Self-Timed Pipeline Register Operating at Near-Threshold Voltage

Tomoki OGAWA and Makoto IWATA
Graduate School of Engineering, Kochi University of Technology,
Kami, Kochi, 782-8502 Japan

Abstract—In recent years, steady improvement on both performance and energy efficiency of LSI systems by virtue of the miniaturization of CMOS process has been harder. However, lower power devices are still demanded, especially in growing IoT market. Although near-/sub-threshold voltage operation is one of promising ultra-low-power techniques, there are critical issues on exponential performance degradation, huge performance variations, and high probability of functional failures. In order to solve those issues, this study focuses on the self-timed pipeline circuit which has inherent robustness against circuit delay variation.

In this study, a self-timed data-transfer control circuit and pipeline register operable under sub-threshold voltage is proposed. Through SPICE simulation of self-timed control circuit, called C-element, an optimal CMOS transistor sizing method is discussed to help C-element and pipeline register more robust under sub-threshold voltage region. SPICE simulation results of 65nm CMOS circuit reveal that the circuit can correctly operate in 0.1 V steps under nominal voltage, variations of process, voltage, and temperature.

Keywords: Near-/Sub-threshold voltage, C-element, self-timed pipeline

1. Introduction

Miniaturization of semiconductor process along with Moore’s Law has brought great benefit to modern computing devices, but the increase of power consumption of CMOS circuits has become a serious problem with denser integration of semiconductor transistors. However, lower power devices are still demanded, especially in growing IoT market.

In order to reduce the power consumption significantly, near-threshold computing (NTC) in which the circuit operates at near or less threshold voltage has been recently studied [1], [2]. This is because power consumed in CMOS circuit is proportional to the square of the voltage. However, the problem is often caused by lowering power supply voltage (VDD). Major critical problems on NTC are as follows.

(1) Performance Degradation.

The performance of CMOS circuit decreases in proportion to the supply voltage when it is over the threshold voltage (Vth) of the CMOS transistor. However, the performance is degraded exponentially when near- or sub-threshold voltage.

(2) Increase performance Variation.

Drivability of CMOS transistor usually depends on Vth, VDD, and temperature, especially in the near-threshold region. As a result, NTC will bring considerable variations in circuit performance.

(3) Function Failure.

Similar to the performance variation, variations in process, voltage, and temperature have a significant impact on the occurrence probability of function failures in the circuit, especially at the near-threshold voltage region. In order to eliminate the function failures, circuit design margin have to be considered to guarantee plenty of space at the sacrifice of its performance. As a result, it may induce the increase of leakage energy per operation.

In case of the modern synchronous circuit, clock distribution to the whole chip area tends to lengthen the wires. The longer wiring delay might induce more malfunction due to the clock skew. In contrast, the self-timed pipeline (STP) circuit [3] controls data-transfer within the pipeline by using hand-shake signals between adjacent pipeline stages. STP is hence designed to wire only adjacent pipeline stages, so that the influence of the wire delay variation induced by various process and environmental variations could be localized. The design margin could be also minimized for sub-threshold voltage operation. It can be said that the STP itself has an inherent robust feature against various variability of the CMOS transistor. This smart feature implies that the STP has a potentiality to operate without malfunction under the dynamic voltage scale control, from super- to near-/sub-threshold voltage. This is because the voltage scaling could be regarded as a sort of voltage variation, as long as the supply voltage is gently scaled [4].

In order to achieve significant power reduction, this study aims to investigate an optimal circuit design of the STP operating at near-threshold voltage. At first, the behaviors of typical STP circuit is characterized under various voltage and temperature conditions. After that, optimal transistor sizing method for self-timed data-transfer control circuit, called C-element, is proposed. Finally, widely operating conditions of the self-timed pipeline register including the C-element which is designed by 65nm CMOS transistors, will be revealed through SPICE simulation results.
2. Low-Voltage Characteristics of STP

The influences of PVT variations appear to be worse relatively when lowering the supply voltage of general CMOS circuit to sub-threshold region, in comparison with recommended supply voltage operation. The self-timed pipeline circuit (STP) has a potentiality to moderate those influences because of its localized wiring between pipeline stages.

Figure 1 shows a basic structure of self-timed pipeline where every pipeline stage is composed of a data latch, function logic, and data-transfer control circuit called C-element shown in figure 2.

![Figure 1: Basic structure of STP.](image1)

![Figure 2: Basic logic circuit of C-element.](image2)

Figure 3 illustrates a typical timing sequence of the STP. The STP circuit usually operates as follows.

1. Master reset: At first, a master-reset signal initializes all states of C-elements. At that time, every handshake (send and ack) signal is set to 1 (high).

2. Ready to transfer data: When a data packet hold at the stage $i-1$ is ready to be transferred, its data-latch open signal $CP_{i-1}$ is asserted and send$_i$ signal is set to 0 (low) to begin to transfer the data. Then, the stage$_i$ negates ack$_i$ as an acknowledge signal.

3. Completion to transfer data: Then, the stage$_{i-1}$ asserts send$_i$ for completion of the data transfer. After that, the stage$_i$ asserts the $CP_i$ to receive the data and the stage then asserts ack$_i$. It implies that the stage$_i$ is ready to transfer the data to the next stage$_{i+1}$ as well.

Every pipeline stage in the STP operates with the above hand-shake protocol. Therefore, even if an adjacent pipeline stage might be delayed due to some PVT variation, the data can be adaptively transferred based on the hand-shake protocol, where the time from the positive edge of the $CP_i$ to that of the $CP_{i+1}$ represents the data-transfer time between the two stages and where the time between a positive edge of the $CP_i$ to the next positive edge of that represents the period of consecutive data transfer at the stage$_i$. Furthermore, it is noted that the STP stage realized by the CMOS transistors consumes active power only when the stage is transferring and processing the data. This means that the STP does not require additional clock gating technique to save its dynamic power of clock distribution.

The article [1] reported critical issues under near-threshold voltage, where the nominal voltage is 1.1 V and the near-threshold voltage is 400 mV. In this case, performance of the fanout-of-four inverter in industrial 40 nm CMOS is degraded in 1/10th. Performance variations of transistor switching speed increases in 20x, in which process variations increase in 5x, the sensitivity due to temperature and supply ripple increase in 2x respectively. Function failure rate increases in 5 orders of magnitude.

Those issues are related to characteristics of individual CMOS gate composing the CMOS circuit. In terms of circuit design for the pipeline operating at near-threshold voltage region, there are two counter methods, synchronous and asynchronous. As for the synchronous pipeline, its clock cycle must be adjusted to the longest delay in the pipeline, i.e., it might be occurred under the worst case in PVT variations. The longest path such as clock distribution tree or forwarding mechanism might exist among several pipeline stages. In such case, the worst case design requires larger margin to reduce its function failure rate in sacrifice of the performance.

As for the self-timed pipeline, a kind of asynchronous pipeline, the influence of performance variations and the function failure rate can be limited to a single pipeline stage and suppressed in comparison of that of the synchronous pipeline. Thanks to the hand-shake protocol of the STP, the operating speed of the C-element alters autonomously even if the PVT variations affect the delay time of the CMOS cells.
composing the C-element. As a result, the average pipeline throughput is restricted by the slowest pipeline stage in the STP. As with the synchronous pipeline, setup and hold time of data latches must be guaranteed in the STP so that each delay time of send and ack signals have to be adjusted to the longest latency of the critical path in its corresponding stage. Furthermore, the SR latch in the C-element might oscillate or fall into a meta-stable state when both input signals of the RS latch change within a short time. A systematic design to avoid those malfunction under near-threshold voltage will be required with consideration of PVT variations. Furthermore, in order to realize the dynamic voltage scaling including sub-threshold voltage region, lower limit of supply voltage must be settled. This is because the energy minimum voltage around the sub-threshold region must exist due to the exponential increase of leakage power beyond the reduction of switching power.

3. NTC-Oriented C-Element

In order to ensure correct operation of the C-element under near-threshold voltage, there are two major issues to be solved. The first one is that the setup and hold time violation of the data latch must be avoided even if the transistor delay increases exponentially. The second one is that the electric potential on cyclic paths in the C-element must be kept steady at high potential, VDD, or low potential, VSS. The first problem is solved by adjusting the delay buffers, D_{c}, D_{b}, D_{s}, and D_{d}, of the C-element.

In general, forwarding latency required to transfer valid data from one set of data-latches to a set of data-latches in the succeeding stage is calculated by the sum of response time of the data latch τ_{q}, delay time of a critical path in the logic τ_{cp}, and setup time of the data latch τ_{setup}. Thus, handshake time of STP must be adjusted to the forwarding latency. As explained in Section 2, the original STP circuit operates based on the 4-phase handshake protocol so that latency time T_{f} required from the first to the third phase of the protocol has to exceed the forwarding latency. After completing the data-transfer, the data latch has to receive the next data correctly. Therefore, backward latency time T_{r} required for the fourth phase must exceed hold time of the data latch τ_{hold}.

\[
T_{f} \geq \tau_{q} + \tau_{cp} + \tau_{setup} \quad (1)
\]

\[
T_{f} \geq \tau_{hold} \quad (2)
\]

Using the two parameters and, it is then possible to define pipeline throughput as \(1/(T_{f} + T_{r})\) and pipeline efficiency as \(T_{f}/(T_{f} + T_{r})\). Pipeline throughput is a measure of packet flow rate through the pipeline. Pipeline efficiency is the proportion of net processing time spent on packet processing in terms of pipeline throughput.

The lower the supply voltage is set, the weaker the driving power of the output port of CMOS logic cell becomes. Thus, it is hard for output electric potential of the CMOS cell to keep steady VDD or VSS under the near-threshold voltage. The following techniques can be introduced as a candidate solution to tune the driving power of the related CMOS cells.

1) Replacement: to replace the target cell to a stronger cell which generates more driving current.
2) Multiplication: To multiplicate nMOS/pMOS transistors constructing the target cell.
3) Widening: To widen gate-width of nMOS/pMOS transistors of the target cell.

In terms of the technological difficulty of those tuning techniques, the first one is tried to be applied to C-element circuit design by using 65nm SOI-CMOS process. At first, the basic logic of the C-element is synthesized with zero capacitive load for every cell. After that, the critical cells are replaced to suitable stronger cells.

Figure 4 is a diagram showing the drivability of CMOS cells in the C-element which is synthesized with zero capacitive load for every cell. In the figure, each number described on individual cell denotes the relative drivability of the cell.

![C-element synthesized with zero capacitive load](image)

Figure 5 shows a C-element CMOS circuit modified with large cell replacement.

![C-element modified with large cell replacement](image)

4. Evaluation

The robust operation of the STP register with the modified C-element is verified by using SPICE simulation with the 65nm SOI-CMOS process libraries. As a preliminary verification, the two kinds of single C-element CMOS circuit designed by using standard threshold voltage transistor (SVT) and low threshold one (LVT) are
simulated respectively. The operating conditions are assumed as listed in table 1. The supply voltage range is set from 0.1 V to 1.2 V in 0.1 V steps, the temperature is $-20^\circ$C, $25^\circ$C, and $75^\circ$C, the process corner of nMOS and pMOS transistors is slow-slow (SS), typical-typical (TT), and fast-fast (FF).

Table 1: Transistors and PVT conditions in SPICE.

<table>
<thead>
<tr>
<th>Process Corner</th>
<th>Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS, TT, FF</td>
<td>$-20^\circ$C, $25^\circ$C, $75^\circ$C</td>
</tr>
</tbody>
</table>

In the SPICE simulation, the data transfer time ($T_i$) from receiving a data packet to sending the packet is measured as well as the average power during the data-transfer time ($P_i$). Precisely describing that in the figure 3, it is the time from when send$_{i-1}$ is negated to when ack$_i$ is asserted. That is, $T_i$ indicates about $T_f \times 2$. Therefore, the energy ($E_i$) consumed in the C-element to operate a data packet can be calculated by $E_i = P_i \times T_i$. In this measurement, the effective signal level is set to over 90 % of VDD as the high level signal and to under 10 % of VDD as the low level signal. The used SPICE simulator is Hspice, Synopsys Inc.

At first, the measurement results of the SVT-based C-element circuit at $-20^\circ$C, $25^\circ$C, and $75^\circ$C is shown in figure 6, 7, and 8 respectively. The SVT-based C-element can operate well without function failure, except for case of 0.1 V and the SS process condition. As discussed in section 2, the energy/operation decrease in proportion to the square of VDD and the data-transfer time in case of 0.1 V is degraded about 5 order of magnitude in comparison with the 1.2 V.

At first, the measurement results of the LVT-based C-element circuit at $-20^\circ$C, $25^\circ$C, and $75^\circ$C is shown in figure 9, 10, and 11 respectively. Compared with the SVT circuit, the LVT-based C-element can hardly operate in some cases of 0.1 V.

As discussed in section 2, there is the energy minimum point around the near-threshold voltage due to the exponential increase of leakage power. However, the minimum point cannot be observed in the preliminary measurements. Therefore, the supply voltage range around the sub-threshold is finely set from 0.1 V to 0.2 in 10 mV steps and the energy minimum voltage is searched in case of $25^\circ$C. Figure 12 shows the energy/operation around the energy minimum voltage, 0.16 V, in both SVT and LVT circuits. This means that the NTC operating at the supply voltage lower than 0.16 V can bring no advantageous effect to the C-element.

5. Conclusion

Near-threshold computing (NTC) technique for operating the circuit below the threshold or less called is expected to be a promising scheme to reduce power consumption significantly. There is challenging issues on the near-/sub-threshold voltage operation, in which function failure must be eliminated and performance degradation and variations must be suppressed.

This paper focuses on the self-timed pipeline (STP) circuit which has an inherent robust feature against the PVT variations and discusses the low voltage characteristics of the CMOS STP circuit. After that, an optimal transistor sizing method for the STP operating at near-threshold voltage is proposed and verified by using SPICE simulation of 65 nm CMOS process libraries (SVT and LVT) in various conditions, where VDD is 0.1 V to 1.2 V in 0.1 V steps, temperature is $-20^\circ$C, $25^\circ$C, and $75^\circ$C, and the process corner is SS, TT, and FF. The lower limit voltage for minimizing the energy/operation was 160 mV for the C-element realized by either SVT or LVT transistor. The simulation results reported in the paper are limited to the single C-element CMOS circuit, so that the practical STP circuit such as DDMP should be designed by applying the proposed method in the paper and evaluated under near-threshold voltage region. Those works are still remained as further investigations in our research project.
Fig. 7: Performance characteristics of SVT-based C-element (25°C).

Fig. 8: Performance characteristics of SVT-based C-element (75°C).

Fig. 9: Performance characteristics of LVT-based C-element (−20°C).
Acknowledgement

Although it is impossible to give credit individually to all those who organized and supported our project, the authors would like to express their sincere appreciation to all the colleagues in the project.

The circuit design work was supported by VLSI Design and Education Center (VDEC), the University of Tokyo in collaboration with Synopsys, Inc. and Cadence Design Systems, Inc.

References


Fig. 10: Performance characteristics of LVT-based C-element (25°C).

Fig. 11: Performance characteristics of LVT-based C-element (75°C).

Fig. 12: Energy minimum voltage of C-element.