ASSESSING MERGED DRAM/LOGIC TECHNOLOGY

Yong-Bin Kim* Tom Chen**
*Engineering Systems Lab. MS–55, Fort Collins, CO 80525
ybk@hpesybk.fc.hp.com
**Department of Electrical Engineering, Colorado State Univ. Fort Collins, CO 80523
chen@longs.lance.colostate.edu

ABSTRACT

This paper describes the impact of DRAM process on the logic circuit performance of Memory/Logic Merged Integrated Circuit and the alternative circuit design technology to offset the performance penalty. Three state-of-the-art logic processes(0.5 μm, 0.6μm, and 0.8μm) and two state-of-the-art DRAM(64Mb and 256Mb) processes have been selected for the study. The simulation results show that the logic circuit performance is degraded about 22% on DRAM process including the increased interconnect delay due to less interconnect layers available in DRAM process. The silicon area increased up to 80% depending on the number of net and components when implementing a logic circuit in a DRAM process.

1 INTRODUCTION

Since the actual memory bandwidth of a system is limited by off-chip interconnects, the memory access time is improved dramatically by utilizing the bandwidth available from internal arrays if they are integrated into a single chip. For the densest DRAM technologies the interconnects provide typically 50MB/s of bandwidth per part even though the actual bandwidth is much higher. The core of most modern microprocessor need 1 to 2 GB/s bandwidth. DRAM/Logic merged technology now permits very significant amount of logic to be placed on DRAM chips, meaning that he bandwidth available from internal memory arrays can be utilized directly by one or more CPUs placed directly on the chip. The second advantage is power dissipation. If memory bus can be eliminated, almost of memory bus related power dissipation which is major portion of total power dissipation of the system can be removed.

In fact, Neo Magic Corp. announced revolutionary graphics controller chip that embedded 1 Mb frame buffer on the die alongside the controller logic using an available 16Mb DRAM process[1]. And a chip that merged 100K custom circuits and 4.5Mb DRAM onto a single die has been announced by Loral Federal Systems. However, any quantitative investigation has not been done about many open issues which deals with the slower speed of the DRAM transistors, routing area penalty due to lack of metal layers of typical DRAM process, and circuit technology to offset those device performance penalty. In this paper the logic device performance penalty of DRAM process has been investigated, penalty model has been extracted, the gate delay penalty model has been verified with realistic critical path circuit simulations of the state of the art 64 bit microprocessor, and routing area overhead has been predicted.

2 DIFFERENCES OF LOGIC AND DRAM PROCESSES

Logic and memory processes have been developed in different directions. Therefore, there are so many things different about a DRAM process that DRAM array can not be implemented on logic process. In DRAM transistor it is extremely important to reduce the leakage current of storage capacitor through pass transistor, and threshold voltage is increased by applying substrate bias. Therefore, leakage current is the big obstacle in implementing DRAM using logic process, and the leakage current of logic process including all the leakage currents such as tunneling and subthreshold currents and its impact on DRAM refresh cycle have been investigated. The leakge currents through gate oxide and junction have been investigated using the following equations [2]:

$$J_{ox} = J_0 E_{ox}^2 e^{-kt_{ox}} \quad (1)$$

$$J_j = G_0 V \frac{E_i}{E_0} e^{-\frac{E_0}{E_j}} \quad (2)$$

where J_0 = 6.5 x 10^{10} (A/cm^2), k is the imaginary part of the wave vector,E_0 = 2.9 x 10^7 V/cm from Fair and Wivell[3], and the preexponential factor G_0 = 3 x 10^9A/V/cm^2. The leakage currents through gate oxide and drain junction are 0.0018fA and 0.003fA, respectively for W=8.25μm, L=0.4μm for gate oxide of 95Åand 3V power and 1.73286 x $10^{17}/cm^3$ doping concentration. Those are very negligible. However, the subthreshold current on logic process is a big issue to look at. Using the SPICE BSIM model the leakage current has been calculated considering the continuity of drain current and the first order equation in the subthreshold voltage region[4]. The subthreshold current for 0.5μm/0.5μm of 0.5μm logic process is 2.53 x 10^{-11}A. The total amount of leakage current in the logic process requires 3.19ms refresh cycle time because 10% leakage of the total stored charge is regarded as a failure. 3.19ms refresh cycle is faster than commercially available 64M DRAM refresh cycle by factor of twenty. This is the main reason that the merged systems have to be implemented on DRAM process. The other issue in using logic process for DRAM is reliability issue. In the current DRAM process, the DRAM cell leakage current should be less that 1fA in order to pass reliability specification. The

0-7803-3073-0/96/$5.00 ©1996 IEEE

breakdown voltage of the storage capacitance is another problem. DRAM processes uses a special dielectric structure(Oxide/Nitride/Oxide) to increase the breakdown voltage of the storage capacitor. The last problem of using logic process for DRAM is contact resistance and capacitance of the bit line. In order to secure enough sensing margin, the ratio of CB(bit line capacitance) to CS(cell capacitance) should be not more than 15 to 1. As die size increases by using logic process the routing capacitance will be increased and it will be another challenging issue to keep that ratio.

Considering those issues described above, it would be reasonable that DRAM/Logic merged integrated circuit would be built on a DRAM process.

3 PERFORMANCE IMPACT OF MERGED LOGIC/DRAM ON STATIC LOGIC GATES

Since logic and memory processes are on diverging paths, there are so many differences between DRAM and logic process parameters. Three industry leading logic processes and two industry leading DRAM processes(64Mb and 256Mb) have been used for comparison, and very extensive circuit simulations have been performed using the device parameters to find out gate performance penalty of using DRAM process for logic components. The threshold voltage of DRAM process, therefore, is increased more than 0.6V comparing to that of logic process due to the substrate voltage applied to reduce the leakage current. This threshold voltage shift causes performance degradation of logic circuits The three logic processes are 0.5μm, 0.6μm, and 0.8μ CMOS processes. The two DRAM processes are 0.5μm 64Mb and 0.4μm 256Mb DRAM process. Even though minimum channel lengths of those processes are different, the processes are contemporary ones.

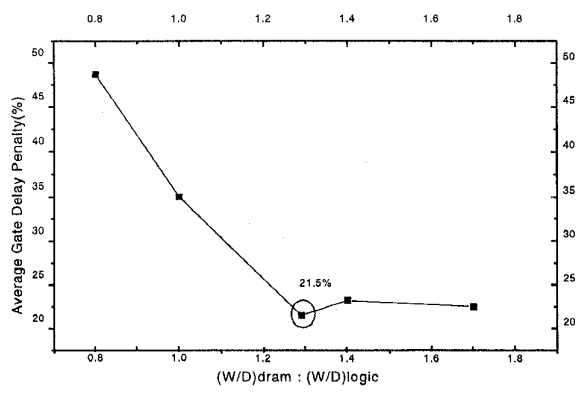

Fig.1 Average Gate Penalty VS. (MD:ML)

In order to find out the minimum penalty, different device geometries have been tested for five different gates. Fig.1 shows the average penalty as a function of (MD:ML). MD:ML is the ratio of $(W/L)_{DRAM}$ to $(W/L)_{LOGIC}$. As shown in Fig.1 the average logic gate performance degradation is about 21% as minimum when (W/L) of the transistors on DRAM process is increased 29% comparing to the (W/L) of the transistors on logic process for the same gate.

Table 1. Logic gate performance comparison

–	*L-Process*	*D-Process*	*Perf.Penalty*
Inverter	0.186nS	0.231nS	24.2%
2-Nand	0.279nS	0.344nS	23.3%
3-Nand	0.443nS	0.514nS	16.0%
2-Nor	0.389nS	0.484nS	24.4%
3-Nor	0.670nS	0.803nS	19.8%
AVERAGE	–	–	21.5%

Table.1 shows the detail comparison of the average logic gate delay of five different gates for both DRAM and logic process. The simulation was done for the worst case process corner at room temperature with fanout of four which is average fanout in realistic designs. The device geometry of those gates were chosen so that the effective β_p/β_n ratio becomes two and effective driving strength becomes comparable. Fig.2 shows the average driving capability of the DRAM and logic transistors with different fanout. The average gate delay was taken out of the five gates used for the Table.1 simulation results. DRAM process device tends to have more capacitance loading because the devices are optimized not for performance but for device density and yield. And also as mentioned earlier device geometry is increased 29% from logic process transistor to achieve minimum gate delay penalty.

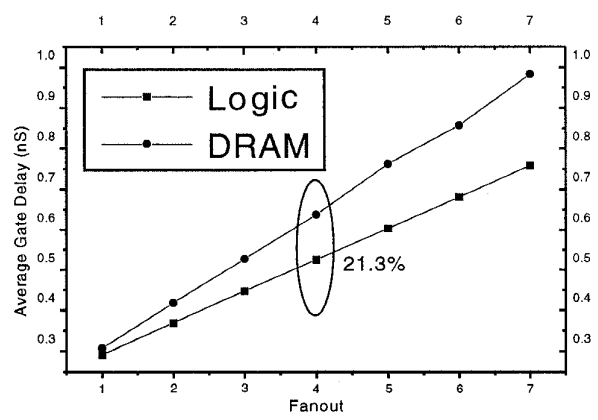

Fig.2 Driving capabilities of DRAM and logic gates

The input capacitance C_{IN} of a MOS transistor consists of the following components[5].

1. CGBO – Gate to bulk parasitic capacitance
2. CGDO, CGSO – Gate to drain and source parasitic capacitance
3. C_{OX} – Gate Oxide capacitance

$f_B(v)$, $f_D(v)$, $f_S(v)$, $F_O(v)$ are voltage dependent functions. Combining those individual terms, C_{IN} is given by

$$C_{IN} = CGBO \times L \times f_B(v) + CGDO \times W \quad (3)$$
$$+ \; CGSO \times W + C_{OX} \times W \times L \times F_O(v) \quad (4)$$

In DRAM process, all of those CGBO, CGDO, CGSO, and C_{OX} factors are larger than those of logic process. Consequently, those factors increase the input gate capacitance up to 60% for the same fanout gates comparing to logic devices. The 21.3% gate delay performance degradation in Fig.2 matches with the data in Table.2

4 PERFORMANCE PENALTY ON REAL CMOS CIRCUIT

In real circuit design, device sizes are optimized and ratioed so that a particular edge is faster than the other for critical path circuit. If the critical path circuit has to drive a long distance, the circuit may have to be designed so that it is less sensitive to any noise and ground bouncing. Therefore, most of the critical path circuits are non-standard type of gates. In order to investigate the realistic circuit performance penalty on DRAM process, those non-standard type circuit's performance has to be studied to find out if the circuit performance penalty model can be applied to the non-standard type circuits. Total twenty real critical path circuits have been selected to verify the gate delay penalty model from state-of- the-art 64 bit microprocessor. The simulation results and gate delay penalty are shown in Table 3. As shown in Table.3 the logic gate delay penalty is well matched with the model.

5 ROUTING AREA PENALTY

Memory processes have many poly interconnects but only one or two metal layers. Logic processes have only one or two poly layers, but it usually has four or more layers of metal interconnect. With two metal layers, a penalty has to be paid in routability, size and performance. The longer average metal line will add more capacitive and resistive load to the driver. In order to investigate area penalty of CMOS circuits implemented in DRAM process due to less routability and wiring capacitance increase per net, nine blocks have been selected out of state-of-the-art 64 bit microprocessor design, and routing simulations have been performed with two metal layers and four metal layers using automatic routing and placement tool. The average wiring length per net can be found by dividing the total metal interconnection length of all the metal layers by the total number of nets for both two and four metal layer cases. Table.3 shows the area and wiring capacitance per net of those nine test circuits for both two and four metal layer process. The results show that the wiring capacitance per net is increased by 20.85% if the same circuit is routed using two level metal layer process. But the absolute capacitance increase per net is only 7.6ff per net. When the wiring capacitance increase is reflected on the total gate delay performance penalty, it is additional 1.5%. But this data is valid for intra-block routing. Considering global clock and power distribution, the performance penalty due to routability will be worse than this at chip level depending on chip size and block sizes. Therefore, block partition is critical in order to minimize the routing penalty, and extra efforts should be taken for cell library and macro cell development so that they meet synthesis requirements by leaving more routing channels inside the cell. It is obvious that the routing channel area will be increased as total number of nets and components in a circuit increases. Using the routing simulation results in this research, an area penalty model is derived by fitting the curve using the simulation results. Fig.3 shows the trend of area penalty as a function of total component used and total number of net.

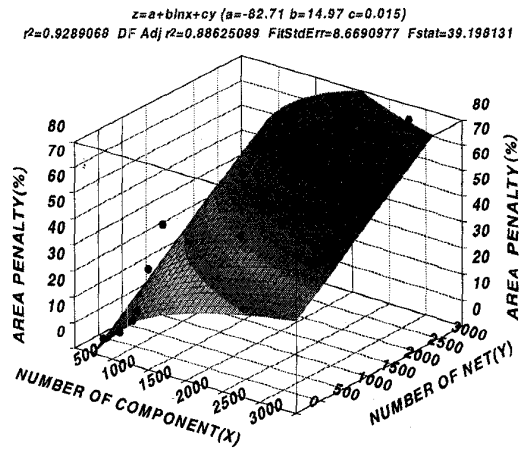

Fig.3 Routing area overhead

As shown in Fig.3 The area penalty is more sensitive to the number of net than to the number of components. The fitting equation representing the area penalty trend is given by

$$Z = a + b \times ln(X) + c \times Y \quad (5)$$

$$a = -82.7, b = 15, c = 0.015 \quad (6)$$

,where Z is the area penalty and X and Y are number of component and number of net,respectively,

6 ALTERNATIVE CIRCUIT TECHNOLOGY TO OFFSET THE PERFORMANCE PENALTY

It is important for circuit designers to select appropriate logic family to implement logic function and to meet the design specification. Low power logic design uses static, dynamic and pass transistor logic, and special logic families. Pass transistor logic offers reduced transistor count by eliminating the PMOS transistors and it provides improved speed. However, complementary pass networks are desirable to achieve good logic levels. The number of drain-source connections at the output is effectively doubled, hence doubling the output capacitance and making CMOS pass transistor networks inherently slow. As a result, this pass logic offers only moderate area and speed improvements. Special logic circuits include adiabetic circuits, self timed circuits. The primary disadvantage of adiabetic logic is the requirement of the ramp clock. The power supply for this circuit requires a power resonant circuit to provide the ramp clock signal. Self timed circuits fall into the general class of adaptive clocked circuits. The disadvantages of this circuits are that they are complex to design and reduce system performance if blocks are allowed to go into a "sleep" mode. On the other hand dynamic logic gates use only one of the two networks which form a static gate, taking into account the simplicity of interconnection, for complex blocks of gate with several inputs a saving of more than 50% in the layout area is expected.

Table 2. Performance difference of critical paths for DRAM and logic process

–	*Logic Depth*	*No.of Gates*	*Logic Process(nS)*	*DRAM Process(nS)*	*Performance Penalty(%)*
circuit1	12	516	3.02	3.68	21.85
circuit2	14	308	3.43	4.05	18.07
circuit3	19	308	4.45	5.54	24.50
circuit4	11	308	3.35	4.02	20.00
circuit5	15	308	5.07	6.20	22.28
circuit6	17	186	5.76	7.45	29.34
circuit7	7	186	2.29	2.79	21.83
circuit8	8	186	2.65	3.28	23.77
circuit9	9	186	3.63	4.60	26.72
circuit10	16	141	5.36	6.53	21.82
circuit11	5	141	1.69	2.02	19.52
circuit12	9	101	1.59	2.00	25.78
circuit13	6	101	1.31	1.57	19.85
circuit14	7	101	1.61	1.92	19.25
circuit15	6	101	1.32	1.65	25.00
circuit16	6	101	1.43	1.78	24.48
circuit17	9	240	2.49	3.13	25.70
circuit18	7	240	1.56	1.94	24.36
cricuit19	8	240	1.88	2.30	22.34
circuit20	5	16	1.00	1.22	22.00
AVERAGE	–	–	–	–	22.91

Table 3. Routing and wiring simulation results

–	*Block Size*		*Area*			*Cap . per net*		
Block	♯ of Net	♯ of Comp.	4 layer(sq.μm)	2 layer(sq.μm)	penalty(%)	4 layer(ff)	2 layer(ff)	penalty(%)
block1	587	409	219,340.8	238,280	8.63	27	31	14.81
block2	2334	2566	1,033,881.6	1,736,582.4	67.90	50	64	28.00
block3	962	962	241,990.4	346,032	42.99	15	23	53.33
block4	580	476	323,251.2	364,540.8	12.77	40	59	47.50
block5	651	651	341,990.4	441,686.4	29.15	41	55	34.14
block6	468	468	215,424	225,836	4.83	48	49	2.08
block7	1321	1400	1,108,224	1,576,872	42.28	53	63	18.86
block8	241	231	147,993.6	152,716.8	3.19	21	14	-33.33
block9	3111	2519	811,008	1,455,648	79.48	27	33	22.22
Average	–	–	–	–	–	–	–	20.85

The CMOS dynamic gates offer much faster gate delay than CMOS static gates not only because it has less gate loading but also its gate threshold is actual device threshold.(Static gate's threshold is generally (VDD-VSS)/2.) The dynamic gate performance benefit was looked into using the same methodology used for static gate performance penalty. The same critical path circuits used for the logic gates performance analysis have been implemented using dynamic gates and simulated with DRAM process parameters. In most of the cases the penalty of using DRAM process for logic circuits is overcome by using dynamic gates. Furthermore, since dynamic gates are positive logic gates(AND, OR, etc) the number of logic level could be reduced by taking advantage of the logic polarity.

7 CONCLUSION AND DISCUSSION

This paper presented the circuit performance, routing penalty of using DRAM process for logic components of DRAM/Logic merged integrated circuits, and alternative circuit technology to offset the penalty. The simulation results show that there is 22% logic circuit performance degradation including routing capacitance penalty when they are implemented using DRAM process. That amount of penalty can be offset by using dynamic circuit technologies. This paper has demonstrated the penalty of using DRAM process in a quantitative manner and will be a good reference for DRAM embedded circuit design in the future.

References

[1] EE TIMES, "Embedded-DRAM arrival stirs debate" ,*EE. Times*, pp.4, April 10, 1995.

[2] Carver A. M., "Scaling of MOS Tech. to Submicrometer Feature Sizes", *Journal. of VLSI Signal Processing*, vol 8,pp.9-25, July 1994.

[3] R.B. Fair and H.W. Wivell, "Zener and avalanche breakdown in As-implanted and low-voltage Si n-p junctions", *IEEE Trans. Electron Dev.*, vol. ED-23, pp.1527-1534,1983

[4] Meta-software, "HSPICE User's manual",*Meta-software*, vol.2 pp7-71 ~ 7-80, 1995

[5] Mohamad I. E., "Digital MOS Integrated Circuits: A Tutorial", *Digital MOS Integrated Circuits 2.* , *IEEE Press* pp.3-33.