

VOLTAGE AND TIMING ADAPTATION FOR
VARIATION AND AGING TOLERANCE IN NANOMETER VLSI CIRCUITS

A Dissertation

by

KYU-NAM SHIM

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Jiang Hu
Committee Members,	Duncan M. Walker
	Gwan Choi
	Jose Silva-Martinez
Head of Department,	Costas Georgiades

December 2012

Major Subject: Computer Engineering

Copyright 2012 Kyu-Nam Shim

ABSTRACT

Process variations and circuit aging continue to be main challenges to the power-efficiency of VLSI circuits, as considerable power budget must be allocated at design time to mitigate timing variations. Modern designs incorporate adaptive techniques for variation compensation to reduce the extra power consumption. The efficiency of existing adaptive approaches, however, is often significantly attenuated by the fine-grained nature of variations in nanometer technology such as random dopant fluctuation, litho-variation, and different rates of transistor degradation due to non-uniform activity factors. This dissertation addresses the limitations from existing adaptation techniques, and proposes new adaptive approaches to effectively compensate the fine-grained variations.

Adaptive supply voltage (ASV) is one of the effective adaptation approaches for power-performance tuning. ASV has advantages on controlling dynamic and leakage power, while voltage generation and delivery overheads from conventional ASV systems make their application to mitigate fine-grained variations demanding. This dissertation presents a dual-level ASV system which provides ASV at both coarse-grained and fine-grained level, and has limited power routing overhead. Significant power reduction from our dual-ASV system demonstrates its superiority over existing approaches.

Another novel technique on supply voltage adaptation for variation resilience in VLSI interconnects is proposed. A programmable boostable repeater design boosts switching speed by raising its internal voltage rail transiently and autonomously, and

achieves fine-grained voltage adaptation without stand-alone voltage regulators or additional power grid. Since interconnect is a widely recognized bottleneck to chip performance and tremendous repeaters are employed on chip designs, boostable repeater has plenty of chances to improve system robustness.

A low cost scheme for delay variation detection is essential to compose an efficient adaptation system. This dissertation presents an area-efficient built-in delay testing scheme which exploits BIST SCAN architecture and dynamic clock skew control. Using this built-in delay testing scheme, a fine-grained adaptation system composed of the proposed boostable repeater design and adaptive clock skew control is proposed, and demonstrated to mitigate process variation and aging induced timing degradations in a power as well as area efficient manner.

DEDICATION

To my wife

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Jiang Hu, for his direction and support over the last 5 years. I would also like to thank Dr. Silva-Martinez, Dr. Choi, and Dr. Walker, for their guidance and support throughout the course of this research.

Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.

Finally, special thanks to my wife for her patience and love, and to my parents for their support and encouragement.

NOMENCLATURE

ABB	Adaptive Body Bias
ASV	Adaptive Supply Voltage
BIST	Built-In Self Test
CMOS	Complementary Metal Oxide Semiconductor
DFT	Design For Testability
DVFS	Dynamic Voltage Frequency Scaling
HCI	Hot Carrier Injection
LFSR	Linear Feedback Shift Register
MPR	Mini Programmable linear voltage Regulator
NBTI	Negative Bias Temperature Instability
OAB	Online Adjustable Buffer
PTM	Predictive Technology Model
VLSI	Very Large Scale Integration

TABLE OF CONTENTS

	Page
I. INTRODUCTION OF RESEARCH.....	1
1. Introduction	1
2. Existing Methods and Their Limitations.....	2
A. Adaptive Body Bias vs. Adaptive Supply Voltage	2
B. Granularity of ASV	3
C. Dual Static Supply Voltage	4
D. Overhead on Supply Voltage Generation.....	4
E. Online Adjustable Buffer	6
3. Proposed Adaptation Techniques	6
A. A Dual-Level Adaptive Supply Voltage (ASV) System	6
B. Boostable Repeater for Variation Tolerance in Interconnects	8
C. Built-in Delay Testing Scheme	10
II. A DUAL-LEVEL ADAPTIVE SUPPLY VOLTAGE (ASV) SYSTEM.....	13
1. Rationale and Overview of Dual-ASV System.....	13
2. Voltage Tapping.....	17
3. Progressive Voltage Enhancement.....	28
4. Experimental Results.....	32
5. Conclusions	40
III. BOOSTABLE REPEATER DESIGN FOR VLSI INTERCONNECT	42
1. Boostable Repeater Design	42
A. Main Idea and Overview	42
B. Design and Operations	45
2. Experimental Validation.....	52
A. Experiment Setup	52
B. Impacts of Design Parameters	53
C. Simulations on Global Interconnect.....	57
D. Comparison with Online Adjustable Buffer.....	61
E. Simulations on Logic Circuit	62
F. Peak Current and Leakage.....	65
G. Impacts on Noise Immunity	70
H. Temperature Variations.....	71
I. Area Overhead.....	73
3. Conclusions	75

IV. BUILT-IN DELAY TESTING SCHEME	77
1. Adaptive System Design	77
A. Area-Efficient Built-In Delay Testing Scheme	79
B. Variation Compensation Procedures	84
2. Experimental Validation.....	89
A. Simulation Configurations	89
B. Overheads from Delay Fault Detection Modules.....	90
C. Simulations on Sequential Logic Circuits	94
3. Conclusions	98
V. CONCLUSIONS AND FUTURE WORK	99
1. Conclusions	99
2. Future Work	101
REFERENCES.....	102

LIST OF FIGURES

FIGURE	Page
II.1 Overview of dual-ASV system.	14
II.2 Static power leakage due to the increase of $V_f - V_{DD,L}$	15
II.3 Circuit of Mini Programmable linear voltage Regulator.	18
II.4 Whole circuit power consumption at multiple V_{DD} levels of conventional ASV system.	22
II.5 Waveforms of signals at different voltage levels.	23
II.6 V_f variation vs. the size of driver transistor M_{dr}	23
II.7 Bode plots of MPR considering process variation and aging effect.	25
II.8 Load transient response of V_f	25
II.9 Design topologies to supply a circuit block from a neighboring high $V_{DD,H}$ island.	27
II.10 Illustration for progressive voltage enhancement.	30
II.11 Effects from the granularity of MPRs on circuit resources.	34
II.12 Power consumption comparison for the dual-ASV system.	37
II.13 Power-delay curves for the dual-ASV system.	39
III.1 Overview of the proposed boostable repeater.	43
III.2 Schematics of boostable repeater design.	46
III.3 Charging and boosting operations of boostable repeater.	47
III.4 Transition waveforms from the proposed boostable repeater compared to a conventional repeater.	50
III.5 Waveforms from boostable repeater when it is turned on and off.	52

III.6	Increased voltage at node 2 beyond V_{DD} ($V_2 - V_{DD}$) and signal speed-up acquired through boosting w.r.t. major design parameters.	54
III.7	Power and signal delay variations due to parasitic capacitance.	55
III.8	Signal speed-up w.r.t. the rising time (T_{rise}) of input signal.	56
III.9	Power-delay curves for interconnect networks.....	60
III.10	Power-delay curves to compare the proposed boostable repeater with online adjustable buffer (OAB).	61
III.11	Power consumption comparison for boostable repeater design.....	63
III.12	Power-delay curves for boostable repeater design.	64
III.13	Power reduction vs. V_{DD} levels from boostable repeater compared to adaptive supply voltage (ASV).....	65
III.14	Peak current / leakage current vs. delay.	66
III.15	Performance enhancement degradation due to leakage.....	68
III.16	Noise characteristics of the proposed boostable repeater and conventional repeater.....	69
III.17	CMOS inverter noise margins on 45nm PTM [35].	71
III.18	Circuit performance variations w.r.t. temperature increase.....	72
III.19	Effects from the sizing of C_{pump} on circuit resources.	74
IV.1	Overview of the proposed built-in delay testing scheme with SCAN chain architecture.....	78
IV.2	A common structure of built-in self test (BIST) SCAN chain architecture with muxed-D SCAN cells.	81
IV.3	Control signals and internal waves of the proposed built-in delay testing scheme.....	83
IV.4	Variation mitigation through dynamic clock skew control.	85
IV.5	Area and power overhead comparison between the three delay fault	

detection approaches.....	91
IV.6 Power consumption comparison for the proposed adaptive system.	97

LIST OF TABLES

TABLE	Page
II.1 V_{ref} and V_f variations due to process variations and aging effects.	19
II.2 Power consumption of MPR vs. whole circuit power in dual-ASV system.	20
II.3 V_f fluctuation while powering dynamic load of S526 and S1423 ISCAS'89 circuits with process variation and aging effects.	26
II.4 Circuit performance comparison between design topologies to power a circuit block.	27
II.5 Circuit characteristics and experimental setup for dual-ASV system.	33
II.6 Voltage configurations and average results on power dissipation.....	35
III.1 Benchmark circuit configurations with deployment of the proposed boostable repeaters.....	58
III.2 Voltage configurations and average results on power dissipation from simulations on interconnect networks.....	59
III.3 Voltage configurations and average results on power dissipation from simulations on ISCAS'89 benchmark circuits.....	63
IV.1 Signal delay variations of sequential benchmark circuits due to process variation.	87
IV.2 Estimated maximum power reduction achievable through δ_D gain in T_{CLOCK} through dynamic clock skew control.	87
IV.3 Circuit characteristics and experimental setup for the proposed adaptive system.....	89
IV.4 Die area occupied by one individual testing module of three delay fault detection approaches.....	92
IV.5 Average power consumption during testing operation period.....	93

IV.6	Energy consumption during testing operation period.....	93
IV.7	Power reduction acquired though the proposed adaptive system.	96

I. INTRODUCTION OF RESEARCH

1. Introduction

In nanometer regime, process variations and circuit aging [1] cause remarkable, and often unwanted, uncertainty in circuit system characteristics. How to efficiently harness the variation effects remains a major challenge to be solved. A straightforward approach for variation tolerance is over-design. If guard-bands are large enough, they can ensure that performance specifications are satisfied in presence of the worst case variations. However, they often entail large design overhead and waste substantial resources, especially power, in typical cases, especially during initial lifetime of circuits. Statistical techniques are primarily to reduce the pessimism of guard-bands, but cannot reduce the variations therefore they do not solve the fundamental inefficiency of the over-design. A more fundamental strategy is to equip circuits with the capability of tuning themselves and thereby compensating the variations, i.e., adaptive circuit design.

This thesis presents efforts on adaptive system designs for power-efficient resilience to process and aging induced timing variations. In general, an adaptive circuit contains two major components: variation detection and variation compensation. We will focus on developing both variation detection and compensation techniques in order to provide a complete adaptive system. Various existing adaptive approaches will be reviewed, and their limitations to fine-grained variation compensation will be analyzed as well. To effectively address the limitations, we will propose voltage and frequency adaptation techniques, and an area-efficient delay variation detection scheme.

2. Existing Methods and Their Limitations

A. Adaptive Body Bias vs. Adaptive Supply Voltage

There are two main approaches for variation compensation: adaptive body bias (ABB) [2, 3] and supply voltage adaptation [4, 5]. Adaptive body bias (ABB) is a well-known adaptive technique which tunes body voltage to control transistor threshold voltage. If variation effect is strong, ABB can either lower transistor threshold voltage to restore performance or increase the threshold voltage to reduce leakage power. Supply voltage adaptation, a.k.a. adaptive supply voltage (ASV), can compensate variations by changing supply voltage. ASV has several advantages over ABB. First, ASV can be applied to almost any kind of circuits while ABB is difficult to be applied on SOI (Silicon-On-Insulator) circuits. Second, the tuning range of ABB is limited because of junction leakage current [6]. Third, ABB affects only leakage power while ASV can change both leakage and dynamic power. The leakage power (P_{LEAK}) and dynamic (switching) power (P_D) can be expressed as,

$$P_{LEAK} = I_{LEAK} \bullet V_{DD}$$
$$P_D = \alpha f C V_{DD}^2$$

[7], where I_{LEAK} stands for the cumulative leakage current through circuit, α is activity factor, f is operating frequency, and C is for load capacitance. ASV reduces P_D as well as P_{LEAK} by tuning V_{DD} . Overall, ASV is a stronger and more sustainable leverage than ABB.

On the other hand, the implementation of ASV is more difficult than ABB. In ABB, once a body voltage reaches the desired level, it experiences only small perturbation from leakage current. Thus, the power supply for the body bias does not need to be strong. People can use simple and small circuit, like voltage divider, to achieve ABB at fine granularity of individual wells [3]. In contrast, ASV needs to accommodate large and quick current withdraw from transistors. Hence, the implementation of ASV usually requires voltage regulators, which are either complex or bulky, and is restricted to coarse granularity of chip-level or large-block-level [4, 5].

B. Granularity of ASV

The process variations in nanometer technology are often fine-grained. Two transistors a few microns apart may have different litho-variations and random doping fluctuations. Likewise, two neighboring transistors may have different degree of aging due to different switching activities. In typical designs, circuit optimizations are performed and result in many equally critical paths. Due to their probabilistic nature, the variations may manifest strongly on either a few or most of the critical paths.

The fine-grained variations together with the large number of critical paths present a dilemma for deciding the granularity of conventional ASV. If ASV is applied at a coarse-grained level, it is likely that the supply voltage is raised due to variations on a few paths. Then, the extra power spent on those paths without strong variations is largely wasted. If one uses fine-grained ASV, i.e., many small ASV domains, there

could be large overhead on supply voltage generation, which will be further discussed in Section I.2.D.

C. Dual Static Supply Voltage

Recently, dual static supply voltage based adaptation techniques are reported [8, 9, 10]. They assume that two power supply (V_{DD}) lines are available to a circuit block. The circuit block can be adaptively connected to either high or low V_{DD} through sleep transistors. The difference between the two supply voltages is small so that there is no need to use level shifters as for voltage islands [11]. These works did not show details on how to obtain the two different power supply lines at the same place, which is a difficult task. When the block sizes are small, i.e., in fine granularity, this approach implies nearly doubling of power grid wires. Power grid in modern chip designs is already huge, complex and has very limited room for additional overhead.

D. Overhead on Supply Voltage Generation

Both fine-grained ASV and dual static supply voltage require more than one supply voltages. Conceivably, there are only two options to obtain a supply voltage:

- Option 1: generate it from an off-chip voltage regulator and delivery it to on-chip destinations;
- Option 2: generate it locally using on-chip voltage regulator.

Both involve voltage regulator which has two categories: switching regulator and linear regulator. Linear regulator is compact, relatively easy to be integrated on-chip and has

fast response [12, 13]. However, its energy-efficiency can be low compared to switching regulators, especially when voltage difference between input supply and regulated output is large [14]. Switching regulator can be divided into two categories, switched-mode regulator [15] and switched-capacitor converter [16]. Both types have better energy-efficiency [15], but usually entail very large passive elements such as inductors or capacitors, which make them very difficult to be integrated on-chip.

For dual static supply voltage, if one obtains the additional voltage through option 1, there would be a large power delivery network overhead due to duplicated supply lines. For example, if a half of the entire circuit is powered by the dual static supply voltage, the size of power delivery network would increase by 50%. In current chip designs, the power supply network for even single voltage level is already very complex and heavily loaded. Hence, the room for additional power delivery lines is very small. If one chooses option 2, there are also problems. Switching regulator is difficult to implement on-chip. For linear regulators, small ones with limited output load current capacity are not sufficient to compensate large scale variations while large ones supporting high output load current cause too much power waste as well as stability issues which in turn increase complexity of regulator design.

For fine-grained ASV, there is no obvious good solution either. If one goes with option 1, the overhead on off-chip regulators would be huge. Consider a chip with a half million gates. If each block of 5k gates has its own ASV, then the chip needs 100 regulators each of which should supply all the 5k gates in each block. Option 2 is far from being practical as well since a large number of on-chip regulators each of which

powers 5k gates cause either huge area overhead (from switching regulators) or large power waste (from linear regulators).

E. Online Adjustable Buffer

Online adjustable buffer (OAB) [17] is a static CMOS buffer augmented by a tri-state buffer in parallel. The tri-state buffer is turned on (off) when the OAB operates for high-speed (low power). In order to significantly reduce delay at high-speed-mode, the size of the tri-state buffer needs to be close to that of the static CMOS buffer. This means that an OAB at low-power-mode has intrinsic load of twice as much as that of a single static CMOS buffer. In other words, this technique improves one mode at the cost of degrading another mode. Another related work is dual- V_{DD} buffer [18]. However, it requires two V_{DD} lines like [10].

3. Proposed Adaptation Techniques

A. A Dual-Level Adaptive Supply Voltage (ASV) System

In this work, we propose a new Adaptive Supply Voltage (ASV) system for circuits with many timing critical paths. In typical design flows, circuits are optimized to suppress the delay of timing critical paths and reduce the power on non-critical paths. Consequently, path delays tend to be equalized and there are many paths with similar timing criticality. This phenomenon implies a wide range of possibilities for variation-induced timing degradation: from a few paths to many paths. This wide range makes

ASV very difficult to implement. If one adopts coarse-grained ASV, there could be considerable power waste when the degradation actually occurs only on a few paths. If one chooses fine-grained ASV and prepares for degradations on many paths, a large overhead on power routing or voltage regulators is usually incurred.

Our system solves the aforementioned difficulty by simultaneously providing coarse-grained and fine-grained ASV. In this dual-level ASV (dual-ASV) system, power routing overhead is largely avoided by a new technique of voltage tapping in the context of voltage island based designs [11, 19]. A progressive voltage enhancement method is suggested to further improve the efficiency of this system. In experiments, we compared the dual-ASV system with the over-design and conventional ASV method [4, 5]. SPICE simulations are performed on benchmark circuits with consideration of process variations and NBTI (Negative Bias Temperature Instability) induced aging effect [1]. We also considered the impact of the variations on the dual-ASV system embedded in the circuits. The results indicate that our approach can achieve similar performance and robustness with significantly less power dissipation. The average power reductions are 40% and 21% compared to over-design and conventional ASV, respectively. A weakness of our approach is that it is limited to voltage island based designs. However, multi-supply-voltage designs become increasingly popular for the sake of power management. Therefore, the limitation of our approach will become less severe. Overview and detailed implementation of the proposed dual-ASV system, and its corresponding analysis and verification will be presented in chapter II.

B. Boostable Repeater for Variation Tolerance in Interconnects

In this work, we propose a new approach for fine-grained voltage adaptation – boostable repeater, which can boost its switching speed through transiently and autonomously enhancing its internal voltage rail. The boosting feature can be turned on and off at runtime and therefore can adaptively compensate delay variations. It does not require voltage regulator or additional power grid. Since its adaptivity is per repeater, it is inherently fine-grained. A previous technique in the same direction is online adjustable buffer (OAB) [17]. Roughly speaking, an OAB is a static CMOS buffer augmented by a tri-state buffer in parallel. The tri-state buffer is turned on (off) when the OAB operates for high-speed (low power). In order to significantly reduce delay at high-speed-mode, the size of the tri-state buffer needs to be close to that of the static CMOS buffer. This means that an OAB at low-power-mode has intrinsic load of twice as much as that of a single static CMOS buffer. In other words, this technique improves one mode at the cost of degrading another mode. Unlike OAB, which relies on dynamic device sizing, our boostable repeater exploits transiently higher voltage rail for speed improvement. Therefore, the extra load presented to timing path from a boostable repeater is significantly smaller. Another related work is dual- V_{DD} buffer [18]. However, it requires two V_{DD} lines like [10].

The boostable repeater design enables fine-grained circuit adaptation and therefore power-efficient resilience to variations. It presents limited load overhead to timing paths and thus has small timing penalty in low power mode. The weakness is that it has significant device area overhead. However, this overhead can be largely alleviated

through careful applications. That is, boostable repeaters are placed in only a small number of judiciously selected nets so as to limit the area overhead. When the variations on a fabricated circuit are widespread, a global supply voltage tuning will be more effective. Hence, the boostable repeater design is a complement rather than a replacement to coarse-grained adaptation. Since interconnect is a well-known bottleneck for chip performance and a modern chip design may contain hundreds of thousands of repeaters [20], there are plenty of opportunities for boostable repeaters to find critical places and exert significant impact.

Besides variation resilience, boostable repeaters can be applied for dynamic power management due to its online programmability. By turning on/off the boosting, one can tune circuits to work in high performance or low power mode.

We composed an adaptive system which combines the proposed boostable repeaters with global ASV system. This is because we propose boostable repeater to compensate fine-grained variations, in other words, severe degradations on a limited number of critical paths. In conjunction with ASV, we maintain global V_{DD} low while significantly degraded paths can still meet their timing constraints with the boostable repeaters.

The proposed adaptive system with boostable repeaters integrated with global ASV system is validated through SPICE simulations on various test cases. Compared to ASV alone, our approach achieves an average of more than 25% power reduction with the same performance and robustness. When applied in a logic circuit, the device area overhead of our approach is about 4~5% which takes all the additional devices from the

proposed approach into account. Our approach also significantly outperforms online adjustable buffer [17]. Main idea of implementing the proposed boostable buffer, and detailed analysis and verification of its benefits will be provided in chapter III.

C. Built-in Delay Testing Scheme

This work presents an area-efficient built-in delay testing scheme utilizing BIST SCAN architecture [21] and dynamic clock skew control. A low cost scheme for delay fault detection is essential to provide an efficient adaptive circuit design for mitigating fine-grained variation, since the granularity of adaptive components critically depends on the number of variation detection components to be deployed throughout circuit blocks. To meet the stringent area as well as power constraints of modern circuit design, the built-in testing components are expected to consume as small amount of circuit resources as possible, especially die area.

Previous works provided various schemes of on-chip delay measurement to detect circuit variations. Vernier Delay Line presented in [22] senses delay perturbation by digitizing a path delay through chains of FFs and buffers. Although this type of measurement yields a fine resolution of delay value of each path, it induces huge area overhead and is not adequate for fine-grained variation detection. Razor, a double-sampling flip-flop, presented in [23] is a well known and simple technique which compares output value from master latch against one from shadow latch fed by delayed clock to detect delay error. This approach, however, doubles the size of regular FF in order to embed shadow latch which is not a trivial overhead as well. The shadow latch

even significantly entails extra capacitive load to the timing path. Another approach to predict delay fault is proposed by [24] which senses signal transitions within a guard-band. This technique also requires a huge area overhead due to its complex structure of stability checker and delay element.

Adjusting system clocks with tunable delay buffers [25] is a well studied approach to adapt clock periods for system tuning. K. Nagaraj [26] proposed a clock tuning technique to compensate process variations through adjusting clock arriving times to each FF. They addressed methodologies to select FFs to be tuned and to determine their required clock tuning range in accordance with a faulty system response from SCAN chain output. However, they focused merely on clock period reductions, and omitted discussions over any achievable power gains or extra resource overheads necessitated by their system. Dynamic clock retiming approach utilizing Razor FF is presented in [27]. In their work, Razor FF detects timing error, and local clock controller tunes clock skews between pipeline stages to minimize clock period. Their discussion, however, does not cover circuit variation handling, and they applied Razor FF which may also cause significant area overhead and capacitive extra load as discussed above. A pipeline adaptation technique to mitigate process variation is proposed by [28], but they also presume overheads from variation sensors are negligible which may not be true in fine-grained variation compensation.

In this work, we propose an adaptive system with very low overheads on area as well as power. Main idea of our system to minimize the overheads is to multi-use its components as much as possible. The proposed built-in delay testing scheme exploits

SCAN chain system which is a standard testing component in modern VLSI architectures, therefore we can reuse it for delay fault detection circuitry as well. Unlike the error detection scheme through SCAN BIST architecture which needs complex procedure to shift the SCAN results out and post-process the results, the proposed approach flags error signal indicating delay fault directly from each testing module. This error signals are directly connected to the compensation circuitry to handle variations easily. We also embed dynamic clock skew control in the proposed system to conduct two distinctive roles simultaneously: detection operation and compensation function to mitigate delay fault. Through these multiple use of adaptive components, circuit designs utilizing the proposed system can significantly reduce extra resources required by the adaptive circuitry.

We demonstrates test bench circuits equipped with the proposed built-in delay testing scheme combined with boostable repeater design and adaptive clock skew control. According to the results from SPICE simulations, the proposed adaptive system detects and mitigates process and aging induced timing degradations in a power as well as area efficient manner. When applied in a logic circuit, the proposed adaptive system causes 7.2% of extra area overhead from the adaptive circuitries, while an adaptive design with Razor FF induces 12.7% of area overhead. When it comes to compensating the variations in both fresh circuit and aged circuit, the proposed system achieves 26% less power consumptions in comparison with circuit design with global adaptive supply voltage (ASV) system alone.

II. A DUAL-LEVEL ADAPTIVE SUPPLY VOLTAGE (ASV) SYSTEM

1. Rationale and Overview of Dual-ASV System

The proposed solution addresses the two aforementioned problems:

- 1) How to handle the granularity uncertainty in variations?
- 2) How to reduce the supply voltage generation overhead?

We propose a dual-ASV system to solve the problems. In this system, each selected circuit block has two power supply lines – one belongs to a global and coarse-grained ASV and the other is a local and fine-grained ASV. This system has the following two advantages:

- 1) A few paths of variations can be handled by the fine-grained ASV and widespread variations can be taken care of by the coarse-grained ASV. Thus, this system can solve the granularity uncertainty problem.
- 2) The overhead of the coarse-grained ASV is no greater than conventional ASV. With the help from the coarse-grained ASV, the fine-grained ASV can be focused on a small number of gates and therefore allows low overhead of power and delivery network for supply voltage generation.

In a dual-ASV system, usually one supply voltage is higher than the other. Then, which one is delivered at global (coarse-grained) level? We choose to use the lower supply voltage at global level and the higher voltage at local level. The main reason is that local voltage generation is mostly obtained on-chip and difficult to have high energy-efficiency. If low supply voltage is applied locally, the power savings from the

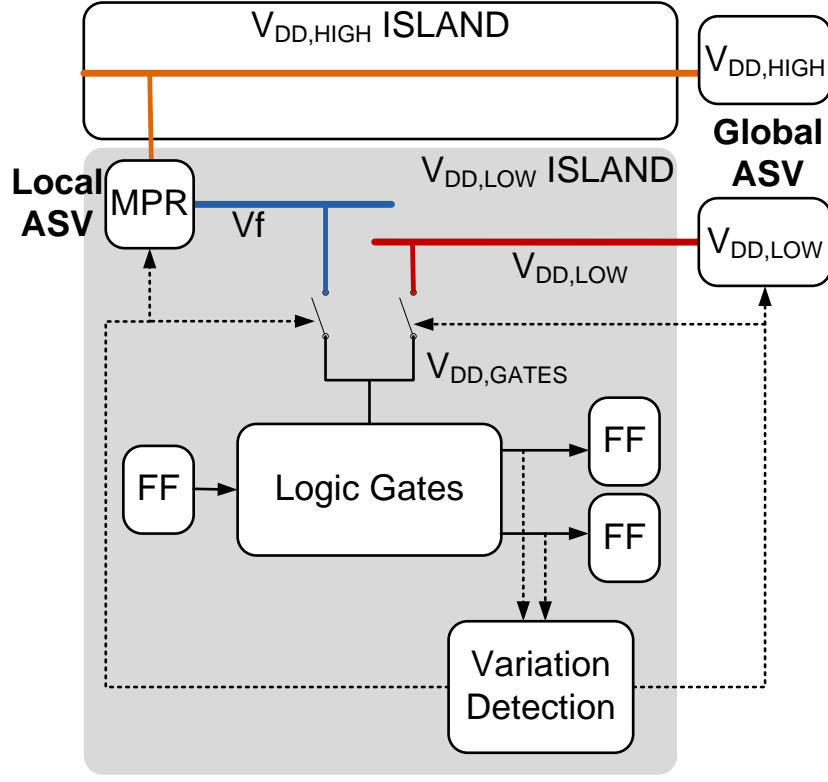


Figure II.1. Overview of dual-ASV system. MPR is the proposed mini programmable linear voltage regulator.

low supply voltage is largely cancelled by the waste in the supply voltage generation. In contrast, if high voltage is applied locally, its power waste can be overshadowed by the power savings from global application of low voltage.

The proposed system is illustrated in Figure II.1. It is in the context of voltage island based designs [11, 19]. In the low V_{DD} island, an additional power supply is obtained by tapping off an intermediate voltage level V_f from $V_{DD,H}$ of its neighboring high V_{DD} island. The level of V_f is somewhere between $V_{DD,H}$ and $V_{DD,L}$. V_f is supplied only to the critical paths in the low V_{DD} island. In a rare but possible case when our

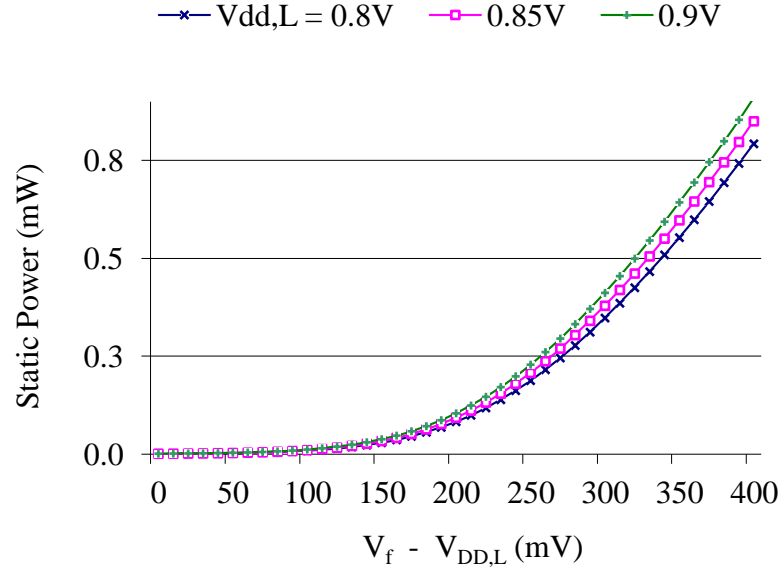


Figure II.2. Static power leakage due to the increase of $V_f - V_{DD,L}$.

ASV system is temporarily unable to find a nearby $V_{DD,H}$ island for tapping V_f due to dynamic voltage control of each voltage island, our system can simply utilize global ASV to raise $V_{DD,L}$.

The difference of $V_f - V_{DD,L}$ is small so that V_f and $V_{DD,L}$ can be applied to the same circuit block without using level shifters. This voltage difference is critical because it determines the efficiency of the dual-ASV system. Larger difference will allow more power room for the overhead of generating V_f . The difference, however, will also increase the static current leakage at a gate supplied by V_f whose logical input ‘1’ from $V_{DD,L}$ cannot turn the pMOS of the gate off completely. Figure II.2 exhibits that static power leakage exponentially ascends in accordance with the increase of the difference of supply voltages, $V_f - V_{DD,L}$, between two inverters. Based on the graph, we can

determine that the voltage difference, $V_f - V_{DD,L}$, acceptable in our system considering static power leakage should be less than 150mV. In case of strong variations where V_f needs to be raised higher than $V_{DD,L} + 150\text{mV}$, we increase $V_{DD,L}$ through the global ASV system as depicted in Figure II.1 in order to maintain $V_f - V_{DD,L}$ less than 150 mV. Similar issues of static leakage current due to multiple supply rails in a circuit block are discussed and addressed in previous researches regarding voltage interpolation [10].

A delay variation prediction circuit [24] is employed. It can generate a warning signal if a delay variation is large and close to timing error. With the warning signal, we can efficiently use V_f to save overall power consumption as well as to minimize load to V_f so as to restrict overhead from generating V_f . Suppose V_f is designed to deliver power to totally up to N critical paths in a circuit block. When less than N critical paths flag variation warnings, these paths are switched from $V_{DD,L}$ to V_f . If more than N critical paths have warnings, $V_{DD,L}$ is raised like in conventional ASV. Therefore, we can achieve ASV at two levels: $V_{DD,L}$ at coarse-grained level and V_f at fine-grained level. Detailed descriptions about how we efficiently control V_f in conjunction with variation warning signals are discussed in Section V. In our experiment circuits presented in Section VI, we adjust $V_{DD,L}$ through coarse-grained ASV if more than 10~15% of paths flag warning.

When tap V_f from $V_{DD,H}$, we use linear regulators [12]. Since V_f is generated locally, there is no significant overhead on power delivery. Although the energy-efficiency of linear regulators is can be low, the overall power overhead is small as they are applied in a restricted manner. The small power waste at the linear regulators is

exchanged for a large power reduction at the rest of the circuits which operate at a lower supply voltage. Overall, the power-efficiency is improved compared to conventional methods.

We focus on process variations and circuit aging. Since the former is static and the latter is a very slow change, the voltage adaptation can be performed offline either at power-on or periodically. For the offline tuning, a test pattern generator is needed. One can either employ LFSR (Linear Feedback Shift Register) to generate test vectors or use test vectors saved in memory.

2. Voltage Tapping

Voltage tapping is to obtain an intermediate voltage level from a high voltage island and supply it to its neighboring low voltage islands. Then, the low voltage islands have dual voltage supplies. The voltage tapping is achieved by voltage down converting using a voltage regulator. We choose to use linear regulator since it is easy to be integrated on-chip and has fast response. Since it is applied locally and at small scale, the inefficiency of linear regulator does not reverse the global efficiency of the entire dual-ASV system. There are various designs for on-chip linear regulator [13]. Most of them are complex since they aim to accommodate very large current (about 100mA). Since our application is at local supply voltage powering small current load of few milliamperes with assistance from a global ASV, we prefer a simple design.

We choose a design similar to [12] and propose a Mini Programmable linear voltage Regulator (MPR) for a relatively small load current. The transistor level

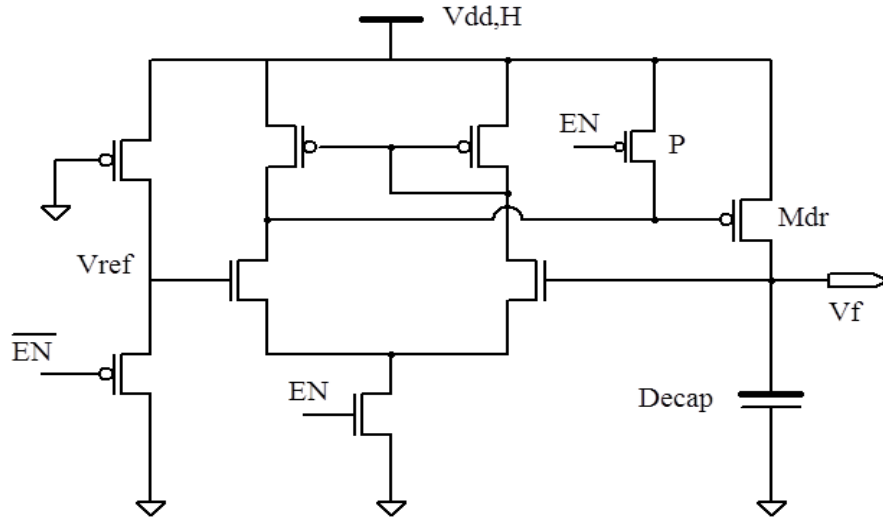


Figure II.3. Circuit of mini programmable linear voltage regulator (MPR).

schematic of MPR is depicted in Figure II.3. This circuit consists of 3 stages. The first stage is a voltage divider which generates the reference voltage. Compared to [12], the reference voltage generator is greatly simplified. In Figure II.3, a single voltage is generated. In practice, one can easily extend it for generating multiple reference voltages by parallelizing or cascading the voltage divider. By switching among different reference voltages, the output V_f level can be dynamically changed and therefore provides more options on variation compensation. Certainly the voltage divider is composed of transistors which are also vulnerable to the variations we are now considering. Table II.1 shows V_{ref} and V_f variations of test cases caused by process variations and aging effects. The deviations of V_{ref} and V_f with respect to averages are maximum 3.41% and 3.02% respectively, and those are reasonably acceptable. For more reliability against variations, one can easily replace this voltage divider circuit with a bandgap reference [29, 30] which generates a temperature as well as process variation

Table II.1. V_{ref} and V_f variations due to process variations and aging effects.

V_{th} degra- dation	$V_{\text{DD,L}}$	$V_{\text{DD,H}}$	V_{ref}						V_f				
			designed V_{ref}	Average	Max	$+\Delta$ (%)	Min	$-\Delta$ (%)	Average	Max	$+\Delta$ (%)	Min	$-\Delta$ (%)
Fresh Circuit	0.75	1.1	0.85	0.852	0.880	3.24	0.829	2.70	0.849	0.869	2.41	0.832	1.94
	0.8		0.9	0.896	0.920	2.68	0.877	2.18	0.894	0.910	1.88	0.880	1.51
	0.85		0.95	0.949	0.968	1.99	0.934	1.59	0.943	0.954	1.14	0.934	0.94
	0.9		1	1.014	1.026	1.13	1.005	0.92	0.993	0.995	0.26	0.989	0.35
Aged Circuit	0.75	1.1	0.85	0.866	0.896	3.41	0.840	3.02	0.858	0.878	2.34	0.840	2.03
	0.8		0.9	0.908	0.934	2.83	0.885	2.51	0.900	0.915	1.66	0.886	1.51
	0.85		0.95	0.958	0.978	2.10	0.940	1.87	0.945	0.952	0.73	0.938	0.79
	0.9		1	1.019	1.031	1.20	1.008	1.07	0.990	0.994	0.35	0.985	0.55

Table II.2. Power consumption of MPR vs. whole circuit power in dual-ASV system.

V_{th} degra- dation	V_f	Total (mW)	MPR			
			On (mW)	Overhead (%)	Off (μ W)	Overhead (%)
Fresh Circuit	0.85	4.137	0.191	4.60	0.127	0.003
	0.9	4.502	0.194	4.31	0.126	0.003
	0.95	5.272	0.198	3.75	0.124	0.002
	1	6.197	0.204	3.29	0.122	0.002
Aged Circuit	0.85	4.137	0.191	4.60	0.127	0.003
	0.9	4.502	0.194	4.31	0.126	0.003
	0.95	5.272	0.198	3.75	0.124	0.002
	1	6.197	0.204	3.29	0.122	0.002

independent reference voltage. The middle stage of MPR is an opamp-based voltage follower. The last stage includes an output driver Mdr and a decap. The driver provides current to load and the decap is to reduce supply voltage noise.

A main change compared to [12] is that we make it programmable. This is achieved by the EN signal and a control transistor P in Figure II.3. MPR is turned off when EN is low. By applying the EN signal, one does not need to worry about the standby current like in the conventional regulator designs [12]. The standby power of MPR when it is turned off is presented in Table II.2. The column ‘Off’ is for the standby power, and it is almost negligible compared to the whole circuit power. This is one reason that we use such simple design. In the dual-ASV system, the EN signal is not an

overhead since we need to have it to control the power supply switching between V_f and $V_{DD,L}$ anyway. In other words, the EN signal is already in our system regardless of the design of the regulator. Additional benefit of the EN signal is that the switch between V_f and the logic circuit in Figure II.1 can be skipped. Usually such switch is implemented by pMOS sleep transistors. Removing the sleep transistors not only reduces area/power overhead but also allows performance improvement. This is because the voltage drop across the sleep transistor is avoided and the logic circuit can be powered by V_f directly instead of degraded V_f .

Despite the inefficiency of each MPR, the global efficiency of the entire dual-ASV system can be achieved through the supply voltage difference, $V_f - V_{DD,L}$. In accordance with the equations given in the section I.2.A from [7], $P_{TOTAL} (= P_{LEAK} + P_D)$ reduces if we lower down V_{DD} . Supplying a whole circuit block with $V_{DD,L} (< V_f)$, and powering V_f only to the degraded critical paths in the block will naturally consume less power than supplying $V_f (> V_{DD,L})$ throughout a whole circuit block, provided that the power difference allows room for MPR operating power. (Please note that, without adaptive technique, this circuit block requires V_f supply in order for all the degraded paths to meet timing constraint.) In addition, the efficiency of each MPR can be enhanced if we choose the level of $V_{DD,H}$ as close to V_f as possible among nearby voltage island, since the efficiency of linear regulator raises with low-dropout voltage [14]. The power overhead of one MPR vs. whole circuit power in dual-ASV system is given in Table II.2. When it is turned on, MPR itself consumes around 0.2mW, 3~5% of power overhead, which is not negligible. However, the global power saving from

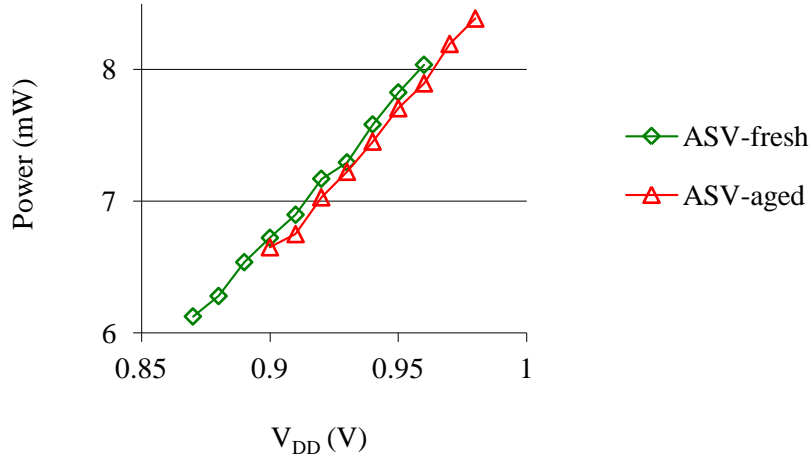


Figure II.4. Whole circuit power consumption at multiple V_{DD} levels of conventional ASV system. Dual-ASV system is not applied here.

lowering down $V_{DD,L}$ can overshadow the power overhead. Figure II.4 shows the tendency of power consumption in conventional ASV system in accordance with adjusting V_{DD} level. In the slopes, if we decrease V_{DD} level by 100mV, the circuit consumes almost 2mW less than when it is supplied by higher voltage. Though the 2mW of power saving is not totally what we can solely use for our purpose only, we can still say that lowering $V_{DD,L}$ of the majority of circuit by 100~150mV will allow enough power budget to generate few but sufficient number of V_f rails through several MPRs. This is the main idea of our dual-ASV system to achieve global efficiency.

Although the design of the regulator MPR here is simple, it works very well for the voltage tapping where the typical current is around 1mA. In Figure II.5, we show transient waveforms from SPICE simulations. In this case, $V_{DD,L}=0.8V$, $V_{DD,H}=1.1V$ and $V_f=0.95V$. One can see that the logic signal powered by V_f switches very well, not

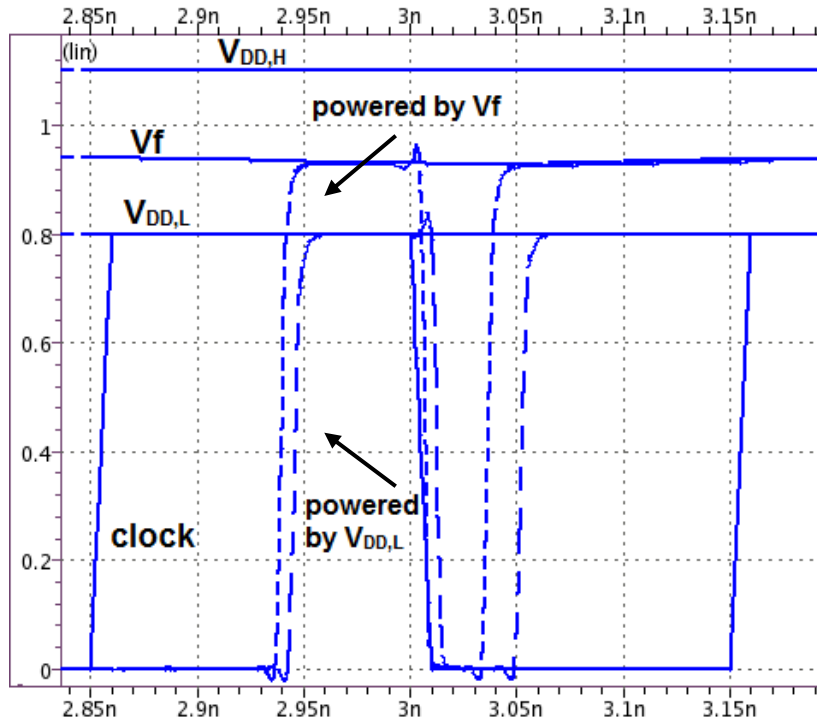


Figure II.5. Waveforms of signals at different voltage levels.

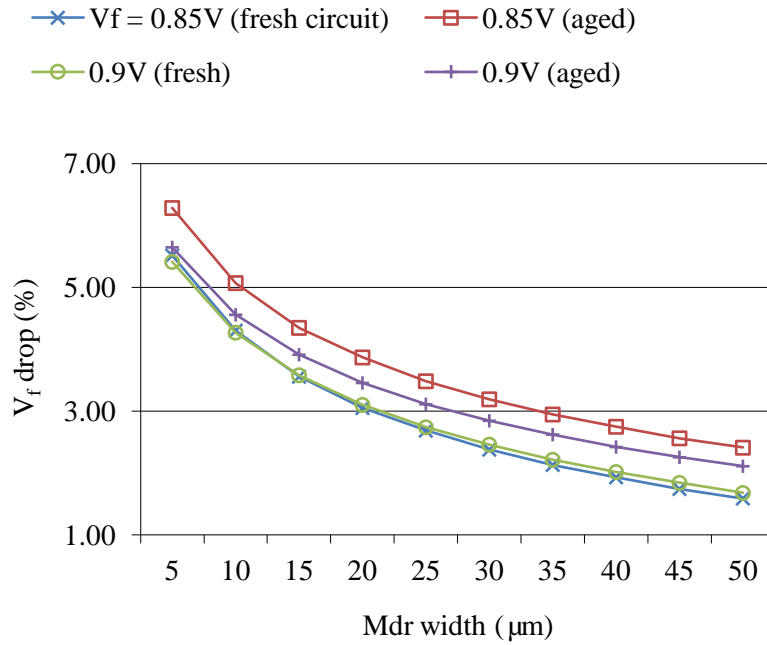


Figure II.6. V_f variation vs. the size of driver transistor M_{dr} .

much different from the signal powered by $V_{DD,L}$. A key role of linear regulator is to stabilize V_f while supplying power to dynamic load. In considering this point, we investigated the impact of the size of driver transistor M_{dr} of MPR design given in Figure II.3. The results are depicted in Figure II.6. We observe the effect on the voltage drop of V_f , which is very important to the performance and predictability of the logic circuits. The V_f drops decrease exponentially with the increase of the size of M_{dr} . In order to safely guarantee that all the circuits under the variations have V_f drops less than 5%, the transistor width of M_{dr} is set as 20 μm . Figure II.7 shows bode plots of the regulator MPR where we considered both process variations (gate length variation: $3\sigma=15\%$, V_{th} variation: $3\sigma=20\%$) and aging effects (additional V_{th} variation: $\mu=10\%$ $3\sigma=10\%$) through 30 Monte-Carlo iterations. The worst case phase margin of 71° confirms that our proposed MPR design is stable. A load transient response from MPR for a transient load current of $50\mu A \sim 1\text{ mA}$ with 50ps of rising and falling times is given in Figure II.8, and MPR shows 3.5% of load regulation and 16mV of undershoot. Table II.3 shows actual V_f voltage drops (V_{drop}) in our test cases with S526 and S1423 ISCAS'89 benchmark circuits, considering both process variation (gate length variation: $3\sigma=15\%$, V_{th} variation: $3\sigma=20\%$) and aging effects (additional V_{th} variation: $\mu=10\%$ $3\sigma=10\%$). In those cases, MPR regulates V_f for dynamic loads varying from $50\mu A$ to $500\mu A$ range. In consideration of circuit variations, the worst case drop is less than 4% compared to average which is smaller than the typical specification of 5%.

An alternative approach to supply V_f to an individual circuit block is to power it directly through $V_{DD,H}$ and attach level shifters at the inputs and outputs of the block.

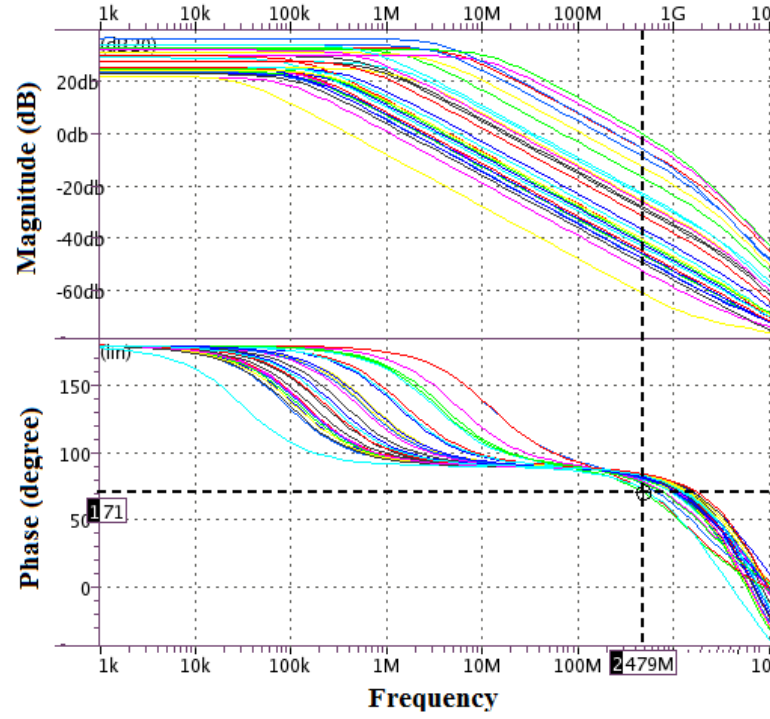


Figure II.7. Bode plots of MPR considering process variation and aging effect. In all cases, phase margin is higher than 71° which confirms the stability of MPR.

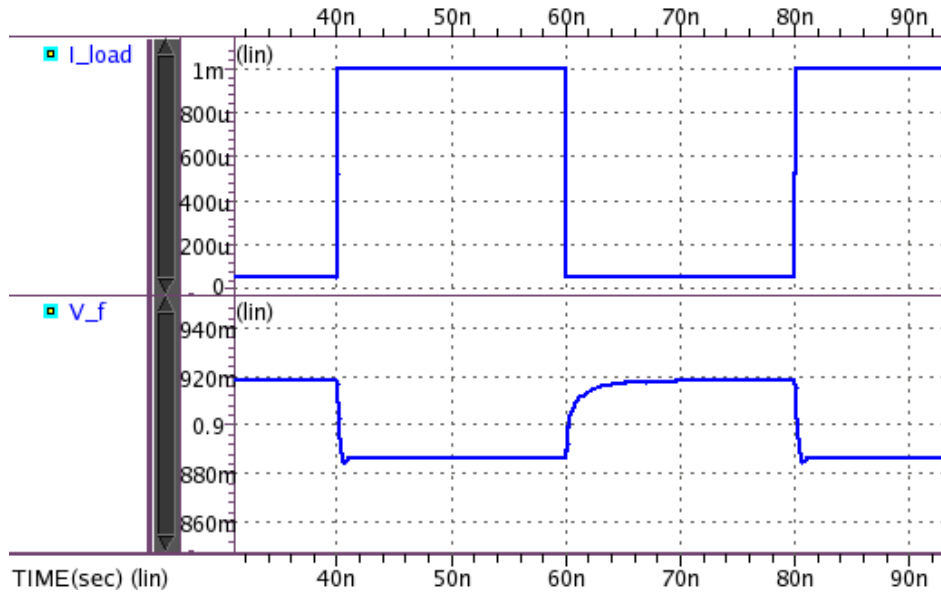


Figure II.8. Load transient response of V_f ($I_{load} = 50\mu A \sim 1\text{ mA}$).

Table II.3. V_f fluctuation while powering dynamic load of S526 and S1423 ISCAS'89 circuits with process variation and aging effects. V_f is regulated voltage output from proposed Mini Programmable linear voltage Regulator (MPR).

V_{th} degradation	designed	V_f				
	V_{ref}	Average	Vpeak	Δ (%)	Vdrop	Δ (%)
Fresh Circuit	0.85	0.849	0.875	3.11	0.817	-3.76
	0.9	0.894	0.916	2.56	0.862	-3.51
	0.95	0.943	0.960	1.76	0.914	-3.12
	1	0.993	1.001	0.88	0.967	-2.59
Aged Circuit	0.85	0.858	0.883	2.97	0.824	-3.87
	0.9	0.900	0.920	2.27	0.869	-3.46
	0.95	0.945	0.958	1.31	0.919	-2.81
	1	0.990	1.000	0.97	0.964	-2.72

This approach, however, inevitably wastes extra power if it boosts up the block performance faster than its target timing constraints. Figure II.9 shows design topologies to supply V_f from $V_{DD,H}$, a high voltage rail from a neighboring voltage island: one is powering V_f through MPR so that $V_f < V_{DD,H}$, and the other topologies are attaching $V_{DD,H}$ directly to V_f , therefore $V_f = V_{DD,H}$. For the second case of attaching $V_{DD,H}$ directly to V_f , the effects from level shifters on the circuit performances are also examined and compared. Table II.4 analyzes power consumptions from each block as well as from total circuit, and power dissipations from MPR and level shifters are taken into account. According to Table II.4, power consumption reduces if a block is powered by MPR which lowers V_f voltage level down to 0.9V, even though MPR extracts power from $V_{DD,H}$ as well. This is because lowering down V_f reduces voltage swings at the

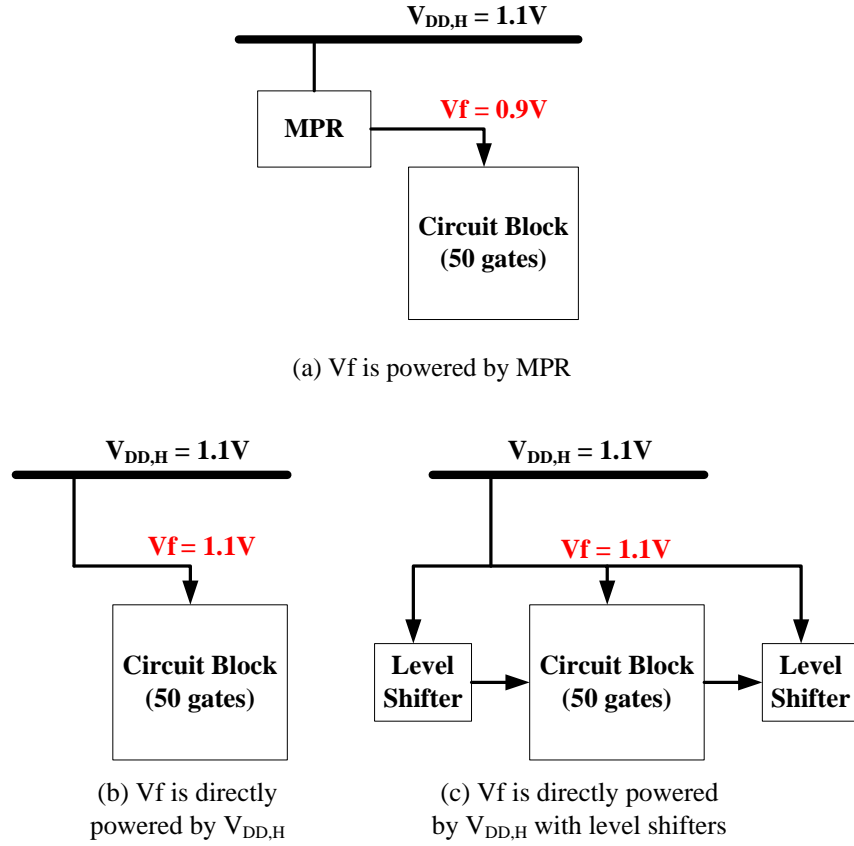


Figure II.9. Design topologies to supply a circuit block from a neighboring high $V_{DD,H}$ island. The top topology (a) shows that MPR reduces V_f down below $V_{DD,H}$ so that block power consumption can decrease. Power dissipations from MPR and level shifters are all considered in the power comparison presented in Table II.4.

Table II.4. Circuit performance comparison between design topologies to power a circuit block ($V_{DD,H} = 1.1V$).

Circuit	V_f is powered by	V_f (V)	Delay (ps)	Power consumption	
				Block (μW)	Total (mW)
S1423	MPR	0.9	182.0	459.8	4.50
	$V_{DD,H}$	1.1	163.4	760.8	4.83
	$V_{DD,H}$ + level shifters	1.1	172.2	794.9	4.85
S5378	MPR	0.9	183.7	554.9	11.25
	$V_{DD,H}$	1.1	179.9	891.4	14.44
	$V_{DD,H}$ + level shifters	1.1	182.5	909.6	14.46

logic gates in the circuit block which in turn consumes less power. This proves that, in terms of fine-grained adaptation, power can be further saved by using MPR if we only need to meet a certain timing constraint (e.g., 182ps for S1423, and 184ps for S5378), since attaching $V_{DD,H}$ directly as a supply source will waste extra power. Besides, due to power overhead and intrinsic delay from level shifters, inserting level shifters slightly degrades performance as well, according to the rows for “+ level shifters” in Table II.4.

3. Progressive Voltage Enhancement

This section introduces how to utilize the two dynamic voltage levels for variation resilience in different scenarios: from a few paths to many paths of variation assertions. We propose two different ways of connections with the voltage tapping output V_f .

- 1) Selective connection with single-path block. A single-path block is a set of logic gates which are on one or only a few (e.g., less than 3) critical paths. One MPR is connected to multiple single-path blocks through switches (sleep transistors). But, only the connections to one or two blocks can be turned on at the same time due to the limited power supplying capacity of MPR. Through this connection, an MPR can selectively focus its capacity on 1 or 2 paths to supply extra power for variation compensation.
- 2) Direct connection with multi-path block. A multi-path block is a group of logic gates which are shared by many (e.g., more than or equal to 3)

critical paths. One MPR is directly connected with a multi path block without using switches. Please note that the direct connection does not necessarily mean that the multi-path block is powered by the MPR since it can be turned off by the EN signal. This connection is to widely provide all the shared paths with additional power for compensation.

Every block consists of the same number of gates, regardless of whether it is a single-path block or a multi-path block. In our case of experiment, around 50 gates are allocated in one block.

In terms of powering each one path, a single-path block can supply one path with more power than a multi-path block can provide, since every block supplies the same number of gates. In other words, a single-path block is to focus power from MPR on a few paths when variation is sparse, while a multi-path block is to share the power with many paths, therefore to handle widespread variation.

For example, in Figure II.10, the three circles of block B, C, and D on right correspond to single-path blocks and they are connected with one voltage regulator MPR2 through sleep transistors. MPR2 can drive at most one block at a time. The tall circle of block A on the left indicates a multi-path block and is connected with MPR1 directly. As mentioned above, every block is composed of the same number of gates. These two kinds of connections to MPR, together with global ASV of $V_{DD,L}$, handle variations in a progressive enhancement manner:

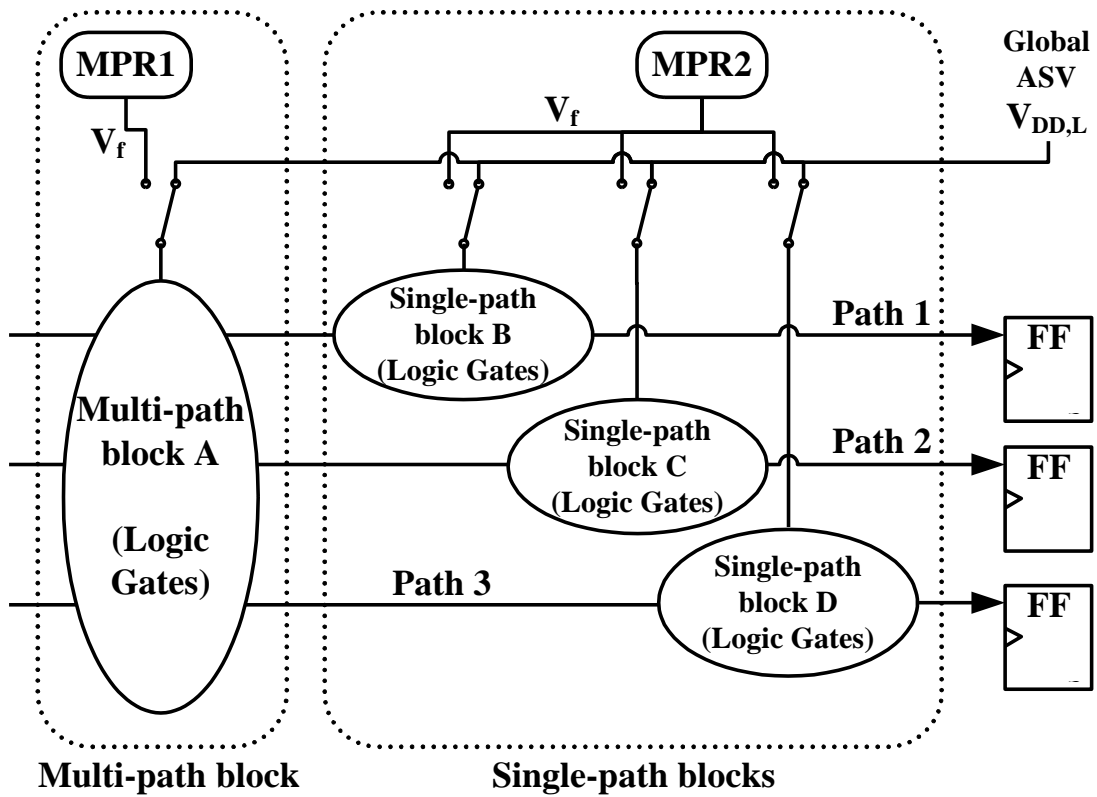


Figure II.10. Illustration for progressive voltage enhancement. The circles are blocks of logic gates and FF is a flip-flop. Every block consists of the same number of gates. MPR is the proposed mini programmable linear voltage regulator. All blocks are powered by $V_{DD,L}$ from global ASV system by default.

- 1) When variation is scattered so that the paths with large delay degradations do not share MPR through single-path blocks, their corresponding single-path blocks are switched from $V_{DD,L}$ to voltage tapping output V_f (MPR2).
- 2) If degradation spreads out, and the number of single-path blocks requiring V_f exceeds the capacity of MPR2, the multi-path block involving these paths is switched from $V_{DD,L}$ to V_f of MPR1. This is to assist single-path

blocks with a multi-path block so as not to overuse the MPR2 shared by single-path blocks.

- 3) When variation extremely intensifies, and 2) is eventually insufficient to handle the number of critical paths with strong delay variations, $V_{DD,L}$ is raised according to conventional ASV method. MPRs are now turned off again.

For the example in Figure II.10, all logic circuits are powered by $V_{DD,L}$ by default. If only the path 1 has strong variation, its corresponding single-path block B is switched to MPR2, while the other blocks C and D are still powered by $V_{DD,L}$. If the path 2 as well as the path 1 has large variations simultaneously, the multi-path block A is switched to MPR1 in addition. If all three paths 1, 2, and 3 have strong variations, then $V_{DD,L}$ is raised by global ASV to compensate for the widespread variations, and all the MPRs are turned off. This approach provides a large flexibility to accommodate different delay variation scenarios with the regulator MPR of limited capacity (and therefore with small overhead).

The delay variation prediction at each path is implemented by a sensor circuit proposed in [24]. Although it is originally designed for predicting circuit aging, it actually works for general delay variations including process variations. Its main idea is to add a transition edge detector to a flip-flop. A data switching in a small time window before the setup time constraint implies that the delay variation is large and close to the point of setup time violation. The sensor circuit [24] can detect such switching and generate a warning signal. Alternatively, the sensor circuit of [24] can be replaced by a

double-sampling flip-flop [31]. The sensors from feedback loops to a control block which tunes V_f and $V_{DD,L}$ in accordance with our proposed progressive voltage enhancement scheme described in this section. The block counts the number of warning signals flagged, and determine whether to activate single-path block, multi-path block, or disable them and raise $V_{DD,L}$.

The proposed dual-ASV system exploits voltage island design which is a very common practice and has been well studied in the context of a standard VLSI-CAD design flow [32, 33]. With partitioning and cuts on sets of logic gates in accordance with their delays, and assigning the regulator MPR as one type of power supply pads in placement [34], our proposed dual-ASV system will be well incorporated with standard VLSI processes, i.e., auto placement and route.

4. Experimental Results

We tested the proposed dual-ASV system by SPICE simulations on benchmark circuits and compared with conventional methods. The benchmark circuits are S526, S1423, and S5378 from ISCAS 89 suite in 45nm technology. The device models of 45nm technology are from PTM [35]. The characteristics of the three circuits are shown in Table II.5. The 4th column of Table II.5 tells the number of gates which have dual power supply lines in the dual-ASV system. The 5th column indicates the number of flip-flops monitored by the variation sensor circuit [24]. The rightmost column lists the number of MPRs employed.

Table II.5. Circuit characteristics and experimental setup for dual-ASV system.

Circuit	#gates	#FF	#gates D-ASV	#FF monitored	#MPR
S526	193	21	52	12	4
S1423	657	74	144	23	3
S5378	2779	179	600	24	12

Two kinds of variations are considered in the experiment: manufacturing process variations and NBTI-induced pMOS performance degradation. All these variations are assumed to follow Gaussian distribution. For the process variations, we focus on gate length variation and threshold voltage variation. The standard deviations of gate length variation and threshold voltage variation are 5% and 6.7% of their nominal values, respectively. For an aged circuit, additional threshold voltage degradation on pMOS transistors is considered. The mean and standard deviation of the degradation are 10% and 3.3% of the nominal values, respectively. The process variations and threshold voltage degradations are also applied to all the components of the dual-ASV system including MPR during simulation. Due to overly long simulation time, we ran 10 random Monte-Carlo iterations considering process variation (gate length variation: Gaussian $\mu=45\text{nm}$, $3\sigma=15\%$ of mean; V_{th} variation: Gaussian $\mu_{\text{NMOS}}=0.3423\text{V}$ $\mu_{\text{PMOS}}=-0.23122\text{V}$, $3\sigma=20\%$ of mean) and aging effects (additional V_{th} variation: Gaussian $\mu=10\%$ of μ_{NMOS} or μ_{PMOS} , $3\sigma=10\%$ of mean), instead of full-fledged Monte Carlo test on each circuit.

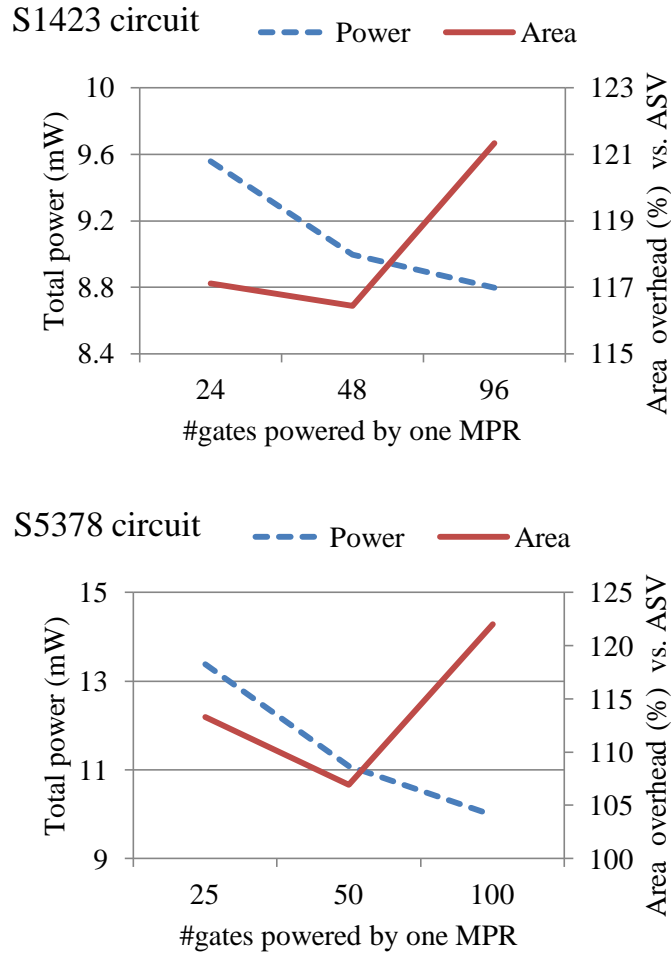


Figure II.11. Effects from the granularity of MPRs on circuit resources. All cases are adjusted to yield an equal circuit delay.

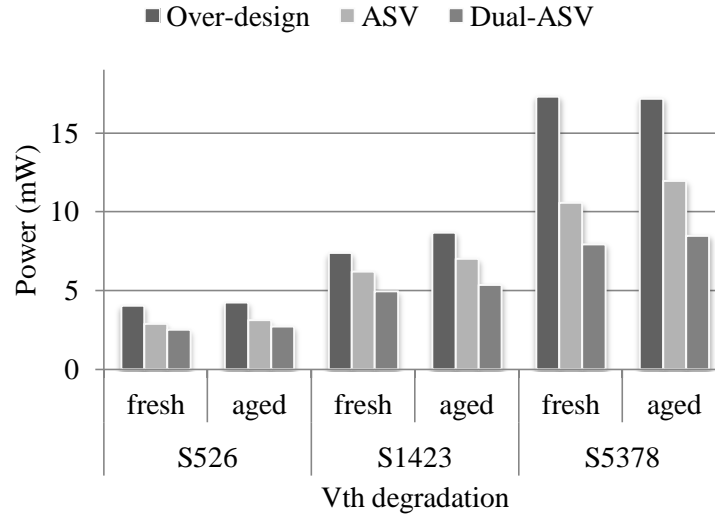
An important parameter to determine the number of MPRs deployed in circuit is the number of gates supplied by one MPR. A finer granularity of dual-ASV system requires more MPRs each of which supplies fewer gates. Due to intrinsic overheads from each MPR, deploying more MPRs in a circuit consumes larger die area as well as power. On the contrary, the more gates one MPR supplies, the larger area the MPR

Table II.6. Voltage configurations and average results on power dissipation. All cases have the same worst case delay.

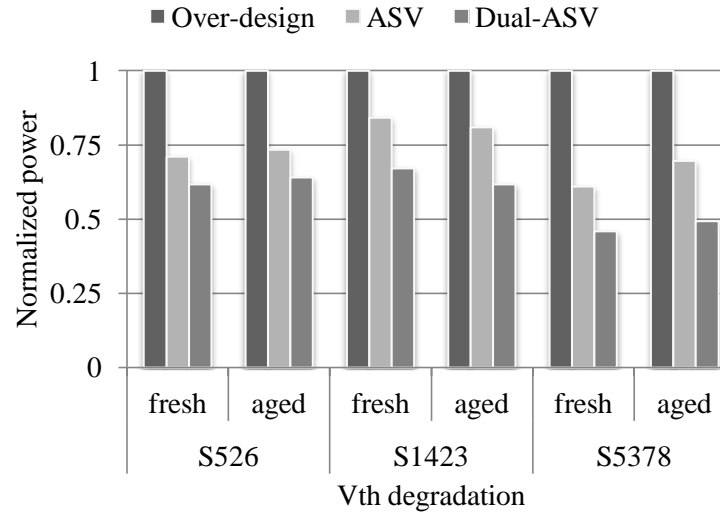
		Over-design		ASV		Dual-ASV					
		V _{DD} (V)	Power (mW)	V _{DD} (V)	Power (mW)	V _{DD,L} (V)	V _f (V)	V _f - V _{DD,L}	Power (mW)	Reduction vs. over-design	Reduction vs. ASV
S526	Fresh	1.00	4.05	0.78 - 0.98	2.88	0.70 - 0.90	0.80 - 1.00	100mV	2.50	38.3%	12.8%
	Aged	1.00	4.25	0.81 - 1.00	3.12	0.725 - 0.925	0.825 - 1.025	100mV	2.72	36.0%	13.0%
S1423	Fresh	0.99	7.37	0.83 - 0.93	6.20	0.725 - 0.825	0.875 - 0.975	150mV	4.94	33.0%	20.4%
	Aged	0.99	8.68	0.86 - 0.99	7.02	0.75 - 0.85	0.90 - 1.00	150mV	5.35	38.4%	23.8%
S5378	Fresh	1.02	17.30	0.80 - 0.95	10.56	0.85 - 0.95	0.7 - 0.80	150mV	7.94	54.1%	24.2%
	Aged	1.02	17.18	0.83 - 1.02	11.96	0.85 - 1.00	0.7 - 0.85	150mV	8.46	50.8%	29.0%

occupies since its pass transistor size is proportional to load current demand from the gates. The curves in Figure II.11 show how the granularity of MPRs affects major circuit resources: power and die area. The dashed lines represent power overhead increase according to finer granularity of MPR deployment where each MPR supplies fewer gates. The solid lines show that supplying more gates incurs increase in die area overhead from MPRs due to larger pass transistors. The solid lines for area overhead also tell that area overhead would not ever decrease even though each MPR supplies fewer gates. Not only pass transistor in each MPR, but also its amplifier circuit contributes area overhead which does not ever shrink while current demand decreases. Area overhead is, therefore, dependent on the number of MPRs deployed in circuit as well as the number of gates each MPR supplies.

The main experimental results are obtained from SPICE simulations and summarized in Table II.6. For each circuit, we report results on both fresh circuits where only process variations are considered, and aged circuits where NBTI induced threshold voltage degradation is considered additionally. The power dissipations in Table II.6 are the average results over the 10 random instances of each design. The 4th column of Table II.6 includes the power dissipation from over-design method. The power estimation here is also from SPICE simulations and includes both dynamic and leakage power. In over-design method, all instances of a design are applied with the same supply voltage for both fresh and aged cases. The V_{DD} level is 1.0V for S526, 0.99V for S1423, and 1.02V for S5378. This level is chosen such that the worst case delay satisfies timing constraint. The 5th and 6th columns are the results of conventional ASV. There is a single supply



(a) Total circuit power consumptions are compared.



(b) Total circuit powers are normalized to over-design circuit. Significant power reductions are achieved by dual-ASV system throughout all the cases examined.

Figure II.12. Power consumption comparison for the dual-ASV system.

voltage in this system and the V_{DD} level can be finely tuned with the step size of 0.01V. The ASV tuning is performed such that the power is minimized while the delay target same as that of over-design is reached. The V_{DD} tuning ranges are in the 5th column and the power dissipations are in the 6th column. One can see that ASV uses less power than over-design.

The results from the proposed dual-ASV system are in the rightmost 6 columns of Table II.6. The 7th column shows the tuning range of $V_{DD,L}$, which is the global (coarse-grained) supply voltage. The 8th column indicates the range of V_f , which is the local (fine-grained) supply voltage. The 9th column shows the difference of voltages between the dual supply lines V_f and $V_{DD,L}$ which should be less than 150mV in order not to use level shifter. The power dissipations including the power overhead of all the components of dual-ASV system are listed on the 10th column. Here, the dual-ASV also achieves the same delay as that of over-design and conventional ASV cases. The power reductions from the dual-ASV are summarized in the right-most two columns. On the average, the dual-ASV system can reduce power consumption by 42% and 21% compared to over-design and conventional ASV, respectively.

The bars of Figure II.12 show more clear comparison in power consumptions between three systems. The leftmost bar is for the power consumed in over-design system, the middle is for conventional ASV, and the rightmost bar is for the dual-ASV system. The proposed dual-ASV system shows the best power efficiency among all the systems.

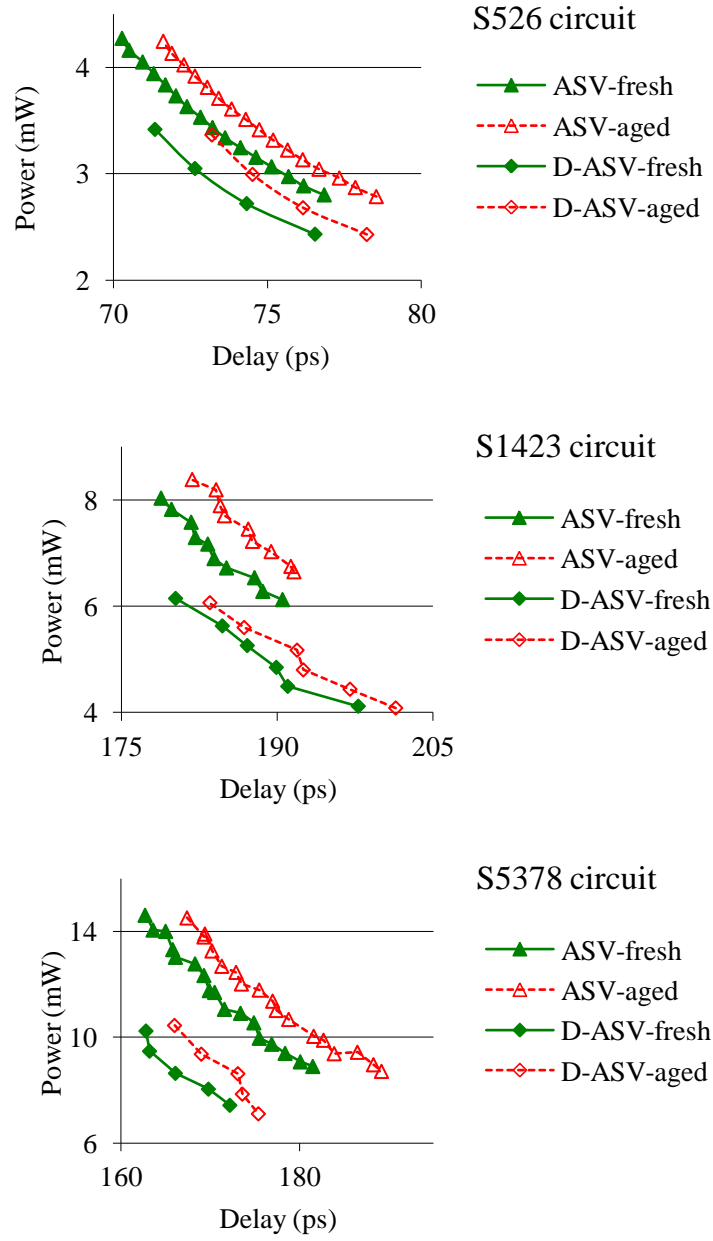


Figure II.13. Power-delay curves for the dual-ASV system.

We further compared the power-delay tradeoff curves of the conventional ASV and the dual-ASV in Figure II.13. The solid curves are for the fresh circuits while the dashed curves are for the aged circuits. The triangles indicate results from the conventional ASV and the diamonds represent results from the dual-ASV system. One can see that solutions from the dual-ASV are superior to those from conventional ASV in terms of the entire power-delay tradeoff.

The dual-ASV system causes area overhead due to the regulator MPR including decap and sleep transistors, and the other control circuits. For S526, S1423, and S5378 circuits, the area overheads are about 14.6%, 16.4%, and 6.9% of the conventional circuits, respectively. Since the dual-ASV implementation here is manually designed without optimizations, we believe the area overhead can be reduced if the design is performed in an optimized manner. Delay sensors [24] used in our benchmark circuits incur additional 10~20% of area overhead. The sensors, however, are simply adopted in our system only for delay sensing functionality, and can be easily replaced with other types of techniques to reduce overhead, e.g., SCAN chain which is a standard component of modern architecture so that we can even ignore area overheads from delay variation sensing.

5. Conclusions

In order to handle different scale of variations in an efficient manner, we propose a dual-level adaptive supply voltage (dual-ASV) system. This system allows fine-grained adaptive supply voltage with simple regulator designs and low power routing

overhead. This system includes a progressive voltage enhancement scheme, which has large flexibility for accommodating different scenarios of variations. The effectiveness of this system is validated by SPICE simulations on benchmark circuits.

In the future, we will continue the study of the dual-ASV system to further reduce area overhead and remove the dependence on voltage-island based designs. Future work will also include extended experiments to cover more complex configurations of dynamic voltage and frequency scaling (DVFS) over multiple voltage islands, stable reference voltage generation with band-gap reference for enhancing reliability of MPR design, and further discussions on design automation approaches to incorporate our proposed dual-ASV system with standard digital design process.

III. BOOSTABLE REPEATER DESIGN FOR VLSI INTERCONNECT

1. Boostable Repeater Design

A. Main Idea and Overview

I attempted to achieve fine-grained circuit adaptation by making repeaters adaptable. Adaptability means that a repeater can be controlled to operate in either high performance or low power mode. The proposed boostable repeater design is to use a transiently high voltage rail to assist the V_{DD} in high performance mode and simply turn off the high voltage in low power mode. In other words, the low power mode is the same as conventional static CMOS circuits. Since we do not want to incur stand-alone voltage regulator or additional power grid, the transiently high voltage must be somehow obtained from the existing V_{DD} . Perhaps the only well-known solution for voltage up-conversion is charge pump. The main drawback of charge pump is that it usually requires a large area of capacitor. Since the high voltage is needed transiently, the demand on the charge pump capacitor size is reduced. Since the high voltage here plays an auxiliary role on pulling-up signals, instead of being the sole supply voltage, the demand for the pump capacitor size is further reduced.

Using transiently high voltage has another advantage. In fine-grained voltage adaptation, logic gates with different supply voltages are mingled together [10]. Specifically, a gate with low V_{DD} may drive another gate with high V_{DD} . Then, the PMOS transistor of the high V_{DD} gate cannot be completely turned off and leakage

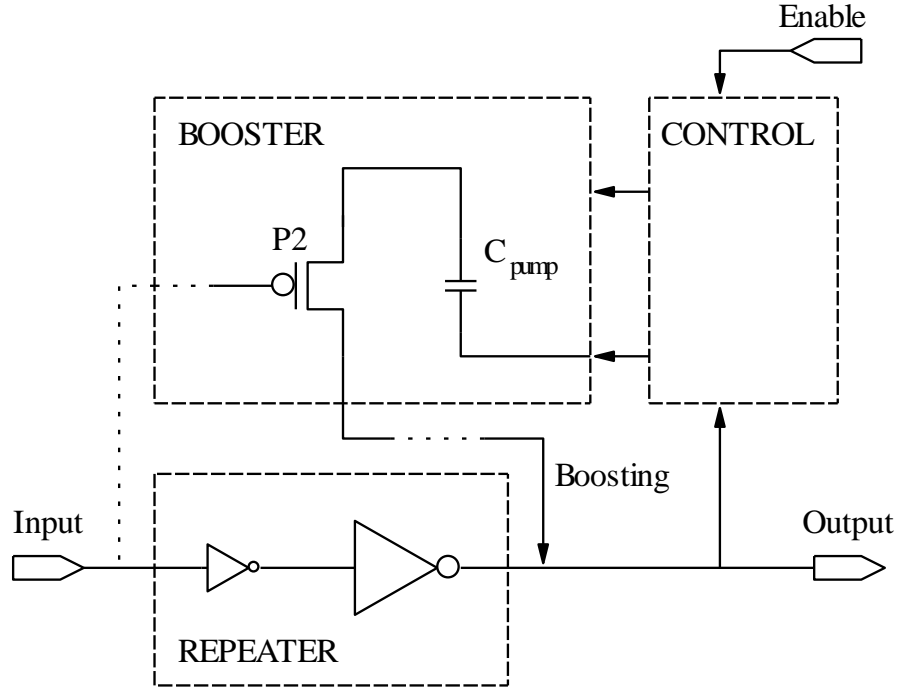


Figure III.1. Overview of the proposed boostable repeater.

current consequently increases. Since our high voltage is enforced transiently, such leakage increase does not exist and no level shifter is needed in our case.

An overview of our boostable repeater design is depicted in Figure III.1. It is composed by three parts: a conventional repeater, booster and control circuit. The core part is the booster, which is a capacitive charge pump. When the repeater is in steady state, the pump is charged. When the repeater has a rising switching, the pump discharges and provides a transiently voltage that is higher than V_{DD} . This high voltage is applied together with the V_{DD} to the repeater and therefore boosts the switching speed. Although capacitive charge pump is also employed in voltage regulator designs, our design has a key difference from that of stand-alone regulator. A stand-alone regulator needs to provide a steady voltage output and its design is much more complex and its

on-chip implementation causes huge overhead. In contrast, our charge pump only needs to provide a transiently high voltage. Therefore, the design can be greatly simplified and the overhead can be largely reduced. The control circuit activates or disables this boosting function according to an “enable” signal, which is usually obtained from variation detection circuit. The implementation of variation detection can be any among [2, 24, 22, 23].

The booster adds extra load capacitance to the input and output node. Since boostable repeaters are usually inserted at timing-critical paths, we do not want the low power mode to be slower than conventional repeaters. As such, it is important to restrict the size of the extra load. Here, the boosting is applied to reduce only the rising delay. We focus on rising edge of signal since NBTI (Negative Bias Temperature Instability) effect [1] on pMOS device is severe than HCI (Hot Carrier Injection) [1] on nMOS device. However, the falling delay can also be reduced in the same manner with additional design and area cost only from another booster for falling edge. In our design, since the load presented to timing paths is limited, the rising delay reduction overshadows the small falling delay increase.

In circuit design, one can judiciously select repeaters on timing critical paths and replace them with boostable repeaters. Since only a small portion of repeaters are replaced, the overall area overhead to an entire chip is very limited. The area overheads induced by the boostable repeaters will be further discussed in section III.2.F.

B. Design and Operations

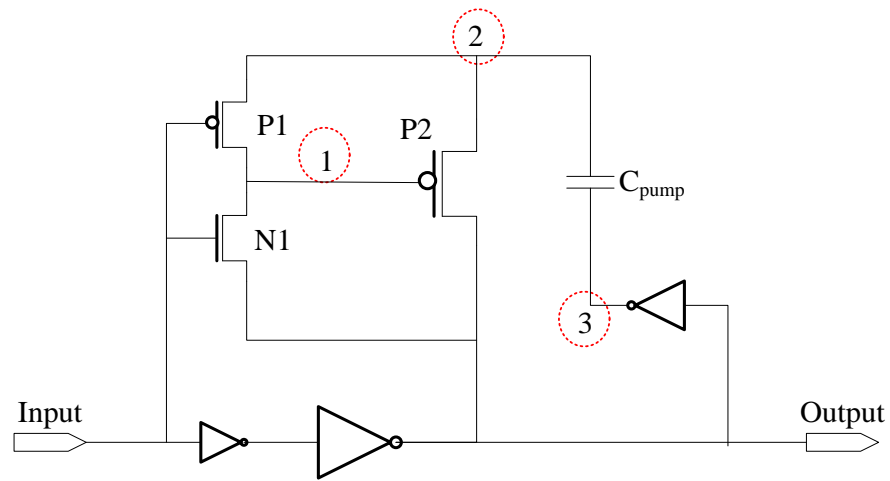
This section will first introduce the design and operations of the boostable repeater when the boosting feature is always on. Then, it will show the design on how to turn it off.

1) Boosting On

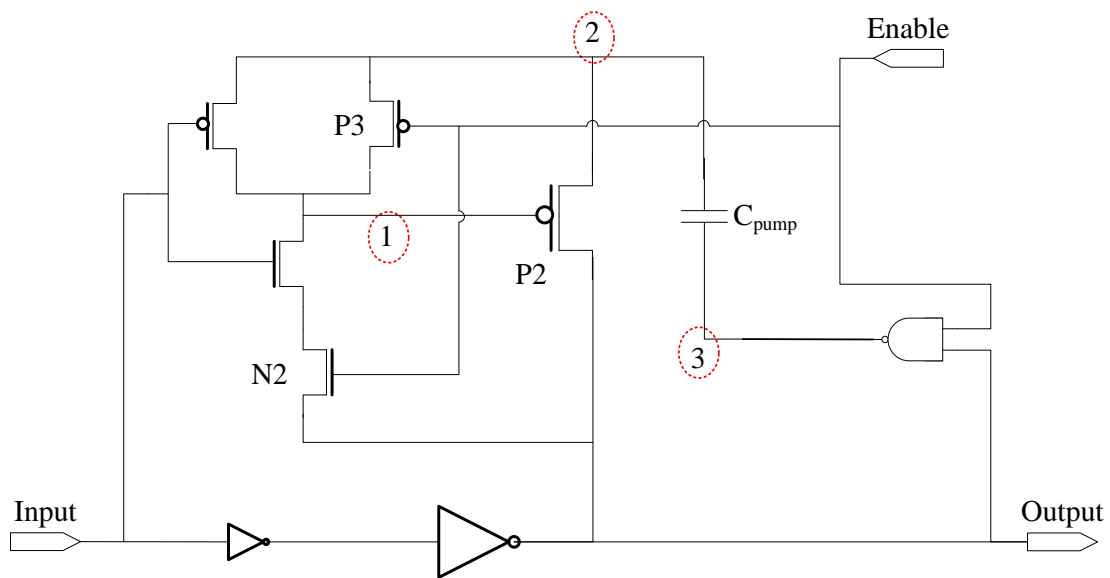
If the boosting feature is always on and cannot be turned off, i.e., not programmable, the design is a simplified version shown in Figure III.2(a). Transistor P2 is the pass transistor that delivers current from C_{pump} to the output node. Transistor P1, N1 and the inverter between the output and node 2 coordinate the operations. The operations mainly include two phases: charging and boosting.

a. Charging Phase

Charging to the capacitor C_{pump} takes place when both the input and the output are stabilized to high, or V_{DD} . Figure III.3(a) shows the waveforms in this phase. When the input and the output are high, P1 is off, N1 is on and V3 (voltage at node 3) is low. Since N1 is on, V1 (voltage at node 1) is at $V_{\text{DD}} - V_{\text{th},N1}$, where $V_{\text{th},N1}$ is the threshold voltage of N1. Then, the pass transistor P2 is partially on and V_{DD} at the output charges node 2 through P2. In other words, C_{pump} is charged. In Figure III.3(a), the voltage across C_{pump} is indicated by the difference between V2 and V3, which increases during the charging.

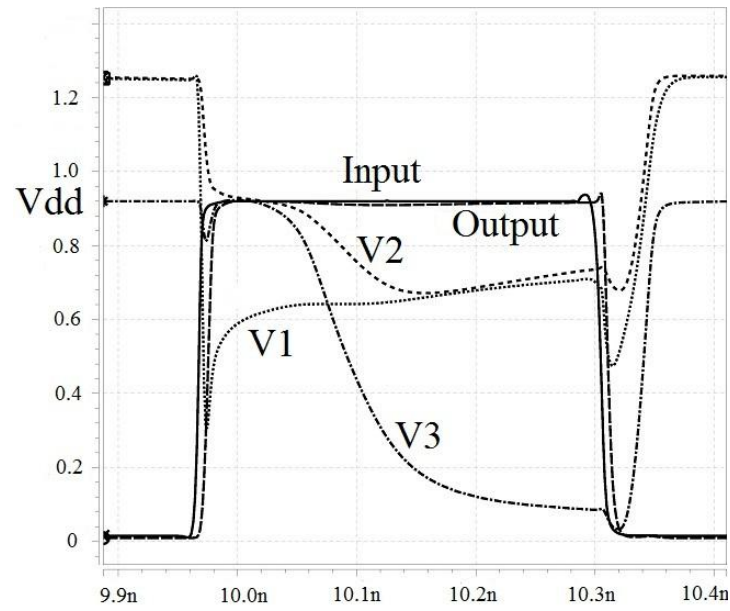


(a) Simplified schematic

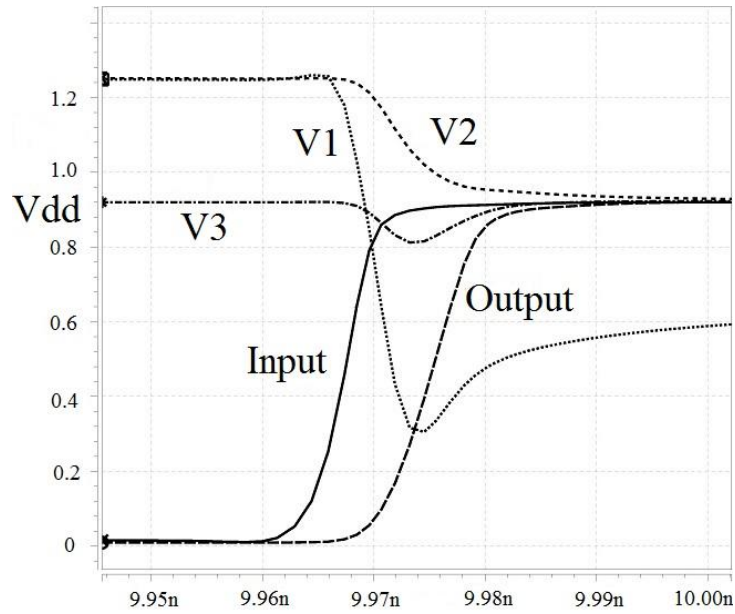


(b) Complete schematic

Figure III.2. Schematics of boostable repeater design.



(a) Charging phase



(b) Boosting phase

Figure III.3. Charging and boosting operations of boostable repeater. Figure (b) is an expended view of a region in (a) where input switches from low to high.

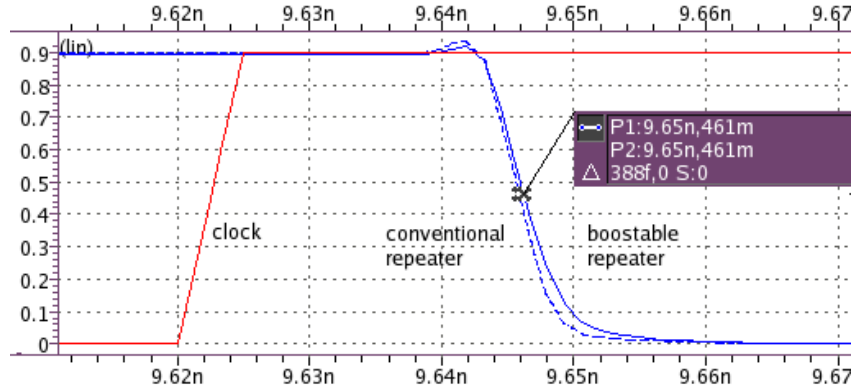
Voltage V_2 at C_{pump} is raised higher than V_{DD} when the input and the output switched to low. When the output is low, V_3 goes to high due to the inverter. While the voltage difference across C_{pump} remains, V_2 is raised beyond V_{DD} . In Figure III.3(a), V_2 is raised to about 1.25V which is higher than 0.9V of V_{DD} . When the input is low, N_1 is turned off and P_1 is turned on. Then, node 1 is connected with node 2 through P_1 , and V_1 also reaches the same level as V_2 . Consequently, P_2 is turned off and node is not connected with the output node any more. In Figure III.3(a), one can clearly see that both V_1 and V_2 reach 1.25V at time of 10.4ns.

b. Boosting Phase

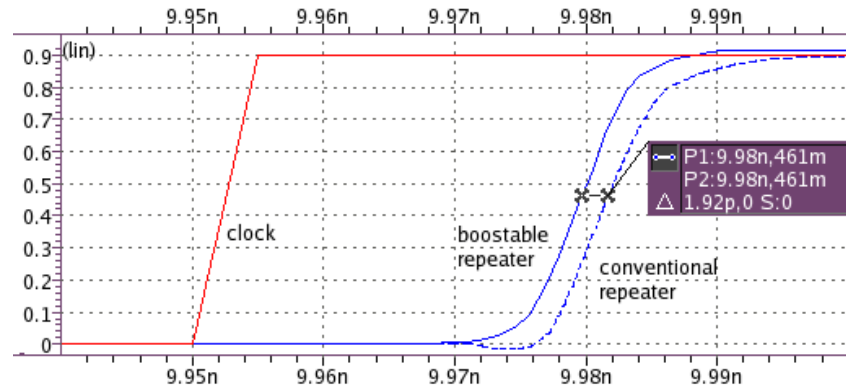
The boosting occurs when there is a rising switching at the input. The corresponding waveforms are shown in Figure III.3(b). Due to the gate delays between the input and the output, there is a short time period when the output is still low even the input goes high. During this period, N_1 is turned on and pulls V_1 toward low. The low voltage at node 1 turns on pass transistor P_2 and C_{pump} starts to discharge and pull up the output voltage, i.e., the boosting starts. At the same time, the input rising is propagated to the output and the output is pulled up by V_{DD} as well. Evidently, the boosting accelerates the rising transition of the output and therefore improves switching speed. The main area overhead is due to pass transistor P_2 and capacitor C_{pump} . C_{pump} can be implemented using trench cap [36], which is very area-efficient. The sizes of P_1 , N_1 and the inverter between the output and C_{pump} are small. Thus, the extra capacitive load presented to the timing path is limited. Further discussions on the area overheads from pass transistor P_2 and C_{pump} will be presented in section III.2.F.

In view of output load, boosting operation naturally demands extra strength from repeater output which may, if not properly addressed, hamper timing advantages from the proposed boostable repeater. According to Figure III.2(a), at the rising edge of the repeater output, the inverter driving C_{pump} contributes output load of repeater, whereas charges to the output are supplied by P2 and N1 (gate charge from node 1) as well. The inverter supports V3 (voltage at node 3) while C_{pump} is providing extra charges to the output for boosting. The inverter, therefore, needs to be sized wide sufficient to supply the current drain through C_{pump} , which in turn entails output load increase. When sizing the inverter, however, its switching speed is not a concern since the switching is only for raising V2 (voltage at node 2) up in charging phase which is not a timing critical operation. Accordingly, to minimize the extra output load from the inverter, we compose the inverter with three inverter gates where the 1st inverter driven by the repeater output can be size as small as possible, and the 3rd inverter driving C_{pump} is large enough to support V3.

While rising transition gains speed-up from boosting, falling transition from the repeater output in Figure III.2(a) can be negatively affected by the extra components of the proposed boostable repeater design attached to its output. Figure III.4 provides falling transition as well as rising transition waveforms from the boostable repeater and a conventional repeater. Figure III.4(b) shows that 1.92ps of rising edge speed-up is achieved through boosting operation. According to Figure III.4(a), however, the boostable repeater slows down falling transition of its output by 0.39ps compared to a conventional repeater output which is reasonably small timing penalty with respect to



(a) Falling edge waves. Switching from boostable repeater is 0.39ps slower than switching from a conventional repeater.



(b) Rising edge waves. Switching from boostable repeater is 1.92ps faster than switching from a conventional repeater.

Figure III.4. Transition waveforms from the proposed boostable repeater compared to a conventional repeater.

the timing gain of 1.92ps in rising transition. The rising timing benefit (1.92ps) from boosting operation which is 5 times larger than falling timing loss (0.39ps) can sufficiently diminish extra resources needed for addressing the falling transition timing penalty.

In order to further clearly compare accumulated timing benefits in rising transition as well as timing losses in falling transition due to the boostable repeaters, we constructed chains of 9 repeaters composed of either the proposed boostable repeaters only or conventional repeaters only. In rising transition through 9 repeaters, the boostable repeater boosts signal edge by 28.12ps, while it lags falling transition 5.65ps slower than a chain of 9 conventional repeaters. One can easily see that the accumulated timing gain in rising edge is about 5 times larger than timing loss in falling edge, just like the timing benefit from individual boostable repeater as discussed in the above paragraph. In the chain of 9 buffers, each one boostable repeater consumes 3.69 μ W more power compared to one conventional repeater. Simulation results conducted on logic circuits given in section 3.5, power reductions achieved through the proposed boostable repeaters are around 2~5mW which tells that the 3.69 μ W of power overhead from one boostable repeater is well overshadowed by significant global power savings.

2) Boosting Off

In Figure III.2(a), the boosting function cannot be turned off, i.e., not programmable. In order to allow it to be turned off, additional transistors must be added. The complete design with programmability is shown in Figure III.2(b). When the “Enable” signal is high, the circuit operates in the same way as that in Figure III.2(a). When “Enable” is low, the NAND2 gate outputs constant high. At the same time, transistor N2 is off and P3 is on. Therefore, pass transistor P2 is turned off. The booster part is disconnected from the output node.

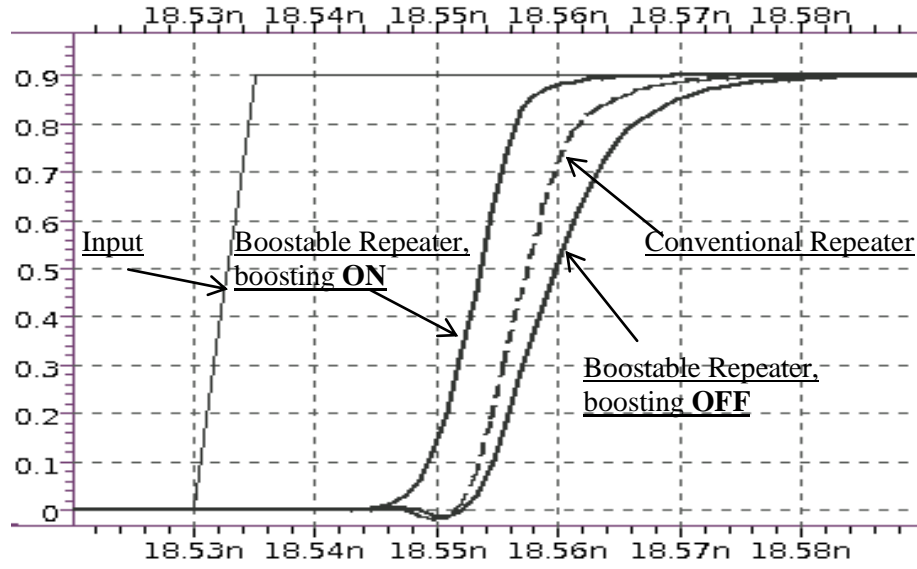


Figure III.5. Waveforms from boostable repeater when it is turned on and off.

Figure III.5 displays the waveforms of rising transitions at the output when the boosting is on and off. Additionally, the waveform from a conventional repeater is included. The conventional repeater is a couple of cascaded inverters, which are identical to the two inverters in Figure III.3(b). One can see that turning the boosting on can significantly reduce the rising delay. When the boosting is off, the rising delay has a small degradation compared to conventional repeater.

2. Experimental Validation

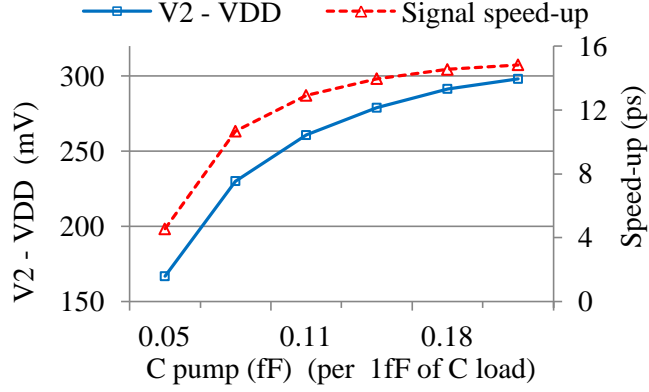
A. Experiment Setup

The boostable repeater design was validated by various SPICE simulation-based experiments. In these experiments, Predictive Technology Model [35] of 45nm was

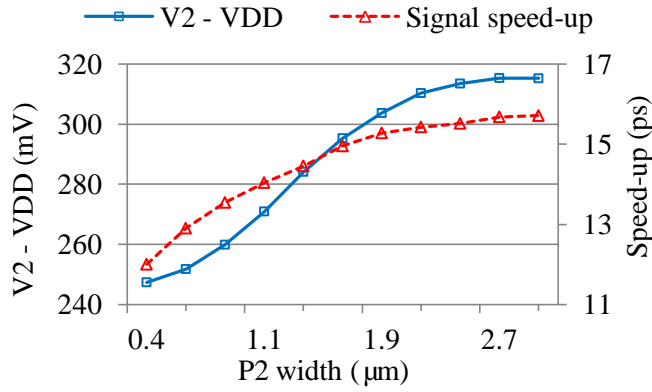
employed for the transistor models in the SPICE simulations. When variation resilience was tested, process variations and NBTI-induced pMOS aging are considered and both of them are assumed to follow Gaussian distribution. For the process variations, we focus on gate length variation and threshold voltage variation. The standard deviations of gate length variation and threshold voltage variation are 5% and 6.7% of their nominal values, respectively (gate length variation: Gaussian $\mu=45\text{nm}$, $3\sigma=15\%$ of mean; V_{th} variation: Gaussian $\mu_{\text{NMOS}}=0.3423\text{V}$ $\mu_{\text{PMOS}}=-0.23122\text{V}$, $3\sigma=20\%$ of mean). For an aged circuit, additional threshold voltage degradation on pMOS transistors is considered and added on top of process variation. The mean and standard deviation of the degradation are 10% and 3.3% of the nominal values (additional V_{th} variation: Gaussian $\mu=10\%$ of μ_{PMOS} , $3\sigma=10\%$ of mean). We added the additional V_{th} degradation due to circuit aging to the V_{th} of each device which is already determined for process variation of the device. The process variations and threshold voltage degradations are also applied to all the components of boostable repeaters during simulation.

B. Impacts of Design Parameters

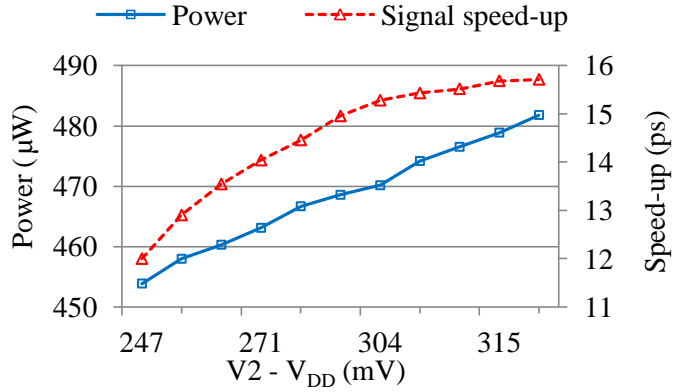
We investigated the impacts of two major design parameters, the size of pass transistor P2 and the capacitance C_{pump} . The simulation results are presented in Figure III.6. One can see that when either of P2 width or C_{pump} increases, the signal speed-up boosted through boostable repeater (the dotted line) also increases. This speed-up is achieved with the increase of voltage V2 (voltage at node 2) change (the solid line). Figure III.6(c) indicates that the V2 change and the speed-up gain naturally induce more



(a) V2 boost and signal speed-up w.r.t. C_{pump} size



(b) V2 boost and signal speed-up w.r.t. P2 width ($C_{\text{load}} = 70\text{fF}$)



(c) Signal speed-up and power consumption increase w.r.t. V2 boost

Figure III.6. Increased voltage at node 2 beyond V_{DD} ($V2 - V_{DD}$) and signal speed-up acquired through boosting w.r.t. major design parameters.

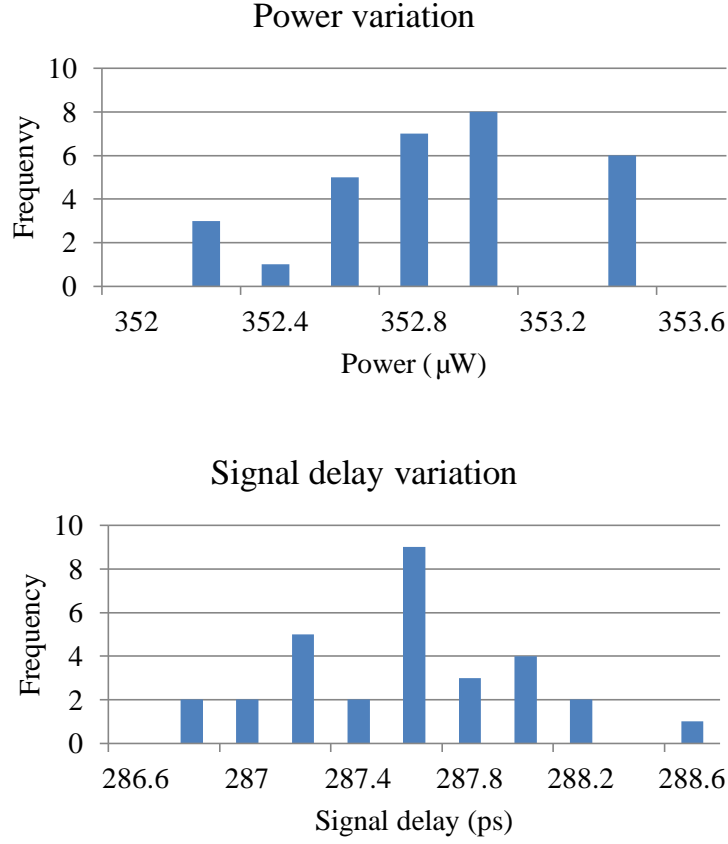


Figure III.7. Power and signal delay variations due to parasitic capacitance. Source and drain parasitic capacitance variations (Gaussian $\mu=100\text{pF}$, $3\sigma=15\%$ of mean) of P2 transistor are considered.

power consumption. These curves provide guidelines to choose the sizes of P2 and C_{pump} .

Typically, we use $1.7\mu\text{m}$ width for P2 and 10fF for C_{pump} .

Variations in the parasitic capacitances of devices in boostable repeater design may affect its boosting operations thereby cause variations in performance enhancement from signal boosting. We investigated effects from variations of source and drain parasitic capacitances of the pass transistor P2 which is the biggest and critical device in

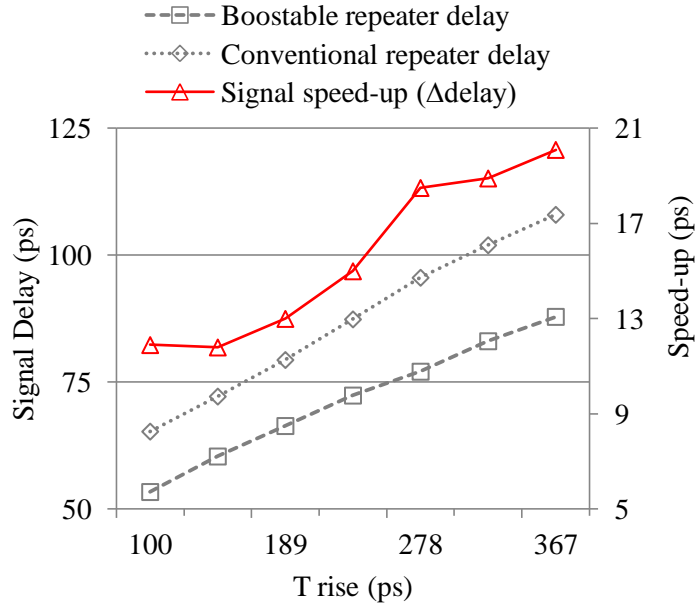


Figure III.8. Signal speed-up w.r.t. the rising time (T_{rise}) of input signal. The dotted line and dashed line show repeaters slow down as T_{rise} increases.

boostable repeater design. The variations of parasitic capacitances are assumed to follow Gaussian distribution ($\mu=100\text{pF}$, $3\sigma=15\%$ of mean). Boostable repeaters with parasitic capacitance variations are deployed in a interconnect network circuit, and Figure III.7 presents corresponding power and critical path delay variations from 30 Monte-Carlo runs. Mean of power consumption is $352.8\mu\text{W}$, and 3σ of power variation is $0.98\mu\text{W}$ which is merely 0.28% of its mean. Distributions of signal delay variation also show very small deviations ($3\sigma=1.28\text{ps}$) from its mean (287.5ps) which is only 0.45% of the mean. This implies that effects from the parasitic capacitance variations on the overall performance of the proposed boostable repeater design are limited.

The rising time of input signal also determines signal delay through a repeater. The dotted line and dashed line in Figure III.8 show how the signal rising time affects the signal delays from both repeater designs. One can see that the speed-up acquired through boostable repeater also rises as the delays increase. This speed-up gain in accordance with the extended input rising time occurs since the more charges in C_{pump} can be supplied to output before repeater in Figure 1 begins to pull the output up. Along with the speed-up, compared to circuit design with conventional repeaters, overall power overhead from the proposed boostable repeaters is maintained around 4% increase which can be overshadowed by 13~20ps of speed-up. This implies that boostable repeater design is more advantageous under situations with slow rising time of input, e.g., interconnects with larger loads.

C. Simulations on Global Interconnect

We tested the boostable repeater design in the context of global interconnect, which usually has long wires and a significant number of repeaters. The experiments were conducted on 4 global nets. Each net is tested for two different delay specifications and with/without aging. Hence, totally 16 cases are used. Table III.1(a) presents circuit configuration of each net informing the number of total repeaters in each net and the number of the proposed boostable repeaters replacing conventional repeaters.

Our approach is compared with over-design and ASV (Adaptive Supply Voltage) [4] on variation/aging resilience. We conducted 15 runs of random test for each method on each case. Each run emulates a post-silicon instance. We compared the proposed

Table III.1. Benchmark circuit configurations with deployment of the proposed boostable repeaters.

(a) Global interconnects

# sinks	# total repeaters	# boostable repeaters
10	34	7
13	44	10
16	73	15
19	85	18

(b) ISCAS '89 sequential benchmark circuits

Circuit	# logic gates	#FF	# total repeaters	# boostable repeaters
S526	141	21	68	9
S1423	490	74	168	23
S5378	1004	179	1779	101

boostable repeater design with over-design and conventional adaptive supply voltage (ASV). Conventional repeaters were employed in both the over-design and ASV. The over-design applied the same V_{DD} level in all runs according to the worst case variations. For ASV, each run found its own V_{DD} level that minimized the power and satisfied the delay specification. In our approach, a part of the repeaters were replaced by the boostable repeaters. The boosting function was turned on/off depending on the variations of each run. The boostable repeater design is applied jointly with ASV, since widespread variation can be covered by ASV while fine-grained variations on critical paths are handled with the proposed boostable repeaters.

Table III.2. Voltage configurations and average results on power dissipation from simulations on interconnect networks.

# sinks	# total repeaters	Maximum Delay (ps)	Circuit Aging	Over-design		ASV		Boostable Repeater + ASV				
				V _{DD}	Power (μ W)	V _{DD}	Power (μ W)	# boostable repeaters	V _{DD}	Power (μ W)	Power Reduction (%) vs. Over-design vs. ASV	
10	34	300	fresh	0.9	616	0.73-0.89	358	7	0.68-0.77	244	60.2	29.4
			aged		526	0.74-0.9	346		0.7-0.78	242	53.9	27.9
		333	fresh	0.75	279	0.63-0.73	179		0.61-0.67	144	48.1	18.0
			aged		248	0.66-0.75	181		0.63-0.69	146	40.9	18.0
13	44	320	fresh	0.9	652	0.74-0.88	387	10	0.69-0.78	298	54.5	21.3
			aged		587	0.76-0.9	392		0.71-0.78	291	50.5	23.0
		350	fresh	0.77	337	0.64-0.74	211		0.62-0.68	188	44.4	10.3
			aged		302	0.68-0.77	222		0.64-0.7	191	36.9	12.6
16	73	400	fresh	0.92	836	0.72-0.89	417	15	0.64-0.73	288	65.6	29.3
			aged		760	0.75-0.92	443		0.66-0.75	300	60.5	30.5
		435	fresh	0.81	498	0.66-0.78	285		0.6-0.68	225	54.8	19.5
			aged		453	0.69-0.81	300		0.62-0.7	232	48.7	21.1
19	85	400	fresh	0.98	1147	0.75-0.95	555	18	0.67-0.77	371	67.7	31.0
			aged		1045	0.79-0.98	591		0.68-0.79	390	62.7	32.0
		455	fresh	0.8	509	0.65-0.77	299		0.6-0.68	250	50.9	15.4
			aged		465	0.68-0.8	311		0.61-0.7	258	44.4	15.9

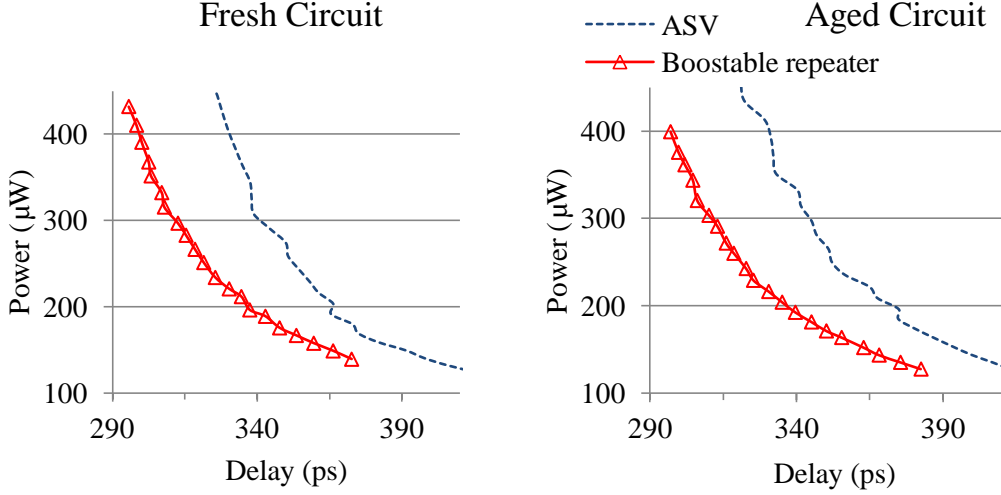


Figure III.9. Power-delay curves for interconnect networks. Boostable repeater design is compared with adaptive supply voltage (ASV) when circuit is fresh or aged.

The results are listed in Table III.2. For all methods, we compared power dissipation for the same maximum sink delay. The average power dissipations are the average results over the 15 random instances of each design. The power estimation here is from SPICE simulations and includes both dynamic and leakage power. On the average, the boostable repeater design combined with ASV can reduce power by 52.8% and 22.2% compared to over-design and conventional ASV, respectively. We further compared the power-delay tradeoff curves of our approach with the conventional ASV. The curves are depicted in Figure III.9. One can see that solutions of our approach are superior to those from conventional ASV in terms of the entire power-delay tradeoff.

D. Comparison with Online Adjustable Buffer

We also compared the proposed boostable repeater design with online adjustable buffer (OAB) [17]. The results are depicted in Figure III.10. Simulations were performed on the global net with 10 sinks. The OAB is implemented to have the same area overhead as the boostable repeater design. In Figure III.10, the power-delay curves from different repeaters are presented. The dashed lines represent results from OAB, and the solid lines with x markers indicate results from the boostable repeater. The performance of conventional repeater design is given through the dotted lines. In all cases when the repeaters and buffers are ON or OFF, the proposed boostable repeater design exhibits better power efficiency than OAB. Boosting off for OAB means its tri-state buffer is turned off. One can see that the performance-power of boostable repeater in off mode is

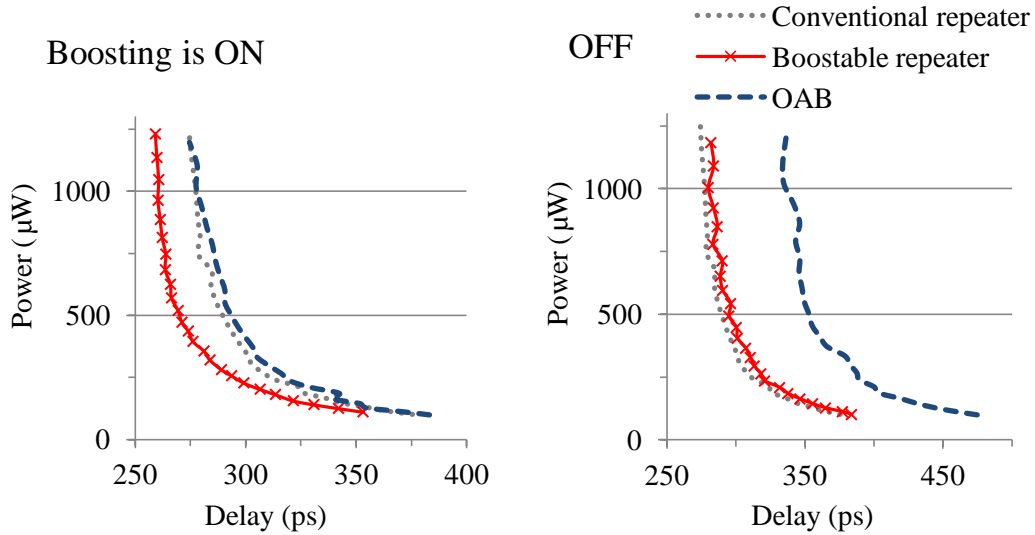


Figure III.10. Power-delay curves to compare the proposed boostable repeater with online adjustable buffer (OAB).

very close to the conventional repeater. This means its extra load to timing path is significantly less than that of OAB.

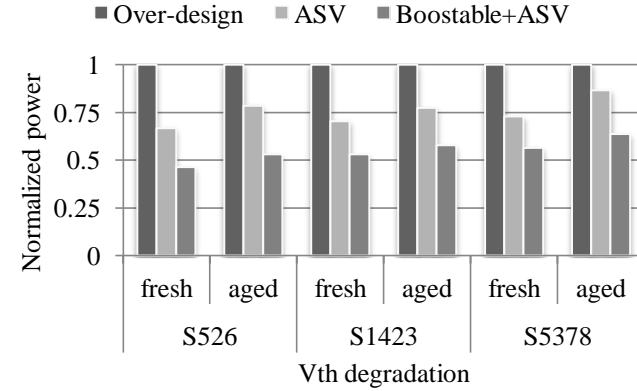
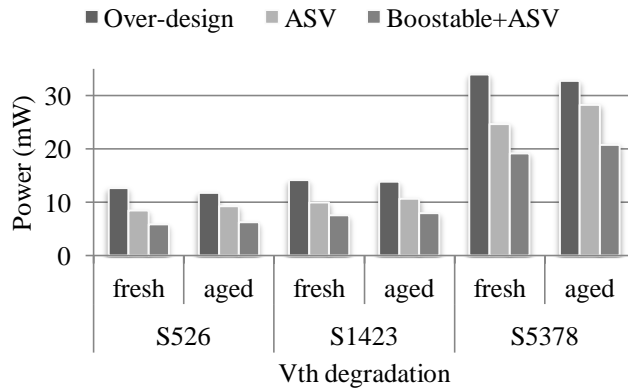
E. Simulations on Logic Circuit

The boostable repeater design is also tested in logic circuit. We employed ISCAS'89 benchmark circuits S526, S1423, and S5378 as test beds. Similar to the experiment of global interconnect, which is described in Section III.2.3, we compared methods of over-design and ASV only system against the proposed approach which integrates boostable repeater with ASV system. Table III.1(b) describes the number of repeaters and boostable repeaters deployed in each circuit. The results on the testbeds are summarized in Table III.3. The 6th and 8th columns show power dissipations from over-designed and conventional ASV system respectively. The power dissipation from our approach is presented in the 11th column. The power reductions achieved with the boostable repeater are listed on the 12th and 13th columns of Table III.3, and the average of reductions are 45.9% and 25.8% compared to over-design and conventional ASV, respectively.

The bars of Figure III.11 show more clear comparison in power consumptions between three systems. The leftmost bar is for the power consumed in over-design system, the middle is for conventional ASV, and the rightmost bar is for our approach with boostable repeaters. Our proposed system shows the best power efficiency among all the systems. Figure III.12 depicts the power-delay curves comparing ASV and our

Table III.3. Voltage configurations and average results on power dissipation from simulations on ISCAS'89 benchmark circuits.

	# total repeaters	Max. Delay (ps)	Circuit Aging	Over-design		ASV		Boostable Repeater + ASV				
				V _{DD}	Power (mW)	V _{DD}	Power (mW)	# boostable repeaters	V _{DD}	Power (mW)	Power Reduction (%) vs. Over-design	Area over-head
S526	68	165	fresh	0.93	12.70	0.74-0.89	8.48	9	0.7-0.78	5.88	53.6 %	4.4%
			aged		11.80	0.78-0.93	9.27		0.73-0.8	6.25	47.2 %	
		176	fresh	0.86	9.70	0.7-0.82	6.33		0.66-0.73	4.70	51.6 %	
			aged		9.06	0.73-0.86	6.86		0.68-0.76	4.98	45.1 %	
S1423	168	89	fresh	1.06	14.19	0.84-1.04	10.00	23	0.79-0.89	7.52	47.0 %	5.3%
			aged		13.84	0.88-1.06	10.69		0.81-0.95	7.98	42.4 %	
S5378	1779	216	fresh	1.07	33.96	0.90-1.01	24.71	101	0.83-0.91	19.15	43.6 %	4.7%
			aged		32.74	0.94-1.07	28.29		0.86-0.94	20.80	36.5 %	



(a) Total circuit power consumptions are compared.

(b) Circuit powers are normalized to over-design circuit.

Figure III.11. Power consumption comparison for boostable repeater design. Significant power reductions are achieved through boostable repeater design throughout all the cases examined.

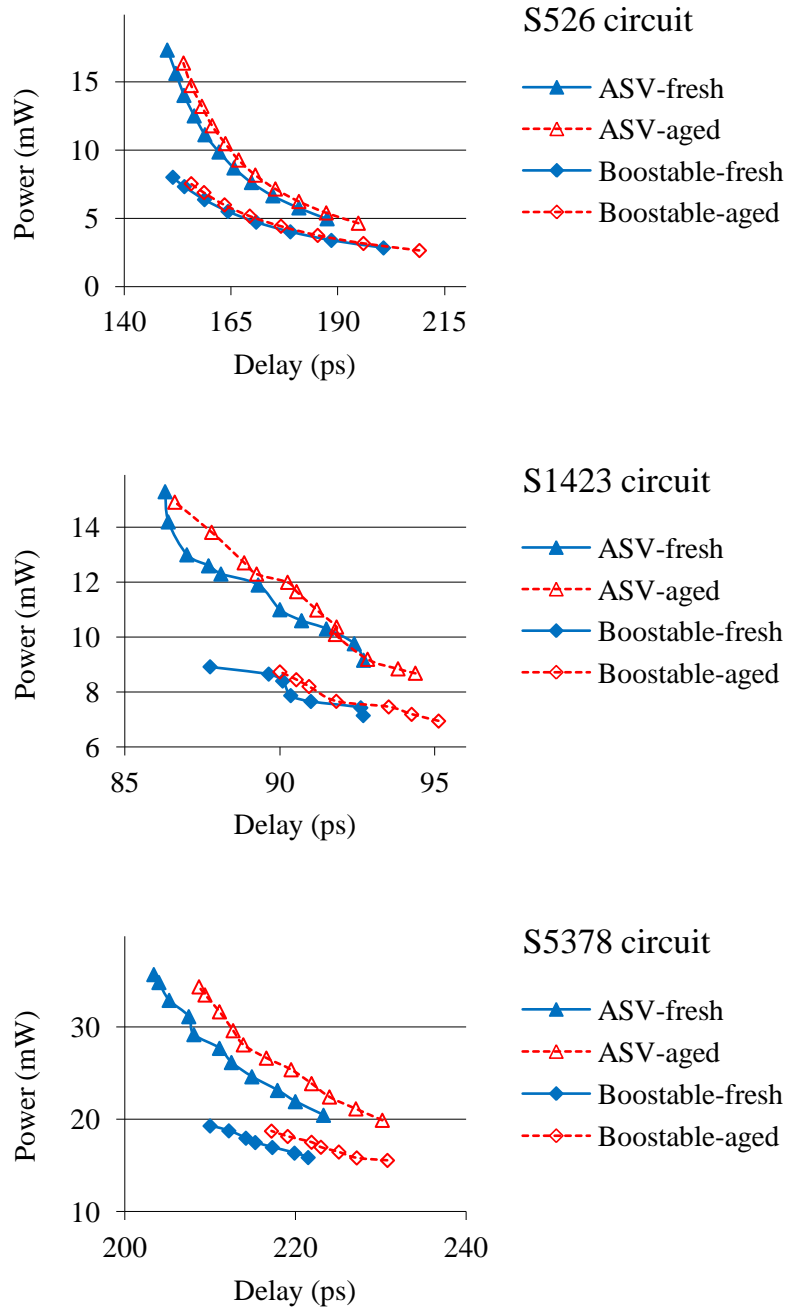


Figure III.12. Power-delay curves for boostable repeater design. Boostable repeater design is compared with adaptive supply voltage (ASV) when circuit is fresh or aged.

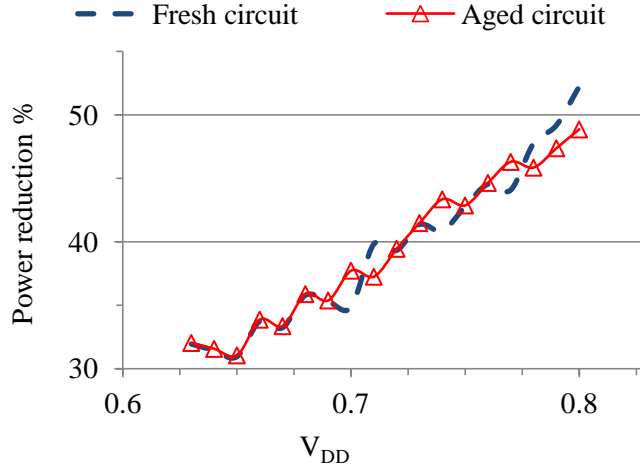
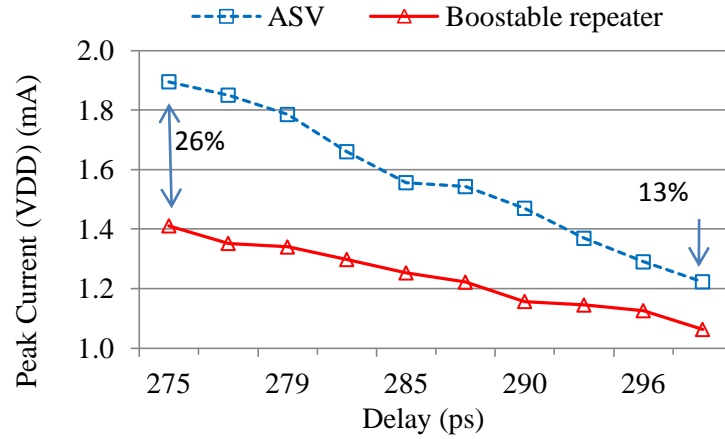


Figure III.13. Power reduction vs. V_{DD} levels from boostable repeater compared to adaptive supply voltage (ASV).

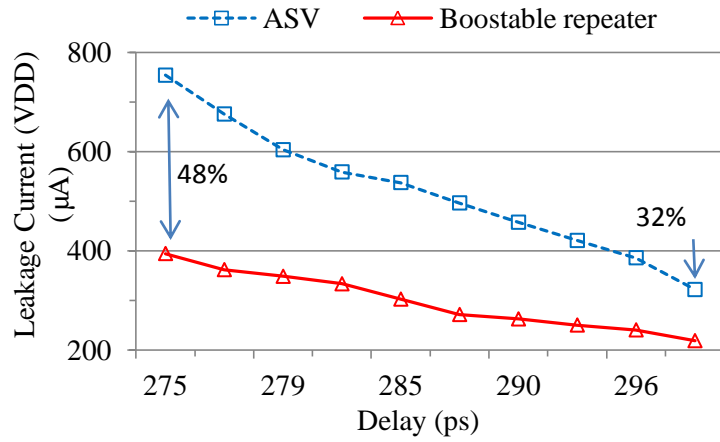
approach. One can see that our approach outperforms conventional ASV only system in all cases. Figure III.13 presents iso-delay power reductions of our approach versus ASV only system when V_{DD} varies. One can see that the power reduction is more when the circuit runs in high performance mode.

F. Peak Current and Leakage

We also examined the impacts of the boostable repeater design on peak current and leakage current. The peak current drawn from repeaters are critical since it exaggerates power supply noise. More decaps, hence larger area overhead, are required if the noise worsens. The peak currents demands from repeater designs are given in Figure III.14(a). The boostable repeater design shows up to 26% reduction in peak current flow compared to conventional repeaters in ASV system with similar signal speeds. Leakage current from repeaters is also one of the most significant issues due to



(a) Peak current demands from repeaters



(b) Leakage current through repeaters

Figure III.14. Peak current / leakage current vs. delay.

the vast number of repeaters and their huge current demands in modern circuits. According to Figure III.14(b), the leakage current also reduces from 32% up to 48% if boostable repeater design is applied. The reductions on both peak and leakage current are mostly achieved by lowering down the global V_{DD} levels while providing equal signal speed through boostable repeaters. Actually, each boostable repeater flows leakage current 45% ~ 80% more than a conventional repeater. However, the boostable

repeaters replace only a small portion (up to 22% in our cases) of all repeaters in a circuit so that their contribution on whole circuit current flow is limited. In our cases, all boostable repeaters in a circuit draw current from 2.8% up to 4.5% among whole leakage current flow throughout a circuit. In accordance with the curves in Figure III.14, the extra current demand from individual boostable repeater is well overshadowed by huge reduction on overall current flow due to the lowered V_{DD} level.

The highly raised voltage level V_2 at net 2 in Figure III.2(a) causes forward junction bias at the transistors attached to the net, since V_2 increases over V_{DD} level. Forward biased source-body junction induces leakage current from source to substrate. We observed that, when V_2 is higher than V_{DD} , the leakage current to substrate terminal from source connected to V_2 is around 1nA while leakage from the source to its drain is around 0.7 μ A. Since substrate leakage is merely 0.15% of drain leakage, the forward junction bias does not seriously affect and/or worsen current leakage issues caused by the proposed boostable repeater design.

When both input and output are low, the charges in C_{pump} leaks through P_2 transistor in Figure III.2(a), and this leakage finally causes depletion of C_{pump} . If C_{pump} is fully depleted, therefore the voltage level V_2 at net 2 in Figure III.2(a) is not higher than V_{DD} any longer, the boosting capability of the proposed boostable repeater disappears. In our experiments of interconnect networks and logic circuits presented in section III.2.C and III.2.E, one boostable repeater maintains its boosting capability up to around 100ns on average. This means, in giga-hertz system, tens of clock cycles with low input status are permitted for the proposed boostable repeater to wait for a rising edge to boost.

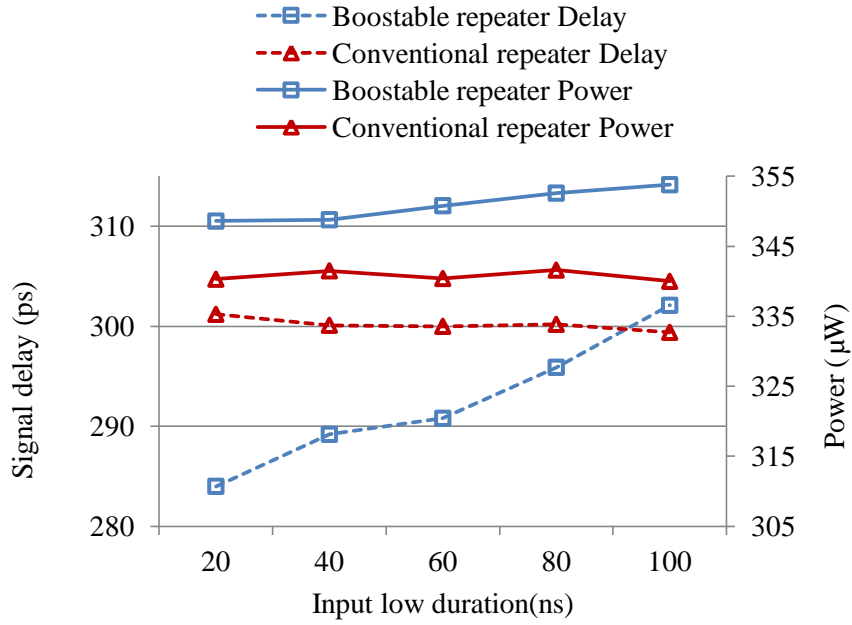
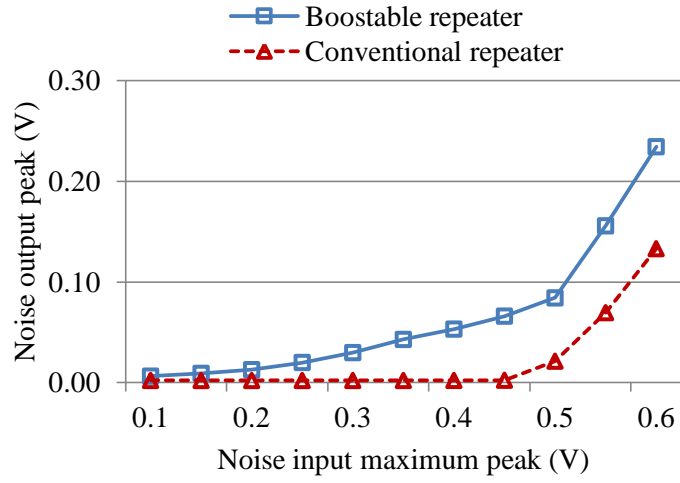
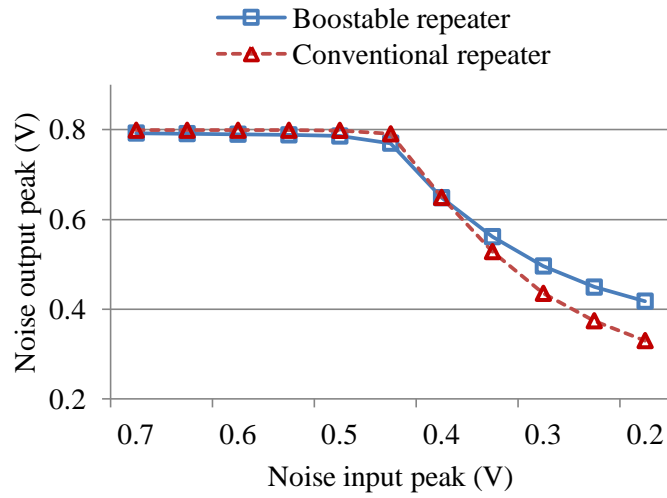


Figure III.15. Performance enhancement degradation due to leakage. Performance enhancement from the proposed boostable repeater degrades in accordance with longer duration of low input state due to charge leakage from C_{pump} .

Figure III.15 presents further analysis on how performance enhancement from the proposed boostable repeater degrades when input low status lengthens. If input stays low logic level longer than 90ns, then C_{pump} is depleted so that boostable repeater cannot boost its output. Due to the charge leakage from C_{pump} , average power consumption also slightly increases as low input status extends. This leakage phenomenon may limit applicability of the proposed boostable repeaters, while it achieves significant performance enhancement in critical paths with frequent switching activities, i.e., high activity factors. The higher activity factor in turn aggravates circuit aging effect so that the proposed boostable repeater can more effectively handles the variation.



(a) Propagated noise output peak for low to high rising glitch noise input seed when input is low



(b) Propagated noise output peak for high to low falling glitch noise input seed when input is high

Figure III.16. Noise characteristics of the proposed boostable repeater and conventional repeater.

G. Impacts on Noise Immunity

As discussed in section III.2.F, each individual boostable repeater causes larger leakage current than conventional repeater which may induce noise immunity degradation, even though the leakage is well compensated by significant block-wide benefits. Figure III.16 compares propagated noise peaks of the proposed boostable repeater against noise propagations of conventional repeater. A triangular noise seed determined by its peak voltage level with fixed 200ps duration is given into each repeater input, and the peak voltage of its corresponding propagated output noise is observed. Two types of noise seeds are simulated, one is for rising glitch noise when input is low, and the other is for falling glitch noise when input is high. According to Figure III.16(a), the proposed boostable repeater worsens noise propagation of rising glitch noise because of leakage current through P2 in Figure III.2(a). However, noise propagation of falling glitch noise is reduced from boostable repeater output as given in Figure III.16(b), due to the extra output load from boostable repeater design. In all cases, $V_{DD} = 0.8V$.

Although the proposed boostable repeater degrades noise immunity for rising glitch noise, the noise propagation can be well addressed on account of CMOS noise immunity. Figure III.17 reports noise margins of CMOS inverter implemented on 45nm Predictive Technology Model [35] utilized in our experiments. In accordance with the noise margins of CMOS inverter, the maximum allowed input low noise V_{IL} is 0.347V. As given in Figure III.16(a), even with noise seeds with significantly high peak values, the peaks of propagated noise from boostable repeater are well lower than V_{IL} .

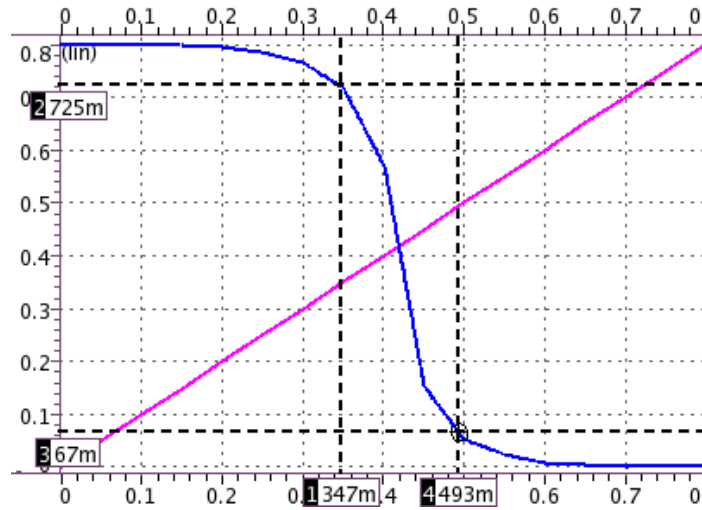
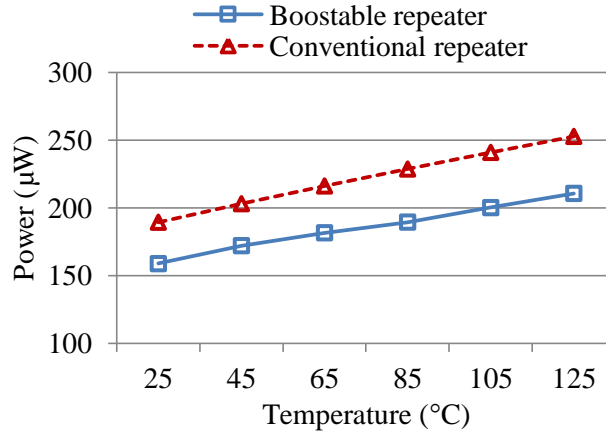


Figure III.17. CMOS inverter noise margins on 45nm PTM [35]. $V_{OH}=0.725V$, $V_{OL}=0.067V$, $V_{IL}=0.347V$, and $V_{IH}=0.493V$. $NML=V_{IL} - V_{OL}=0.28V$, $NMH=V_{OH} - V_{IH}=0.232V$

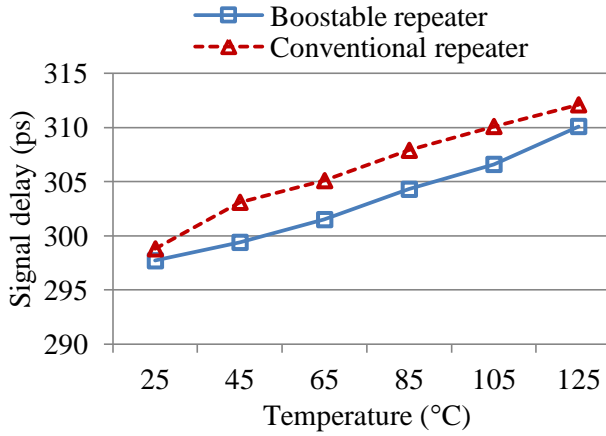
H. Temperature Variations

Circuit performance degrades with temperature increase. We analyzed temperature dependency of the proposed boostable repeater design in comparison with conventional repeaters. Although this work mainly focuses on threshold voltage variation and gate length variation, temperature dependency of boostable repeater design needs to be also addressed to propose a predictable device design approach. If device performance degrades irregularly and unpredictably with respect to temperature, the device may not be easily applicable to standard design processes.

Figure III.18 presents performance (power and signal delay) variations of circuit designs with boostable repeaters or conventional repeaters in accordance with temperature increase. V_{DD} of each type of circuit design is initially adjusted to meet the same timing constraints at 25 °C. Simulation linearly sweeps temperature from 25 °C to



(a) Power consumption variation w.r.t. temperature



(b) Signal delay variation w.r.t. temperature

Figure III.18. Circuit performance variations w.r.t. temperature increase. V_{DD} of circuits are initially configured to show the same signal delay at 25°C.

125 °C. According to Figure III.18 (a) and (b), both power and signal delay from the circuit design with the proposed boostable repeaters degrade linearly and predictably, very similar to the circuit design with conventional repeaters.

I. Area Overhead

Compared to conventional repeater design, the boostable repeater design has overhead on both device area and metal area. The device area overhead is due to the extra transistors employed in the design. The experimental results on the global interconnect show that the boostable repeaters cause about 20% additional device area in these global nets. Since boostable repeaters are needed only in timing critical nets, which is usually a small portion of entire netlist, the overall device area overhead for a chip is very limited. According to [20], typically 5-10% of all nets are timing critical in industrial designs. Hence, the chip-wise device area overhead is expected to be around 2%. When boostable repeaters are applied in logic circuits, S526, S1423, and S5378, the device area overheads are only 4.4%, 5.3%, and 4.7% respectively, according to Table III.3.

C_{pump} is one of the major components of boostable repeater design which dominates area overhead as well as determines performance enhancement. A larger capacitance allows faster rising edge, while it incurs increased die area overhead as well. Figure III.19 shows how C_{pump} sizing impacts circuit resources, power and area. The curves are all aligned to meet a certain delay constraints. The dotted curves with rectangles represent the whole circuit power consumption trends in accordance with the size of C_{pump} capacitance. Tendency of area overheads (%) compared to the circuits with conventional repeaters are shown with the solid curves with triangles. Through the solid curves, one can easily see that enlarging charge pump keeps linearly adding extra area

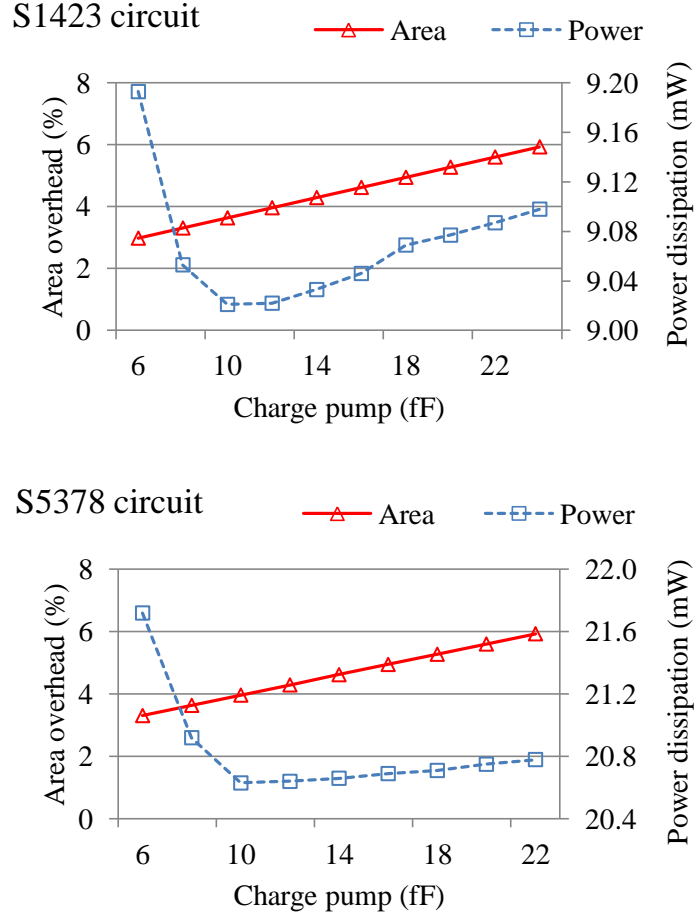


Figure III.19. Effects from the sizing of C_{pump} on circuit resources. All cases are adjusted to yield an equal circuit delay.

overhead which comes from resizing internal devices of boostable repeaters as well as resizing C_{pump} itself.

However, ever enlarging the capacitance does not mean that circuit performance keeps enhancing, according to the dotted curves in Figure III.19. One can see that the power efficiency touches its maximum at around 10fF of capacitance, and it begins to degrade. This is because of limitations such as saturated boosting of internal high

voltage rail (voltage at node 2 in Figure III.2(b)), high leakage from the boosted rail, etc. Furthermore, the larger a C_{pump} is sized, the bigger other internal devices of boostable repeater grow as well to support current through C_{pump} , and this induces extra load on the signal path.

In order to further minimize area overheads from C_{pump} in the boostable repeater design, capacitor C_{pump} is implemented by trench cap [36], which is very area-efficient and has low leakage. Trench cap has been employed as decap in industrial designs [36]. Capacitor C_{pump} of 10-20fF can be easily implemented by a single trench decap. Instead of the trench capacitor technique which is available in several processes [37], one can implement C_{pump} with metal-insulator-metal (MIM) capacitor [38] which can be integrated in standard CMOS technologies, and has high capacitance density as well. Since a chip design typically includes a tremendous number of decaps, the overall area overhead from C_{pump} is negligible.

3. Conclusions

In this work, we proposed a new technique of boostable repeater design, which can transiently boost its switching speed. This technique can be applied to achieve variation and aging resilience in a power efficient manner. Experimental results show that when the boostable repeater design is integrated with ASV, the same performance and robustness with over 25% less power than using ASV alone. Our approach also significantly outperforms the previous work on online adjustable buffer. In future

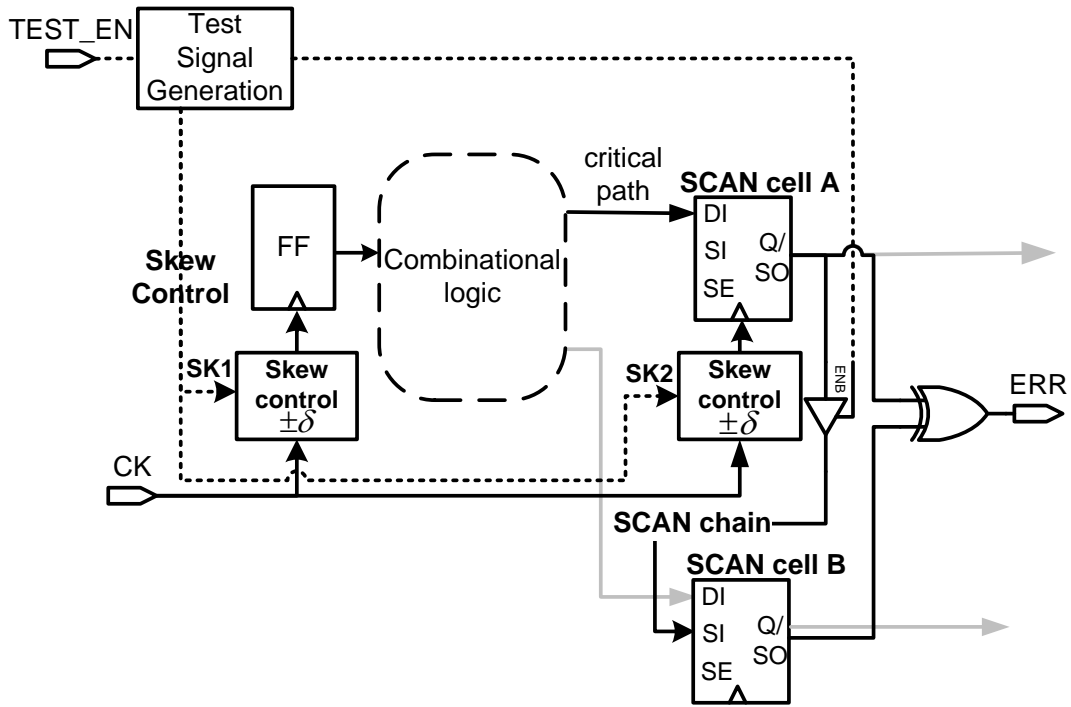
research, we will further investigate the application of boostable repeaters in dynamic power management.

IV. BUILT-IN DELAY TESTING SCHEME

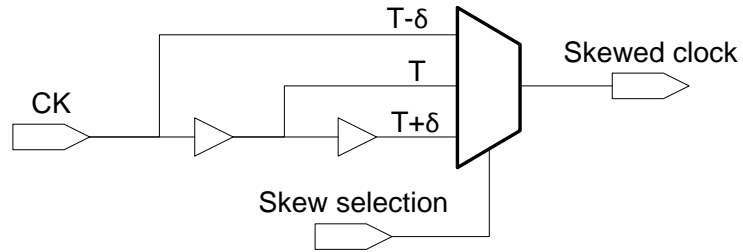
1. Adaptive System Design

This work presents an adaptive variation handling system composed of delay error detection and variation compensation. The proposed system is illustrated in Figure IV.1(a). A built-in delay testing scheme runs to monitor process variation or circuit aging induced delay faults. The proposed testing scheme uses dynamic clock skew control which manipulates timing constraints of a critical path so as to gauge the actual signal delay of the path. If a delay fault is detected, the dynamic clock skew control is now applied again to tune clock periods in order to mitigate the fault.

The proposed built-in delay testing scheme is initiated in a similar way that a BIST system checks circuit blocks. When circuit is fresh, and it shows variations due to process variations only, testing procedure runs once after chip fabrication. In the course of normal circuit operations, circuit aging proceeds slowly as transistors are switching for a long period of time. To monitor aging induced timing variations, testing procedure regularly runs according to scheduled intervals which are short enough to catch delay fault in advance, and long as well enough not to interrupt normal operation of the circuit. Since we exploits SCAN architecture as shown in Figure IV.1(a), the proposed delay testing scheme can share the testing vectors generated through ATPG (automatic test pattern generator) which perform other BIST procedures for fault detections at the same time.



(a) Proposed adaptive system with area-efficient built-in delay testing scheme



(b) Dynamic clock skew control block

Figure IV.1. Overview of the proposed built-in delay testing scheme with SCAN chain architecture. Dynamic clock skew control is applied to diagnose delay variation as well as to compensate the variation.

We propose a built-in delay testing scheme which is area-efficient and consumes low average testing power. Design and operation of the proposed built-in delay testing scheme is provided in section IV.1.A. In compensation procedure after variation is detected, the proposed system exploits dynamic clock skew control which is already embedded in the built-in delay testing system. In combination with dynamic skew control, boostable repeaters proposed in chapter III are limitedly deployed to focus on NBTI induced circuit aging effect. Section IV.1.B describes the proposed variation handling approach further in detail.

A. Area-Efficient Built-In Delay Testing Scheme

The overview of the proposed built-in delay testing scheme is given in Figure IV.1(a). The system can be divided into two major tasks, clock skew control and circuit response comparison. The skew control block in Figure IV.1(a) tunes clock period (T_{CLOCK}) to the combinational logic block under test in order to diagnose delay variation from the block. If a response from the block with $T_{CLOCK,SHORT}$ (reduced T_{CLOCK} through clock tuning) is different from $T_{CLOCK,LONG}$ (extended T_{CLOCK}), then we can presume that the circuit block cause delay fault with $T_{CLOCK,SHORT}$. Circuit responses from the block under test with variable T_{CLOCK} are checked through comparison system which is based on SCAN chain as shown in Figure IV.1.(a), and it flags delay fault signal if necessary.

Dynamic clock skew tuning allows flexible clock period (T_{CLOCK}) borrowing and lending between adjacent logic blocks connected through flip-flops. Figure IV.1.(b) presents a simple implementation of the skew control block whose MUX selects clock

arriving time to each flip-flop. If a block delay ($T_{BLOCK,FAST}$) is faster than T_{CLOCK} , the block can lend a portion of its clock period to adjacent blocks whose $T_{BLOCK,SLOW}$ is closer to T_{CLOCK} than $T_{BLOCK,FAST}$. This flexible clock tuning functionality is employed in the proposed built-in delay testing scheme in order to diagnose delay variation as well as to mitigate the variation. The skew control block given in Figure IV.1(b) tunes clock arrival time with δ resolution. Suppose T is the global clock period without dynamic clock skew control. The maximum available clock period to the circuit block under test ($T_{CLOCK,MAX}$) is $T + 2\delta$ through skew tuning, and the minimum ($T_{CLOCK,MIN}$) is $T - 2\delta$. The proposed built-in delay testing scheme assumes that a circuit response from $T_{CLOCK,MAX}$ is the correct response from the block. Then we iteratively subtract δ from $T_{CLOCK,MAX}$ through skew tuning (from $T_{CLOCK,MAX} - \delta$ down to $T_{CLOCK,MIN}$), and observe corresponding circuit responses to check if current response with reduced T_{CLOCK} differs with the correct response with $T_{CLOCK,MAX}$. If current response is different from the correct response, variation compensation system takes an action to mitigate the situation. If the circuit is fresh, and its variation is induced only by process variation, then the dynamic clock skew control has already determined a proper clock skew to handle the variation. For example, if $T_{CLOCK} = T_{CLOCK,MAX} - 3\delta$ yields a different response compared to the correct response with $T_{CLOCK,MAX}$, then $T_{CLOCK} = T_{CLOCK,MAX} - 2\delta$ is the shortest tuned clock period ($T_{CLOCK,TUNED}$) to the block for compensating process variation effect. Under the circuit aging effect, we begin the built-in delay testing procedure from $T_{CLOCK,TUNED}$, not from $T_{CLOCK,MAX} - \delta$. If, due to circuit aging effect, $T_{CLOCK,TUNED}$ yields

a response which differs with the correct response from $T_{CLOCK,MAX}$, a delay fault is flagged, and circuit aging compensation procedure operates.

The proposed built-in delay testing scheme exploits SCAN chain architecture to implement circuit response comparison part of testing procedure. SCAN chain is a well known and widely applied BIST DFT (Design for Testability) technique in VLSI circuit design. BIST system allows a fabricated chip by itself to diagnose any defects and degradations of its internal circuit blocks after production. SCAN chain in BIST is to

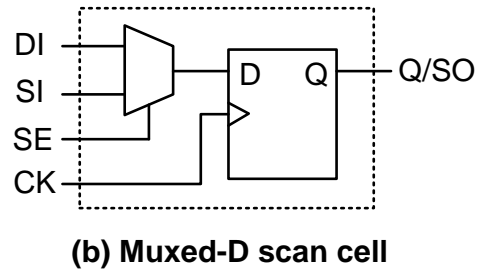
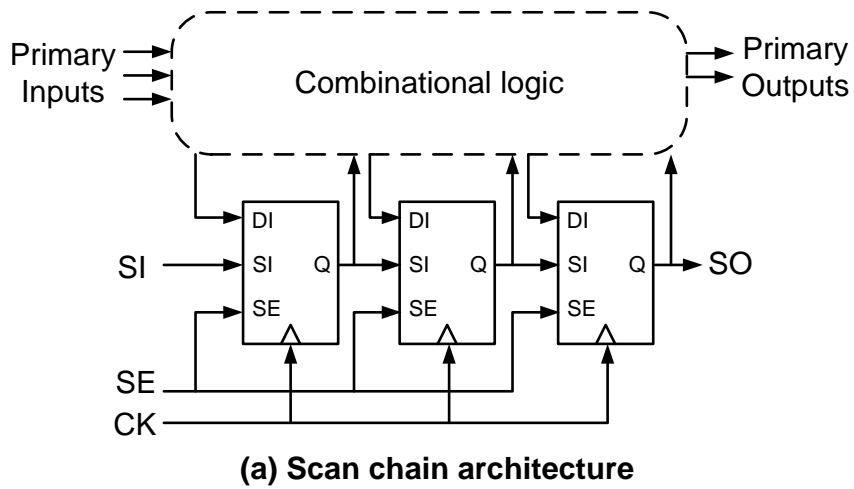


Figure IV.2. A common structure of built-in self test (BIST) SCAN chain architecture (a) with muxed-D SCAN cells (b). Figure (b) is the simplest SCAN cell design.

deliver testing vectors from pattern generator to logic block under test, to inject the vectors into the block, to capture response from the block, and finally to shift out the responses for diagnosis of the block's current status. Figure IV.2 depicts a basic structure of SCAN chain architecture with Muxed-D SCAN cell. Flip-flops in circuit design are selectively replaced with the SCAN cells to form multiple shift registers named SCAN chains as described in Figure IV.2(a). The Muxed-D SCAN cell in Figure IV.2(b) is the simplest design of SCAN cell which is composed of a regular D type flip-flop and a MUX. The MUX selects its input signal to the D flip-flop between data input (DI) from combinational logic circuit or scan input (SI) from SCAN chain in accordance with scan enable (SE) signal. The scan out (SO) of the SCAN cell drives scan input (SI) of its next SCAN cell, and this connection forms a shift register of SCAN chain architecture. BIST system with SCAN chain is a generally used component in modern VSLI design, and well supported by various CAD tools.

In the proposed built-in delay testing scheme, we implement response comparison part with the SCAN chain connection from SO to SI between adjacent SCAN cells. In the overview of the proposed system shown in Figure IV.1(a), the SCAN cell A is monitoring the critical path from the combinational logic block, and the cell A is connected to the next SCAN cell B through SCAN chain connection from SO to SI. In the proposed testing scheme, the cell B is working as a storage to save the correct response from the combinational logic block with $T_{CLOCK,MAX}$ as described in the clock skew tuning procedure above. The correct response captured by the cell A is transferred to the cell B through SCAN chain between them. The SCAN chain is separated by a tri-

state buffer to retain the correct response saved in the cell B through testing procedure. With the reduced clock period ($T_{CLOCK,MAX} - i\delta$, where $i = 1, 2, 3, 4$), the circuit block yields responses, and the XOR compares them against the correct response saved in the cell B.

Figure IV.3 depicts control signals and internal waves of the proposed built-in delay testing scheme described above. The waves show how our system adjusts clock skews until it detects delay fault. At first, the combinational circuit is tested with the maximum clock skews ($T_{CLOCK,MAX}$) which is known to be fault-free, and the correct

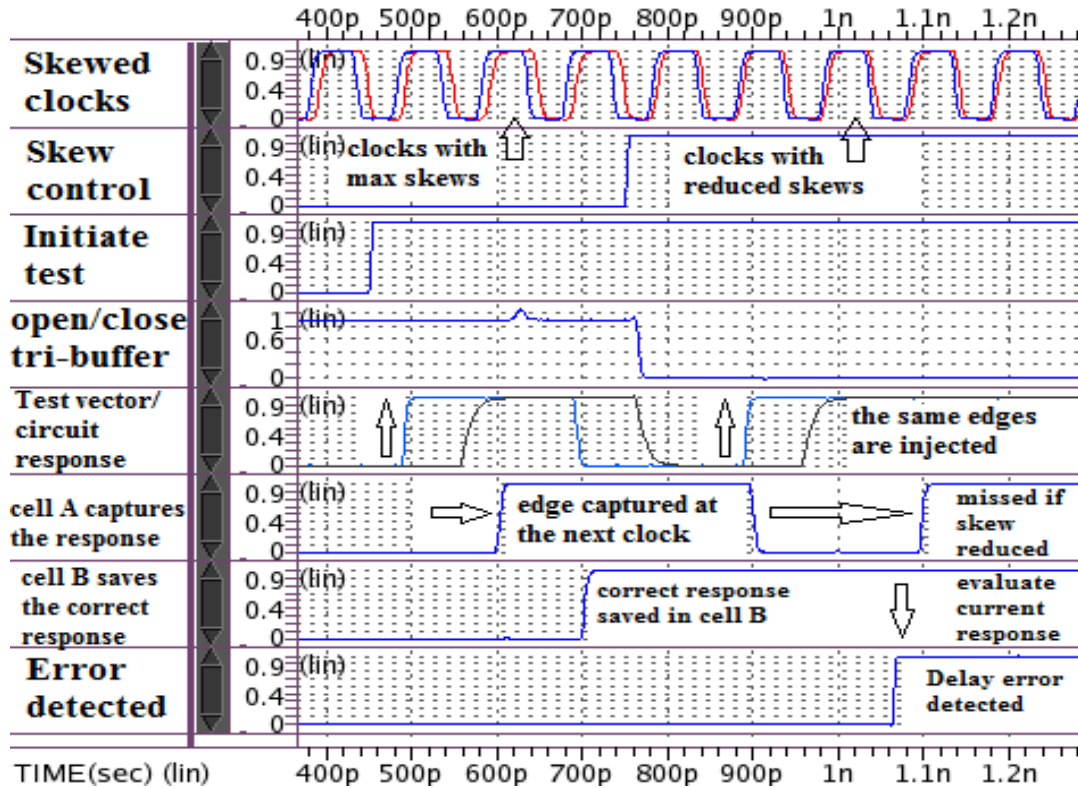


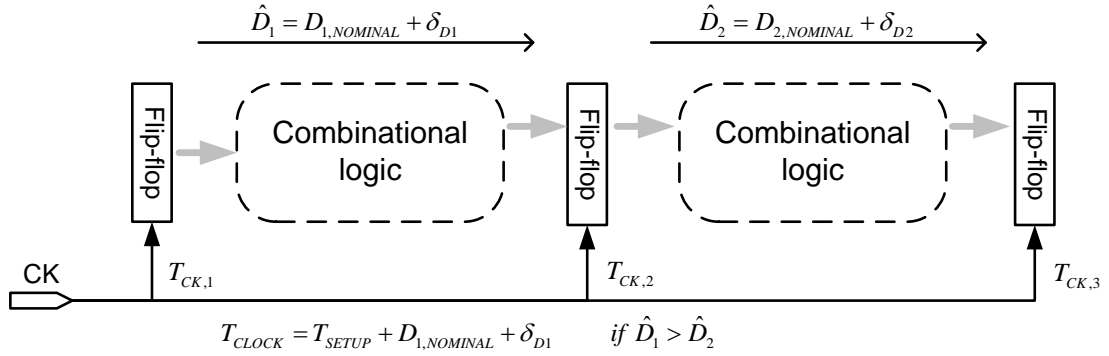
Figure IV.3. Control signals and internal waves of the proposed built-in delay testing scheme.

circuit response is saved into the next SCAN cell connected through SCAN chain. Then, the same vector is given with reduced clock skew which may cause delay fault due to tighter timing constraint. The circuit response from the second case is compared with the saved correct response, and the error signal is flagged if the two responses are different.

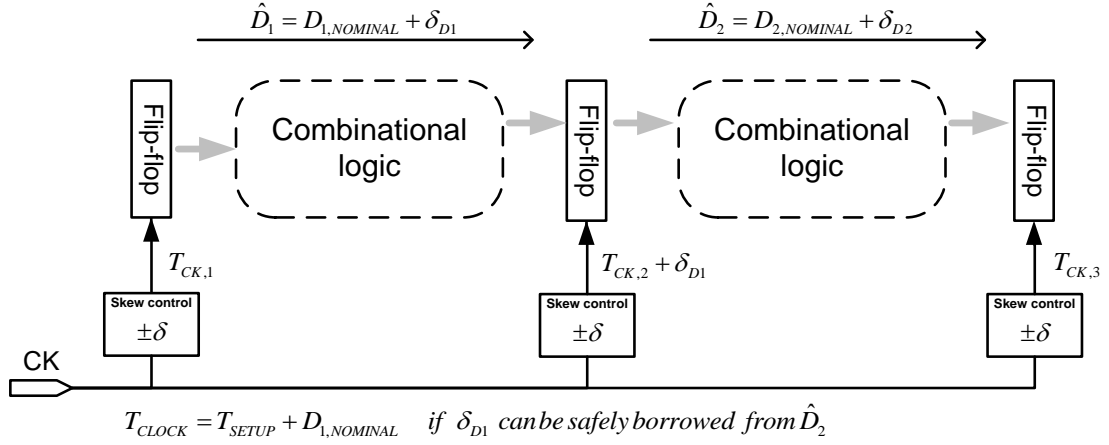
The proposed built-in delay testing scheme is very area efficient, since most of its components are designed to be shared by multiple functionality. In the common SCAN chain architecture, we simply added a tri-state buffer in the chain. Dynamic clock skew control block contributes to delay variation testing procedure as well as to the variation compensation system. Testing signal generation block given in Figure IV.1 (a) is implemented as a globally shared module so as to give the control signals to all testing modules simultaneously. Detailed analysis on area overhead from our approach compared to existing previous testing techniques will be given in section IV.2.B.

B. Variation Compensation Procedures

We present a combined variation handling system to handle both process variation and NBTI induced circuit aging effect. The proposed system employs dynamic clock skew control to mitigate process variation. The skew control block is already embedded in the delay testing scheme described in section IV.2.A. As discussed above, tuning clock skews against process variation is performed and determined simultaneously with the proposed delay testing procedure. On top of process variation, when circuit aging effect degrades performance, we activate boostable repeaters proposed in chapter III. In this variation compensation system, boostable repeaters are



(a) T_{CLOCK} with zero-skew clocking



(b) T_{CLOCK} with dynamic clock skew control

Figure IV.4. Variation mitigation through dynamic clock skew control. Both \hat{D}_1 and \hat{D}_2 may have variations with respect to their nominal values. When $\hat{D}_1 > \hat{D}_2$, dynamic clock skew control (b) can maintain T_{CLOCK} under the variation with borrowed skews, while zero skew clocking (a) needs to add δ_{D1} to T_{CLOCK} to evade delay fault due to \hat{D}_1 variation.

secondary modules which assist dynamic clock skew control to efficiently mitigate aging effects. This approach is to maximize usage of dynamic clock skew control because it is a basic component of the proposed built-in delay testing module. In this reason, boostable repeaters are deployed in a restricted manner enough only to assist circuit aging mitigation.

Figure IV.4 presents how dynamic clock skew control in the proposed adaptive system handles variation. Each combinational logic block in the Figure IV.4 takes delay of D (D_1 or D_2) for signals to pass its critical path, and the delay may also be affected by circuit variation, δ_D . \hat{D}_1 and \hat{D}_2 are degraded delays from the circuit blocks with respect to their nominal values. Suppose $\hat{D}_1 > \hat{D}_2$ so that global clock period (T_{CLOCK}) is determined by \hat{D}_1 . According to Figure IV.4(a), conventional clocking system with zero skew requires to increase T_{CLOCK} by δ_{D1} ($T_{CLOCK} = T_{SETUP} + D_{1,NOMINAL} + \delta_{D1}$) to avoid any delay fault induced by δ_{D1} variation. However, in the proposed adaptive system with dynamic clock skew control given in Figure IV.4(b), δ_{D1} variation can be compensated without increasing T_{CLOCK} . Suppose we control the clock skew at $T_{CK,2}$, and \hat{D}_2 has room for δ_{D1} such that $D_{1,NOMINAL} > \hat{D}_2 + \delta_{D1}$. Now, we can borrow δ_{D1} from $T_{CK,2}$ through dynamic skew control while \hat{D}_2 cause no delay fault. Therefore, in the proposed adaptive system, we do not need to increase T_{CLOCK} to mitigate δ_{D1} variation, and still $T_{CLOCK} = T_{SETUP} + D_{1,NOMINAL}$ which is δ_{D1} smaller than the zero-skew clocking case in Figure IV.4(a).

Table IV.1. Signal delay variations of sequential benchmark circuits due to process variation. Through Monte-Carlo simulations, each row represents one case of process variation, i.e., one individual fabricated chip. $\delta_D = 0.5 \times \{\max(\text{delay}) - \min(\text{delay})\}$. The maximum delay and the minimum delay of each circuit are highlighted.

#chip	S526 circuit			S1423 circuit			S5378 circuit		
	V _{DD}	Power (mW)	Path delay (ps)	V _{DD}	Power (mW)	Path delay (ps)	V _{DD}	Power (mW)	Path delay (ps)
1	0.9	3.245	68.5	0.9	8.88	92.4	0.9	20.32	214.4
2		3.244	74.1		8.85	87.4		20.51	213.2
3		3.25	74.8		8.91	86.5		20.76	212.8
4		3.251	70.8		8.96	94.4		20.46	225.2
5		3.245	68.9		8.91	88.4		20.85	225.5
6		3.25	70.6		8.89	85.0		20.73	223.2
7		3.243	72.3		8.91	87.0		20.46	223.3
8		3.25	73.9		8.87	91.0		20.72	226.1
9		3.257	77.5		8.98	93.7		20.29	225.4
10		3.243	70.4		8.88	90.0		20.49	220.1
		$\delta_D =$	4.5 ps		$\delta_D =$	4.7 ps		$\delta_D =$	6.6 ps

Table IV.2. Estimated maximum power reduction achievable through δ_D gain in T_{CLOCK} through dynamic clock skew control. One chip (one process variation case) per circuit is examined, and power reduction is measured through lowering V_{DD} down until the relaxed timing constraints by δ_D are still satisfied.

S526 circuit			S1423 circuit			S5378 circuit		
V _{DD}	Power (mW)	Path delay (ps)	V _{DD}	Power (mW)	Path delay (ps)	V _{DD}	Power (mW)	Path delay (ps)
1	4.27	65.2	0.98	11.3	85.6	0.96	24.53	213.3
0.87	2.975	69.9	0.9	8.88	90.0	0.9	20.49	220.1
Power reduction = 30.3% $\delta_D=4.7ps$			Power reduction = 21.4% $\delta_D=4.4ps$			Power reduction = 16.5% $\delta_D=6.8ps$		

In terms of power consumption, the proposed adaptive system with dynamic clock skew control can tune logic circuit blocks to consume less power while providing the same circuit performance thanks to the relaxed timing constraints compared to conventional zero-skew clocking as discussed above. The δ_{D1} difference in T_{CLOCK} between conventional zero-skew clocking and dynamic clock skew control can be converted to power reduction through reducing V_{DD} . The actual δ_D variation under process variation is analyzed through simulations on sequential benchmark circuits, and their delay variations (δ_D) are given in Table IV.1. S526, S1423, and S5378 ISCAS'89 benchmark circuits are simulated with process variation (gate length variation: Gaussian $\mu=45\text{nm}$, $3\sigma=15\%$ of mean; V_{th} variation: Gaussian $\mu_{NMOS}=0.3423\text{V}$ $\mu_{PMOS}=-0.23122\text{V}$, $3\sigma=20\%$ of mean). Monte-Carlo iterations yielded 10 random cases which imply 10 fabricated chips, and their critical path delay are measured. Their path delay varies in accordance with process variation, and δ_D of each circuit is one half of the maximum variation, $\delta_D = 0.5 \times \{\max(\text{delay}) - \min(\text{delay})\}$. Table IV.2 estimates the maximum achievable power reduction through the relaxed timing constraints by δ_D . In each circuit, V_{DD} is lowered down until it still satisfies the relaxed timing constraints. For example, in S526 circuit, when $V_{DD} = 1\text{V}$, the critical path delay is 65.2 ps. Suppose we can give $\delta_D=4.7\text{ps}$ of relaxed timing constraint to the path through dynamic skew control, then V_{DD} can be reduced down to 0.87V only to meet 69.9 ps ($65.2\text{ps} + \delta_D$), thereby 30.3% less power consumption is achieved. Theoretically, this is the maximum estimation of power reduction where we assumed that δ_D can be certainly borrowed in total. If

Table IV.3. Circuit characteristics and experimental setup for the proposed adaptive system.

circuit	#gates	#FF	Built-in delay test module		#FF with dynamic clock skew control	Boostable-repeater	
			#SCAN FF	# critical paths monitored with test module		# total repeaters	# boostable repeaters
S526	193	21	14	4	8	68	4
S1423	657	74	50	13	26	168	14

adjacent blocks have not enough timing margins, adaptive system merely borrows skews less than δ_D which in turn yields power reduction less than the maximum estimation.

2. Experimental Validation

A. Simulation Configurations

The proposed built-in delay test scheme in conjunction with dynamic skew control and boostable repeater is tested by SPICE simulations on benchmark circuits and compared with conventional methods. The benchmark circuits are S526 and S1423 from ISCAS 89 suite in 45nm technology. The device models of 45nm technology from PTM [35] are employed. The characteristics of the two benchmark circuits are shown in Table IV.3. The 5th column of Table IV.3 tells the number of critical paths as well as the number of built-in delay testing modules monitoring the paths. The number of FFs with dynamic clock skew control is given in the 6th column. The 7th column indicates the number of repeaters in each circuit, and the rightmost column lists the number of repeaters which are replaced with boostable repeaters. In this configuration, process

variation is compensated by dynamic clock skew control, and circuit aging effect is handled by both boostable repeaters and dynamic clock skew control.

Two kinds of variations are considered in the experiment: manufacturing process variations and NBTI-induced pMOS performance degradation. All these variations are assumed to follow Gaussian distribution. For the process variations, we focus on gate length variation and threshold voltage variation. The standard deviations of gate length variation and threshold voltage variation are 5% and 6.7% of their nominal values, respectively (gate length variation: Gaussian $\mu=45\text{nm}$, $3\sigma=15\%$ of mean; V_{th} variation: Gaussian $\mu_{NMOS}=0.3423\text{V}$ $\mu_{PMOS}=-0.23122\text{V}$, $3\sigma=20\%$ of mean). For an aged circuit, additional threshold voltage degradation on pMOS transistors is considered. The mean and standard deviation of the degradation are 10% and 3.3% of the nominal values, respectively (additional V_{th} variation: Gaussian $\mu=10\%$ of μ_{NMOS} or μ_{PMOS} , $3\sigma=10\%$ of mean). The process variations and threshold voltage degradations are also applied to all the components of the proposed built-in delay testing scheme, dynamic clock skew control, as well as boostable repeaters. Due to overly long simulation time, we ran 10 random Monte-Carlo iterations considering process variation and aging effects, instead of full-fledged Monte Carlo test on each circuit.

B. Overheads from Delay Fault Detection Modules

Area overhead from delay error detection system is an important factor to compose adaptive design to mitigate fine-grained variation. Compensation components can effectively focus extra resources on severely degraded paths only if sensors are

deployed in a widespread way so as to monitor as many critical paths as possible. If delay fault detection system occupies large die area, adaptive approach loses its advantage against over-design approach which simply enlarges devices. In Figure IV.5 (a), area overhead from our proposed built-in delay testing scheme is compared with two other techniques, one is a well-known double-sampling flip-flop design of Razor [23], and the other is a built-in aging sensor presented in [24] which sets a guard-band to detect/prevent delay error. We deployed sensors in ISCAS89 benchmark circuits, and in each case, the sensors are monitoring the same number of critical paths for fair comparison. According to Figure IV.5 (a), the proposed built-in delay testing scheme

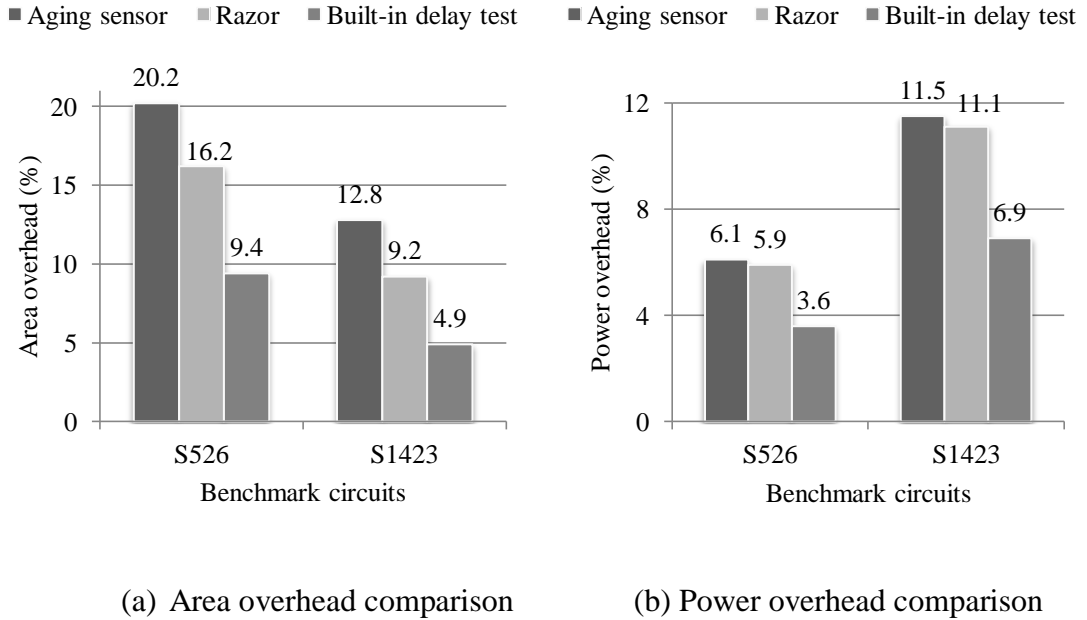


Figure IV.5. Area and power overhead comparison between the three delay fault detection approaches. The rightmost bars represent overheads from the proposed built-in delay testing scheme, and show clear benefits against the other two approaches.

Table IV.4. Die area occupied by one individual testing module of three delay fault detection approaches (normalized by the summation of width of all transistors in one module).

	Razor		Aging sensor		Built-in delay testing module
	Area (μm)	Area vs. built-in delay test	Area (μm)	Area vs. built-in delay test	Area (μm)
Area of one testing module	16.24	2.1 \times	25.04	3.2 \times	7.84 (1 \times)

represented by the rightmost bars shows significantly low area overhead compared to the other two techniques. This considerably low overhead can be achieved since, in the proposed approach, each individual module watching one path delay does not consist of any duplicated component or complex structure to detect error, while the Razor FF needs a duplicated latch and the aging sensor [24] has elaborate circuitries for sensing. Table IV.4 compares die areas occupied by each of delay testing modules. When calculating one module area of our approach, we considered area for shared control block as well which generates testing control signals to multiple modules. According to Table IV.4, each one module of Razor is 2.1 times larger than one module of the proposed approach due to its shadow latch and control logics, and the aging sensor [24] is even 3.2 times larger than ours.

Average power consumption is a critical factor in testing strategy in terms of circuit reliability. Increase in average power during testing incurs additional thermal load

Table IV.5. Average power consumption during testing operation period.

	Conventional whole circuit (w/o testing module)	Aging sensor (+Dynamic skew control)	Razor (+ Dynamic skew control)	Built-in delay test module (+ Dynamic skew control)
	Power (mW)	Power (mW) overhead	Power (mW) overhead	Power (mW) overhead
S526	12.85	13.63 6.1%	13.60 5.9%	13.32 3.6%
S1423	27.99	31.20 11.5%	31.09 11.1%	29.91 6.9%

Table IV.6. Energy consumption during testing operation period.

	Conventional whole circuit (w/o testing module)	Aging sensor (+Dynamic skew control) Operation takes 1 clock	Razor (+ Dynamic skew control) Operation takes 1 clock	Built-in delay test module (+ Dynamic skew control) Operation takes 3 clocks
	Energy (pico J)	Energy (pico J) overhead	Energy (pico J) overhead	Energy (pico J) overhead
S526	2.25 (in 1 clock) 6.75 (in 3 clocks)	2.39 6.1%	2.38 5.9%	6.99 3.6%
S1423	4.48 (in 1 clock) 13.44 (in 3 clocks)	4.99 11.5%	4.97 11.1%	14.36 6.9%

and hot spot which may cause structural damage to the silicon [21]. In terms of average power consumption during testing operation period, the proposed approach also incurs less power overhead than the other two techniques, as presented in Figure IV.5 (b) and Table IV.5. In Table IV.5, each case with delay fault testing system is compared with a conventional whole circuit power without testing modules. The proposed built-in delay testing scheme consumes around 40% less average power than other two techniques. However, since the proposed approach takes multiple clock periods for detecting

operation, total energy consumption from the proposed systems becomes larger than other two approaches both of which detect error in one clock cycle, as given in Table IV.6. Higher energy consumption, however, has limited effects only on the battery-operated devices where battery life will be impacted [21]. Even in the case with battery-operated systems, we can easily evade the high energy consumption issue simply by conducting testing procedures when external power is available for charging their battery. Since the proposed approach exploits BIST SCAN architecture, the testing vectors generated through ATPG (automatic test pattern generator) can perform other BIST procedures for fault detections as well as the proposed built-in delay testing scheme simultaneously, therefore the vectors can make the best use of its energy usage.

C. Simulations on Sequential Logic Circuits

The proposed built-in delay testing scheme in conjunction with variation compensation systems, dynamic clock skew control and boostable repeater, is tested in ISCAS'89 sequential benchmark circuits S526 and S1423, as described in section IV.2.A. In fresh circuits with process variation only, dynamic clock skew system handles delay fluctuations. Under the circuit degradations due to aging effect in addition to process variation, boostable repeater as well as dynamic clock skew cooperates to mitigate circuit aging. In the proposed adaptation system, boostable repeater is a supplemental module to help dynamic clock skew efficiently compensate circuit variation, in order to maximize usage of dynamic clock skew control which is a basic component of the proposed built-in delay testing module as well. In this reason,

boostable repeaters are deployed in a restricted manner solely enough to handle circuit aging.

The results are listed in Table IV.7. For all circuit configurations, clock periods are tuned to meet the same timing constraint in each benchmark circuit. T_{CLOCK} in S526 circuit is 175ps, and T_{CLOCK} in S1423 circuit is 160ps. The power dissipations are the average results over the 10 random instances of each design. The power estimation here is from SPICE simulations and includes both dynamic power and leakage power from logic circuit, the proposed built-in delay testing module, and compensation system as well. The power reductions from the proposed adaptive system in fresh circuit considering process variation are compared with overly-designed circuit and conventional ASV only system, and the reductions are listed on 9th and 10th columns of Table IV.7 (a). For fresh circuits under processing variations, our system with dynamic skew control achieves 47.5% average power reduction compared to over-design, and 17.9% average power reduction compared to conventional ASV only system. To mitigate circuit aging effect as well as process variation in aged circuits, the proposed adaptive system turns on boostable repeaters and tunes clock skews simultaneously. In Table IV.7 (b), we compared power consumptions from circuits with the proposed adaptive system against conventional systems, and we separately analyzed power reduction capability from each combination of the two techniques applied in the proposed system: dynamic clock skew control and boostable repeaters. Listed in Table IV.7 (b) are the three combinations: dynamic clock skew control only system, boostable repeaters only system, and the proposed system which cooperatively uses two techniques

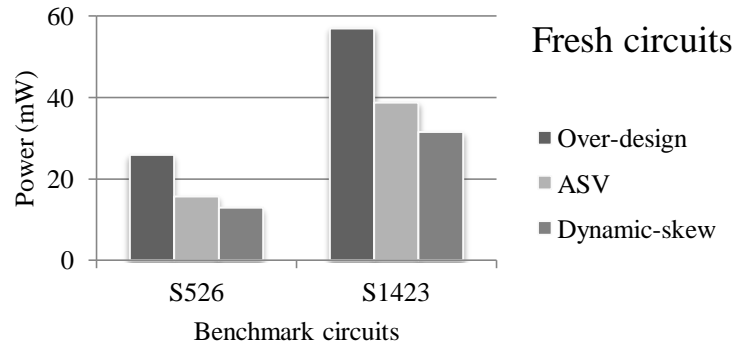
Table IV.7. Power reduction acquired though the proposed adaptive system. The proposed system is compared to over-design and conventional global ASV system only.

Circuit	Over-design		Conventional ASV		Dynamic skew control only			
	V _{DD}	Power (mW)	V _{DD}	Power (mW)	V _{DD}	Power (mW)	Power reduction (%) vs. over-design	Power reduction (%) vs. ASV
S526	1.08	25.90	0.87-1.01	15.73	0.82-0.93	12.89	50.2%	17.8%
S1423	1.16	57.04	0.93-1.09	38.72	0.88-0.99	31.49	44.8%	17.9%

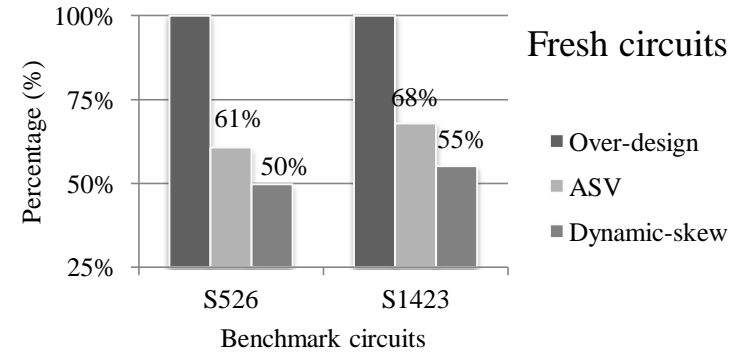
- (a) Power consumptions and achieved power reductions when circuit is fresh. Delay variations due to process variation are only considered. The proposed adaptive system uses dynamic skew control to mitigate process variation in fresh circuit. Boostable buffers are turned off.

	Over-design		Conventional ASV		Dynamic skew control only				Boostable repeater only				Dynamic skew control + Boostable repeater			
	V _{DD}	Power (mW)	V _{DD}	Power (mW)	V _{DD}	Power (mW)	Power reduction (%) vs. over-design	Power reduction (%) vs. ASV	V _{DD}	Power (mW)	Power reduction (%) vs. over-design	Power reduction (%) vs. ASV	V _{DD}	Power (mW)	Power reduction (%) vs. over-design	Power reduction (%) vs. ASV
S526	1.08	24.73	0.92-1.08	17.73	0.87-1.02	15.30	38.1%	13.6%	0.85-0.91	14.24	42.4%	19.5%	0.82-0.92	12.15	50.8%	31.1%
S1423	1.16	56.31	0.95-1.16	44.81	0.91-1.06	35.93	36.2%	18.6%	0.88-1.02	32.49	42.3%	26.4%	0.86-0.92	27.28	51.6%	37.3%

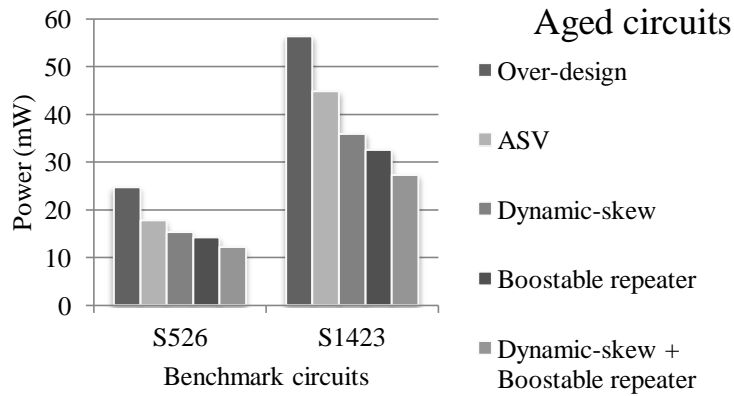
- (b) Power consumptions and achieved power reductions when circuit is aged. Both process variation and circuit aging are considered in delay variations in the benchmark circuits. The proposed adaptive system turns on boostable repeaters as well as tunes clock skews when handling circuit aging.



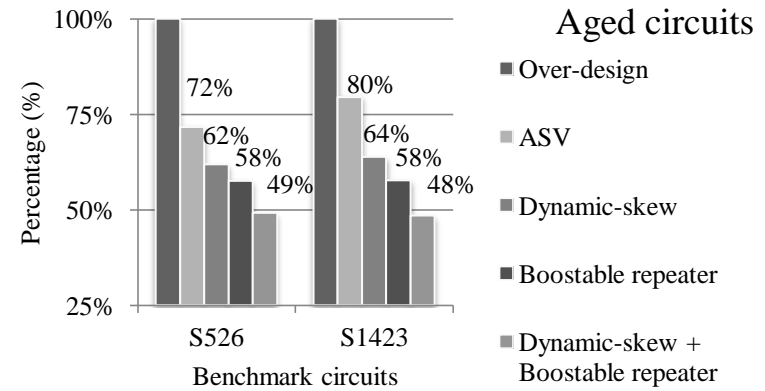
(a) Total circuit power consumptions in fresh circuits.



(b) Percentages of circuit powers are compared.



(c) Total circuit power consumptions in aged circuits.



(d) Percentages of circuit powers are compared

Figure IV.6. Power consumption comparison for the proposed adaptive system. Power consumptions in fresh circuits and aged circuits are compared against conventional systems.

at the same time. Among the three combinations, the proposed combined approach achieves the best power reductions against conventional systems, according to the rightmost 4 columns in Table IV.7 (b). On the average, in aged circuits, the power reductions from the proposed adaptive approach are 51.2% and 34.2% compared to over-design and conventional ASV system only, respectively.

The bars of Figure IV.6 show more clear comparison in power consumptions from the proposed adaptive system against other systems. The leftmost bars are for the power consumed in overly-designed circuits, and the second bars represent circuits with conventional ASV only system. The rightmost bars are for the power consumption from circuits equipped with the proposed approach with dynamic clock skew in fresh circuits, and combined with boostable repeaters in aged circuits. The proposed system shows the best power efficiency among all the systems.

3. Conclusions

In this work, we proposed a built-in delay testing scheme which significantly reduces area overhead from delay variation detection system, and consumes low average testing power. This testing scheme is applied to the proposed adaptive system combined with dynamic clock skew control and boostable repeater design in order to mitigate delay fluctuations due to process variation and circuit aging effect. The proposed system is validated with experimental results which show that, compared to conventional approaches, over 26% less power consumption is achieved through the proposed adaptive system.

V. CONCLUSIONS AND FUTURE WORK

1. Conclusions

As technology scales down in modern VLSI design, mitigating circuit variations becomes increasingly challenging due to fine-grained nature of process variations and circuit aging effects. In this dissertation, we studied critical issues and considerations to be addressed for building a fine-grained variation adaptation system, and discussed about limitations of existing technologies which hinder effectively handling the variations. Several contributions in voltage adaptation approaches as well as efficient variation sensing system have been proposed in this dissertation to solve the problems.

A dual-level Adaptive Supply Voltage (ASV) system based on voltage islands design is presented to handle the fine-grained variations in circuits with many timing critical paths. In conjunction with conventional global ASV in voltage islands, the proposed system generates a lightweight local supply with limited load capacity. Only the logic gates on severely degraded paths are selectively powered by the local supply, while all the remaining gates of the island still consume less power with global supply whose voltage level is lower than the local supply. A tiny and effective voltage regulator, Mini Programmable linear voltage Regulator (MPR), is proposed to provide the local supply voltage whose efficiency can be further enhanced through the proposed progressive voltage enhancement method. Compared to over-design and conventional ASV only system, the proposed system achieves significant power saving while satisfying equal timing constraints.

A novel boostable repeater design is proposed to efficiently compensate fine-grained variations in VLSI interconnects which are critical bottlenecks in modern circuit design. The proposed boostable repeater internally raises its switching speed by itself without need of external source for boosting. The boosting feature of the proposed repeater can be turned off for low power operation. Combined with global ASV system which provides coarse-grained variation adaptation, the proposed boostable repeaters deployed in interconnects selectively enhance strongly degraded critical paths due to fine-grained variations. We validated the proposed adaptive system equipped with boostable repeaters through experiments which compare the proposed adaptive system against existing techniques including over-design, conventional ASV only system, and online programmable finger design. Experimental results show that significant power reductions are achieved through the proposed adaptive system with boostable repeaters.

An area-efficient built-in delay testing scheme is suggested to construct a complete system of fine-grained variation adaptation. Mitigating fine-grained variation inevitably needs to monitor a large number of signal paths, therefore requires sensing system to be minute in size. The proposed scheme minimizes its area overhead in virtue of sharing most of its components between multiple functionalities and common architectures in modern VLSI circuit. A variation adaptation system composed of the proposed testing scheme and dynamic skew control is proposed, and tested through simulations. Compared to conventional approaches, the same performance and robustness with less power consumption are achieved through the proposed system. In terms of testing power and energy, while the proposed scheme consumes more testing

energy than counterparts, average testing power consumption is less in the proposed scheme which is more critical issue in system testing strategy.

2. Future Work

In modern processors, numerous configurations of dynamic voltage frequency scaling (DVFS) are already activated. Combining nowadays complex coarse-grained adaptive system with fine-grained variation compensation inevitably requires efficient adaptation architectures and procedures as well as feasibility to incorporate the compensation system with existing standard digital design process. Power saving capability of the proposed voltage adaptation techniques which tunes circuit performance per critical path resolution can be further applicable to implement enhanced dynamic power management systems.

REFERENCES

- [1] S. Borkar, "Designing Reliable Systems from Unreliable Components: The Challenges of Transistor Variability and Degradation," IEEE Micro, Vol. 25, No. 6, pp. 10-16, November-December 2005.
- [2] J. W. Tschanz, J. T. Kao, S. G. Narendra, R. Nair, D. A. Antoniadis, A. P. Chandrakasan, and V. De, "Adaptive Body Bias for Reducing Impacts of Die-to-Die and Within-Die Parameter Variations on Microprocessor Frequency and Leakage," IEEE Journal of Solid-State Circuits, Vol. 37, No. 11, pp. 1396-1402, November 2002.
- [3] J. Gregg and T. Chen, "Post Silicon Power/Performance Optimization in the Presence of Process Variations Using Individual Well-Adaptive Body Biasing," IEEE Transactions on VLSI Systems, Vol. 15, No. 3, pp. 366-376, March 2007.
- [4] T. Chen and S. Naffziger, "Comparison of Adaptive Body Bias (ABB) and Adaptive Supply Voltage (ASV) for Improving Delay and Leakage Under the Presence of Process Variations," IEEE Transactions on VLSI Systems, Vol. 11, No. 5, pp. 888-899, October 2003.
- [5] J. W. Tschanz, S. Narendra, R. Nair and V. De, "Effectiveness of Adaptive Supply Voltage and Body Bias for Reducing Impact of Parameter Variations in Low Power and High Performance Microprocessors," IEEE Journal of Solid-State Circuits, Vol. 38, No. 5, pp. 826-829, May 2003.

- [6] C. Neau and K. Roy, "Optimal Body Bias Selection for Leakage Improvement and Process Compensation Over Different Technology Generations," IEEE International Symposium on Low Power Electronics and Design, pp. 116-121, August 2003.
- [7] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand, "Leakage Current Mechanisms and Leakage Reduction Techniques in Deep-Submicrometer CMOS Circuits," Proceedings of IEEE, Vol. 91, No. 2, pp. 305-327, February 2003.
- [8] K. Agarwal and K. Nowka, "Dynamic Power Management by Combination of Dual Static Supply Voltages," IEEE International Symposium on Quality Electronic Design, pp. 85-92, March 2007.
- [9] R. Samanta, G. Venkataraman, N. Shah and J. Hu, "Elastic Timing Scheme for Energy-Efficient and Robust Performance," IEEE International Symposium on Quality Electronic Design, pp. 537-542, March 2008.
- [10] X. Liang, G.-Y. Wei and D. Brooks, "ReVIVaL: A Variation-Tolerant Architecture Using Voltage Interpolation and Variable Latency," IEEE International Symposium on Computer Architecture, pp. 191-202, June 2008.
- [11] D. E. Lackey, P. S. Zuchowski, T. R. Bednar, D. W. Stout, S. W. Gould and J. M. Cohn, "Managing Power and Performance for System-on-Chip Designs Using Voltage Island," IEEE/ACM International Conference on Computer-Aided Design, pp. 195-202, November 2002.
- [12] S.-J. Jou and T.-L. Chen, "On-Chip Voltage Down Converter for Low-Power Digital System," IEEE Transactions on Circuits and Systems – II: Analog and Digital Signal Processing, Vol. 45, No. 5, pp. 617-625, May 1998.

- [13] P. Hazucha, T. Karnik, B. A. Bloechel, C. Parsons, D. Finan and S. Borkar, "Area-Efficient Linear Regulator with Ultra-Fast Load Regulation," IEEE Journal of Solid-State Circuits, Vol. 40, No. 4, pp. 933-940, April 2005.
- [14] E. Sanchez-Sinencio, "Low Drop-Out Linear Regulators : Design Considerations and Trends for High Power Supply Rejection (PSR)," IEEE Santa Clara Valley (SCV) Solid State Circuits Society, February 2010.
- [15] W. Kim, M. S. Gupta, G.-Y. Wei and D. Brooks, "System Level Analysis of Fast, Per-Core DVFS Using On-Chip Switching Regulators," IEEE International Symposium on High Performance Computer Architecture, pp. 123-134, February 2008.
- [16] D. Maurath and Y. Manoli, "A Self-Adaptive Switched-Capacitor Voltage Converter with Dynamic Input Load Control for Energy Harvesting," European Solid-State Circuits Conference, pp. 284-287, September 2009.
- [17] A. B. Kahng, S. Reda and P. Sharma, "On-Line Adjustable Buffering for Runtime Power Reduction," IEEE International Symposium on Quality Electronic Design, pp. 550-555, March 2007.
- [18] H. Kaul and D. Sylvester, "A Novel Buffer Circuit for Energy Efficient Signaling in Dual-Vdd Systems," ACM/IEEE Great Lakes Symposium on VLSI, pp. 462-467, April 2005.
- [19] E. Talpes and D. Marculescu, "Toward a Multiple Clock/Voltage Island Design Style for Power-Aware Processors," IEEE Transactions on VLSI Systems, Vol. 13, No. 5, pp. 591-603, May 2005.

- [20] P. J. Osler, "Placement Driven Synthesis Case Studies on Two Sets of Two Chips: Hierarchical and Flat," ACM International Symposium on Physical Design, pp. 190-197, March 2004.
- [21] L. Wang, C. Stroud, and N. Touba, "System On Chip Test Architectures," Morgan Kaufmann Publishers, Burlington, MA, 2007.
- [22] R. Datta, J. A. Abraham, A. U. Diril, A. Chatterjee and K. Nowka, "Adaptive Design for Performance-Optimized Robustness," IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems, pp. 3-11, October 2006.
- [23] S. Das, D. Roberts, S. Lee, S. Pant, D. Blaauw, T. Austin, K. Flautner and T. Mudge, "A Self-Tuning DVS Processor Using Delay-Error Detection and Correction," IEEE Journal of Solid-State Circuits, Vol. 41, No. 4, pp. 792-804, April 2006.
- [24] M. Agarwal, B. C. Paul, M. Zhang and S. Mitra, "Circuit Failure Prediction and Its Application to Transistor Aging," IEEE VLSI Test Symposium, pp. 277-286, May 2007.
- [25] Y. Su, W. Hon, C. Yang, S. Chang, and Y. Chang, "Clock Skew Minimization in Multi-Voltage Mode Designs Using Adjustable Delay Buffers," IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Vol. 29, No. 12, pp. 1921–1930, December 2010.
- [26] K. Nagaraj and S. Kundu, "Process Variation Mitigation via Post Silicon Clock Tuning," ACM Great Lakes Symposium on VLSI, pp. 227–232, May 2009.

- [27] S. Lee, S. Das, T. Pham, T. Austin, D. Blaauw, T. Mudge, “Reducing Pipeline Energy Demands with Local DVS and Dynamic Retiming,” International Symposium on Low Power Electronics and Design, pp. 319-324, August 2004.
- [28] A. Tiwari, S. R. Sarangi, J. Torrellas, “ReCycle: Pipeline Adaptation to Tolerate Process Variation,” ACM International Symposium on Computer Architecture, pp. 323-334, June 2007.
- [29] R. J. Widlar, “New Developments in IC Voltage Regulators,” IEEE Journal of Solid-State Circuits, Vol. SC-6, pp. 2–7, February 1971.
- [30] A. P. Brokaw, “A Simple Three-Terminal IC Bandgap Reference,” IEEE Journal of Solid-State Circuits, Vol. SC-9, pp. 388–393, December 1974.
- [31] M. Nicolaidis, “Time Redundancy Based Soft-Error Tolerance to Rescue Nanometer Technologies,” IEEE VLSI Test Symposium, pp. 86-94, April 1999.
- [32] D. Sengupta and R. Saleh, “Application-Driven Floorplan-Aware Voltage-Island Design,” ACM Design Automation Conference, pp. 155-160, June 2008.
- [33] J. Hu, Y. Shin, N. Dhanwada, and R. Marculescu., “Architecting Voltage Islands in Core-Based System-on-a-Chip Designs,” IEEE International Symposium on Low Power Electronics and Design, pp. 180-185, July 2004.
- [34] M. Zhao, Y. Fu, V. Zolotov, S. Sundareswaran, and R. Panda, “Optimal Placement of Power Supply Pads and Pins,” IEEE Transactions Computer-Aided Design Integrated Circuits System, Vol. 25, No. 1, pp. 144–154, January 2006.
- [35] Y. Cao, Predictive Technology Model, <http://www.eas.asu.edu/~ptm>, 2008.

- [36] C. Pei, R. Booth, H. Ho, N. Kusaba, X. Li, M. Brodsky, P. Parries, H. Shang, R. Divakaruni, and S. Iyer, "A Novel, Low-Cost Deep Trench Decoupling Capacitor for High-Performance, Low-Power Bulk CMOS Applications," IEEE International Conference on Solid-State and Integrated-Circuit Technology, pp. 1146-1149, October 2008.
- [37] L. Chang, R.K. Montoye, 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/mm²," IEEE Symposium on VLSI Circuits, pp. 55-56, June 2010.
- [38] P. Zhou, K. Sridharan, and S. Sapatnekar, "Congestion-Aware Power Grid Optimization for 3D Circuits Using MIM and CMOS Decoupling Capacitors," IEEE Asia and South Pacific Design Automation Conference, pp. 179-184, January 2009.