

# Level Shifter Design for Voltage Stacking

Elnaz Ebrahimi, Rafael Trapani Possignolo and Jose Renau

Dept. of Computer Engineering, University of California, Santa Cruz, Santa Cruz, CA, 95064

Email: {eebrahim,rpossign,renau}@ucsc.edu

**Abstract**—As chips increase in complexity with ever increasing power consumption, pressure in efficient power delivery mechanism such as multi-VDD, voltage stacked and DVS continues to rise. The main objective is to reduce the overall current delivered to the chip. For instance, in voltage stacking, if the circuit is stacked in 2 levels and supply voltage is doubled, the current drawn will be reduced by half. Hence, the same amount of power is delivered, but with half the current. With the prevalence of systems using those techniques, level shifters will have to be optimally designed to perform fast with low power. As the number of level shifters grows, area consumption becomes another design factor. This study explores different types of existing level shifters for voltage stacking application, their optimal sizing and energy, delay and area trade-offs. It includes effect the of PVT variation as another design factor and its impact on delay and energy consumption. We will also propose modifications to the best energy-delay level shifter to reduce its area overhead.

## I. INTRODUCTION

As device size scales down and number of transistors and frequency increase, power consumption becomes a critical issue in System-On-Chip design. Since dynamic power is cubically proportional to supply voltage, one prevalent technique to reduce power is scaling down the supply voltage, which impacts performance by reducing the frequency at which the design can run. To avoid performance degradation, one solution is to use multiple supply voltages to reduce the power consumption. The critical path components will continue running at the VDD level while non-critical path components run at a scaled down VDD [10].

Using a multi-VDD system is an alternative to voltage scaling technique. It counteracts the negative impact on performance, because the critical path components will continue running at VDD level while non-critical components run at a scaled down VDD [14]. In a multi-VDD system, when the DC current flows from a low voltage gate to a high voltage gate, the voltage is not sufficient to turn the PMOS “ON” and therefore, the PMOS in the high voltage gate is weakly “ON” conducting static current from the power supply to the ground. The level shifters will remove the static current and restore the full voltage swing from VddL to VddH [7].

Designing a multi-VDD system is inherently complex as there are a few challenges in using level shifters (LS) in the system. They dissipate power and add propagation delay. It is necessary to optimize the LS circuit for minimum energy-delay product to obtain the potential benefit of using multiple power supply domains. As an LS includes both high voltage and low voltage gate, it will require more area and routing resources. For example, when each functional block on a die needs a different voltage for its desired performance, the number of level converters can easily grow and become a design area overhead. Techniques such as Dynamic Voltage scaling (DVS) has been widely used in digital signal processing elements for reducing energy consumption [17]. And future low-power systems-on-chips (SoCs) are likely to consist of many scalable voltage domains. This requires level shifters to be able to perform at a high speed with low power [14], [19].

Another more recent approach to reduce the current required by a chip is voltage stacking [5], [11]. Voltage stacking

is connecting logic blocks in a series configuration, rather than parallel configuration [2], and thus delivering the same amount of power by increasing voltage and reducing current by a factor of  $n$  (the number of stack levels). Voltage stacking has been proposed between cores [11], within a core [2], and more recently in GPUs [18] and SRAMs [4]. Voltage stacking reduces the number of pins dedicated to power, increases the voltage regulator efficiency and reduces voltage noise and droop [2].

As in the case of multi-VDD systems, voltage stacked systems require level shifters for inter-level communication [5], [11]. Traditional level shifters are inserted to translate or shift the logic levels from the level supplied by one domain to another level supplied by the second domain. In the context of voltage stacking, the level shifters will have a primary voltage rail which sits at the top and a secondary voltage rail which sits in the middle. When placed in a voltage stacked design, they will shift the both rails, either from GND-midrail to midrail-toprail (low to high level shifters) or from midrail-toprail to GND-midrail (high to low level shifters). Although many designs for level shifters exist, an evaluation of different designs in the context of voltage stacking has not been made, so the trade-offs of different designs are not clear. We evaluate existing approaches of LSs for voltage stacking applications. We are especially interested in delay and power, but area and sensitivity to PVT (Process, Voltage and Temperature) variations are also considered. Each of those parameters may have different priority in different designs. For instance, CoreUnfolding [2] allows an entire clock cycle for level shifting, thus delay is less important. However, it requires a large amount of shifters, which makes area a critical design factor. On the other hand, a voltage stacked SRAM [4] requires minimal impact on timing, but due to the small number of shifters, can tolerate more area overhead per shifter.

The contributions of this paper are:

- Overview of different LS designs.
- Energy, delay, and area comparison of LS designs.
- PVT tolerance evaluation of LS designs.

## II. OVERVIEW OF LEVEL SHIFTER DESIGNS

This study explores some of the LS designs that are suitable for a stacked architecture. We focus on converting in a stacked architecture where the primary/top voltage rail is 2V and the middle voltage rail is 1V. The signals are shifted from 0-1V voltage domain to operate in 1-2V voltage domain and vice versa. The schematics for the LS designs evaluated are in Figure 1, where each circuit shows how Vin “low” is converted to Vout “high”, all of the chosen level shifters are bidirectional.

**Capacitive-Coupling-based (Conventional)** (Figure 1(a)) is a capacitive-coupling-based LS for a multi-story or voltage stacked power delivery scheme. This LS has a driving inverter, a coupling capacitor, and a receiver with gain stages. Two diodes are connected back to back in order to constrain the voltage swing at the output node of the coupling capacitor (gate and drain are shorted in the NMOS transistors). Since it always settles near the inverter trip point, a signal transition takes place with a minimal size coupling capacitor [5].

**Two-Stage Cross-Coupled (TSCC)** (Figure 1(b)) uses two cross-coupled stages. The first stage is a differential cascade voltage switched logic gate, using a cross-coupled PMOS half latch operating at the higher supply voltage. To overcome the leakage of weakly conducting PMOS transistors, drive strength of NMOS transistor is enhanced. Low  $V_{in}$  input voltage turns  $mn1$  on, which discharges node A to ground and activates  $mp2$ . Node B will be pulled up to  $V_{ddH}$  and the output voltage will be low. Subsequently, when  $V_{in}$  is asserted,  $mn2$  and  $mp1$  are activated shifting the output voltage up to  $V_{ddH}$ . The drive of the pull-down transistors needs to be much larger than the PMOS transistors to overcome its latch action driven with a higher supply voltage. It is a simple design suited for super-threshold conversion [6], [8], [12]. The second stage is intended to achieve full voltage swing. The diode-connected NMOS transistor is employed to weaken the pull-up network (Figure 1(c)), which expands the convertible input voltage. The operating range is determined by the transistor size and threshold voltage [8], [9], [19].

**Wilson Current Mirror (WCM)** (Figure 1(c)) is based on the traditional Current Mirror (CM), a unity gain current amplifier which provides output current proportional to input current at its high impedance output. It maintains the output current constant regardless of load [1], [12]. The high drain-to-source voltage of PMOSs facilitates the construction of a stable current mirror, which offers an effective on-off current comparison at the output. However, for super-threshold input voltage, a high amount of quiescent current occurs, limiting its use [8]. In WCM, this current is cut off by a feedback PMOS ( $mp3$ ), reducing standby power. However, as the source current is cut off, the mirror current through  $mp2$  is largely reduced, weakening pull-up strength and dropping the voltage at node A. Although the voltage drop increases the source current through the feedback control, the current increase is too small to pull the voltage at node A back to  $V_{ddH}$ . The output finally stabilizes at a voltage below  $V_{ddH}$ , which causes large static current and standby power in the output buffer [19].

**Stacked Wilson Current Mirror (Stacked)** (Figure 1(d)) is an enhancement to the WCM design and Kumar *et al.* use a stacking technique to reduce the leakage power consumption [13]. The technique adds three NMOS transistors in the pull-down network.

**Switched-Capacitance (Tong)** (Figure 1(e)) is a capacitive-coupled design for voltage stacking [11]. The voltage across the capacitor depends on the difference between the two domains, but it can be higher than the gate-oxide breakdown voltage. Hence, this approach requires metal-oxide-metal (MOM) capacitors [16]. The original design has one 25fF capacitor on each side of the back to back inverters. If we translate each fF to  $\approx 1\mu m^2$ , the LS area is considerably large. Our experiments show that the 25fF capacitors are over-designed for an LS, and we were able to reduce that number to  $\approx 2.6$ fF (details in the experimental section), considering a 30% margin over the minimum operational point. Even with this size reduction, the area is still large.

**Modified Switched-Cap (Mod-Tong)** (Figure 1(f)): To prevent the use of the MOM capacitors in Tong, we replace each capacitor with two NMOS transistors connected such that the drain and gate are shorted. This reduces the area, but is expected to increase the power consumption due to the resistive effect added.

### III. CHARACTERIZATION

#### A. Transistor Sizing

To setup LSs for energy, delay, and area comparison, we begin by determining the optimal size for each LS. Our

experiments use the NCSU FreePDK-45nm [15]. We perform HSpice simulations varying the width of each transistor in an LS between 90nm and 720nm. Energy and average propagation delay of the output signal for an input signal transitioning at 1GHz is calculated. The results are plotted for low-to-high and high-to-low conversion (Figure 2).

There is a Pareto frontier for each LS in Figure 2 and any of the frontier points could be of interest for a specific LS depending on its application. The minimum energy-delay product (ED) point is our point of interest for optimal sizing of an LS (It is circled in the plot and listed in Table I. The power and delay are measured for a 1GHz input pulse as *active* energy and *idle* energy, *i.e.*, when the input voltage is kept constant. We estimate the area considering it is proportional to the sum of widths of transistors in the design and the area of one transistor (45nm $\times$ 90nm). The capacitors in Tong LS dominate the area, 7x the area of Mod-Tong. The top Pareto point is not shown for TSCC down shifter in Figure 2, because the delay is more than 250ps.

Looking at the minimum ED Pareto points (in Table I), there is not a single level shifter that is superior in area, delay, and power. For instance, if minimum delay in the level up shifters is the main design concern, Conventional and Tong are best choices, however, they impact the area significantly as Tong uses two 2.6fF capacitors and Conventional uses a decoupling capacitor and a diode [3]. Also, when down converting, Mod-Tong presents the best delay, and Tong has the best energy with largest area. Thus, it is not possible to pick the optimal LS although Tong, Conventional and Mod-Tong have overall best numbers. TSCC has the worst delay in this case.

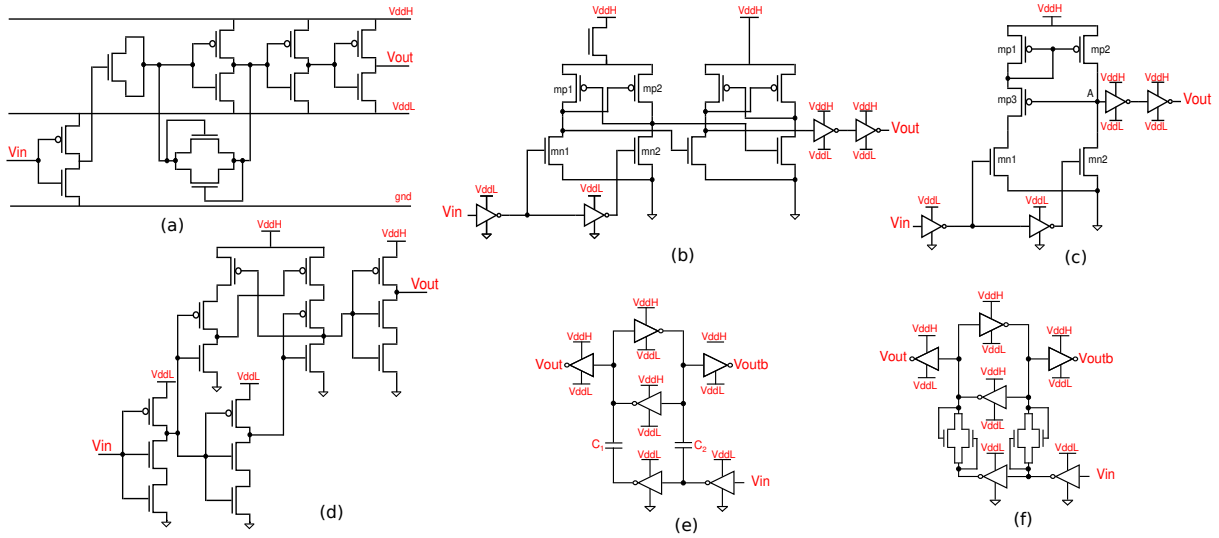
TABLE I: Minimum ED points for each LS.

name	Area ( $\mu m^2$ )	Low to high conversion		
		Delay (ps)	Active Power (pW)	Idle Power (pW)
Conv	0.62	12.37	3.07	0.53
TSCC	0.50	190	6.42	0.53
WCM	0.58	30.69	6.04	0.41
Stacked	0.68	27.64	5.04	0.37
Tong	5.74	14.36	0.48	0.001
Mod-Tong	0.74	20.09	6.57	0.75
name	Area ( $\mu m^2$ )	High to low conversion		
		Delay (ps)	Active Power (pW)	Idle Power (pW)
Conv	0.62	13.99	1.79	0.24
TSCC	0.44	56.39	8.89	1.11
WCM	0.58	21.51	6.04	0.41
Stacked	0.44	26.72	8.69	1.43
Tong	5.68	14.13	0.36	0.01
Mod-Tong	0.82	9.06	14.7	3.40

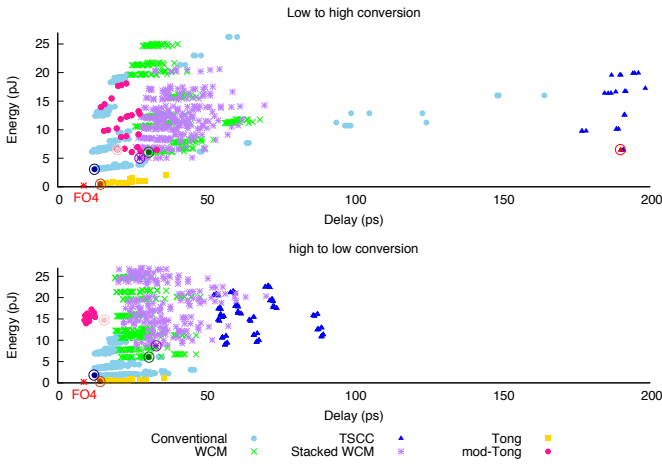
#### B. PVT Variation Effect

An integral deciding factor is LS robustness in presence of PVT variation. To see how temperature affects the delay and energy consumption, we perform an HSpice temperature sweep from 10°C to 90°C. Figure 3 shows the trend when converting from high to low.

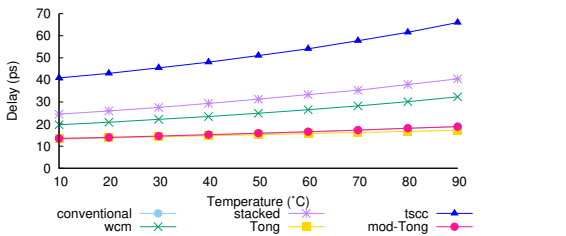
When shifting down, Tong delay line has a slope of 0.046ps/°C and there is less than 5ps difference in delay as the temperature rises up to 90°C. Conventional and mod-Tong have close slopes of 0.068 and 0.072ps/°C respectively which translates to less than 10ps delay difference across different temperatures. WCM and Stacked WCM each have 3.4x and 4.3x the slope of Tong LS, which is equivalent to delay range of up to 16ps and up to 13ps respectively. When shifting up, as the temperature increases, so does the average propagation delay, however, the delay increase is minimal for Conventional, Tong, and mod-Tong: 0.065, 0.053, and 0.082ps/°C. TSCC follows an inverse trend of decreasing delay (22 times the slope of Tong's), where rise and fall delay vary  $\approx 100$ ps from start to finish. The delay itself is large as the whole circuit is slower when up shifting than it is when down shifting (Table I). WCM



**Fig. 1:** (a) Capacitive-coupling (Conventional) (b) Two-Stage Cross-Coupled (TSCC) (c) Wilson Current Mirror (WCM) (d) Stacked Wilson Current Mirror (Stacked) (e) Switched-Capacitance (Tong) (f) Modified Switched-Capacitance (Mod-Tong).



**Fig. 2:** Up shifters active ED for transistor widths 90nm-720nm. and Stacked WCM each have a slope twice as steep compared to Tong which translates to  $\approx 10$ ps delay range.

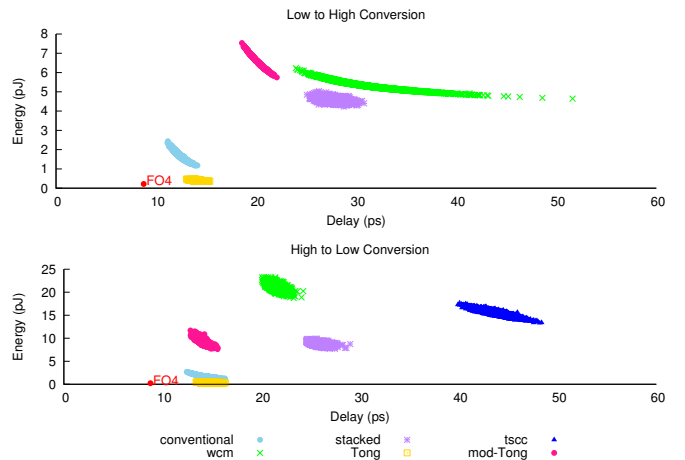


**Fig. 3:** Delay vs. temperature in high to low conversion.

Active energy has a small decreasing trend during the up and down conversion in all the level shifters except for Tong where the line slope is  $\approx 0$  pJ/°C. For up conversion, Tong and Conventional with slopes of 0.003 and  $-0.058$  pJ/°C are the best candidates and mod-Tong has the highest slope,  $-0.023$  pJ/°C. However, for all the level shifters the range that energy varies is less than 2pJ. During the down conversion as temperature increases, TSCC and Stacked WCM are affected the most with slopes of  $-0.065$  and  $0.093$  pJ/°C which translates to an energy range of 7pJ and 8pJ. The least sensitive to varying temperature are Tong and Conventional with slopes

of 0 and  $-0.006$  pJ/°C. Overall, as the temperature increases up to 90°C, taking delay sensitivity into consideration takes priority over the power sensitivity.

Continuing the PVT variation effect analysis, we use HSpice Gaussian distribution function with absolute variation to vary the threshold voltage  $\pm 6\%$  with  $3\sigma$  value or 99.7% yield and run 5000 Monte Carlo simulations. In this experiment, we vary NMOS and PMOS threshold voltage by  $\pm 6\%$  (from 0.3V) and measure delay and energy in the active mode. Figure 4 is the final plot and an FO4 delay point has been included as a point of reference. When up shifting, the energy variation is a few pJs for all the converters, however, the delay variation is not small. For example, WCM delay varies  $\approx 30$ ps. TSCC points have been removed from the up shifters plot, where the delay varies from 100ps to 450ps whereas energy varies from 4.3pJ to 6pJ and is comparable to that of WCM and Mod-Tong. Tong seems to be the least sensitive to the Vth variation. There is an energy-delay trade-off between Mod-Tong and Stacked. Mod-Tong has a smaller delay variation whereas Stacked power consumption varies less. When downshifting, the energy variation for all is less than 5pJ. And again, TSCC is the most sensitive with the largest delay range of  $\approx 10$ ps. TSCC is a two-stage LS, and other LS types have nearly half the number of transistors as TSCC. Consequently, their delay range is almost half as TSCC.



**Fig. 4:** Active Energy vs. Delay:  $\pm 6\%$  Vth variation.

We repeat the same experiment and vary the supply voltages by  $\pm 5\%$  (Figure 5). Unlike previous experiment, Tong seems to be sensitive as the delay varies both in up conversion and down conversion with an 11ps and 18ps range respectively. In up conversion, mainly the delay difference separates the choices. The most unpredictable delay belongs to TSCC, 100ps, and Tong comes in second. However, the energy variation is less than 2pJ for TSCC. Conventional or Mod-Tod might be better choices for down conversion as both energy and delay vary a few units. When downshifting, WCM and Stacked are more sensitive to variation as their energy consumption differs from their Pareto frontier values (Table I). Devices based on capacitance have lower variation on energy, since they tend to not dissipate power.

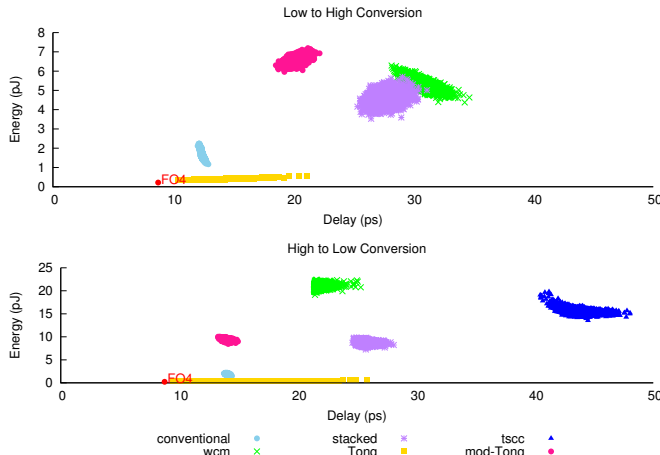


Fig. 5: Active Energy vs. Delay:  $\pm 5\%$  VDD variation.

Since Tong uses 2 capacitors, variation could affect its operational behavior. We repeat the experiment by applying variation to the capacitors. Similar to VDD, the transient sweep is done by using  $\pm 5\%$  variation with  $3\sigma$ . However, it hardly has any effect on the delay and energy consumption of the LS 6. Overall Tong immunity Tong to PVT variation comes at a cost of large area compared to other LS types.

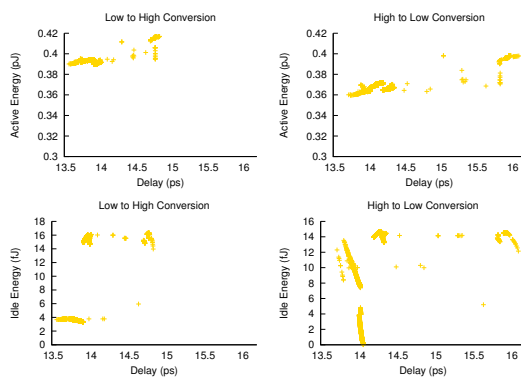


Fig. 6: Active and Idle Energy vs. Delay:  $\pm 5\%$  Capacitance variation.

#### IV. CONCLUSION

Voltage stacking is a promising technique to reduce the overall current required to power chips. It delivers the same amount of power by multiplying the supply voltage and dividing the current by  $n$  (the number of stack levels). Nevertheless, it requires level shifters for inter-level communication.

We study the trade-offs between different designs of level shifters existing in the literature, considering applications in voltage stacked systems. The performance, power, and area of

those designs greatly vary depending on the sizing architecture. In terms of power and delay, Tong [16] offers the best design points, but requires large MOM capacitors, which makes it unsuitable for applications with a considerable number of shifters, therefore, we analyze various designs to show how critical each design factor becomes in different contexts.

#### ACKNOWLEDGMENTS

This study is supported in part by the National Science Foundation under grants CNS-1059442-003, CNS-1318943-001, CCF-1337278, and CCF-1514284. Any opinions, findings, and conclusions expressed herein are those of the authors and do not necessarily reflect the NSF views.

#### REFERENCES

- [1] B. Aggarwal, M. Gupta, and A. Gupta, "A comparative study of various current mirror configurations: Topologies and characteristics," *Microelectronics Journal*, 2016.
- [2] E. K. Ardestani, R. T. Possignolo, J. L. Briz, and J. Renau, "Managing mismatches in voltage stacking with coreUnfolding," *ACM Trans. Archit. Code Optim.*, vol. 12, no. 4, pp. 43:1–43:26, Nov. 2015.
- [3] L. Chang, R. K. Montoyo, B. L. Ji, A. J. Weger, K. G. Stawiasz, and R. H. Dennard, "A fully-integrated switched-capacitor 2:1 voltage converter with regulation capability and 90% efficiency at 2.3a/mm<sup>2</sup>," in *2010 Symposium on VLSI Circuits*, 2010.
- [4] E. Ebrahimi, R. T. Possignolo, and J. Renau, "SRAM voltage stacking," in *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2016, pp. 1634–1637.
- [5] J. Gu and C. H. Kim, "Multi-story power delivery for supply noise reduction and low voltage operation," ser. ISLPED '05. ACM.
- [6] A. Hasanbegovic and S. Aunet, "Low-power subthreshold to above threshold level shifter in 90nm process and 65nm process," in *NORCHIP*, 2009.
- [7] R. Jakushokas, M. Popovich, A. V. Mezhiba, S. Kse, and E. G. Friedman, *Power Distribution Networks with On-Chip Decoupling Capacitors*, 2nd ed. Springer Publishing Company, Incorporated, 2010.
- [8] T. S. Joshi and P. M. R. Nerkar, "A wide range level shifter using a self biased cascode current mirror with ptl based buffer," 2015.
- [9] K.-H. Koo, J.-H. Seo, M.-L. Ko, and J.-W. Kim, "A new level-up shifter for high speed and wide range interface in ultra deep sub-micron," in *2005 IEEE International Symposium on Circuits and Systems*.
- [10] V. Kursun and E. G. Friedman, "Multi-voltage cmos circuit design," 2006.
- [11] S. K. Lee, T. Tong, X. Zhang, D. Brooks, and G.-Y. Wei, "A 16-Core Voltage-Stacked System with an Integrated Switched-Capacitor DC-DC Converter," in *VLSI Circuits (VLSI Circuits), 2015 Symposium on*, June 2015, pp. C318–C319.
- [12] S. C. Luo, C. J. Huang, and Y. H. Chu, "A wide-range level shifter using a modified wilson current mirror hybrid buffer," *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2014.
- [13] S. K. A. M. Kumar and S. Pandey, "Level Shifter Design for Low Power Applications," *Computer Science & Information Technology, International Journal of*, Oct. 2010.
- [14] K. Pasupathy and B. Bindu, "Low power, high speed carbon nanotube fet based level shifters for multi-vdd systems-on-chips," *Microelectron. J.*, 2015.
- [15] J. E. Stine, J. Chen, I. Castellanos, G. Sundararajan, M. Qayam, P. Kumar, J. Remington, and S. Sohoni, "Freepdk v2.0: Transitioning vlsi education towards nanometer variation-aware designs," in *Microelectronic Systems Education, 2009. MSE '09. IEEE International Conference on*, July 2009, pp. 100–103.
- [16] T. Tong, S. K. Lee, X. Zhang, D. Brooks, and G. Y. Wei, "A fully integrated reconfigurable switched-capacitor dc-dc converter with four stacked output channels for voltage stacking applications," *IEEE Journal of Solid-State Circuits*, 2016.
- [17] B. Zhai, D. Blaauw, D. Sylvester, and K. Flautner, "Theoretical and Practical Limits of Dynamic Voltage Scaling," in *Proceedings of the 41st Annual Design Automation Conference*, ser. DAC '04, 2004.
- [18] Q. Zhang, L. Lai, M. Gottscho, and P. Gupta, "Multi-story power distribution networks for GPUs," in *Design, Automation, and Test in Europe (DATE), Proceedings of*, Mar 2016.
- [19] J. Zhou, C. Wang, X. Liu, and M. Je, "Fast and energy-efficient low-voltage level shifters," *Microelectron. J.*, 2015.