Journal of

CIRCUITS, SYSTEMS, AND COMPUTERS

Volume 14 • Number 2 • April 2005

Multi-GHz SiGe BiCMOS FPGAs with New Architecture and Novel Power Management Techniques

K. Zhou, J.-R. Guo, C. You, J. Mayega, R. P. Kraft,

T. Zhang, J. F. McDonald and B. S. Goda

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Journal of Circuits, Systems, and Computers
Vol. 14, No. 2 (2005) 179–193
© World Scientific Publishing Company

World Scientific
www.worldscientific.com

MULTI-GHz SiGe BiCMOS FPGAs WITH NEW ARCHITECTURE AND NOVEL POWER MANAGEMENT TECHNIQUES

K. ZHOU*, J.-R. GUO, C. YOU, J. MAYEGA, R. P. KRAFT, T. ZHANG
and J. F. McDONALD

Department of Electrical, Computer and Systems Engineering,
Rensselaer Polytechnic Institute,
Troy, NY 12180, USA
**zhouk@rpi.edu*

B. S. GODA

Department of Electrical Engineering and Computer Science,
United States Military Academy,
West Point, NY 10996, USA

Received 23 December 2002
Revised 14 April 2004

The availability of Silicon Germanium (SiGe) Heterojunction Bipolar Transistor (HBT) devices has opened a door for GHz Field Programmable Gate Arrays (FPGAs).[1,2] The integration of high-speed SiGe HBTs and low-power CMOS gives a significant speed advantage to SiGe FPGAs over CMOS FPGAs. In the past, high static power consumption discouraged the pursuit of bipolar FPGAs from being scaled up significantly. This paper details new ideas to reduce power in designing high-speed SiGe BiCMOS FPGAs. The paper explains new methods to reduce circuitry and utilize a novel power management scheme to achieve a flexible trade-off between power consumption and circuit speed. In addition, new decoding logic is developed with shared address and data lines. A SiGe FPGA test chip based on the Xilinx 6200 architecture has been fabricated for demonstration.

Keywords: FPGA; SiGe; CML; power management; X-pattern decoding.

1. Introduction

Field Programmable Gate Arrays (FPGAs) have gained popularity due to their flexibility and wide range of applications. A FPGA consists of multiple copies of a basic programmable logic element or cell. Logic cells are arranged in a column or matrix on the chip. To perform more complex operations, logic cells can be connected to other logic elements on the chip using a programmable interconnection network. The operating speeds of current CMOS FPGAs are around 50–600 MHz. These clock rates are not comparable to contemporary microprocessor clock rates, much less in providing access to X- or K-band applications. These slow operating speeds prevent the use of CMOS FPGAs in high-speed digital system applications.

High-speed FPGAs find applications in many research and commercial fields such as Digital Signal Processing, where digital filters need fast multipliers, adders, shifters, flip-flops, etc.[3] They can also be used in applications which involve high-speed broadband networks,[4] high-speed inline processing, image recognition, and the area of genome analysis.[5] The top-level architecture of a Silicon Germanium (SiGe) FPGA is shown in Fig. 1(a). The block diagram of a single logic cell is shown in Fig. 1(b). It consists of a Configurable Logic Block (CLB) and routing multiplexers.[6]

This paper describes the design of a SiGe FPGA (with new features), which is compatible with the Xilinx 6200 architecture. Changes have been made to make it work optimally in the design environment. The Xilinx 6200 is selected as the test-bench because its design has already been made public. The same ideas presented hereafter for the Xilinx 6200 can also be applied to other FPGA families. In spite of operating at high frequencies, the switching noise is less because Current Mode Logic (CML) has been selected for the logic cell design. CML is very similar to Emitter Coupled Logic (ECL).[7] The only difference is that differential pairs are used for all signals and there is no need for a reference voltage. The SiGe 5HP Heterojunction Bipolar Transistors (HBTs) are high-speed transistors with cutoff frequencies around 50 GHz.[8] All simulations were done using Cadence 4.4.6 and IBM's 0.5 μm three-metal layer SiGe BiCMOS 5HP technology.[9] IBM has announced even faster SiGe HBT technologies, such as 7HP which offers 0.18 μm minimum feature sizes

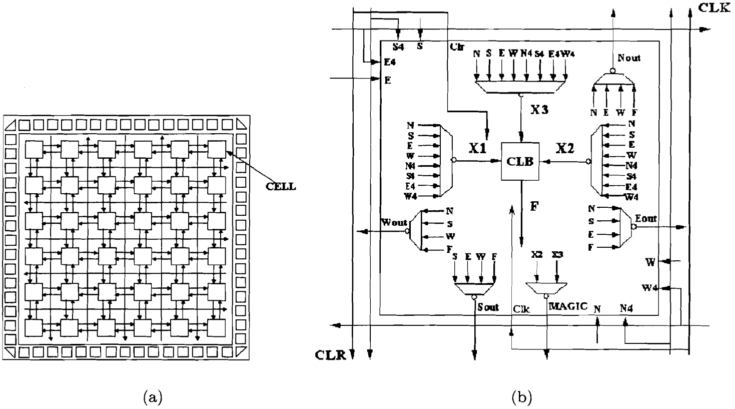


Fig. 1. (a) Top-level architecture of the SiGe FPGA. (b) Schematic of a logic cell: the CLB and routing multiplexers.

and an f_T of 120 GHz,[10] and 8HP which offers 0.13 μm minimum feature sizes and an f_T of 210 GHz.[11]

2. The SiGe HBT Structure and Its Advantages

Si is widely used in radio frequency (RF) and microwave circuit applications because of its assets such as high-quality dielectric, excellent thermal property, extreme abundance and easy purification, etc. However, Si is not ideal from a device designer's point of view because the carrier mobility is rather small for both electrons and holes, and the maximum velocity that these carriers can attain is limited to about 1×10^7 cm/s under normal conditions. Hence, Si is regarded as a slow semiconductor.[12]

The SiGe HBT is one of the most successful bandgap engineering devices. It has comparable performance to Gallium Arsenide (GaAs) RF devices, while it can be fabricated at a significantly lower cost. In order to achieve higher performance, Ge is selectively introduced into the base region of the transistor. The Ge mole fraction in typical profiles varies from 3~9%. From Fig. 2, it can be seen that there exists a drift field in the base which aids in the faster movement of minority carriers. This reduces the base transit time and hence increases the cutoff frequency. The smaller base bandgap of SiGe compared to Si enhances electron injection, producing higher current gain for the same base doping level compared to Si devices. SiGe HBT and Si CMOS can be grown over the same substrate because the process has strict processing compatibility with existing CMOS tool sets and metallization schemes. This technology is referred to as BiCMOS technology. Due to all these advantages, the SiGe 5 HP technology (provided by IBM) was chosen for the FPGA design.

3. Design Description

Figure 3 shows a simple CML Exclusive-OR (XOR) gate with input and output waveforms as well as a Binary Decision Diagram (BDD).[13,14] The rise and fall times of the XOR gate are approximately 17 ps and 13.6 ps, respectively. The current is

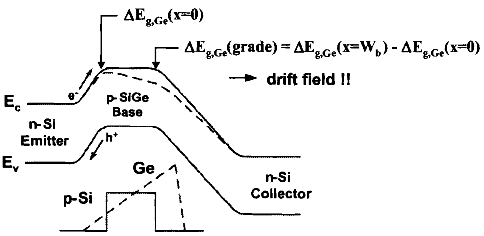


Fig. 2. Band diagram of SiGe HBT.[12]

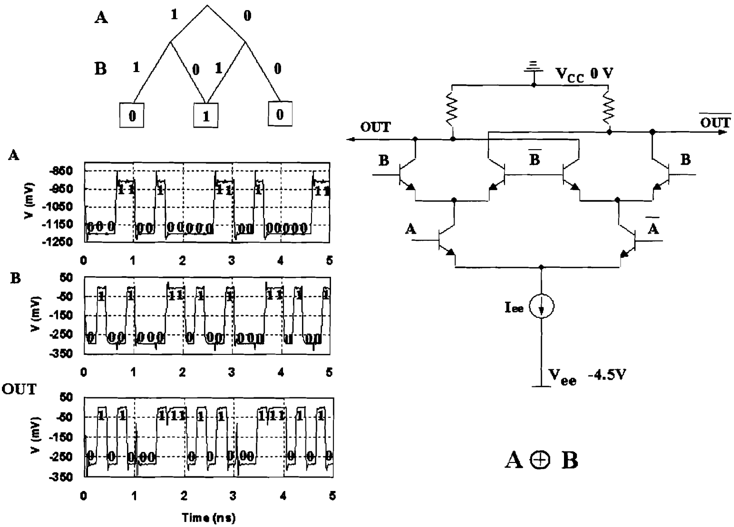


Fig. 3. A CML XOR gate with input and output waveforms.

maintained constant at 0.6 mA by using a current mirror at the bottom of the circuit. The XOR gate structure can be immediately determined from the creation of the BDD. The BDD in Fig. 3 is basically a two-level binary tree with some leaves merged. Hence, CML circuits are also called CML trees. The two-input XOR gate, therefore, requires correspondingly two levels of transistors to realize the logic function. The maximum number of levels is determined by the power supply (0 and −4.5 V) and the difference between levels is slightly more than one V_{BE} (0.85 V). Hence there can be up to four stacked transistors in every branch of the tree although the two-input XOR gate implementation only uses two levels. Correspondingly this would lead to four levels of logic — (0 to −0.25 V)(level1), (−0.95 to −1.2 V)(level2), (−1.9 to −2.15 V)(level3) and (−2.85 to −3.1 V)(level4) in a single tree. The 250 mV peak-to-peak voltage swings were found to be appropriate for the design.

A level1 output can be converted to other level signals by using an emitter follower. Current mostly flows through one of the branches of the tree structure. This branch pulls either $\overline{\text{OUT}}$ or OUT low depending on which path is conducting according to the logic. Considering the XOR gate, if $A = 1$ and $B = 0$, $\overline{\text{OUT}}$ has to be pulled low. So the branch through which the current flows when $A = 1$ and $B = 0$ should be connected to $\overline{\text{OUT}}$. The resistor which is the top of the tree structure in conjunction with the current source determines the swing.

A disadvantage of CML is its high-level of DC power consumption. The power consumption is directly proportional to the number of trees used. In CML designs, there is a constant current flowing in all the trees, so there is always a constant power level even if a tree is not being used. This is why power management techniques play an important role in all CML designs.

We propose two approaches to reduce power consumption. The first one is to reduce the speed of operation dynamically and thus trade off speed for power. The second approach is the use of intelligent Computer Aided Design (CAD) software which generates the configuration stream to optimally manage the power states of the FPGAs. The detailed discussion is presented as follows.

The first approach is realized through multiple power states in the logic cell: Fast, Noncritical, Slow and Off. Before the logic cell is configured to operate in one of the four states, it should be determined whether that particular cell is used or not. If it is to be used, the cell is put into one of the three "ON" states. The cell is in the Fast mode when it is required to run at high-speed. When the cell operates at moderate speed, it can be configured to the Noncritical mode, which reduces the power consumption. When the cell is only used at low-speed, it can be configured to the Slow mode to reduce the power consumption even further. Finally, when a cell is "OFF", it does nothing and consumes no power. The CAD software modified from XACT6000 of the Xilinx 6200 manages the power state configurations during programming.

Power management brings up a number of issues that the circuit designer must be aware of. How long does it take for the cell to switch from one mode to another? How much power is saved in the Slow and Noncritical modes? Can part of the FPGA work in the Fast mode while another part work in the Slow mode? How fast can FPGAs work in these three modes respectively? Of course, the CAD tool must also be extended to optimize the design for these issues.

A Widlar current mirror can be easily redesigned to output multiple reference voltages as shown in Fig. 4. It can safely take 15 loads without obvious loading effects. Figure 4 also shows a simple CML tree with three transmission gates on top of it. The current in the current mirror is 0.6 mA, 0.3 mA, 0.1 mA and 0 in the Fast, Noncritical, Slow and Off modes, respectively.

The transmission gates control the mode in which the CML tree operates. Only one transmission gate is on at a time. The Schottky diodes are used to prevent shorts between V_{cc} and V_{ee}.

A main issue here is how fast the circuit can switch from one mode to another. The configuration switching speed is mainly limited by the Schottky diodes and transmission gates on top of the CML tree due to their introduced higher parasitic capacitances. An example of the current response is shown in Fig. 4 when the mode switches from Slow to Fast. The switching time of the current is around 37 ps, which is several orders of magnitude faster than the configuration time.

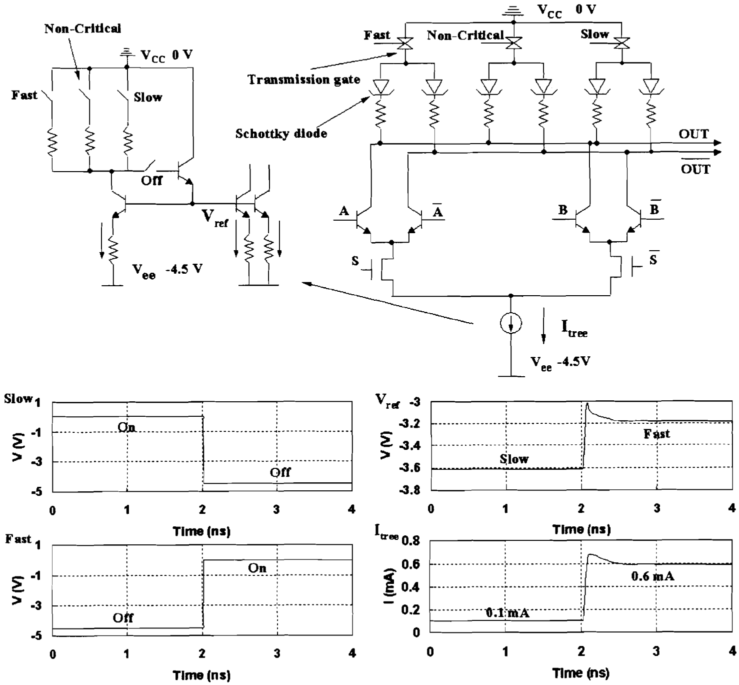


Fig. 4. Current response of the tree structure when the mode is switched.

The circuit technique in Fig. 4 does not rely on specific architectures. Hence other FPGA architectures can use the same technique to adjust power and speed.

3.1. *The CLB structure*

Figure 5 shows the CLB structure in the Xilinx 6200. There are two paths in the structure. They are:

(1) The sequential path which passes through the input multiplexers and then through the flip-flop.
(2) The combinational path that involves only the input multiplexers.

The X1 input controls whether Y2 or Y3 will be selected. The inputs to Y2 and Y3 can be outside inputs X2 or X3, their complements, the signal stored in

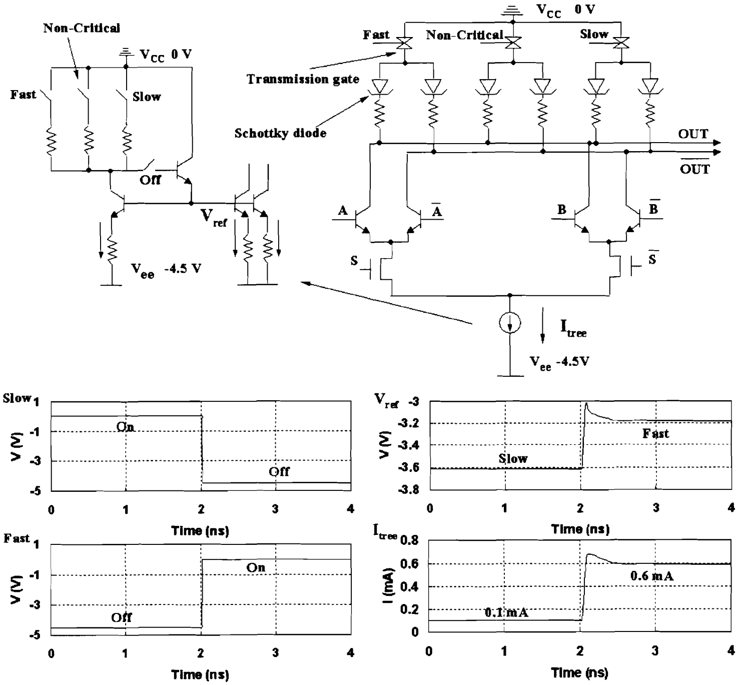


Fig. 4. Current response of the tree structure when the mode is switched.

The circuit technique in Fig. 4 does not rely on specific architectures. Hence other FPGA architectures can use the same technique to adjust power and speed.

3.1. *The CLB structure*

Figure 5 shows the CLB structure in the Xilinx 6200. There are two paths in the structure. They are:

(1) The sequential path which passes through the input multiplexers and then through the flip-flop.
(2) The combinational path that involves only the input multiplexers.

The X1 input controls whether Y2 or Y3 will be selected. The inputs to Y2 and Y3 can be outside inputs X2 or X3, their complements, the signal stored in

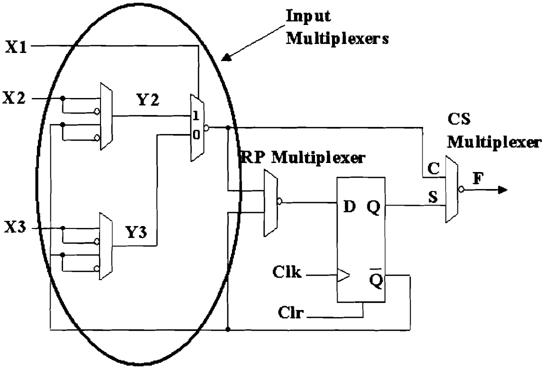


Fig. 5. Original Xilinx 6200 structure.[6]

the flip-flop, or its complement. The Chip Select (CS) multiplexer determines which path to choose and the Register Protect (RP) register controls what signal gets into the D flip-flop. The Clear is an asynchronous signal that resets the D flip-flop. The clock controls when the bit will be stored. Only during the rising edge of the clock can the bit be stored in the D filp-flop. The selected bits for all the multiplexers come from the configuration memory.

A simple implementation would be to design each multiplexer separately and then join them, but the power dissipation in such an implementation would be large. The original structure has been modified to make it suitable for CML. The objective is to achieve the same logic using fewer trees in order to reduce the power and propagation delay. Figure 6 shows the redesigned structure.

This structure can be implemented in just seven trees (3-logic with two pairs of emitter followers), whereas the original structure required 11 trees. This 36% reduction in the number of trees leads to at least a 36% reduction in power dissipation. Figure 7 shows the schematics of all the three blocks.

Since the number of trees in both the combinational and sequential paths have been reduced, the propagation delay is also reduced. The propagation delay for the previous CLB was 120 ps[1] and that of the new CLB with power management techniques is around 175 ps with the same number of loads in the Fast mode. The new architecture is slightly slower but has better performance in terms of power. The larger propagation delay arises from the parasitic capacitances introduced by the Schottky diodes as well as reduced current from 0.8 mA to 0.6 mA in the current mirror.[1] The design is multiplexer-based which drives the interconnect by emitter followers, thus gaining a much higher speed than CMOS FPGAs. The power

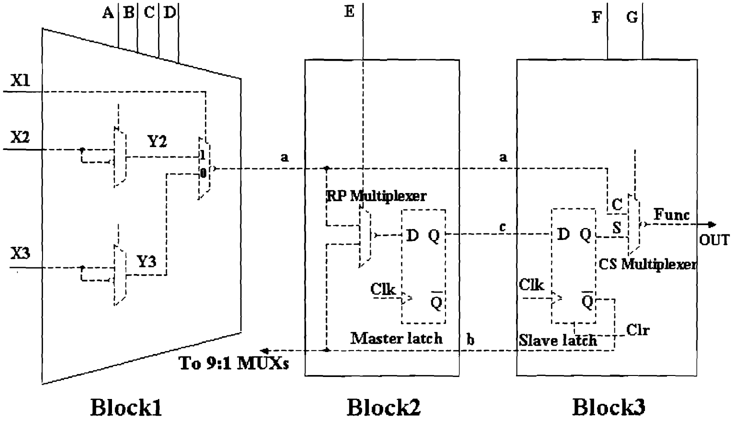


Fig. 6. Redesigned CLB structure.

management scheme results in 27% area overhead for each CLB. It is possible to eliminate some levels of speeds if area overhead is limited.

The same idea can be applied to other FPGA architectures too. CML circuits can realize any Boolean logic function. The layered structure of the CML tree allows an efficient implementation of complex logic functions, so that simple gates in other FPGA architectures can also be merged to reduce circuitry and save power when enough layers are presented.

3.2. *Configuration memory*

Figure 8 shows the schematic of a complete logic cell with configuration memory. The logic cell consists of a CLB and four routing multiplexers. Each multiplexer can be used to route the output signal of the CLB to the nearest logic cells (North, South, East and West). It also has a special multiplexer called the Magic multiplexer which is useful for corner turning (all the other four routing resources are straight). The previous architecture required 8:1 multiplexers for selecting the correct inputs to the CLB. These have been replaced by 9:1 multiplexers which implement the feedback from the Master–Slave flip-flop to the input multiplexers. The redesigned architecture can realize all the functions of the XC6200 architecture with fewer circuit trees.

The FPGA has two memory planes for the configuration data. More memory planes are possible when needed. In Fig. 8, each memory plane contains a different set of configuration bits for the FPGA as well as the state of the latch in the CLB. It configures the state of the routing multiplexers and the function of the CLB. The

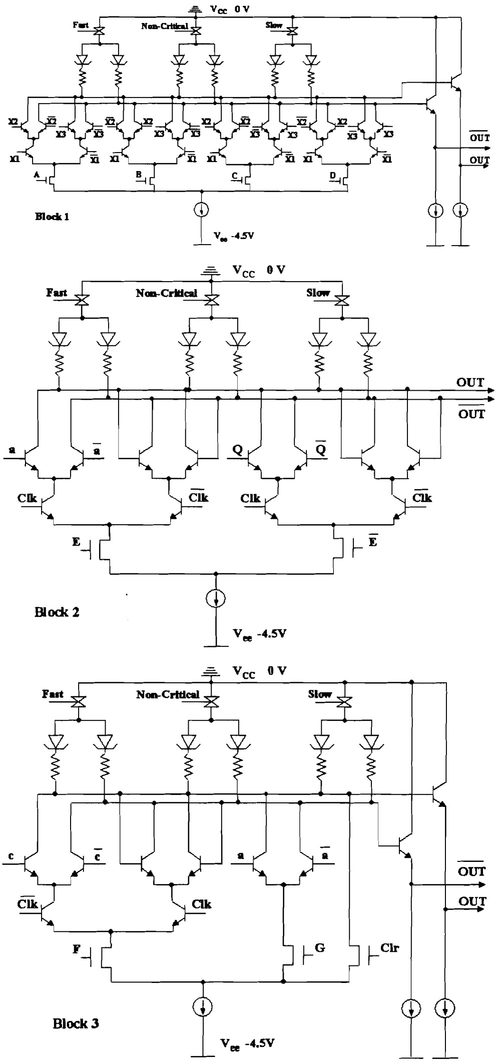


Fig. 7. Schematics of all three blocks in the new structure.

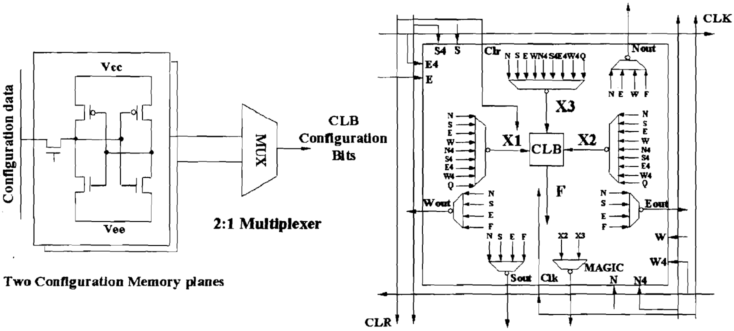


Fig. 8. Schematic of the new logic cell (three 8:1 input multiplexers in Fig. 1 have been replaced by 9:1 multiplexers).[15]

CLB can change functionality by loading in a different set of configuration bits. The switching can be done dynamically. Each memory plane has 52-bits to program the logic cell (18-bits for Routing multiplexers, 7-bits for CLB functionality and 27-bits for 9:1 multiplexers). One part of the FGPA can work in one mode on one application while another part can be configured to work in another mode on another application. The power management scheme requires three more programming bits in each logic cell. The configuration time is 5.7% longer as a consequence.

3.3. *X-pattern decoding*

A CAD software utility generates the binary data to configure the FPGA. This configuration is stored into the memory, which in turn makes every cell behaves as desired. Unless an efficient decoding scheme is in place, programming may result in many long address and data lines. Long address/data lines will increase congestion. Figure 9 shows a new decoding scheme which is more symmetric and has shared address and data lines. For a 4 × 4 FPGA design in Fig. 9(a), there is one main-decoder and four sub-decoders. When the global enable line is set, the main decoder enables one of the sub-decoders based on the least significant 2 bits of the control signals. The enabled sub-decoder in turn enables one of the cell decoders based on the most significant 2-bits of the control signals. The cell whose decoder is enabled will get programmed by the data coming on the address/data lines. This is shown in Fig. 9(b).

There are four address/data lines reaching each cell decoder. The address of the memory location into which the data has to be written is first sent over to the address/data bus. The address is registered by the rising edge of the enable signal. Each decoded address enables four SRAMs simultaneously. Next, the data which is to be written into the four enabled SRAMs is sent over to the shared bus.

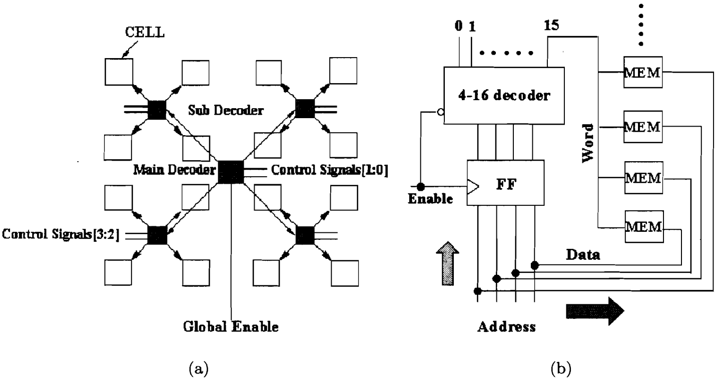


Fig. 9. (a) X-pattern decoding. (b) Cell decoder structure inside a single cell.

By this method, it is possible to write into $2^4 \times 4 = 64$ memory locations by using only four lines plus an enable signal. Using straightforward decoding scheme would require six address lines and one data line. This reduction is significant because all the address/data lines go through all the logic cells in the FPGA. Moreover, the normal decoding scheme would require a 6–64 decoder for each cell. Apart from this, there would be 64 lines going into the memory, which makes the layout denser. The decoding logic has been implemented in CMOS to save power since speed is not critical here.

4. Measured Results

A main concern for the new CLB is how fast it can run. In order to demonstrate the feasibility of the proposed circuits, a four-stage ring oscillator (RO) was designed to test the propagation delay of the CLB in the Fast mode with a three-stage and a two-stage ROs built for the Noncritical and Slow modes, respectively. Figure 10 shows the testing mechanism. The CLB is programmed to be an inverter by writing the appropriate bits into the configuration memory. Several CLBs are connected from end to end to make a RO. The 50-Ω terminated pad driver outputs the signal to

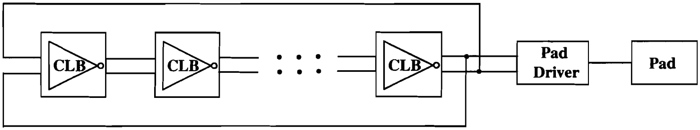


Fig. 10. Test structure.

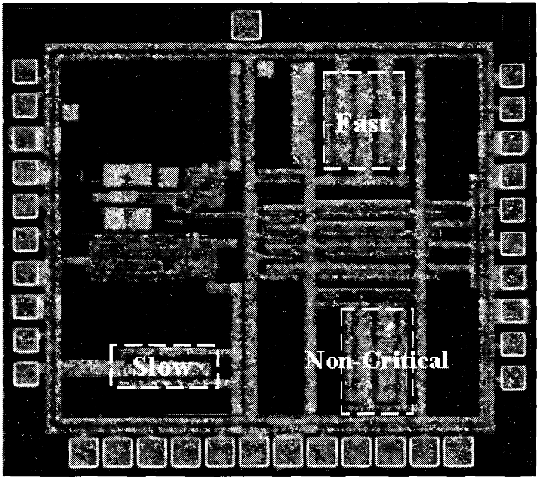
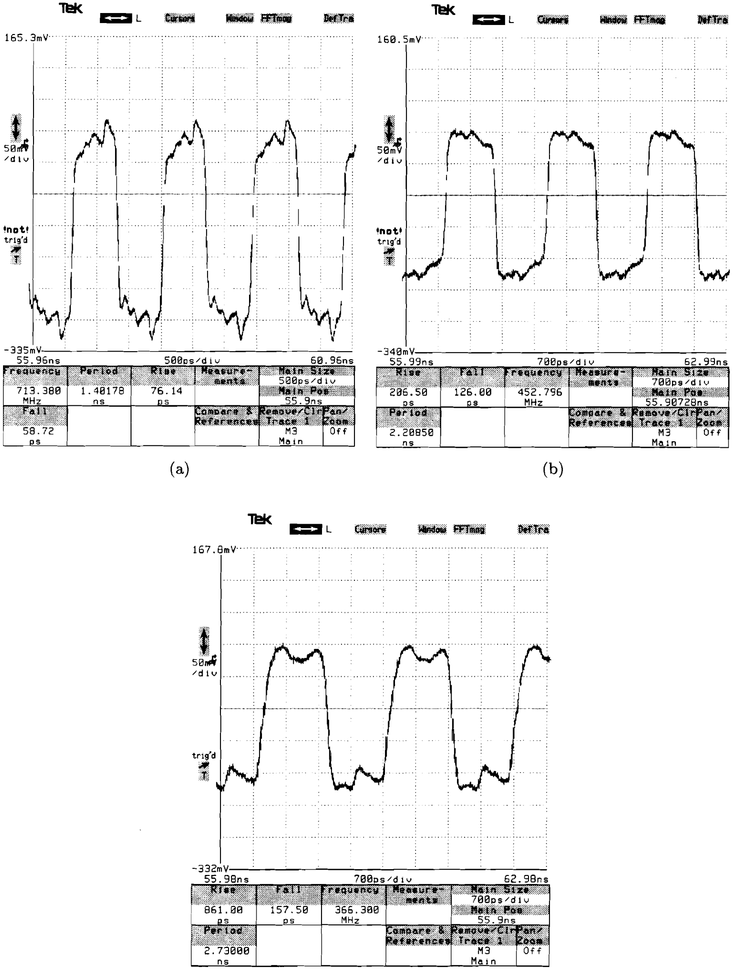


Fig. 11. Die photo of the test chip (2075 μm × 2300 μm).

the oscilloscope. It should be noted that all the outputs of the CLB are differential, hence the RO can consist of any number of CLBs.

These three ROs have been fabricated in IBM's 0.5 μm three-metal layer SiGe BiCMOS 5 HP technology. The switches in the Widlar current mirror are fixed to facilitate the testing. Figure 11 shows a die photograph of the prototypes. The CLB test chip die was tested as a bare die on a Techtronix probe station using two GGB Picoprobe Multi-Contact Wedge probes. Each probe contains two sets of power and ground pins and six signal pins. The output signals of the test chip were measured using a Tektronix 11801C digital sampling oscilloscope with a SD32 sampling head through 50-Ω cables. The periods of the output waveforms are only determined by the internal test chip circuits, hence the exclusion of packaging parasitics as a result of testing the bare die is not an important factor in the measurements. All the circuits have been tested with a supply voltage of 4.5 V.

Figure 12 shows the output waveforms from the test chip. The performance for major parts of the SiGe FPGA have been summarized in Table 1. It is obvious from the table that this is a very high-speed FPGA compared to a CMOS FPGA with significantly reduced power consumption. According to the measurements, the operating frequency of the new CLB in the Fast mode is 5.7 GHz. The main consideration for a bipolar FPGA is power dissipation. By using the new architectures, which were discussed earlier, at least 36% of the power can be conserved in the Fast mode. If the FPGA runs in the Slow mode, it can save even more power (at least



Frequency	Period	Rise	Measure-ments	Main Size 500ps/div Main Pos 55.9ns	
713.380 MHz	1.40178 ns	76.14 ps			
Fall			Compare & References	Remove/Clr Trace 1	Pan/Zoom
58.72 ps				M3 Main	Off

(a)

Rise	Fall	Frequency	Measure-ments	Main Size 700ps/div Main Pos 55.90728ns	
206.50 ps	126.00 ps	452.796 MHz			
Period			Compare & References	Remove/Clr Trace 1	Pan/Zoom
2.20850 ns				M3 Main	Off

(b)

Rise	Fall	Frequency	Measure-ments	Main Size 700ps/div Main Pos 55.9ns	
861.00 ps	157.50 ps	366.300 MHz			
Period			Compare & References	Remove/Clr Trace 1	Pan/Zoom
2.73000 ns				M3 Main	Off

(c)

Fig. 12. Measured waveforms out of the test chip: (a) measured waveform of the four-stage RO with CLBs in the Fast mode, (b) measured waveform of the three-stage RO with CLBs in the Noncritical mode, and (c) measured waveform of the two-stage RO with CLBs in the Slow mode.

Table 1. SiGe HBT CLB performance with all trees on.

Circuit type	CLB[1,2] (Previous)	CLB (Fast)	CLB (Noncritical)	CLB (Slow)
Propagation delay (ps)	120	175	368	689
Current (mA)	8.8	4.2	2.1	0.7
Power (mW)	39.8	18.9	9.45	3.15

83% of the power can be saved). The simulated propagation delay of the CLB in Fast, Noncritical and Slow modes are 171 ps, 329 ps and 641 ps, with the differences being 2.3%, 10.5% and 6.9%, respectively, between the measured and simulated results.

5. Conclusion

Low-power and high-speed is an eternal goal for circuit designers. SiGe is an obvious solution that combines low-power CMOS and high-speed bipolar together. However, in order to scale up the FPGA significantly, aggressive power management schemes must be in place. This paper presented several ideas such as the novel power control scheme, X-pattern decoding and architectural changes. The multiple power states allow the CLBs on the critical path to run in the Fast mode while other CLBs can be configured to operate in the Noncritical, Slow or Off mode without jeopardizing the throughput. The FPGA design can also be made more efficient with a smaller layout by using the new decoding logic and reduced circuitry. All these techniques make GHz FPGAs viable with reasonable sizes, rendering them suitable for high-speed digital applications. There are many other ideas yet to be implemented and tested as part of this research effort, such as improving the FPGA architecture, reducing the tree height and using faster BiCMOS technologies as they become available.

Acknowledgments

This work was supported in part by the Defense Advanced Research Project Agency (DARPA) under Contract N66001-02-1-8919 and in part by the National Science Foundation under Grant ECS-9980251.

References

1. B. Goda, J. F. McDonald, S. Carlough, T. Krawczyk and R. Kraft, SiGe HBT BiCMOS for fast reconfigurable computing, *IEE Proc. Comp. Digit. Tech.* **147** (2000) 189–194.
2. B. Goda, J. F. McDonald, R. Kraft, S. Carlough and T. Kwawczyk, Gigahertz reconfigurable computing using SiGe HBT BiCMOS FPGA, *11th Int. Conf. Field Programmable Logic and Applications* (2001), pp. 59–69.
3. B. V. Herzen, Signal processing at 250 MHz using high-performance FPGAs, *IEEE Trans. VLSI Syst.* **6** (1998) 238–246.

4. J. T. McHenry, P. W. Dowd, F. A. Pellegrino, T. M. Carrozzi and W. B. Cocks, An FPGA-based coprocessor for ATM firewalls, *IEEE Symp. FPGAs for Custom Computing Machines* (1997), pp. 30–39.
5. K.-P. Lam and S.-T. Mak, On computing transitive-closure equivalence sets using a hybrid GA-DP approach, *12th Int. Conf. Field Programmable Logic and Applications* (2002), pp. 935–944.
6. Xilinx series 6000 user guide, Xilinx, Inc., San Jose, CA (1997).
7. J. Rabaey, *Digital Integrated Circuits: A Design Perspective* (Prentice Hall, NJ, 1996).
8. SiGeHP (BiCMOS 5HP) design manual, IBM, Inc. (2001).
9. D. C. Ahlgren, G. Freeman, S. Subbanna, R. Groves, D. Greenberg, J. Malinowski, D. Nguyen-Ngoc, S. J. Jeng, K. Stein, K. Schonenberg, D. Kiesling, B. Martin, S. Wu, D. L. Harame and B. Meyerson, A SiGe HBT BiCMOS technology for mixed signal RF applications, *Proc. Bipolar/BiCMOS Circ. Tech. Meeting* (1997), pp. 195–197.
10. G. Freeman, D. Ahlgren, D. R. Greenberg, R. Groves, F. Huang, G. Hugo, B. Jagannathan, S. J. Shen, J. Johnson, K. Schonenberg, K. Stein, R. Volant and S. Subbanna, A 0.18 μm 90 GHz f_T SiGe HBT BiCMOS, ASIC-compatible, copper interconnect technology for RF and microwave applications, *Tech. Dig. IEEE Int. Elect. Dev. Meeting* (1999), pp. 569–572.
11. B. Jagannathan, M. Khater, F. Pagette, J.-S. Rieh, D. Angell, H. Chen, J. Florkey, F. Golan, D. R. Greenberg, R. Groves, S. J. Jeng, J. Johnson, E. Mengistu, K. T. Schonenberg, C. M. Schnabel, P. Smith, A. Stricker, D. Ahlgren, G. Freeman, K. Stein and S. Subbanna, Self-aligned SiGe NPN transistors with 285 GHz $f_{\max}$ and 207 GHz f_T in a manufacturable technology, *IEEE Electron. Dev. Lett.* **23** (2002) 258–260.
12. J. Cressler, SiGe HBT technology: A new contender for Si-based RF and microwave circuit applications, *IEEE Trans. Microw. Theor. Tech.* **46** (1998) 572–589.
13. H. Greub, J. F. McDonald and T. Yamaguchi, High performance standard cell library and modeling technique for differential advanced bipolar current tree logic, *IEEE J. Solid-State Circuits* **26** (1991) 749–762.
14. G. D. Micheli, *Synthesis and Optimization of Digital Circuits* (McGraw-Hill, 1994).
15. M. Smith, *Application Specific Integrated Circuits* (Addison-Wesley, MA, 1997).