

## Gigahertz SiGe BiCMOS FPGAs with new architecture and novel power management techniques

K. Zhou\*, J. -R. Guo, C. You, J. Mayega, R. P. Kraft and J. F. McDonald  
*Department of Electrical, Computer and System Engineering, Rensselaer Polytechnic Institute,  
Troy, New York 12180, United States of America*  
*\*Email: zhouk@rpi.edu*

B. S. Goda  
*Department of Electrical Engineering and Computer Science, United States Military Academy,  
West Point, NY 10996, USA*

Received (received date)  
Revised (revised date)  
Accepted (accepted date)

Typical Field Programmable Gate Arrays (FPGAs) are generally used in signal processing, image processing and rapid prototyping applications. The integration of Silicon Germanium (SiGe) Heterojunction Bipolar Transistor (HBT) devices with CMOS allows a new family of FPGAs to be created. This paper elaborates new ideas in designing high-speed SiGe BiCMOS FPGAs based on the Xilinx 6200. The paper explains new methods to reduce circuitry and utilize a novel power management scheme to reduce power consumption. In addition, new decoding logic has been developed where the address and data lines are shared. These ideas have improved the performance of a SiGe FPGA to run in the 5~6 GHz range.

### 1. Introduction

Field programmable Gate Arrays (FPGAs) have gained popularity due to their flexibility and wide range of applications. An 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 automatically connected to other logic elements on the chip using a programmable interconnection network. The operating speeds of current CMOS FPGAs are around 70-250 MHz. These slow operating speeds prevent their use 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, subtractors, flip-flops etc <sup>1</sup>. They can also be used in applications which involve high-speed broadband networks <sup>2</sup>, high-speed inline processing, image recognition and in the area of genome analysis <sup>3</sup>. The top-level architecture of a Silicon Germanium (SiGe) FPGA is shown in Fig. 1(a). The block diagram of a single logic cell

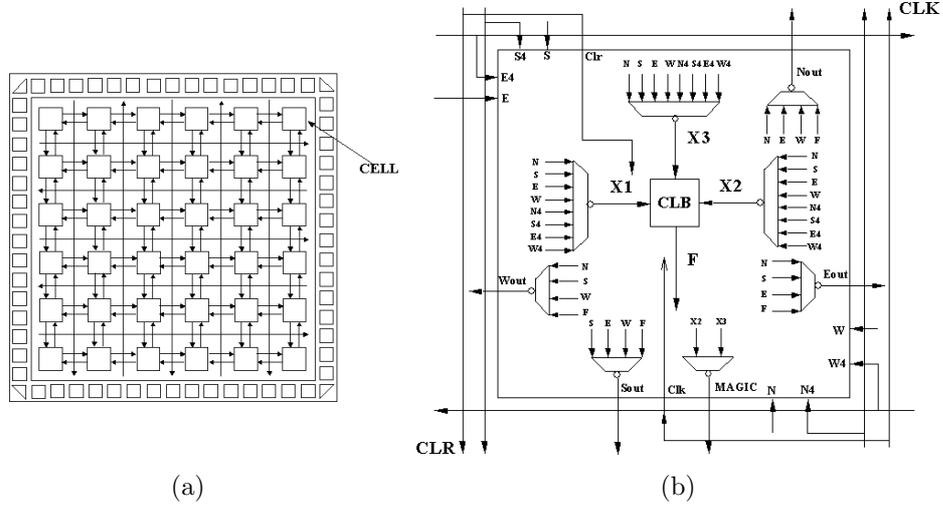


Fig. 1. (a) Top level architecture of the SiGe FPGA (b) Schematic of a logic cell: the CLB and Routing Multiplexers.

is shown in Fig. 1(b). It consists of a Configurable Logic Block (CLB) and routing multiplexers<sup>4</sup>. 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. 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)<sup>5</sup>. The only difference being 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<sup>6</sup>. All simulations were done using the IBM 5HP design kit and Cadence 4.4.6.

## 2. The SiGe HBT structure and its advantages

Si is widely used in radio frequency (RF) and microwave circuit applications because of its advantages 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  $1 \times 10^7$  cm/s under normal conditions. Hence, Si is regarded as a slow semiconductor<sup>7</sup>.

The SiGe HBT is one of the most successful bandgap engineering devices. It has the comparable performance to 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,

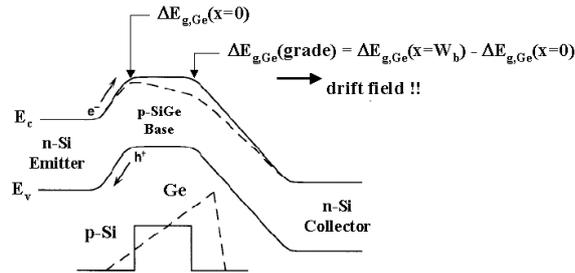


Fig. 2. Band diagram of SiGe HBT <sup>7</sup>.

the SiGe 5HP technology (provided by IBM) was chosen for the FPGA design.

### 3. Design Description

Fig. 3 shows a simple CML Exclusive-OR (XOR) gate with input and output waveforms <sup>8</sup>. The rise and fall times of the XOR gate are approximately 17 ps and 13.6 ps respectively. The current is maintained constant at 0.6 mA by using a current mirror at the bottom of the tree. There can be up to 4 transistors in every branch of the tree. Correspondingly this would lead to 4 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 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). 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  $\overline{OUT}$  or  $OUT$  low depending on which path is conducting according to the logic. Considering the XOR gate, if  $A=1$  and  $B=0$ ,  $\overline{OUT}$  has to be pulled low. So the branch through which the current flows when  $A=1$  and  $B=0$  should be connected to  $\overline{OUT}$ . 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

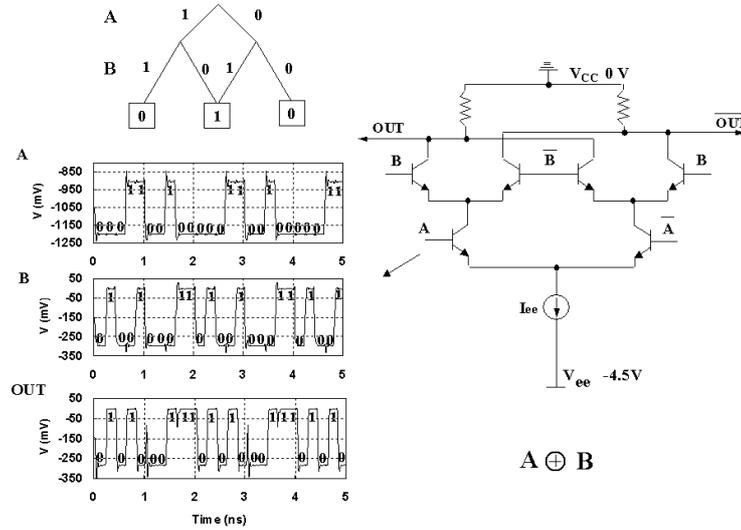


Fig. 3. A CML XOR gate with input and output waveforms.

play an important role in all CML designs.

There are two general approaches to reducing power consumption. The first one being the use of a smart Computer Aided Design (CAD) software which generates the configuration stream to turn off parts of the FPGAs (such as I/O devices and those unused cells). The second approach is to reduce the speed of operation dynamically and thus reduce the power. The following discussion presents the hardware design of the power management strategies.

The logic cell is designed to have multiple power states: Fast, Non-Critical, 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 supposed to be used, the cell is put into one of the three “ON” states. When the cell is not required to operate at high speed, it can be configured to the Non-Critical mode, which reduces the power consumption. When the cell is only used at low speed, it can be configured to reduce the power consumption even further. Finally, when a cell is “OFF”, it does nothing and consumes no power. The CAD software manages the state configurations during programming. The trade-off here is the CAD software is made more complicated and may need more time for compiling.

Power management brings up a number of issues that the circuit designer must be aware of. How much power is saved in the Slow and Non-Critical modes? How long does it take for the cell to switch from one mode to another? 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 tools must also be extended to optimize the design for these issues.

A Widlar current mirror can be easily redesigned to output multiple reference

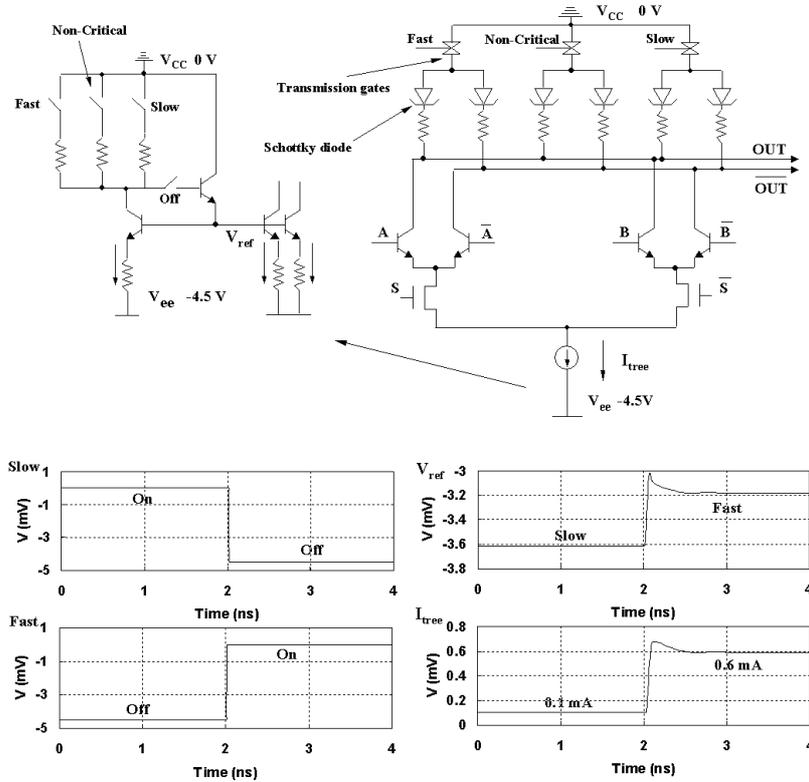


Fig. 4. Current response of the tree structure when the mode is switched.

voltages as shown in Fig. 4. It can safely take 15 loads without obvious loading effect. Fig. 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, Non-Critical, 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 concern 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.

### 3.1. The CLB structure

Fig. 5 shows the CLB structure in the Xilinx 6200. There are two paths in the

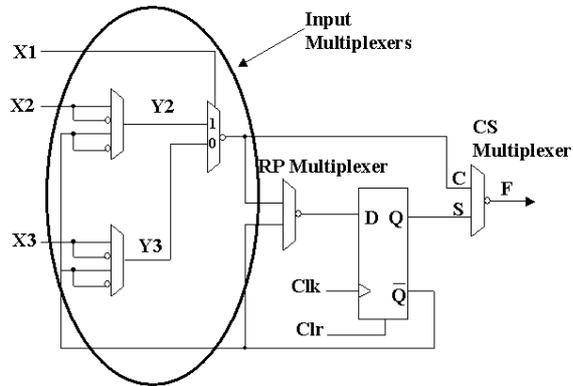


Fig. 5. Original Xilinx 6200 structure <sup>4</sup>.

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, its complement, the signal stored in 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 flip-flop. The select bits for all the multiplexers come from the configuration memory.

A simple implementation would be to design each multiplexer separately and

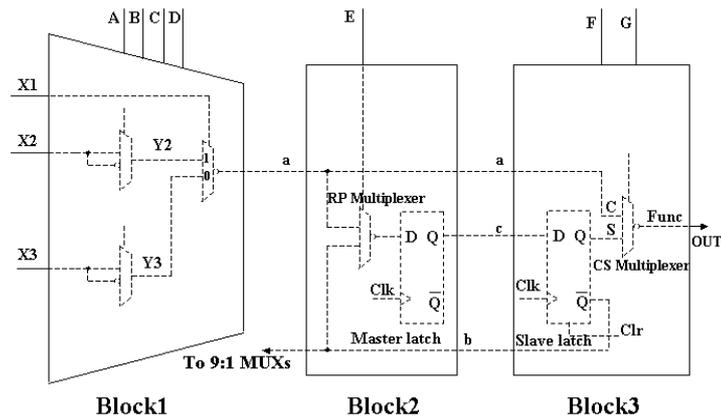


Fig. 6. Redesigned CLB structure.

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. Fig. 6 shows the redesigned structure.

This structure can be implemented in just 7 trees (3-logic with 2 pairs of emitter

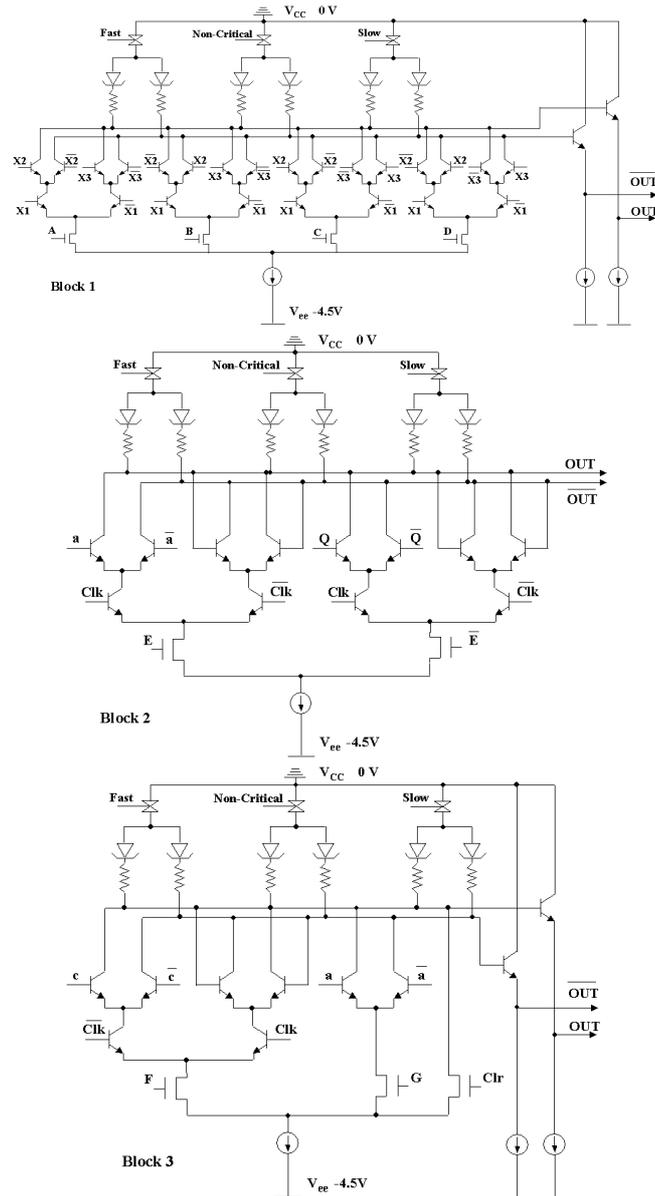


Fig. 7. Schematics of all three blocks in the new structure.



configures the state of the routing multiplexers and the function of the CLB. The 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 FPGA can work in one mode on one application while another part can be configured to work in another mode on another application.

### 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 behave 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. Fig. 9 shows a new decoding scheme which is more symmetric and has shared address and data lines. For a  $4 \times 4$  FPGA design in Fig. 9(a), there is one main-

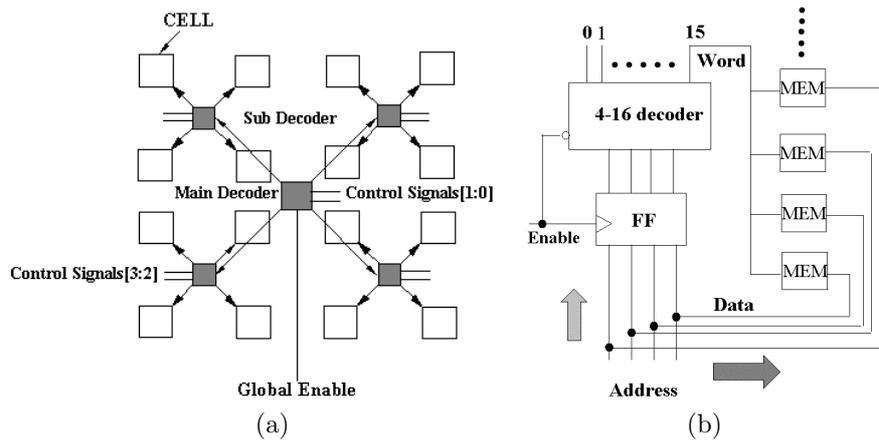


Fig. 9. (a) X-pattern decoding (b) Cell decoder structure inside a single cell.

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 sent first over the address/data bus. The address is registered by the rising edge of the enable signal. Each decoded address enables 4 SRAMs simultaneously. Next, the data which is to be written into the 4 enabled SRAMs is sent over the shared bus. By this method, it is possible to write into  $2^4 \times 4 = 64$  memory locations by using only 4 lines. Using straightforward decoding scheme would require 6 address lines

and 1 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 more dense. 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 RO built for the Non-Critical and Slow modes respectively. Fig. 10 shows the testing mechanism. The CLB is programmed to be an inverter by writing the correct bits into the configuration memory. Several CLBs are connected from end to end to make a RO. The 50- $\Omega$  terminated pad driver outputs the signal to the oscilloscope. It should be noted that all the outputs of the CLB are differential, hence the RO can be consisted of any number of CLBs.

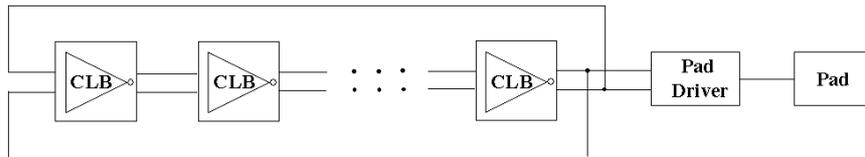
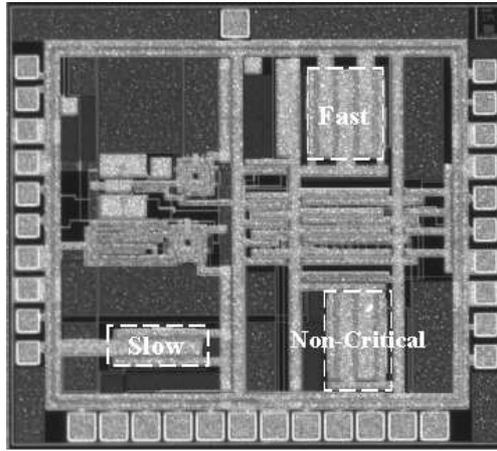


Fig. 10. Test structure.

These three ROs have been fabricated in IBM's 0.5  $\mu\text{m}$  three metal layer SiGe BiCMOS 5HP technology<sup>11</sup>. The switches in the Widlar current mirror are fixed to facilitate the testing. Fig. 11 shows a die photograph of the prototypes. The CLB test chip die was tested as bare die on a Techtronix probe station using two GGB Picoprobe Multi-Contact Wedge probes. Each probe contained two sets of power and ground pins and six signal pins. Test chip output signals were measured using a Tektronix 11801C digital sampling oscilloscope with a SD32 sampling head through 50- $\Omega$  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 bare die is not an important factor in the measurements. All the circuits have been tested with a supply voltage of 4.5 V.

Fig. 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 with significantly reduced propagation delay in the Fast mode. 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 new architectures, which were discussed earlier, at least 36% of the power can be conserved in the Fast mode. If the FGPA runs in the Slow mode, it can save even more power (at least 83% of the

Fig. 11. Layout of the new logic cell ( $2075 \mu\text{m} \times 2300 \mu\text{m}$ ).

power can be saved). The simulated propagation delay of the CLB in Fast, Non-Critical 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.

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

Circuit type	CLB <sup>9,12</sup> (Previous)	CLB (Fast)	CLB (Non-Critical)	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

## 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, serious power management scheme must be in place. This paper presented several ideas such as the novel power control scheme, X pattern decoding and architectural changes etc. The multiple power states allows the CLBs on the critical path to run in the Fast mode while other CLBs can be configured to operate in the Non-Critical, 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 it viable for gigahertz FPGAs with reasonable sizes, rendering it applicable 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.

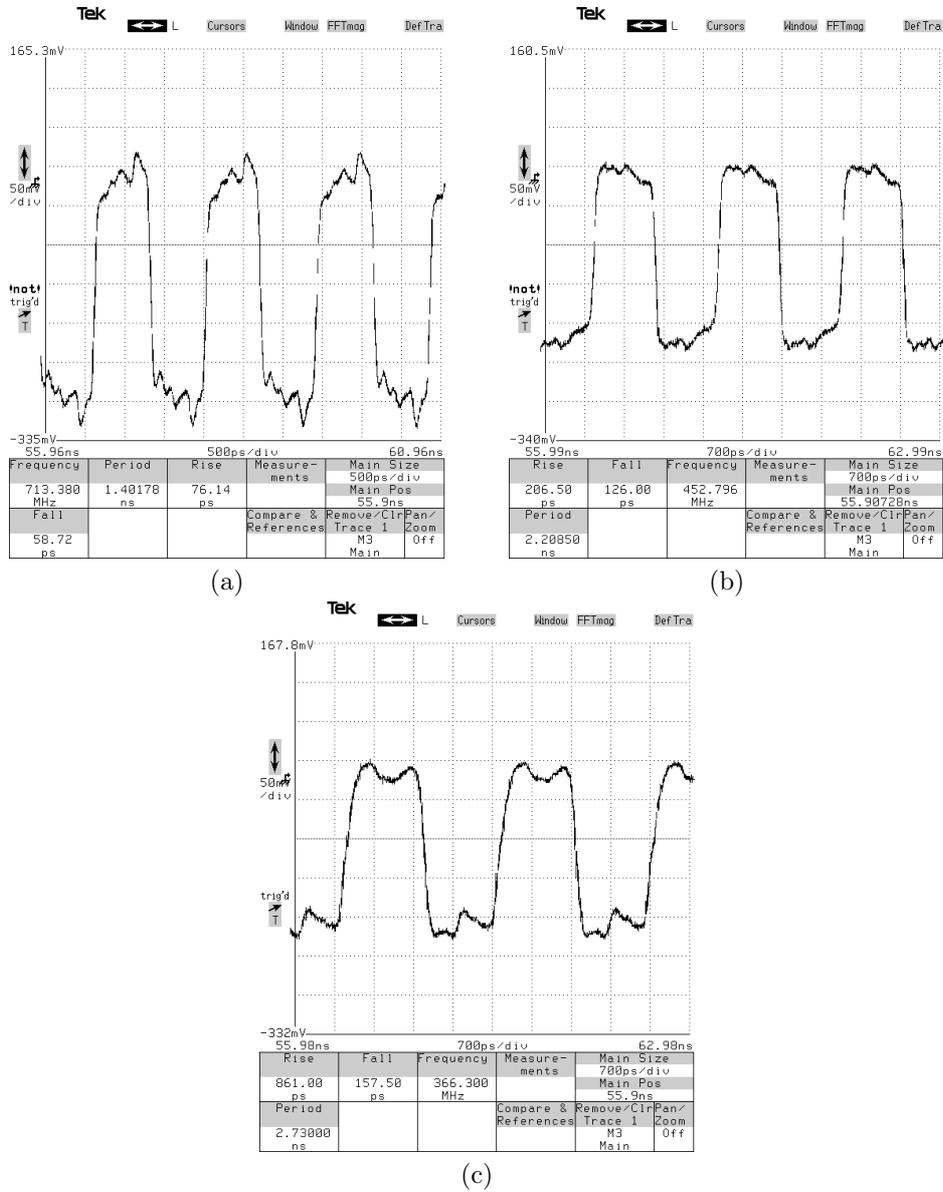


Fig. 12. Measured waveforms out of the test chip (a) measured waveform of the 4-stage RO with CLBs in the Fast mode (b) measured waveform of the 3-stage RO with CLBs in the Non-Critical mode (c) measured waveform of the 2-stage RO with CLBs in the Slow mode.

## References

1. B. Herzen, "Signal processing at 250 MHz using high-performance FPGAs", *IEEE Trans. VLSI Systems* **6** (1998) pp. 238–246.
2. J. McHenry, P. Dowd, et al., "An FPGA-based coprocessor for ATM firewalls", *IEEE*

*Symp. on FPGAs for Custom Computing Machines*, April 1997, pp. 30–39.

3. K.-P. Lam, S.-T. Mak, “Computing transitive closure equivalence sets using a hybrid ga-dp approach”, *12th International Conference on Field Programmable Logic and Applications*, Springer-Verlag, Sep. 2002, pp. 935–944.
4. “Xilinx series 6000 user guide”, Xilinx Inc., San Jose, CA., 1997.
5. J. Rabaey, “Digital integrated circuits: a design perspective”, Prentice Hall, NJ, 1996.
6. “SiGeHP (BiCMOS 5HP) design manual”, IBM Inc., May 2001.
7. J. Cressler, “SiGe HBT technology: a new contender for Si-based RF and microwave circuit applications”, *IEEE Trans. Microwave Theory and Techniques* **46** (1998) pp. 572–589.
8. H. Greub, J. F. McDonald, T. Yamaguchi, “High performance standard cell library and modeling technique for differential advanced bipolar current tree logic” *IEEE J. Solid-State Circuits* **26** (1991) pp. 749–762.
9. B. Goda, J. F. McDonald, S. Carlough, T. Krawczyk, R. Kraft, “SiGe HBT BiCMOS for fast reconfigurable computing”, *IEE Proc. Computers and Digital Techniques* **147** (2000) 189–194.
10. M. Smith, “Application specific integrated circuits”, Addison Wesley, MA, 1997.
11. 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. Harnome, B. Meyerson, “A SiGe HBT BiCMOS technology for mixed signal RF applications”, *Proc. Bipolar/BiCMOS Circuits and Technology Meeting*, 1997, pp. 195–197.
12. B. Goda, J. F. McDonald, R. Kraft, S. Carlough, T. Kwawczyk, “Gigahertz reconfigurable computing using SiGe HBT BiCMOS FPGA”, *11th International Conference on Field Programmable Logic and Applications*, Sep. 2001, pp. 59–69.