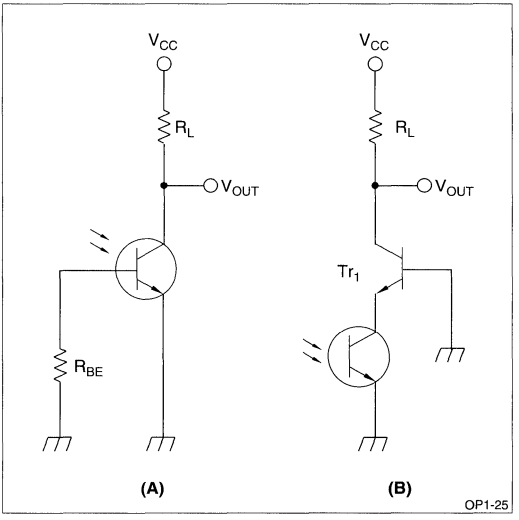


Figure 25. Fundamental Phototransistor Circuit (II)

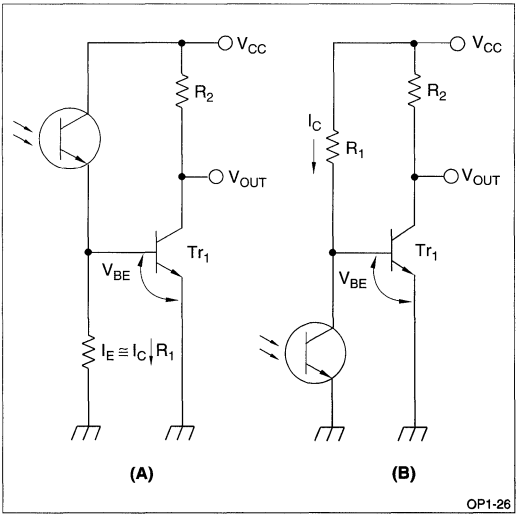


Figure 26. Amplifier Circuit Using Transistor

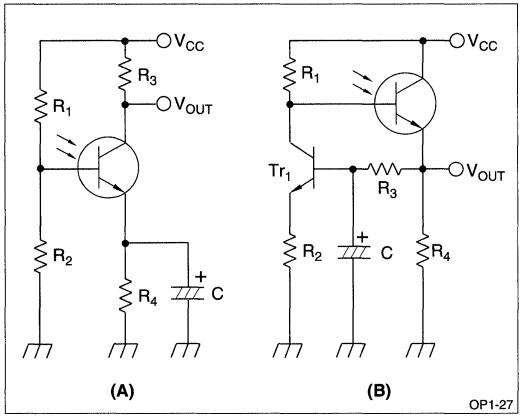


Figure 27. Modulated Signal Detection Circuit

Amplifier Circuit Using Operational Amplifier

Figure 28 shows a current-voltage conversion circuit using an operational amplifier. Its output voltage (V_{OUT}) is expressed as $V_{OUT} = I_C \times R_1$.

The current-voltage conversion circuit for the phototransistor is basically identical to that of the photodiode, except that the phototransistor requires a bias. The circuit shown if Figure 28 (A) has a negative bias (-V) for the emitter against the virtually grounded collector potential. Figure 28 (B) shows the output voltage vs. irradiation intensity characteristics.

Auto-stroboscope Circuit

Figure 29 shows the auto-stroboscope circuit of the current cut type. This circuit is most frequently used because of advantages such as continuous light emission and lower battery power consumption.

When the switch is in the ON-state, the SCR_2 and SCR_3 turn on to discharge capacitor C_4 so that the xenon lamp is energized to emit light. The anode of the SCR_2 is then reverse-biased, causing it to turn off and light emission of the xenon lamp ceases. The irradiation time is set automatically in response to variations in the collector current of the phototransistor. This follows the intensity of reflected light from the object and the value of C_1 in the circuit. In other words, the irradiation time is long for a distant object, and short for a near object.

PHOTOCOUPLER/PHOTOTHYRISTOR COUPLER/PHOTOTRIAC COUPLER APPLICATION CIRCUITS

For the effective use of photocouplers, the usage utilizing the features and fundamental circuits using photocouplers are described below.

Logic Gate Circuit Using Photocouplers

Figure 30 shows logic gates using photocouplers and their associated truth tables. The circuit of Figure 30 (A) forms an AND gate while the circuit of Figure 30 (B) forms an OR gate. These circuits are converted to a NAND gate and NOR gate, respectively, when the R_L load resistor is connected to the collector.

Level Conversion Circuit

Figure 31 shows simple level converters using a photocoupler. The circuit simple level converters using a photocoupler. The circuit shown in Figure 31 (A) converts the MOS level to the TTL level. Because of the small output current from the MOS IC, a photocoupler with a high current transfer ratio (CTR) at low input is required.

The circuit shown in Figure 31 (B) is a Schmitt trigger arranged using a photocoupler and transistor and a convert signal into an arbitrary level.

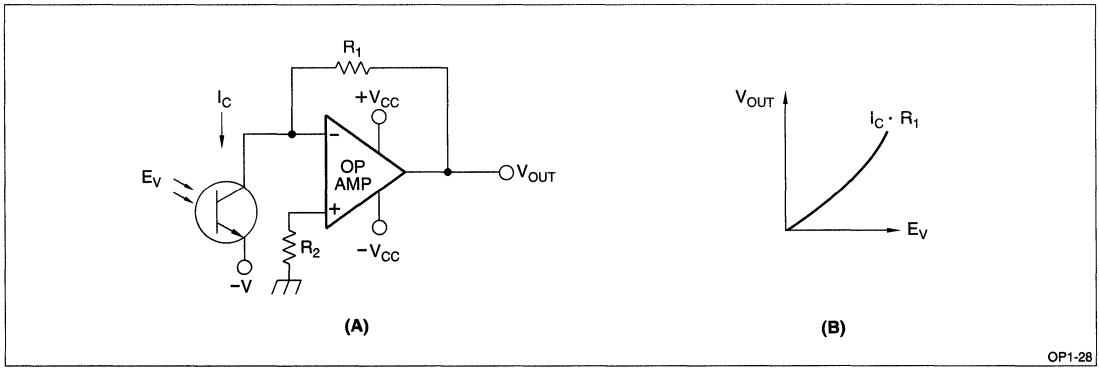


Figure 28. Amplifier Circuit using an Operational Amplifier

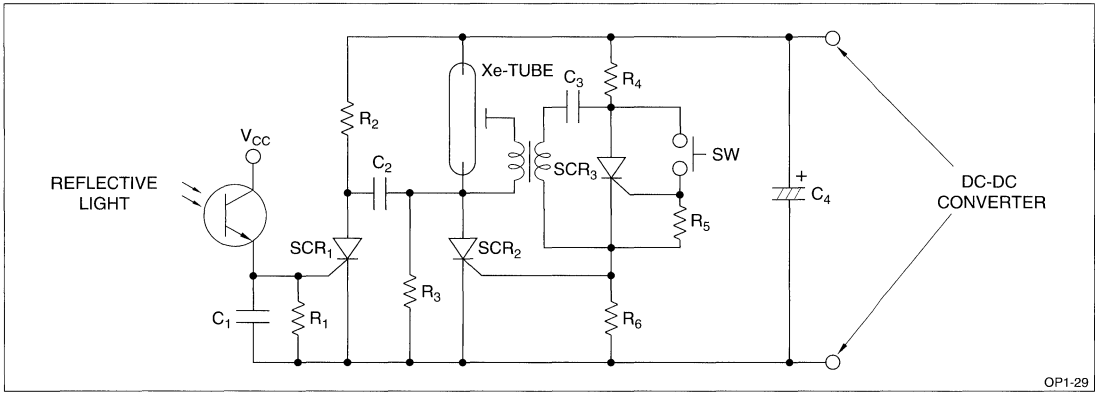
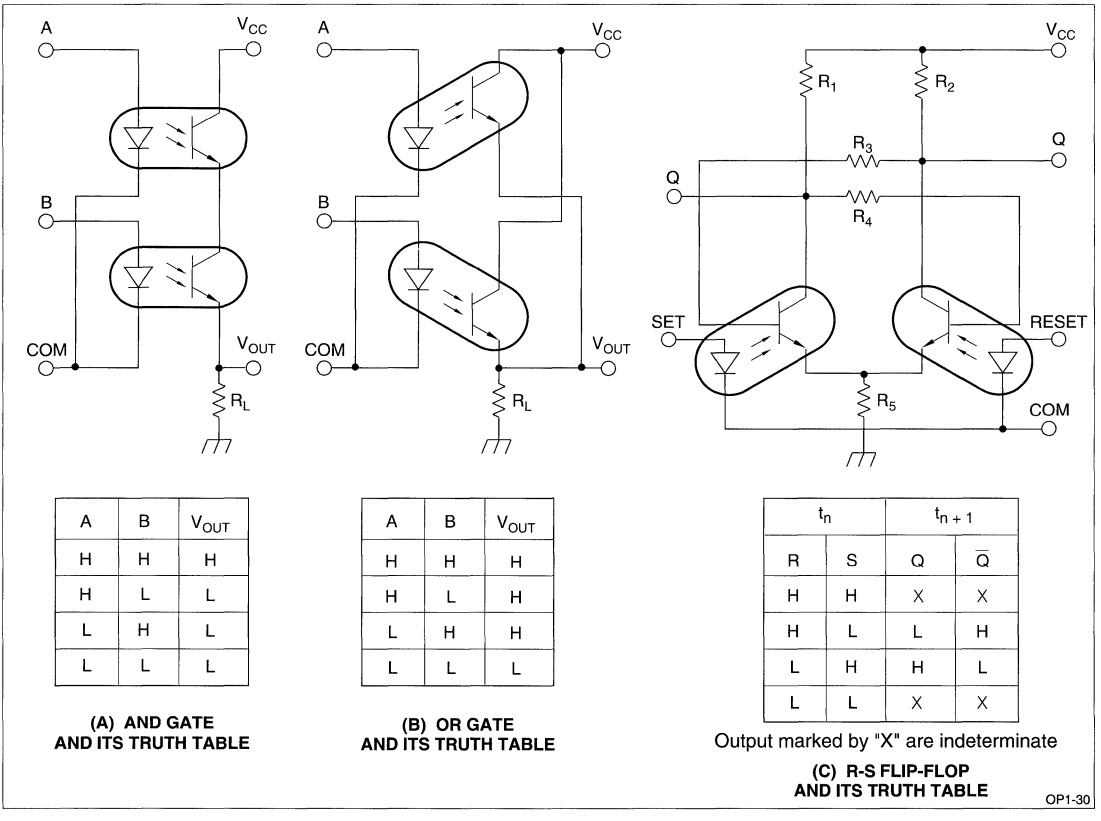


Figure 29. Auto-Stroboscope Circuit



A	B	V_{OUT}
H	H	H
H	L	L
L	H	L
L	L	L

(A) AND GATE AND ITS TRUTH TABLE

A	B	V_{OUT}
H	H	H
H	L	H
L	H	H
L	L	L

(B) OR GATE AND ITS TRUTH TABLE

t_n		t_{n+1}	
R	S	Q	$\overline{Q}$
H	H	X	X
H	L	L	H
L	H	H	L
L	L	X	X

Output marked by "X" are indeterminate

(C) R-S FLIP-FLOP AND ITS TRUTH TABLE

Figure 30. Logic Gate Circuits using Photocouplers

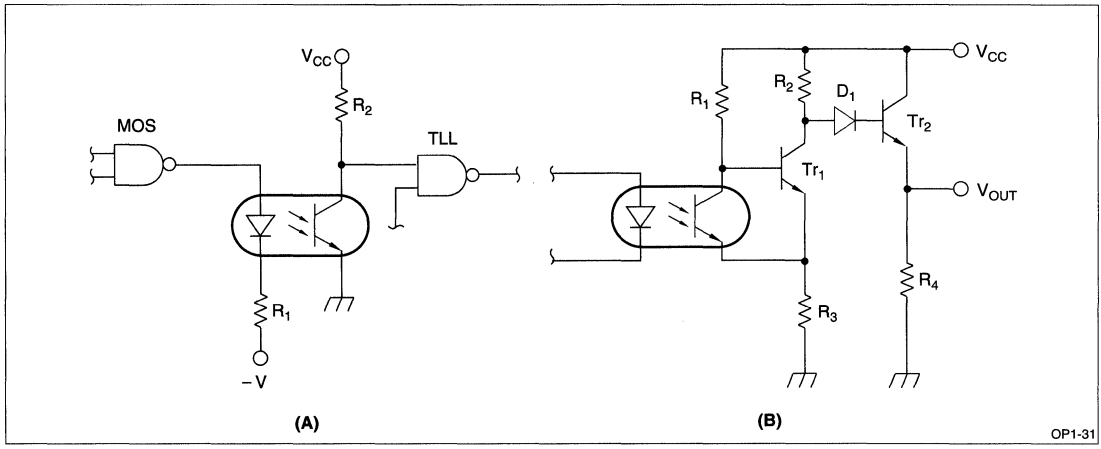


Figure 31. Level Conversion Circuit

Isolation Amplifier

Figure 32 shows a non-modulated isolation amplifier operable with low-frequency signals. In the arrangement, the photocoupler input is biased by DC forward current which is superimposed by a low-frequency signal. This gives the operating region of the good linearity of photocoupler. The DC bias current is adjusted by VR_1.

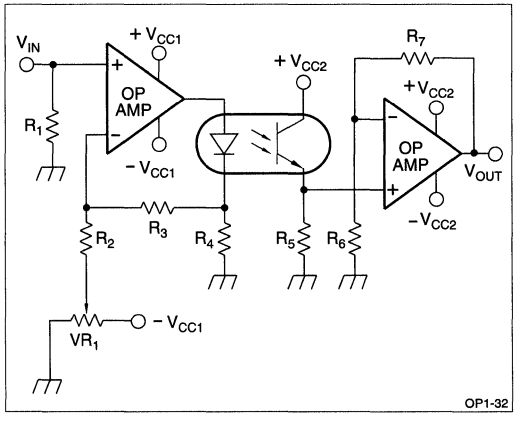


Figure 32. Isolation Amplifier

Noise Protection

Figure 33 shows some noise protection examples. The example shown in Figure 33 (A) includes the parallel connection of a capacitor (C_1) and resistor (R_1) across the input of the photocoupler where relatively long signal lines are connected for example where a computer and a terminal unit. The larger the capacitance of C_1, the greater the effect is expected, although signal propagation time is sacrificed.

The examples in Figure 33 (B) and 33 (C) use a photocoupler with a base terminal. Example (B) is effective against noise, but only in exchange for the response time, while example (C) tends to have low current transfer ratio (CTR).

However, when the photocoupler is operated in the switching mode, the base terminal tends to be affected by noise. Therefore, the use of photocouplers without a base terminal is recommended.

Lamp Driving Circuit and Relay Driving Circuit

Figures 34 and 35 show circuits for driving a lamp and relay, respectively, directly at the output of the photocoupler.

For this purpose, a suitable photocoupler includes a Darlington transistor providing a high CTR. The circuit shown in Figure 34 includes an R_2 resistor for supplying a preheating current to the lamp so as to prevent a rush current in lighting the lamp. The circuit in Figure 35 includes a diode D_1 for suppressing a counter-electromotive voltage produced when the relay is in the OFF-state.

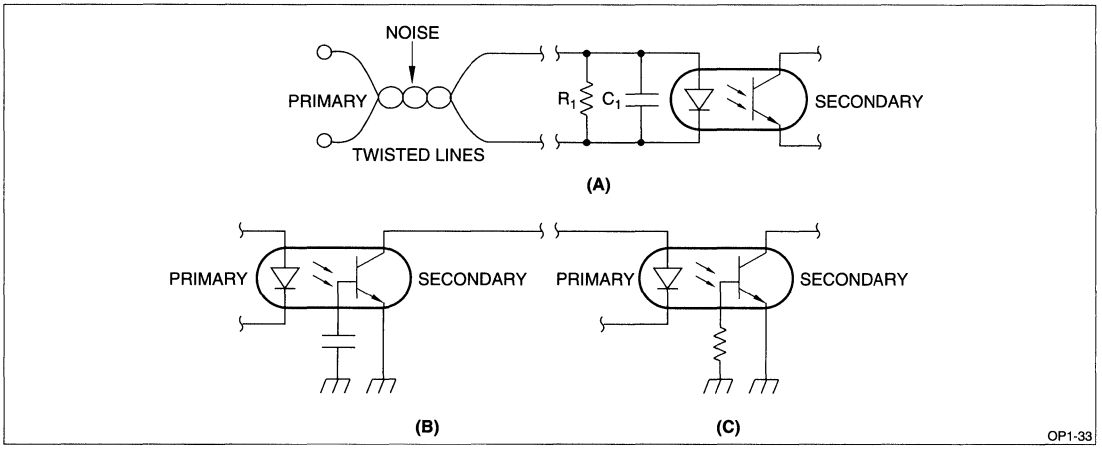


Figure 33. Noise Protection Example

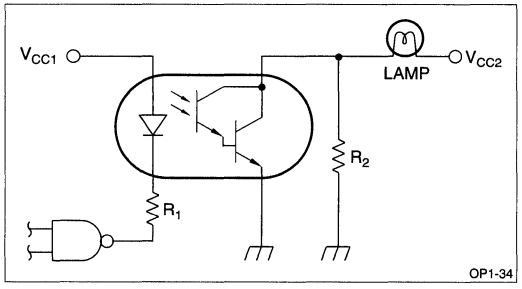


Figure 34. Lamp Driving Circuit

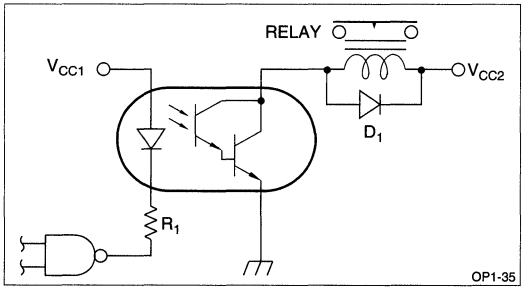


Figure 35. Relay Driving Circuit

Current Monitoring Circuit

The current monitoring circuit shown in Figure 36 is designed to detect and indicate leak current in a circuit using a photocoupler. The LED indicator lights off if the leak current exceeds the V_F/R_1 value.

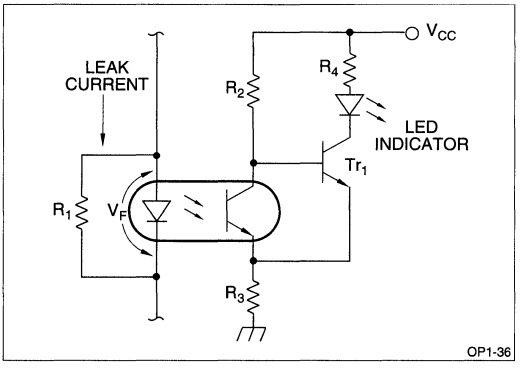


Figure 36. Current Monitoring Circuit

Solid State Relay

Solid State Relay Using Photocoupler

Figure 37 shows a solid state relay circuit using a photocoupler. Figure 37 includes an input circuit, photocoupler, thyristor for triggering, rectifying diode bridge, snubber circuit, and high power triac. In operation, the photocoupler turns on the thyristor for triggering and its ON-current activates the high power triac to drive the load. Because of a low collector withstand voltage and the low output current of the photocoupler, a thyristor for triggering is needed to interface it with power control devices such as a power triac or power thyristor.

By appropriately choosing the R_1 and R_2 values, a high sensitive solid state relay having a wide range of input signal of the photocoupler type is realized. The zero-cross voltage is determined from the voltage division ratio by R_4 and R_5.

Solid State Relay Using Photothyristor Coupler

Figure 38 shows the drive circuit of thyristor using a half-wave control type photothyristor coupler.

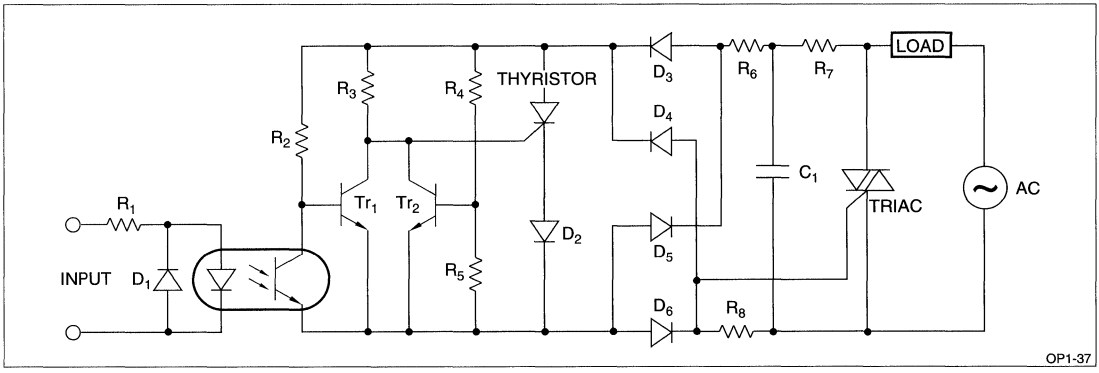


Figure 37. Solid State Relay with Built-in Zero-Cross Circuit

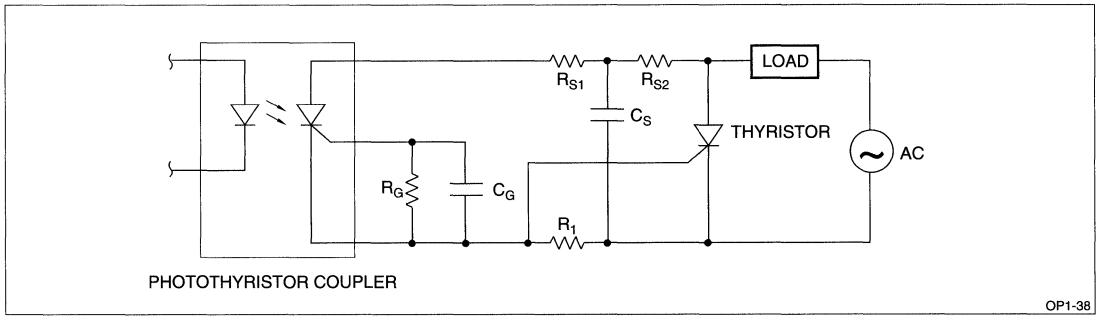


Figure 38. Large Power Thyristor Drive Circuit

Figure 39 shows the drive circuit of triac using a half-wave control type photothyristor coupler. In this circuit, D_1 ~ D_4 rectifying bridges are required for AC control using a half-wave control type photothyristor coupler.

Figure 40 shows the drive circuit of triac using a full-wave control type photothyristor coupler.

In each figure, R_1 is a resistor used to prevent mistriggering of a large power thyristor and triac by leak current (I_{DRM}) when the photothyristor coupler is OFF. Therefore, the setting is required by checking the photothyristor coupler (I_{DRM}) and gate trigger current (I_{GT}) of a large power thyristor and triac. R_{S1}, R_{S2} and C_S form a snubber circuit.

Solid State Relay Using Phototriac Coupler

Figure 41 shows the basic operating circuit of a triac using a phototriac coupler.

Figure 42 shows a circuit example of controlling forward and reverse rotation of the motor, using a control signal as one example of phototriac coupler application circuit.

Input Drive Circuit

Figure 43 shows the input drive circuit of a solid state relay (SSR). (A) and (B) operate with a positive signal, and (C) and (D) operate with a negative signal. (B) and (C) are effective when the output current of control circuit is small.

(E) is a drive circuit using IC (TTL/DTL), which operates when IC is in the "L" state.

(F) and (G) are drive circuits using CMOS IC, each of which cannot drive the primary side of SSR with CMOS IC only; it therefore drives via a transistor.

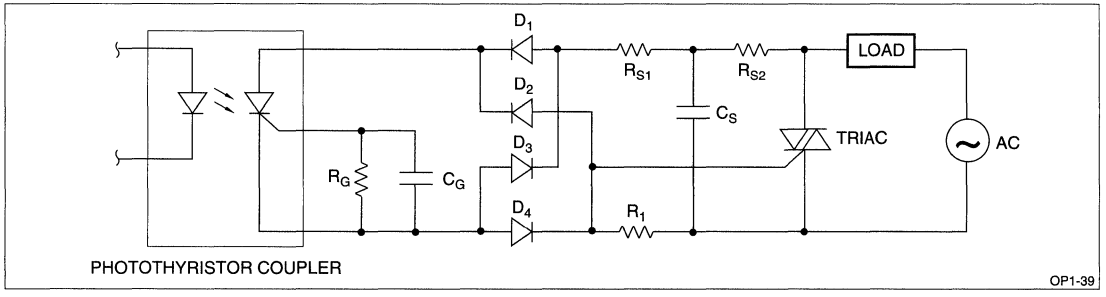


Figure 39. Triac Drive Circuit (I)

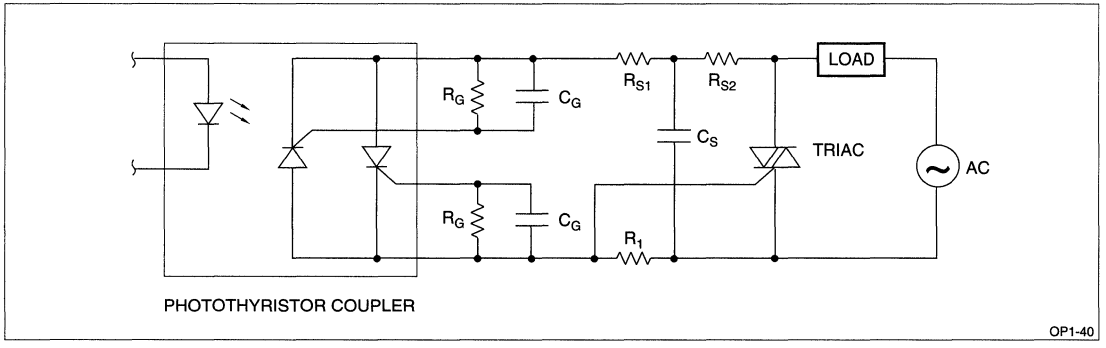


Figure 40. Triac Drive Circuit (II)

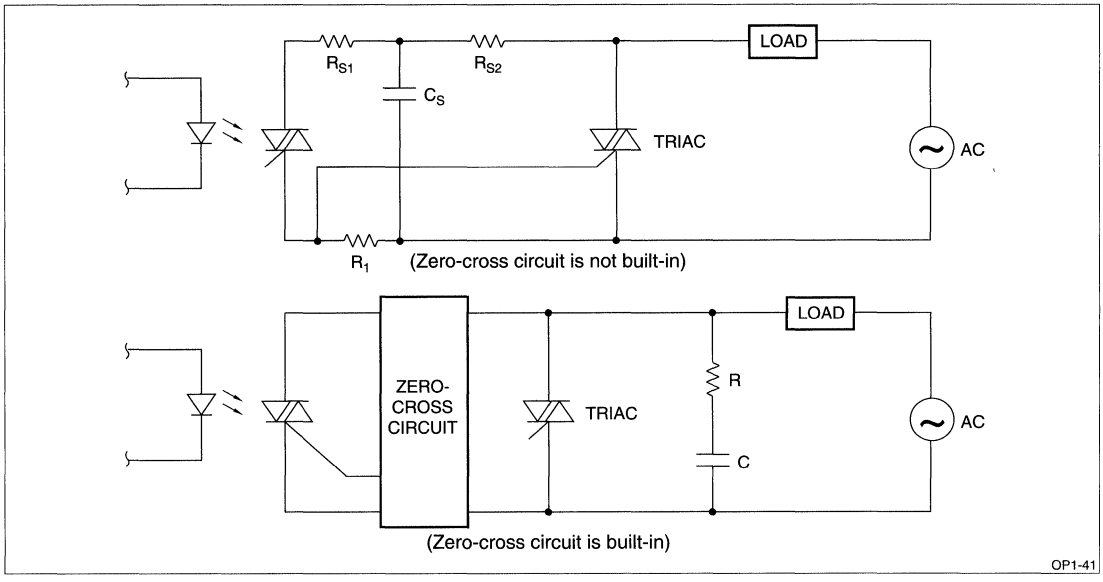


Figure 41. Triac Drive Circuit (III)

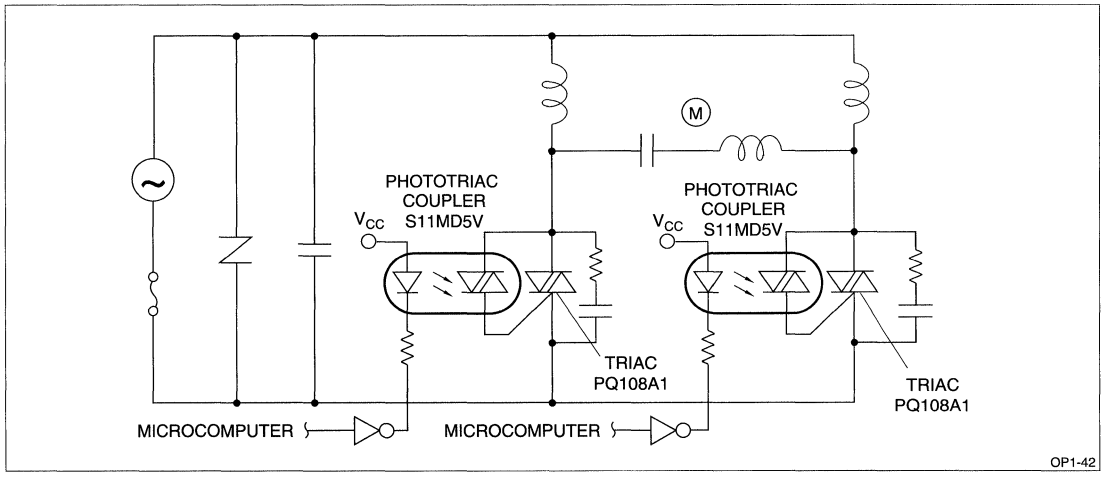
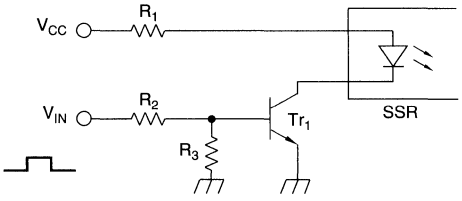
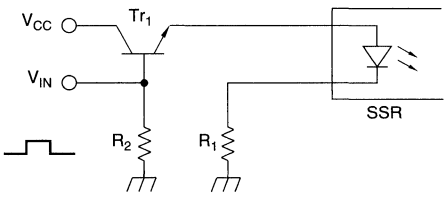


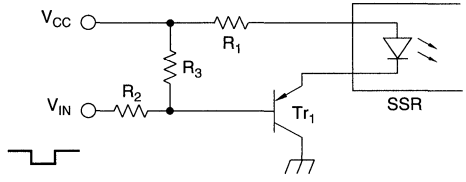
Figure 42. Motor Drive Circuit



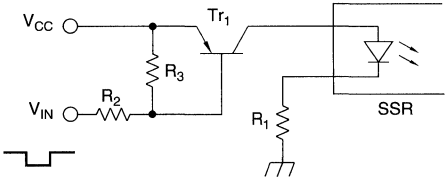
(A) NPN TRANSISTOR DRIVE (I)



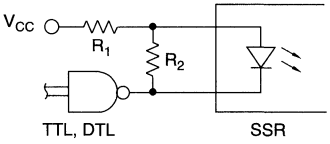
(B) NPN TRANSISTOR DRIVE (II)



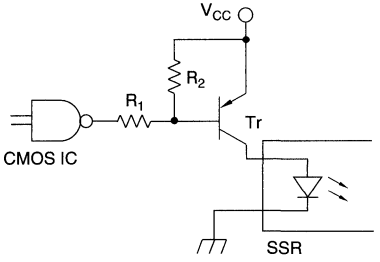
(C) PNP TRANSISTOR DRIVE (I)



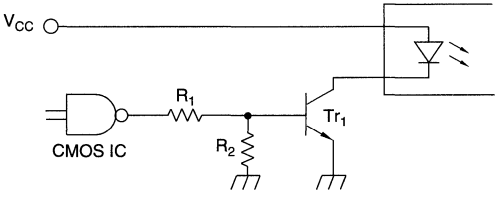
(D) PNP TRANSISTOR DRIVE (II)



(E) IC (TTL, DTL) DRIVE



(F) CMOS IC DRIVE (I)



(G) CMOS IC DRIVE (II)

OP1-43

Figure 43. Input Drive Circuit

Arrival Bell Signal Detection Of Telephone

Figure 44 shows a circuit for transmitting an arrival bell signal to a telephone related device while maintaining the electrical isolation between the device and the telephone subscriber line. The ring signal is an AC signal (75 Vrms, 16 Hz) superimposed on the 48 V line.

A non-polarized photocoupler (designed for AC input response) is suited for this purpose.

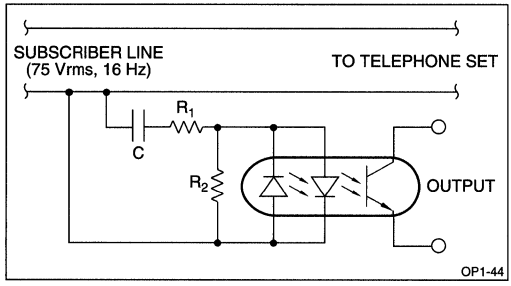


Figure 44. Telephone Arrival Bell Signal Detection

Telephone Line Interface

Figure 45 shows an interface circuit used to link a telephone related device to the telephone line. Through parallel connections of photocouplers, telephone related devices can be linked to the telephone line.

Telephone Line Polarity Detection (Ring Counter)

Figure 46 shows an example of a photocoupler used for the polarity detecting circuit in a telephone line.

Dial Pulse Monitor Circuit

Figure 47 shows an example in which a photocoupler is actuated due to dial pulse current if the circuit is connected to the telephone line, the light detector side of photocoupler operates as a dial pulse monitor circuit.

Power Control Circuit By Bell Signal

Figure 48 shows an application example for ON/OFF switching of the power supply of a particular equipment by a telephone bell signal.

Servo Motor Driving Circuit

Figure 49 shows an inverter-type AC servo motor speed control circuit. A transistorized inverter is featured to readily control an AC motor in a wide speed range. It is increasingly used in appliances such as air conditioners.

The photocoupler is used to drive the power transistor base amplifier so that it interfaces with a microcomputer. Because of the high surge voltage applied to the PWM base signal circuit (input) and driver circuit (output) at the switching of magnetic polarity, a high noise resistance (high dv/dt) photocoupler is used.

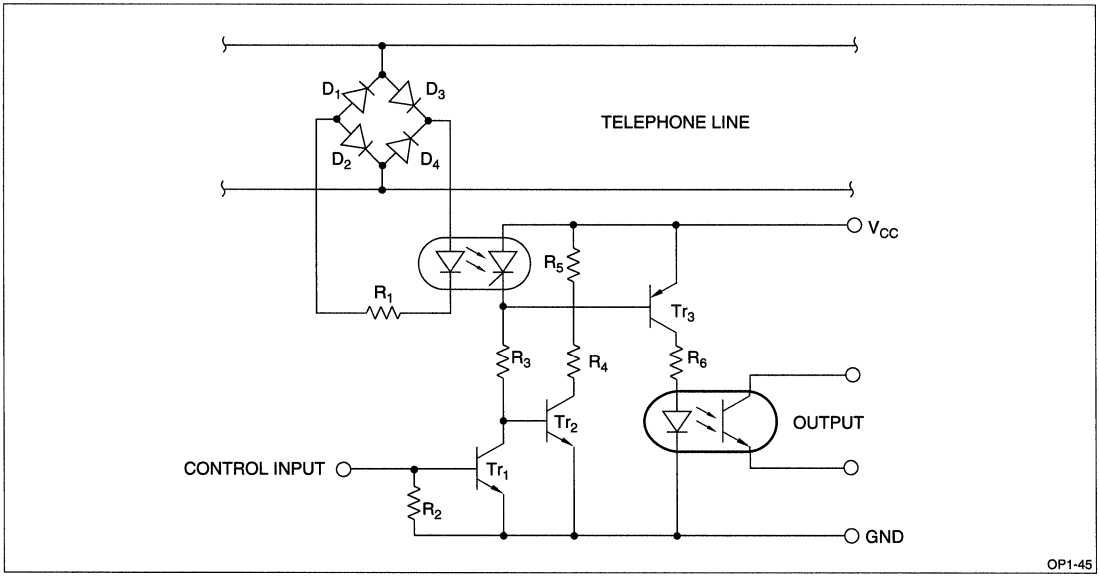


Figure 45. Telephone Line Interface

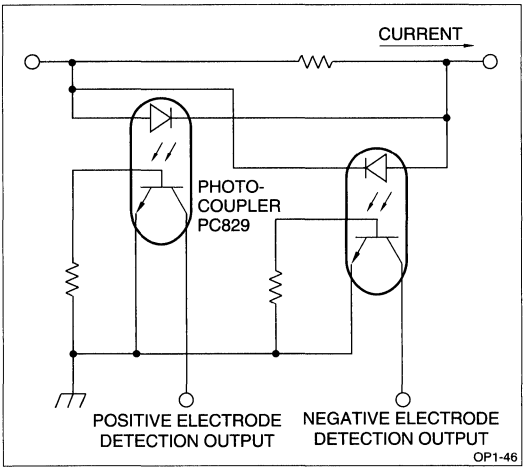


Figure 46. Telephone Line Polarity Detection Circuit

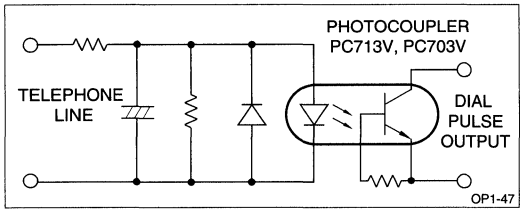


Figure 47. Dial Pulse Monitor Circuit

Servo Motor Braking Control Circuit

Figure 50 shows a servo motor braking control circuit in which a photocoupler is used to separate the control circuit from the brake driving circuit. A serial connection of C_2 and R_7 across the coil is designed to absorb the inductive current by the coil. C_1 is used to absorb high frequency noise on the DC power line.

Switching Regulator Circuit

Figure 51 shows a switching regulator circuit using a photocoupler.

In operation, the AC power line voltage is rectified into a DC voltage and is inverted into an AC voltage of around 50 kHz. It is then converted back to a DC voltage by a choke-input rectifying circuit. The output voltage is determined by the values of R_1, R_2, and ZD.

Chopper Circuit

Figure 52 shows a chopper circuit featuring high response and low signal amplification.

Conventional choppers are formed by FETs and transistors and create problems by switching spike noise which adversely affects the output signal.

Use of a photocoupler allows electrical isolation of the control and amplifying circuits. A small signal can then readily be amplified with no affect from spike noise.

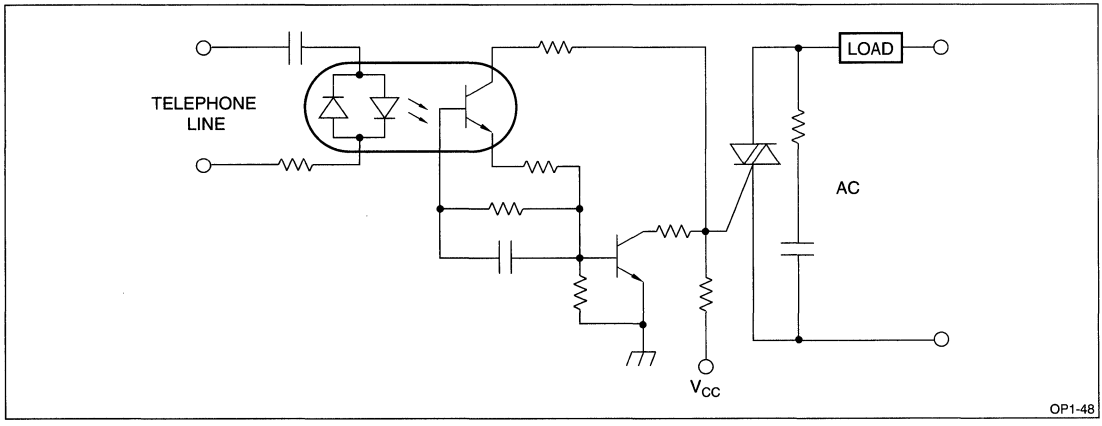


Figure 48. Power Control Circuit by Bell Signal

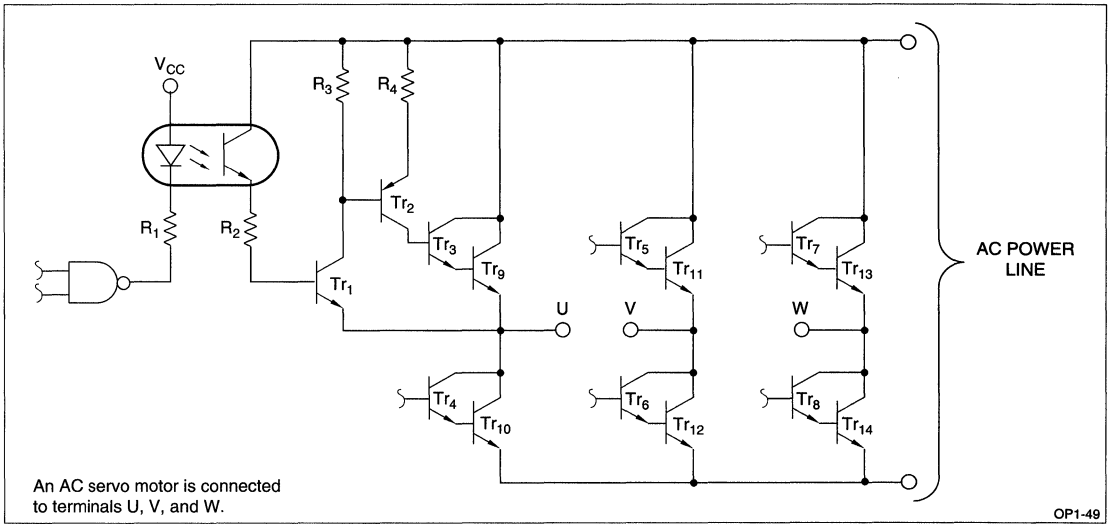


Figure 49. Servo Motor Driving Circuit

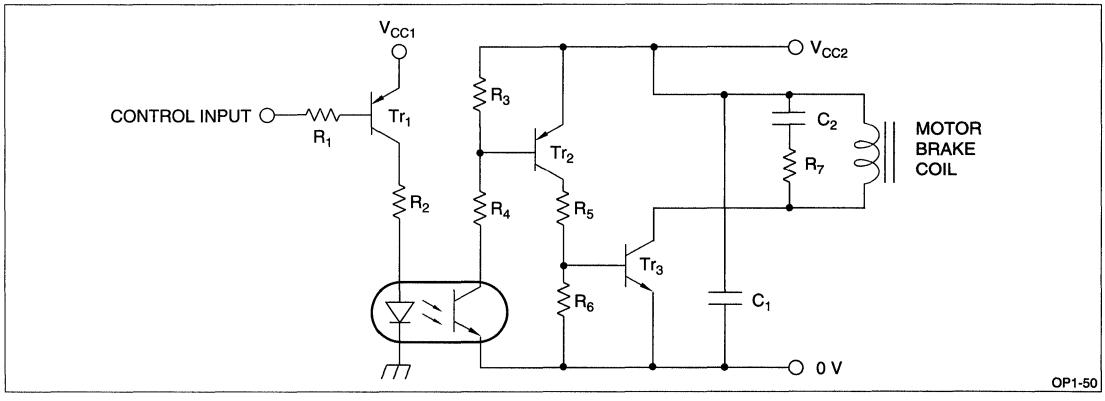


Figure 50. Servo Motor Brake Control Circuit

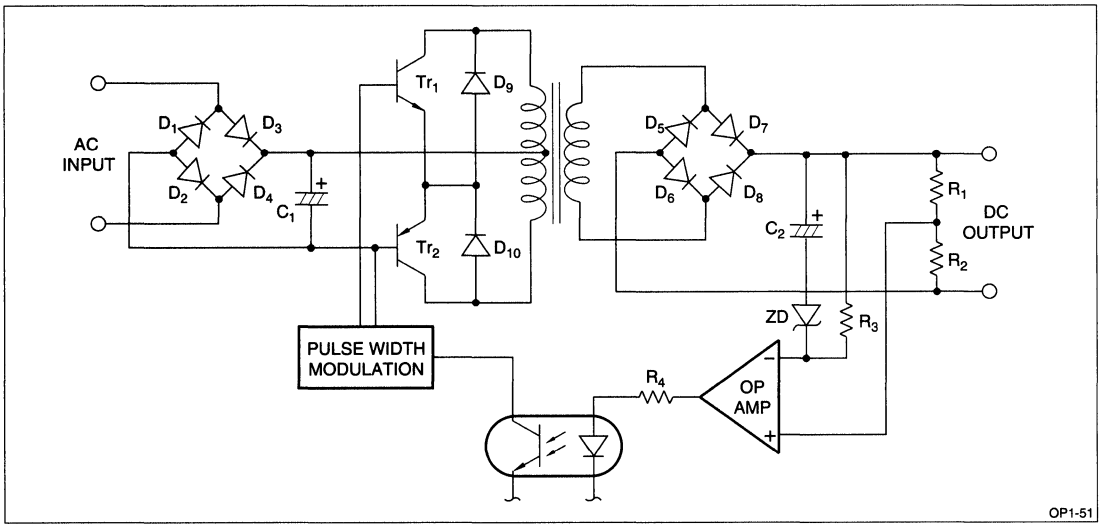


Figure 51. Switching Regulator Circuit

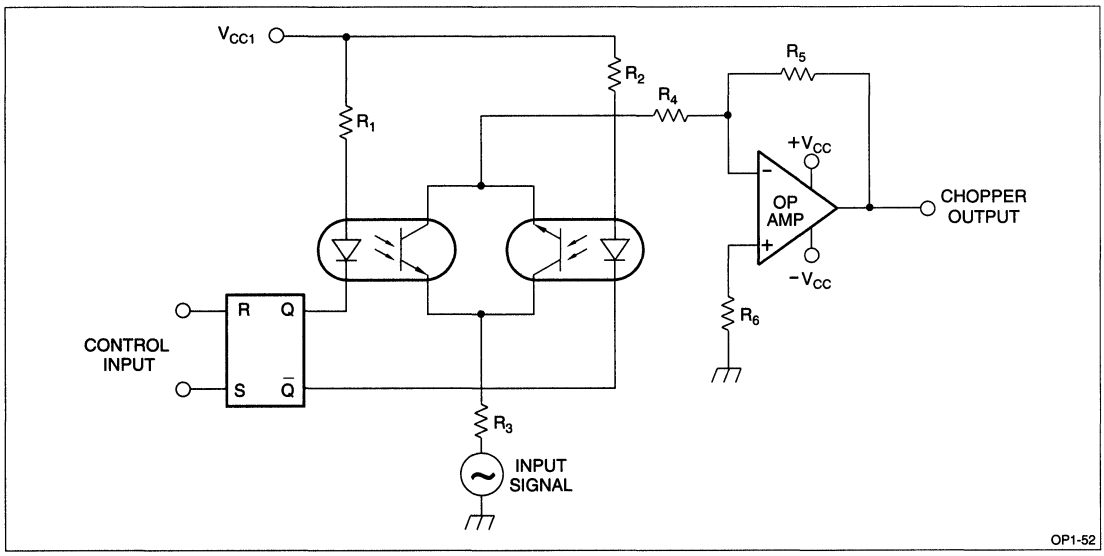


Figure 52. Chopper Circuit

Electrostatic Printer Control Circuit

In an electrostatic printer, the print head driving circuit, which operates at high voltage, is separated from the control circuit.

A photocoupler with a high isolation voltage between the input and output is useful in the electrostatic printer control circuit. Figure 53 shows an electrostatic printer control circuit using a photocoupler.

Photocoupler Application Fields

Table 1 summarizes the industrial applications of the photocoupler.

PHOTOINTERRUPTER APPLICATION CIRCUITS

Photointerrupters are used to detect the passage or existence of an object. Accordingly, an output digital signal, i.e., high or low, is required in most applications.

Photointerrupters fall into two categories: transmissive and reflective. Many variations in performance, such as detecting gap width, resolution and focal distance are provided to meet various application needs.

The fundamental circuits for operating a photointerrupter are described below.

DC Signal Processing Circuits

Figure 54 shows signal processing circuits for a transmissive type photointerrupter.

This type of photointerrupter provides an output signal with a relatively large S/N ratio, allowing detection of an object with a simple circuit as shown below, Circuits (A), (B) and (C) are used with a logic circuit, while (D) is used in the case of a large current leaks. The voltage division by resistors R_3 and R_4 determine the threshold level of detection.

AC Signal Processing Circuits

Figure 55 shows the signal processing circuits of a reflective type photointerrupter.

This type of photointerrupter provides a very small output signal with an inferior S/N ratio. Therefore, the AC signal processing circuit eliminates disturbing light and amplifies only a varying signal.

Circuit (A) is fixed level slicing, (B) is a floating level slicing, and (C) is light level compensating circuit.

Detection Of Moving Objects

Figure 56 shows a circuit example including a transmissive type photointerrupter for detecting a moving object, such as a coin. In operation, the output (V_{OUT}) is high when the object is absent and low for a certain period as the object passes through the gap of the photointerrupter. The duration of a "low" signal is determined from values C_2 and R_4.

Detection Of Paper Moving Direction And Speed

The circuit in Figure 57 includes two photointerrupters disposed at a certain interval so that the moving direction and speed of a sheet of paper are detected by the outputs of the circuit.

Detection Of A Tape Edge

The circuit in Figure 58 is designed to correct the edge position of tape when wound on a reel. Two photointerrupters are located at both edges of the tape. The circuit provides differential output V_{OUT1} and V_{OUT2}. The reel position is controlled in the traversing direction so that the difference between V_{OUT1} and V_{OUT2} is zero.

Detection Of Disk Rotation

Figure 59 shows a circuit designed to detect the number of rotations of a disk using a transmissive type photointerrupter. The slit signal is converted into digital signal as shown in Figure 60 and is used for motor speed control.

Detection Of Arm Angle In Record Player

Figure 61 shows a mechanism for correcting the angular deflection of the player tone arm using the differential outputs of two photointerrupters so that the tone arm is always parallel to the tangent of the grooves in the disk.

Detection Of Cassette Tape End

Figure 62 shows a circuit designed to detect the end of cassette tape using a reflective type photointerrupter so that the tape driving motor is deactivated at the end of the tape.

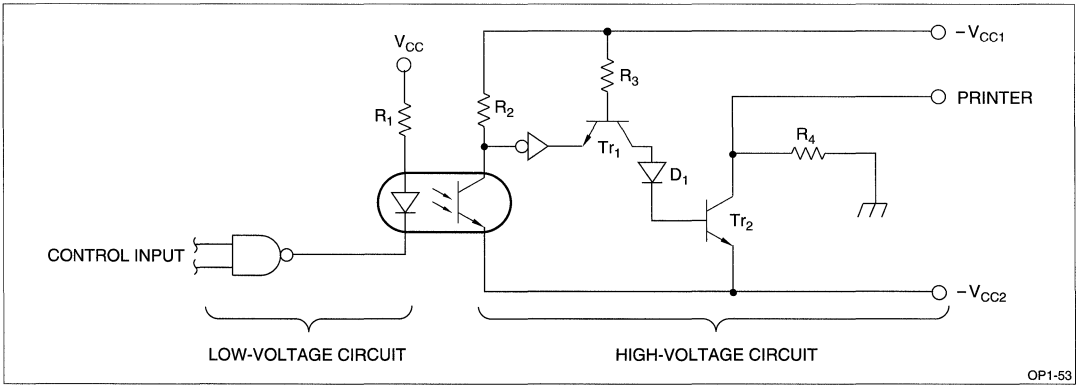


Figure 53. Electrostatic Printer Control Circuit

Table 1.
Photocoupler Application Fields

FIELD		EQUIPMENT	APPLICATIONS
A	Computer peripheral	Computer peripherals and I/O units	Interface circuit between computer and peripheral
			Battery backup circuit
B	Control equipment	Programmable controllers, numerical control machines	Isolation circuit in input unit
			Contact input circuit for small signal
			Isolation of signal transmission system
			Servo motor control circuit
		Power control, distribution board	Current monitoring circuit
			Contact input circuit
			Noise protection circuit
			AC power line monitoring circuit
			Ground isolation
		Elevator, auto-door	Isolation of signal transmission system
			Auto-door control circuit
		Others	Self-hold switch circuit
			Lamp and relay driving system
C	Instrument	Measuring, testing instruments	Isolation of signal transmission system (line driver, line receiver)
			I/O isolation of isolation amplifier
			Inductive noise protection circuit
			Level conversion circuit
D	Office equipment	Copiers, facsimiles	Ground isolation
			Power circuit (primary-secondary isolation)
		Printers	High voltage control circuit of electrostatic printer
			Printer drive circuit
			I/O interface
E	Automatic vendors		I/O interface
			Commodity/ticket selection circuit
F	Home appliances	Television sets	Audio multiplexing circuit isolation
			R - G - B signal interface
			Power circuit
		Electronic sewing machines	Motor control circuit
		Microwave oven, room heaters	Ground isolation
			I/O interface
		Air-conditioners	Inverter control base amplifier circuit
			Over-current detection circuit
G	Audio equipment	Players, cassette tape recorders	Power circuit (primary-secondary isolation)
			Isolation of signal transmission system
H	Telephone system	Telephone sets	Dial pulse monitoring circuit
			Ring signal counter circuit
			Chiming circuit
			Modem switching circuit
		Exchangers	Subscriber line/control system separation
I	Power supply unit	Switching regulators	Pulse width modulation circuit
			Feedback circuit
			Isolation between primary and secondary
J	Others	SSR	Isolation between primary and secondary
		Motor control	Over-current detection circuit of induction motor
			Braking control circuit of servo motor
			Chopper circuit

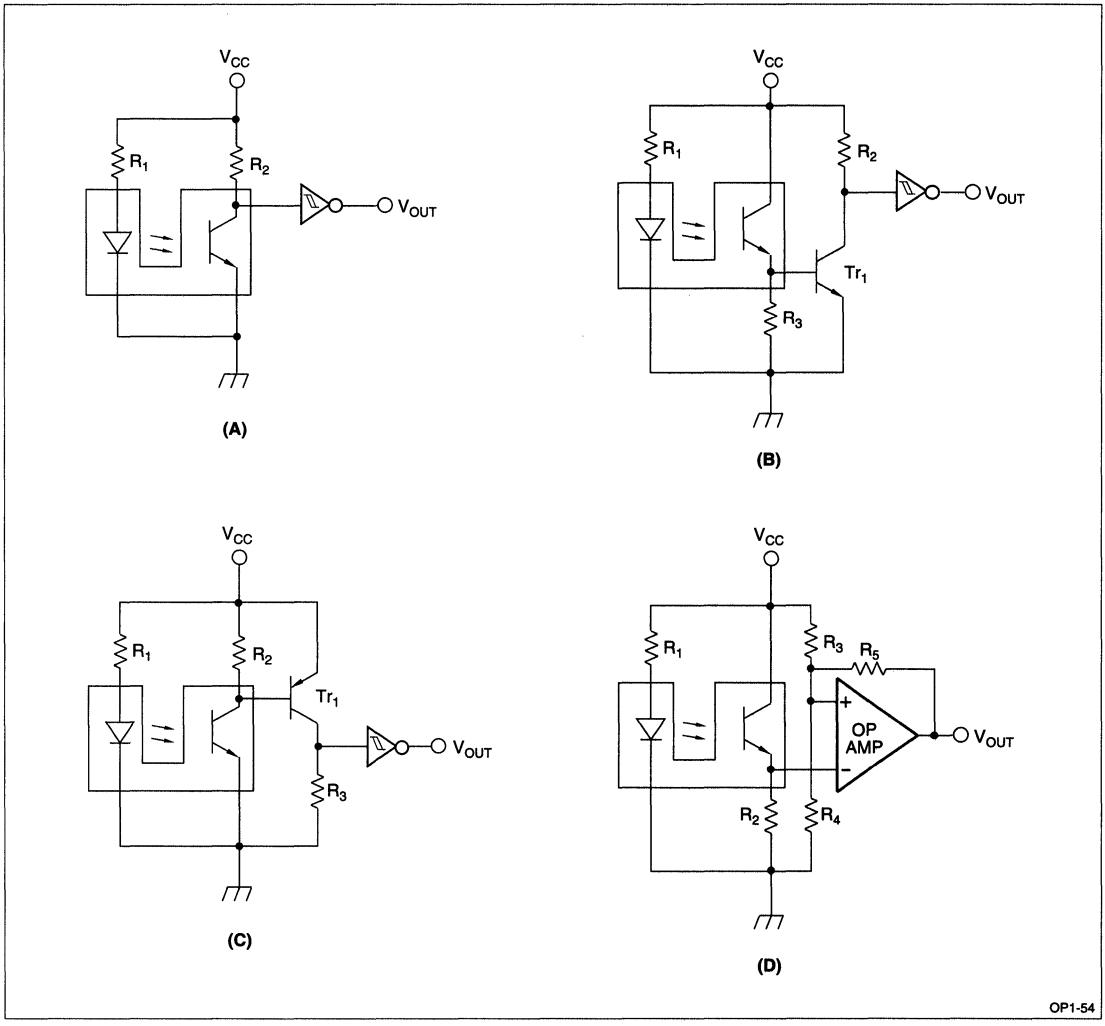


Figure 54. DC Signal Processing Circuit

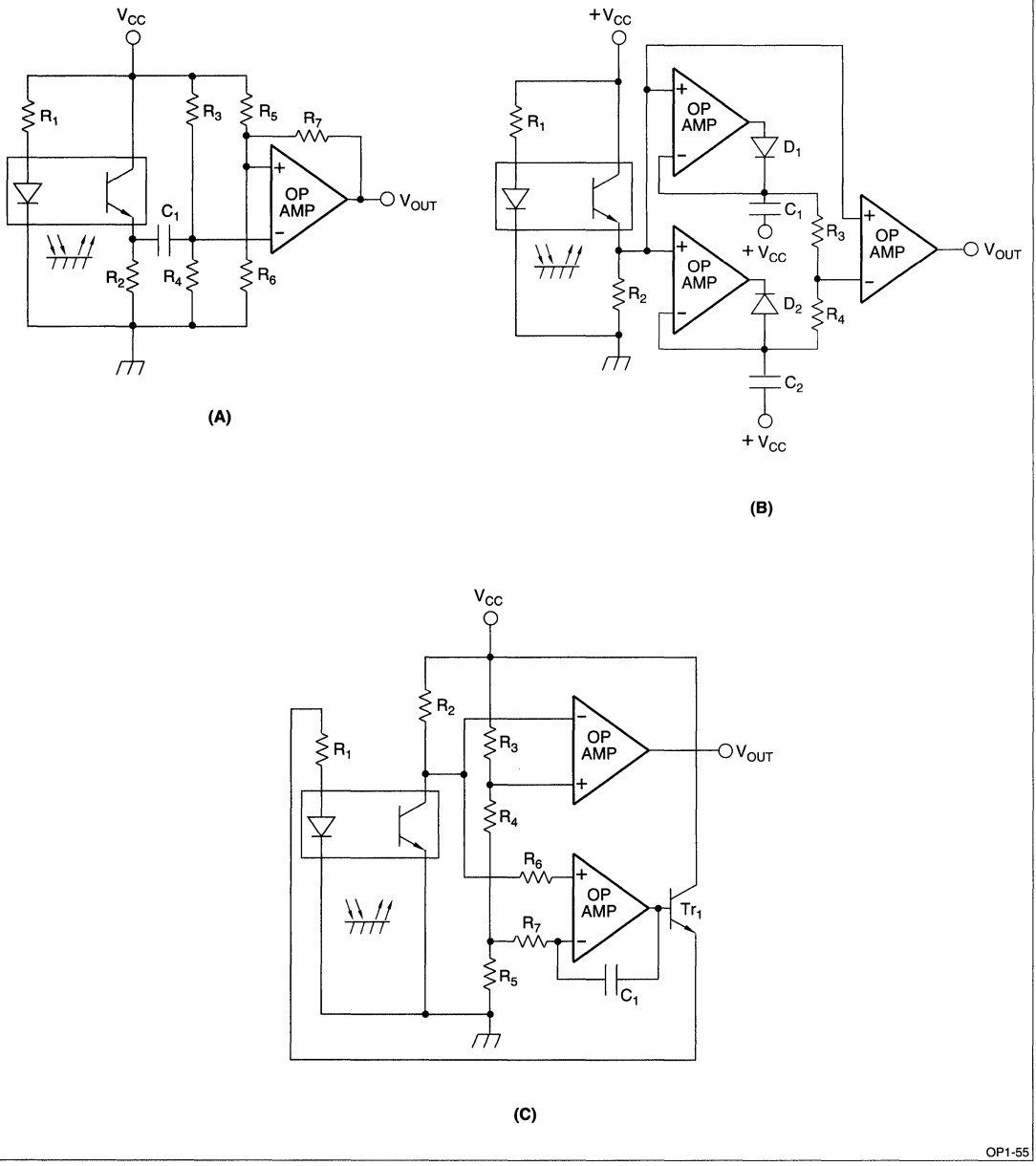


Figure 55. AC Signal Processing Circuits

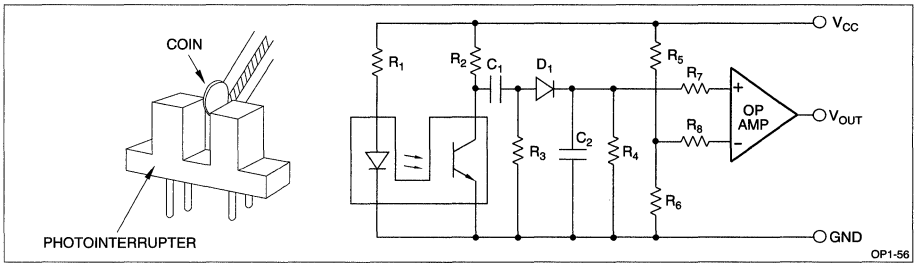


Figure 56. Detection of Moving Objects

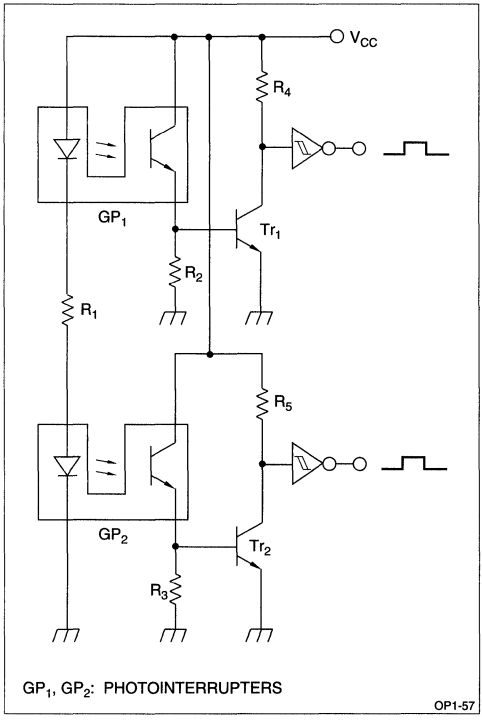


Figure 57. Paper Moving Direction and Speed Detection

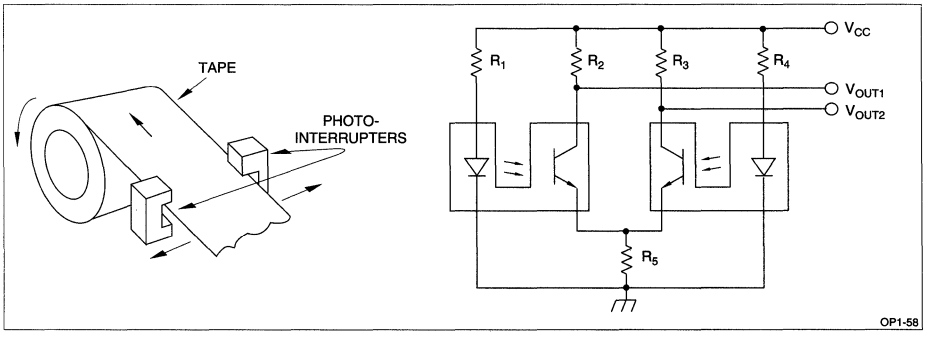


Figure 58. Tape Edge Detection

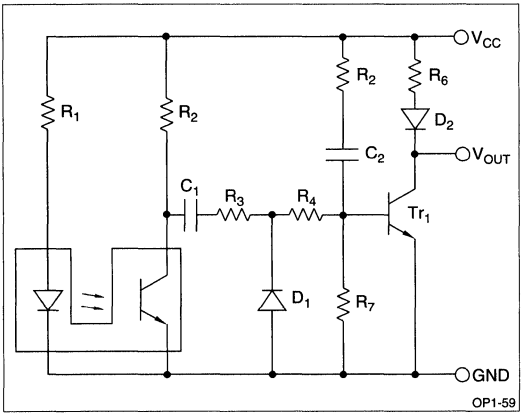


Figure 59. Detection of Disk Rotation

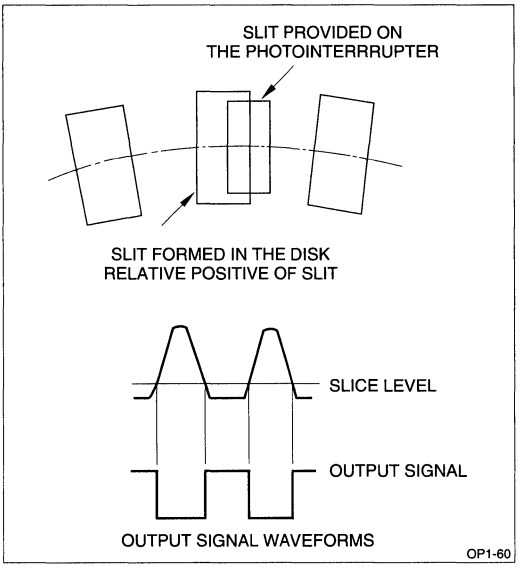


Figure 60. Detected Signal Waveform using Disk

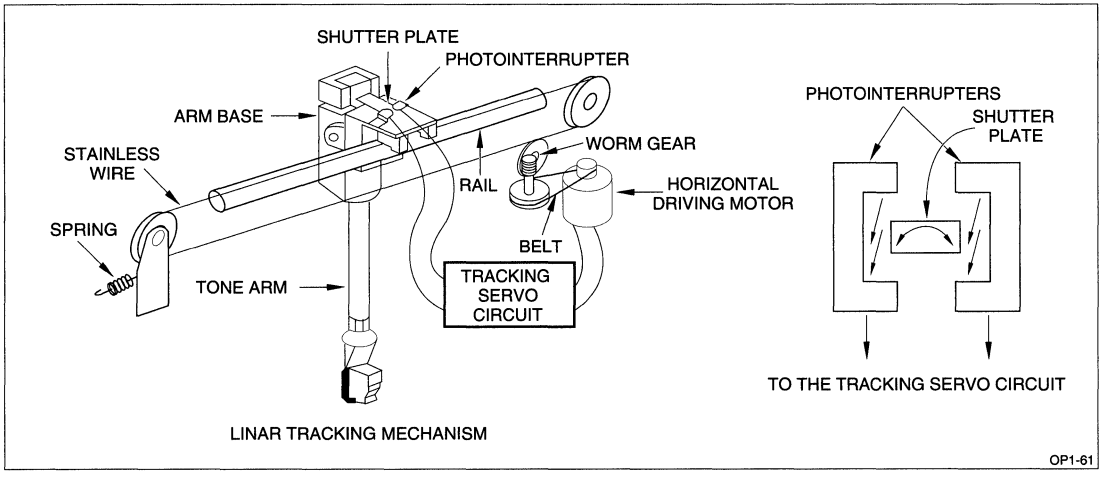


Figure 61. Arm Angle Position Detection in Record Player

Weight Detection In Electronic Balance

Figure 64 shows a weight detection example in electronic balance. Two photointerrupters are used together to count the number of slits formed in the disk and to discriminate the upward and downward movement of the tray.

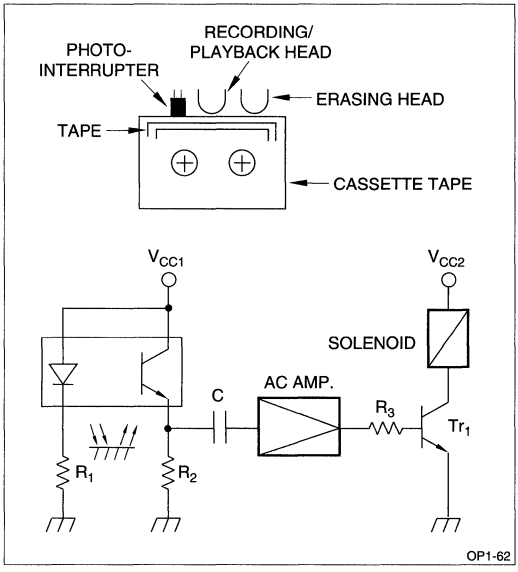
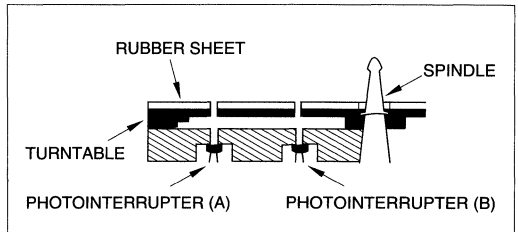


Figure 62. Tape End Detection in Cassette Desk



DISK SIZE	PHOTOINTER-RUPTER (A)	PHOTOINTER-RUPTER (B)	INSTRUCTION
None	OFF	OFF	Do not lower the tone arm
30 cm disk	ON	ON	30 cm disk position detecting circuit: ON, 33 1/3 rpm
17 cm disk	OFF	ON	17 cm disk position detecting circuit: ON, 45 rpm

OP1-63

Figure 63. Disk Size Detection

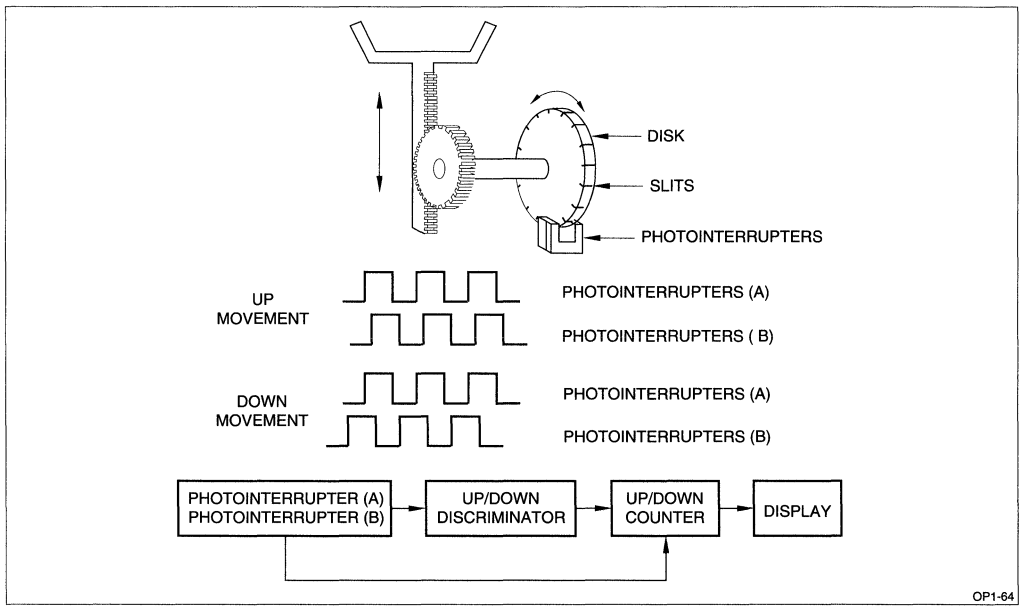


Figure 64. Weight Detection in Electronic Balance

Speed Alarm Device

Figure 65 shows the mechanism and circuit of a speed alarm device in which a disk is provided coaxially on the speedometer pointer and the associated photointerrupter responds to the output section of the disk to activate the acoustic element.

Upper/lower Limit Detection In Measuring Instrument

Figure 66 shows a mechanism and circuit for detecting the upper and lower limit on the instrument scale indicated by the pointer. The instrument pointer is provided with a mask tab which moves in the gap of the photointerrupters located at the upper and lower limit positions of the scale. The circuit lights a red LED when the pointer reaches the upper limit, a yellow LED when the lower limit is reached, or a green LED when the pointer is within the proper range.

Card Mark Reader

Figure 67 shows a card mark detecting circuit operating according to the floating level slice system using the envelope circuit. The circuit is sensitive and less affected by paper quality and smears. Yet it performs consistently even if the sensor output falls by half.

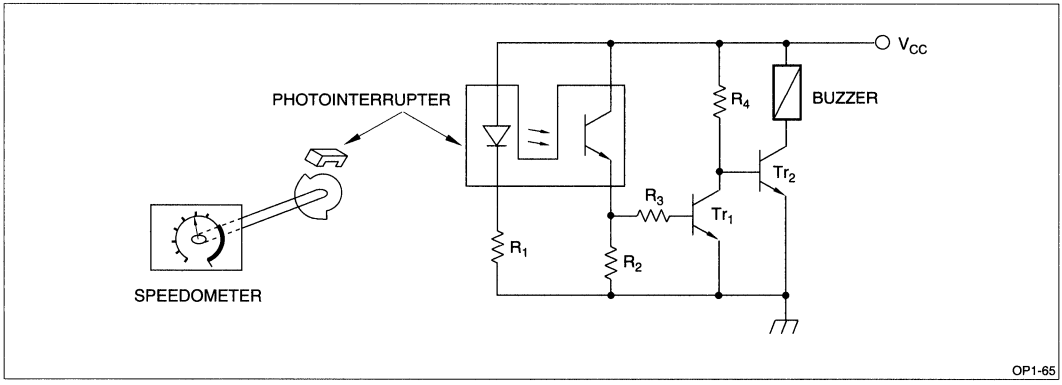


Figure 65. Speed Alarm Device

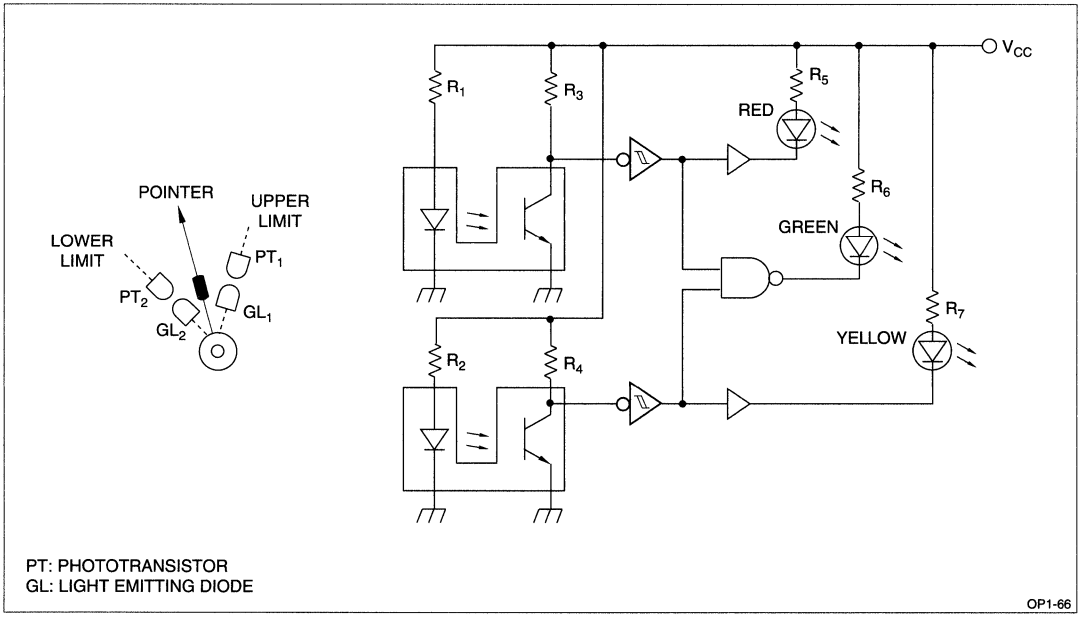


Figure 66. Upper/Lower Limits Detection in Instrument

Although the marking must be done carefully, particularly to prevent faint or thin marks and imperfect erasure, any spot or mark outside the specified entry position must be disregarded by referring to the timing signal.

Figure 67 shows the circuit for a mark reader.

Figure 68 shows the signal waveforms observed at various parts of the circuit.

Copy Paper Feed Detection In Copier

Figure 69 shows circuit example using a reflective type photointerrupter designed to detect paper fed from the paper stocker.

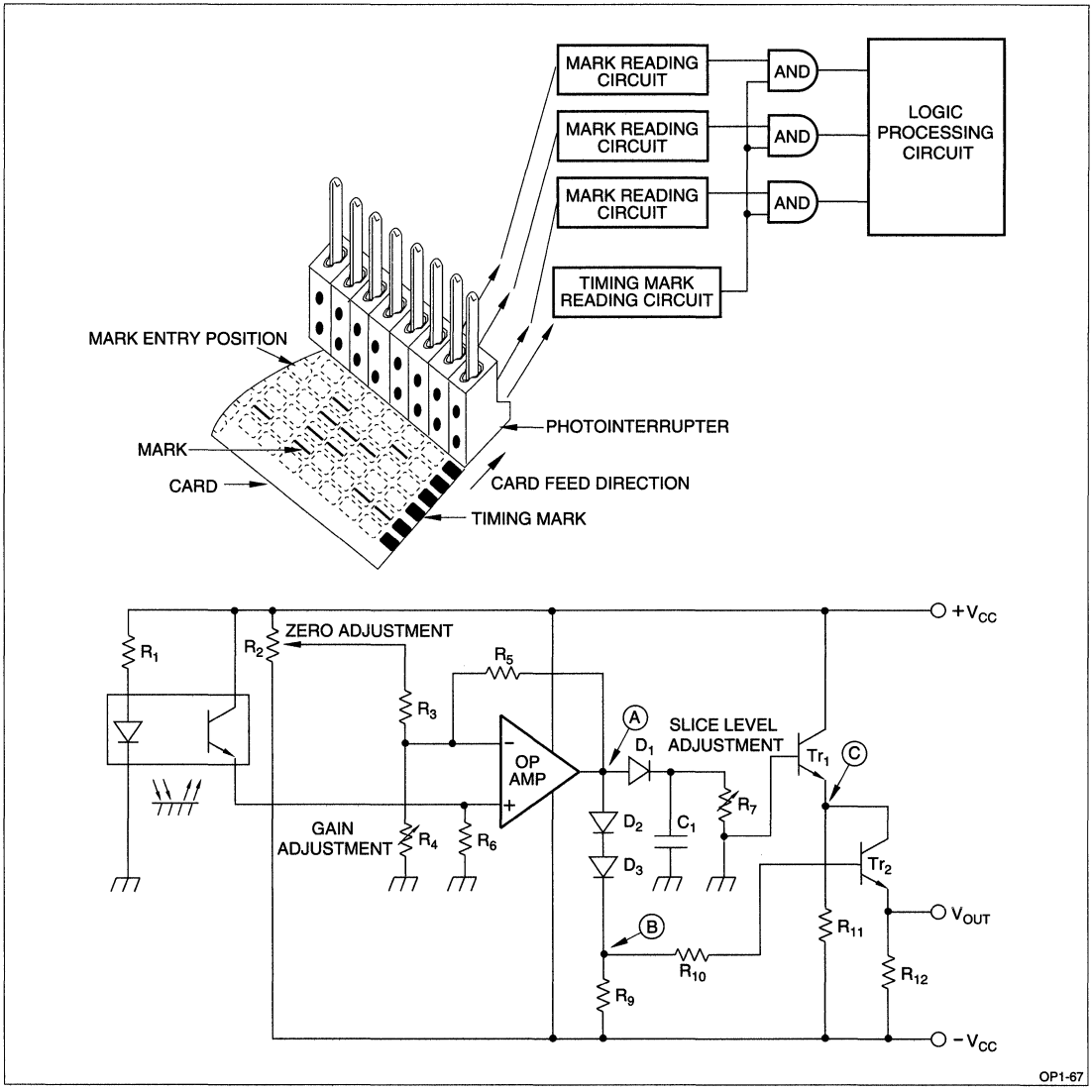


Figure 67. Card Mark Reading Circuit

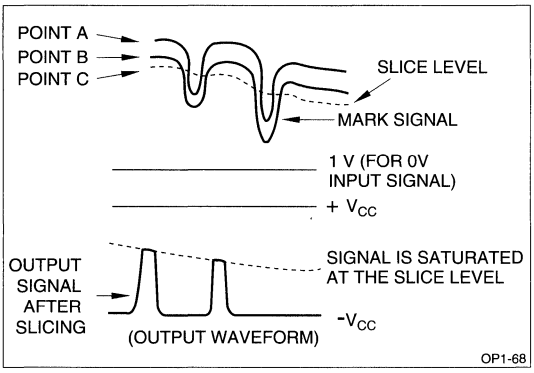


Figure 68. Card Mark Detecting Signal

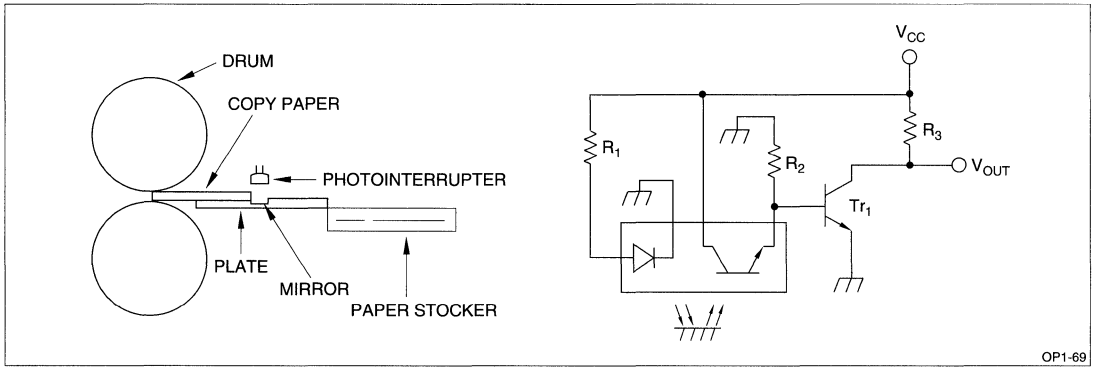


Figure 69. Copy Paper Feed Detection in Copier

Print Paper Detection

Figure 70 and 71 show the circuit example using a reflective type photointerrupter designed to detect the existence of print paper (detection of a mark on the paper).

The circuit in Figure 70 employs the floating level slicing system in which the threshold level is varied depending on the white level. The key of this circuit is the setup of load resistor R_2 so that the voltage of point A is not saturated by the collector current of the phototransistor at the white level.

The circuit in Figure 71 uses a peak hold circuit so as to detect the peak of the white level in advance. The threshold level is set in accordance with the peak value. The value of R_2 is also determined in this circuit so that the operating point of the phototransistor is not saturated.

Smoke Detector

Figure 72 shows the example of a smoke detector using a reflective type photointerrupter.

The infrared light emitted from the LED of the photointerrupter is dispersed by particles of smoke. The dispersion of light is sensed by the light detector of the photointerrupter.

Thread-cut Detection

Figure 73 shows the example of thread-cut detection, in which a cut thread falls across the gap of the photointerrupter, and circuit detects a slight variation in the output of the photointerrupter. A high resolution photointerrupter is used to detect such slight signal.

Photointerrupter Application Fields

Table 2 summarizes the application fields of photointerrupters.

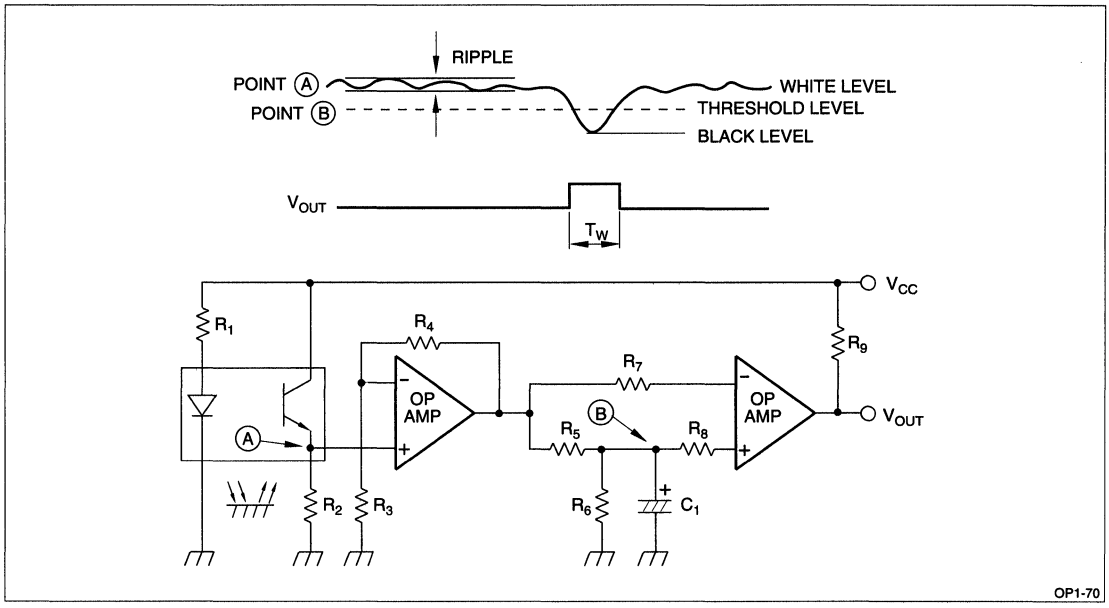


Figure 70. Print Paper Detection

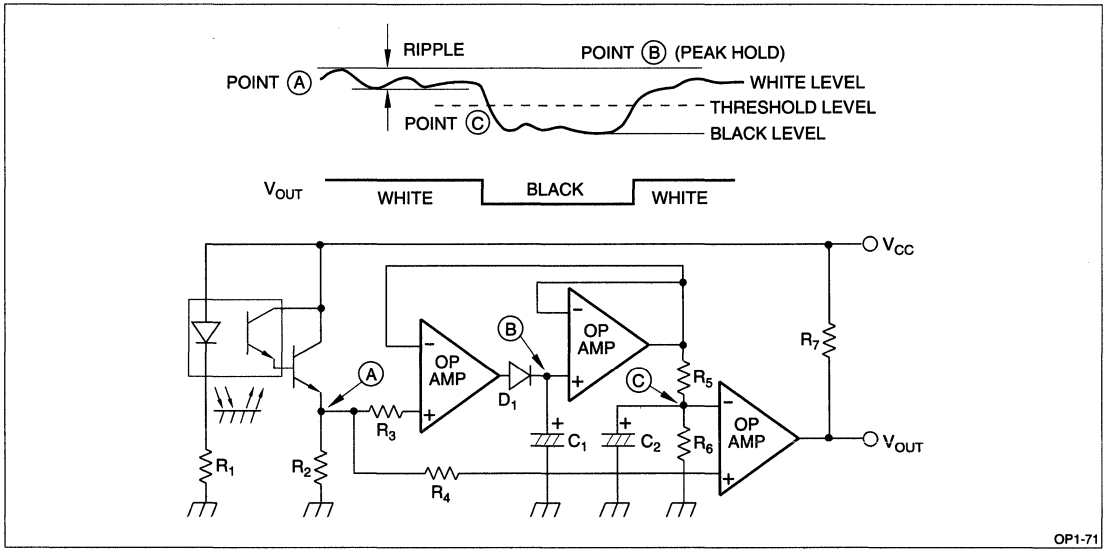


Figure 71. Print Paper Detection (II)

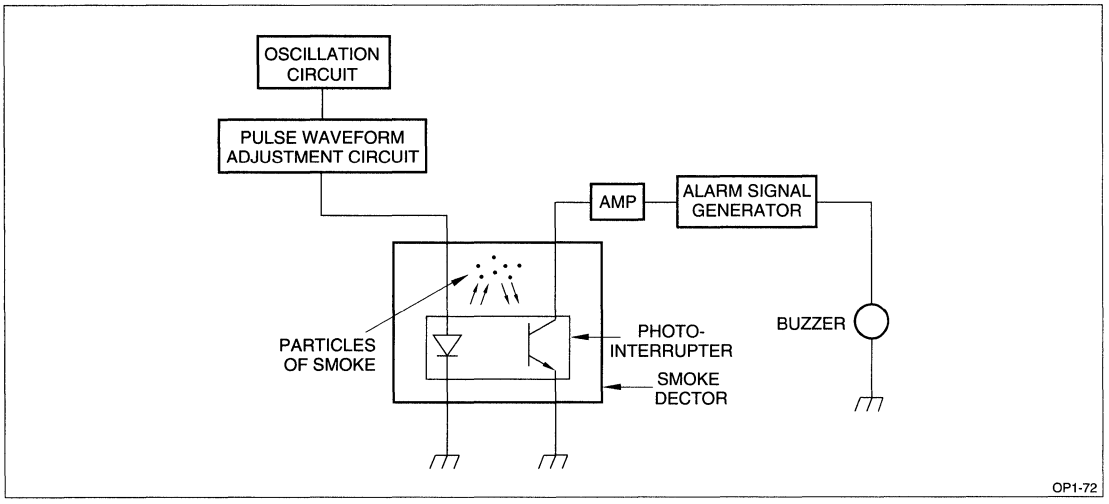


Figure 72. Smoke Detector

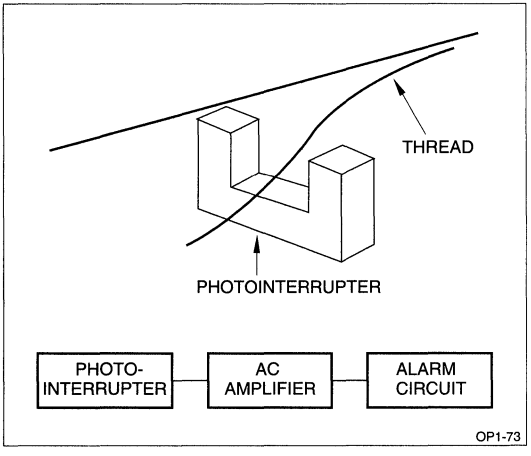


Figure 73. Thread-Cut Detection

Table 2.
Photointerrupter Application Fields

	FIELD	EQUIPMENT	APPLICATIONS
A	Computer peripheral devices	Input/output units	Paper tape reader, punched card reader
		Magnetic disk unit	Head start position detection (heading)
			Disk write-protection notch detection (erasure prevention)
B	Measuring instruments	Industrial measuring instruments	Upper of panel meter/lower limit detection
			Counter circuit
		Electronic micrometer, caliper	Detection of the number of rotations, and rotational direction
		Electronic balances	Weight counter
		Water meters, gas meters	Flow meter
C	Office equipment	Copiers	Paper feed timing detection
			Paper feed direction detection
			Drum timing detection
			Toner quantity detection
		Facsimiles	Paper feed timing detection
			Paper passage detection
		Printers	Timing detection of print drum
			End detection of inked ribbon
			Timing detection of paper-out and paper feed
		ECRs	Bar code reading
D	Automobiles	Speedometers	Speed alarm, power steering, automatic doorlocks, auto-driving controls
E	Automatic vending machines Ticket vending machines		Coin passage detection
			Ticket paper detection
F	Home appliances	VCRs	End detection of magnetic tape
			Slack detection of magnetic tape
			Rotation control of tape reel
		Electronic sewing machines	Needle up/down position detection
			Thread-cut detection
			Lower thread quantity detection
			Pattern recognition
		Room heaters	Fuel quantity detection
G	Audio equipment	Record players	Rotational speed monitoring of turntable
			Disk size detection
			Tone arm angle detection
			Unrecorded portion detection
		Cassette tape recorders	Tape end detection/Auto-return position detection
			Counter circuit
H	Medical equipment	Pulse counter	Pulse detection
I	Communications equipment	Transceivers	Digital tuning circuits
J	Others	Fuel pumps	Automatic shut off devices
		Telephone responders	Tape end detection
		Smoke detectors	Smoke particle detection

SOLID STATE RELAY APPLICATION CIRCUITS

Solid state relays (SSR) have extensive applications, from industrial equipment to home appliances, including SSRs for triggering for activating high power thyristors and triacs and SSRs for power control for directly switching AC loads. The following describes the fundamental circuit example of SSRs.

(1) Snubber circuit

Application of a voltage above the rating at the output side of SSR would result in a malfunction or even destruction of the device due to overcurrent. The snubber circuit is designed to absorb and suppress impulse noise.

Figure 74 shows some examples of a snubber circuit.

Circuit (A) is most commonly used for CR absorbers.

Circuit (B) is more effective for noise absorption since it can have a smaller Rg value.

Circuit (C) uses a varistor which can absorb high energy noise such as that caused by lightning.

The values of the resistors and capacitors in the snubber circuit depend on the kind of load and power capacity.

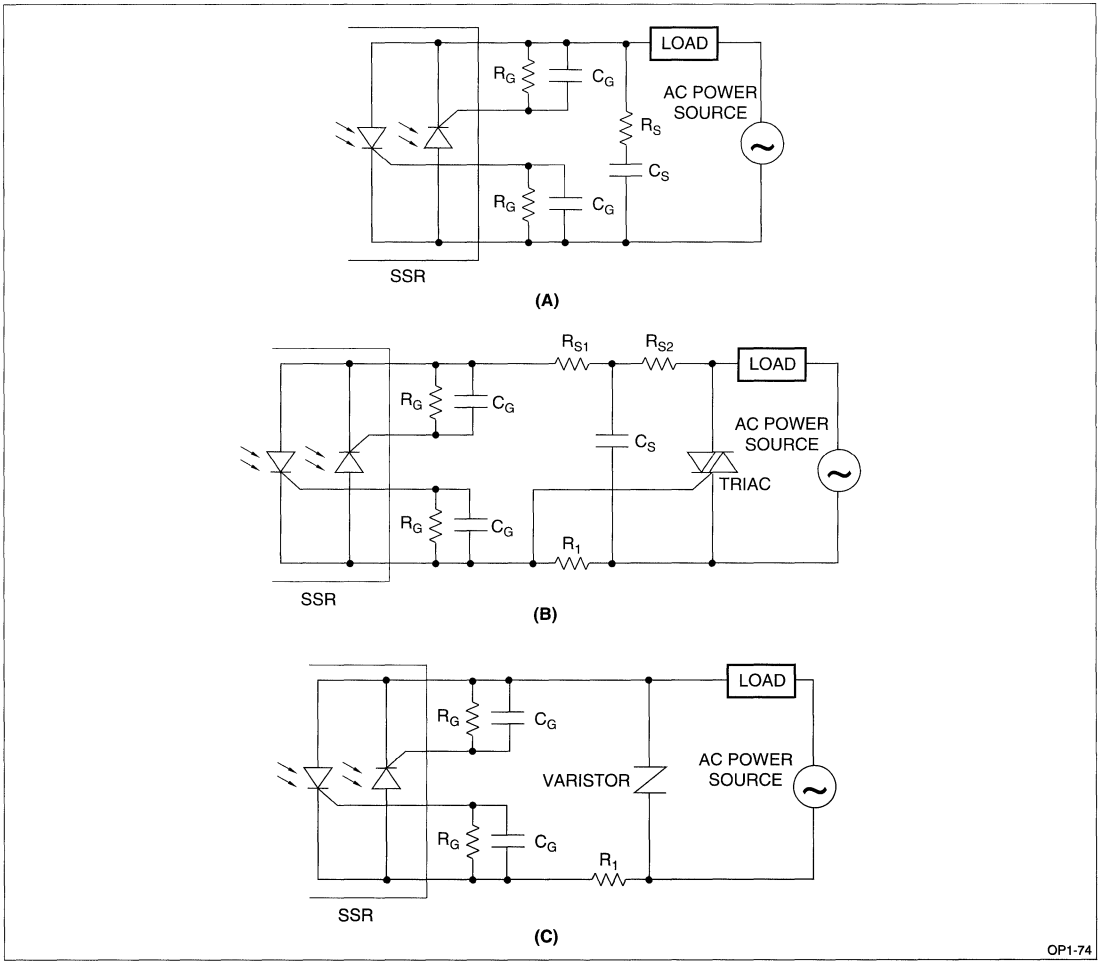


Figure 74. Snubber Circuits

Serial/Parallel Connections

Figures 75 and 76 show the circuit of two SSRs connected in series and parallel, respectively. The following precautions should be considered when operating two photothyristors type SSRs connected in series or parallel.

For Serial Connections:

Due to dispersion of leak current in forward and reverse of photothyristors, there can be a difference in the voltage across each device. For this reason, a resistor (R_1, R_2) is connected in parallel with each device so as to minimize the voltage difference.

There are dispersion in the critical rate of rise of off-state voltage (dV/dt) among photothyristors depending on the junction capacity and sensitivity of devices. For this reason, the serial connection of resistor and capacitor (R_S, C_S) is connected in parallel to each device so that both SSRs are balanced.

For Parallel Connections:

(1) Two SSRs connected in parallel must be turned on simultaneously. If one SSR turns on first, the other SSR has both its terminals short-circuited through the first one and possibly cannot turn on even if the device has low on-voltage characteristics. For this reason, adjustments must be done through the gate resistor or gate capacitor so that both SSRs have an equal turn-on time.

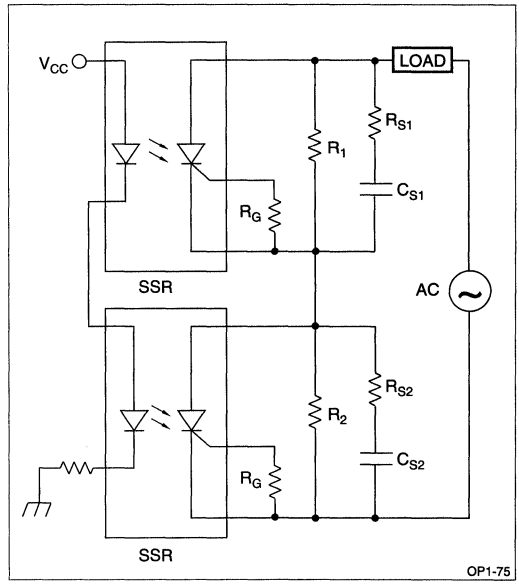


Figure 75. Serial Connection Example

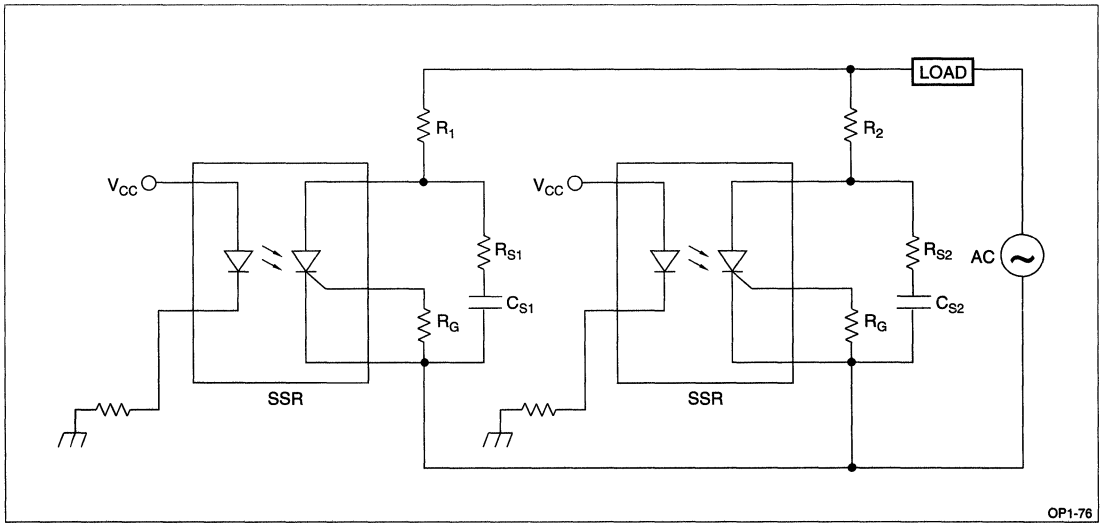


Figure 76. Parallel Connection Example

(2) There are slight differences in the turn-on voltage according to devices. This causes an unbalanced current distribution among the photothyristors, resulting in the possibility of an overcurrent. On this account, a resistor of low resistance is connected serially to the photothyristor so as to adjust the on-voltage. This equalizes the current distribution in both devices.

Zero-cross circuit

Figure 77 shows a zero-cross circuit using photocouplers, As shown in waveforms A and B, both photocouplers are off around the zero voltage level of the AC power voltage. One of the photocouplers is on in the remaining time. Accordingly, a zero-detect signal can be produced by taking the logical product of signals A and B. A zero-cross made SSR is thus arranged through activation of the SSR by the zero-detect signal Q.

This system is particularly useful in operating multichannel SSRs in a zero-cross mode in a programmable controller etc.

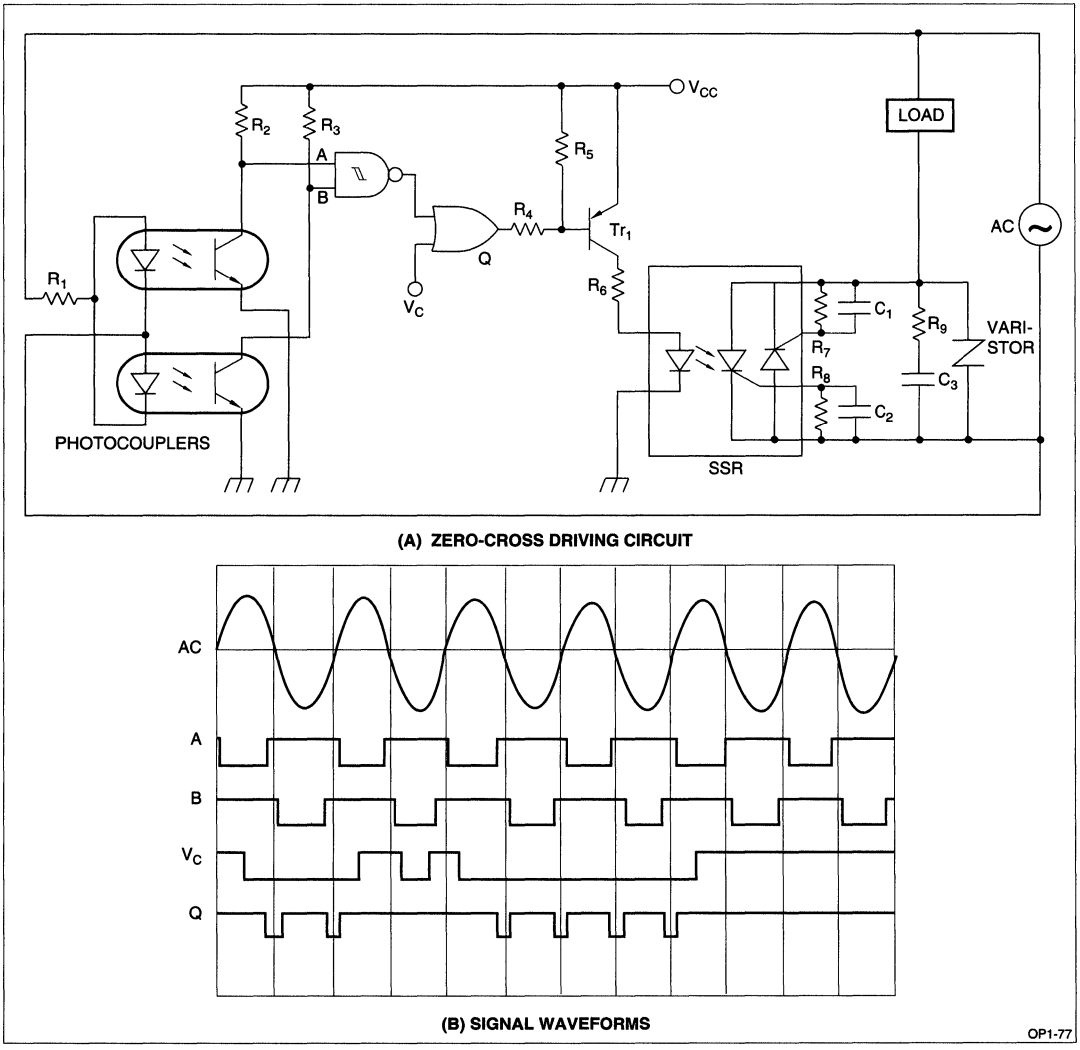


Figure 77. Zero-Cross Circuit

Application Circuits Of SSR

Programmable Controller

A programmable controller is a sequence controller which takes the place of conventional relay sequence controllers. It consists of a CPU, memory, I/O interface and peripheral devices for programming. The input/output unit of the programmable controller employs photocouplers for the input post and photocouplers or SSRs for the output in place of conventional relay contacts.

Figure 78 is a block diagram of a typical programmable controller.

Figure 79 shows the circuit of an output unit including 8 or 16 SSRs. The SIP-type SSR features a compact, built-in snubber circuit and input current limiting resistors and is frequently used in modern programmable controllers.

Copier

Figure 80 (A) shows an internal view of a copy machine in which SSRs are used. Figure 80 (B) shows a circuit example of the copy lamp control circuit using SSRs.

The copy lamp has a start-up period of several power cycles in which a rush current 10 times or more than the steady-state current flows as shown in Figure 80 (C), causing control circuits of Figure 80 (B), the SSR is shunted by resistor R_1 so as to supply a small current to the lamp for preheating. This reduces the rush current when starting.

Reversible Motor Driver

Figure 81 (A) shows the circuit of a reversible motor driver using an SSR. The circuit operates in response to the input signal shown in Figure 81 (B) to produce the motor current and voltage shown in Figure 81 (C).

The reversible motor in driven in the forward or reverse direction by one of two SSRs. If both SSRs are made conductive simultaneously, the motor will overheat. To prevent this, a time length of 1/2 cycle or more is used in switching the rotational direction as shown if Figure 81 (B). Each SSR is applied across its output terminals with a voltage twice the peak-to-peak voltage of the power line. Therefore, for a 100 VAC power source, SSRs with a withstand voltage of 300 V or more must be used.

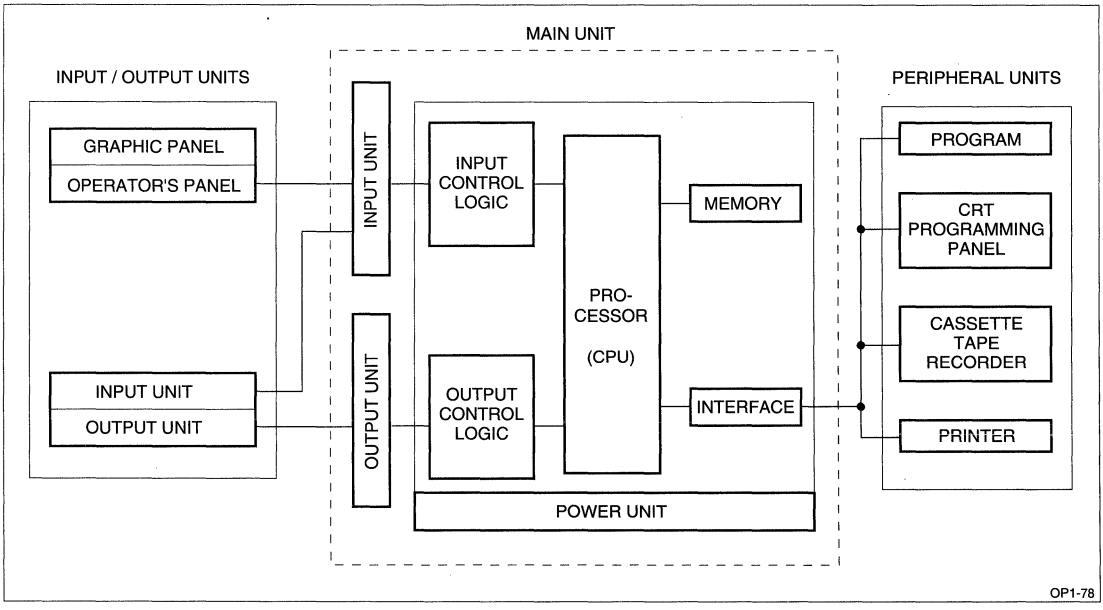


Figure 78. Block Diagram of Programmable Controller

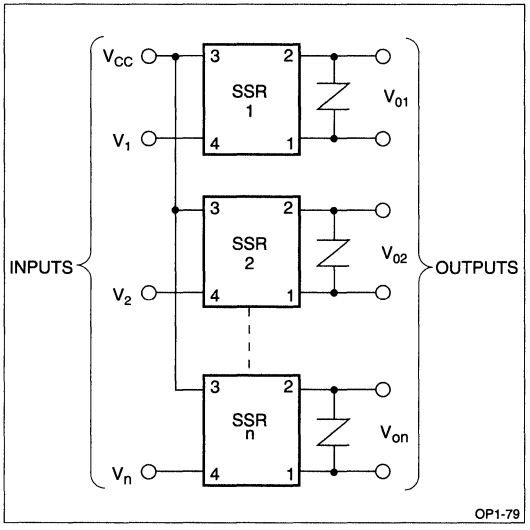


Figure 79. Programmable Controller Output Unit

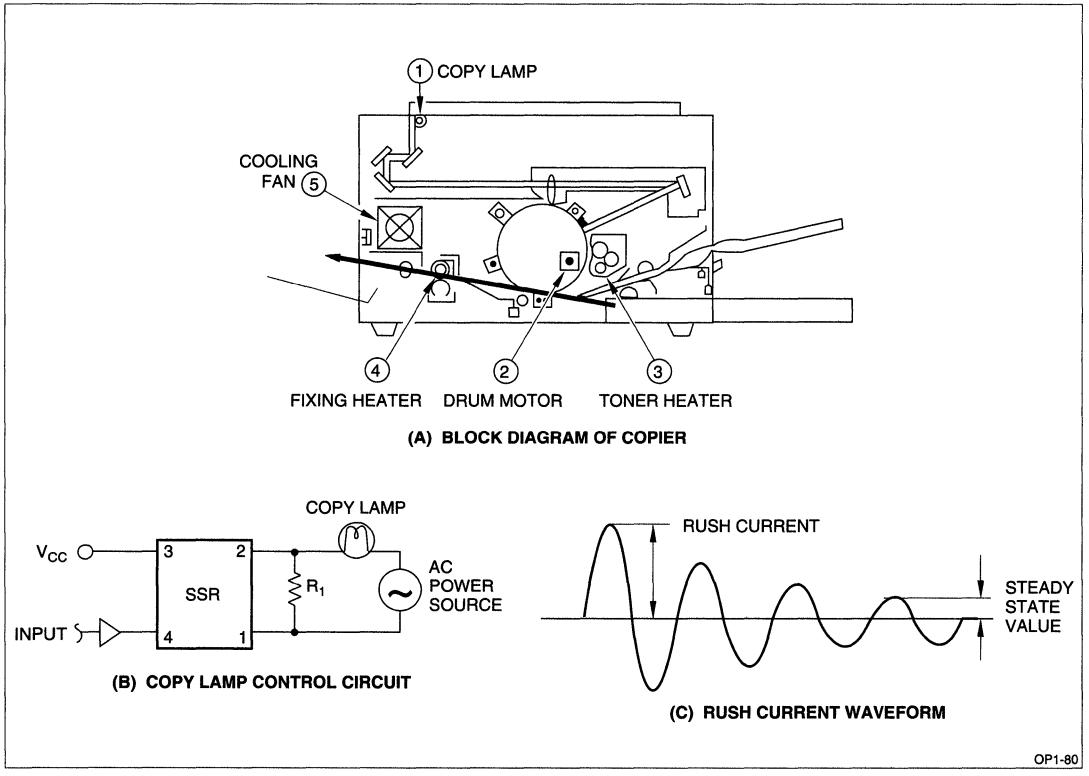
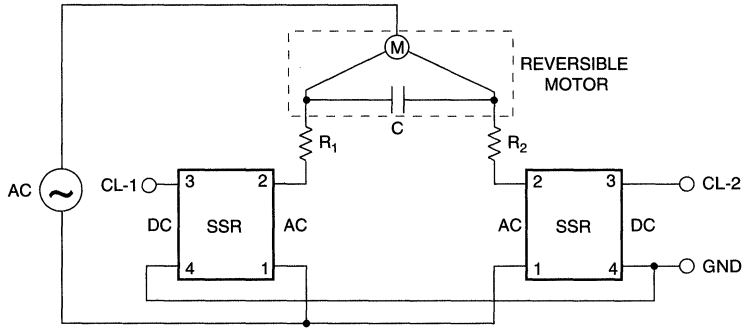
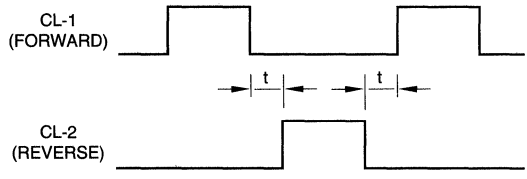


Figure 80. Copier

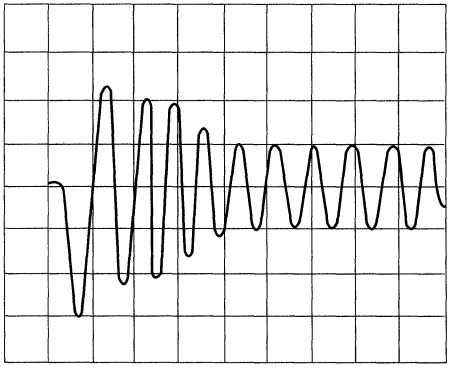


(A) REVERSIBLE MOTOR DRIVING CIRCUIT



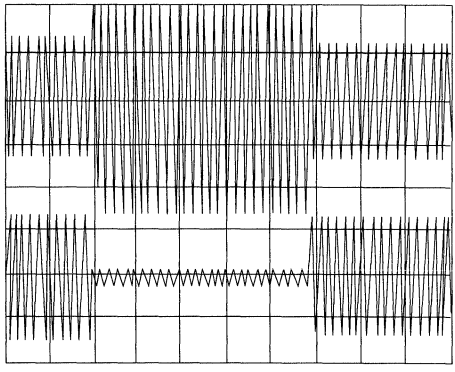
(B) TIMING CHART

HORIZONTAL AXIS: 20 ms/ DIV.
VERTICAL AXIS: 0.5A/DIV.



MOTOR CURRENT

HORIZONTAL AXIS: 0.1 s/DIV.
VERTICAL AXIS: 100 V/DIV.



VOLTAGE WAVEFORM BETWEEN OUTPUT TERMINALS

(C) OUTPUT WAVEFORMS

OP1-81

Figure 81. Reversible Motor Driver

Heater Control Circuit

Figure 82 shows a heater control circuit designed to maintain a constant temperature. A copper-constantan thermocoupler is used as a temperature sensor. An SSR is used to switch the power to the heater. The circuit shown features a wide temperature setting range and high control accuracy. Setup is made by a variable resistor (VR).

Microcomputerized Rice Cooker

In the circuit arrangement of Figure 83, three heaters (a main heater, side heater and lid heater) installed in a rice cooker are controlled by two SSRs. The three heaters are connected in series. The side and lid heaters are short-circuited by the cooking SSR during the cooking, while the warming SSR is turned on and off cyclically on completion of cooking so as to maintain a constant temperature inside the rice cooker.

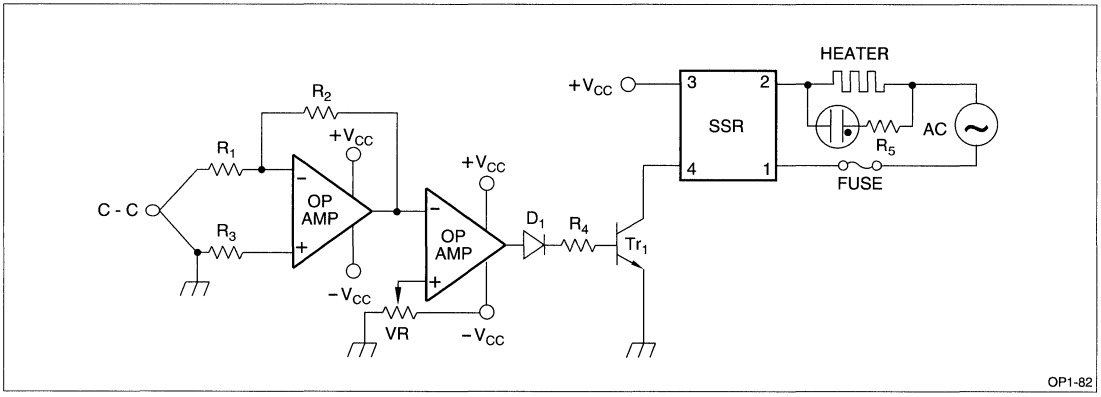


Figure 82. Heater Control Circuit

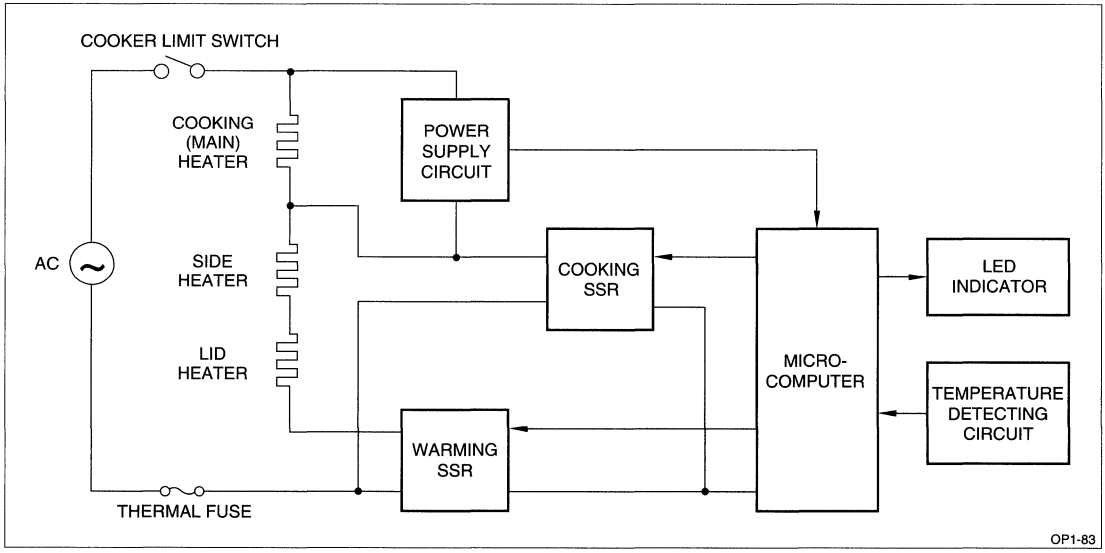


Figure 83. Microcomputerized Cooker

Table 3.
SSR Application Fields

FIELD		EQUIPMENT	APPLICATION
A	Home appliances	Air conditioners	On/off control of compressor, Speed control of blower motor
		Washing machine	Speed control pulsator
		Refrigerators	On/off control of compressor, defrosting circuit
		Electric blankets	Automatic temperature control
		Electric carpets	Automatic temperature control
		Electric jars	Automatic temperature control
		Electric powered tools	Motor speed control
		Electric sewing machines	Motor speed control
B	Office equipment	Copiers	On/off control of copy lamp, heater control
		Facsimiles	Speed control of motor (drum)
		Computers	Power switching of peripheral equipment
		Printers	On/off and speed control of motor
		Photograph processors	Exposure control
C	Automobiles	Ignition system	Switching of discharge circuit
		Generators	Output voltage control
		Others	On/off control of wiper motor and side mirror motor
D	Automatic vending machines	Coin sensors	Interface between coin sensor and indicator
		Vendors	On/off control of solenoid and indicator
E	Control equipment	Electric furnaces	On/off and temperature control of heater
		Process controllers	On/off and speed control of motor, an/off control of solenoid
		Programmable controllers	Output board (interface)
		Numerical control machines	On/off control of motor and solenoid
		Elevators	On/off control of indicator lamp, open/close control of door, on/off control of fan motor
F	Illuminators and others	Traffic signals	Flickering control of lamp
		Electric sign boards	On/off control of road information lamp
		Fluorescent lamps	Lighting circuit
		Illumination controllers	Phase control circuit

'VISIBLE' LIGHT-EMITTING DIODE APPLICATION CIRCUITS

Serial Connection and Parallel Connection

Figure 1 shows the most basic and commonly used circuits for driving light-emitting diodes.

In Figure 1(A), a constant voltage source (V_{CC}) is connected through a current limiting resistor (R) to an LED so that it is supplied with forward current (I_F). The I_F current flowing through the LED is expressed as $I_F = (V_{CC} - V_F)/R$, providing a radiant flux proportional to the I_F. The forward voltage (V_F) of the LED is dependent on the value of I_F, but it is approximated by a constant voltage when setting R.

Figures 1(B) and 1(C) show the circuits for driving LEDs in serial connection and parallel connection, respectively. In arrangement (B), the current flowing through the LED is expressed as $I_F = (V_{CC} - V_F \times N)/R$, while in arrangement (C), the current flowing through each LED is expressed as $I_F = V_{CC} - V_F)/R$ and the total supply current is $N \times I_F$, where N is the number of LEDs.

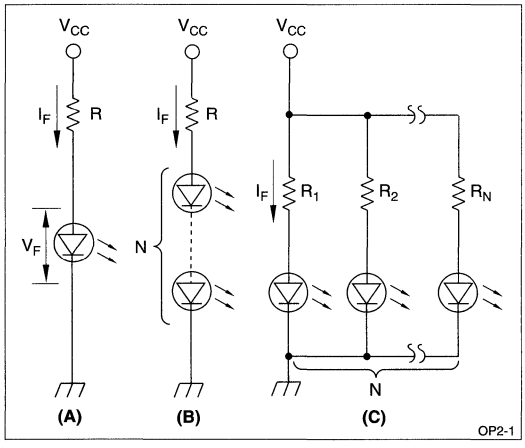


Figure 1. Driving Circuit of Light-Emitting Diode (LED)

The V_F of an LED has a temperature dependency of approximately -1.9 mV/°C. The operating point for the load R varies in response to the ambient temperature as shown in Figure 2.

Constant Current Drive

To stabilize the radiant flux of the LED, the forward current (I_F) must be stabilized by using a constant current source. Figure 3 shows a circuit for constantly driving several LEDs using a transistor. The transistor (Tr_1) is biased by a constant voltage supplied by a zener diode (ZD) so that the voltage across the emitter follower loaded by resistor R_E is constant, thereby making the collector current ($I_C = I_F$) constant. The I_C is given as $I_C = I_E = (V_Z = V_{BE})/R_E$. If too many LEDs are connected, the transistor enters the saturation region and does not operate as a constant current circuit. The number of LEDs (N) which can be connected in series is calculated by the following equations.

$$V_{CC} - N \times V_F - V_E > V_{CE}\ (sat)$$

$$V_E = V_Z - V_{BE}$$

These equations give:

$$N < (V_{CC} - V_Z + V_{BE} - V_{CE}(sat))/V_F$$

Figures 4 and 5 show other constant current driving circuits that use diodes or transistors, instead of zener diodes.

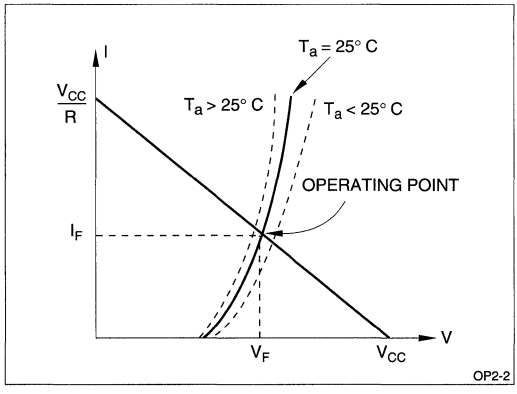


Figure 2. Current vs. Voltage of Light-Emitting Diode (LED)

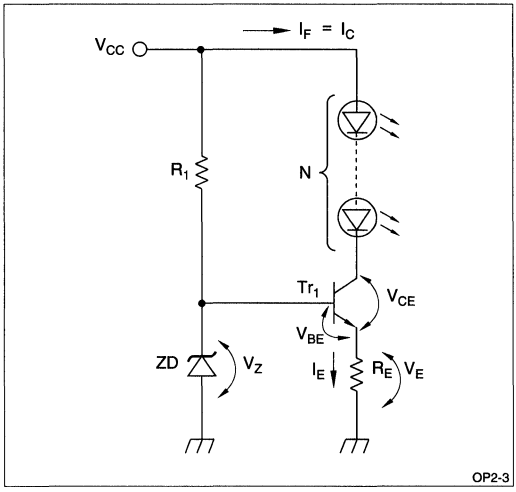


Figure 3. Constant Current Driving Circuit (1)

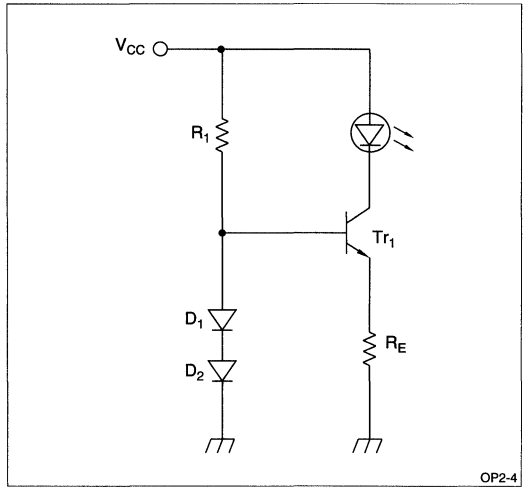


Figure 4. Constant Current Driving Circuit (2)

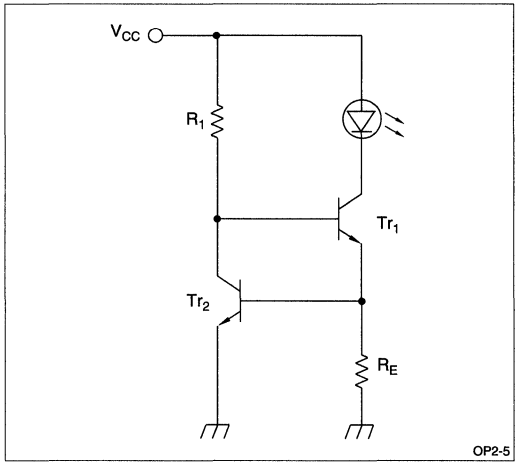


Figure 5. Constant Current Driving Circuit (3)

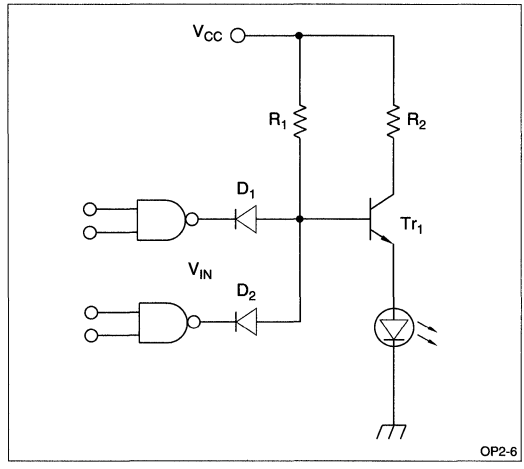


Figure 6. Connection with the TTL Logic Circuit (1)

Driving Circuit Activated by a Logic IC

Figures 6 and 7 show LED driving circuits that operate in response to digital signals provided by TTL or CMOS circuits.

Figure 8 shows a driving circuit connected with a high level logic circuit.

In Figure 6, a high input signal V_{IN} from a TTL circuit makes the NPN transistor (Tr_1) conductive so that the forward current (I_F) flows through the LED. Accordingly, this circuit operates in the positive logic mode, in which a high input activates the LED.

In Figure 7, a low input signal V_{IN} from a TTL circuit makes the PNP transistor (Tr_1) conductive so that the forward current flows through the LED. This circuit operates in the negative logic mode, in which a low input activates the LED.

In Figure 8, the circuit operates in the positive logic mode, and current I_F is stabilized by constant current driving so that the radiant flux of LED is stabilized against variations in the supply voltage (V_{CC}).

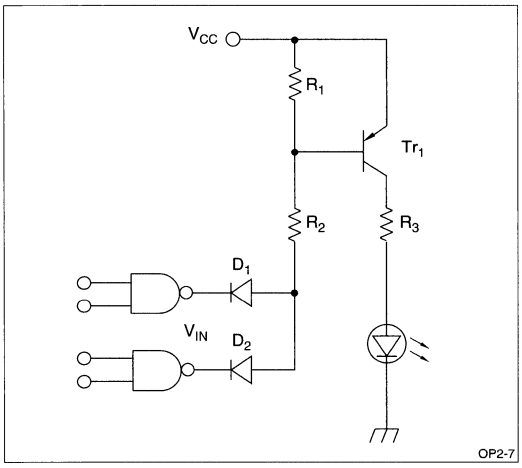


Figure 7. Connection with the TTL Logic Circuit (2)

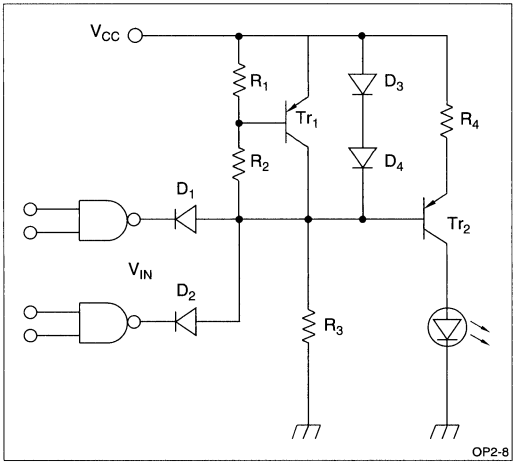


Figure 8. Connection with the TTL Logic Circuit (3)

Driving Circuit with an AC Signal

Figure 9 (A) shows a circuit in which an AC power source supplies the forward current (I_{F1}) to an LED. A diode (D_1) in inverse parallel connection with the LED protects the LED against reverse voltage, suppressing the reverse voltage applied to the LED lower than V_{F2} by using a reverse voltage protection diode of an LED. The LED provides a radiant flux proportional to the applied AC current, (emitting only in half wave).

Figure 9 (B) shows the driving waveform of the AC power source.

Figure 10 (A) shows a driving circuit which modulates the radiant flux of LED in response to a sine wave or modulation signal. Figure 10 (B) shows modulation operation.

If an LED and light detector are used together in an environment of high intensity disturbing light, it is difficult for the light detector to detect the optical signal. In this case, modulating the LED drive signal alleviates the influence of disturbing light and facilitates signal detection.

To drive an LED with a continuous modulation signal, it is necessary to operate the LED in the linear region of the light-emitting characteristics. In the arrangement of Figure 10, a fixed bias (I_{F1}) is applied to the LED using R_1 and R_2 so that the maximum amplitude of the modulation signal voltage (V_{IN}) lies within the linear portion of the LED characteristics. Moreover, to stabilize the radiant flux of the LED, it is driven by a constant current by the constant current driving circuit shown in Figure 3. The capacitor (C) used in Figure 10 (A) is a DC signal blocking capacitor.

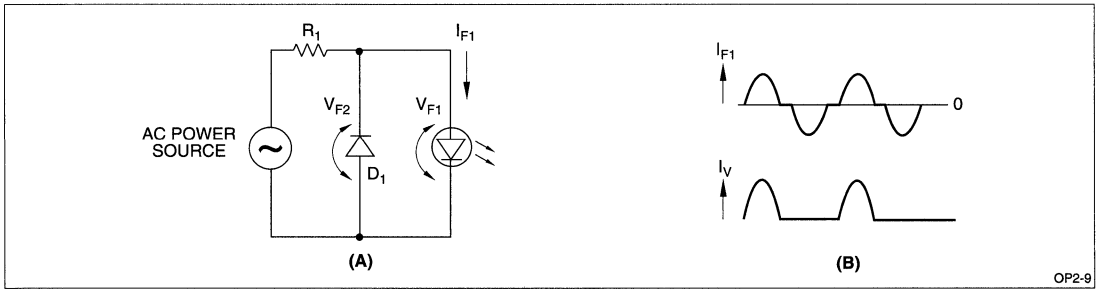


Figure 9. (A) Driving Circuit with AC Power Source (B) Driving Waveform

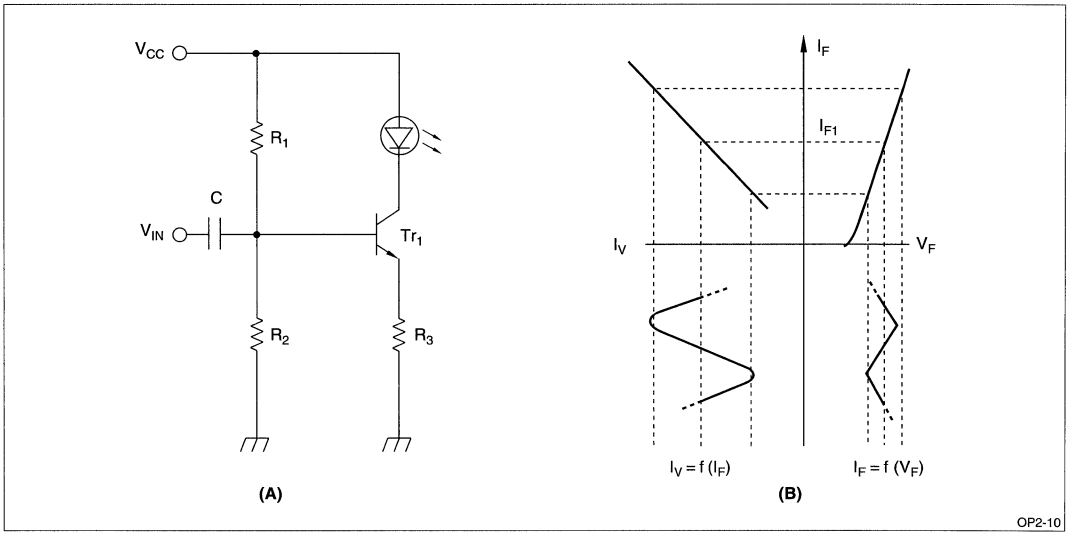


Figure 10. (A) Modulation Driving Circuit (B) Modulation Operation

Pulse Driving

LED driving systems fall into three categories: DC driving system, AC driving system (including modulation systems), and pulse driving system.

Features of the pulse driving system:

1. Large radiant flux
2. Less influence of disturbing light
3. Information transmission

The radiant flux of the LED is proportional to its forward current (I_F), but in reality a large I_F heats up the LED by itself, causing the light-emitting efficiency to fall and thus saturating the radiant flux. In this circumstance, a relatively large I_F can be used with no risk of heating through the pulse drive of the LED. Consequently, a large radiant flux can be obtained.

When an LED is used in the outdoors where disturbing light is intense, the DC driving system or AC driving system which superimposes an AC signal on a fixed bias current provides low radiant flux, making it difficult to distinguish the signal (irradiation of LED) from disturbing light. In other words, the S/N ratio is small enough to reliably detect the signal. The pulse driving system provides high radiant flux and allows the detection of signal variations at the rising and falling edges of pulses, thereby enabling the use of LED-light detector where disturbing light is intense.

Transmission of information is possible by variations in pulse width or counting of the number of pulses used to encode the LED emission.

Figures 11 through 15 show typical pulse driving circuits. The circuit shown in Figure 11 uses an N-gate thyristor with voltage between the anode and cathode oscillated at a certain interval determined by the time constant of $C \times R$ so that the LED emits light pulse. To turn off the N-gate thyristor, resistor R_3 must be used so that the anode current is smaller than the holding current (I_H), i.e., $I_H > V_{CC}/R_3$. Therefore, R_3 has a large value, resulting in a large time constant ($\tau \pm C \times R_3$) and the circuit operates for a relatively long period to provide short pulse widths. The circuit shown in Figure 12 uses a type 555 timer IC to form an astable multivibrator to produce light pulses on the LED. The off-period (t_1) and the on-period (t_2) of the LED are calculated by the following equations:

$$t_1 = 1n2 \times (R_1 + R_2) \times C_1$$

$$t_2 = 1n2 \times R_2 \times C_1$$

The value of R_1 is determined so that the rating of I_{IN} of a 555 timer IC is not exceeded, i.e. $S_1 > V_{CC}/I_{IN}$.

This pulse driving circuit uses a 555 timer IC to provide wide variable range in the oscillation period and light-on time. It is used extensively.

The circuit shown in Figure 13 uses transistors to form an astable multivibrator for pulse driving an LED. The off-period (t_1) of the LED is given by $C_1 \times R_1$, while its on-period (t_2) is given by $C_2 \times R_2$. For oscillation of this circuit, resistors must be chosen so that the R_1/R_3 and R_2/R_5 ratios are large.

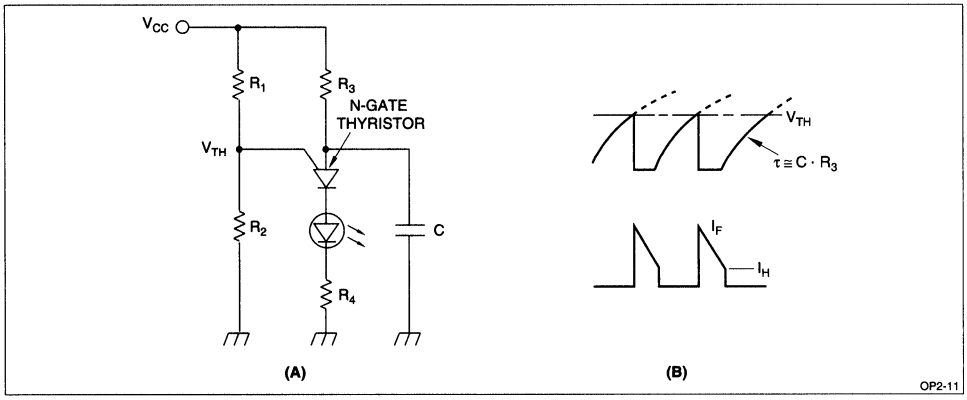


Figure 11. (A) Pulse Driving Circuit using N-Gate Thyristor (B) Operating Waveform

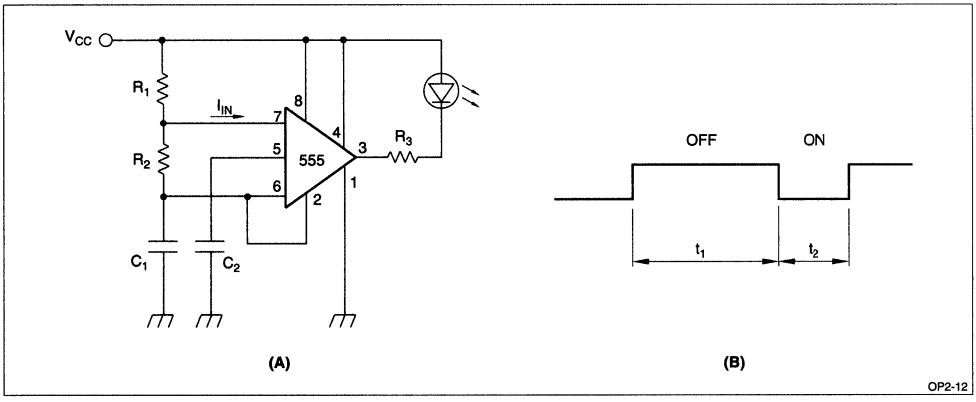


Figure 12. (A) Pulse Driving using a 555 Timer IC (B) Output Waveform

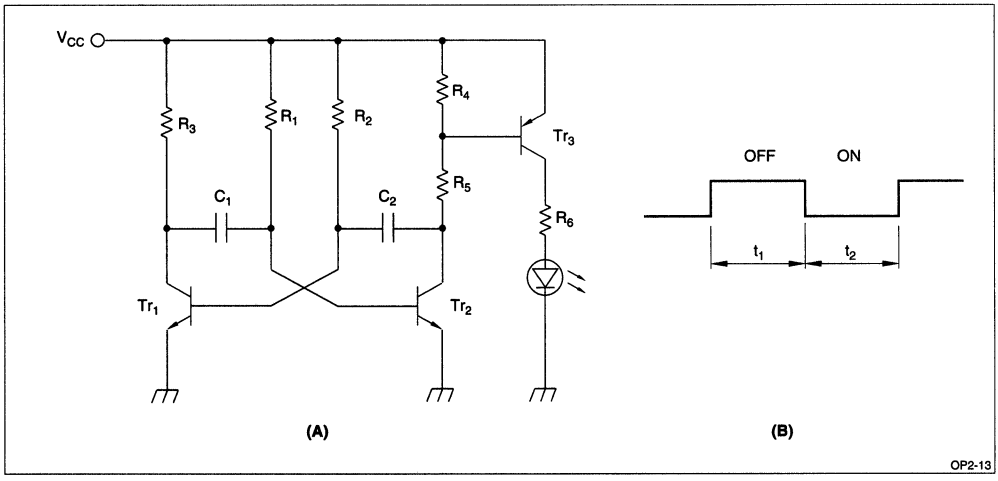


Figure 13. (A) Pulse Driving Circuit using Astable Multivibrator (B) Output Waveform

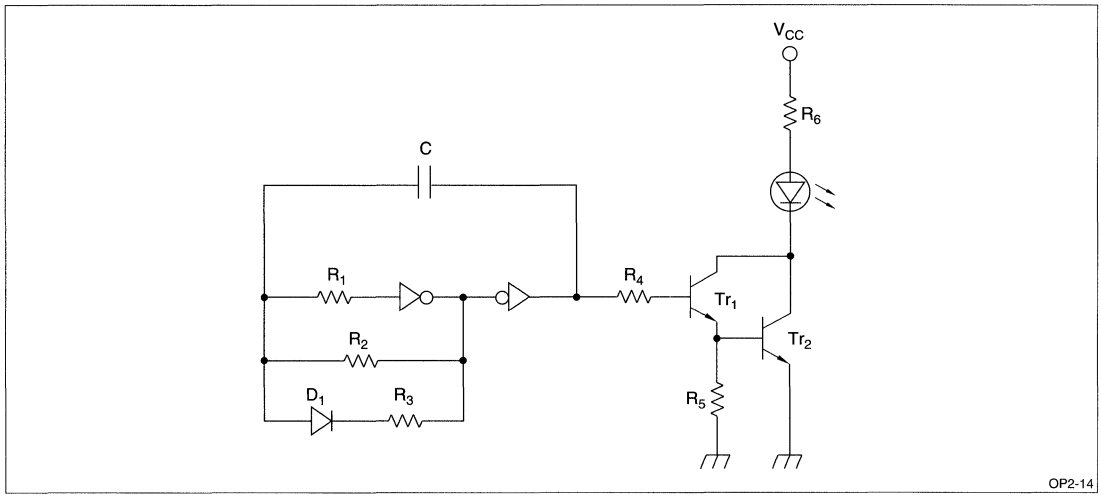


Figure 14. Pulse Driving Circuit using CMOS Logic IC

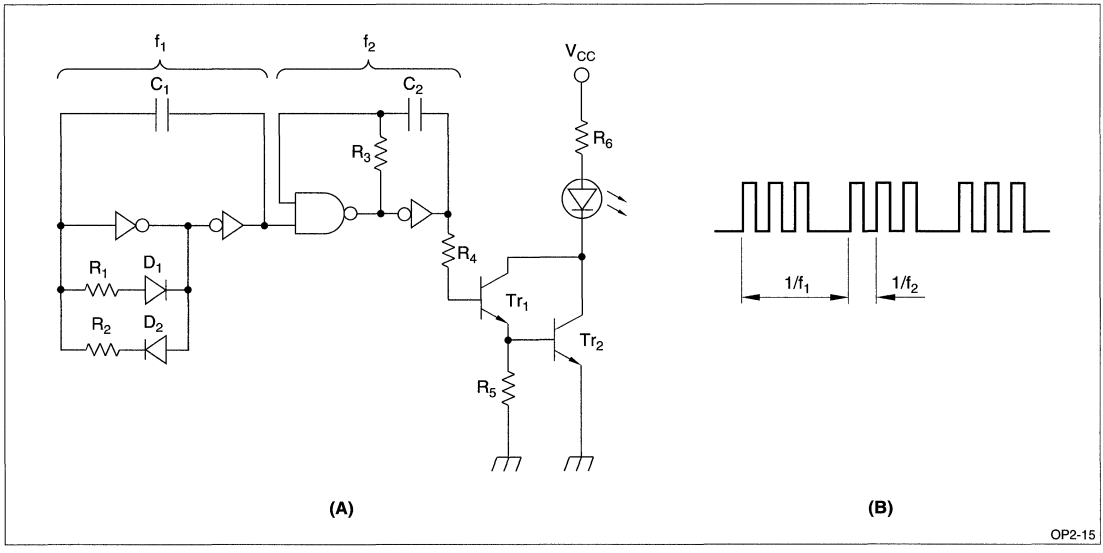


Figure 15. (A) Pulse Driving Circuit
(B) Output Waveform

The circuit shown in Figure 14 uses a CMOS logic IC (inverter) to form an oscillation circuit for pulse driving an LED. The pulse driving circuit using a logic IC provides a relatively short oscillation period with a 50% duty cycle.

Figure 15 (A) shows an LED pulse driving circuit used for the light projector of the optical remote control and optoelectronic switch. The circuit is arranged by combining two different oscillation circuits i.e., a long period oscillation (f_1) superimposed with a short period oscillation (f_2) as shown in Figure 15 (B). Frequencies f_1 and f_2 can be set independently.

Numeric and Symbolic LEDs Driving Circuit

Figure 16 shows an example of a static lighting circuit for a single-digit display. There are two types of pin connections for numeric and symbolic LEDs: the common cathode type, and the common anode type. Driving circuits differ depending on the type.

In practice, the digital display usually has multiple digits in a row, with dynamic lighting by time-sharing. Figure 17 shows an example of the dynamic lighting circuit for a four-digit display, Input signals A, B, C, and D determine numbers 0 to 9; input signals DS1 to DS4 select digit positions.

Figure 18 shows a basic circuit for dynamic lighting. The following equation gives the current (I_{FP}/seg.) that flows through each segment.

$$I_{FP}/seg. = \frac{V_{CC} - V_{CE(sat)}s - V_F - V_{CE(sat)}d}{R}$$

where,

V_{CC}: Supply voltage

$V_{CE(sat)}s$: Segment driver saturation voltage

V_F: LED forward voltage

$V_{CE(sat)}d$: Digit driver saturation voltage

R: Current limiting resistor

Although pulse peak current increases for greater luminous intensity, care should be taken to set the current limiting resistor to an appropriate value so as not to exceed the maximum rating.

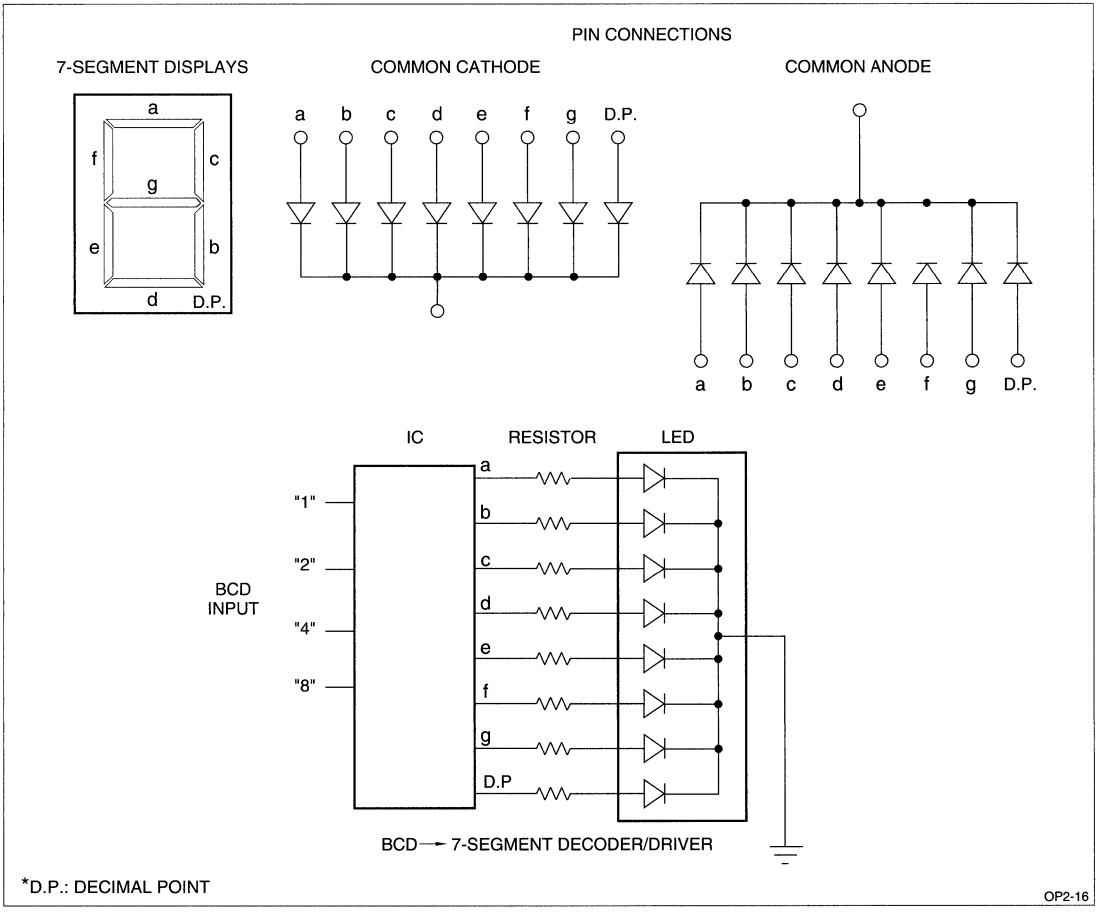


Figure 16. 7-Segment Display Driving Circuit

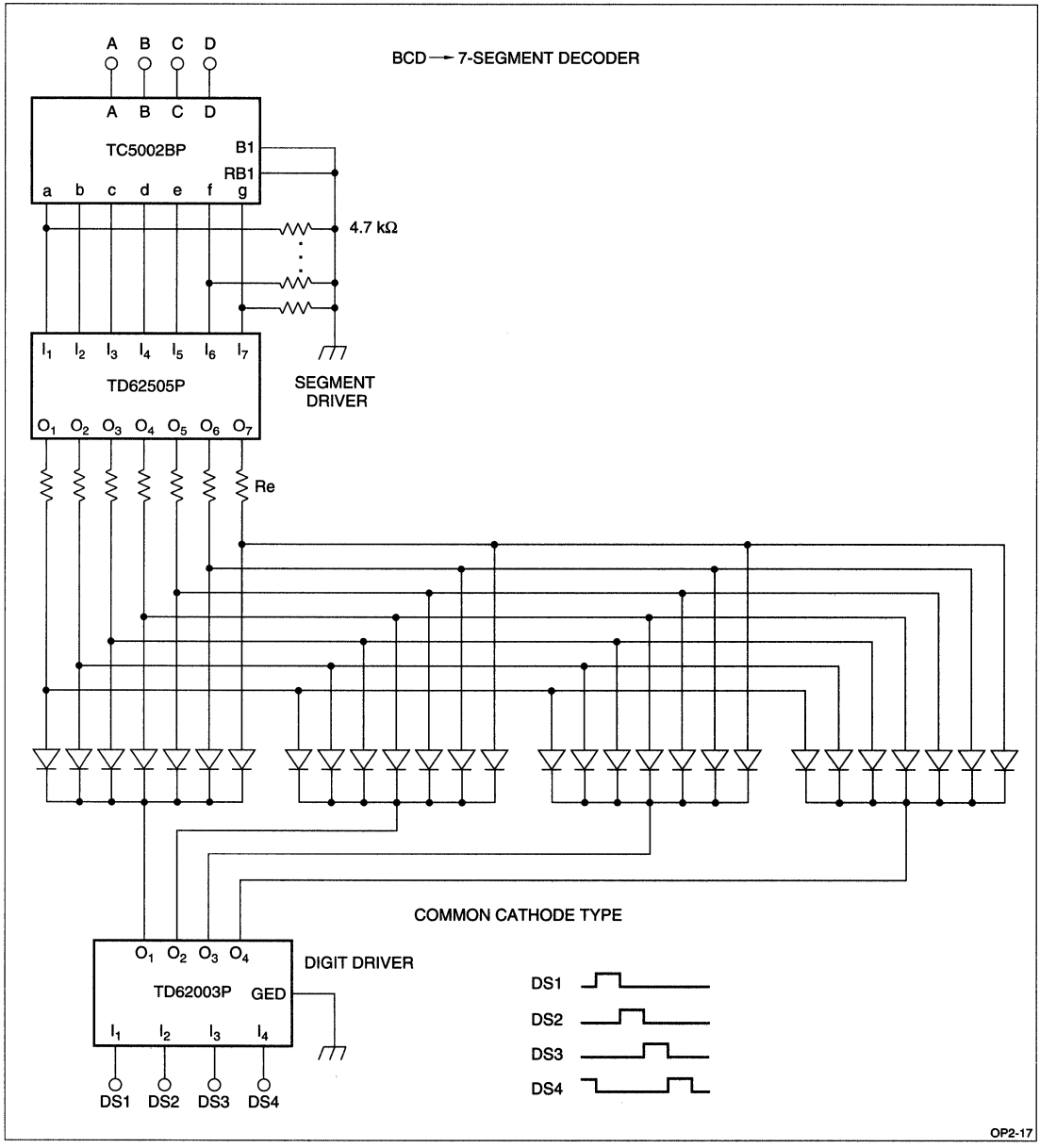


Figure 17. 4-Digit Dynamic Lighting Circuit

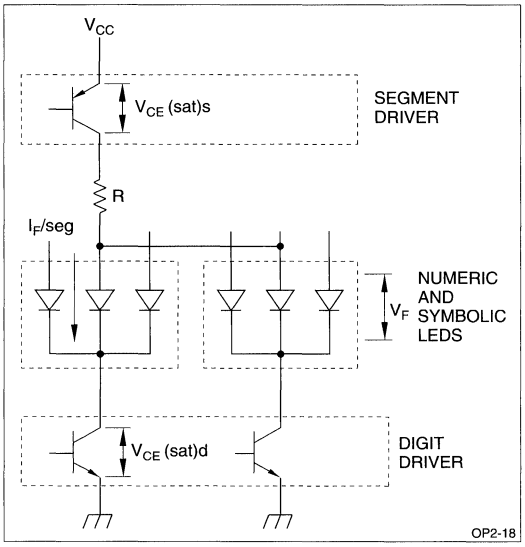


Figure 18. Basic Circuit For Dynamic Driving

Since recent advances in microcomputer technology have brought many types of microcomputer capable of serial signal output, suitable LED drivers were developed. Figure 19 shows an example of such a driver. It has built-in a 32 to 35 bit shift register, a latch circuit and a constant-current output circuit driver for LEDs. One external resistor can vary the current value for all bits. By static operation and with only one chip, this IC can drive a four-digit display. This is very useful. In addition to the digital display, the IC can be used for 5 x 7 dot matrix displays.

Application of Level Meter Circuit

Level meters that indicate voltage, current, temperature humidity or sound volume by bar graphs have advantages such as color coding displacement, quick response and long life. They are used in many fields in place of needle meters. Figure 20 shows an example of a circuit for a 12-dot single row level meter used instead of VU meters in audio systems. It also has a constant-current circuit for output. Therefore, it needs few external components. This makes it easy to use.

In addition, an IC for 5-dot, 2-row and 7-dot, 2-row level meters is also on the market.

Dot Matrix Circuit

Recently, many large information boards using LEDs can be seen on station platforms, in plazas, and on building walls.

In general, 16 x 16 dot or 24 x 24 dot matrix circuits are used for indoor information boards. In many cases, a dynamic driving system is adopted. Figure 21 shows an example. For outdoor use, because display panels must be of greater brightness, static driving systems, as shown in Figure 19, are mainly used.

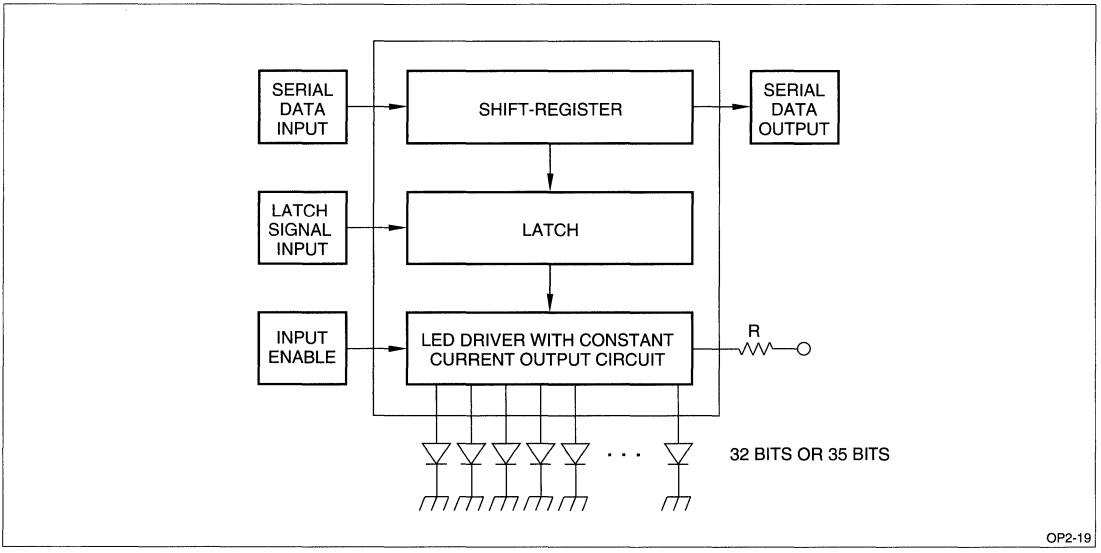
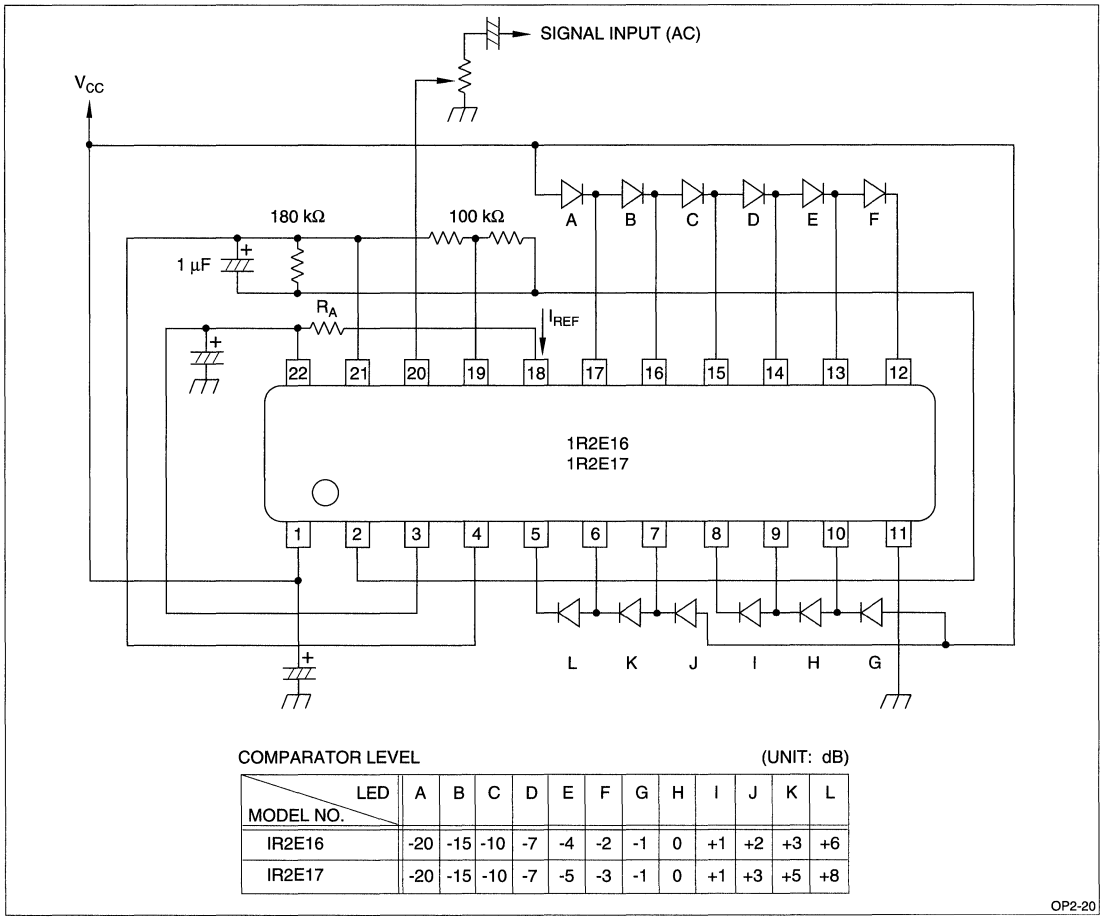


Figure 19. LED Driver with Constant Current Output Circuit



COMPARATOR LEVEL (UNIT: dB)

LED / MODEL NO.	A	B	C	D	E	F	G	H	I	J	K	L
IR2E16	-20	-15	-10	-7	-4	-2	-1	0	+1	+2	+3	+6
IR2E17	-20	-15	-10	-7	-5	-3	-1	0	+1	+3	+5	+8

Figure 20. Driving Circuit using 12-Dot LED

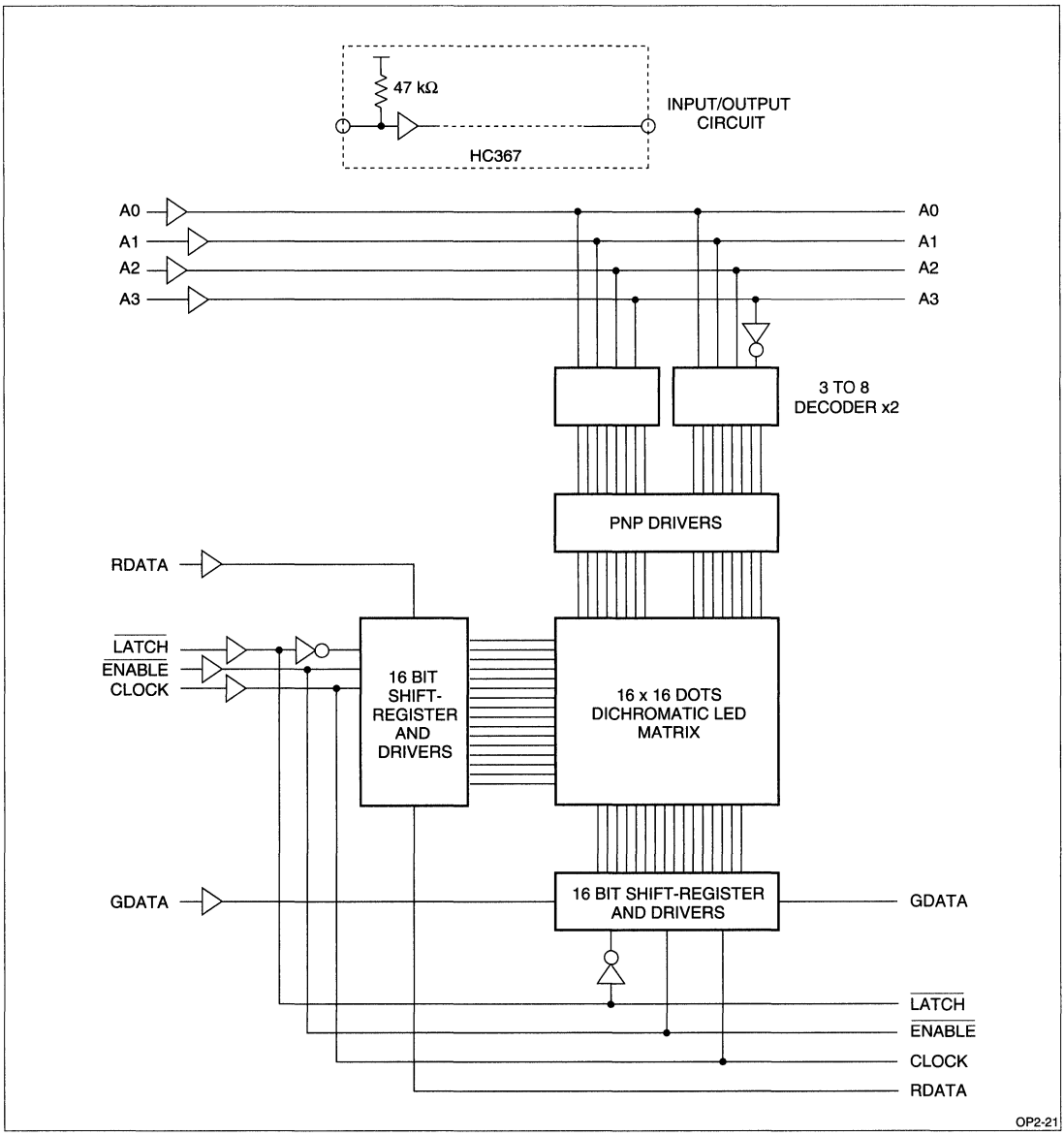


Figure 21. Indoor use of Dot Matrix LED Driving Circuit

DEVELOPERS CONTINUE TO REFINE BLUE LED TECHNOLOGIES FOR DISPLAY USE

Thanks to ongoing progress, light-emitting diodes (LEDs) today are much brighter and less expensive than the first models, which entered the market 20 years ago. To achieve these gains, developers selected the proper light-emitting materials and improved device technologies. Applications for LEDs now expand rapidly among indoor and outdoor display devices. Until recently, LEDs mainly emitted red, yellow and green. However, developers have improved the brightness of blue LEDs by using silicon carbide. This makes it possible to introduce full-color displays with the three primary colors of red, green and blue (RGB).

Various kinds of materials work in LEDs. Full-color LEDs ideally should exhibit three features: their constituent material should emit monochromatic colors as close as possible to true RGB, the LEDs should use common electrodes to simplify driving circuits, and the materials should exhibit excellent luminosity and linearity with respect to current.

MATERIALS AND STRUCTURE OF RGB LED CHIPS

During attempts to fabricate full-color LEDs, makers run into difficulties with the poor emission efficiency of blue LEDs. In correcting this problem, materials selection is the most important consideration. Existing LEDs for blue light emission use SiC, GaN or ZnSe; of these, SiC, by its nature, has an indirect transition band structure and a low emission efficiency.

Recent developments in substrate crystal growth technology and control technology for impurity concentrations, however, enable stable production of fine-quality P-N junctions. With such developments, light emission efficiency has improved steadily over the years.

GaN-based blue LEDs have direct transition band structures and light emission efficiencies one order of magnitude greater than that of SiC models. With GaN LEDs, though, formation of P-N junctions is difficult and the forward voltage is strong, from 5 to 9 V, creating problems with power supplies. Additionally, it is difficult to fabricate excellent-quality GaN crystal on a sapphire substrate, and makers have concerns about the devices' reliability. These conditions bar the mass production of GaN LEDs for now.

ZnSe LEDs are under study at various research laboratories for use in blue lasers. These LEDs offer emission efficiencies one order of magnitude lower than SiC models, however, so they are not yet in practical use.

After studying all the considerations, Sharp selected SiC for LED chips (Figure 1A) because this material is useful in stably producing crystals of excellent quality for use in blue LEDs. In the past, production of large single crystals of SiC was problematic. However, with the introduction of the Lely method, it is possible to produce ingots of superior quality with diameters larger than one inch.

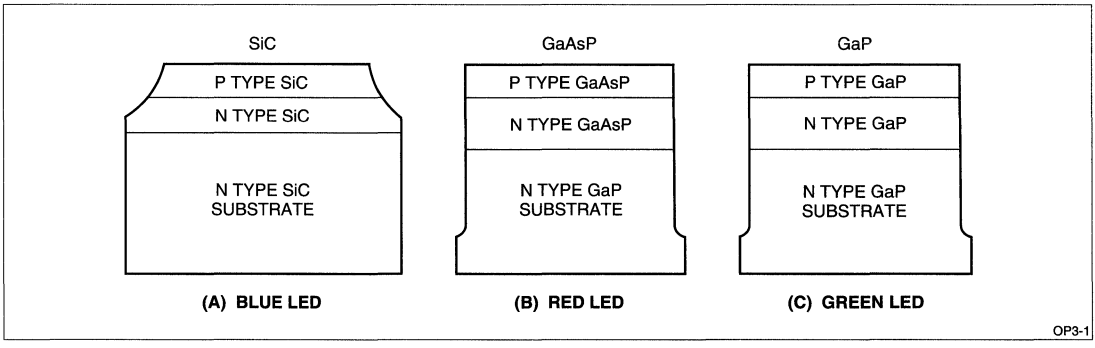


Figure 1. LED Structures

Generally, engineers fabricate SiC LEDs with a liquid phase epitaxial method using Si solution on an N-type SiC substrate. Recently, though, some makers have attempted to use the vapor phase epitaxial growth method [1], which is suitable for mass production. The emission efficiencies of LEDs using the vapor phase epitaxial method cannot compare to those of the LEDs using the liquid phase epitaxial method, but makers expect much of the vapor phase method for the future.

Nitrogen serves as the impurity in the N-type substrates, while aluminum works in P-type substrates. The addition of a small amount of aluminum to an N-type layer enhances the emission efficiency [2]. P-layers usually measure 4 to 5 μm thick. To separate the P-N junction, engineers perform mesa etching after attaching the electrodes. The die-bonded surface is the N-type SiC substrate side, which forms the cathode. The wire-bonded surface is the P-type SiC side, forming the anode. For polarity matching that of SiC LEDs, engineers selected LEDs whose die-bonded surface is the cathode, for the red and green LEDs.

Red LEDs use GaP, GaAsP, and GaAlAs. GaP LEDs have poor linearity with current, while GaAlAs LEDs use P-type crystal substrates. Preferring to use devices in which the die-bonded surface forms the anode, Sharp selected GaAsP LEDs (Figure 1B).

The only green LED in practical use is the GaP LED. Here, doping the N into the emission layer improves the emission efficiency. However, doping N shifts the emission wavelength to the longer side, changing the color to a yellowish green. To avoid this condition, Sharp selected pure green LEDs that do not use N (Figure 1C).

STRUCTURE OF FULL-COLOR LED LAMP

To create the lamp, the engineers mount the RGB LED chip on a four-leg metal stem and resin-mold it into a lamp with a 5 mm diameter (Figure 2A). The metal frame enhances the heat radiation. Silica filler in the resin facilitates color mixing. Electrically, the device adopts common-cathode wiring (Figure 2B).

ELECTRICAL AND OPTICAL CHARACTERISTICS

In the current-voltage characteristics of R, G, and B LEDs (Figure 3), red LEDs require 1.7 to 1.8 V, green ones require 2.1 to 2.2 V and blue requires 3.3 to 3.5 V. Additionally, in R, G, and B LEDs, luminosity rises linearly with the current (Figure 4). For the emission spectrum (Figure 5), the red LEDs offer a peak wavelength of 66 nm with a half-value width of 35 nm. For the green LEDs, this number is 555 nm, with a half-value width of 25 nm. Among blue LEDs, the peak wavelength reaches 470 nm and the half-value width is 70 nm. Makers hope to shorten the peak wavelengths of blue and green LEDs, ideally by trimming 10 nm or so from the existing figures.

In the CIE chromaticity diagrams for R, G, and B LEDs (Figure 6), white occurs at X = 0.33 and Y = 0.33, indicating that the devices can emit white light. In Sharp's RGB LED, the engineers created a 12 mcd white LED by applying 4 mA to the red LED, 10 mA to the green LED and 50 mA to the blue LED.

Tests of blue LEDs in continuous operation (Figure 7) show that even after 2,000 hours of operation at 50 mA. the device does not exhibit any substantial optical deterioration. Red and green LEDs, in comparison, have operated for 20,000 to 30,000 hours. These performance levels for the three colors of LEDs indicates that full-color LEDs have a reliability level that is adequate for practical use (Table 1).

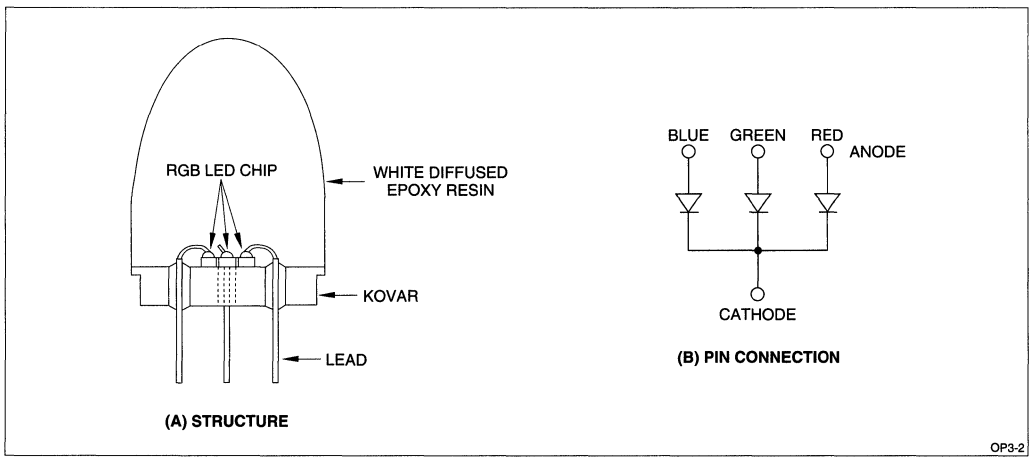


Figure 2. Full Color LED

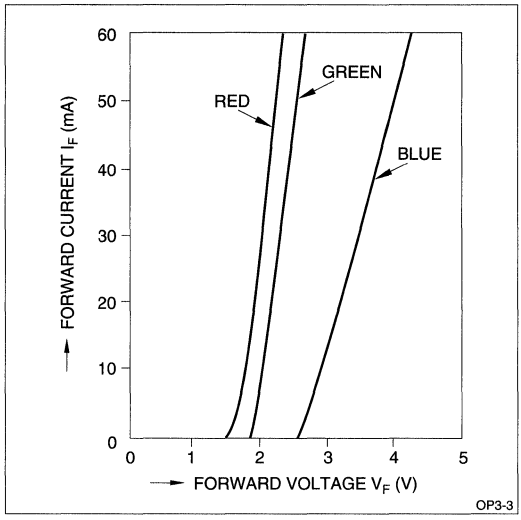


Figure 3. Forward Current vs. Forward Voltage

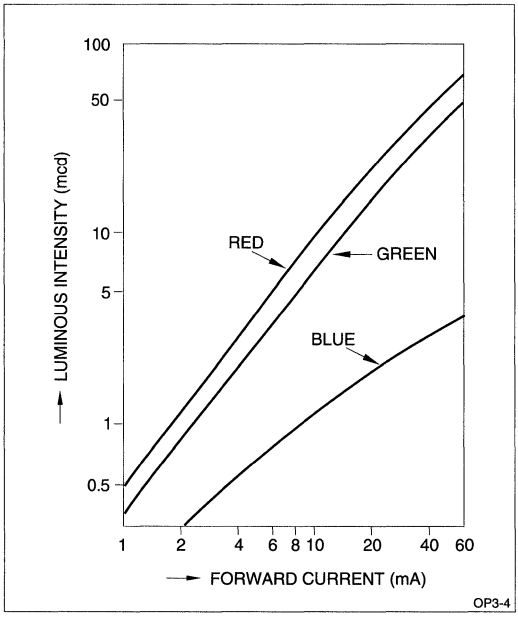


Figure 4. Luminous Intensity vs. Forward Current

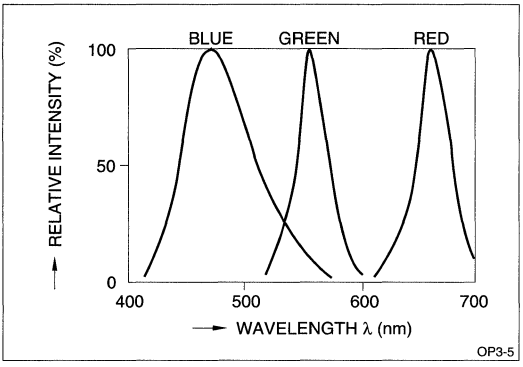


Figure 5. Wavelength

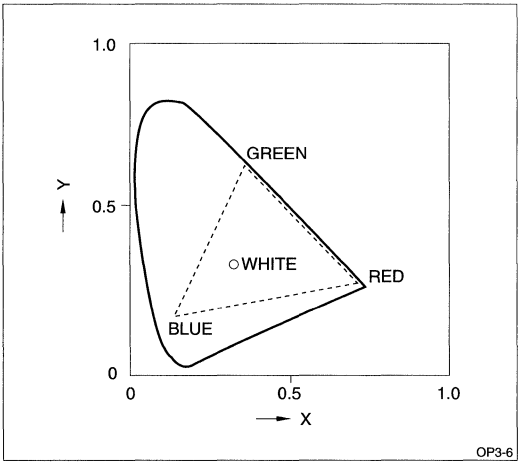


Figure 6. CIE Chromaticity Diagram

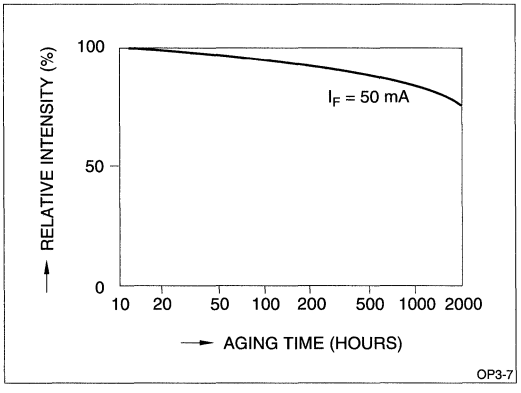


Figure 7. SiC LED Degradation Curve

Table 1. Full-color LED Characteristics

COLOR	MATERIALS	WAVELENGTH (nm)	CHROMATICITY		LUMINOSITY I_F = 20 mA (mcd)
			X	Y	
Red	GaAsP	660	0.71	0.28	23
Green	GaP	555	0.35	0.63	15
Blue	SiC	470	0.14	0.16	2

APPLICATION OF FULL-COLOR LED LAMP

Using RGB LEDs as indicators requires a transistor to drive each LED (Figure 8). Simultaneously lighting two colors enables a mixed-color display. Thus these devices will display even a seven-color rainbow.

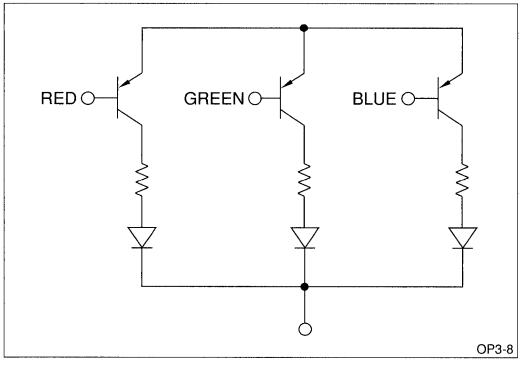


Figure 8. Full Color LED Driving Circuit

When full-color LEDs work as display panels, the most suitable arrangement is a dot-matrix unit consisting of 256 LEDs with a configuration of 16 vertical dots and 16 horizontal dots. Arranging multiple units vertically and horizontally allows construction of a large display panel. For displaying TV pictures, the arrangement must consist of at least 256 dots × 160 dots (16 horizontal × 10 vertical units).

Sharp's LED panel uses static driving (Figure 9). The system inputs data to a 256-bit shift register sequentially, and when the input process is complete, the shift register latches. The system holds the data and the output transistors turn on or off. Users input characters and patterns to computers before operation. Lighting the desired LEDs permits users to draw arbitrary figures. When each dot has gradations, the frame frequency (the number of frames that appears each second) rises and the duty of the On/Off ratio alters.

In a unit consisting of 16-dot × 16-dot, 5 mm-diameter lamps in a 6 mm pitch arrangement, the display emits a brightness of 300 cd/m^2 when the red LED receives 4 mA, the green receives 10 mA and the blue receives 50 mA. With all LEDs lit, the system's total power consumption is 85 W. An aluminum hood promotes heat radiation, but because the system generates quite a lot of heat, it requires forced-air cooling if multiple units are in use. Examples of configurations include connections of two vertical units by four horizontal units.

FULL-COLOR ASSEMBLY LAMPS

Even though developers have improved the luminosity of blue LEDs, these devices have not yet attained sufficient luminosity for daytime use. To overcome this problem, developers are conducting trials, using conventional red, green and yellow LEDs in the daytime and blue LEDs at night. Such display panels are effective as advertising bulletins in busy commercial districts. Because large outdoor displays so far have offered insufficient reliability, demand will grow substantially for these displays in the future.

STRUCTURE OF FULL-COLOR ASSEMBLY LAMPS

Large displays light up many separate RGB LEDs, but from a distance, the human eye perceives the overall effect of these lights, rather than distinguishing the individual lamps. This is because the human eye has a resolution of only 1/60°. Therefore, makers devise waterproof lamps scaled with silicon (Figure 10) for outdoor displays. Normally, large displays use only strong-luminosity red and green LEDs measuring 5 mm in diameter, but it is possible to replace these devices with LEDs measuring 3 mm in diameter and to mix blue LEDs with red and green LEDs.

To obtain a sufficient level of luminance for daytime recognition, LED displays must offer a luminosity of more than 2,000 cd/m^2. The luminosity changes with 3 mm-diameter lamps, as does the number of lamps necessary for use in assembly lamps (Table 2). A 26 mm-diameter assembly of four red LEDs, eight green LEDs and nine blue LEDs produces a luminosity of 2.4 cd.

Arranging this assembly lamp at a 30 mm pitch enables a brightness of 2,700 cd/m^2. A combination of 80 mcd luminosity for red, 160 mcd for green and 54 mcd for blue produces white light at night, at a 320 cd/m^2 brightness, sufficient for nighttime use.

Table 2. Full-color Solid-State Lamp

COLOR	LUMINOSITY	NUMBER OF LAMPS	DAYTIME MODE TOTAL LUMINOSITY (mcd)	NIGHTTIME MODE TOTAL LUMINOSITY (mcd)
Red	350 (20 mA)	4	1,400	80
Green	130 (20 mA)	8	1,040	160
Blue	6 (30 mA)	9	–	54

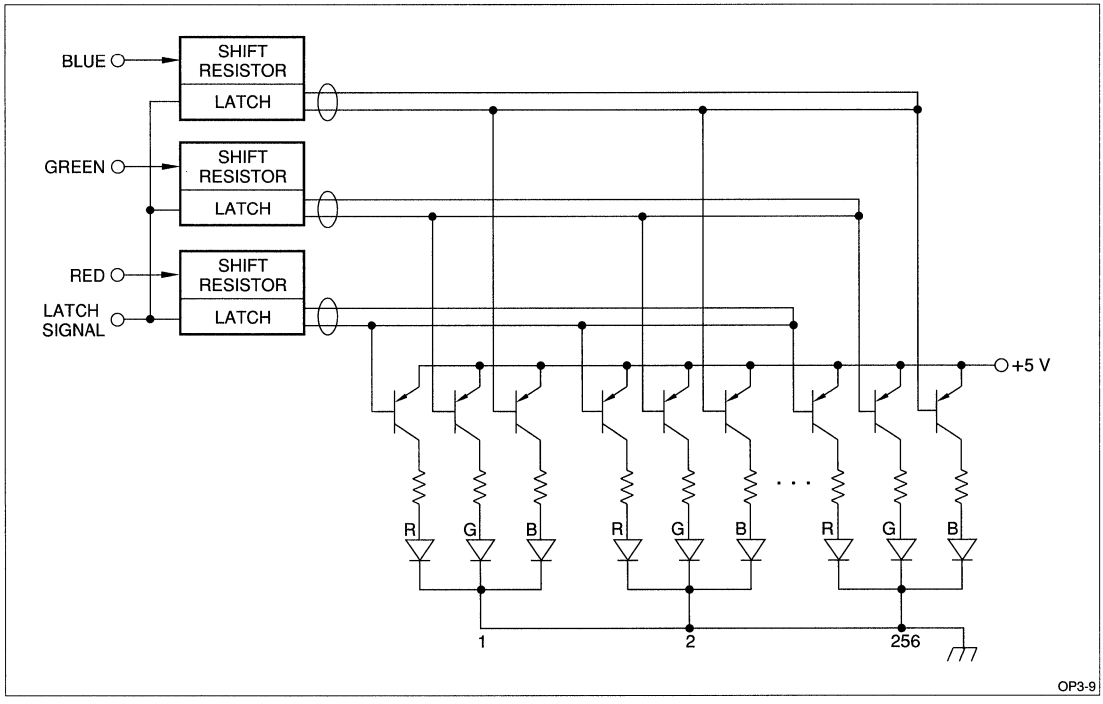


Figure 9. Full Color LED Matrix Driving Circuit

APPLICATIONS FOR FULL-COLOR ASSEMBLY LAMPS

It is possible to drive a full-color assembly lamp with a static circuit (Figure 9). In large outdoor displays, the panels fully light the red and green LEDs. At night, the system lowers the luminosity of the red and green LEDs by controlling the duty ratio of the pulse widths, and it turns on the blue LEDs. To display TV pictures, it is necessary to attain individual gradation control over RGB colors. Normally, display panels conduct gradation control by governing the duty ratio of each dot. For instance, if the system constructs images with one second duration in 60 frames, obtaining 16 gradations requires 16 times the original number of frames, or a total of 960 frames. To produce this quantity, the system turns each dot on and off, enabling the representation of 4,096 colors. For neat TV image reproduction, each dot must be able to display 64 gradations. Achieving this goal and to diminish the variability of luminosity in LEDs requires developers to cultivate rapid resistors faster than 20 MHz. In the near future, manufacturers will introduce large, full-color LED displays.

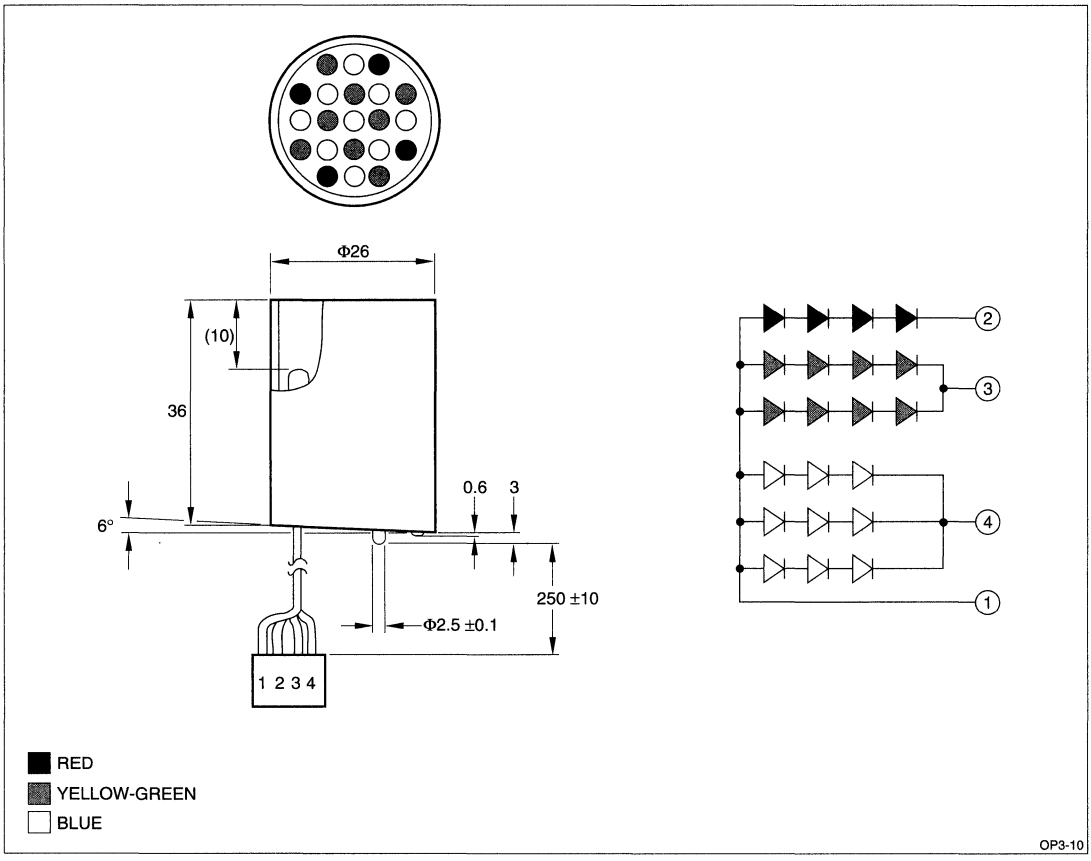


Figure 10. Full Color Solid State Lamps

CONCLUSION

LEDs feature long service lives, the ability to emit the three primary colors, red, green and blue, for full-color displays, and sharp emission spectra with excellent visibility. Except for blue lamps, LED displays work anywhere, indoors or outdoors, because they feature excellent brightness. They respond rapidly, making them suitable for displaying dynamic pictures, and they use low operating voltages, so they will incorporate into other devices and will offer excellent reliability.

In the future, a new market will develop that will make use of the characteristics of LEDs. This market also will introduce blue LEDs to join the red and green ones; together these devices will replace neon signs. For now, though the recently developed blue LEDs do not provide sufficient brightness, at least for daytime use. Large display systems use more than one million LEDs, so they must attain reliability levels of less than one frame interline transfer (FIT). Thus developers will have to find ways to improve LED quality. Step-by-step solution of these problems will pave the way for LEDs to play an important role as displays devices in information and other fields.

REFERENCES

[1] Suzuki et al., *Journal of Crystal Growth 115*, pp. 623-627, 1991.

[2] Matsushita et al. *Japanese Journal of Applied Physics*, Vol 29, No. 2, pp. L343-L345, February, 1990.

LOW-CURRENT DRIVE, HIGH OUTPUT POWER ARE THE KEY FACTORS TO IMPROVE 780 nm SEMICONDUCTOR LASERS

Among the short-wavelength laser devices now in use, 780 nm semiconductor laser diodes are the most reliable. They hold an established position as important devices in the field of optical information processing equipment. To improve equipment performance, however, manufacturers will have to reinforce the characteristics of the lasers. Additionally, it is important to expand the applications for laser diodes. Toward this end, manufacturers are seeking ways to incorporate advanced and multiple functions, which are not possible with conventional lasers.

As CDs and CD-ROMs have gained popularity, the electronics industry has promoted the development of compact, lightweight, portable systems with low power consumption. The introduction of miniature equipment causes a rise in the light noise that returns to the laser. Thus there is a growing need for low-noise lasers. To lower power consumption in CD players, it is necessary to drive lasers at low currents. Responding to these demands, developers are making progress in cultivating lasers compatible with low-current drives and that offer low-noise characteristics. Thus the industry follows specific trends in developing 780 nm laser diodes (Figure 1).

In applications of the lasers for erasable optical disks, makers seek to improve the data transfer rates, because the rates for erasable optical disks are slower than those for magnetic disks. One way to achieve this goal is to accelerate the rotation speed, but this involves raising the optical output power of the laser diode. Conventionally, lasers with 30 to 40 mW optical output powers have served users with these needs. However, it is possible to double the rotation frequency of optical disks by pushing the output power to 50 mW or more. Consequently, in its efforts to develop strong power lasers, Sharp focuses on increasing the optical output power.

In efforts to explore new applications for semiconductor laser diodes, researchers have studied applications in holographic equipment, which incorporates diffraction gratings. For this kind of use, makers prefer lasers that remain unaffected by changes in external temperature and that do not exhibit skipping characteristics under direct modulation. One solution calls for the development of a distributed feedback (DFB) laser in which an incorporated diffraction grating inside the laser cavity suppresses wavelength fluctuation. Other proposed solutions involve developing lasers with innovative structures for application in parallel optical information processing and optical interconnection systems.

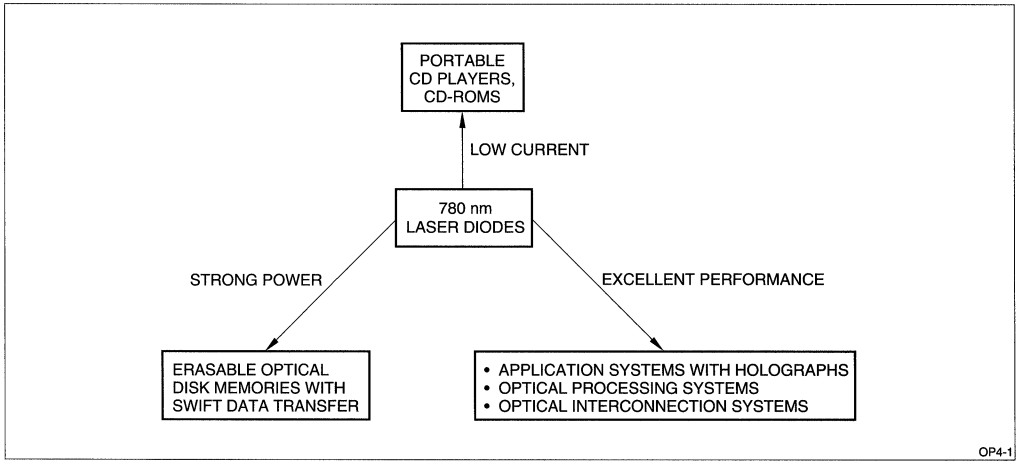


Figure 1. Development Trends Among 780 nm Lasers

EXAMPLES OF RECENT TECHNOLOGY

Low-Current Lasers

Lowering coherence by inducing self-oscillation is an effective means for inhibiting the noise that arises from the return of light to semiconductor diodes. However, self-oscillation lasers with a low current drive pose a problem, because the threshold current rises with the gains in the ineffective current components. Sharp cut the threshold current by adopting a graded-index separate confinement hetero structure (GRIN-SCH structure) in layers near the active layer. This structure provides large optical confinement in the active layer than that in conventional double-hetero structure (DH structure). While the GRIN-SCH structure poses a problem of expanding radiation beams from the augmented light confinement, Sharp solved this by improving the layer structure.

Specific features of this structure become apparent upon examination of a cross-section of an improved GRIN-SCH structure (Figure 2) and of the layer structure of the center of the stripe (Figure 3). To grow the layers near the active layer, the engineers used molecular beam epitaxy (MBE), which offers excellent properties for controlling the layer thickness. Then they fabricated a layer segment with a low Al composition ratio at both ends of the GRIN-SCH structure. This suppresses the expansion of the radiation beams. After forming the stripe ditches, the engineers re-grew the layer using liquid phase epitaxy (LPE), thereby refilling the ditch. They incorporated a 250 μm cavity length and applied a facet coating of a half a wavelength.

The improved GRIN-SCH structure produced a threshold current of 39.6 mA and sufficient self-oscillation to suppress the return light noise. Conventional DH structures offer a 45 mA threshold current, so the GRIN-SCH structure provides a 5 mA drop in this figure. Also, because of the incorporation of low Al composition ratio layers on both sides of the GRIN-SCH structure, the engineers diminished the spread of the radiation beam to the same level as that in the DH structure. In addition, the optimum structure of the improved GRIN-SCH laser makes it possible to decrease the threshold current to less than 30 mA.

High-Power Lasers

In high-output-power operation of semiconductor lasers, there is a problem with deteriorating beam emissions, because of the gains in output intensity within the device. For output power greater than 50 mW, it is essential to control the deterioration of the laser facet. Some developers proposed a window structure laser with a non-absorbing region near the facet. This region would not absorb the laser light at the oscillation wavelength. However, this involved a complicated structure and fabrication process.

Instead, Sharp developed a window-structure laser by growing on the facet surface a thin aluminum gallium-arsenide (AlGaAs) crystal layer that does not absorb at the oscillation wavelength. To fabricate this window-grown-on-facets (WGF) laser, the engineers employed a regrowth technology to grow a crystal on the cleaved facet of the laser, using metal organic chemical vapor deposition (MOCVD) [1].

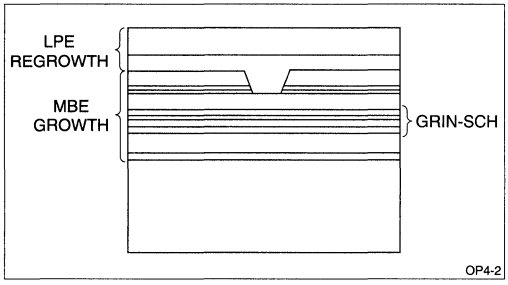


Figure 2. Cross-Section of Improved GRIN-SCH Structure

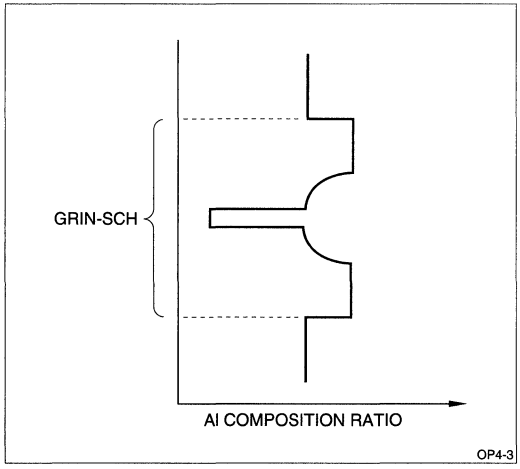


Figure 3. Improved GRIN-SCH Structure

With the structure of the WGF laser (Figure 4), the internal laser is a V-channeled substrate inner stripe (VSIS), one that features excellent transverse-mode stability and reliability under high output power conditions. A thin AlGaAs window layer is re-grown on the laser facet using MOCVD processes.

During optical output power of 150 mW or more, the facet did not exhibit any observable deterioration. The observation took place with a device that had a 375 μm cavity length and a 15 to 95 percent asymmetrical coating on the facet. Furthermore, the researchers could not observe any fluctuation of the radiation beam with output power of up to 150 mW. During acceleration tests of the WGF laser at a 780 nm oscillation wavelength, a 60°C ambient temperature and a 70 mW output power, the device operated safely for more than 2,000 hours without deterioration (Figure 5) [2]. When the researchers applied the WGF structure to an 830 nm laser with a 480 μm cavity length and a 3 to 95 percent asymmetrical facet coating, the device performed stably for more than 2,500 hours at optical output power of 100 mW and 150 mW under specific conditions (Figure 6). In all the tests, the WGF laser demonstrated excellent reliability during high power operations greater than 50 mW.

CONCLUSIONS

In refining 780 nm semiconductor lasers, developers aim to improve the performance of lasers. The efforts to improve these lasers focus on achieving low-current drive and high power operations. Developers are making progress along these lines.

Advanced-performance features not available in conventional lasers are necessary for new laser applications. Laser development has progressed to a certain degree as makers have started using lasers for applications in holographic equipments. Developers will continue their rigorous pursuit of lasers for application in parallel optical information processing equipment and optical interconnection.

REFERENCES

[1] K. Sasaki, M. Matsumoto, M. Kondo, T. Ishizumi, T. Takeoka, S. Yamamoto and T. Hijikata, *Japan Journal of Applied Physics 30*, L904 (1991).

2] K. Sasaki, M. Matsumoto, M. Kondo, T. Ishizumi, T. Takeoka, H. Nakatsu, M. Watanabe, O. Yamamoto, S. Yamamoto and T. Hijikata, SPIE, Vol. 1634, "*Laser Diode Technology and Applications IV*," p. 204 (1992).

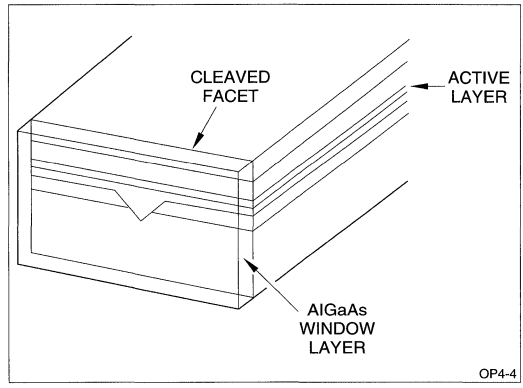


Figure 4. WGF Laser

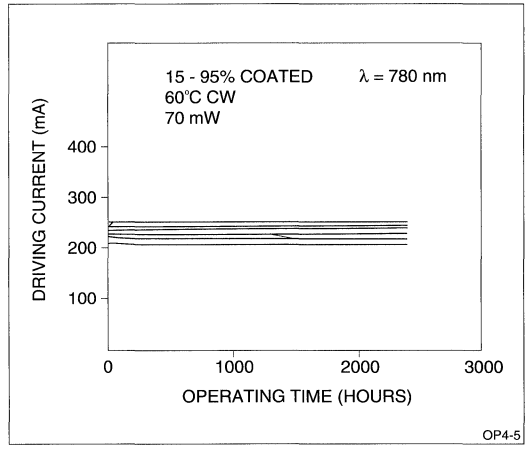


Figure 5. Results of 780 nm WGF Laser Aging Tests

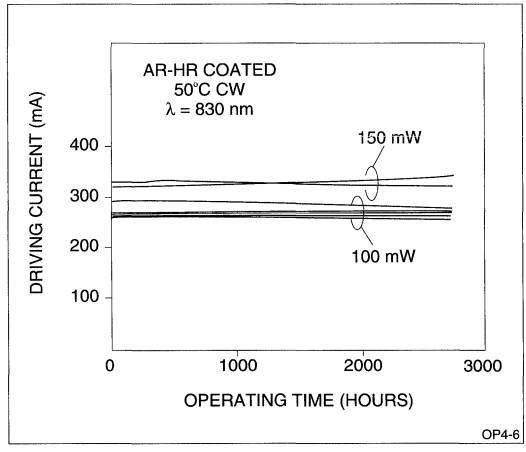


Figure 6. Results of 830 nm WGF Laser Aging Tests

CONSIDERATIONS WHEN DESIGNING WITH PHOTOCOUPLERS

Chuck Bohac, Field Applications Engineer

PHOTOCOUPLER PRIMER

Photocouplers (Optocouplers, or Optoisolators) are compound semiconductors that are used to transmit signals, data, or DC levels across a circuit interface while providing high isolation between input and output. Typically, photocouplers are used: to eliminate ground loops (in circuits of differing voltages or ground planes), to reduce noise within a circuit (such as the feedback to an FET in a switching power supply), to communicate between circuits with different operating voltages (such as a 5 V programmable controller and higher voltage I/O module), as well as a variety of other uses. Fundamentally, a photocoupler consists of a light emitter and a light detector as shown in Figure 1.

An equivalent circuit of a photocoupler is shown in Figure 2. The light emitter consists of a GaAs Infrared (IR) Light Emitting Diode (LED). The light detector consists of a photo transistor (a photodiode connected to the collector/base of a transistor of a common silicon substrate). The structure of the phototransistor is shown in Figure 3.

There are differing ways of manufacturing photocouplers. Some manufacturers mount the emitter adjacent to the detector while other manufacturers mount the emitter above the detector. SHARP mounts the IR LED die on the upper arm of a stamped, lead frame using a conductive epoxy paste such that the emitter is located directly above the detector (the advantage of placing the emitter directly above the detector

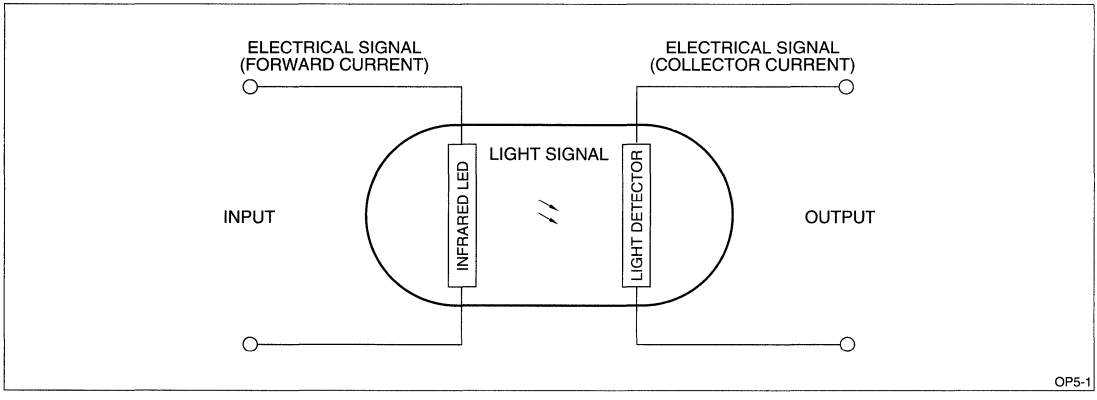


Figure 1. Basic Components of a Photocoupler

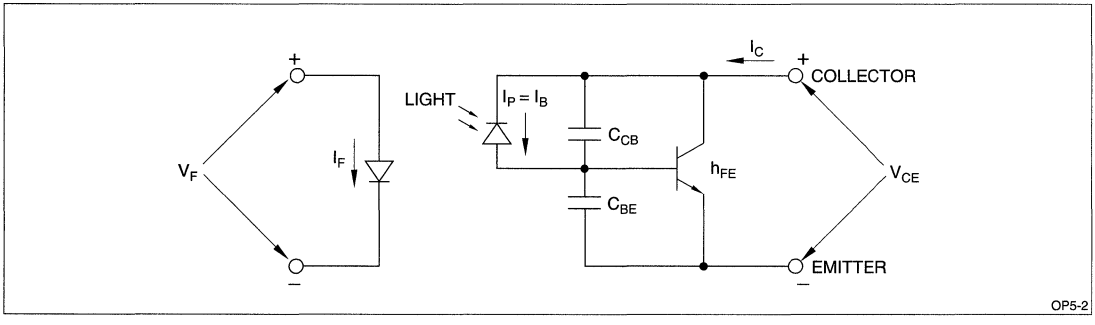


Figure 2. Equivalent Circuit of a Photocoupler

allows for the most efficient coupling of IR energy from the emitter to the detector). A gold bond wire is then attached to both the die and the lead frame. An identical process is used to mount the Photo transistor die to the lower lead frame. A minute deposit of transparent silicon epoxy is deposited around the LED die, and ball bond. This transparent silicon epoxy has a higher coefficient of expansion than the surrounding transparent epoxy which allows the LED die, ball bond and bond wire to expand and contract during temperature cycling without imposing physical stress on the ball bond (weakest point). A cross-section of the SHARP PC700 and PC800 series of photocouplers is shown in Figure 4.

Once the die are mounted, bond wires attached and silicon epoxy applied to the LED, SHARPs patented double transfer mold process is used. The emitter and detector are encapsulated into a single IR transparent epoxy mold. Then a second transfer mold encapsulates the transparent mold with an opaque epoxy. Both epoxy resins are of the same basic chemical compound. The advantage of this process lies in the fact that since both epoxy resins are of the same chemical makeup (with the exception of a proprietary darkening agent added to the opaque resin), they adhere to each other as a homogenous compound. The net result is a chemically homogenous mold with a uniform expansion coefficient. The homogenous compound eliminates the potential for the creation of a void that would

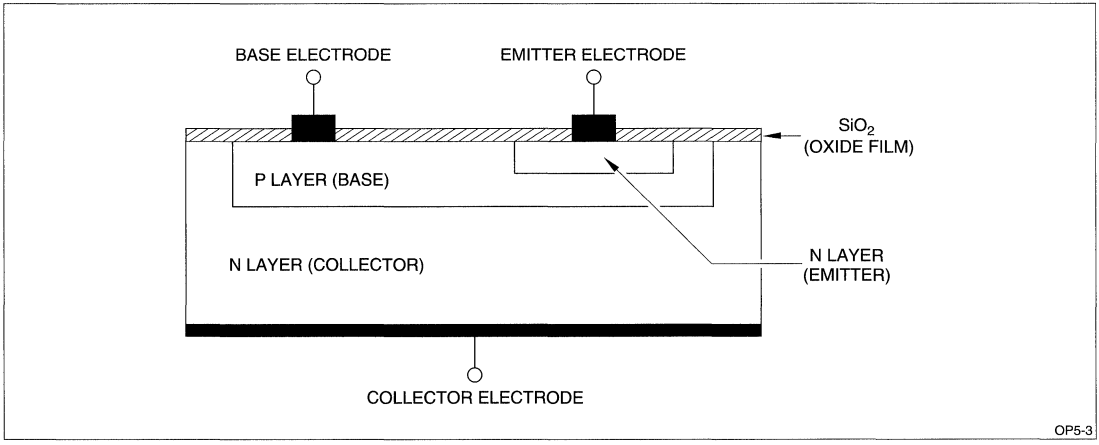
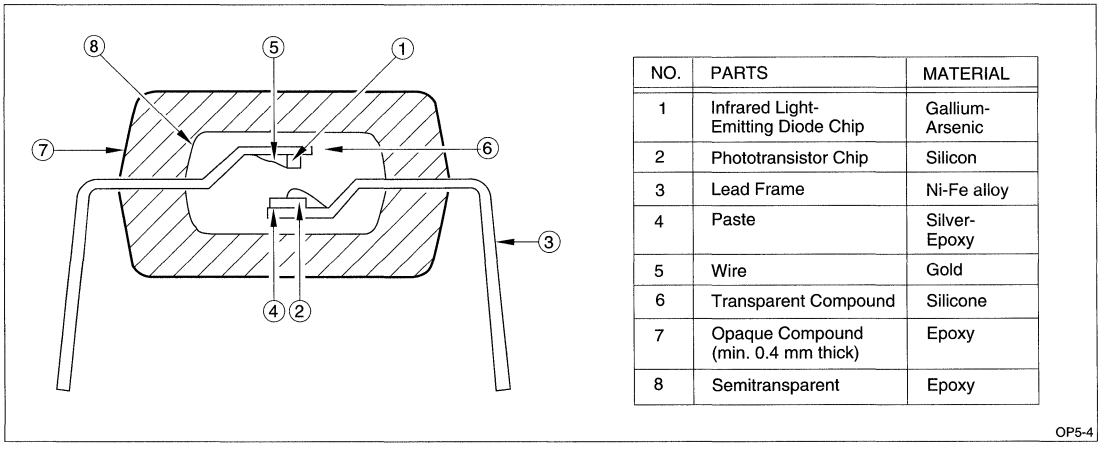


Figure 3. Phototransistor Structure



NO.	PARTS	MATERIAL
1	Infrared Light-Emitting Diode Chip	Gallium-Arsenic
2	Phototransistor Chip	Silicon
3	Lead Frame	Ni-Fe alloy
4	Paste	Silver-Epoxy
5	Wire	Gold
6	Transparent Compound	Silicone
7	Opaque Compound (min. 0.4 mm thick)	Epoxy
8	Semitransparent	Epoxy

Figure 4. Cross-Section of the PC700 and PC800 Series Photocouplers

exist between two chemically different compounds. As a result, the minimum discharge distance is transferred from the inside of the package (labeled 1) to the outside of the package (labeled 2) as shown in Figure 5. With the type of transparent epoxy resins used today coupled with an internal spacing of between 0.3 and 0.4 mm between the emitter and detector, the Isolation voltage becomes a function of the ambient air and is approximately 10 kVrms in dry air or 7 kVrms in normal humidity.

DESIGN CONSIDERATIONS

Photocoupler Model

It is useful to create a model of the photocoupler in order to better understand its transfer characteristics. Figure 6 can be used to model a photocoupler.

$$I_O = I_F\eta(I_F,t)K_1K_2K_3R\beta(I_P)$$

where:

I_O = Output current of the photocoupler

I_F = Forward current of the LED

$\eta(I_F,t)$ = Efficiency of the emitter (photons per electron)

K_1 = Transmissivity of transparent silicone resin

K_2 = Transmissivity of the transparent epoxy resin

K_3 = Sensitivity of the photo diode

$\beta(I_P)$ = Gain of transistor

As shown by the model above, the output current I_O is a function of the efficiency of the LED (which we will show degrades with time, temperature, and forward

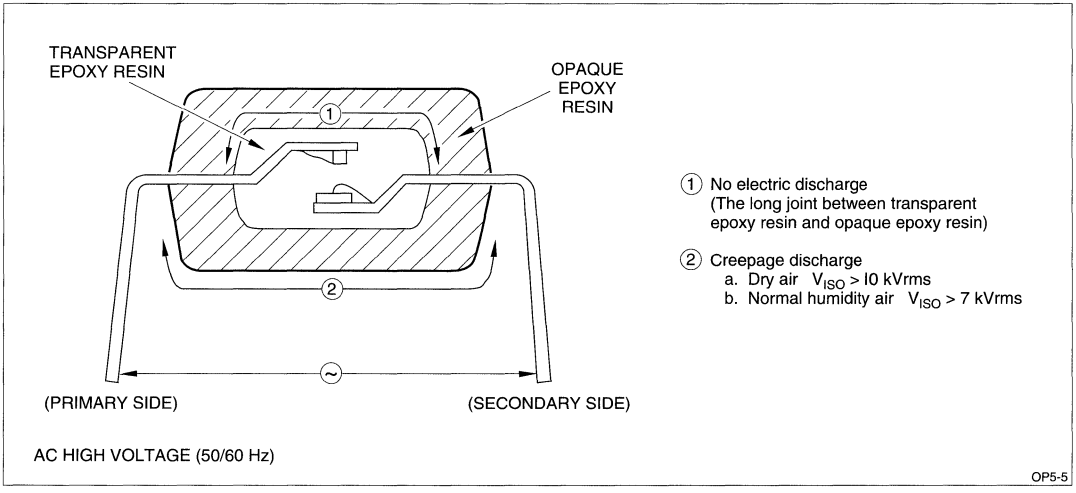


Figure 5. Double Transfer Mold Process of the PC700 and PC800 Series Photocouplers

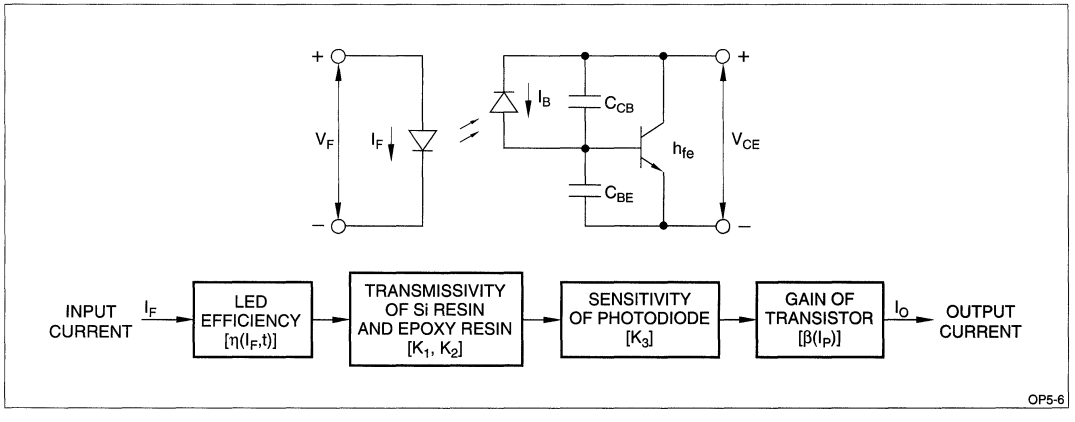


Figure 6. Phototcoupler Model

current), the transmissivity of the optical link, the sensitivity of the photo diode and the gain of the transistor. The transmissivity of the transparent silicone resin, the transmissivity of the transparent epoxy resin, the sensitivity of the photo diode and the gain of the transistor can be considered to be constant with time and temperature. Therefore, the most critical variable to the designer is the Current Transfer Ratio (CTR) as it changes with time, temperature, and Forward Current.

CTR

The CTR is defined as the ratio of the output current of the photocoupler (or more specifically the collector current of the photo transistor) to the input current of the IR LED expressed as a percentage value.

$$CTR = \frac{I_C \times 100}{I_F}$$

As an LED ages, its efficiency degrades with time at an approximate logarithmic rate for a given junction temperature and forward current (Figures 7 and 8). Since the CTR is a function of I_F, time and temperature, it will decrease as the LED's efficiency degrades. This degradation must be considered when designing a product for a long life cycle.

Degradation of Forward Current vs. Lifetime

Figure 7 shows a plot of the degradation of a generic photocoupler LED operated at +25°C over time as a function of forward current. As can be expected, a greater current density yields a more rapid degradation. For this reason, long life products should be designed with a lower forward current.

Degradation of Forward Current vs. Temperature

Comparing Figures 7 and 8, one can see the impact of ambient temperature on Lifetime vs. Forward Current. The designer should always consider the anticipated ambient temperature when selecting the forward current. It should be noted that Figures 7 and 8 allow the designer to graphically interpolate the degradation of CTR for a given temperature. The interpolation of these graphs provides sufficiently accurate plots from which an estimate of the CTR degradation due to lifetime and forward current may be obtained. The graphical interpolation method allows the designer to forgo the more tedious method of calculating the junction temperature for both the emitter and detector using junction resistance's, etc.

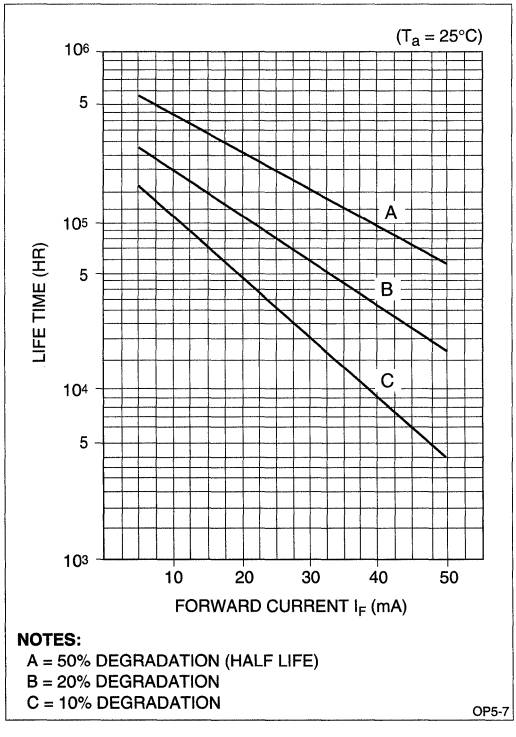


Figure 7. Lifetime vs. Forward Current at 25°C

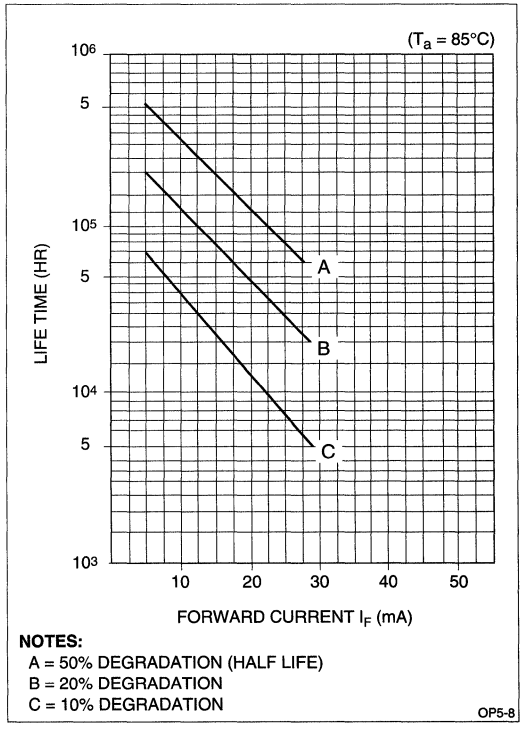


Figure 8. Lifetime vs. Forward Current at 85°C

Once the percent of degradation is determined by interpolation between Figures 7 and 8, the CTR vs. Forward Current curve (Figure 9) can be derated for time and temperature of the application. As an example, if the lifetime and temperature were interpolated to a curve that suggested a 20 percent decrease in CTR over the life of the product, then the CTR curve shown in Figure 9 should be reduced by an equal amount.

RESPONSE TIME

In time critical designs or when several photocouplers are used in parallel (such as in a data bus) a race condition could exist. The response time of the signal is the summation of the delay time through the photocoupler added to the rise time (T_R) or fall time (T_F) of the signal to guarantee a logic "high" or "low" respectively. The rise and fall time is typically a function of the load capacitance, load resistance, phototransistor, and V_{CE}. Curves such as the one shown in Figure 10 indicate response times with given Load Resistance's. As can be seen, additional capacitive effects (created by adding external capacitors) are not shown. By observing the Response Time vs. Load Resistance curve, an output load resistor can be selected to meet the response time. Response Time vs. Load Resistance curves are generally given with minimum collector-emitter voltage. If the collector-emitter voltage is greater than 2 V, the response time will be faster than that provided by the graph

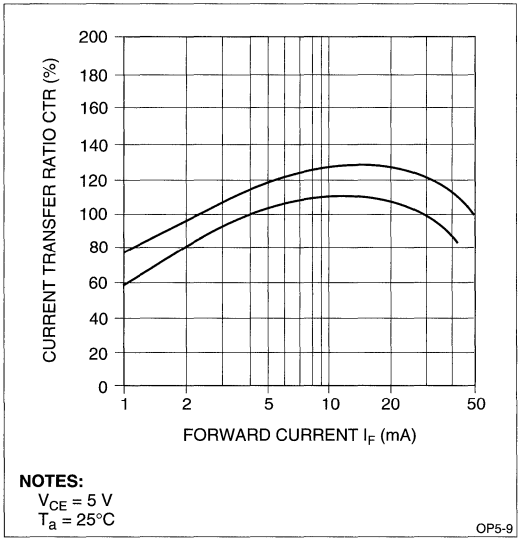


Figure 9. CTR vs. Forward Current Curve

ONE DESIGN APPROACH

When selecting a photocoupler, the obvious concerns such as the Isolation Voltage, Safety Approval rating, package type, Operating Temperature, response time, Emitter-Collector Voltage and Collector-Emitter Voltage should be addressed. Other factors such as product lifetime and ambient operating temperature must be considered to ensure reliable operation for the life of the product. The following is one design approach that aids the designer in addressing these issues. This approach does not consider all potential factors since each application is unique, however it does address those crucial to ensuring operation for the life of the product over the expected temperature range response time.

1. Determine the lifetime of the product in hours.
2. Determine the expected operating temperature range.
3. Interpolate between Figures 7 and 8 to obtain a CTR derating for product temperature, lifetime and Forward Current (the CTR can also be derated by using the Relative CTR vs. Ambient Temperature curves, however, these curves typically do not provide multiple Forward Current curves).
4. Derate the CTR vs. Forward Current curve to accommodate the degradation attributed to time and temperature.
5. Add engineering guardband of 5% to CTR by again derating the CTR vs. Forward Current curve.
6. Determine the response time needed for proper circuit operation.
7. From the Load Resistance vs. Response Time curves a load resistor can be selected.
8. Generally, the maximum collector-emitter voltage is determined by the interface. Therefore, with the load resistor selected and collector-emitter determined, the maximum collector current is easily calculated. The designer should consult the Collector Current vs. Collector-Emitter voltage curves to ensure that the photo transistor is not operated outside the maximum power envelope. (See Figure 11)
9. With the maximum collector current identified, and worst case CTR determined, the minimum Forward Current can be obtained by the minimum CTR vs. I_F. Figure 12 Assumes a typical digital interface interconnecting two logic gates with a photocoupler.

$$I_{F(Min.)} = \frac{I_O \times 100}{CTR_{(Min.)}}$$

$$R_{(In)} = \frac{V_{CC1} - V_F - V_{OL}}{I_{F(Min)}}$$

If the supply voltage is unstable, the below equations should be used to determine the variation of Forward Current as they relate to Figure 12.

$$I_{F(Min.)} = \frac{V_{CC1(Min.)} - V_{F(Max.)} - V_{OL}}{R_{in(Max.)}}$$

$$I_{F(Max.)} = \frac{V_{CC1(Max.)} - V_{F(Min.)} - V_{OL}}{R_{in(Min.)}}$$

VERIFY THE DESIGN

The designer should review the design to ensure that no Absolute Maximum Rating will be exceeded. Particular attention should be paid to the following:

- **Power Dissipation of the Emitter.** Review the Diode power vs. Ambient temperature graph and ensure the diode power is within the allowed envelope.
- **Power Dissipation of the Collector.** Review the Collector power dissipation vs. Ambient temperature graph and ensure the Collector power is within the allowed envelope.
- **Total Power.** It is quite possible that the Collector Power and Diode Power can operate within its the allowable range for each independently, however, the package as a whole is not capable of dissipating the maximum power of each. This measurement should be verified by the Total Power Dissipation vs. Ambient Temperature plot.

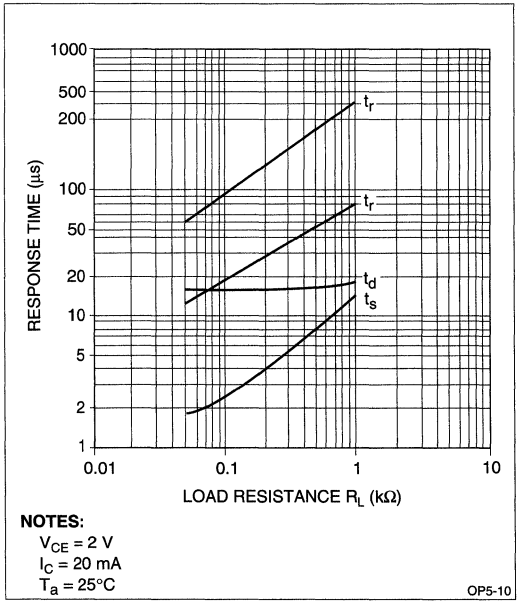


Figure 10. Response Time vs. Load Resistance Curve

- **Peak Forward Current.** If the photocoupler is not operated in continuous mode but is pulsed, particular attention should be paid to the Peak Forward Current vs. Duty Ratio curve. Operating outside this curve could cause permanent damage to the device.
- **Forward Current.** If operated in continuous mode, the Forward Current vs. Ambient Temperature curve should be verified to ensure operation is within this envelope.

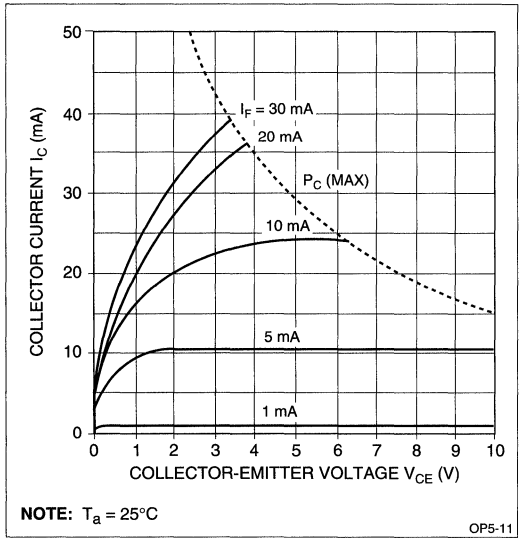


Figure 11. Collector Current vs. Collector-Emitter Voltage Curve

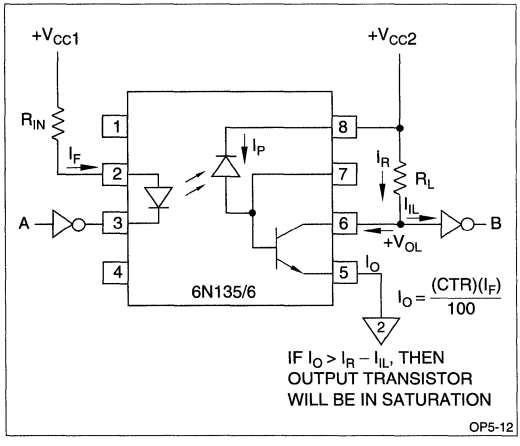


Figure 12. Digital Interface Interconnecting Two Logic Gates with a Photocoupler

PHOTOCOUPLERS

Bob Laird, Field Applications Engineer

INTRODUCTION

Sharp, as the world's leading producer of optoelectronic devices, currently has the broadest offering of photocouplers on the market. The number of different models offered easily exceeds 500 if you include all variations of a given base model. For example, the base model of the PC817 is offered in a standard DIP package and as the PC817I with formed leads for surface mounting. For each of these package types, there are multiple current transfer ratio (CTR) ranked versions offered, and there are also specially tested versions available that meet the TUV-0884 safety rating. The standard DIP packaged version is also available as a dual channel (PC827), triple channel (PC837), or quad channel (PC847) with multiple CTR ranked versions of each of these. When including all variations, the PC817 is offered as 76 different models. The purpose of this technical note is to offer some clarification of Sharp's photocoupler offering and to help simplify the photocoupler selection process.

Part Numbering Scheme The Sharp photocoupler part number scheme uses the prefix "PC" followed by a three digit model number. A breakout of the model number classification is shown below.

- PC100 Series DIP (long creepage distance type)
- PC300 Series Surface Mount (Phototransistor output type)
- PC400 Series Surface Mount (OPIC output type)
- PC700 Series DIP, 6 pin
- PC800 Series DIP, 4 pin or multi-channel type
- PC900 Series DIP (OPIC output type)

NOTE: OPIC is a trademark of Sharp and is an acronym for OPtical Integrated Circuit.

Special alphanumeric designators are used as a suffix to denote such things as CTR rank, lead forming, package type, special testing, etc. For example, a "V" suffix was added to several models when a packaging change was made by Sharp. Previously these models were all molded individually during assembly. A new assembly process was implemented that involves molding several photocouplers together in a long strip. The strip is then diced up with a diamond saw. As a result, the "V" type package is a little shorter than the older package. There are a few exceptions to the part number scheme described above. First, a general purpose 6N series is offered. These are standard, industry compatible devices that are still used in older designs. Second, Sharp offers a full line of phototriacs and photothyristors that use a different part number scheme.

DESIGN CONSIDERATIONS

Mechanical Requirements

The market demand today for photocouplers is still dominated by through-hole devices, but surface mount packages are gaining rapidly in popularity. As mentioned above, some of Sharp's standard DIP models are offered with a lead forming option. In this case, the leads are cut and formed in a gull wing to achieve a surface mount package (see Figure 1A). The lead forming option is denoted by an "I" suffix on the part number except in the case of the PC725 in which it is denoted by a "W".

Sharp is a pioneer in the development of new, ultra small surface mount packages (see Figure 1B). These packages are much smaller than the DIP packages with formed leads. For new surface mount designs, the photocoupler selection can usually be narrowed to the PC300 series or PC400 series. The main drawback of the small surface mount package is a reduced isolation voltage. Due to the shorter distance between leads, the standard isolation voltage of 5000 Vrms on the DIP packages drops to either 2500 Vrms or 3750 Vrms, depending on the model.

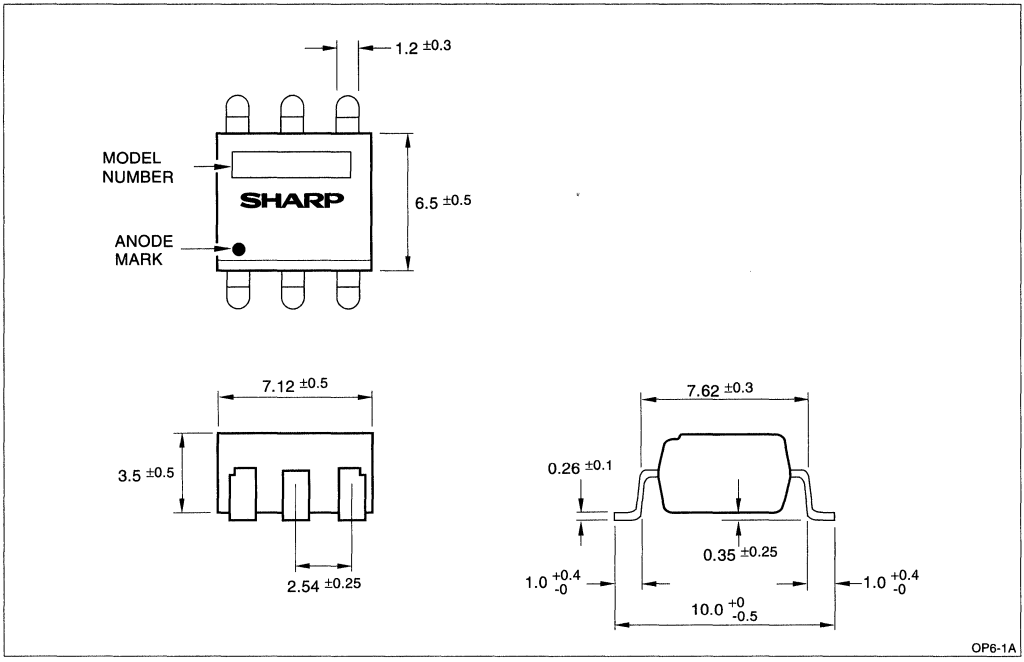


Figure 1A. Surface Mount Package

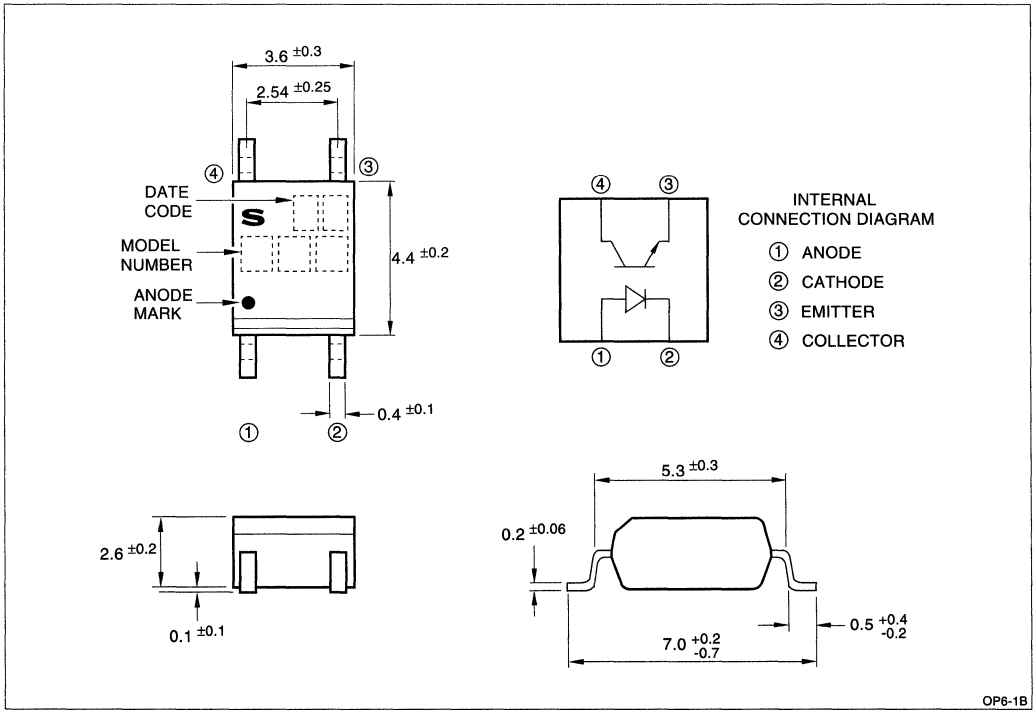


Figure 1B. Ultra-Small Surface Mount Package

There are various safety ratings that photocouplers must meet depending on the country and product that they will be used in. Agencies such as UL in the U.S., CSA in Canada, BS in Britain, and TUV/VDE in Germany maintain and update the safety ratings. Sharp's photocouplers are approved against many of the key safety ratings, so they generally do not represent an obstacle. The most important parameter for meeting the various ratings is isolation voltage (Viso), and Sharp's patented double transfer mold assembly process provides the best performance to Viso in the industry. Certain safety ratings specify a minimum external creepage distance and internal insulation distance that must be met by the photocoupler. Sharp offers the PC100 series with long creepage distances to meet these safety ratings. Also, on several models, Sharp is offering specially screened versions that comply with the new TUV-0884/VDE-0884 safety rating. A "Y" suffix is used to denote the special screening.

Electrical Requirements

The most common reason for requiring a photocoupler is to transmit a signal between two circuits while maintaining a very high level of isolation between these circuits. Since a photocoupler transmits a signal from its input side to its output side via light instead of an electrical conductor, the input and output are isolated from each other. The electrical requirements of each of the circuits being isolated help define which of the various photocouplers is best suited to the design.

Photocouplers offer only a few variations on their inputs. Most models have a standard IR LED with a maximum DC forward current of 50 mA. One variation is an AC input type, in which two IR LEDs are used. The two LEDs are connected in the opposite direction (see Figure 2), so that one of the LEDs will always turn on regardless of the polarity of the input signal. The PC814 series are the most popular AC input photocouplers offered by Sharp. AC input photocouplers are commonly used in industrial controllers or telecommunications equipment. Another less common variation on the input pertains to the forward current or input current of the LED. A high input current model such as the PC724V is available in which the maximum DC forward current is 150 mA. Or several low input current models are available such as the PC866 in which the electrical characteristics of the device are specified with an input current of 1 mA.

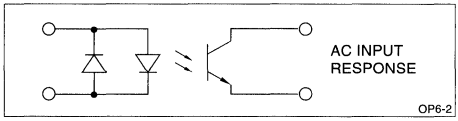


Figure 2. AC Input LED

Many different variations are offered on the output side of a photocoupler. A selection tree is shown in Figure 3. When selecting an output configuration, there are a few different decisions to be made. The OPIC output offers simplified design and improved reliability since a lot of external circuitry has been integrated onto one chip. The fact that they offer a digital output makes them ideal for interfacing to standard digital circuitry. They are an excellent choice when a photocoupler is used to clock digital data. One common application where our OPIC photocouplers are used is electronic musical instruments that support the Musical Instrument Digital Interface (MIDI) standard. The PC900V or PC400 are typically the models used in the MIDI designs. There are several applications where the output of the photocoupler is used as a simple switch. In this case, a simple phototransistor output will work fine, so the higher cost of the OPIC type photocoupler is not justified. Phototransistor outputs are also good for interfacing to circuits operating at non standard voltage levels. Some of the issues that must be addressed when selecting a phototransistor output are discussed below.

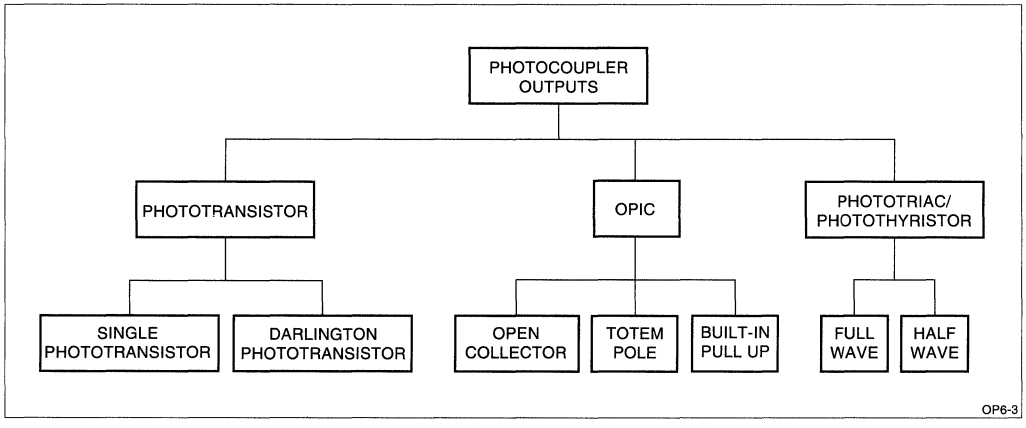


Figure 3. Photocoupler Output Selections

PHOTOTRANSISTOR OUTPUT

Collector-Emitter Voltage

Most commonly the phototransistor has a maximum collector-emitter voltage (Vceo) rating of 35 V. For those applications where the load circuit will be supplying a collector-emitter voltage in excess of 35 V, models are available with a Vceo of 70 V or higher. Such models as the PC729 and PC829 have a Vceo of 300 V for interfacing to telephone lines where large electrical disturbances commonly occur.

Base Terminal

Older photocouplers, such as many of the PC700 series photocouplers, offer a base terminal connection to allow the design engineer more control over the output current. With todays more advanced process controls, the variation in common emitter current gain from one phototransistor to the next is small enough that it is generally not necessary to use the base terminal to control output current.

Noise Immunity

Phototransistors that are fabricated to have large junction capacitances offer improved immunity to common mode noise over standard phototransistors because of the larger inherent time constant. Models such as the PC812 and the PC813 series utilize phototransistors with larger junction capacitances. The larger junction capacitances do result in slower response times.

Response Time

Phototransistors do not have a fast response time in comparison to photodiodes. Therefore, for true high speed requirements a photodiode type output is required. An OPIC uses a photodiode, so high speed OPIC photocouplers are available such as the PC910 and PC911 with 10 Mb/s response times. The 6N series also use photodiodes, and are capable of high speeds. For those specialized applications in which it is necessary to drive a load that is highly resistive relative to typical loads, Sharp offers the PC810 and PC818. Both devices are rated at relatively high speeds with 1k Ω loads.

Collector Current

The available collector current is represented by the CTR specification which is a ratio of collector current to input current specified as a percentage. The minimum and maximum ratings for the CTR are specified for a given input current. A plot of typical data for CTR versus input current is included in the specification to provide for extrapolation for the various input currents that different designers will be using. The difference in the minimum and maximum ratings for CTR can be very large depending on the model. Many designers cannot tolerate such a large potential variation; therefore, some models are offered with different CTR rankings. For example, the PC817 is specified with a CTR of 50% minimum and 600% maximum. The other rankings available are shown in Table 1.

Table 1. CTR Rankings

MODEL NO.	CTR (%)
PC817A	80 to 160
PC817B	130 to 260
PC817C	200 to 400
PC817D	300 to 600
PC817AB	80 to 260
PC817BC	130 to 400
PC817CD	200 to 600
PC817AC	80 to 400
PC817BD	130 to 600
PC817AD	80 to 600
PC817	50 to 600

Some circuits require a higher output current drive than what can be achieved with a single phototransistor. For these applications, photodarlington outputs are used. In the photodarlington configuration (see Figure 4), a second transistor is used for amplification purposes resulting in increased sensitivity. The improved sensitivity is represented by large CTR values in the specification. The photodarlington transistors do have a slower response time than single phototransistors.

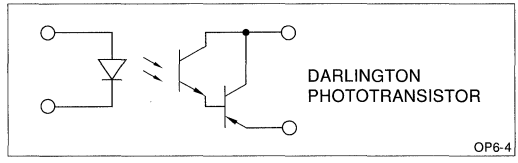


Figure 4. Darlington Phototransistor

SUMMARY

Sharp's line-up of photocouplers were designed for a multitude of different applications. This technical note has touched on many of the issues that should be addressed when selecting a photocoupler. Since many of the photocouplers are very similar in performance, it is possible that several different models would function properly in a given application. Therefore, it is important for the designer to work closely with the appropriate Sharp salesperson and Field Applications Engineer during the selection process.

SHARP

NORTH AMERICA
SHARP Electronics Corporation
Microelectronics Group
5700 NW Pacific Rim Blvd.
Camas, WA 98607, U.S.A.
Literature: 800-642-0261
Phone: 206-834-2500
Telex: 49608472 (SHARPCAM)

JAPAN
SHARP Corporation
IC Sales Department
International Sales & Marketing Group
IC/Electronic Components
Integrated Circuits Group
1, 2613-banchi, Ichinomoto-cho
Tenri-shi, Nara Pref. 632, Japan
Phone: (07436) 5-1321
Telex: LABOMETA-B J63428
Facsimile: (07436) 5-1532

EUROPE
SHARP Electronics (Europe) GmbH
Microelectronics Division
Sonninstraβe 3
20097 Hamburg, Germany
Phone: (49) 40 2376-2286
Telex: 2161867 (HEEG D)
Facsimile: (49) 40 2376-2232

Reference No. SMT93009-B