

Ultra-Low Power Delay-Insensitive Circuit Design

Andrew D. Bailey¹, Jia Di¹, Scott C. Smith², and H. Alan Mantooth²

¹Department of Computer Science and Computer Engineering

²Department of Electrical Engineering

University of Arkansas, Fayetteville, AR 72701

{adbaile, jdi, smithsco, mantooth}@uark.edu

Abstract—This paper presents a design methodology incorporating multi-threshold CMOS (MTCMOS) into delay-insensitive asynchronous circuits in order to solve the problems of the synchronous counterpart, e.g., *sleep* signal generation, storage element data loss during sleep mode, and sleep transistor sizing. Significant leakage power reduction has been demonstrated by simulation. Due to the flexible timing requirement feature of delay-insensitive circuits, sub-threshold operation can be achieved, which allows for further supply voltage scaling for ultra-low power.

I. INTRODUCTION

With the extreme scaling of transistor feature size and the growing demand of battery-powered mobile communication and computing devices, power consumption has become a major concern of integrated circuit designers. Among the three components of CMOS digital circuit power consumption, dynamic power has been the dominant figure in the past. As threshold voltage (V_t) lowers with shrinking transistor size, leakage power becomes more important. For applications requiring ultra-low power consumption, leakage power must be under control. Various techniques have been proposed for this purpose. Multi-Threshold CMOS (MTCMOS) uses low- V_t transistors for logic function and high- V_t transistors to gate leakage current in sleep mode. Unfortunately, three serious drawbacks hinder the widespread use of MTCMOS, as detailed in Section II. This paper presents a design methodology utilizing delay-insensitive asynchronous logic in conjunction with the MTCMOS technique to eliminate all these drawbacks. In addition, sub-threshold operation can be achieved to further reduce power. Promising results and analysis, as well as future work, are discussed.

II. BACKGROUND

A. MTCMOS

It is well-known that lowering V_{DD} is the most efficient approach to reducing power consumption. However, a lower V_{DD} causes insufficient gate overdrive, which in turn causes increased delay. In order to reduce power while maintaining performance, transistor threshold voltage, V_t , needs to be scaled with V_{DD} . Although current fabrication technology has significantly reduced threshold voltage, this decrease causes an exponential increase in leakage power, due to the increase in sub-threshold leakage. To reduce the leakage, MTCMOS

involves using high- V_t transistors to gate power and ground of a low- V_t logic block, as shown in Figure 1 [1]. When the high- V_t transistors are turned on, the low- V_t logic is connected to virtual ground and power, and switching is performed through fast, low- V_t devices. When the circuit enters sleep mode, the high- V_t gating transistors are turned off, resulting in a very low sub-threshold leakage current from V_{DD} to ground. MTCMOS is a very attractive technique for reducing sub-threshold leakage current during standby mode because existing designs can be easily modified into MTCMOS blocks by simply adding high- V_t power supply switches/transistors; and the semiconductor fabrication process only requires an additional implant processing step to provide the extra threshold voltage level [1]. However, the three drawbacks of synchronous MTCMOS circuits stated above need to be eliminated before this technique can be widely adopted [1]: 1) the generation of “sleep” signals is timing critical, often requiring complex logic circuits; 2) sequential circuits will lose data when the power transistors are turned off; and 3) proper sizing of the sleep transistors in a large circuit is a very difficult task.

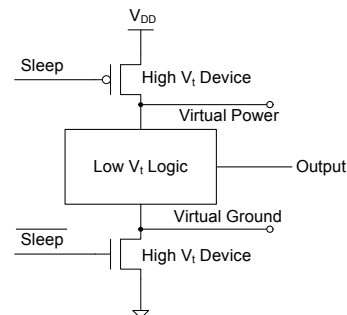


Figure 1: MTCMOS architecture

B. Delay-Insensitive Circuit and NULL Convention Logic

Digital logic can be classified into two categories: synchronous logic and asynchronous logic. Although for the past four decades synchronous logic has been dominating the IC market, recently asynchronous logic has drawn more attention. One of the major reasons is the clock management problem for the increasingly complex synchronous circuits.

Asynchronous circuits are clock-free. Delay-insensitive asynchronous logic uses handshaking protocols rather than

clocks to control the circuit behavior. After each functional block finishes the current operation, it sends a request signal to its previous stage indicating it is ready for the next operation. Without receiving this signal, the previous block will not assign the next operation to the succeeding block. On the other hand, if the previous block receives the request signal before it finishes its current operation, it will hold the assignment until its current operation is done. Theoretically, delay-insensitive circuits have the ability to function correctly as long as the transistors can switch properly. The advantages of delay-insensitive circuits over their synchronous counterparts include no clock skew, average case performance, high energy efficiency, flexible timing requirement, and low noise/EMI.

NULL Convention Logic (NCL), a quasi-delay-insensitive paradigm, is a symbolically complete logic, which expresses a process completely in terms of the logic itself. Logic signals in NCL circuits are usually encoded in a multiple-rail format. The simplest encoding scheme is dual-rail logic, which uses two wires to interpret one signal value. The truth table of dual-rail logic is shown in Table 1 [2]. There are two valid states: DATA state, including DATA 0 and DATA 1, and NULL, represented by both wires being logic low.

TABLE 1: DUAL-RAIL TRUTH TABLE

	Wire 1	Wire 0
DATA 1	1	0
DATA 0	0	1
NULL	0	0
Invalid	1	1

The NCL logic family is composed of threshold gates. Figure 2 shows an m -threshold, n -input threshold gate, denoted as TH mn , where the output is asserted when at least m of the n inputs are logic high. Because NCL threshold gates are designed with hysteresis, once the output is asserted, it remains asserted until all n inputs are logic low. NCL threshold gates are an extension of the C-element (equivalent to a TH nn gate) commonly used in other delay-insensitive circuit designs. The operation of NCL circuits includes DATA-NULL cycles: after a DATA state, all signals in a stage go to a NULL state before the next DATA state [2]. This DATA-NULL cycle operation, together with the hysteresis property of threshold gates, fits the needs of a feedback request/acknowledge signal based handshaking protocol, which eliminates the effect of gate delay on circuit function, making NCL circuits speed-independent.

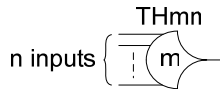


Figure 2: TH mn gate

A general circuit structure of threshold gates is shown in Figure 3 [3]. The “Go-to-NULL” block forces the output to ‘0’ when activated; the “Hold-NULL” block makes sure the output keeps ‘0’ when the number of logic high inputs doesn’t meet or exceed the gate’s threshold; the “Go-to-DATA” block forces the output to ‘1’ when the gate’s threshold is met; and the “Hold-DATA” block makes sure the output keeps ‘1’ until all inputs return to ‘0’.

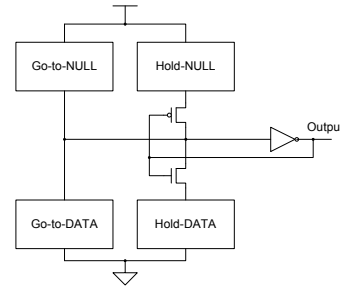


Figure 3: General circuit structure of threshold gates

III. DESIGN METHODOLOGY

A. Incorporate MTCMOS to NCL Threshold Gates

In the proposed approach, the MTCMOS technique has been incorporated into NCL circuits at the gate-level, by modifying the threshold gate library for ultra-low power operation. There are two reasons for this decision. First, threshold gates are basic building blocks of NCL circuits, which are more complex and powerful than Boolean gates, requiring fewer gates to implement the same logic function; hence, implementing MTCMOS in threshold gates will not result in a large area overhead. This also facilitates an automated NCL design flow. Second, implementing MTCMOS in each threshold gate eliminates the MTCMOS drawback of proper transistor sizing for large logic blocks, since it is easier to size transistors for each gate, and once the NCL MTCMOS library has been completed, it can be utilized for any arbitrary design.

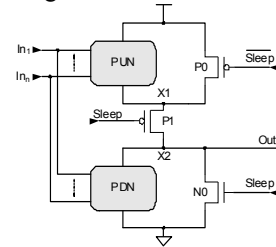


Figure 4: Generic CMOS ultra-low power gate

One method of designing MTCMOS Boolean gates is shown in Figure 4 [4], where there are n inputs feeding the Pull-Up Network (PUN) and Pull-Down Network (PDN). The sleep circuitry consists of three transistors – high- V_t PMOS P1, low- V_t PMOS P0, and low- V_t NMOS device N0, controlled by *Sleep* signal and its complement. In normal operating mode, *Sleep* is off (logic 0), causing transistors P0 and N0 to be OFF and transistor P1 to be ON. The circuit behaves exactly as a normal CMOS circuit. In sleep mode, *Sleep* is ON (logic 1), causing transistors P0 and N0 to be ON and transistor P1 to be OFF. Since P0 is ON, the PUN is now between two points at V_{DD} , and hence no leakage current should flow through the PUN. Similarly, the PDN is now between two points at equal voltage potential, GND , and hence no leakage current should flow through the PDN. Since *Out* is connected to X2, during sleep mode the output value will always be logic 0. The leakage loss occurring during

sleep mode will only be through transistor P1, which is turned OFF and is a high- V_t device; therefore, leakage will be minimized [4].

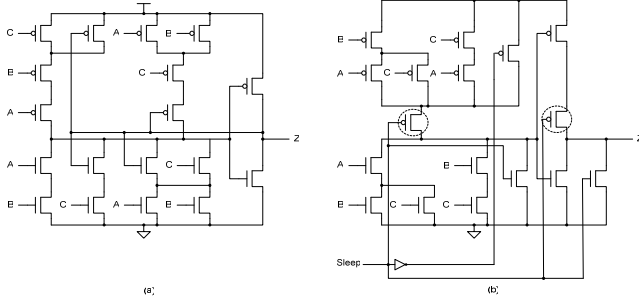


Figure 5: Static TH23 gate (a) original version (b) ultra-low power version (circled transistors are high- V_t ones).

A similar circuit structure can be applied to NCL threshold gates. Figure 5 shows the transistor schematic of an original static TH23 gate (a) and its ultra-low power version (b). The basic concept of applying MTCMOS to threshold gates is very similar to Boolean gates, with the exception of the following: 1) since threshold gates have inverters at the outputs, the MTCMOS structure needs to be applied to the output inverter as well as the pull-up and pull-down logic; 2) since the threshold gate output will be forced to logic 0 during sleep mode, the “Go-to-NULL” and “Hold-DATA” blocks are no longer needed, thus reducing area overhead of the MTCMOS NCL technique. In fact, for some threshold gates, e.g., TH24, the gate size is actually smaller after converting to MTCMOS.

B. Ultra-Low Power NCL Circuit Architecture

To build a NCL circuit with the ultra-low power NCL threshold gates, it is very important to consider how to generate the *Sleep* signal for each gate. *Sleep* signal generation is a major concern for synchronous MTCMOS circuits, requiring additional logic with carefully analyzed timing to avoid glitches and possibly circuit malfunction. However, this is not a problem for NCL circuits.

Figure 6 shows the general NCL circuit architecture, consisting of pipeline stages with combinational logic, asynchronous registers/latches, and completion detection circuitry. 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 Detection circuitry to produce the request signal(s) to the previous register stage. When all current register outputs are DATA, the corresponding completion detection signal will be logic 0, indicating a “request-for-NULL”; and when all current register outputs are NULL, the corresponding completion detection signal will be logic 1, indicating a “request-for-DATA”. After receiving the request signal, the previous register will allow the corresponding NULL/DATA wavefront to pass to the combinational logic block between the two registers. This handshaking protocol coordinates

NCL circuit behavior, analogous to coordination of synchronous circuits by a clock signal.

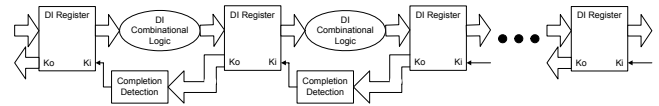


Figure 6: NCL circuit architecture

During a NULL cycle, the outputs of all threshold gates in the combinational logic block become logic 0; therefore, it is appropriate to force these gates to logic 0 using the sleep mechanism, in order to reduce switching power and leakage. **The corresponding completion detection signal naturally serves as the inverted *sleep* signal to the NCL gates, without requiring additional sleep generation hardware.** Moreover, due to the delay-insensitive nature of NCL circuits, there is no timing requirement as to what order the completion detection signal arrives at the previous register versus the combinational logic block. Therefore, hardware overhead as well as timing analysis requirements is greatly reduced when applying MTCMOS to NCL versus synchronous circuits. Additionally, NCL circuits do not lose data during sleep mode, since sleep mode is applied in lieu of the NULL cycle, which has the same effect, causing all gates to return to zero. Note that the data has already been latched by the subsequent register prior to the NULL cycle / sleep mode. Thus, the remaining two drawbacks in MTCMOS synchronous circuit are eliminated.

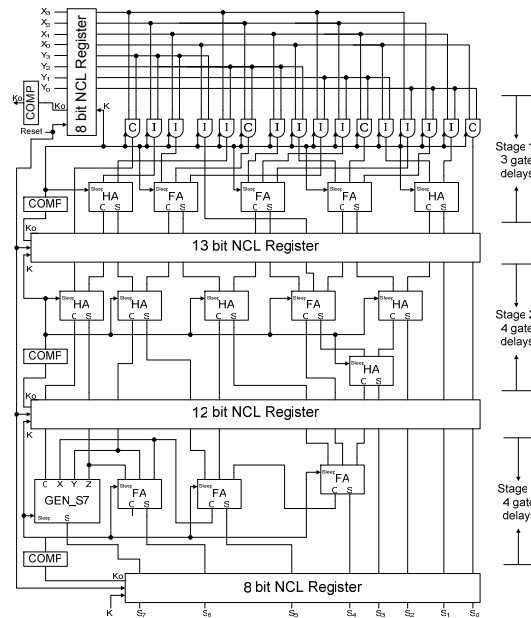


Figure 7: 3-stage NCL 4x4 multiplier

IV. RESULTS AND ANALYSIS

A. Testing Circuits and Setup

The testing circuit is a 4x4 NCL array multiplier, pipelined into 3 stages, consisting of full adders, half adders, input-complete ANDs, input-incomplete ANDs, registers, completion-detection units, and a S_7 generation block [5].

Both original version (uses low V_t transistors only) and MTCMOS version (as shown in Figure 7) were designed. The process applied is IBM 8RF-LM 0.13 μ m available at MOSIS. The threshold voltages of high-/low- V_t NMOS and PMOS transistors are around 0.30V/0.16V and -0.36V/-0.28V, respectively. The original supply voltage is 1.2V. Both circuits were simulated in Cadence Ultra-Sim.

B. Results

A series of simulations has been performed for the comparison of leakage power, propagation delay, and the trends of these two parameters during extreme V_{DD} scaling. The results are shown in the figures below.

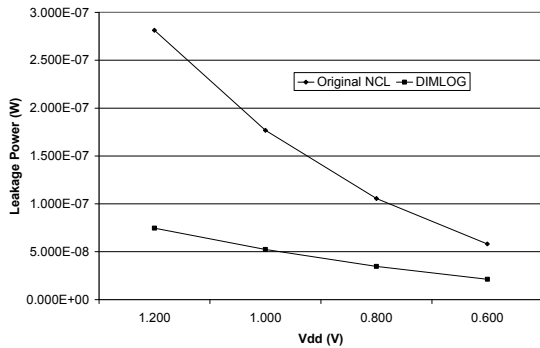


Figure 8: Leakage power comparison

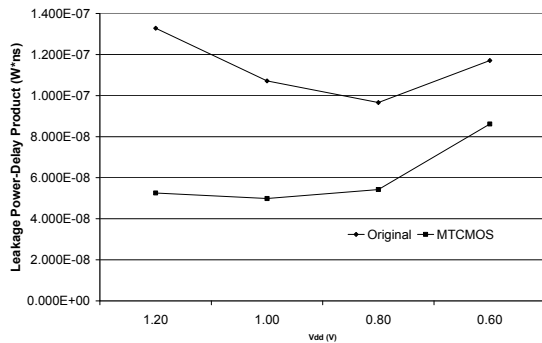


Figure 9: Leakage power - delay product comparison

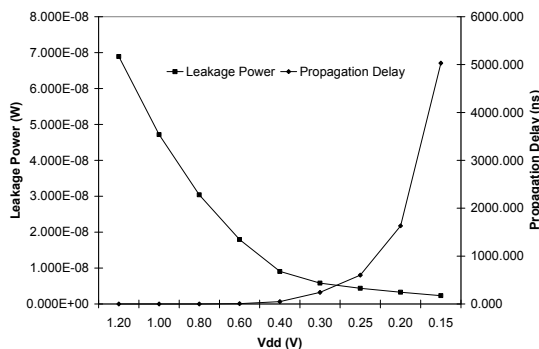


Figure 10: Delay and leakage power trends with V_{DD} scaling to sub-threshold operation

C. Analysis

From Figures 8-10 it is clear that MTCMOS NCL design has significant leakage power savings over the original version. In Figure 8, the leakage power of the original design is on average 3.5 times higher than the MTCMOS version. This is due to the use of high- V_t sleep transistors with bypassing low- V_t transistors. Since adding these additional devices will cause higher propagation delay, leakage power - delay product is a comprehensive measurement of the efficiency of MTCMOS NCL. As shown in Figure 9, MTCMOS design is still better than the original version, by a factor of two on average.

Figure 10 shows the extreme supply voltage scaling of the MTCMOS NCL circuit to sub-threshold operation. It shows that the MTCMOS NCL design is able to function correctly at 0.15V V_{DD} , which is already in sub-threshold operation region. This leads to 25 \times further leakage power saving. The propagation delay, as predicted, increases exponentially as V_{DD} lowers.

V. CONCLUSION

A design methodology incorporating MTCMOS into the delay-insensitive NCL paradigm for ultra-low power consumption is presented. Due to the handshaking protocol and delay-insensitive nature, drawbacks in the MTCMOS synchronous counterpart can be eliminated. Results show significant leakage power reduction and reasonable delay overhead, which result in an improved leakage power - delay product. In addition, delay-insensitivity allows for sub-threshold operation, which enables further capability to trade delay for low power consumption. Once the MTCMOS NCL cell library is built, the proposed design methodology can be applied to design ultra-low power ASICs. Future work includes reducing area/power and improving the circuit speed by using the bit-wise completion scheme [5], eliminating delay sensitivities introduced when applying MTCMOS by utilizing Early Completion [6], and tuning the sizes of transistors in each gate to achieve optimum performance.

REFERENCES

- [1] J. Kao and A. P. Chandrakasan, "Dual-Threshold Voltage Techniques for Low-Power Digital Circuits," *IEEE Journal of Solid-State Circuits*, Vol. 35, NO. 7, July 2000.
- [2] K. Fant and S. A. Bradt, "Null Convention Logic: a complete and consistent logic for asynchronous digital circuits synthesis". *Proceedings of International Conference on Application Specific Systems*, 1996, pages 261-273.
- [3] G. E. Sobelman and K. Fant, "CMOS Circuit Design for Threshold Gates with Hysteresis," *IEEE International Symposium on Circuits and Systems*, Vol. 2, page 61-64, May 1998.
- [4] P. Lakshmikanthan, K. Sahni, and A. Nunez, "Design of Ultra-Low Power Combinational Standard Library Cells Using A Novel Leakage Reduction Methodology," *2006 IEEE International SoC Conference*, Sep. 2006.
- [5] 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.
- [6] S. C. Smith, "Speedup of Self-Timed Digital Systems Using Early Completion," *IEEE Computer Society Annual Symposium on VLSI*, pp. 107-113, April 2002.