DESIGN OF A MULTILEVEL DRAM WITH ADJUSTABLE CELL CAPACITY [†]

Yunan Xiang, Bruce F. Cockburn, Duncan G. Elliott

Department of Electrical and Computer Engineering
University of Alberta
Edmonton, AB T6G 2G7, Canada
E-mail: {ynxiang|cockburn|elliott}@ee.ualberta.ca

Abstract

A multilevel DRAM (MLDRAM^{TM}) increases the per-cell storage capacity over conventional DRAM by using more than two cell signal levels. The key challenge when designing an MLDRAM is to ensure reliable operation using the more closely spaced signal levels despite the presence of on-chip noise and the inevitable small variations in circuit parameters that occur in integrated circuit (IC) production. This paper describes a circuit architecture that implements an inherently balanced and robust MLDRAM scheme proposed by Birk et al. The design has an adjustable cell capacity that can be selected from among 1, 1.5, 2 and 2.5 bits per cell. Fractional bits arise when groups of two or more cells are considered together. Thus if each cell in a pair stores one of six possible levels, then each cell has a capacity of 2.5 bits. The test chip implementation of our design should facilitate experimental characterization of the proposed MLDRAM scheme.

1. INTRODUCTION

The per-area storage density of dynamic random-access memories (DRAMs) has been increased dramatically in recent years by reducing the physical size of the storage cells and by adopting complex three-dimensional cell capacitor structures [1]. By storing more than one bit per memory cell, multilevel DRAM exploits an additional dimension to increase the storage density without requiring a further reduction in the feature size. Figure 1 shows how two bits (MSB and LSB) can be encoded in a single cell by using four equally-spaced voltages from the ground potential *GND* up to the array supply voltage V_{dd}. The digital data stored in such a cell are sensed by comparing the cell signal with reference signals that lie midway between the expected cell signal levels. (At the sense amplifiers, the cell signals that must be sensed are attenuated towards the bitline precharge voltage because of charge sharing between the storage node capacitance C_c of the cell and the larger bitline capacitance C_b. In our design, for example, C_b is roughly 10 times the value of C_c.) In a 2-bits-per-cell MLDRAM, the fastest way of sensing the stored data is to simultaneously compare the three (suitably attenuated) reference signals with three copies of the attenuated cell signal.

The major challenge when designing a practical MLDRAM is to ensure sufficiently high reliability despite the reduced noise margins implied by the more closely spaced signal levels. This challenge must also be met in the face of the reduced power supply voltages required by deep sub-micron technologies. Reduced supply voltages may act to reduce some of the on-chip noise sources; however, the smaller geometries may also tend to increase the parasitic coupling between conductors and may make other formerly minor noise sources more significant. The introduction of low dielectric constant K insulators would reduce the strength of many sources of parasitic capacitive coupling. Also, high-K dielectrics could become available for the cell capacitors, and this would permit the cell signal to noise ratio to be raised. The net consequences of all these changes on the practicality of MLDRAM are therefore difficult to predict.

By using different numbers of cell signal levels, it is possible to design MLDRAMs that store other than two bits per cell. For example, eight signal levels could be used to store three bits per cell. By using three and six levels per cell and by considering cells in pairs, it is possible to store 1.5 and 2.5 bits per cell, respectively. Note that, in these two cases, the number of distinct signal level com-

---

[†] This work was supported by Micronet R & D, by MOSAID Technologies Inc., and by the Natural Sciences and Engineering Research Council of Canada.

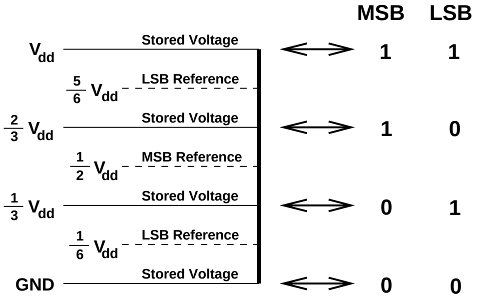

Figure 1: Positive Encoding for Two Bits Per Cell

Table 1: Cell Capacities Using Cell Pairs

Number of Levels, n	n^2	$\log_2 n^2$	$\lfloor \log_2 n^2 \rfloor$	Bits per cell
2	4	2	2	1
3	9	3.170	3	1.5
4	16	4	4	2
5	25	4.644	4	2
6	36	5.170	5	2.5
7	49	5.615	5	2.5
8	64	6	6	3

binations is 3×3 = 9 $\geq 2^3$ and 6×6 = 36 $\geq 2^5$. Table 1 shows how various cell capacities could be obtained by using pairs of cells with increasing numbers of signal levels. The second column shows the maximum number of distinct combinations that could be encoded in each cell pair; the third and fourth columns show the corresponding asymptotically achievable and practical (rounded down) numbers of information bits per cell pair. Finally, the fifth column shows the resulting number of bits per cell. Table 2 shows how one could generalize the number of cells in each group beyond two to obtain a variety of possible fractional per-cell storage capacities. If g denotes the number of cells per group, then the storage density per cell is given by $\frac{1}{g}\lfloor \log_2 n^g \rfloor$. For example, if $n = 5$ signal levels are used, then eight bits plus a parity bit could be stored together in a group of $g = 4$ cells with $2\frac{1}{4}$ bits stored in each cell.

Several MLDRAM schemes have been described in previous papers and some of them have been implemented in experimental chips [2, 3, 4, 5]. Recently, Birk et al. [6, 7] proposed an inherently robust MLDRAM that combines the speed advantages of [3] with the noise cancellation advantages of [8]. This paper describes an MLDRAM test chip design that is more flexible than Chan's 2-bit-per-cell implementation of Birk's MLDRAM [9]. The design can be operated either as a conventional 1-bit-per-cell DRAM or as a 1.5, 2 or 2.5-bits-per-cell MLDRAM. Regardless of the selected cell capacity, data storage and sensing follow the same mechanisms described in [6]. Thus the results of sensing parallel cell arrays are combined and decoded to recover the binary data. The ability to adjust the cell capacity will be useful for characterizing a test chip implementation and for experimentally determining the practical limits of one particularly promising MLDRAM technology. The adjustable cell capacity feature could also potentially be used as a form of graceful static redundancy in a production IC: if a particular memory is found in testing to have overly-leaky cells or overly-small sensing noise margins, then a lower-capacity working die could possibly still be sold by operating some or all of the cells at a reduced per-cell capacity.

Table 2: MLDRAM Cell Capacities

Group Size, g	Number of Signal Levels, n						
	2	3	4	5	6	7	8
1	1	1	2	2	2	2	3
2	1	$1\frac{1}{2}$	2	2	$2\frac{1}{2}$	$2\frac{1}{2}$	3
3	1	$1\frac{1}{3}$	2	2	$2\frac{1}{3}$	$2\frac{2}{3}$	3
4	1	$1\frac{1}{2}$	2	$2\frac{1}{4}$	$2\frac{1}{2}$	$2\frac{3}{4}$	3
5	1	$1\frac{2}{5}$	2	$2\frac{1}{5}$	$2\frac{2}{5}$	$2\frac{4}{5}$	3
6	1	$1\frac{1}{2}$	2	$2\frac{1}{6}$	$2\frac{1}{2}$	$2\frac{2}{3}$	3
7	1	$1\frac{4}{7}$	2	$2\frac{2}{7}$	$2\frac{4}{7}$	$2\frac{5}{7}$	3
8	1	$1\frac{1}{2}$	2	$2\frac{1}{4}$	$2\frac{1}{2}$	$2\frac{3}{4}$	3

2. MLDRAM ARCHITECTURE

One of the major advantages of the MLDRAM scheme introduced in [6] is that proven conventional DRAM blocks, such as the densely packed cell array core and the pitch-matched sense amplifiers and wordline drivers, are reused essentially unchanged. Most of the changes involve how the various blocks are interconnected. In addition, the number of sense amplifiers required for a fixed number of storage cells must increase as the number of reference levels increases. Special reference and generate cells, which are identical in layout to data storage cells, must also be included in the core. These extra cells require extra control circuitry in the periphery. Of course, the additional area overhead of MLDRAM reduces its achievable cell density advantage over DRAM. Given cell reliability requirements and a particular layout style, there will be an optimal number of signal levels that maximizes the per-area storage density and hence minimizes the per-bit cost. The conventional wisdom is that the optimal number of signal levels is only two. Our goal is to clarify the reliability trade-offs when using more than two cell signal levels in a theoretically robust MLDRAM scheme.

Figure 2 gives a simplified diagram of the MLDRAM test chip architecture. At the core are five cell array sections containing equal numbers of cells. Five sections, labelled A, . . . , E, are required for 2.5-bits-per-cell operation. Only three, two and one section(s) are required for 2, 1.5 and 1-bit-per-cell operation, respectively. Each sec-

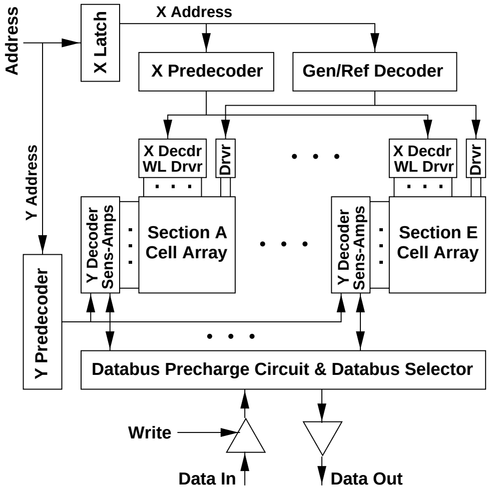

Figure 2: Simplified Test Chip Architecture

tion is surrounded, as in conventional DRAM cell arrays, with X and Y decoders, and with pitch-matched sense amplifiers in the Y direction. A difference from conventional DRAMs is that our sub-bitlines can be interconnected in the horizontal direction through pass transistor switches (not shown in Fig. 2) going from one section to the next. As well, groups of sub-bitlines can be interconnected within each section going in the vertical direction. When operating in 2.5-bits-per-cell mode, the five sections are used together in charge sharing operations, and hence five sub-bitlines are required in each vertical sub-bitline group. The 1.5 and 2-bits-per-cell modes use two and three sections, respectively, so the vertical groups in those modes include two and three sub-bitlines each. In all three MLDRAM modes, the appropriate voltages required for writing and sensing are created locally using the robust, fully-balanced charge sharing technique from [6]. The charge sharing operations require the sub-bitline signals to interact via the horizontal and vertical switches.

The cell array in our simulated design includes 24 wordlines in the X-address direction and 250 sub-bitlines in the Y-direction. Figure 3 shows the schematic of one *folded bitline* composed of two sub-bitlines. The *true sub-bitline* is connected to the positive input of a conventional DRAM sense amplifier (denoted by S-A) while the *complement sub-bitline* is connected to the negative input. Five true (complement) sub-bitlines can be connected horizontally, end-to-end, via the *horizontal switches* controlled by SW0 (SW1) signals. Only four sets of horizontal switches are required between the five sections. A *full-length horizontal bitline* is formed by connecting together corresponding sub-bitlines from each section going in the horizontal direction. Within each section, groups of five adjacent true (complement) sub-bitlines can be connected at one end in the vertical direction by closing the *vertical switches* controlled by REF0 (REF1) signals. A *full-length vertical bitline* is thus formed when the four corresponding vertical switches are closed.

Storage cells are of the conventional one transistor, one capacitor design [1]. Our access transistors are of type PMOS. In our simulation study, we assumed that the test chip would be implemented in 0.18-μm CMOS and that the storage capacitor would be formed using the gate capacitance of a PMOS transistor. The source and drain terminals of all of these PMOS transistors are connected together (and to the n-well) and tied externally to +2.5V, well above the 1.8 V array supply voltage. This ensures that the channels of the PMOS transistors are strongly inverted to create a relatively large, voltage-insensitive gate-to-channel cell storage capacitance. The nominal cell capacitance in our simulation model is 50 femtofarads, which is larger than in most commercial commodity DRAMs.

Four of the 24 wordlines in each section are not connected to data-storing cells. Two of these wordlines, called *reference wordlines*, go to *reference cells* at some sub-bitlines and empty cell locations at other sub-bitlines; the other two special wordlines, called *generate wordlines*, go to *generate cells* and “empty” cell locations. When the memory is in storage stand-by mode, the reference cells contain signals corresponding to the unattenuated reference voltages that lie mid-way between the allowed data signal levels. For example, in 2.5-bits-per-cell mode, the reference cells in sections A, B, C, D and E will normally contain $0.1V_{dd}$, $0.3V_{dd}$, $0.5V_{dd}$, $0.7V_{dd}$ and $0.9V_{dd}$, respectively. The generate cells are provided to ensure that the sub-bitline capacitances can be made exactly equal when a cell is being accessed on one sub-bitline. This feature permits the cancellation of unwanted voltage offsets that would otherwise arise during charge sharing operations. The gate-to-drain capacitances of the access transistors of the generate cells are also used to cancel out noise injected onto the sub-bitlines by wordline signal transitions [6, 7]. The folded bitline in Fig. 3 is formed from *reference sub-bitlines* because the reference cells are present while the generate cells are absent (an empty cell location is formed from a cell by omitting the cell access transistor). In the case of *generate bitlines*, the generate cells are present while the reference cells are empty.

Cell data are written into the test chip over a serial data input bus and read out over a serial data output bus. Memory accesses are sequenced under the control of the signals listed at the top of Fig. 3 as well as other signals not shown (such as the sense amplifier isolation and enable signals). The control signals are assumed to be controlled by an external tester to simplify the test chip design and to provide a great degree of flexibility during characterization. The control signal sequence is detailed in [6, 7, 9] and in the next section.

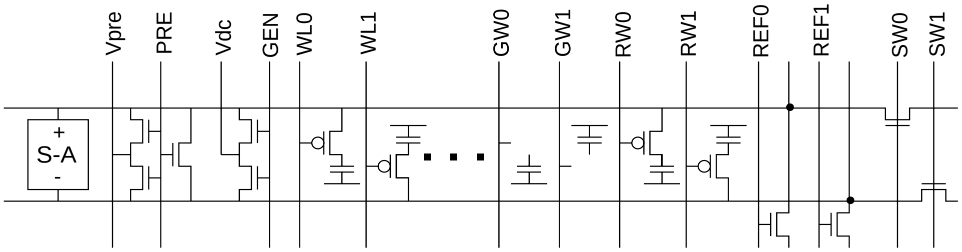

Figure 3: Sub-bitline Schematic

3. SIMULATION RESULTS

Figures 4, 5, 6 and 7 give the simulated sub-bitline signals for the 1, 1.5, 2 and 2.5-bits-per-cell operating modes, respectively. In each figure, all of the cell and reference sub-bitline signals are shown superimposed on top of each other to more clearly show the cell signal and reference signal spacings.

Consider the 2.5-bits-per-cell waveforms in Fig. 7. Two cells must be used together in this mode, with each cell storing one of six possible signal levels. Assume that the first cell in such a cell pair lies on a true sub-bitline in section A. In 2.5-bits-per-cell mode we will be using the sense-amplifiers in all five sections. Initially, all sub-bitlines are precharged to the array mid-voltage of $V_{pre} = \frac{1}{2}V_{dd}$ = 0.9 V. The sense amplifiers are powered up at time t = 300 ns and the folded sub-bitline signals are driven to the power supply rails with random states. The six possible cell signal levels are stored using a five-bit unary code (also called a thermometer code) with codewords 00000, 10000, 11000, 11100, 11110 and 11111. From time t = 300 ns to 840 ns, the states of the one addressed sense amplifier (SA) in each section are overwritten in succession to load in the five code bits. Meanwhile, the wordline of the addressed cell is held asserted in section A. To ensure equal sub-bitline capacitances during multilevel signal generation, the reference and generate signals are all de-asserted in the addressed section (A in this scenario) and asserted in the other sections (B, C, D and E). All sub-bitlines are kept isolated until the horizontal switches for the true sub-bitlines, controlled by SW0 signals, are closed at time 930 ns. The resulting charge sharing creates one of the six equally-spaced cell signal levels, which are clearly visible on the superimposed plots in Fig. 7 from t = 930 ns to 990 ns. At time t = 960 ns the wordline of the addressed cell is deasserted to trap the multilevel cell signal on the storage capacitance. Note that this positive-going wordline transition injects a small positive offset onto the sub-bitlines (and thus also into the cells).

During the interval from t = 1080 ns to 1140 ns, control signal GEN is asserted to cause the true and complement sub-bitlines to be precharged variously to 0 V, 0.9 V, and 1.8 V (depending on the particular voltage connected to V_{dc} for that folded bitline) to prepare for the creation of the five required reference signals. In section A, four sub-bitlines in each group of five are precharged to *GND* while one sub-bitline is precharged to V_{pre}. The resulting voltage after charge sharing is $0.1V_{dd}$, that it, the smallest required reference voltage. In section B, three sub-bitlines are precharged to *GND* while two other sub-bitlines are precharged to V_{dd} and V_{pre}; when charge shared they will produce the $0.3V_{dd}$ reference. In section C, three sub-bitlines are precharged to V_{pre} while the other two sub-bitlines are connected to V_{dd} and *GND*; when charge shared they produce the $0.5V_{dd}$ reference. In section D, three sub-bitlines are precharged to V_{dd} while the other two sub-bitlines are precharged to *GND* and V_{pre}; they produce the $0.7V_{dd}$ reference. Finally in section E, four sub-bitlines are precharged to V_{dd} while one sub-bitline is precharged to V_{pre} to produce the $0.9V_{dd}$ reference. The sub-bitlines are then disconnected from their respective V_{dc} signals by de-asserting GEN. Charge sharing is initiated by asserting the REF0 and REF1 signals, which control the vertical switches. The full-length vertical bitlines in the five sections will then be at the five required reference potentials. The reference signals are then trapped in the reference cells by de-asserting RW0 and RW1 in all five sections. This last step injects a small positive offset that will closely match the offset injected into the data-storing cells.

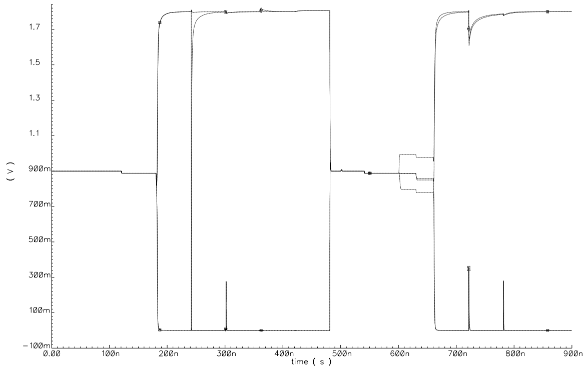

Figure 4: Bitline and Reference Signals in 1-bit-per-cell DRAM

From time t = 1320 ns to 1440 ns all of the sub-bitlines are connected together via the horizontal and vertical switches and held at the precharge voltage while the

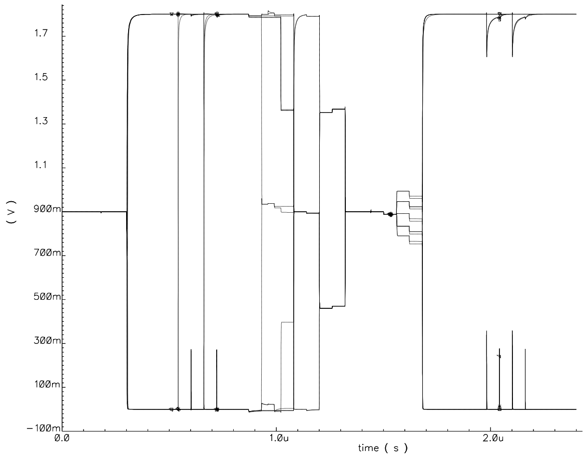

Figure 5: Bitline and Reference Signals in 1.5-bits-per-cell MLDRAM

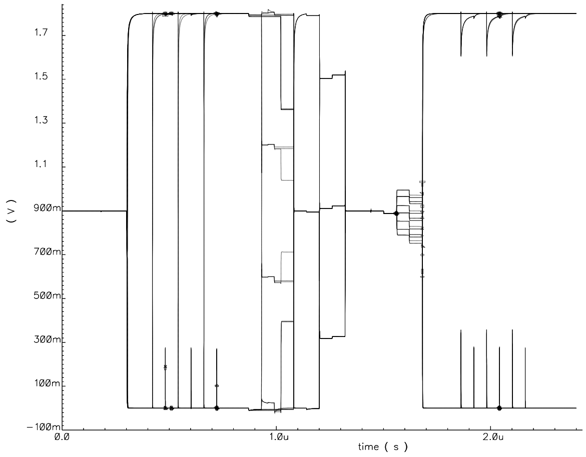

Figure 6: Bitline and Reference Signals in 2-bits-per-cell MLDRAM

cells are in storage mode. Cell sensing is a simple and fast operation. First, some of the sub-bitline switches are opened at time t = 1440 ns. Since the addressed cell is on a true sub-bitline, all SW0's are kept asserted while all SW1's are de-asserted. In the vertical direction, however, the REF1's are held asserted while the REF0's are de-asserted. The sub-bitlines are still precharged to V_{pre} until t = 1560 ns to eliminate the small negative offset injected otherwise by the switches when they are opened. Then at time t = 1560 ns the addressed wordline is asserted to dump the signal of the addressed cell onto its full-length horizontal (true) bitline; simultaneously, the reference wordlines RW1 in sections A, . . . , E are asserted to dump their reference signals onto the full-length vertical (complement) bitlines. These control signal transitions will inject small negative offsets onto the sub-bitlines, but these offsets will be matched and the hence they will be effectively rejected as common-mode components by the differential-mode sense amplifiers.

At time 1620 ns the sub-bitlines are all fully isolated by de-asserting the SW0's and REF1's; each folded bitline in sections A, . . . , E now has a copy of the cell signal on the true sub-bitline and the appropriate reference signal on the complement sub-bitline. The negative-going transitions of the SW0's and REF1's also inject a significant negative noise offset onto the relatively weak sub-bitline signals, an offset that is clearly visible in Figs. 4-7; however, the matched arrangement of the two sub-bitlines in each pair ensures that the noise signals will be common-mode and thus effectively rejected by the SAs. The SAs are then powered up and the data are sensed, amplified, and driven back onto the sub-bitlines. In our simulations the states of the five addressed SAs are read out serially over the databus from time t = 1740 to 2280. The 2.5 bits recovered from the addressed cell are at that time encoded using the five-bit unary code. A second five-bit codeword must be retrieved from a second cell (possibly at the same time that the first cell is accessed) before the original five bits of data can be obtained through a simple decoding step.

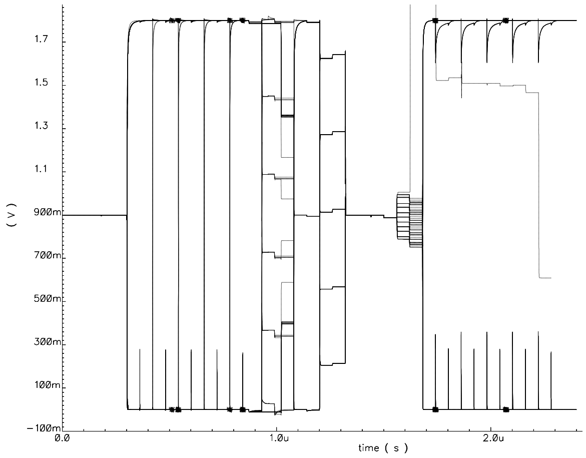

Figure 7: Bitline and Reference Signals in 2.5-bits-per-cell MLDRAM

It should be pointed out that while a cell at one Y-address in one section is being accessed, all other cells with the same X-address in the same section as the addressed cell are also being accessed in parallel using the same reference signals, but on different folded bitlines in the same section. If two independent data buses are provided into the SAs, then the two cell accesses required in 1.5 and 2.5-bit-per-cell mode can be performed at the same time in parallel.

The signals in an MLDRAM cell, as in a conventional DRAM cell, weaken over time as a result of charge leakage from the cell storage capacitance. The problem is more serious in an MLDRAM because of the more closely spaced cell signal levels. Therefore to restore the cell signals and avoid data loss, all cells in the MLDRAM must be refreshed by accessing, sufficiently often, cells on each X-address. A disadvantage of MLDRAM versus conventional DRAM is that the refresh rate is going to be higher, corresponding to greater standby power consumption. For some applications, however, the data access patterns of the application will by themselves be sufficient to fully refresh all cells without requiring many accesses devoted exclusively to refreshing.

The 1.5 and 2-bits-per-cell control sequences are similar to that of the 2.5-bits-per-cell mode. In 1.5-bits-per-cell mode, two of the sections (C and D) are used to create the two required reference signals, $\frac{1}{4}V_{dd}$ and $\frac{3}{4}V_{dd}$. The addressed SAs in sections C and D are used to recover one of the three possible unary codes 00, 10 and 11. Sensing a second cell yields a second two-bit unary code. Only eight of the nine possible combinations of two two-bit codes are required to represent the eight three-bit data words 000, $\ldots$, 111. The control sequence used to produce the plots in Fig. 5 was modified slightly from the 2.5-bit-per-cell sequence: just before powering up the sense amplifiers, both the cell and reference signals were charge shared over full-length horizontal and vertical bitlines, respectively, instead of only two connected subbitlines each. This was done to simulate the situation of a 1.5-bit-per-cell chip that uses the same bitline length as in the actual 2.5-bit-per-cell test chip.

In 2-bits-per-cell mode, the three middle sections are used to create the three required reference signals ($\frac{1}{6}V_{dd}$, $\frac{1}{2}V_{dd}$ and $\frac{5}{6}V_{dd}$), and the SAs in all three sections are used to recover one of the four possible 3-bit-long unary codes from one cell. The control sequence used to produce the plot in Fig. 6 was also modified to charge share the cell and reference signals over full-length bitlines. The three-bit unary code recovered by sensing must be decoded in a simple operation into the corresponding two-bit data combination.

The 1-bit-per-cell plot in Fig. 4 was also made using a control sequence that charge shared the cell signals over full-length bitlines. Note that the potential difference from the greatest to least cell signal is about the same for all four operating modes plotted in Figs. 4-7. Thus it will be possible to use the one test chip to directly compare the four operating modes with each other during characterization experiments.

4. CONCLUSIONS

This paper described the features and operation of a multilevel DRAM architecture with adjustable cell capacity. The design has been verified to function correctly in circuit simulation; however, only by characterizing a test chip in silicon will the feasibility of storing six signal levels in a cell, or even only four or three levels, be convincingly demonstrated. We are implementing a modified version of the MLDRAM described above in a commercial 0.18-μm CMOS process. The test chip will have 15000 cells and thus will have a maximum storage capacity of 22500 bits when operated in 2.5-bits-per-cell mode. Demonstrating correct operation in silicon is still a long way from proving the practicality of using MLDRAM in production ICs. However, the adjustable cell capacity of our test chip should provide the ideal means for experimentally exploring the feasibility and reliability issues, which are necessary first steps on the way to commercial introduction.

5. REFERENCES

[1] B. Keeth and R. J. Baker, *DRAM Circuit Design: A Tutorial*, IEEE Press, 2000.

[2] M. Aoki et al., “A 16-Level/Cell Dynamic Memory”, *IEEE J. Solid-State Circuits*, vol. 22, no. 2, pp. 297-299, April 1987.

[3] T. Furuyama et al., “An Experimental 2-bit/Cell Storage DRAM for Macrocell or Memory-on-Logic Application”, *IEEE J. Solid-State Circuits*, vol. 24, no. 2, pp. 388-393, April 1989.

[4] E. Lee and P. G. Gulak, “Current-Mode Multivalued Dynamic MOS Memory with Error Correction”, *IEE Electronics Letters*, v. 28, May 1992, pp. 1067-1069.

[5] T. Okuda et al., “A Four-Level Storage 4-Gb DRAM”, *IEEE J. Solid-State Circuits*, vol. 32, no. 11, pp. 1743-1747, Nov. 1997.

[6] G. Birk et al., “A Comparative Simulation Study of Four Multilevel DRAMs”, *1999 IEEE Int. Workshop on Memory Tech., Design and Testing*, Aug. 9-10, San Jose, CA, U.S.A., pp. 102-109.

[7] G. Birk, “Evaluation, Design, and Implementation of Multilevel DRAM”, M.Sc. thesis, Dept. of Elect. and Comp. Eng., U. Alberta, Aug. 1999.

[8] P. Gillingham, “A Sense and Restore Technique for Multilevel DRAM”, *IEEE Trans. Circuits and Systems-II: Analog and Digital Signal Processing*, vol. 43, no. 7, pp. 483-486, July 1996.

[9] A. Chan, “Design and Implementation of a Multilevel DRAM”, M.Eng. report, Dept. of Elect. and Comp. Eng., U. Alberta, Apr. 2000.