A High-Speed Variation-Tolerant Interconnect Technique for Sub-Threshold Circuits Using Capacitive Boosting

Jonggab Kil
Intel Corporation
1900 Prairie City Road
Folsom, CA 95630
+1-916-356-9968
jonggab.kil@intel.com

Jie Gu
University of Minnesota
200 Union Street SE
Minneapolis, MN 55455
+1-612-625-1029
jiegu@umn.edu

Chris H. Kim
University of Minnesota
200 Union Street SE
Minneapolis, MN 55455
+1-612-625-2346
chriskim@umn.edu

ABSTRACT
This paper describes an interconnect technique for sub-threshold circuits to improve global wire delay and reduce the delay variation due to PVT fluctuations. By internally boosting the gate voltage of the driver transistors, operating region is shifted from sub-threshold region to super-threshold region enhancing performance and improving tolerance to PVT variations. A clock distribution network using the proposed drivers shows an 89% reduction in 3σ clock skew value. A 0.4V test chip has been fabricated in a 0.18μm 6-metal CMOS process to demonstrate the effectiveness of the proposed scheme. Measurement results show 2.6X faster switching speed and 2.4X less delay sensitivity under temperature variations.

Categories and Subject Descriptors
B.7.1 [**Integrated Circuits**]: Types and Design Styles - *Microprocessors and microcomputers; VLSI (very large scale integration)*

General Terms
Performance, Design, Measurement

Keywords
Global interconnect, sub-threshold circuit, capacitive boosting, clock skew, variation tolerance

1. INTRODUCTION
With aggressive CMOS scaling, on-chip global interconnects have become the bottleneck for high-speed operation due to the increase in RC per length of minimum wires and near-constant die size. To mitigate the global interconnect delay problem, metal wires have been scaled in a selective fashion. The upper layer metals are remained thick and wide to reduce the wire resistance. As such, the wire pitch is not scaled as aggressively in these layers to maintain a low interwire capacitance. This ensures a low RC value for global signals and power networks. The lower layer metals on the other hand, are scaled at approximately the same rate as the devices for the local interconnects. Low-k inter-dielectric materials and copper wires have been deployed for a one time improvement in RC delay. The wire delay can be made proportional (instead of quadratic) to wire length using tapered wires and efficient buffer insertion techniques [1]-[3]. Despite the various process and circuit techniques for wire RC reduction, global wire delay will continue to become the performance limiter as the delay of logic and short interconnects continue to scale faster than that of global interconnects. Fig. 1 illustrates this trend where the proportion of global interconnect delay with respect to total system delay rapidly increases with technology scaling for a constant die size.

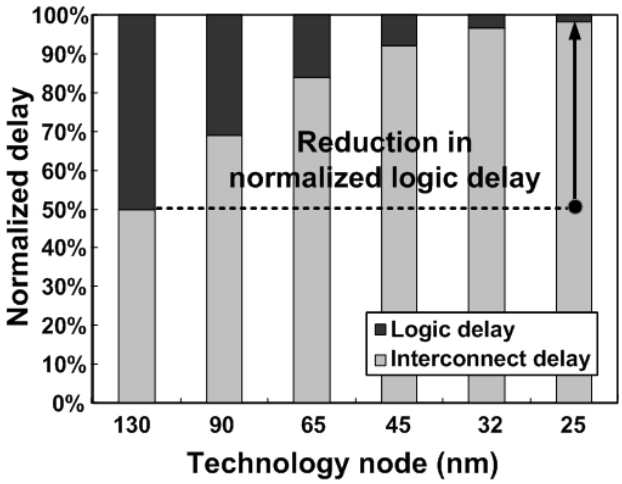

Figure 1. Scaling trend of logic delay and interconnect delay (source: International Technology Roadmap for Semiconductors [9]).

Interconnect delay becomes even more problematic in VLSI systems operating at low supply voltages. The extreme case for this is sub-threshold circuits where the supply voltage is lower than the threshold voltage. Sub-threshold operation can achieve orders of magnitude lower power consumption compared to conventional super-threshold operation and can be used in applications such as medical devices, portable electronics, and sensor networks where performance is of secondary importance [4]-[7]. By simply scaling down the supply voltage, undesirable characteristics of scaled CMOS, such as Drain Induced Barrier

ISLPED'06, October 4–6, 2006, Tegernsee, Germany.

Lowering (DIBL), quantum mechanical gate tunneling, and punch through can also be alleviated. In the sub-threshold region, the MOS gate capacitance is significantly less than that in the super-threshold region due to the channel depletion capacitance that appears in series with the oxide capacitance [8]. Unlike MOS gate capacitance, the wire capacitance value is independent of the supply voltage. As a result, the CV/I delay of wires increase more steeply than that of logic gates in sub-threshold circuits, exacerbating the global interconnect delay problem. Fig. 2 shows the logic delay in proportion to interconnect delay as the supply voltage is reduced from 1V to 0.2V. Evidently, the interconnect delay will dominate the overall system delay as the relative logic delay decreases at sub-threshold voltages due to the reduced MOS gate capacitance.

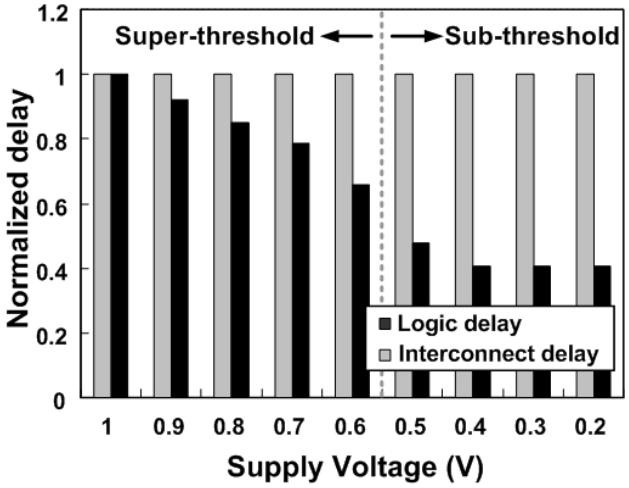

Figure 2. Simulation results on logic delay and interconnect delay at different supply voltages (0.18µm CMOS process).

With global wire delay dictating the overall system performance, its variation also has a larger impact on system performance in sub-threshold circuits. Due to the exponential relationship between weak-inversion current and the PVT parameters, performance and power dissipation of sub-threshold circuits are exceedingly sensitive to PVT fluctuations. Note that random dopant fluctuation becomes the main source of parameter fluctuation in sub-threshold since DIBL is much reduced [10]. To build efficient sub-threshold circuits with operating frequencies up to the MHz range, it is crucial to minimize the global interconnect delay and its variation.

In this paper, we propose a high-speed variation-tolerant interconnect technique for sub-threshold circuits based on capacitive boosting. The central idea is to shift the operating region of the final interconnect drivers from the sub-threshold to super-threshold region via bootstrapping techniques. Boosting techniques such as the one proposed in this work are extremely effective for improving interconnect performance in sub-threshold circuits since current is an exponential function of gate voltage. For example, a 100mV boost in gate voltage can offer a 10X increase in drive current. Circuit techniques to achieve high boosting efficiency, minimum leakage, and no start-up time constraints were deployed. The proposed scheme was applied to a clock distribution network to verify the effectiveness in reducing the clock skew. A 0.4V, 0.18µm test chip was successfully fabricated and tested. Measurement results show 2.6X higher switching speed and 2.4X reduced delay sensitivity with a 12% leakage power overhead at 0.4V.

2. PROPOSED SUB-THRESHOLD INTERCONNECT TECHNIQUE

2.1. Conceptual idea

Fig. 3 shows the principle of the proposed sub-threshold interconnect technique based on capacitive boosting. The 0.4V input signal is boosted to 0.8V for the NMOS driver and to -0.4V for the PMOS driver using internal gate capacitors. The supply voltage is 0.4V while the threshold voltages of the PMOS and NMOS are 0.51V and -0.51V, respectively. Owing to the exponential behavior of current in the sub-threshold region, 100X higher operating current can be achieved with a 0.4V boost in driver voltage as shown in Fig. 3. Note that for the same amount of voltage boost, the increase in drive current will be significantly less in super-threshold circuits. Hence voltage boosting offers greater speed benefits in the sub-threshold region where current is an exponential function of the gate-to-source voltage. The improvement in drive current comes at the expense of leakage current since the gate input to N5 (P5) must be preset to 0.4V (0V) before the boosting occurs. However, this has minimal impact on total chip power dissipation due to the limited number of speed-critical interconnects (buses and clock buffers). The leakage power can be further reduced by using minimum sized transistors P5 and N5 in Fig. 3 since the operating current is significantly increased via the boosting technique.

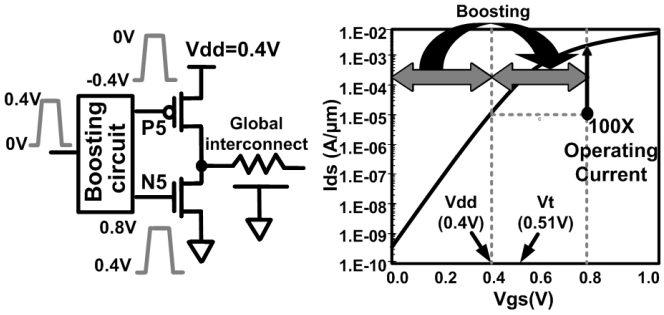

Figure 3. Concept of proposed sub-threshold interconnect driver using boosting technique (left). Effectiveness of boosting technique in sub-threshold region (right).

2.2. Circuit design

Circuit implementation and operating waveforms of the proposed interconnect driver operating at 0.4V are shown in Fig. 4. $V_{BOOST-N}$ and $V_{BOOST-P}$ are boosted by capacitors C_1 and C_2 to increase the operating currents of N5 and P5 which drive the long RC wire. Circuit operation for boosting the gate voltage of N5 is as follows (boosting operation of P5 is similar to that of N5). In order for V_{IN-BAR} to offset the voltage level of V_{BOOST_N} using C_1, V_{BOOST_N} is preset to 0.4V before V_{IN-BAR} makes the low-to-high transition. This is realized by P4 which connects $V_{BOOST-N}$ to 0.4V while V_{IN-BAR}=0V. After the low-to-high transition of V_{IN-BAR}, P4 is cut off so that the boosted voltage V_{BOOST_N}, stays at 0.7V while N5 is driving the RC interconnect. Due to the parasitic capacitance on the node V_{BOOST_N}, the boosting voltage does not reach the ideal 0.8V value. The Preset Signal Generator (PSG) circuit generates a $V_{PRESET-N}$ of -0.25V using C_4 during the preset of $V_{BOOST-N}$. This enables a fast preset by overdriving P4, minimizing the start-up time despite the low drive current. On the other hand, a V_{PRESET_N} of 0.7V is generated by the PSG circuit during the boosting operation by connecting V_{BOOST_N} to V_{PRESET_N} via P3. This eliminates the reverse current through P4 which can adversely discharge the boosted voltage.

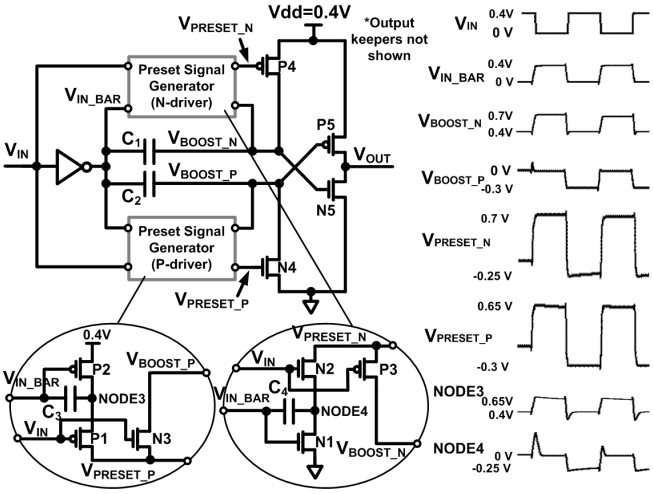

Figure 4. Circuit implementation and operation waveforms of the proposed interconnect driver.

2.3. Boosting efficiency in sub-threshold region

The boosting efficiency is defined in Fig. 5 as the ratio between the boosting capacitance (C_{boost}) and the total capacitance (C_{boost}+C_{node}) which consists of the boosting capacitance and the node capacitance. The boosting capacitance is implemented using a MOS capacitor and the node capacitance consists of the gate capacitances of P5, N5, P4, and N4, as well as the junction capacitances of all the other devices attached to the boosted node. To obtain a high drive current via efficient boosting, the boosting capacitance implemented using a MOS capacitor must be significantly larger than the node capacitance. Unfortunately, the boosting MOS capacitance reduces in the sub-threshold region since the depletion capacitance appears in series with the oxide capacitance in a weak-inversion device. Fig. 6 (left) shows the reduction in boost capacitance in the sub-threshold region. To achieve a high boosting efficiency in our design, 40% of the total driver area is dedicated to boosting capacitors. N5 and P5 which can make up 50% of the total node capacitance are also minimized. Fig. 6 (right) verifies the boosting efficiency of the proposed circuit at different voltages. The boosting efficiency is maximized at 0.6V and reduces to 59% at 0.3V mainly due to the reduction in boosting capacitance in the weak-inversion region. At 0.4V operation, boosting efficiency of 70% was achieved which is sufficient for a significant drive current boost.

	Boosting cap	*Node cap*
Pull-up	C_2	*P5, N4, N3, P1*
Pull-down	C_1	*N5, P4, P3, N2*

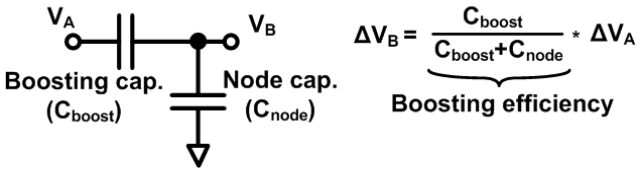

Figure 5. Definition of boosting efficiency.

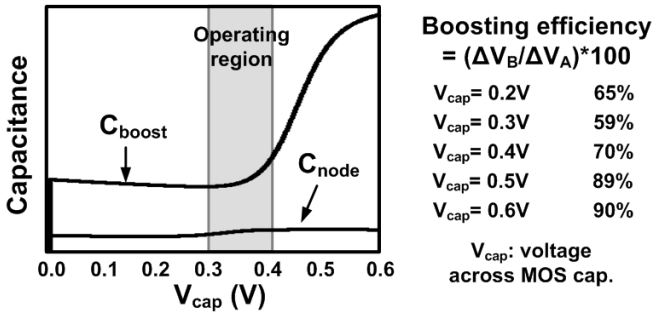

Figure 6. Boosting capacitance, node capacitance, and boosting efficiency at different voltages across MOS capacitors.

2.4. Sensitivity to PVT variation

In addition to the speed benefit, the proposed interconnect technique reduces the impact of PVT variation on global interconnect performance. Fig. 7 shows the rise delay and fall delay under supply (0.4V±5%) and temperature (0~40°C) variations. The conventional driver in the sub-threshold region is highly sensitive to voltage and temperature variation since the drive current is an exponential function of the PVT parameters. This results in a 167.5ns (161.2ns) variation in rise (fall) delay. The switching speed of the proposed interconnect varies significantly less even for the worst case corner conditions because the driver transistors are no longer in the sub-threshold region. Delay variation is reduced from 167.5ns to 50.2ns for the rise delay, and from 161.2ns to 45.1ns for the fall delay.

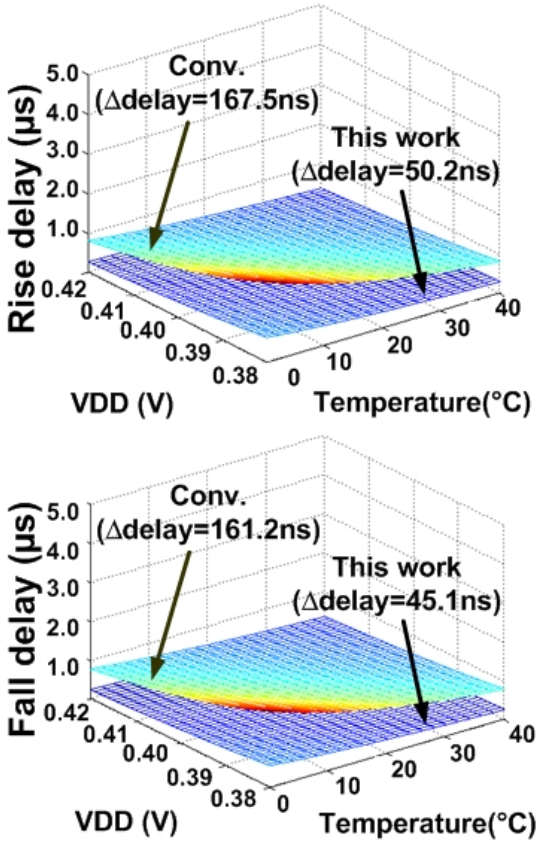

Figure 7. Rise and fall delay with respect to voltage and temperature variation.

3. CLOCK SKEW REDUCTION USING PROPOSED INTERCONNECT TECHNIQUE

The proposed interconnect technique is applied to a realistic clock distribution network [11] to validate the effectiveness in reducing the clock skew caused by PVT variations. Fig. 8 shows the simplified H-tree clock network topology where the clock signal paths are symmetrically routed across the chip for identical delays from clock source to the final load. The hierarchical topology combines four clock buffer stages with each buffer driving four clock buffers as well as the long interconnects. Clock skew is the signal arrival time difference between two different locations in a die. An ideal H-tree should be perfectly matched and will have zero clock skew without any within-die PVT variations. The worst case clock skew occurs when two clock paths are experiencing the extreme opposite PVT conditions. However, this would lead to unrealistically pessimistic estimates on clock skew [12, 13]. Hence in this work, we follow the clock skew distribution analysis based on Monte Carlo simulations assuming that the local supply voltage and threshold voltages are Gaussian random variables [11].

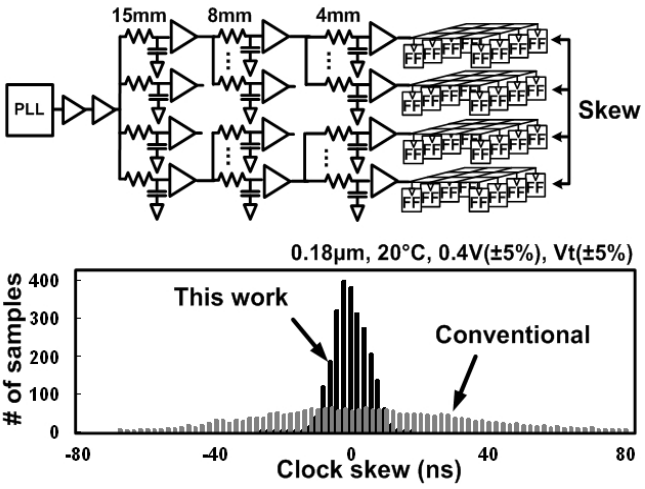

Figure 8. Sub-threshold clock network and clock skew distribution comparison.

Simulation results in Fig. 8 indicate a 9X reduction in 3σ clock skew value using the proposed driver. The 3σ value for VDD and Vt was assumed to be 5% of their nominal values. The average clock tree delay reduced from 251ns to 51ns and the standard deviation reduced from 30.7ns to 5.1ns.

4. TEST CHIP MEASUREMENTS

A test chip was fabricated in a 0.18µm 6-metal triple-well CMOS process to demonstrate the effectiveness of the proposed sub-threshold interconnect scheme. Threshold voltage of the NMOS and PMOS were 0.51V and -0.51V, respectively. The die photo of the test chip and the layout comparisons are shown in Fig. 9. Although the device count for the proposed driver has gone up to 18 (device count for conventional repeater is 4) the increase in layout area was only 32% since the transistors can be minimized thanks to the higher operating current using the proposed boosting technique. The 4 MOS capacitors for the boosting operation occupy 40% of the total driver layout area. Eight stages of conventional drivers and proposed drivers were implemented with each stage driving a 10mm long on-chip wire (Fig. 10). To improve the accuracy of the delay measurement, a differential method was used to obtain the delay difference between the two paths and a bypass path. The peripheral and I/O circuit delay is cancelled out by subtracting the delay of the bypass path (t_{bypass}) from the delay of the other two paths ($t_{conv}+t_{bypass}$, $t_{proposed}+t_{bypass}$). The core, peripheral, and I/O circuits operate at 0.4V, 1.0V, and 1.8V respectively. Level converters were employed for the interface between sub-threshold and super-threshold circuits. The sub-threshold level-down converter contains a pull-up NMOS N6 to speed up the low-to-high transition in the internal node, A1. A dual-rail level-up converter was designed with an extra PMOS switch P7 (or P8) to reduce the contention current between P9 (or P10) and N7 (or N8).

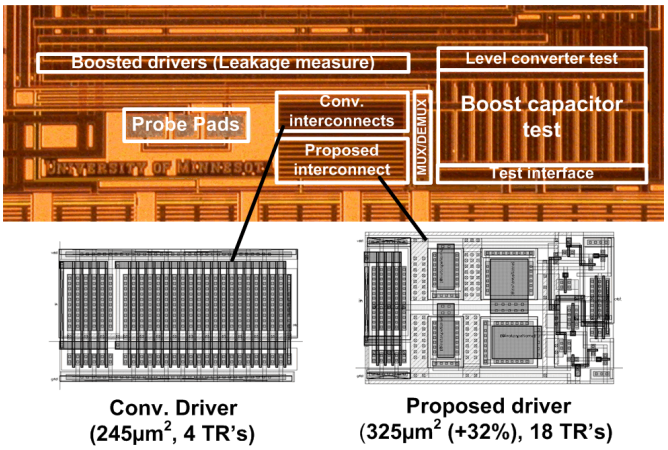

Figure 9. Chip microphotograph and driver layout comparison. The test chip was fabricated in a 0.18µm 6-metal CMOS process.

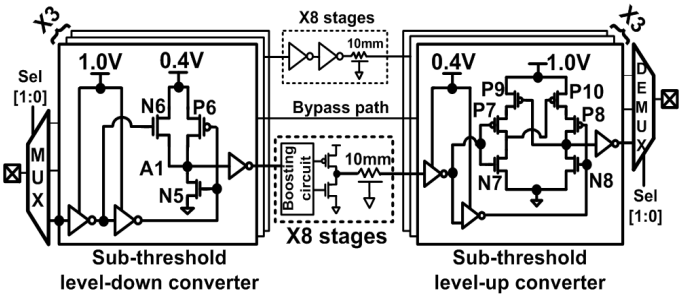

Figure 10. Organization of the test chip for differential delay measurements. Level converters are designed for the interface between sub-threshold and super-threshold circuits.

Fig. 11 shows the measured waveforms from the three different paths together with the input trigger signal. Delay of the proposed driver was 0.18µs which is 2.6X shorter than that of a conventional driver. Delay of the bypass path was 0.13µs which is comparable to the core interconnect delay. The delay improvement is less than the ideal amount of boost in operating current shown in Fig. 3 due to the following reasons: (i) minimum size driver transistors for reducing leakage power, (ii) extra number of logic stages required in the boosting circuit and (iii) reduced boosting effect because of the node capacitance. Fig. 12 shows measured delay improvements of the proposed interconnect technique at different core voltages. The delay improvement is 1.7-1.8X at 0.5V and 2.6-2.9X at 0.4V. This confirms that the proposed boosting technique becomes more

efficient at low supply voltages due to the exponential current behavior in the sub-threshold regime.

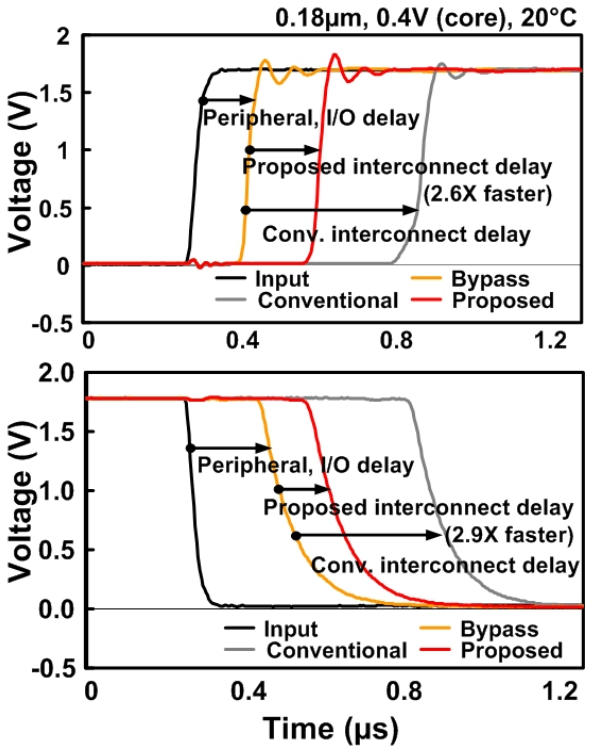

Figure 11. Measured waveforms (input trigger signal and output waveforms from the 3 interconnect paths) from the test chip.

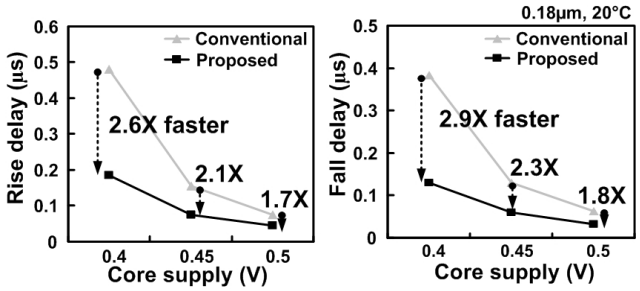

Figure 12. Rise and fall delay measurement data for different core voltages indicating 2.6-2.9X improvement in interconnect speed.

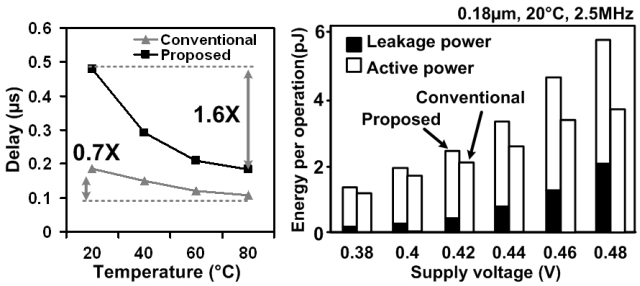

Figure 13. Measurement data from test chip. Delay variation with respect to temperature (left). Energy per operation comparison between conventional and proposed interconnect driver (right).

Fig. 13 shows the measured delay variation and power dissipation of the proposed and conventional interconnect drivers. The delay variation of the conventional and proposed driver were 1.6X and 0.7X, respectively for a temperature range of 20-80°C. Delay of the proposed driver is less sensitive to PVT variations since the driver transistors are no longer operating in the sub-threshold region due to the boosted gate voltages. Measured energy per operation is shown in Fig. 13 (right) for the conventional and proposed drivers. Due to the leakage power during the preset of the gate voltages, energy per operation of the proposed driver increases by 12% when operated at 0.4V. The leakage current of the proposed boosting technique reduces exponentially in the deep sub-threshold region as shown in Fig. 13. Power dissipation comparison between conventional and proposed buffer is shown in Fig. 14 for different operating frequencies. Measurement results show good agreement with the simulation data. The proposed driver consumes 41% less power (or 49%+ higher performance) compared to the conventional driver for 4MHz (or 5.1µW) operation since the boosting technique allows the proposed driver to run at a lower supply voltage for the same frequency of operation.

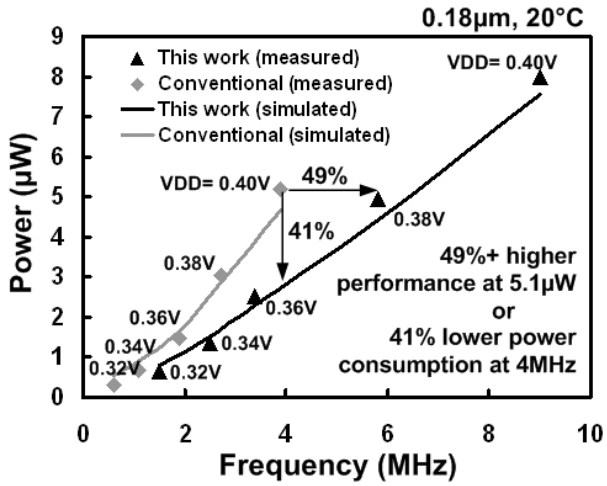

Figure 14. Power consumption versus frequency. (Measurement data is shown in dots and simulation data is shown in lines)

5. CONCLUSION

Digital sub-threshold logics are becoming increasingly popular for ultra-low power applications where performance is of secondary importance. Global interconnect drivers used for on-chip buses and clock distribution networks significantly suffer from performance degradation. This is because unlike gate capacitance, wire capacitance does not scale as the supply voltage is lowered. Another issue with sub-threshold interconnects is the large variability in performance under PVT variations; drive current in the sub-threshold region is an exponential function of threshold voltage, supply voltage, and temperature. In this paper, we proposed a capacitive boosting technique that can mitigate the performance and variability issues in sub-threshold interconnects. Owing to the exponential relationship between current and gate-to-source voltage in the sub-threshold region, 100mV boost in gate voltage can offer 10X improvement in drive current. This makes boosting techniques extremely effective for global interconnect drivers. A test chip was fabricated in a 0.18µm 6-metal CMOS process to demonstrate the proposed ideas. Measurement results show 41% less power consumption for same performance (or 49%+ higher performance for same power consumption) with 2.4X reduced delay sensitivity under temperature variations.

6. REFERENCES

[1] K. Yamashita and S. Odanaka, "Interconnect scaling scenario using a chip level interconnect model," *IEEE Trans. Electron devices,* vol. 47, no.1, pp. 90-96, Jan. 2000.

[2] D. Sylvester, C. Hu, O. S. Nakagawa, and S.-Y. Oh, "Interconnect scaling: Signal integrity and performance in future high-speed CMOS designs," *Symposium on VLSI Technology Dig. Tech. Papers*, pp. 42-43. Jun. 1998.

[3] S. Dhar and M. A. Franklin, "Optimum buffer circuits for driving long uniform lines," *IEEE Journal of Solid-State Circuits*, vol. 26, pp. 32-40, Jan. 1991.

[4] C.H. Kim, and K. Roy, "Ultra-low power DLMS adaptive filter for hearing aid applications," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol.11, no.6, pp. 352-357, Dec. 2003.

[5] A. Wang, and A.P. Chandrakasan, "A 180mV FFT processor using subthreshold circuit techniques," *International Solid-State Circuits Conference*, pp. 292-293, Feb. 2004.

[6] B.H. Calhoun, and A.P. Chandrakasan, "Ultra-dynamic voltage scaling using sub-threshold operation and local voltage dithering in 90nm CMOS," *International Solid-State Circuits Conference*, pp. 300-302, Feb. 2005.

[7] B.H.Calhoun, A.Wang, and A.P. Chandrakasan, "Modeling and sizing for minimum energy operation in subthreshold circuits," *IEEE Journal of Solid-State Circuits,* vol. 40, no. 9, pp. 1778-1786, Sept. 2005.

[8] Y. Taur, and T. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge University Press, 1998.

[9] International Technology Roadmap for Semiconductors, http://public.itrs.net/

[10] B. Zhai, S. Hanson, D. Blaauw, and D. Sylvester, "Analysis and mitigation of variability in subthreshold design," *International Symposium on Low Power Electronics and Design*, pp. 20-25, Aug. 2005.

[11] D. Harris and S. Naffziger, "Statistical Clock Skew Modeling With Data Delay Variations," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems,* vol.9, no.6, pp. 888-898, Dec. 2001.

[12] M. Shoji, "Elimination of process-dependent clock skew in CMOS VLSI," *IEEE Journal of Solid-State Circuits*, vol.21, no.5, pp. 875-880, Oct. 1986.

[13] G. Geannopoulos, and X. Dai "An adaptive digital deskewing circuit for clock distribution networks," *International Solid-State Circuits Conference*, pp. 400-401, Feb. 1998.