A 16 KILOBIT HIGH DENSITY CCD MEMORY

S.D. Rosenbaum J.T. Caves S. Poon C.H. Chan

BELL-NORTHERN RESEARCH LTD., Ottawa, Canada.

ABSTRACT

The organization and performance of a 16,384-bit charge-coupled device (CCD) memory are described. The aim was to provide lower cost per bit of storage consistent with high data transfer rates, moderate access times and relatively low system operating power. The chip has dimensions of 136 x 169 mil^2, to fit a standard 16-pin package, and is organized as four separate shift registers of 4096 bits, each with its own data input and data output terminals. A two-level polysilicon gate process was used for device fabrication. Three functional modes are provided: recirculate, read and recirculate, read and write, controlled by chip select and write enable control inputs which apply to all four registers together.

A condensed serial-parallel-serial (CSPS) organization was found to provide the highest packing density and lowest system cost per bit, but requires various clock waveforms. Most of these are generated on-chip, some operating at the serial transfer frequency and some at the parallel transfer frequency. Only two external clocks are required, driving capacitances of 60 pF each at one-half the data transfer rate. Operation at data rates of 100 KHz to 10 MHz have been demonstrated experimentally, the overall operating power at 10 MHz being less than 20μW/bit.

1. INTRODUCTION

For the design of systems requiring digital memory, including memory hierarchies for large and small computers, the principal performance parameters, in order of importance are: access time (also called latency in serial memories), reliability, maximum frequency of data transfer, power per bit during normal operation, and standby power per bit in the case of volatile memories.

Comparison of the various available technologies with regard to cost per bit as a function of access time leads to a relationship as shown in Fig. 1, illustrating the "access gap" in which there has been no technology which is both available and cost-effective within the computer memory hierarchy. Magnetic cores, bipolar and MOS RAMs have been fast but too expensive, while drums and disks were cheaper but too slow. MOS shift registers have found wide applications where small amounts of memory were

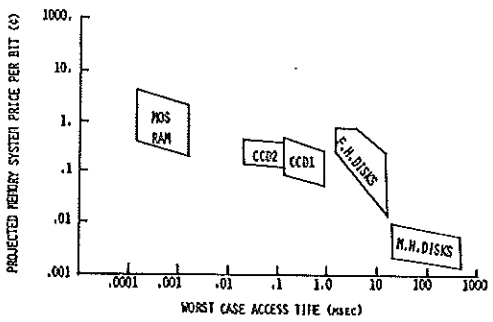

Fig. 1 - MEMORY SYSTEM PRICE VS ACCESS TIME

needed, for very small systems or for buffering mechanical devices such as disks whose speed cannot be usefully varied, but their cost per bit has been greater than for RAMs by a factor of two or more, when the access gap really needs something less expensive.

Discussions with many systems designers have indicated that the CCD serial memory technology can play a useful role in filling the access gap, eventually at lower system cost per bit than RAMs, yet offering performance that cannot be achieved using drums or disks. Many of the uncertainties which are inherent in cost predictions can be avoided by taking the MOS RAM technology as a benchmark, and taking into account the relative packing densities, process complexities, and expected yields. It is not to be expected that CCDs will compete directly with RAMs or with moving head disks, but the arrival of CCD technology will encourage re-evaluation of the trade-offs which determine the relative amounts of mainframe and disk storage in present-day systems. To allow for residual uncertainties, and to overcome inertia in establishing new products, a projected factor-of-two cost advantage over RAMs has been suggested as the minimum which is necessary to induce semiconductor manufacturer and systems designers to take CCDs seriously. The development of the 16K CCD memory to be described here was aimed at projecting a fourfold system cost advantage over RAM systems, even if this required the acceptance of longer access times than those already achieved in CCDs, including block-addressable memories [1]. A number of such devices have been developed by various manufacturers and laboratories; most of them provide access to relatively small blocks of serial data, of typically 256 or 128 bits, although differing in maximum data rates, operating power, and system overheads for clocking, data transfer and data refreshing. The new 16K device described here, designated CC16M1, achieves a higher packing density by using an improved form of serial-parallel-serial organization using relatively large arrays of 2048 bits each, two such arrays being paralleled for each block of 4096 bits. Despite this large block size, the worst-case access time is only 410 μ sec at a data transfer rate of 10 MHz. However, this organization requires a larger number of clock waveforms, which therefore must be generated on-chip in order to minimize the number of external connections and to keep the system overheads low.

II. ORGANIZATION OF CCD CHIP FOR HIGH DENSITY STORAGE

Comparison of various alternatives led to the selection of a serial-parallel-serial (SPS) organization, which offers several advantages. Especially, these include greatly reduced clock power at high data transfer rates, since the SPS structure reduces the number of transfers experienced by each charge packet to a small fraction of what it would be for a purely serial shift register. Also, for any given set of geometrical layout rules, serial-parallel-serial (SPS) organizations offer high packing density because only one sense amplifier is required for a large array, such as the 2048-bit array chosen for the present design. This allows space for designing the sense amplifier to detect extremely small amounts of charge, and the CCD storage electrodes can be small, both for this reason and because the electrode dimensions are not limited by any need to pack the sense amplifier into the repeat spacing of every one or two parallel rows.

The packing density in the CC16M1

was considerably increased by the introduction of two improvements to the SPS structure. One of these was a two-phase interlaced organization of the serial-parallel and parallel-serial transfers, by which the serial registers are loaded and unloaded twice during every clock period of the main parallel array. Alternate rows of the parallel array are serviced on one occasion, and the alternate remaining rows are serviced on the other. Each serial register is required to store only half as many bits as there are parallel rows, which permits a reduction in the pitch spacing of the rows, as well as halving the number of high-speed transfers.

An even more significant increase in density was obtained by adapting the basic principle of the multiplexed electrode-per-bit (ME/B) organization, [2]. As originally described, serial strings of data are allowed to occupy adjacent storage locations, rather than occupying every other location, as in two-phase clocking, or every third location, as in three-phase clocking. The bits in each string must then be moved one at a time, making use of a single vacant location. The operation requires the application of a sequence of individual clock waveforms (ripple clocks) to the set of storage electrodes, and although such a sequence can be generated by ring counters situated along the edges of the array, this tends to limit the pitch repeat spacing of the electrodes of the parallel array to the density which can be achieved in MOS ring counters. Also, although the repetition frequency of each clock is low, the clock edges must be fast, because the transfers must take place one at a time.

In the CC16M1 these problems were avoided by identifying an optimum organization, adapting the ripple clock principle to short data strings of only three bits each, each string being separated from the next by a vacant location. A set of four clocks is then sufficient to drive all the storage electrodes of the main parallel array, and each transfer gate could also be driven from the same waveform as applied to the following adjacent storage electrode, as is done in ordinary two-phase clocking. However, another set of four clocks is actually used for the transfer electrodes, since this enables "full-bucket" charge storage, increasing the amount of charge reaching each sense amplifier, with resulting increase in noise margins. All eight clock waveforms are distributed on buses along the edges of the storage arrays, from centrally-located clock drivers.

This combination of an interlaced serial-parallel-serial organization with a ripple-clock scheme characterized by the storage of short strings of data provides a high density of storage, for which the name "condensed serial-parallel-serial" or "CSPS" organization is suggested.

The choice of CSPS as the optimum organization for CCD memories is only justified if the array size is sufficiently large, otherwise the area occupied by the serial registers and sense amplifiers, and the power they consume, become disproportionately large. Fig. 2 illustrates the

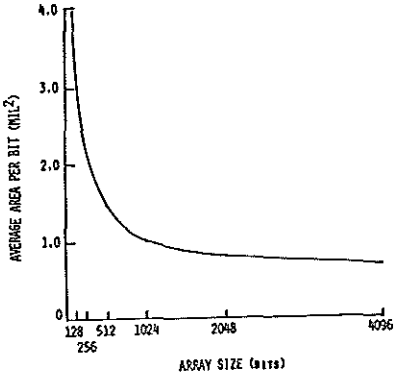

Fig. 2 - AVERAGE AREA PER BIT VS ARRAY SIZE IN CSPS ARRAYS

relationship between array size and average area per bit, for CSPS arrays having the same ripple spacing and multiplexing factors, cell dimensions, sense amplifier, and input and output buffers as used in the CC16M1. The averaging takes into account the areas occupied by the above circuit features, but does not

include the area occupied by the clock drivers, clock distribution buses, and other support circuitry, since these facilities can be shared by a number of arrays and therefore do not greatly influence the choice of size for each individual array. Fig. 3 illustrates the

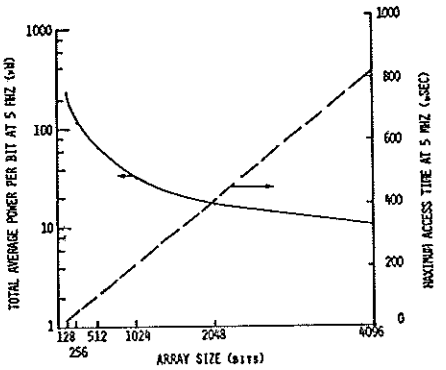

Fig. 3 - TOTAL AVERAGE POWER BIT AND MAXIMUM ACCESS TIME AT 5 MHZ RELATIVE TO ARRAY SIZE

corresponding effect of array size on average power consumption per bit. Once again, the contribution made by the shared support circuitry is not included. It may be seen that the CSPS organization is more favourable for larger arrays, containing at least 1 kilobit each.

In the geometrical layout which was chosen, the area of each storage location is 0.43 mil^2, including the area of the associated transfer electrode and isolation between adjacent parallel rows. Allowing for the absence of a data bit in every fourth location, the effective area occupied by each bit stored in the parallel part of the array is 0.57 mil^2. A multiplexing factor of 32 was chosen for the number of parallel registers in each array. Since two arrays are paralleled to make up each block, the effective multiplexing factor is 64 relative to the external data stream, so that at a data rate of 10 MHz the parallel clock frequency is only 156 KHz. Providing two paralleled arrays is considerably more favourable than providing a single array with 64 parallel paths and the same overall block size, since the number of serial transfers experienced by each bit is halved, and the serial transfer frequency is also halved. An array size of 2048 bits was chosen, providing an average area per bit of 0.65 mil^2 for each CCD array, equivalent to 0.8 mil^2/bit when the array support circuitry is included. Only a small advantage in area per bit and power per bit would have been achieved by choosing a larger array size than 2048 bits, while the access time would have lengthened in proportion. However, another reason for this choice was that it led to a quad-4K configuration on the 16K chip, which was convenient since a standard 16-pin package provides enough pins to permit individual input and output connections to each block, in addition to those required for the d.c. power rails, two external clocks and two control inputs.

Fig. 4 shows the organization of one

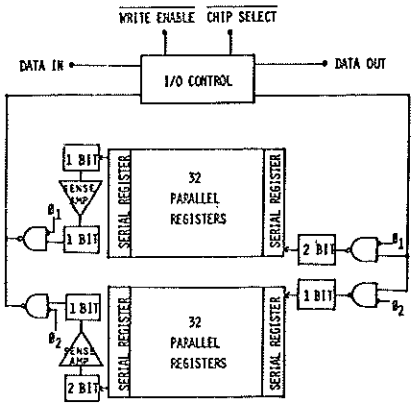

Fig. 4 - ONE 4096-BIT MEMORY BLOCK

4K block of memory, including various bit delays which lie outside the SPS arrays. These external bit delays were included for various reasons: to ensure proper control of the input charge, to transfer output charge from the arrays to the sense amplifiers, and to allow data

interleaving between the two arrays while enabling them to operate internally in the same phase with common clocks. The bit delay which follows each sense amplifier forms part of each data output buffer amplifier, and permits driving the output buffer at high speeds, while the sense amplifier is not loaded by excessive capacitance.

III. SUPPORT CIRCUITRY AND CHIP LAYOUT

A block schematic diagram of the on-chip clock generation circuitry is shown in Fig. 5. The external clocks ϕ_1, ϕ_2

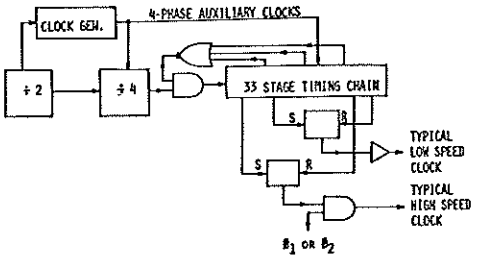

Fig. 5 - ON-CHIP CLOCK GENERATION CIRCUITRY

drive a divide-two counter, which in turn generates a set of four-phase clocks for operating the internal circuitry. A divide-4 circuit and feedback logic provide an impulse at the parallel clock rate into a 33-stage timing chain, which in turn provides impulses at the proper times to each of the clock waveform generators. Because of the three feedback tappings on the timing chain, the divide-4 circuit actually functions as a divide-16 counter. Each clock generator can be represented as a set/reset flip-flop, which is set by an impulse from one point on the timing chain, and reset from another point. Those clock generators which operate on the main parallel arrays must drive high capacitances, but are not required to switch at high speed. The clock generators which are associated with transfers into and out of the serial registers drive only small capacitances, but must generate fast edges to synchronize with the serial transfer processes. Therefore, the MOS transistors employed in the output driver of each clock generator are all of comparable size, with channel breadth about 30 times the channel length. These provide more than adequate driving capability for the 16K memory. The counters, timing chain and clock generators have been designed to clear rapidly any illegal states which might be picked up as a result of external clock irregularities. The circuitry also recognizes a "start-up mode" whenever d.c. power is applied while both external clocks are low, which does not occur during normal operation. Following recognition of the start-up mode, the internal clocks will commence with predictable phases relative to the first turn-on of either of the external clocks, and input data can be supplied within 43 or 45 clock cycles, depending on which external clock turns on first. The sense amplifier at the output of each 2K array is a differential-input flip-flop, which was specially designed for this task because the requirements differ from those in MOS RAMs.

The CC16M1 is provided with two control inputs, CHIP SELECT and WRITE ENABLE, and associated circuitry designed for operation up to 10 MHz data rate in three functional modes: RECIRCULATE, READ and RECIRCULATE, READ and WRITE. In READ and WRITE mode, the 4K blocks function as "straight-through" digital delay lines, which may be serially interconnected to form larger blocks with no penalty in maximum data rate. It was considered to be important that the CHIP SELECT would operate from ordinary 5-volt TTL signals, since these signals must be generated individually for all the memory devices which are OR-tied to common data buses. WRITE ENABLE requires a full clock level, but this is not a serious disadvantage because this signal can be supplied to a group of devices from a common driver. The physical layout of the memory arrays and support circuitry is shown to scale in Fig. 6

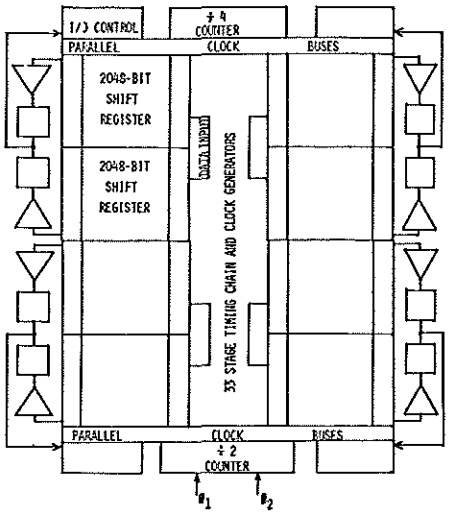

Fig. 6 - CHIP LAYOUT DIAGRAM OF THE 16,384 BIT CCD SHIFT REGISTER

APPLICATION IN SYSTEMS

The CC16M1, by requiring only two clocks, and handling continuous data streams at high speeds, has been designed for simplified application in memory systems. Suitable applications are likely to take advantage of the reduced access time compared to all fixed-head disks, with which a memory system built around the CC16M1 could, in volume production, be cost-competitive. Other applications, such as in video or radar signal processing, require large amounts of serial memory and in recognition of such applications the device was designed to permit serial interconnection between blocks on the same and on other chips, with no loss in maximum data rate. The 4x4K organization of the device is expected to be optimum for smaller systems. For large systems, there would be a preference for a decoded, block addressing (one out of four) scheme, so that only one bit of any word is stored on any chip. This would best be implemented using block addressing into a larger number of smaller blocks, which will be a logical extension of the present scheme. However, an alternative logical extension to chips of 64K size, with or without block-addressing, will be seen by many as being of greater importance in reducing system cost.

Acknowledgement

The authors wish to thank D.R. Colton, J.J. White and C.R. Robinson for valuable contributions. This work was supported in part by the Defence Research Board of Canada, under its Defence Industrial Research Program.

References

1. S.D. Rosenbaum and J.T. Caves, "8192-bit block-addressable CCD memory", IEEE JSSC, Vol. SC-10, Oct. 1975.

2. D.R. Collins, J.B. Barton, D.C. Buss, A.R. Kmetz and J.E. Schroeder, "CCD Memory Options", Tech. Digest, 1973 ISSCC, pp. 136-137, Feb. 1973.