Reinventing Computing

Dynamic Embedded Memories (Abstract)

Derrick Chen
kuang@ai.mit.edu
NE43-770

MIT Artificial Intelligence Laboratory
545 Technology Square, Cambridge, MA 02139
April 30, 1995

Motivated by the consumer electronics market, a great body of research has been accomplished to date contributing to our understanding of large-scale memory architectures. In comparison, with the exception of cache designs, little work has been publicized covering small-scale memories. The rapid growth of the FPGA market over the past decade, however, has created new incentives for better understanding the properties of embedded memories. These memories are the core elements which make reconfigurable logic possible.

Embedded memories are not simply scaled-down versions of megabit architectures; on the contrary, they need to be designed to satisfy a different set of requirements. Because of their tight integration with logic elements on the same die, embedded memories are likely to be clocked and multi-ported for rapid, parallel access to stored information. They typically have wide datapaths of 32, 64, or more bits, rather than the byte arrangements found in megabit memories. For embedded memory design, density and speed are the primary concerns, with power consumption a secondary issue. In comparison, density and power are the driving factors in large-scale memory designs, with speed being of secondary importance.

To help address the gap in our understanding of small-scale designs, I present in this paper a selection of embedded memories focusing on dynamic structures implemented in CMOS technology. In particular, the sizings, timings, and layout considerations which I discuss are all in reference to the HP26 0.8u 3-layer metal N-well CMOS process. I present one memory, based on a 3T cell, which has been incorporated into a prototype DPGA chip engineered by the MIT Transit Project [TEC+95]. [BCE+94] presents more in-depth documentation on the design and implementation of the prototype. The other memory structures discussed in this paper are being considered for possible use in future versions of the DPGA and MATRIX arrays.

The common characteristic that all the memory designs in this paper share is the heavy use of N devices. Since embedded memories are designed for speed, it is desirable to focus on N devices with their higher carrier mobility, and to minimize as much as possible the use of slower P devices. From a manufacturing point of view, the heavy use of one type of device is also significant because it bypasses the need for implanting into a common silicon substrate numerous distinct wells with large spacing requirements for both N and P diffusions, thereby allowing higher cell densities. Additionally, for the sake of reliability, using only one type of device ensures that there can be no failures due to CMOS latchup.

Conventional small-scale memories have focused almost exclusively on static designs as the most desirable density and speed solution. Dynamic memories have been neglected because of the perception that refresh is a difficult problem to handle. My work shows that dynamic embedded memories can match the speed of static devices, while allowing for higher densities. In particular, memory organizations based on the 3T DRAM cell being used in the DPGA prototype can be half the size of a conventional 6T SRAM design, but still maintain the high-performance characteristics shown below. In this paper, I also discuss several simple techniques for managing the refresh requirements associated with dynamic memory cells.

Acknowledgments: This work is conducted under the direction of Principal Research Scientist Thomas Knight Jr. This research is supported by the Advanced Research Projects Agency of the Department of Defense under Rome Labs contract number F30602-94-C-0252.

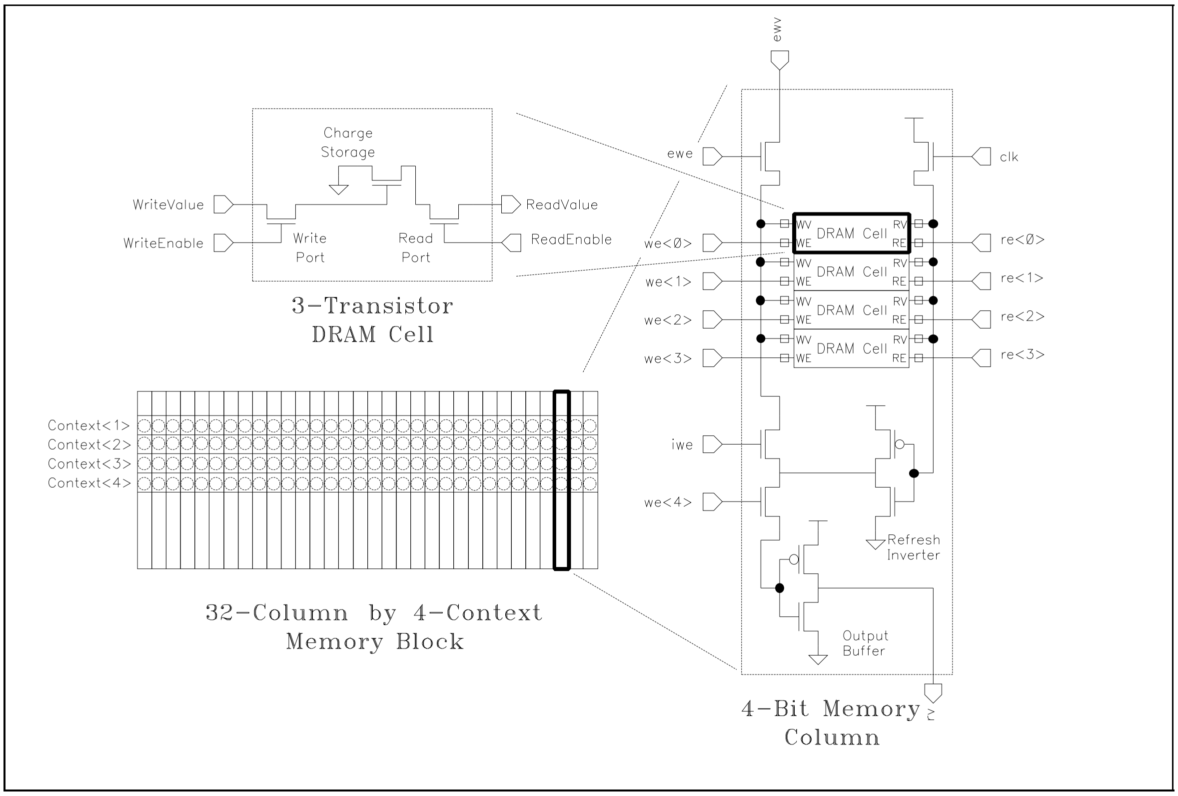

Figure 1: 3T DRAM organization used in DPGA chip

Read Delays	Worst Case	Nominal
High->Low (CLK to valid 0.45V output)	4.07 ns	2.13 ns
Low->High (CLK to valid 4.05V output)	3.55 ns	1.76 ns

Programming Delays	Worst Case	Nominal
High->Low (CLK to valid 0.45V output)	7.14 ns	4.40 ns
Low->High (CLK to valid 3.00V output)	7.23 ns	4.19 ns

References

[BCE+94] Jeremy Brown, Derrick Chen, Ian Eslick, Edward Tau, and André DeHon. A 1μ CMOS Dynamically Programmable Gate Array. Transit Note 112, MIT Artificial Intelligence Laboratory, November 1994. Anonymous FTP `transit.ai.mit.edu:transit-notes/tn112.ps.Z`.

[TEC+95] Edward Tau, Ian Eslick, Derrick Chen, Jeremy Brown, and André DeHon. A First Generation DPGA Implementation. Transit Note 114, MIT Artificial Intelligence Laboratory, January 1995. Anonymous FTP `transit.ai.mit.edu:transit-notes/tn114.ps.Z`.