Application of Multi-Threshold NULL Convention Logic to Adaptive Beamforming Circuits for Ultra-Low Power

Brent Bell[1], William Bouillon[1], Shiming Li[1], Erica Logan[2], and Jia Di[1]

[1]Computer Science and Computer Engineering
Department
University of Arkansas
Fayetteville, AR 72701
www.csce.uark.edu

[2]Electrical Engineering Department
University of Arkansas
Fayetteville, AR 72701
www.eleg.uark.edu

Contact Author Email: jdi@uark.edu

Abstract: *With the decrease of transistor feature sizes into the ultra-deep submicron range, leakage power becomes an important design challenge for circuit designers. This paper examines the application of an asynchronous design paradigm named Multi-Threshold NULL Convention Logic (MTNCL) to adaptive beamforming circuits. MTNCL and synchronous designs were implemented using the IBM 130nm bulk CMOS process for power and area comparison. The MTNCL design showed substantial improvements in terms of active energy and leakage power compared to the equivalent synchronous design.*

Keywords: Silicon on Insulator; Quasi-Delay Insensitive; Low Power; Multi-Threshold NULL Convention Logic; Adaptive Beamforming; Digital Integrated Circuit;

Introduction

In recent decades, power consumption has become a major consideration in integrated circuit design. In high speed systems, clock switching could use a large portion of power. Additionally, leakage power has come to dominate power consumption as process sizes shrink. Adaptive beamforming circuits have many applications where lower power is highly desirable without sacrificing performance. These systems often require GHz range of throughput to accommodate the fast input data stream, while having long idle periods between sets of activities. In order to reduce power, asynchronous design methods have become increasingly attractive over the past two decades. Quasi-delay-insensitive (QDI) asynchronous circuits, such as NULL Convention Logic (NCL) do not use clock; instead, they incorporate handshaking protocols to control the circuit's behavior [1]. By removing the need for clock, switching power can be reduced and power consumption will be more evenly distributed across the chip.

The Multi-Threshold NCL (MTNCL) design paradigm incorporates the Multi-Threshold CMOS (MTCMOS) power gating mechanism inside every logic gate in order to reduce power even further [2]. This paper presents a fine-grain time delay (FTD) unit and a coarse-grain time delay (CTD) unit for use in an adaptive beamformer designed using the MTNCL paradigm for the DARPA Arrays at Commercial Timescale (ACT) program.

Background

MTNCL is a self-timed asynchronous design paradigm based on NCL. Like NCL, it uses dual-rail encoding to alternate between DATA and NULL phases. Table 1 shows the possible states for dual-rail encoding. During the DATA phase, a data wavefront propagates through combinational logic to the next register set. MTNCL uses early completion detection to determine when the circuit has finished its computation and is ready for a NULL phase. In the NULL phase, a sleep signal generated by the completion detection logic is used to generate NULL for the entire pipeline stage. MTNCL offers several advantages over other architectures:

- Correct-by-construction - as long as transistors switch properly, MTNCL circuits will function correctly without the need for costly timing analysis;
- Low Power - MTNCL circuits use much less leakage power through the application of MTCMOS power gating in each gate;
- Average-case performance - while synchronous systems must be designed for the worst-case delay, pipelined MTNCL systems always exhibit average-case throughput. As a result, MTNCL designs can be faster than many of their synchronous counterparts when ambient conditions change; and
- Enhanced compliance with the commercial digital IC design flow with respect to other asynchronous design paradigms.

Leakage power is reduced by the addition of a high-V_t transistor, controlled by the sleep signal, in the power-ground path of every logic gate. A low-V_t transistor is added to quickly pull the output of every gate to ground during the NULL phase. In addition to reducing leakage

Table 1. Dual-Rail State Encoding

RAIL_0	RAIL_1	STATE
0	0	NULL
0	1	DATA_1
1	0	DATA_0
1	1	INVALID

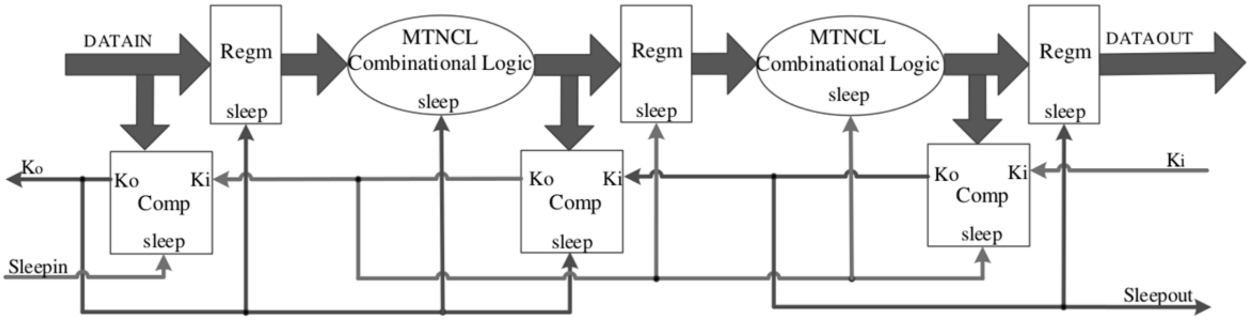

Figure 2. MTNCL Pipeline Structure

power, MTNCL eliminates the two special design requirements of NCL, i.e., input-completeness and observability. Also, MTNCL gates do not need hysteresis, which is required for NCL gates, because the NULL wavefront is generated directly by the sleep signal. Consequently, even with the introduction of two sleep transistors, MTNCL circuits are smaller in area than equivalent NCL circuits.

This paper focuses on the benefits of MTNCL in the delay units of an adaptive beamformer, specifically the FTD and CTD. The remainder of this paper will be divided into three major sections. The Approach section will give an overview of the design of the FTD and CTD. The Results section will provide a comparison between MTNCL and synchronous versions of the FTD and CTD. The Conclusion section will provide final thoughts as well as opportunities for future research.

Approach

Both the FTD and CTD were designed and simulated in ModelSim to verify functionality. Equivalent synchronous designs were also implemented to give a fair power comparison to the MTNCL paradigm.

FTD: The FTD unit itself is in fact a finite impulse response (FIR) filter. FIR filters are widely used in signal processing applications due to their stability and linear phase properties [3]. The FTD unit uses three major components in order to perform the discrete convolution: shift registers to create data taps, Dadda multipliers to multiply the constant coefficients with the data, and carry-select adders to calculate the final output. All numbers are in a fixed-point fractional 2's complement format; therefore, no overflow can occur during multiplication. All bits of precision are kept until the final stage where the output is truncated to 12 bits of precision. Figure 1 shows the FTD structure.

Similar to synchronous designs, MTNCL can be pipelined to increase the throughput of the circuit. The MTNCL pipeline architecture is shown in Figure 2. Each set of combinational logic is separated by MTNCL sleepable registers. MTNCL employs early completion detection; therefore, the data input to each register also connects to the corresponding completion detection unit. When a DATA wavefront passes into the register, the completion detection component outputs a *request-for-NULL* to the previous stage in the pipeline. As shown in Figure 2, there is no additional circuitry required for sleep signal generation because the handshaking signals can be used directly to sleep data when a NULL wavefront is received. In this manner, each combinational logic block alternates between DATA and NULL wavefronts.

The MTNCL shift register is made of alternating registers that are resettable to NULL and resettable to DATA, respectively. Typically, two MTNCL registers (one DATA-resettable and one NULL-resettable) would be sufficient for each tap in a shift register; however, to maximize the throughput, an additional two MTNCL registers were added to each tap. Without this optimization, the shift register must wait for all other data to finish calculating before shifting in new data, thereby reducing the throughput. Adding these registers balances the pipeline stages in the multipliers and adders with the number of stages in the shift register. This allows the DATA wavefront in the FTD to propagate shortly after new data is received.

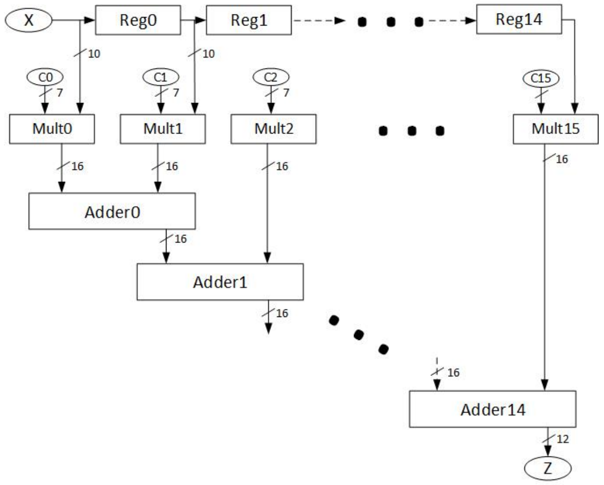

Figure 1. FTD Block Diagram

Table 2. IBM 130nm Design Comparison

Design	Active Energy (pJ/data)	Leakage Power (µW)	Speed (MHz)	Total Gate Width (mm)
MTNCL FTD	122	11.7	378.8	61.8
Syn. FTD	131	19.0	378.2	42.7
MTNCL CTD	12.1	0.871	369.9	4.51
Syn. CTD	8.21	0.665	370.0	1.78
MTNCL FTD+CTD	134.1	12.571		66.31
Synchronous FTD+CTD	139.2	19.665		44.48

The multiplier uses the Dadda algorithm to reduce the number of partial products into two sets which are then added using a carry propagate adder. According to the Dadda algorithm, each reduction stage uses the minimum number of half adders and full adders, resulting in a near optimal reduction [4]. The Dadda multiplier is pipelined into two stages to increase the throughput of the circuit to the GHz range. The first stage contains the partial product reduction, while the second stage contains only the carry propagate adder.

CTD: The CTD is a much smaller design than the FTD, consisting only of 16 shift registers, 4 data MUXes, and 8 MUXes for proper routing of completion detection signals. The MUXes allow any number of shift registers to be bypassed. This in turn allows for a different number of cycle delays between each channel. The input to the CTD is the 12-bit output of the FTD, and the output of the CTD is a 12-bit fractional number shifted by a varying number of cycles. Figure 3 shows the general structure of the CTD.

Results and Analysis

Both the FTD and CTD were implemented using the IBM 8RF 130nm bulk CMOS process. Equivalent synchronous designs were synthesized to run at the same speed as the MTNCL design in order to ensure that the power consumption comparison is fair. Both designs were simulated in Cadence MMSIM using a vector file for each synchronous design and an equivalent Verilog-A controller for the MTNCL designs.

The FTDs were given 100 random inputs (all test cases used the same 100 random values) and a sample set of coefficient values. The CTDs were also given the same 100 random inputs along with values to control the number of registers to skip (1, 2, 4, and 8). The average active energy per data was then found for the period when the pipeline was full.

The MTNCL FTD and CTD were simulated for leakage power by resetting the circuit, giving the circuit all DATA0 as inputs, followed by a NULL wave, and running the simulation for 1 ms. The leakage power of each synchronous design was measured after resetting the circuit and keeping all inputs as 0.

The 2 FTD designs and 2 CTD designs were compared for area, active energy, and leakage power. During logic synthesis, the timing constraint of synthesizing the synchronous designs was set to match the inherent speed of the MTNCL design to avoid overdesign. The same clock speed was also used in simulations. The results are listed in Table 2 above.

The data shows that the MTNCL designs were larger than their equivalent synchronous designs, but the MTNCL FTD is better in terms of active energy (7% less) and leakage power (62% less). The MTNCL CTD was worse in terms of active energy (47% more) and leakage power (31% more). This is due to the fact that MTNCL designs obtain most of their advantages in combinational logic, while the

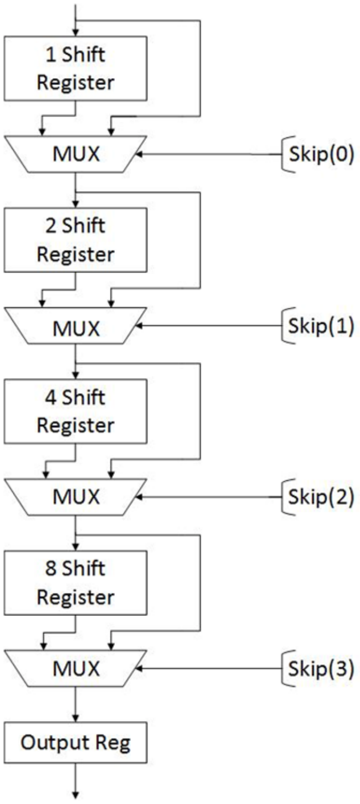

Figure 3. CTD Block Diagram

registers in MTNCL designs are often worse than synchronous registers in terms of power. The CTD is made up of mostly registers, so this design is not able to take full advantage of the benefits of MTNCL. However, the CTD is considerably smaller than the FTD and consumes less than 10% power when compared with the FTD. Therefore, with FTD and CTD combined as a beamformer channel, the power advantage of MTNCL still holds (3.7% overall saving in active energy and 36% saving in leakage power).

In addition, the MTNCL designs were implemented and simulated in the IBM 45nm PDSOI process to find the maximum throughput that could be achieved. For the FTD, the average throughput was 1.22 GHz, while the CTD had an average throughput of 1.89 GHz which would meet the criteria for GHz range throughput required by high-performance beamforming circuits.

Conclusion

In this paper, FTD and CTD digital beamforming circuits were designed using IBM's 8RF 130nm bulk CMOS process and simulated in Cadence MMSIM. Each circuit was designed using an MTNCL architecture and an equivalent synchronous architecture. Results show that MTNCL uses less power while still meeting the high-performance requirements of the adaptive beamformer.

Future work will include adapting the designs for fabrication in IBM's 45nm PD-SOI process. With a smaller process, MTNCL should exhibit an even greater power advantage over the synchronous designs especially in terms of leakage power.

Acknowledgements

This work is supported by the DARPA Arrays at Commercial Timescale (ACT) program.

References

1. K. M. Fant and S. A. Brandt, "NULL Convention Logic: A Complete and Consistent Logic for Asynchronous Digital Circuit Synthesis," Intl. Conf. on Application Specific Systems, Architectures, and Processors, pp. 261-273, 1996.
2. S. C. Smith and J. Di, "Designing asynchronous circuits using NULL convention logic (NCL)," Synth. Lect. Digit. Circuits Syst., vol. 4, no. 1, pp. 1–96, 2009.
3. R. Mehboob, S. A. Khan, and R. Qamar, "FIR filter design methodology for hardware optimized implementation," *IEEE Transactions on Consumer Electronics*, vol.55, no.3, pp.1669-1673, August 2009.
4. V. M. Dhivya, A. Sridevi, A. Ahilan, "A high speed area efficient FIR filter using floating point dadda algorithm," *2014 International Conference on Communications and Signal Processing (ICCSP)*, pp.1640-1644, April 2014.