A 6-b DAC and Analog DRAM for a Maskless Lithography Interface in 90 nm CMOS

David Fang[1,2], Ryan Roberts[1], Borivoje Nikolic[1]
[1]University of California, Berkeley, USA, [2]now with Intel Corp, Hillsboro, OR, USA

***Abstract*- A parallel, 12µm-pitch, low-power 6-b segmented digital-to-analog converter (DAC) array drives an array of 3µm × 3µm analog DRAM cells in a 2.5/1V 90 nm CMOS process, with an application in maskless lithography. An innovative self-calibrating compensation circuit limits the effect of charge leakage and capacitive process mismatch to less than 0.5LSB over 100ms of data hold time. A 2mm x 2mm test chip implements a mixed-signal interface with 32 DACs driving four 32 × 256 analog DRAM arrays.**

I Introduction

As minimum feature sizes in CMOS technology scale, the cost of critical dimension masks dramatically increases. Although the masks are still a small fraction of the overall cost of the large-volume chip development, they present a significant impediment to low-volume ASICs, and increase the cost of prototyping and process development. To alleviate the cost of low volume fabrication, alternatives to mask-based optical lithography have been pursued. Various approaches to maskless lithography have been investigated in the past: e-beam, micro-machined mirror projection, and nano-jet printing. Recent updates of the International Technology Roadmap for Semiconductors (ITRS) predict existence of maskless lithography as an alternative to conventional mask-cased optical lithography in sub-45nm technology nodes [1].

A promising technology for maskless lithography systems is based on spatial light modulators (SLM). In this approach, micro-mirror based system modulates the position of individual mirrors in the array to create the image based on its corresponding diffraction pattern. The advantage of this system is in its compatibility with optical lithography tools, where a fixed mask would be replaced with a programmable one [2]. A feasibility of this approach has been demonstrated on existing mask-writing tools that use the same technology [3]. The mask writer system is based on a 512 × 2048 array of 16µm × 16µm MEMS mirrors. The mirrors are tilted in 256 different steps through analog voltages provided by an external array of 128 DACs. Although acceptable for mask writing, the speed of this system is several orders of magnitude below the requirements for maskless lithography.

To upgrade the SLM-based mask writing process for use in maskless lithography, the throughput needs to be dramatically increased, [4]. To achieve a 1nm edge placement using 22nm pixels targeted for a 45nm process technology, 5-bit grayscale data per pixel representation is needed, [5]. This results in a total of 500Tb of information on a 300mm wafer. Prototyping and low-volume ASIC production can be made economically viable with 1-6 wafer layers per hour [2], requiring data throughputs of up to 1.6Tb/s. Besides overcoming the challenges of handling large volumes of data, the maskless lithography system has to be able to integrate a large array of small-dimension nanomirrors. A key challenge of building this custom silicon chip is the need to bring a large number of DACs on chip and to store the analog control voltages underneath the mirrors. In this paper, we present a mixed-signal interface to the mirror array, designed in 2.5V/1V 90nm CMOS technology satisfying the requirements for exposing up to 3 wafers/hour.

II System Architecture

Figure 1 illustrates a maskless lithography system based on an SLM with a nanomirror array to generate the mask patterns. In this approach, the reconfigurable mask forms a new pattern between consecutive light flashes. The mirror-controlling voltages must be loaded between the flashes and stored in an analog memory array. The memory array connects to the electrostatically controlled mirrors to adjust their individual positions. The 3µm×3µm polysilicon mirrors positioned to a minimum of 32 levels can achieve 22nm feature sizes with a 1nm edge placement through a 140× optical reduction [6]. The mirrors can be either piston or tilting type [7-8], and the size of array requires the integration of driving electronics onto the SLM chip. Our design demonstrates the feasibility of an analog interface to the nanomirror array, capable of storing and maintaining precise voltage based position data between the laser flashes.

To achieve the desired throughput of 3wafers/hour requires 12 million pixels to be exposed in each flash. The mirrors would be placed in a 6,144×2,048 array, driven by parallel DACs from each side, Fig. 2. To accommodate for various errors, the DAC resolution would be 6 bits, and has to be able to load at least one row of 1,024 mirrors in 100µs between the flashes.

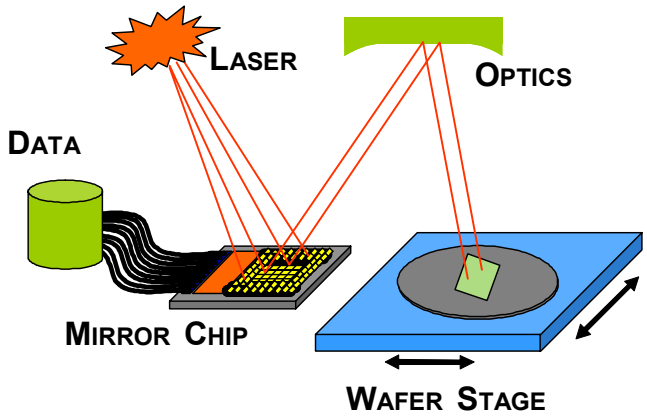


Fig. 1. SLM-based maskless lithography system.

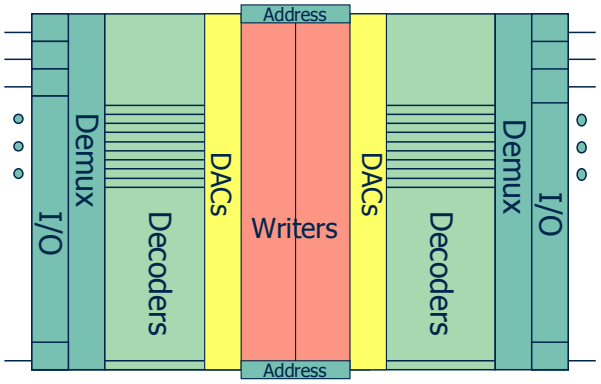


Fig. 2. Concept of the mirror-interface chip.

The main challenge of this design is in the DACs that need to fit into the tight mirror pitch. Being able to load multiple rows of data between the flashes would reduce the DAC pitch requirement. The analog mirror array designed in a deeply scaled technology has to be able to maintain the charge fidelity between the two laser flashes. High-speed parallel I/O interfaces can provide the throughputs of 800Gb/s [9]. Lossless data compression could reduce the data volumes by a factor of 10 [5]; with the decompression implemented on chip, [10-11], the data throughput can be proportionally increased.

III Circuit Design

The test chip contains four 32×256 analog DRAM arrays in an isolated N-well to reduce injected substrate noise into the cells. Figure 3 is the layout view of an analog DRAM cell that would drive the nanomirror array. A half-sized balance transistor is driven complementary to the active word lines to absorb the charge injection onto the DRAM cell to minimize noise.

Figure 4a shows the first design model of the DRAM array, which uses a bottom MOS gate capacitor. This achieves the highest capacitive density by using diffusion-only connections within the cell and benefits from the shielding provided by the grounded polysilicon gate. However, the capacitance of a bottom plate DRAM has a large nonlinear voltage dependency whose final voltage can be affected by junction leakage through its large diffusion surface area. Figure 4b shows a DRAM using the top MOS gate. The capacitance is linear but requires using low metal layers to shield the cell to reduce noise coupling from the bitline and wordlines. The usage of low metal directly over the gates increases the degree of MOSFET mismatch [12] even if each DRAM cell has identical metal coverage.

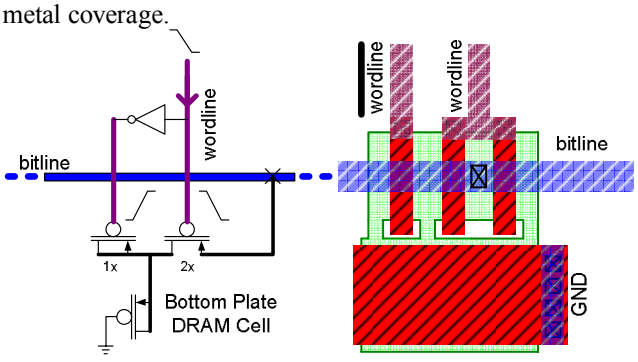


Fig. 3. Balanced access transistor for DRAM writes.

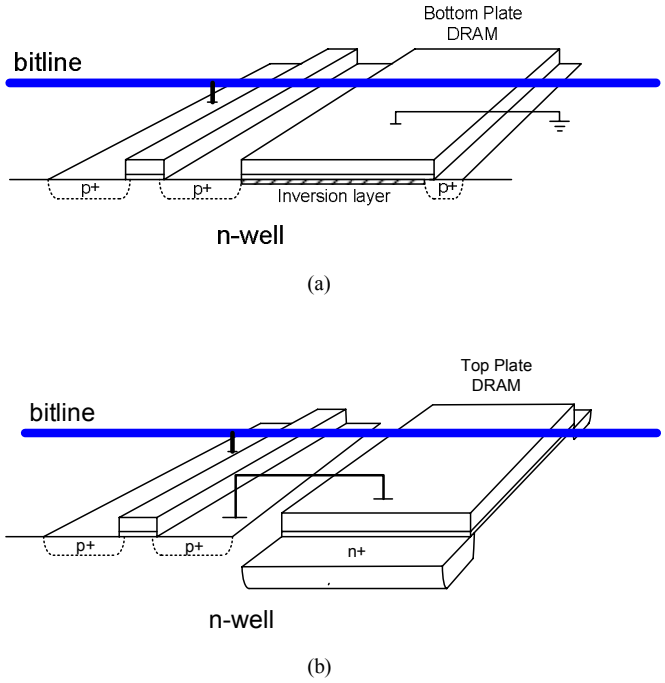


Fig 4. (a) Bottom plate DRAM cell, (b) top plate DRAM cell.

In the time between the flashes, significant charge can escape the DRAM cell through substrate, gate oxide, or source-drain tunneling. Gate oxide tunneling occurs in deep submicron processes when the oxide thickness becomes so thin such that electrons can tunnel from the gate into the substrate. Due to the long retention time needed due to limitations in the slower optical flash rate, this source of leakage over 100ms is capable of discharging the DRAM cell. This design uses a thick oxide layer over the DRAM to mitigate this effect. Source drain leakage across the access transistors can reduce by a combination using non-minimum gate lengths along with a 100mV gate overdrive. The interface also does not drive rail-to-rail to prevent a full VDD drop across the access transistor. This results in body biasing to increase the threshold voltage.

Since the nano-mirrors are electrostatically controlled, the final absolute voltage must be consistent across the entire DRAM array independent of which row and how early or late the cell was written. The interface must be able to measure and compensate this loss of charge. Figure 6 shows the mixed signal interface used to drive the analog DRAM.

A 3-3 segmented DAC with cascode biasing drives the memory array. Uniquely to this design, the DAC is height constrained to four DRAM rows, resulting in a very wide aspect ratio that limits the maximum resolution of the designed DAC. Also unique to this design is the ability calibrate for temporal mismatch due to each cell losing a different amount of charge depending on when that cell was written. A 3-3 segmented approach is used to minimize area for the resolution target.

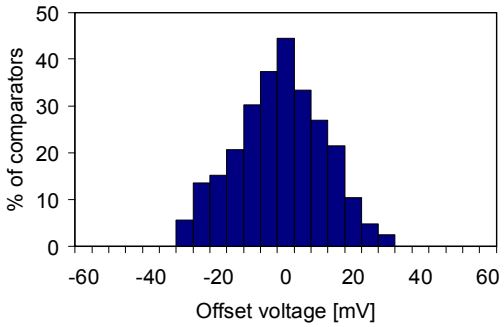
A reference voltage controls the output voltage range. The system uses a two-phase iterative input offset cancellation sequence to calibrate the DACs. In the first phase, the reference voltage directly connects to the feedback amplifier. The final output voltage would deviate from that reference voltage due to resistive variations in the load and transistor mismatch.

The total error is sampled by connecting DAC output when all inputs are high to the same reference voltage across a capacitor. In the second phase, the reference voltage connects to the other side of the capacitor and the correct bias voltage is pushed onto the DRAM used to bias the amplifier. Due to charge sharing between the DRAM and the capacitor storing the error, the sequence iterates multiple times to allow the system to converge to the correct bias voltage.

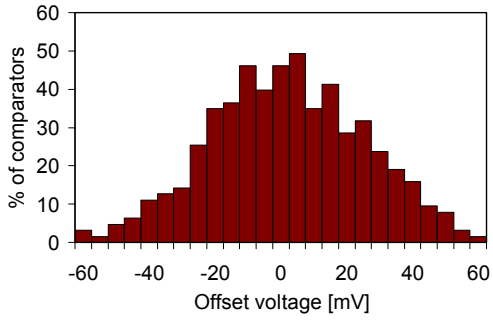
A replica DRAM cell, placed in the bias, monitors the rate of leakage for the die's current operating conditions and adjusts the DAC's internal reference voltage. This is necessary when the interface writes the same digital input to separate DRAM cells at two different times. The earlier written DRAM will be subject to leakage and will hold a different voltage than the DRAM written much later.

The replica DRAM is designed to leak similarly and scale the output of the DAC linearly due to the negative feedback configuration. When the DAC writes the last DRAM cell, the bias voltage would have changed to equal the voltage loss expected on a DRAM in the array. This will cause the output voltage of the DAC for the same digital code to equal to the voltage of any earlier written DRAM with added charge leakage. This methodology allows two cells written apart in time to hold equal voltage even if one cell has been affected by leakage and noise for a longer period than the other.

In maskless lithography applications, the array would be read by observing the diffraction from a laser flash. To test the functionality of the interface, a bank of on-chip voltage comparators [13] is implemented to measure voltages on the DRAM. Figure 5 shows the measured input offset distribution of the comparators for two different common mode voltages.



(a)



(b)

Fig 5: Input offset distribution of the comparators at (a) 1.1V and (b) 2.3V.

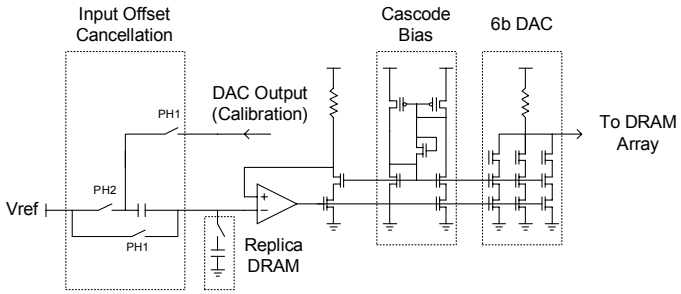


Fig. 6. 6-b segmented DAC with input offset cancellation and leakage compensation.

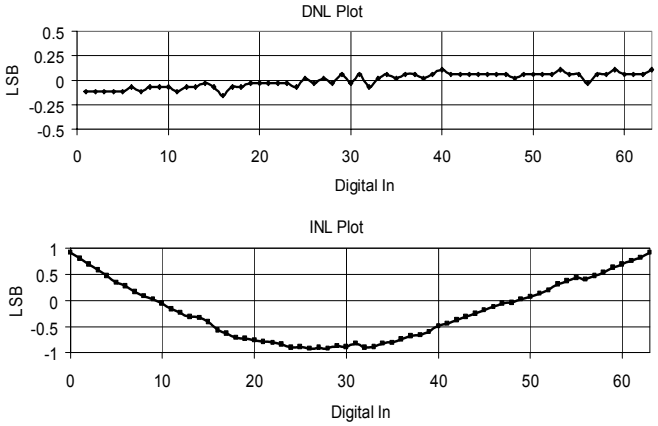


Fig. 7: Measured DNL and INL.

IV Test Results

Figure 7 shows measured DNL and INL of the 6-bit DAC. DNL is less than 0.2LSB and the INL is less than +/-1LSB limited by the large aspect ratio of the DAC (12µm x 320µm). Common centroid layout and dummy cells on the outer current sources minimizes transistor mismatch due to process gradients within the die. Since the DAC is monotonic, its INL can be lumped to the mirror response, and compensated in the input data stream without degradation of edge placement performance or defocus tolerance. The DAC speed is 50MHz limited by the RC constant of the resistive load and bitline capacitance.

The two DRAM cells under test are written apart in time, to compare the final voltages for a cell written early in the sequence and the other written much later. A chain of 64 comparators connects to multiple cell pairs across the entire DRAM array and across rows to ensure the robustness of the design. An identical DAC and corresponding DRAM array that does not contain the leakage compensated bias has been implemented as a baseline for comparison.

The test sequence first calibrates all the DACs simultaneously using the iterative input cancellation method. It then sweeps through various digital input offsets to the DRAMs under test. The ideal output should be identical to the input offset distribution directly measured from the comparators. In the ideal case where leakage was not present, two DRAM cells written with the same digital code should trigger roughly half of the comparators and any offset should mirror the offset distribution function of the on chip comparators.

The test chip reveals that for the uncompensated mixed signal interface, two adjacent DRAMs written 100ms apart with