Conceptual design of an eight megabyte high performance charge-coupled storage device

by B. AUGUSTA and T. V. HARROUN

IBM Components Division
Essex Junction, Vermont

SUMMARY

The design approach suggested to satisfy the conceptual requirements was the use of self-contained, charge-coupled storage chips with on-chip decoding. In this approach, the information on the memory chip is stored in a group of closed-loop shift registers, and random access is provided to any one of the registers by an on-chip dynamic FET decoder. In this way, n-control lines can select one of 2^n shift registers.

Of the many organizations possible in expanding the on-chip decoding concept into a design for a 10^8-bit memory, a bit-per-chip organization was chosen. This proposed organization results in reasonable chip power dissipation and was contained in successively higher levels of packaging. The operating characteristics are summarized in Table I.

The models, limited in capacity to that necessary to show feasibility of the approach used, are intended to demonstrate the operability of the conceptual design. The operations necessary to perform the functions of the storage chip are charge injection, charge transfer efficiency, sensing, absence of channel cross-talk, turn around and charge generation.

These are all performed using a silicon self-aligned gate structure driven by a four-phase electric field, and are basic to the operation of the shift registers. Decoders which select one of the 2^n shift registers have been amply explored by industry.

Two small shift register structures were designed, fabricated and tested to demonstrate these functions and concepts. One, a 480-bit register demonstrating injection, sensing, amplification and turnaround, was designed using 0.3-mil line widths. The second, a 256-bit shift register demonstrating high density, charge retention, and absence of cross talk, was designed using 0.15-mil line widths. Operation is summarized in Table II.

The conceptual design and the feasibility models are described in considerably more detail in the following sections.

GENERAL CCD MASS STORAGE DESIGN

This section describes the conceptual design of a 10^8-bit charge-coupled device mass storage unit. The design principles of the chip, module, system, and error correction and detection are discussed first.

The system is designed to be used in a block-oriented mode, in which random access is provided to blocks of information, which are then read out or written in serially. Physically, this means that the CCD chip is divided into a number of closed-loop shift registers that store the blocks of information. Random access to these shift registers is provided by on-chip FET decoders. Propagation in the shift registers, which occupy the major portion of the chip, is accomplished with a two-level interconnection pattern activated by an external four-phase electric field.

The further organization of such chips into a memory system depends on a number of design criteria. Since the application contemplated here was industrial, low cost and operational life were emphasized. In addition, the storage system was required to be of small size and low power and to operate in typical industrial environments without elaborate support. Consequently, the chip design should give reasonable chip yields, the package should be reasonably easy to manufacture, and the means of achieving the reliability should not be prohibitively expensive fraction of the total storage. These topics are considered in the next several sections. The one remaining task of this section is to specify the access time and data rate of the storage system. After examining present and future needs, it was decided to aim for average access times below 0.5×10^{-3} sec and

TABLE I—Characteristics of the Design of a 10^8-Bit Memory

Capacity	1.0×10^6 words of 73 bits each
Shift frequency	500 kHz
Data rate	32×10^6 bits/sec
Access time	256 μsec to a block of 256 words 512 μsec to a word
Power	237 watts

data transfer rates of at least 3×10^6 bytes/sec. The next few sections describe the influence of these requirements on the design evolution.

Chip, module and card design

Chip size

It is desirable to have as many bits per chip as possible, consistent with reasonable chip yield. In addition, high speed and low cost are easier to obtain as the density of information storage (bits/square cm) increases. These are the primary considerations in choosing the size of the chip.

Based on fabrication experience and on progress in the industry, it seems reasonable to assume that layout ground rules with 0.15-mil minimum line widths will be practical for the mid-seventies. Assuming that a four-phase structure is used to store a bit of information and that information flow is reversed between adjacent parallel channels to close a shift register, topological considerations yield a storage cell area of 15.4 w^2, where w is the minimum line spacing. This corresponds to a storage density of $\sim2.9\times10^6$ bits per square inch. With such densities, a 32,768 bit chip might to be feasible.

Chip layout

The chip organizational layout is uniquely determined by the average access time that the system is to have. Except for some relatively small decoder delays, the average access time T_A is given by

$$T_A = \tfrac{1}{2} B_{SR} T_S$$

where B_{SR} is the number of bits per shift register and T_S is the time period of one shift. The shift frequency is chosen to be 500 kHz to minimize chip power and allow variable speed operation. Hence, $T_S = 2\times10^{-6}$ sec. It was stated earlier that T_A should be less than 0.5×10^{-3} sec. Therefore, B_{SR} must be less than 500 bits. The nearest binary number is 256 bits so the proposed chip design calls for 128 shift registers of 256-bits. This results in an average access time of 256 μsec plus a small amount of decoder and other delays.

The chip layout is shown schematically in Figure 1. Each of the 128 closed-loop shift registers has an input-

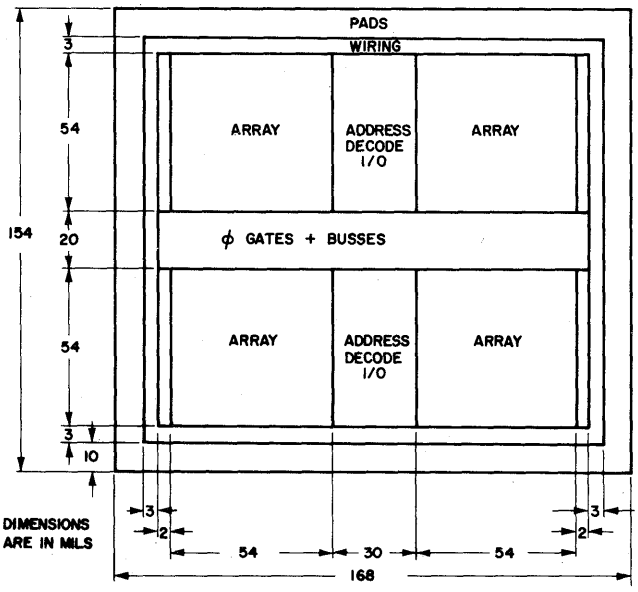

Figure 1—Proposed memory chip layout

output amplifier, four restore amplifiers and a select decoder. These are organized into four nearly square groups of 32 shift registers, allowing shortened phase lines to drive from the center of the chip. Each shift register is folded four times to match I/O amplifier pitch. A restore amplifier follows each 64-bit register segment.

The relative areas occupied by each function on the chip are defined in the following way. A shift register is the product of the vertical and horizontal periods of the cell. The storage section of each array contains 32 shift registers, occupying an area of

$$(64\times5.3\text{ w})\times(32\times4\times2.89\text{ w}) = 12{,}600\text{ w}^2 = 8192\text{ cells.}$$

TABLE II—Measured Operating Characteristics of Feasibility Models

Material	Silicon, n-type substrate
Structure	Modified self-aligned gate, 2-level interconnection
Cell size	0.32 mil^2
Phase voltages	10-12 volts
Output signal	4-volt amplitude
Charge retention	> 0.25 sec
Minimum shift rate	5 kHz

By comparison, each restore amplifier occupies

$$16\ w \times 5.8\ w = 94\ w^2$$
$$= 6 \text{ cells and the decoder I/O amplifier occupies}$$
$$(84\ w \times 11.6\ w) + 2(16\ w \times 2.9\ w) = 1064$$
$$= 69 \text{ cells.}$$

Phase gates and bussing occupy a total area on the chip equivalent to 8200 cells. The resulting chip, including wiring and pad space, is 168 mils by 154 mils.

Register selection, writing, reading and clearing are all accomplished with 14 control lines. Phase drive, reference voltages and supply voltage require 11 lines, resulting in 25 I/O connections per chip.

Module design

The proposed module is a design extension of IBM's widely used logic and monolithic memory module. The proposed module, however, is 35 percent larger than the standard module (0.580 in. square) and provides an increased number of I/O pins. Since only one chip is active at a time, the module thermal characteristics permit two chips to be accommodated per module. Consequently, the module has two stacked ceramic substrates, with one chip per substrate (Figure 2), joined by the IBM controlled collapse solder connection.

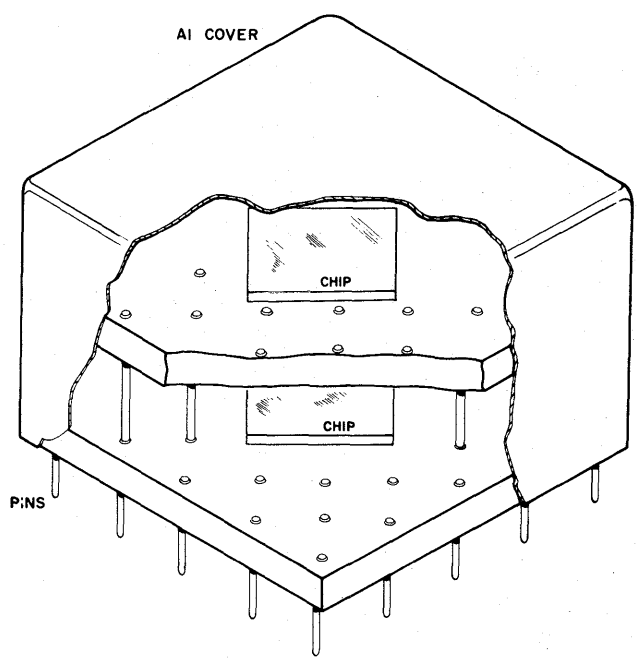

Figure 2—Basic memory module

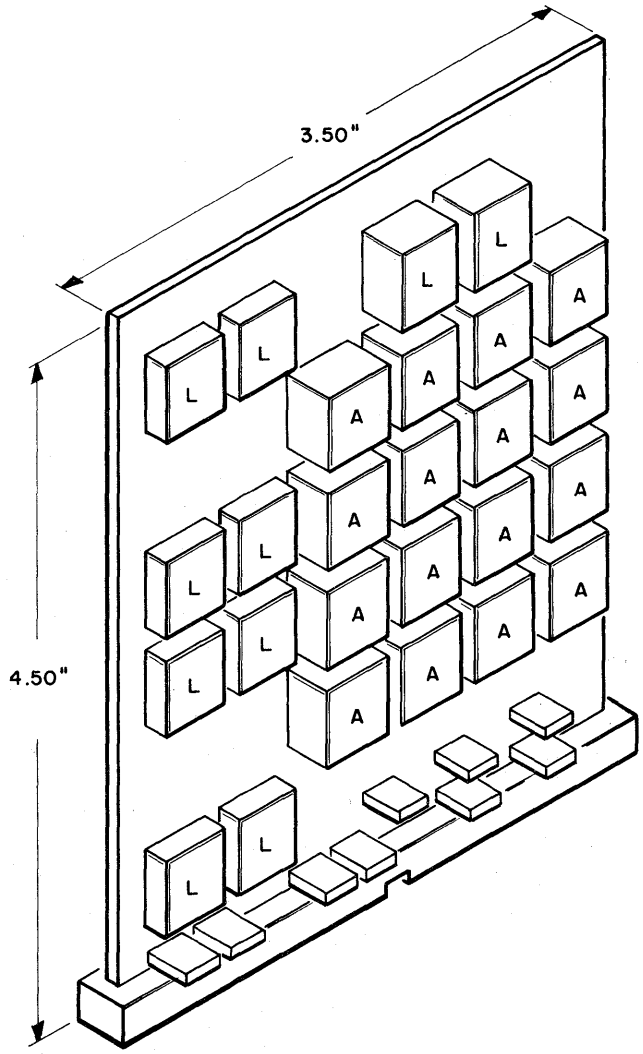

Figure 3—Storage card assembly

To keep the number of interconnections per module down to an acceptable level, the chips are arranged so that address, phase and voltage connections are shared. Input-output and select control lines are separate, resulting in a total of 28 active module connections.

Card design

Further assembly of the memory proceeds by placing 16 of the modules described above on a card as shown in Figure 3. This package or card also contains partial address decode, refresh, control logic, and address interface and phase drivers to provide fan-out for the memory modules. The card or memory subunit is actually a self-sufficient memory in its own right (except for logic and driving circuits), providing a storage capacity of

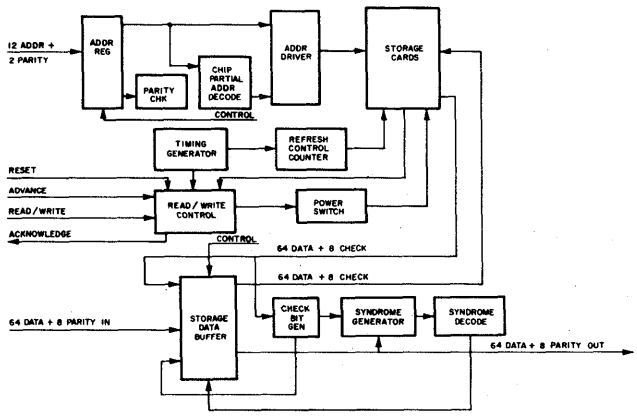

Figure 4—Functional diagram

10^6 bits. The 10^8 bit CCD memory is realized by stacking up the appropriate number of these building blocks. They lend themselves to several different memory organizations, the exact number depending on the organization chosen. In the "bit-per-card" design, a memory word consists of 64 data bits, 8 check bits, plus a position synchronization bit, for a total of 73 cards. The memory is operated in a mode where only one chip per card is selected at a time, providing a substantial savings in card power dissipation. Since the cards are independent sub-memories, serviceability is enhanced.

Summary of chip, module and card designs

Many of the reasons for the particular choice of chip, module and card design have been given in the preceding paragraphs. For further discussion, however, it is necessary to know how the memory will be organized, and this requires a knowledge of the electrical circuitry, packaging and error correction coding needed to implement the possible organizations. Discussion of these topics, comparisons and tradeoffs, however, will be based on the above described chip, a module with two chips and a card with 16 storage modules, logic and interface drivers as described. The storage array card is organized in 1024K word by 1-bit configuration.

TABLE III—Power Requirements For System Control Functions

FUNCTION	EQUIVALENT GATE COUNT	POWER (mw)
Address register	105	3920
Parity check circuit	56	640
Chip partial address	21	637
Address drivers	57	2570
Timing generator	43	2310
Refresh control counter	188	4030
Read/write control	161	6821
Power switch	147	7120
Storage data buffer	588	9500
Check bit generator	845	18120
Syndrome* generator	76	1440
Syndrome decode	180	3240
	2467	60.347 watts

* Syndrome-encoded signals generated as a result of a bit error from which the incorrect bit can be located.

Electrical and mechanical design

Circuit requirements

The circuits designed to implement the functions defined in the memory system design are tentative and have not been optimized. They are only intended to represent a reasonable estimate of complexity and power consumption. For the sake of discussion, all the numbers that follow pertain to the bit-per-card organization.

TABLE IV—Power Requirements (Watts) For Memory System

System logic	60.5
Storage array 73 cards @ 2.185	~160.0
Storage fan-out drivers	15.7
	236.2

The general functional diagram of the system is shown in Figure 4. Twelve address lines provide addressing to each of the 4096 2048-byte blocks of stored data. An eight-byte (64 data bit) parallel interface data bus provides a high data transfer rate with an acceptable investment in power and error correction circuitry. The power and equivalent gate count for each function are presented in Table III. The system power is presented in Table IV.

Packaging design relies upon IBM's standard card and board concepts. The technical objectives of this design concept are (1) minimization of interconnection complexities, (2) adequate environmental protection, (3) sufficient thermal efficiencies to employ forced ambient air cooling, (4) modular flexibility features suitable for expansion, and (5) favorable economic costs.

The overall packaging configuration of the memory system is pictured in Figure 5, resulting in a volume of 2.9 cubic feet.

The unit consists of a gate-structure containing four multilayer circuit boards, covers, interconnecting

cables and pluggable modular card subassemblies containing logic and memory circuits. This unit design does not contain power supplies. It is assumed power and cooling air will be furnished by the host machine.

Power consideration

The CCD cell uses a dynamic charge storage principle and must be periodically regenerated to account for the charge "lost" from the storage node. This regeneration is accomplished by means of restore amplifiers in each shift register. Information is stored as charge in the potential wells, and the ability to differentiate between the amount of charge represents a bit of information. The minimum operating frequency of the shift registers is determined by detectable charge difference, which is a function of time, transfer efficiency and leakage rate. Careful consideration of these factors results in a lower shift rate of 5 kHz with a restore amplifier located every 64 bits. The restore amplifier senses the charge difference remaining at the end of a register and restores the charge levels to those corresponding to an initial one or zero for the next register segment. At this frequency, the chip requires a "standby" power of 19.4 mW. In the standby state, information integrity is maintained but all functions not required to that end are de-powered using pulse power techniques.

The maximum operating frequency is determined by the allowable power dissipation of the chip for forced air cooling. These considerations result in a shift frequency of 500 kHz with a selected chip (i.e., all circuits fully powered) power of 321 mW.

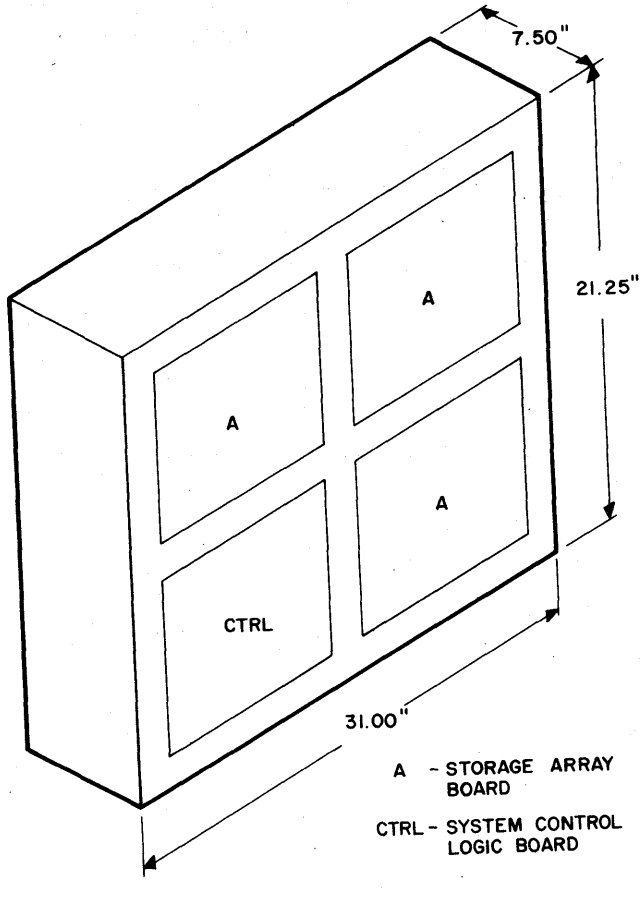

Figure 5—Eight megabyte storage unit

Error detection and correction

A preparatory phase of formulating a reliable storage system requires careful consideration of failure modes for the devices used in the system. Prediction of classes of device failures for the complete memory system is used to impose constraints upon both the chip and system organizations to assure the desired reliability. These considerations are based primarily upon the relative amount of circuitry used to implement the reliability-enhancement features.

To overcome the effects of shift register malfunctions, the memory system is organized so that each bit from a shift register is part of a word located on a different chip.[1] In this memory system organization, the memory device failure modes manifest themselves as either single or, with lessor probability, double bit errors.

The widely-used Hamming-type SEC/DEC* codes, for example, can correct any single-bit error in a memory word, but can only produce an error message in case of a double or higher-order bit error. Such codes, therefore, are most effective in systems organized so that as many failure mechanisms as possible cause only single-bit errors.

The design philosophy leading to the final, recommended, bit-per-chip system was as follows:

1. Offer random access to a block of 256 words of 64 bits each, and serial access to a specific word within that block.
2. Use SEC/DED codes to enhance reliability.
3. Achieve high performance operating characteristics at reduced power by supplying the high speed shift field to only the chips containing the desired shift registers, since only 64 shift registers (representing 0.2 percent of the total memory capacity) are accessed at any one time. This is accomplished by including phase gating logic on the chip, to be activated only when the chip is selected. Additional support electronics are

* Single-Error-Correction/Double-Error-Detection.

TABLE V—Summary of Chip Design Features

CONCEPTUAL DESIGN
- Self-contained CCD chip with on-chip decoding
- n-type substrate SAG technology
- Closed loop shift registers
- Operating speed of 5 to 500 kHz
- 64-bit shift register segments
- Shift register turnaround and restore amplifiers
- Four-phase operation
- 2.9×10^6 bits/in^2
- Operation in a machine environment

OPERATING FEASIBILITY MODEL

High Density Model
- n-type substrate SAG technology
- 256-bit single shift register segment
- 2.1×10^6 bits/in^2
- Operating speed of 500 Hz to 5 MHz
- Four-phase operation

Operating Memory Model
- n-type substrate technology
- 480-bit, 10 segment shift registers
- Shift register turnaround and restore amplifier
- Four-phase operation
- Operation in a machine environment
- Operating speed of 500 kHz
- Wide operating parameters
- I/O amplifier

necessary to insure data retention and regeneration; however, dc driver power dissipation is reduced.

FEASIBILITY MODELS

This section deals with the feasibility models that were built to test and demonstrate the major operating features of the conceptual memory design described above. The models were to be limited in capacity to that necessary to prove feasibility of the approach used in the conceptual design. The essential features of the conceptual chip design and the models are summarized in Table V.

The feasibility model chips differ from the conceptual chip not only in scale but also in that they do not demonstrate on-chip decoding or closed-loop operation. These have been relatively simple functions to accomplish, as shown by workers in this laboratory and industry as a whole. Mounting of the conceptual chip uses techniques well-known to industry.

Critical features of the conceptual design have been demonstrated by the feasibility models. These include the storage cell density, operating speed, sensing and amplification, and sufficient operating parameter tolerance for actual machine environment.

Device structure

The charge-coupled device uses basically MOS technology. The described structure consists essentially of three layers and is a junctionless device except for small diffused junctions that serve as input and output nodes of the shift register. The surface of a semiconductor, such as silicon, is oxidized to form a thin insulating layer. A metal pattern of electrodes is deposited on top of the insulator. In operation, the shift register depends on the transfer of charge from the potential well developed under one electrode to another by application of suitable voltages to these electrodes.

A cross-sectional view of the overlapped electrode devices fabricated in this test chip is illustrated in Figure 6. In the structure shown, each spatial bit has associated with it four independent electrodes. The Ø1, Ø3 electrodes, doped polysilicon, define the storage potential well (node) locations. The Ø2, Ø4 aluminum electrodes serve as transfer/isolation gates between storage nodes.

Device size

The proposed CCD storage cell is designed with an electrode separation of 0.15 mils and 0.05-mil overlap.

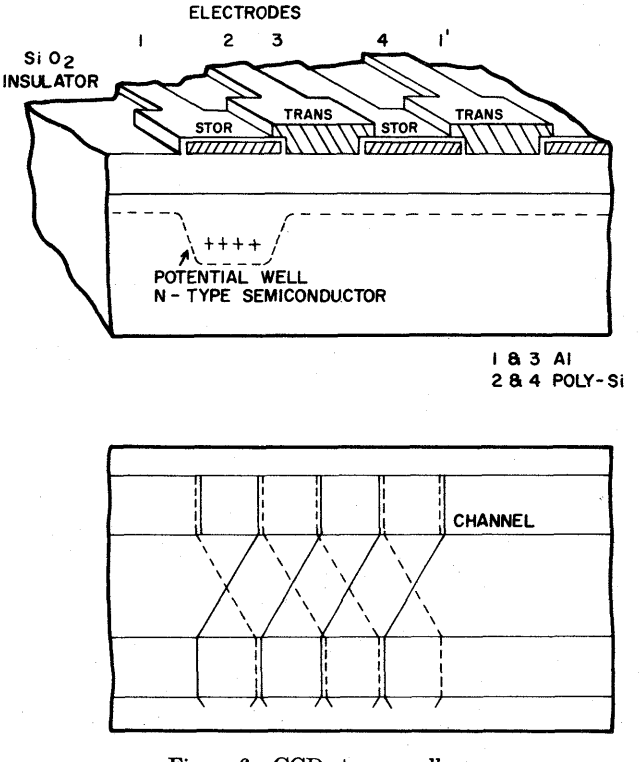

Figure 6—CCD storage cell

Channel width is 0.15 mils with a channel separation of 0.28 mils, resulting in an average area per bit of 0.35 mil.[2] In the high density devices described here,[2] the average area per bit is 0.48 mil[2] corresponding to a channel width and separation of 0.2 and 0.4 mils. Calculations of the potential barrier between channels indicate that the channel separation can be reduced to 0.2 mil, while stilll providing adequate isolation in 1-2 ohm-cm material.

Frequency response

In those pulse powered memory applications where access time minimization and power are important considerations, the frequency response is an important parameter. The frequency response curve for the normalized worst case one/zero difference (ΔD) of a typical 128-bit shift register is presented in Figure 7. Since the one/zero difference is directly related to the charge transfer efficiency, the data imply that the transfer efficiency characterizes device operation over an extremely wide clock frequency range.

The primary mechanism determining the low frequency limit of device operation is the thermal charge generation rate and the tendency of the empty potential wells to fill with thermally generated minority carrier charge. The 500-Hz data point indicates that such room temperature leakage is negligible at information dwell times of at least ¼ second.

The primary parameter affecting the high frequency limit of device operation is the surface mobility and its determination of the maximum transfer times needed to preserve efficient charge transfer between storage nodes. At a clock frequency of 5 MHz, the nominal transfer time duration is 60 nsec at which a slight drop in signal occurs. The signal difference is sufficiently large so that sensing is not impaired.

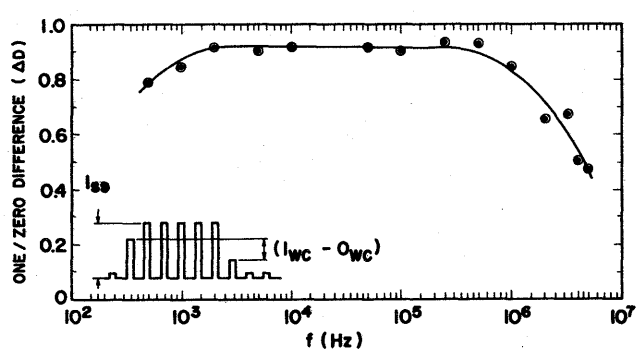

Figure 7—CCD charge transfer characteristics

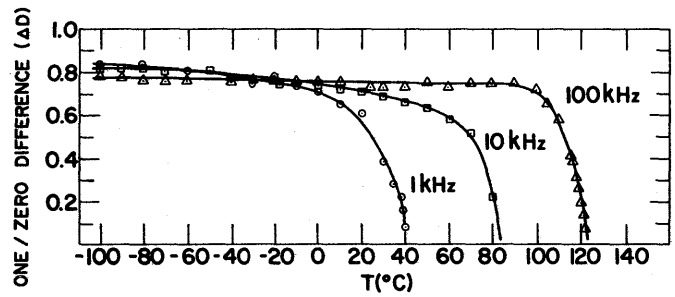

Figure 8—CCD temperature characteristics

There are two primary ways in which temperature can impact CCD operation. First, high frequency response is expected to be lowered with increasing temperature due to a decrease in surface mobility and a consequent decrease in charge transfer efficiency. Secondly, the low frequency response limit is expected to increase with increasing temperature due to the increased rate of thermally generated charge filling an empty potential well. The observed temperature dependence at the three clock frequencies shown in Figure 8 clearly displays the second effect described above. These data were obtained with the substrate biased at a relatively high level (8V), representative of stress conditions. At more moderate substrate bias potentials (~2V) the curves shown shift to significantly higher temperatures. This is due to the fact that a reduction in substrate bias reduces depletion region depth, resulting in a reduction in the rate of carriers filling the potential wells.

Operational memory

All the proposed circuit design concepts were exercised in a fully operational memory system.[3] The chosen memory system used dual 2880-bit shift register memory buffers operating at a fixed clock frequency of 500 kHz. The two 2880-bit shift register buffers were fabricated from six dual 480-bit open-loop shift register memory chips, serially connected to form the memory buffers. Each chip contains two 480-bit shift registers and is fabricated as previously described. Data flow proceeds in one direction for 47 ½ bits and is then amplified and launched in the reverse direction by a sensitive but simple amplifier that introduces an additional ½-bit delay. The restore amplifier, designed to operate in so-called fat-zero mode, consists of three FET's as illustrated in Figure 9. The amplifier inverts the signal and provides a small signal gain of about 80 from channel to channel. In operation, the signal at the launch

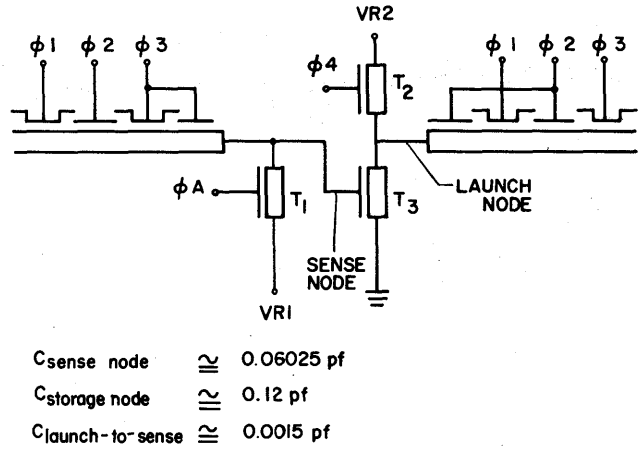

Figure 9—Restore amplifier

node is clamped to either VR_2 or near ground and the gain realized is only that necessary to compensate for the losses in the channels. Input and output support circuitry for the CCD shift registers interface directly to machine logic levels. Machine logic is diode-transistor with logic zero at ground and logic one at plus six volts. The chip output is capable of driving a minimum of one logic load (sink 1.7 mA to ground) and is fully compatible with both machine logic levels and CCD memory chips.

In summary, the work done in fabricating, testing, and designing the feasibility chips has demonstrated that the CCD technology is sufficiently mature and understood so that design and fabrication of a 10^8-bit storage system is possible with an acceptable risk factor.

REFERENCES

1 M Y HSIAO
A class of optimal minimum odd-weight-column SEC-DED codes
IBM Journal of Research and Development Vol 14 No 4 July 1970

2 N A PATRIN
Performance of very high density CCD structures
1972 Device Research Conference University of Alberta Edmonton Canada June 21-24 1972

3 N G VOGL T V HARROUN
Operating memory system using charge-coupled devices
1972 IEEE International Solid-State Circuits Conference February 16-18 1972