

Standby Power Reduction Techniques for Asynchronous Circuits with Indeterminate Standby States

Liang Zhou and Scott C. Smith (contact author)

Department of Electrical Engineering, University of Arkansas, Fayetteville, AR
kingdom701@gmail.com and smithsco@uark.edu

Abstract—In the literature, some schemes were proposed to combine Multi-Threshold CMOS (MTCMOS) with asynchronous circuits to reduce standby power consumption. However, all of these can only be applied to asynchronous circuits in which the values of all signals can be determined in the standby state. As a result, their applications are limited. To solve this problem, this paper develops standby power reduction techniques which not only combine MTCMOS with asynchronous circuits, but also can be applied to asynchronous circuits with indeterminate standby states. Compared to asynchronous circuits implemented with all regular threshold transistors, the proposed implementation has significantly less standby power, reasonable speed penalty, and negligible area overhead.

Keywords—asynchronous circuits; Multi-Threshold CMOS (MTCMOS); NULL Convention Logic (NCL)

I. INTRODUCTION

With the current trend of semiconductor devices scaling into the deep submicron region, design challenges that were previously minor issues have now become increasingly important. Where in the past, dynamic, switching power has been the predominant factor in CMOS digital circuit power dissipation, recently, with the dramatic decrease of supply and threshold voltages, a significant growth in leakage power demands new design methodologies for digital integrated circuits (ICs). The main component of leakage power is sub-threshold leakage, caused by current flowing through a transistor even if it is supposedly turned off. Sub-threshold leakage increases exponentially with decreasing transistor feature size.

Among the many techniques proposed to control or minimize leakage power in deep submicron technology, Multi-Threshold CMOS (MTCMOS) [1], which reduces leakage power by disconnecting the power supply from the circuit during standby mode while maintaining high performance in active mode, is very promising. MTCMOS incorporates transistors with two or more different threshold voltages (V_t) in a circuit. Low- V_t transistors offer fast speed but have high leakage, whereas high- V_t transistors have reduced speed but far less leakage current. MTCMOS combines these two types of transistors by utilizing low- V_t transistors for circuit switching to preserve performance and high- V_t transistors to gate the

circuit power supply to significantly decrease sub-threshold leakage.

Quasi-delay-insensitive (QDI) asynchronous circuits [2] designed using CMOS exhibit an inherent standby behavior since they only switch when useful work is being performed; however, there is still significant leakage power during standby mode. In the literature, some schemes were proposed to combine multi-threshold CMOS (MTCMOS) with asynchronous circuits to reduce standby power consumption, as discussed below.

MTNCL [3-6] combines the MTCMOS technique with full-word pipelined NULL Convention Logic (NCL) asynchronous circuits [7] to sleep the NCL circuit during standby mode. It has significantly less standby power, higher speed, and requires less area than the original NCL circuits implemented with either all low- V_t or high- V_t transistors.

Bit-Wise MTNCL (BWMTNCL) [8] combines the MTCMOS technique with NCL circuits. Compared to original NCL circuits implemented with all low- V_t and high- V_t transistors, respectively, it provides the leakage power advantages of the all high- V_t NCL implementation with a reasonable speed penalty compared to the all low- V_t design, and has negligible area overhead.

Fine-grain leakage power reduction method in [9] combines the MTCMOS technique with asynchronous circuits synthesized with Boolean gates. It utilizes high- V_t transistors for the off-state transistors whose gate input signals are inactive in the standby phase. Compared to original asynchronous circuits implemented with regular- V_t transistors, it has significantly less standby power, reasonable speed penalty, and negligible area overhead.

Static power reduction techniques in [10] combine the MTCMOS technique with Pre-Charge Half Buffer (PCHB) [2] asynchronous circuits. Compared to original PCHB asynchronous circuits implemented with regular- V_t transistors, it has less standby power, reasonable speed penalty, and reasonable area overhead.

However, all of the above can only be applied to asynchronous circuits in which the values of all signals can be determined in the standby state. As a result, their applications

are limited. To solve the problem, this paper develops standby power reduction techniques which not only combine MTCMOS with asynchronous circuits, but also can be applied to asynchronous circuits with indeterminate standby states. Compared to asynchronous circuits implemented with all regular- V_t transistors, the proposed implementation has significantly less standby power, reasonable speed penalty, and negligible area overhead. Although the proposed techniques are illustrated by extending BWMTNCL, they can also be applied to other asynchronous circuit paradigms.

Section II provides an overview of NCL, BWMTNCL, and an NCL unsigned $32+16 \times 16$ Multiply and Accumulate (MAC) unit, as an example with indeterminate standby states. Section III details the proposed techniques to handle indeterminate standby states; Section IV compares the various implementations; and Section V provides conclusions.

II. PREVIOUS WORK

A. Introduction to NCL

NCL circuits utilize multi-rail logic, such as dual-rail, to achieve delay-insensitivity. A dual-rail signal, D , consists of two wires, D^0 and D^1 , which may assume any value from the set {DATA0, DATA1, NULL}. The DATA0 state ($D^0 = 1, D^1 = 0$) corresponds to a Boolean logic 0, the DATA1 state ($D^0 = 0, D^1 = 1$) corresponds to a Boolean logic 1, and the NULL state ($D^0 = 0, D^1 = 0$) corresponds to the empty set meaning that the value of D is not yet available. The two rails are mutually exclusive, such that both rails can never be asserted simultaneously; this state is defined as an illegal state.

NCL circuits are comprised of 27 fundamental gates [11]. These 27 gates constitute the set of all functions consisting of four or fewer variables. The primary type of threshold gate, shown in Fig. 1, is the TH m n gate, where $1 \leq m \leq n$. TH m n gates have n inputs. At least m of the n inputs must be asserted before the output will become asserted. NCL threshold gates are designed with hysteresis state-holding capability such that all asserted inputs must be de-asserted before the output will be de-asserted, as shown in Fig. 2. Therefore, a TH m n gate is equivalent to an n -input C-element [12] and a TH1 n gate is equivalent to an n -input OR gate. NCL threshold gates may also include a reset input to initialize the output. These resettable gates are used in the design of DI registers [13].

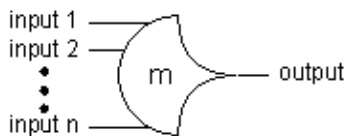


Figure 1. TH m n threshold gate.

NCL systems contain at least two delay-insensitive (DI) registers, one at both the input and at the output, and can be finely pipelined by inserting additional registers, as shown in

Fig. 3. Two adjacent register stages interact through their request and acknowledge signals, K_i and K_o , respectively, to prevent the current DATA wavefront from overwriting the previous DATA wavefront, by ensuring that the two DATA wavefronts are always separated by a NULL wavefront. The acknowledge signals are combined in the Completion Logic to produce the request signal(s) to the previous register stage, utilizing either the full-word or bit-wise completion strategy [13].

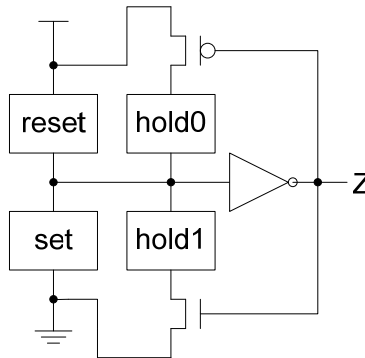


Figure 2. NCL threshold gate design.

To ensure delay-insensitivity, NCL circuits must adhere to the following criteria: Input-Completeness and Observability [14]. Input-Completeness requires that all outputs of a combinational circuit may not transition from NULL to DATA until all inputs have transitioned from NULL to DATA, and that all outputs of a combinational circuit may not transition from DATA to NULL until all inputs have transitioned from DATA to NULL. Observability requires that no orphans may propagate through a gate. An orphan is defined as a wire that transitions during the current DATA wavefront, but is not used in the determination of the output.

B. Introduction to Bit-wise MTNCL

In NCL systems without feedback loops, the inputs of each gate while in the standby state are determinate, since all circuit inputs will be NULL, which causes all Combinational Logic gates and the data inputs and outputs of all registers to be de-asserted, which in turn causes all Completion Logic gates to be asserted (i.e., request-for-data or *rfd*), as shown in Fig. 3.

The *leakage path* is defined as the path formed by “on” transistors and “off” low- V_t transistors in the standby state. To substantially reduce leakage power while degrading speed as little as possible, the following rules should be utilized to determine which transistors should be high- V_t and which transistors should be low- V_t [8]:

1. Determine threshold gate input and output values in standby state.
2. All transistors “on” in standby state should be low- V_t .
3. Replace the minimal number of “off” transistors with high- V_t transistors to eliminate leakage path, and replace the rest with low- V_t transistors.

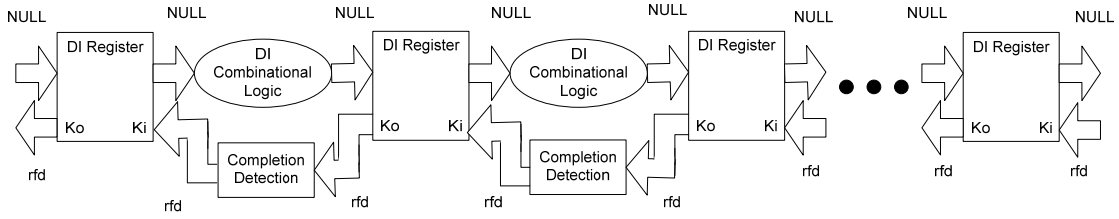


Figure 3. NCL system framework without feedback loops in standby state.

Fig. 4 shows the standby states of a 1-bit dual-rail NCL register, which consists of two TH22 resettable to '0' gates with $A = '0'$, $B = '1'$, reset = '0', and one inverted TH12 gate with both '0' inputs in standby state. After applying those 3 rules, the schematic of the TH22 gate is given in Fig. 5, in which high- V_t transistors are circled and low- V_t transistors are not. T_0 , T_1 , and T_2 are high- V_t because they do not switch except for initialization.

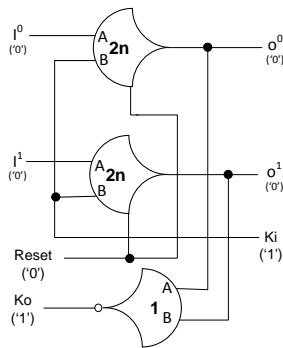


Figure 4. 1-bit NCL register in standby state.

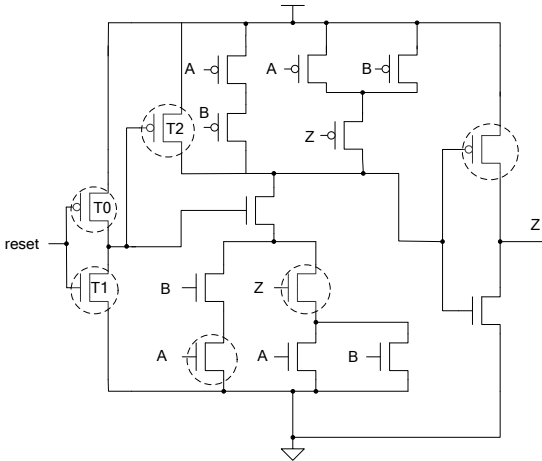


Figure 5. BWMTNCL applied to TH22 resettable to '0' gate with a single standby state: reset = '0', $A = '0'$, $B = '1'$, $Z = '0'$

Compared to original NCL circuits implemented with all low- V_t and high- V_t transistors, respectively, BWMTNCL

provides the leakage power advantages of the all high- V_t NCL implementation with a reasonable speed penalty compared to the all low- V_t design, and has negligible area overhead.

C. Asynchronous Circuits with Indeterminate Standby States

As an example of an asynchronous circuit with indeterminate standby states, an NCL unsigned $32+16 \times 16$ MAC is developed. As shown in Fig. 6, it consists of 3 parts: full-word pipelined 7-stage partial product generation and Wallace tree summation circuit (PP1, PP2), a 4-stage feedback loop which feeds back the accumulator as 2 partial products in carry-save form ($A1$, $A2$), and full-word pipelined 15-stage 30-bit Ripple-Carry-Adder. The architecture is elaborated in [15], except that it is full-word pipelined and extra control functions are removed for simplicity.

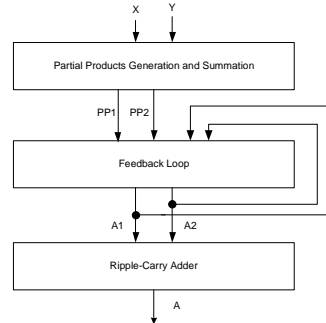


Figure 6. NCL MAC architecture

Fig. 7 shows the standby states of the feedback loop, which has 2-level carry save adders to sum up new partial products, $PP1$ and $PP2$, and old accumulator, $A1$ and $A2$, to generate new accumulator in carry-save form. Partial DATA in Fig. 7 means that some bits of the register are DATA while the other bits are NULL.

In standby state, the old accumulator values are stored in register, REG0, in Fig. 7. Each bit of REG0 can either be DATA0 or DATA1, which cannot be determined at design time. Similarly, register, REG1, Combinational Logic, COMB1 and COMB2, and Completion Logic, COMP0, also have indeterminate standby states. Therefore, BWMTNCL cannot be applied to this feedback loop.

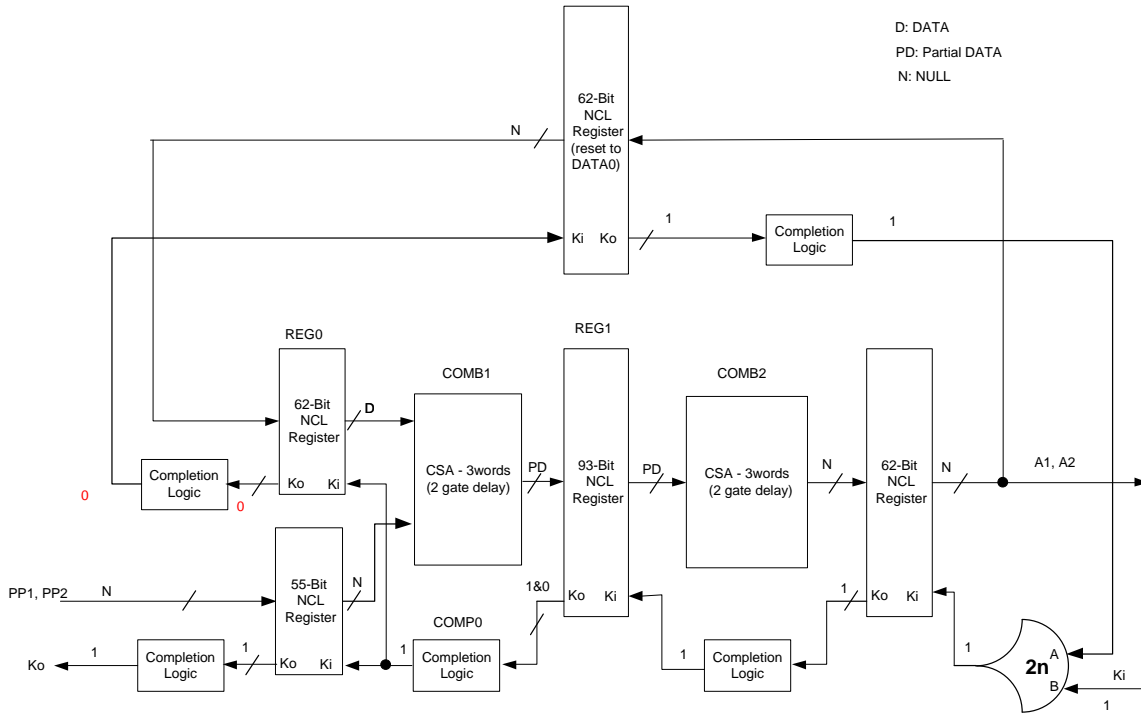


Figure 7. Standby states of the feedback loop without indeterminate states reduction

III. TECHNIQUES TO HANDLE INDETERMINATE STANDBY STATES

In asynchronous circuits with indeterminate standby states, usually not all of the inputs and outputs of a threshold gate are indeterminate in the standby state. For example, TH22 resettable to '0' gates used in REG0 in Fig. 7 have determinate standby states for A, B, and reset inputs (i.e., A = '0', B = '1', reset = '0'), but an indeterminate standby state for output, Z (i.e., Z = '1' or Z = '0'). In other words, it has 2 possible standby states. Instead of using an all high- V_t implementation, those 2 standby states can be analyzed to use the minimal number of high- V_t transistors to eliminate the leakage path in either of the two possible standby states, in order to substantially reduce leakage power while degrading speed as little as possible. The rules used in BWMTNCL can be enhanced as follows:

1. Determine the number of standby states and threshold gate input and output values in each standby state.
2. Replace the minimal number of transistors with high- V_t transistors to eliminate leakage path in any standby state, and replace the rest with low- V_t transistors.

After applying these 2 rules, the schematic of the TH22 gate with 2 possible standby states is given in Fig. 8.

By comparing Fig. 5 with Fig. 8, it can be observed that Fig. 8 has 3 more high- V_t transistors than Fig. 5. These 3 additional high- V_t transistors are required by the extra standby state and make the TH22 gate in Fig. 8 slower than in Fig. 5. Therefore, it is beneficial to reduce the number of gates with indeterminate standby states so that fewer high- V_t transistors are required, to degrade speed as little as possible.

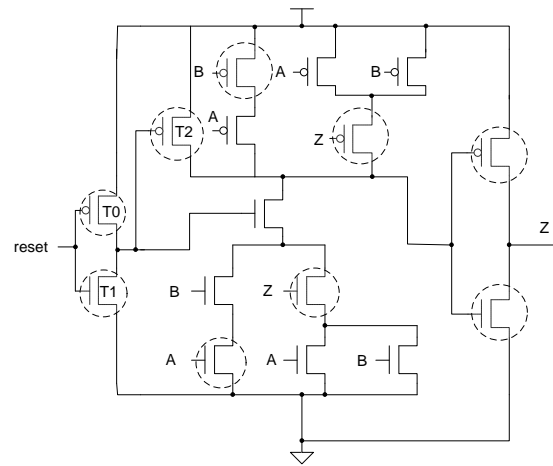


Figure 8. Enhanced BWMTNCL applied to TH22 resettable to '0' gate with 2 standby states: reset = '0', A = '0', B = '1', Z = '0' or Z = '1'

To reduce the number of gates with indeterminate standby states, an inverter U0 and an asymmetric [16] TH22 gate, U1, are added in Fig. 9 to control the K_i input of register, REG0. The input B with '+' of U1 only takes effect in asserting the asymmetric TH22 gate. In other words, U1 will be asserted if both inputs are '1', and de-asserted if A is '0' regardless of the value of B. The preceding partial product generation and Wallace tree summation circuit has output register, REG3, whose K_o output is asserted if PP1 and PP2 are NULL, and de-asserted if they are DATA.

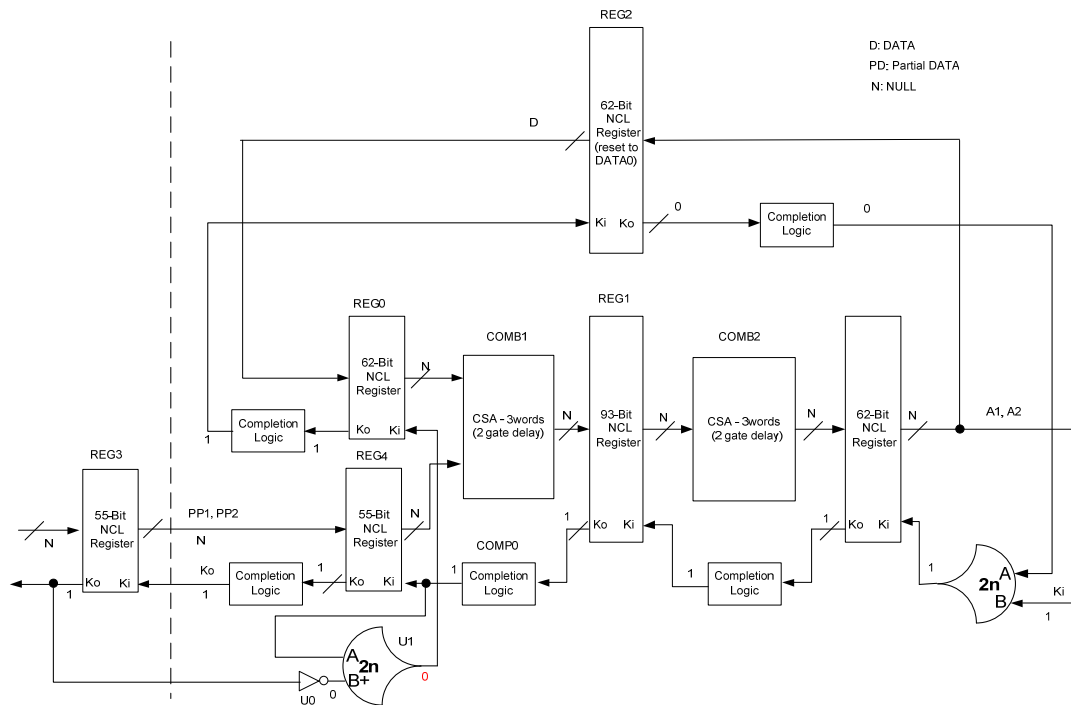


Figure 9. Standby states of the feedback loop with indeterminate states reduction

As a result, the condition for the K_i input of REG0 to be de-asserted remains that Completion Logic, COMP0, is de-asserted, but the condition to assert it is changed from COMP0 is asserted to COMP0 is asserted and $PP1$ and $PP2$ are DATA. The function of the feedback loop in the active state is not changed, because the added condition that $PP1$ and $PP2$ are DATA is always met in the active state. However, in the standby state, $PP1$ and $PP2$ are NULL so that the K_i input of REG0 cannot be asserted, and the old accumulator values are stored in REG2 instead of REG0. In Fig. 9, only REG2 and REG0 have indeterminate standby states, compared to REG0, REG1, COMB1, COMB2, and COMP0 in Fig. 7. As the number of gates with indeterminate standby states is significantly reduced, fewer high- V_t transistors are required, which degrades speed as little as possible.

The area overhead is negligible as only two gates are added. The added condition is equivalent to adding Combinational Logic with 1 gate delay between REG3 and REG4. As the speed of the whole circuit is limited by the slowest stage [13], which is the feedback loop which has much longer delay than the stage between REG3 and REG4, the speed of the whole circuit will not be degraded by adding those two gates.

Fig. 9 requires 4 types of gates with indeterminate standby states. First, REG2 requires TH22 gates resettable to '0' with standby states reset = '0', A = '0', B = '1', Z = '0' or Z = '1', which is already shown in Fig. 8. Second, REG2 requires TH22 gates resettable to '1' with standby states reset = '0', A = '0', B = '1', Z = '0' or Z = '1', which is shown in Fig. 10. Third, REG2 requires inverted TH12 gates with standby states A = '1', B = '0' or A = '0', B = '1', which is shown in Fig. 11. Finally, REG0 requires TH22 gates resettable to '0' with standby states

reset = '0', A = '0' or A = '1', B = '1', Z = '0', which is shown in Fig. 12.

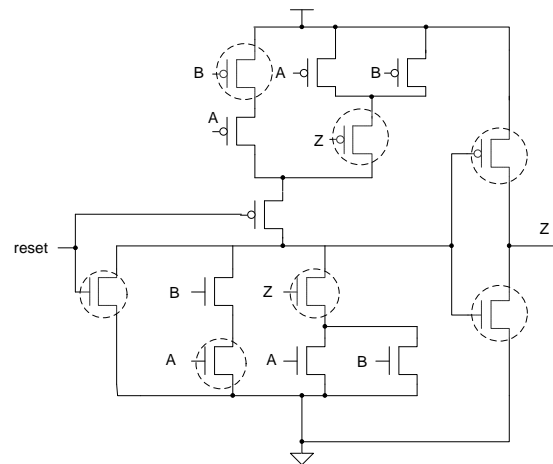


Figure 10. Enhanced BWMTNCL applied to TH22 resettable to '1' gate with 2 standby states: reset = '0', A = '0', B = '1', Z = '0' or Z = '1'.

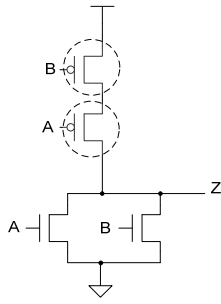


Figure 11. Enhanced BWMTNCL applied to inverted TH12 gate with 2 standby states: A = '0', B = '1' or A = '1' or B = '0'.

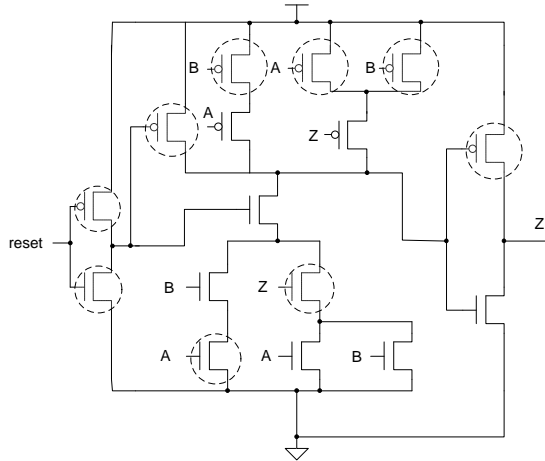


Figure 12. Enhanced BWMTNCL applied to TH22 resettable to '0' gate with 2 standby states: reset = '0', A = '0' or A = '1', B = '1', Z = '0'.

In summary, two techniques to handle indeterminate standby states are proposed. First, add extra gates to reduce indeterminate standby states as much as possible. Second, analyze the standby states of threshold gates and replace the minimal number of transistors with high- V_t transistors to eliminate the leakage path in any possible standby state, and replace the rest with low- V_t transistors.

IV. SIMULATION RESULTS

To compare the proposed BWMTNCL with indeterminate state reduction design to those without, the $32+16 \times 16$ unsigned MAC with feedback loop in Fig. 9 was implemented with enhanced BWMTNCL, and the MAC with feedback loop in

Fig. 7 was implemented with regular- V_t transistors, using the 1.2V IBM 8RF-LM 130nm CMOS process. All designs were simulated at the transistor level using Cadence's UltraSim simulator running a VerilogA controller in mixed-signal mode. Note that all transistors are minimum sized except for the buffers used for high fanout signals.

The first two rows of Table I show the results, in which T_{DD} is the average DATA plus NULL processing time, which is comparable to the clock period in a synchronous system. T_{DD} and Energy/Operation are calculated while the circuit is operating at its maximum speed, while Leakage Power is calculated using Cadence Spectre DC analysis after the pipeline is flushed with all NULL inputs. The results show that the proposed design has 36x standby power reduction over the regular design, with 20% speed penalty and 0.02% area overhead.

The 2 MACs were also implemented with all high- V_t and all low- V_t minimum sized transistors, respectively, as the all high- V_t implementation will give the lower bound on standby power and the all low- V_t implementation will give the lower bound on T_{DD} . The results are shown from row 3 to row 6 of Table I, which prove that adding extra gates to reduce indeterminate states does not increase T_{DD} . It also shows that the proposed design provides the leakage power advantage of the all high- V_t implementations with a reasonable speed penalty compared to the all low- V_t implementations.

Table I also shows that the proposed design is more energy efficient than the regular design, and the all low- V_t implementation is the most energy efficient. The relationship between energy/operation, threshold voltage, and transistor sizing will be investigated in the future.

V. CONCLUSIONS

None of the asynchronous circuit standby power reduction techniques in the literature are able to handle indeterminate standby states. This paper describes enhancements to the asynchronous circuit standby power reduction techniques developed in [8], to make it handle indeterminate standby states. Simulation proves that compared to the regular design, it has significant standby power reduction, reasonable speed penalty, and negligible area overhead.

TABLE I. SIMULATION RESULTS

	Transistor#	T_{DD} (ns)	Energy/ Operation(pJ)	Standby Power(nW)
Enhanced BWMTNCL with indeterminate states reduction	118176	6.2	36	159.312
Regular-V_t without indeterminate states reduction	118158	4.9	37.5	5761.8
All high-V_t with indeterminate states reduction	118176	9.3	35.3	129.096
All high-V_t without indeterminate states reduction	118158	9.3	35.3	130.56
All low-V_t with indeterminate states reduction	118176	4	34.5	13123.8
All low-V_t without indeterminate states reduction	118158	4	34.5	13255.8

References

- [1] S. Mutoh, T. Douseki, Y. Matsuya, T. Aoki, S. Shigematsu, and J. Yamada, "1-V Power Supply High-Speed Digital Circuit Technology with Multithreshold-Voltage CMOS," *IEEE Journal of Solid-State Circuits*, Vol. 30/8, pp. 847-854, August 1995.
- [2] A. J. Martin and M. Nystrom, "Asynchronous techniques for system-on-chip design," *Proceedings of the IEEE*, pp. 1089 – 1120, Vol. 94, No. 6, June 2006.
- [3] A. D. Bailey, J. Di, S. C. Smith, and H. A. Mantooth, "Ultra-Low Power Delay-Insensitive Circuit Design," *IEEE Midwest Symposium on Circuits and Systems*, pp. 503-506, August 2008.
- [4] A. D. Bailey, A. Al Zahrani, G. Fu, J. Di, and S. C. Smith, "Multi-Threshold Asynchronous Circuit Design for Ultra-Low Power," *Journal of Low Power Electronics*, Vol. 4/3, pp. 337-348, December 2008.
- [5] A. Alzahrani, A. D. Bailey, G. Fu, and J. Di, "Glitch-Free Design for Multi-Threshold CMOS NCL Circuits," 2009 *Great Lakes Symposium on VLSI*, May 2009.
- [6] S. C. Smith and J. Di, *Designing Asynchronous Circuits using NULL Convention Logic (NCL)*, Synthesis Lectures on Digital Circuits and Systems, Vol. 4/1, July 2009, Morgan & Claypool Publishers (doi: 10.2200/S00202ED1V01Y200907DCS023).
- [7] K. M. Fant and S. A. Brandt, "NULL Convention Logic: A Complete and Consistent Logic for Asynchronous Digital Circuit Synthesis," *International Conference on Application Specific Systems, Architectures, and Processors*, pp. 261-273, 1996.
- [8] L. Zhou, S. C. Smith, and J. Di, "Bit-Wise MTNCL: An Ultra-Low Power Bit-Wise Pipelined Asynchronous Circuit Design Methodology," *IEEE Midwest Symposium on Circuits and Systems*, pp. 217-220, August 2010.
- [9] M. Imai, K. Takada, T. Nanya, "Fine-grain leakage power reduction method for m-out-of-n encoded circuits using multi-threshold-voltage transistors." *IEEE Int. Symp. on Asynchronous Circuit and Systems*, pp. 209 – 216, 2009.
- [10] C. Ortega, J. Tse, and R. Manohar, "Static power reduction techniques for asynchronous circuits", *IEEE Int. Symp. on Asynchronous Circuit and Systems*, May 2010, pp. 52-61.
- [11] Gerald E. Sobelman and Karl M. Fant, "CMOS Circuit Design of Threshold Gates with Hysteresis," *IEEE International Symposium on Circuits and Systems (II)*, pp. 61-65, 1998.
- [12] D. E. Muller, "Asynchronous Logics and Application to Information Processing," in *Switching Theory in Space Technology*, Stanford University Press, pp. 289-297, 1963.
- [13] S. C. Smith, R. F. DeMara, J. S. Yuan, M. Hagedorn, and D. Ferguson, "Delay-Insensitive Gate-Level Pipelining," *Elsevier's Integration, the VLSI Journal*, Vol. 30/2, pp. 103-131, October 2001.
- [14] S. C. Smith, R. F. DeMara, J. S. Yuan, D. Ferguson, and D. Lamb, "Optimization of NULL Convention Self-Timed Circuits," *Integration, the VLSI Journal*, Vol. 37/3, pp. 135-165, August 2004.
- [15] L. Zhou and S. C. Smith, "Speedup of a Large Word-Width High-Speed Asynchronous Multiply and Accumulate Unit," *IEEE Midwest Symposium on Circuits and Systems*, pp. 499-502, August 2009.
- [16] S. B. Furber and P. Day, "Four-phase micropipeline latch control circuits," *IEEE Trans VLSI Syst.*, vol. 4, pp. 247–253, June 1996.