

Compound semiconductor device requirements for VLSI

R. B. Brown, A. Chandna, T. R. Huff, R. J. Lomax, T. N. Mudge, R. Oettel*, and M. Upton

Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan 48109-2122,
and * Cascade Design Automation, Bellevue, Washington 98006

ABSTRACT: High-performance digital systems, such as a 60,500-transistor RISC microprocessor designed at the University of Michigan, are a growing application area for compound semiconductors. To support such VLSI circuits, a process must have, in addition to good device switching times: high levels of integration, good yields, reasonable power dissipation, and dense, multilevel interconnect. A digital logic family should have gates with good load-driving characteristics, reasonable noise margins, special features for embedded memory, and support by appropriate design automation tools. Device development for digital applications should be guided by all of these requirements.

1. INTRODUCTION

Since the introduction of GaAs circuits, high-performance digital systems have been considered a major application area for compound semiconductor technology. Early overoptimistic predictions of the digital GaAs performance advantage over silicon, followed by the fabrication challenges of an immature technology, caused a negative reaction among system designers which hindered acceptance of GaAs. Only recently has the promise for high-speed digital VLSI in compound semiconductors begun to be fulfilled.

The digital market in GaAs was slower to develop than the microwave and analog areas, but digital GaAs circuits are now being delivered in supercomputers, signal processors, and telecommunication systems. In silicon, digital circuits have come to dominate analog because in many cases they offer lower power, better repeatability, and freedom from dependence on precision passive components. Again in digital silicon circuits, VLSI dominates small-scale integrated circuits, for reasons of better reliability and performance, lower power, and smaller system size. These same forces will be at work in GaAs markets, and can be expected to enlarge the share of GaAs ICs which are digital, and to increase the ratio of large application specific circuits to smaller standard parts. Lande's (1992) current projection for growth in the worldwide GaAs IC merchant market supports this view, with growth rate predictions of 33% for digital circuits in 1992 compared to 20% growth for MMICs, and a 5-year projection of 231% growth for digital, compared to 173% for MMICs. The potential size of the digital market makes it important for compound semiconductor device researchers to understand the device and process requirements of VLSI circuits.

The objective of this paper is not to compare the many compound semiconductor technologies for digital applications, but rather, to illustrate the issues which are important in large digital circuits, so that they will be given appropriate attention by process and device researchers. All of the examples presented are from our work with E/D MESFETS. Most of the observations can also be generalized to other device types and logic styles.

2. EXPERIMENTAL GaAs VLSI CIRCUITS

We have prototyped a simplified version of a RISC (reduced instruction set computer) microprocessor, which is implemented in direct-coupled FET logic (DCFL) in the Vitesse 1.2- μm drawn / 0.8- μm effective channel length HGaAs II process. The circuit and performance details of this chip, called Aurora, will be reported by Brown et al (1992) at the IEEE GaAs IC Symposium. It is composed of 60,500 transistors, and was packaged in a 344-pin multilevel ceramic chip carrier which required a frame size of 12.175 \times 7.941 mm. Aurora includes a control section, a 3-port register file, and a 5-stage pipeline with these functions: instruction fetch, operand read, ALU execution, memory load/store, and register file update. It executes a subset of the MIPS instruction set, including 10 3-operand ALU operations, 8 immediate instructions, and 8 branch and jump instructions.

This first CPU chip was designed to exercise a preliminary version of the GaAs CAD tools described in Section 3.4, and was not optimized for speed. Still, it executes instructions with a clock rate as high as 137 MHz. While this speed is far from what is possible in the technology, it does compare favorably with the CMOS R3000 and ECL R6000 versions of the MIPS architecture, which operate at 40 and 62.5 MHz, respectively; one should note that the CMOS and ECL processors are full implementations of the architecture, whereas Aurora is not.

A 1K \times 32 SRAM and a second version of the CPU have been implemented in the Vitesse 0.6- μm effective channel-length HGaAs III process; this CPU includes more functionality than the first version, and it is optimized for speed. Details of these chips will be reported at a later date. We have also developed automatic, design-rule portable, physical design capability for DCFL circuits, which allows easy comparison of different processes. Our experience designing and testing large digital circuits and implementing layout generators which support various GaAs processes has helped clarify the size, power, and performance dependencies on device and process characteristics.

3. DIGITAL LOGIC ISSUES

Transistor switching time is certainly important for high-performance digital circuits, but over-emphasis on this parameter can obscure other features of a process and logic family which will determine their viability. All of the process parameters are interdependent, and the device and logic family characteristics are intimately related. Among the most important process features are high levels of integration, good yields, reasonable power dissipation, and dense, multilevel interconnect. A logic family should have gates with good load-driving characteristics, reasonable noise margins, tolerance for system-level variations, and support by appropriate design automation tools. One would like to optimize every desirable parameter, but the parameters often present conflicts (such as speed vs. noise margin) that require tradeoffs to be made. Many of the parameters in this optimization have minimum requirements, below which, digital circuits will not be competitive, no matter how attractive other features may be.

3.1 Integration Level

The first issue to be considered is integration level. The 'package delay' associated with getting signals through an output buffer, off-chip interconnect, and an input buffer can account for a large percentage of the clock cycle time in high-performance systems, even when the most advanced packaging is used. For example, Kayssi's (1992) simulations of Aurora, flip-chip mounted on a multichip-module (MCM) with a 4-K word instruction cache, show that the MCM delay is 45% of the total clock cycle. When the cache size is increased to 8-K words, the clock period must be lengthened, and the MCM delay increases to 55% of the clock cycle. These percentages would be even higher with other packaging schemes. The package delay means that a slower technology which has high enough integration levels to keep the critical path on one chip, can outperform a faster technology which has to have chip-crossings in the critical path. Pipelining and judicious partitioning can partially ameliorate the problem, but still, many GaAs applications will require 100,000 to 1,000,000 devices on a chip to be competitive with silicon.

Table 1: Comparison of 8×8 multipliers in three DCFL processes. All parameters are normalized.

	Gate Metal	Metal 1	Metal 2	Metal 3	Total Layout Area	Total Routing Area
Process A	1.00	1.00	1.00	1.00	1.00	1.00
Process B	0.90	0.60	0.50	0.28	0.49	0.21
Process C	0.50	0.97	1.11	1.43	0.97	0.82

Integration level, in turn, is dependent on yield, power dissipation, logic style efficiency, active device area, and interconnect density. Yield is a function of material defect density, process complexity, and other factors in each technology which influence the level of parameter control that can be maintained. Most potential applications for digital GaAs require air cooling. Power dissipation, therefore, becomes the integration-limiting parameter for many high-speed technologies. A true comparison of integration level between technologies would have to be done at the functional level, because of the variation in transistor efficiency from one logic family to another. For example, though DCFL has only $n + 1$ transistors per gate, compared to $2n$ transistors in a complementary CMOS gate, we have found that DCFL requires about 68% more gates (and logic levels) per function because it has low fan-in and fan-out, requires more buffering, provides only limited use of pass gates, and does not support complex gates or dynamic circuits.

The area occupied by active devices is a function of all of the design rules; gate length, the most commonly quoted geometric parameter is not necessarily an accurate indication of cell size. Transistor size is of prime importance in determining the density of RAMs, but much less important than interconnect dimensions in affecting the size of logic circuits composed of datapaths and random logic. This is illustrated in Table 1, a comparison of 8×8 Booth-encoded array multipliers implemented as datapaths by our CAD tools in three real DCFL processes, having drawn gate lengths in the ratio shown. All of these circuits were routed in gate metal plus 3 interconnect levels, but with ground distributed on the top routing level instead of on the fourth level of metal available in two of the processes. As seen in the table, layout area is a much stronger function of interconnect dimensions than of gate length. Even more striking is the difference in total routing area, approximated from the total routing lengths for Metal 1, 2, and 3, times the respective minimum width design rule. Interconnect capacitance, discussed below, is a direct function of routing area.

3.2 Interconnect

The importance of interconnect in a VLSI process cannot be overstated. The switching delay, τ , for any logic family, is related to the difference in charge between states at the output of a logic gate, and to the current available to effect a change of state: $\tau \propto C\Delta V/I$. Sensitivity to parasitic loading varies with process and logic family. In FET technologies, this is the dominant delay mechanism; it calls for small logic swings, high transconductance, and low parasitic capacitance.

Most of the parasitic capacitance comes from interconnect. Of primary importance is keeping the circuit area as small as possible to minimize wire length; this reduces both parasitic capacitance and time-of-flight for signals. Routing capacitance is minimized by using enough levels of interconnect, narrower lines, larger separation between interconnect layers, and lower dielectric-constant insulators. The effect of narrowing the separation between lines is not immediately obvious; while it reduces the circuit area, it does increase horizontal line-to-line capacitance. However, the total-routing-area data shown in Table 1 makes a strong case for reducing interconnect spacing to the fabrication limits.

The importance of minimizing interconnect capacitance is illustrated by Figures 1 and 2, which show the effects of reducing capacitive load (Figure 1) and of reducing unloaded gate delay (Figure 2) on four critical paths in our microprocessor. The logic paths in these plots are from the register file (RF), adder (A1 and A2), and branch logic (BR). (These figures ignore the fact that faster gates would have greater transconductance and therefore drive the capacitive loads more effectively.) The plots do show clearly that performance is

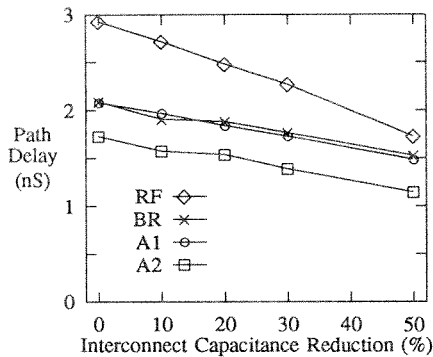


Figure 1: Delay reduction for four critical paths with reduction in interconnect capacitance.

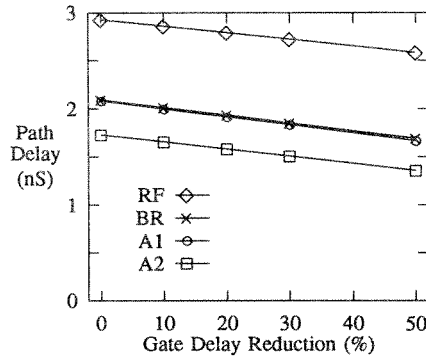


Figure 2: Delay reduction for four critical paths with reduction in intrinsic gate delay.

dominated by interconnect loading, and therefore, reducing interconnect capacitance would be even more effective at increasing circuit speed than would reducing intrinsic gate delay. The sensitivities to these effects vary among the paths simulated. The closest results are for the branch logic, where a 50% reduction in capacitance has a 40% greater effect than a similar reduction in unloaded gate delay. The biggest difference is in the register file, where capacitance reduction has a 248% greater effect.

The importance of having enough layers of interconnect merits further illustration. In the Vitesse HGaAs III process, we use Gate Metal and Metal 1 for wiring inside of leaf cells, and Metal 1, 2, and 3 for datapaths, standard cell blocks, and global routing. Metal 4 is a ground plane, and V_{dd} is distributed on Metal 3. Table 2 shows the improvement in density which we have achieved in moving from the HGaAs II (a 3-metal process) to HGaAs III (a 4-metal process). Of course, geometric design rule changes between the processes and other factors, noted below, also affect the density. The control blocks are different circuits (bypass logic in HGaAs II and stall logic in HGaAs III), but they are about the same size, and both are implemented in standard cells using the same logic synthesis tool (Finesse, from Cascade Design Automation, Bellevue, WA, USA). The register files in Table 2 are both 32-word \times 32-bit, three-port, tree-decoded, pass-gate latch implementations, which differ only in buffering.

The density numbers for the CPUs include all of the unoccupied space in the pad frame — there is actually more of it in the version with 4-metal interconnect. Some of the increase in density is due to the inclusion of additional memory structures for a small on-chip instruction cache on the 4-metal chip. But aside from this, the HGaAs III version of the CPU is still about 2.4 times denser. In our analysis, half of this improvement is due to the third layer of routing; improved circuit structures and layout techniques incorporated into our newer CAD tools account for another 35%; and the remaining 15% of improvement results from smaller line widths in the HGaAs III process.

Adding interconnect layers to a digital process beyond a routeable gate metal, 3 interconnect levels, and a ground plane would result in diminishing returns. Trying to achieve high performance with fewer layers than this or with coarse interconnect pitch or an inefficient design style, though, starts a vicious cycle. A larger layout has more capacitance, therefore requiring larger buffers, which further increase the layout size, parasitic capacitance and power dissipation, requiring larger buffers, yet.

3.3 Memory

To avoid the chip-crossing delay mentioned above, many digital systems will require embedded memory. Our GaAs SRAM work is leading toward on-chip primary cache for the next version of the CPU. Memory

Table 2: Density comparison between 3-metal and 4-metal processes.

Circuit	HGAs II		HGAs III	
	Transistor Count	Density (Trans./mm ²)	Transistor Count	Density (Trans./mm ²)
Largest Control Block	582	1067	516	1364
Register File	21,910	2014	23,278	4253
CPU	60,500	540	160,000	1475

must be dense and power efficient if it is to be embedded. The need to integrate memory with large logic circuits adds to the list of desirable characteristics in a digital process.

For example, subthreshold leakage currents in MESFETs and MODFETs are orders of magnitude larger than those in MOSFETs. In SRAMs, chip size and power are strong functions of leakage current. Though much less attention has been focused on minimizing leakage currents than on increasing transconductance, leakage currents are as important to performance. If too many memory cells are connected to a bit line, the leakage current through the pass transistors connected to unselected memory cells (about 100nA/bit) could corrupt the data of a selected memory cell (about 20 μ A). The total leakage on a bit-line should be an order of magnitude smaller than the active current, so the number of bits that can be safely connected to a column is limited to 32. This constraint requires that a significant portion of the total RAM area be devoted to sense amplifiers and write circuitry. Table 3 shows how SRAM area would decrease if leakage currents could be reduced to allow more memory cells per column, thereby amortizing the column support circuitry over more bits. As can be seen, for this design at 32 bits/column only 70.6% of the total chip area is consumed by the memory cells. A reduction in leakage current by only 1 order of magnitude would increase the percentage of area occupied by the memory cells to 92% of the total area.

In any technology, the pullup of a static RAM cell should provide just enough current to offset the leakage current of the pulldown devices. (Leakage currents, therefore, also set the lower limit for cell power.) In conventional GaAs DCFL processes, long, minimum-width depletion transistors are used to keep this current small. The characteristics of these devices present an area / power tradeoff; for example, in our SRAM, the highest impedance standard-threshold depletion transistor that fits in a 400 μ m² cell provides much more current than is needed to offset the leakage currents. As the area of the cell is decreased, the pullup length

must be decreased, increasing the power. Figure 3 shows the effect of varying the pullup length (cell size) on power dissipation. This plot includes curves for a digital process pullup transistor, a special higher-threshold depletion transistor, and a polysilicon load. The polysilicon load curve was constructed assuming lightly-doped resistors, which can be located above the remaining 4 transistors, adding no additional area. As seen in the figure, poly loads are invaluable to SRAM designs.

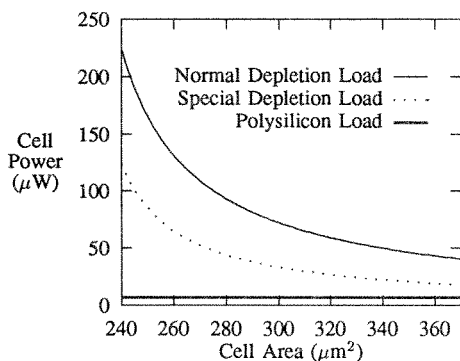


Figure 3: SRAM cell power vs. cell size for three load devices.

Table 3: Effect of reducing leakage currents on area of 1K×8 SRAM.

	Number of Bits / Column				
	32	64	128	256	512
Normalized SRAM Area	1.00	0.87	0.80	0.77	0.75
Cell Area Percentage of Total Area	70.6	81.6	88.4	92.1	93.8

3.4 Computer-Aided Design

In VLSI design, the practicality of a given technology is also dependent on the design tools available for it. Design automation is necessary to manage VLSI complexity. CAD tools can also help designers take advantage of the performance potential of a given technology. As noted above, about 35% of our improvement in density in going from one process generation to the next was due to improved physical design tools. In another experiment, we mapped the 8×8 array multiplier of Table 1 onto a sea-of-gates array provided by one of the three foundries; 100% gate utilization was assumed. The full implementation with our GaAs circuit compiler occupies only 63% of the area taken by the raw gates required in the array. Realistic cell utilization in the array would amplify this difference. Use of appropriate design methodologies and CAD tools is the most efficient way to minimize circuit size and interconnect capacitance, and thereby improve the performance of a circuit in a given technology.

Our design methodology and tools provide the needed support for high-performance design in general, and for DCFI in particular. The tools provide design-rule portability, so that a given design can be evaluated in different rule-sets or easily translated into a newer rule-set. The design methodology creates circuits with physical datapaths organized as one would in a handcrafted design, minimizing chip area and total interconnect length compared to standard cell- or array-based methodologies. The routers support multilevel interconnect, variable width signal routing, multiphase clock distribution, ground planes, and automatic power-rail sizing for IR drop and electromigration. The analysis capability includes a static timing analyzer which handles both single-phase and two-phase clocks, and delay calculations that include interconnect *RC* delay. The CAD tools now include automatic performance-driven placement and buffer-sizing, which we expect to further improve speed and power dissipation in our next generation of chips.

4.0 SUMMARY

It is the performance of compound semiconductors in systems, not ring-oscillator speeds, that will dictate their future in digital applications. Device and process development for digital circuits should be as concerned with potential integration levels, dictated by yield and power dissipation, as with the inherent speed of the nonlinear devices. Without high integration levels, the speed of compound semiconductors is lost to chip-crossing delays. Circuit performance can be improved faster by improving interconnect than by improving device switching speed. Embedded memory will be necessary in the highest-performance systems, so digital processes need to provide memory-specific features. And finally, design methodology and CAD tools have a major influence on the competitiveness of a technology.

ACKNOWLEDGMENT

This work has been supported by the U.S. Defense Advanced Research Projects Agency under DARPA/ARO Contract No. DAAL03-90-C-0028, and by the U.S. Army Research Office under Contract No. DAAL03-87-K-0007.

REFERENCES

- Brown R B, et al 1992 GaAs IC Symposium (IEEE)
- Kayssi A I, et al 1992 Proc. IEEE Int. Symp. on Circ. and Sys., vol. 2 pp 919-922
- Lande S 1992 Gallium Arsenide IC Forecast (Luton: BIS Strategic Decisions Ltd.) pp 7-14