Variation-Aware Supply Voltage Assignment for Simultaneous Power and Aging Optimization

Xiaoming Chen, Yu Wang, Yu Cao, Yuchun Ma, and Huazhong Yang

***Abstract*—As technology scales, negative bias temperature instability (NBTI) has become a major reliability concern for circuit designers. And the growing process variations can no longer be ignored. Meanwhile, reducing power consumption remains to be one of the design goals. In this paper, a variation-aware supply voltage assignment (SVA) technique combining dual V_{dd} assignment and dynamic V_{dd} scaling is proposed on a statistical platform, to minimize circuit power under an aging-aware timing constraint. The experimental results show that our SVA technique can mitigate on average 62% of the NBTI-induced circuit delay degradation. Compared with guard-banding and single V_{dd} scaling approaches, our approach saves more energy.**

***Index Terms*—Dynamic power, leakage power, negative bias temperature instability (NBTI), supply voltage assignment (SVA).**

I. Introduction

With the continuous scaling of CMOS technology, negative bias temperature instability (NBTI) is emerging as one of the major reliability degradation mechanisms [1]. NBTI is an aging effect which gradually increases the threshold voltage (V_{th}) of pMOS transistors when they are negatively biased, thus increasing the gate delay. Meanwhile, leakage power has become a large portion of the total power consumption. Moreover, the growing process and device variations are emerging as key influencing factors of circuit performance. Traditional worst-case design will lead to an over-pessimistic estimation. Instead, statistical static timing analysis (SSTA) is an effective technique to evaluate the increasing variations instead of the traditional STA [2].

Researchers have explored many techniques to mitigate NBTI-induced degradation, such as NBTI-aware synthesis [3], gate and transistor sizing [4], [5], input vector control (IVC) [6]–[8], internal node control (INC) [9], [10]. These techniques are all "one-time" fixed solutions, which give the circuits a high guard-band strength, leading to large positive slacks during the initial time, therefore result in large area and power overhead [11].

Recently some adaptive (dynamic) techniques were proposed. Zhang *et al.* [12] proposed a scheduled voltage scaling technique, which gradually increased V_{dd} to compensate for NBTI-induced degradation. Their technique has the potential to increase IC lifetime by 46%. However, the scheduled method may not be suitable for any circuit, and their approach significantly increases the leakage power. Kumar *et al.* [11] proposed adaptive body biasing (ABB) and adaptive supply voltage (ASV) techniques: they adjusted the supply/bias voltage to recover the circuit performance. ABB and ASV increases the leakage by 23% on average. Mintarno *et al.* [13] proposed an optimized self-tuning method: they adjusted V_{dd} and clock frequency to maximize the lifetime computational power efficiency considering NBTI and DIBL induced degradation. However, the time stamps in [13] are uniform, so it may not be the optimal solution. A recent publication [14] claimed that power savings of dynamic voltage tuning could be significant initially, but were limited after the first few months of operation. However, they only considered the power saving but not energy saving. In this paper, we will show that the energy can be effectively saved by our approach.

For the above reasons, in this paper we attempt to develop a new technique which can mitigate NBTI-induced degradation and reduce power simultaneously. Traditional power reduction techniques, such as dual V_{dd} or dual V_{th}, reduce power by the sacrifice of some positive slacks, thus lead to the increase of the number of critical paths, and make the performance of critical paths degrades. We utilize dual V_{dd} and dynamic V_{dd} scaling, which increases the voltage of critical paths to ensure their performance, and decreases the voltage of non-critical paths to reduce power. Since the partition of dual V_{dd} islands and the scaling method have large impact on both circuit performance and power, reliability and power should be simultaneously considered when partitioning the dual V_{dd} islands and scaling the voltage values. The contribution of this paper can be summarized as follows.

- We propose a variation-aware SVA technique combining dual V_{dd} assignment and dynamic V_{dd} scaling. The high V_{dd} is used to compensate for NBTI-induced degradation; while the low V_{dd} is used to reduce power. During circuit operation, the optimal V_{dd} values are dynamically determined according to the aging-aware timing constraint.
- The experimental results show that our approach can mitigate on average 62% of NBTI-induced degradation. Compared with guard-banding and single V_{dd} scaling approach, our method can effectively save the energy.

The rest of this paper is organized as follows. Section II reviews the models. The variation-aware SVA technique is proposed in Section III. We show the experimental results in Section IV. Finally the paper is concluded in Section V.

II. Model Review

1) Gate Delay Model: The load dependent delay of gate v is given by the alpha-power law [15]

$$D(v) = \frac{KC_vV_{dd}}{(V_{dd}-V_{th})^{\alpha}} \approx KC_vV_{dd}^{1-\alpha}\left(1+\alpha\frac{V_{th}}{V_{dd}}\right) \quad (1)$$

where C_v is the load capacitance of gate v, V_{dd} and V_{th} are the supply voltage and the threshold voltage, respectively, K and α ($1 < \alpha < 2$) are the proportionality constant and the velocity saturation index, respectively.

2) NBTI Degradation Model: NBTI can be described using reaction-diffusion (R-D) mechanism [16]–[20]. The compact model to predict long term NBTI is given by [19]

$$\Delta V_{th} = \left(\frac{\sqrt{K_v^2 T_{\mathrm{clk}}\,\omega}}{1-\beta_t^{1/2n}}\right)^{2n} \quad (2)$$

$$\beta_t = 1 - \frac{2\xi_1 t_e + \sqrt{\xi_2 C(1-\omega)T_{\mathrm{clk}}}}{2T_{ox}+\sqrt{Ct}} \quad (3)$$

$$K_v = \left(\frac{qT_{ox}}{\varepsilon_{ox}}\right)^3 k^2 C_{ox}(V_{gs}-V_{th})\sqrt{C}\exp\left(\frac{2E_{ox}}{E_0}\right) \quad (4)$$

Manuscript received November 16, 2010; revised March 11, 2011; accepted September 03, 2011. Date of publication October 13, 2011; date of current version July 27, 2012. This work was supported by National Key Technological Program of China (2008ZX01035-001, 2010ZX01030-001) and National Natural Science Foundation of China (60870001). The work of Y. Ma was supported by National Natural Science Foundation of China (61076035).

X. Chen, Y. Wang, and H. Yang are with the Department of Electronic Engineering, Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China (e-mail: chenxm05@mails.tsinghua.edu.cn; yu-wang@tsinghua.edu.cn; yanghz@tsinghua.edu.cn).

Y. Ma is with the Department of Computer Science, Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China (e-mail: myc@tsinghua.edu.cn).

Y. Cao is with the Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287-5706 USA (e-mail: yu.cao@asu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVLSI.2011.2168433

1063-8210/$26.00 © 2011 IEEE

where q is the electron charge, k is the Boltzmann constant, C_{ox} is the oxide capacitance per unit area, $E_{ox} = V_{gs}/T_{ox}$, $C = 1/T_0 \exp(-E_a/kT)$, $E_a \approx 0.49eV$, $T_0 = 10^{-8}$ and $E_0 \approx$0.25 V/nm are constants, $T_{\rm clk}$ is the time period of one stress-recovery cycle, ω is the duty cycle (the ratio of the time spent in stress to time period), ξ_1 and ξ_2 are two constants. n is a time exponent and equal to $1/6$ for an H_2 diffusion model, T_{ox} is the oxide thickness.

3) Power Model: Dynamic power is calculated as follows:

$$P_{\rm dyn} = \frac{1}{2} f \sum_{v=1}^{N} \alpha_v C_v V_{dd}^2 \tag{5}$$

where α_v is the switching probability of gate v, f is the clock frequency, N is the gate number in the circuit.

A leakage lookup table is created by simulating all the gates in the standard cell library, under all possible input patterns. Thus the leakage power of gate v can be expressed as

$$P_{lkg}(v) = \sum_{\rm input} P_{lkg}(v, {\rm input}) \times {\rm Prob}(v, {\rm input}) \tag{6}$$

where $P_{lkg}(v, {\rm input})$ and ${\rm Prob}(v, {\rm input})$ are the leakage power and the input signal probability of gate v when the input pattern is ${\rm input}$.

4) Variation Model: Many variations strongly affect the gate delay, such as threshold voltage (V_{th}), channel length $(L_{\rm eff})$, oxide thickness (T_{ox}), and so on. Since gate delay and leakage power both strongly depend on the threshold voltage. We simply consider the V_{th} variations for a fast and approximate statistical analysis, and assume that V_{th} variation is modeled by Gaussian distribution

$${\rm PDF}(V_{th}) = \frac{1}{\sqrt{2\pi}\sigma_{th}} e^{-1/2(V_{th}-\mu_{th}/\sigma_{th})^2} \tag{7}$$

where μ_{th} and σ_{th} are the mean value and standard deviation of V_{th}, respectively. $\sigma_{th}/\mu_{th} = 0.42$, which is obtained from ITRS 2007 [21].

According to (1), the gate delay can also be described as a Gaussian distribution. Then the statistical information of the circuit delay can be calculated using **ADD** and **MAX** operations [22] with Clark's formula [23].

Leakage current which is dominated by sub-threshold effect, depends on V_{gs} and V_{th} according to the sub-threshold leakage formulas [24]

$$I_{\rm sub}(v) = C \frac{W}{L} \mu_e V_T^2 e^{V_{gs}-V_{th}/nV_T} \tag{8}$$

where C is a constant, μ_e is the N-mobility, V_T is the thermal voltage, n is the sub-threshold swing parameter $(n < 3)$.

Gate leakage current (and power) is modeled as a log-normal distribution, so the statistical leakage power of the circuit can be calculated as follows:

$$\mu_{P_{lkg}} = \sum_{v=1}^{N} \mu_{P_{lkg}(v)} \tag{9}$$

$$\sigma^2_{P_{lkg}} = \sum_{v=1}^{N} \sigma^2_{P_{lkg}(v)} + 2 \sum_{u>v} \sigma_{P_{lkg}(u)} \rho_{uv} \sigma_{P_{lkg}(v)} \tag{10}$$

where ρ_{uv} denotes the correlation coefficient between gate u and v. The ρ_{uv}'s are calculated by the method in [25] and [26].

III. Supply Voltage Assignment (SVA) Technique

According to the NBTI model, V_{th} will increase due to the NBTI effect, thus leading to degradation of the circuit speed. Increasing V_{dd} is a feasible method to maintain the circuit performance. However, increasing V_{dd} will directly increase the leakage and dynamic power. We notice that it's not necessary to increase all gates' V_{dd}, because not all the gates are critical, the non-critical gates have positive slack to reduce V_{dd} so that it can reduce power consumption. So we propose to use dual V_{dd}, the high V_{dd} is used to compensate for NBTI-induced degradation on critical gates, while the low V_{dd} is used to reduce power on other gates. Our target is to reduce power as much as possible under an aging-aware timing constraint.

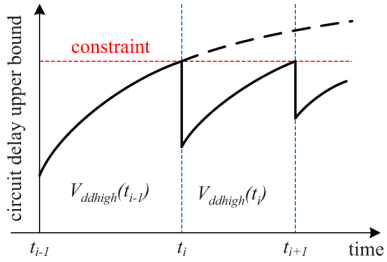

Fig. 1. Impact of V_{dd} scaling on NBTI-induced delay degradation.

There are two steps in our SVA technique: 1) Dual V_{dd} assignment (dual voltage islands partition): divide all the gates into two sets: *HVGS* (high V_{dd} gate set) and *LVGS* (low V_{dd} gate set); 2) Dynamic V_{dd} scaling: dynamically determine the optimal time nodes and the voltage values.

A. Dual V_{dd} Assignment

First, the nominal (nominal means the original circuit, without any NBTI mitigation and power reduction techniques) NBTI-aware delay slack at 10 year time node is calculated, then we determine two gate sets: *HVGS* (High V_{dd} Gate Set) and *LVGS* (Low V_{dd} Gate Set): if the slack of gate i (S_i) is smaller than a given threshold value (S_{th}), then gate i is called "NBTI-aware critical gate". All the NBTI-aware critical gates are included in *HVGS*. In addition, all the predecessors of NBTI-aware critical gates are also included in *HVGS*; *LVGS* is composed of all the rest gates. Once *HVGS* and *LVGS* are determined, they are fixed forever.

B. Dynamic V_{dd} Scaling

In the SVA technique, we set a timing constraint for each circuit. Since the statistical platform is used, delay upper bound (upper bound: $\mu + 3\sigma$, μ is the mean value, and σ is the standard deviation) is used instead of the absolute delay. The timing constraint is chosen as the nominal delay upper bound at a given time node (in this paper, this time node is 10 day) of each circuit. Once the circuit delay upper bound exceeds the constraint, the voltages need to be scaled (as shown in Fig. 1). This means the circuit delay upper bound will never exceed the constraint.

At each time node t_i ($t_0 = 0$, the determination of other t_i's will be described in the below), the mean value and standard deviation of circuit delay and power are calculated, then we determine the optimal voltages: $V_{ddhigh}(t_i)$ and $V_{ddlow}(t_i)$, which will be assigned in the following time interval $[t_i,\ t_{i+1})$. The *HVGS* gates are assigned $V_{ddhigh}(t_i)$, while *LVGS* gates are assigned $V_{ddlow}(t_i)$. The detailed determination of $V_{ddhigh}(t_i)$ and $V_{ddlow}(t_i)$ will be described in the below. Then the V_{th} degradation of each gate after t_i is calculated as the same as the method in [12], and we predict the next time node t_{i+1} at which the circuit delay upper bound will exceed the constraint, so the supply voltages need to be scaled again.

The same procedure including three operations: determine optimal $V_{ddhigh}(t_i)$, $V_{ddlow}(t_i)$, and predict the next time node t_{i+1}, will be repeated at each time node t_i, until the time node achieve the circuit lifetime.

1) Calculating $V_{ddhigh}(t_i)$: Fig. 1 shows how V_{dd} scaling improves the NBTI-induced delay degradation, the delay will be a sudden drop at t_i immediately after assigning a higher V_{dd} according to (1). However, higher V_{dd} directly leads to higher NBTI-induced V_{th} degradation ((2)–(4)) and higher power. Our target is to make the circuit delay and power at t_i achieve optimal values simultaneously. Considering the statistical model, our target is to minimize the following function:

$$F = A \times \left(\mu_{Delay(t_i)} + 3\sigma_{Delay(t_i)}\right) + B \times \left(\mu_{P_{lkg}(t_i)} + 3\sigma_{P_{lkg}(t_i)} + \mu_{P_{\mathrm{dyn}}(t_i)} + 3\sigma_{P_{\mathrm{dyn}}(t_i)}\right). \quad (11)$$

Then a serial search in $V_{ddhigh} \in [1\ \mathrm{V}, 1.4\ \mathrm{V}]$ is performed, with a resolution of 20 mV, to find an optimal $V_{ddhigh}(t_i)$ which leads to the minimum value of (11).

2) Calculating $V_{ddlow}(t_i)$: Consider a gate v in *LVGS*, it has positive delay slack, so its delay can be relaxed and then its V_{dd} could be lower

$$D_{\mathrm{relaxed}}(v) = D_{\mathrm{current}}(v) + R \times D_{\mathrm{slack}}(v) \quad (12)$$

where D_{relaxed}, D_{current}, D_{slack} are the relaxed delay, the current delay and the delay slack of gate v, respectively. R is a constant between 0 and 1 to make sure that the gates in *LVGS* will not change to be critical. According to (1), the new V_{dd} of gate v can be calculated using Newton iteration method. V_{ddlow} of the whole circuit is the maximum supply voltage of the gates in *LVGS* to make sure that each gate in *LVGS* will satisfy the delay requirement.

3) Predicting the Next Time Node: Cao *et al.* pointed out in [27] that the mean value and standard deviation of the circuit delay under both NBTI effect and process variation can be written as

$$\begin{cases} \mu_{\mathrm{Delay}}(t) = & \mu_{\mathrm{Delay}}(0)\left(1 + \zeta_{ti} t^{1/6}\right) \\ \sigma_{\mathrm{Delay}}(t) = & \sigma_{\mathrm{Delay}}(0)\left(1 - \zeta_v t^{1/6}\right) \end{cases} \quad (13)$$

where ζ_{ti} and ζ_v are two constants determined by circuit and variation parameters.

At time node t_i, once the new $V_{ddhigh}(t_i)$ and $V_{ddlow}(t_i)$ are calculated, we use (13) to calculate t_{i+1} according to the timing constraint.

C. Overhead of the SVA Technique

In the traditional dual voltage design, the high and low voltage cells must be isolated by level converters, which introduce extra overheads (power, area, delay, etc.). To avoid these overheads, we propose that **in topological order, the critical gates and all their predecessors are included in *HVGS***. An example is shown in Fig. 2. By using this method, one can simply partition a circuit into two voltage islands without level converters, and place the *HVGS* and *LVGS* cells into two clusters. This method is similar to the Clustered-Voltage-Scaling (CVS) structure proposed in [28]. The researchers of Toshiba [29] have explored the methods of circuit synthesis, placement, routing, and clock tree generation for dual V_{dd} design.

Zhang *et al.* proposed an n-bit digital-to-analog converter (DAC) in [12] with the resolution of 1.37 mV, which is sufficient to implement dynamic voltage scaling. DAC will increase area and power, however, these overheads are small, so the detailed analysis is omitted here.

IV. Simulation Results

Our experiments are implemented by C++ on a PC with Intel Q9550 CPU and 4 GB RAM. Some key parameters are: nominal $|V_{dd}| = 1.0$ V; $|V_{th}| = 0.20$ V for both nMOS and pMOS transistors; $T_{ox} = 1.2$ nm; $T = 378$ K. ISCAS85, ISCAS89 benchmarks and some ALU circuits are synthesized using a 65 nm low power library from industry, and used to evaluate our technique. The circuit lifetime (T_{life}) is set to be 10 years. The timing constraint of each circuit is chosen as the nominal (nominal: the case without any NBTI and power reduction techniques) delay upper bound at 10 day time node, in other words, under this constraint, the circuit performance will always keep the fresh state within 10 days.

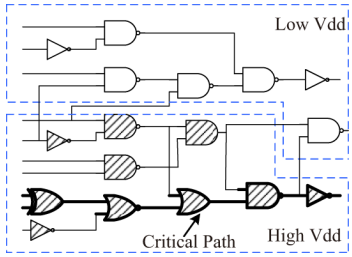

Fig. 2. Our proposed dual voltage islands partition method.

Table I shows the results of our SVA technique and the comparison between our approach and guard-banding approach/single V_{dd} scaling approach. All these approaches are implemented at the same voltage resolution: 20 mV. The power presented in the following is the **average power during the overall circuit lifetime, which is proportional to the energy consumption**.

For guard-banding approach, a fixed supply voltage is set to each circuit, to ensure that the circuit delay upper bound will never exceed the constraint. Since NBTI effect degrades the circuit continuously, the circuit delay at T_{life} should be guaranteed. The results of guard-banding method are shown in the three sub-columns in "Guard-banding" column. Guard-banding approach increases 38% of the average leakage power and 17% of the average dynamic power, to guarantee the timing constraint.

The single V_{dd} scaling approach is similar with our SVA technique, the only difference is that it only uses one supply voltage, which is scaled higher to compensate for the NBTI-induced degradation. This approach can mitigate 63% of the NBTI-induced degradation, with 15% average leakage power increase and 15% average dynamic power increase. Compared with the guard-banding approach, it can save more energy.

The results of our SVA technique are shown in the last 6 columns. Our approach can mitigate on average 62% of the NBTI-induced degradation, while the average leakage power decreases by 12.8% and the average dynamic power decreases by 1.8%. Compared with guard-banding approach, our SVA technique can save 50% more of the average leakage power and 18% more of the average dynamic power; while compared with single V_{dd} scaling technique, it can save 27% more leakage and 16% more dynamic power, respectively.

As pointed out in [14], for dynamic voltage tuning approaches, the rate of NBTI degradation is rapid in the early lifetime and slows down under continued stress. As a result, the supply voltage increases quickly during the early lifetime, so the power benefits are relevant only in the early lifetime, while they swiftly degrade afterward. However, as shown in the comparisons, by using an additional low voltage to reduce power on non-critical gates, **our approach can effectively save the energy consumption, compared with guard-banding or single V_{dd} scaling techniques**.

The comparisons indicate that out SVA technique is more effective and flexible, because we explore the available slack as much as possible, and the optimal voltage values and the time nodes are dynamically decided during circuit operation. The comparisons also well support our motivation. Guard-banding techniques lead to larger energy

TABLE I
RESULTS OF OUR SVA TECHNIQUE AND THE COMPARISONS

Benchmark	$Gate\#$	Nominal design				Guard-banding			Single V_{dd} scaling approach				Our approach					
		D_0 (ns)	D_{Tlife} (ns)	P_{lkg} (μW)	P_{dyn} (μW)	D_{imp} (%)	$\frac{\Delta P_{lkg}}{P_{lkg}}$ (%)	$\frac{\Delta P_{dyn}}{P_{dyn}}$ (%)	D_{imp} (%)	$\frac{\Delta P_{lkg}}{P_{lkg}}$ (%)	$\frac{\Delta P_{dyn}}{P_{dyn}}$ (%)	$Lv\#$	$HVGS\#$	$LVGS\#$	D_{imp} (%)	$\frac{\Delta P_{lkg}}{P_{lkg}}$ (%)	$\frac{\Delta P_{dyn}}{P_{dyn}}$ (%)	$Lv\#$
array4x4	89	2.7	2.8	0.10	0.66	68	39	17	63	15	15	3	76	13	63	-8	8	7
pmult4x4	122	3.4	3.6	0.13	0.81	81	38	17	61	9	12	3	105	17	61	-7	9	7
bkung16	130	2.1	2.2	0.16	0.75	68	38	17	63	15	15	3	70	60	63	-16	2	3
c499	182	2.2	2.3	0.26	1.36	79	38	17	63	14	14	3	174	8	59	6	8	7
kogge16	199	1.4	1.5	0.22	1.20	69	38	17	60	11	13	3	175	24	60	4	12	6
log16	256	2.1	2.2	0.24	0.94	77	38	17	63	10	12	4	188	68	63	5	8	2
bkung32	271	2.6	2.8	0.31	1.60	66	38	17	68	19	17	3	148	123	68	-19	0	3
c432	297	5.2	5.5	0.27	1.57	72	38	17	76	20	17	4	280	17	74	13	12	3
array8x8	401	6.5	6.9	0.44	3.17	64	38	17	62	22	19	3	372	29	59	1	11	7
kogge32	487	1.7	1.8	0.51	3.00	67	38	17	60	12	13	4	439	48	60	3	10	7
pmult8x8	490	6.5	6.9	0.50	3.59	73	38	17	62	10	12	4	449	41	62	-1	9	6
c880	535	3.5	3.8	0.43	2.38	69	38	17	62	17	15	4	110	425	62	-40	-27	5
log32	620	3.9	4.1	0.59	2.40	76	38	17	63	8	11	5	572	68	60	4	7	6
c1355	942	4.0	4.3	0.69	4.52	75	38	17	62	15	14	4	870	72	48	12	11	6
c1908	977	4.8	5.1	0.72	4.78	73	38	17	62	11	12	5	535	442	62	-16	-3	4
c2670	1173	5.0	5.3	0.92	5.88	70	38	17	63	23	18	3	447	726	63	-30	-14	3
booth9x9	1206	5.2	5.6	1.19	7.00	68	39	17	63	23	18	3	1052	154	63	13	15	4
log64	1536	7.8	8.3	1.39	5.84	75	38	17	63	14	14	3	996	540	63	2	7	4
c3540	1743	6.3	6.7	1.31	8.93	69	38	17	63	23	18	3	797	946	63	-20	-9	4
pmult16x16	1934	13.4	14.3	1.98	15.01	68	38	17	62	22	18	3	1845	89	59	6	12	7
c5315	2364	6.5	6.9	1.86	12.17	68	38	17	62	17	15	4	629	1735	58	-55	-40	7
c7552	3912	6.7	7.1	2.98	21.22	70	38	17	63	12	12	5	693	3219	63	-49	-29	4
s38584	19912	7.1	7.6	21.99	112.82	69	38	17	63	8	13	4	886	19026	62	-66	-42	5
s35932	21400	3.8	4.1	22.67	103.87	74	38	17	62	5	12	5	15703	5697	62	-21	-4	5
s38417	22892	7.2	7.7	21.98	118.13	74	38	17	62	15	15	3	4488	18404	62	-41	-19	3
Average						71	38	17	63	15	15	3.6			62	-12.8	-1.8	5.0

D_0 and D_{Tlife} are the nominal circuit delay upper bound at time 0 and T_{life} respectively
P_{lkg} and P_{dyn} are the average leakage power and average dynamic power during the overall lifetime in nominal case
D_{imp} is the delay improvement percentage
ΔP_{lkg} and ΔP_{dyn} are the average leakage power increase and average dynamic power increase caused by the corresponding techniques
$Lv\#$ is the number of voltage levels

increase, dynamic techniques also increase more energy when compensating for aging without taking power into account, compared with our approach.

The delay and energy improvements of different circuits are different, which is mainly caused by the *HVGS/LVGS* partition. The *HVGS/LVGS* partition depends on the circuit structure. From Table I, when there are more gates in *LVGS*, our SVA technique can reduce more energy (for example, c880, s38584, s38417, etc). For some balanced circuits (the delay of most paths is equal, it also means most of the paths and gates are critical), *LVGS* only contains few non-critical gates, so our SVA technique will lead to more energy increase (for example, c432, c1355, booth9x9, etc). Therefore, according to the circuit structure and the value of $HVGS\#/LVGS\#$, we can approximatively predict the effect of our SVA technique.

For the balanced circuits, such as c432, c1355, booth9x9, our SVA technique can still reduce energy compared with the single V_{dd} scaling method, but the improvement is small [for these three circuits, the average power improvement is less than 10%, and the average improvement of all the benchmarks is 27.8% (average leakage power) and 16.8% (average dynamic power)]. Because the dual V_{dd} assignment in our method, reduces power by the sacrifice of some positive slacks on non-critical paths, which is the same as some traditional techniques (such as multi V_{th} [30]). However, in practice, there may not be many perfectly balanced circuits, because they are dangerous under PVT (process, voltage, temperature) variations. Our method can be utilized on the circuit which is not perfectly balanced (even if it's partly balanced). In addition, in this paper, the SVA technique is implemented on gate level, it can also be implemented on module level.

The runtime of our SVA technique ranges from 0.07 to 11.3 s for all the benchmarks, the average runtime is 1.33 s. The computational complexity of our SVA technique is about $O(N)$, where N is the gate number in a circuit.

V. CONCLUSION

Power and reliability have become two key design goals with technology scales. In this paper, a SVA technique combining dual V_{dd} assignment and dynamic V_{dd} scaling is proposed on a statistical platform, to minimize NBTI-induced performance degradation and circuit power consumption. It saves 62% of the NBTI-induced circuit delay degradation. Compared with single V_{dd} scaling and guard-banding techniques, our approach can save more energy. Furthermore, our SVA technique is more flexible, since the optimal results are dynamically decided for any circuit, so the circuit delay exactly meets the performance constraint during circuit operation.

REFERENCES

[1] V. Huard, M. Denais, and C. Parthasarathy, "NBTI degradation: From physical mechanisms to modelling," *Microelectron. Reliab.*, vol. 46, no. 1, pp. 1–23, 2006.
[2] C. Forzan and D. Pandini, "Why we need statistical static timing analysis," in *Proc. ICCD*, 2007, pp. 91–96.
[3] S. V. Kumar, C. H. Kim, and S. S. Sapatnekar, "NBTI-aware synthesis of digital circuits," in *Proc. DAC*, 2007, pp. 370–375.
[4] K. Kang, H. Kufluoglu, M. A. Alain, and K. A. Roy, "Efficient transistor-level sizing technique under temporal performance degradation due to NBTI," in *Proc. ICCD*, 2006, pp. 216–221.

[5] B. Paul, K. Kang, H. Kufluoglu, M. Alam, and K. Roy, "Temporal performance degradation under NBTI: Estimation and design for improved reliability of nanoscale circuits," in *Proc. DATE*, 2006, pp. 1–6.
[6] Y. Wang, X. Chen, W. Wang, V. Balakrishnan, Y. Cao, Y. Xie, and H. Yang, "On the efficacy of input vector control to mitigate NBTI effects and leakage power," in *Proc. ISQED*, 2009, pp. 19–26.
[7] Y. Wang, H. Luo, K. He, R. Luo, H. Yang, and Y. Xie, "Temperature-aware NBTI modeling and the impact of input vector control on performance degradation," in *Proc. DATE*, 2007, pp. 546–551.
[8] J. Abella, X. Vera, and A. Gonzalez, "Penelope: The NBTI-aware processor," in *Proc. MICRO*, 2007, pp. 85–96.
[9] Y. Wang, X. Chen, W. Wang, Y. Cao, Y. Xie, and H. Yang, "Leakage power and circuit aging cooptimization by gate replacement techniques," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 19, no. 4, pp. 615–628, Apr. 2011.
[10] D. Bild, G. Bok, and R. Dick, "Minimization of NBTI performance degradation using internal node control," in *Proc. DATE*, 2009, pp. 148–153.
[11] S. Kumar, C. Kim, and S. Sapatnekar, "Adaptive techniques for overcoming performance degradation due to aging in cmos circuits," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 19, no. 4, pp. 603–614, Apr. 2011.
[12] L. Zhang and R. Dick, "Scheduled voltage scaling for increasing lifetime in the presence of NBTI," in *Proc. ASP-DAC*, 2009, pp. 492–497.
[13] E. Mintarno, J. Skaf, R. Zheng, J. Velamala, Y. Cao, S. Boyd, R. Dutton, and S. Mitra, "Self-tuning for maximized lifetime energy-efficiency in the presence of circuit aging," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 30, no. 5, pp. 760–773, May 2011.
[14] T.-B. Chan, J. Sartori, P. Gupta, and R. Kumar, "On the efficacy of nbti mitigation techniques," in *Proc. DATE*, Mar. 2011, pp. 1–6.
[15] T. Sakurai and A. Newton, "Alpha-power law mosfet model and its applications to cmos inverter delay and other formulas," *J. Solid-State Circuits*, vol. 25, no. 4, pp. 584–594, Apr. 1990.
[16] S. Mahapatra, P. Kumar, and M. Alam, "Investigation and modeling of interface and bulk trap generation during negative bias temperature instability of p-mosfets," *IEEE Trans. Electron Devices*, vol. 51, no. 9, pp. 1371–1379, Sep. 2004.
[17] M. Alam, "A critical examination of the mechanics of dynamic NBTI for PMOSFETS," in *IEDM Tech. Dig.*, 2003, pp. 14.4.1–14.4.4.
[18] R. Vattikonda, W. Wang, and Y. Cao, "Modeling and minimization of PMOS NBTI effect for robust nanometer design," in *Proc. DAC*, 2006, pp. 1047–1052.
[19] S. Bhardwaj, W. Wang, R. Vattikonda, Y. Cao, and S. A. Vrudhula, "Predictive modeling of the NBTI effect for reliable design," in *Proc. CICC*, 2006, pp. 189–192.
[20] W. Wang, V. Reddy, A. Krishnan, R. Vattikonda, S. Krishnan, and Y. Cao, "Compact modeling and simulation of circuit reliability for 65-nm CMOS technology," *IEEE Trans. Device Mater. Reliab.*, vol. 7, no. 4, pp. 509–517, Dec. 2007.
[21] "International Technology Roadmap for Semiconductors," 2007. [Online]. Available: http://public.itrs.net
[22] C. Amin, N. Menezes, K. Killpack, F. Dartu, U. Choudhury, N. Hakim, and Y. Ismail, "Statistical static timing analysis: How simple can we get?," in *Proc. DAC*, 2005, pp. 652–657.
[23] C. E. Clark, "The greatest of a finite set of random variables," *Oper. Res.*, vol. 9, no. 2, pp. 145–162, Mar.-Apr. 1961.
[24] A. Srivastava, D. Sylvester, and D. Blaauw, *Statistical Analysis and Optimization for VLSI: Timing and Power*, no. 1 ed. New York: Springer, 2005.
[25] X. Shen, Y. Wang, R. Luo, and H. Yang, "Leakage power reduction through dual Vth assignment considering threshold voltage variation," in *Proc. ASICON*, 2007, pp. 1122–1125.
[26] X. Chen, Y. Wang, Y. Cao, Y. Ma, and H. Yang, "Variation-aware supply voltage assignment for minimizing circuit degradation and leakage," in *Proc. ISLPED*, 2009, pp. 39–44.
[27] W. Wang, V. Reddy, B. Yang, V. Balakrishnan, S. Krishnan, and Y. Cao, "Statistical prediction of circuit aging under process variations," in *Proc. CICC*, 2008, pp. 13–16, "," *'08*.
[28] K. Usami and M. Horowitz, "Clustered voltage scaling technique for low-power design," in *Proc. ISLPD*, 1995, pp. 3–8.
[29] K. Usami and M. Igarashi, "Low-power design methodology and applications utilizing dual supply voltages," in *Proc. ASP-DAC*, 2000, pp. 123–128.
[30] L. Wei, Z. Chen, K. Roy, M. Johnson, Y. Ye, and V. De, "Design and optimization of dual-threshold circuits for low-voltage low-power applications," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 7, no. 1, pp. 16–24, Mar. 1999.

Direct Compare of Information Coded With Error-Correcting Codes

Wei Wu, Dinesh Somasekhar, and Shih-Lien Lu

***Abstract*—There are situations in a computing system where incoming information needs to be compared with a piece of stored data to locate the matching entry, e.g., cache tag array lookup and translation look-aside buffer matching. If the stored data is protected with error-correcting codes (ECC) for reliability reason, the previous solution is to access the stored information, decode and correct if necessary before it is used to compare with the incoming data. The decoding and correcting step increases the total access time, which is often critical. In this paper, we propose a method to improve the compare latency for information encoded with ECC. We use the cache tag array look-up as an example, and results show that 30% gate count reduction and 12% latency reduction are achieved.**

***Index Terms*—Data comparison, error correcting code (ECC), Hamming distance, set associative cache, tag array lookup.**

I. Introduction

Data comparison circuit is a logic that has many applications in a computing system. For example, to check whether a piece of information is in a cache, the address of the information in the memory is compared to all cache tags in the same set that might contain that address. Another place that uses a data comparison circuit is in the translation look-aside buffer (TLB) unit. TLB is used to speed up virtual to physical address translation.

Error correcting codes (ECC) are widely used in modern microprocessors to enhance the reliability and data integrity of their memory structures. For example, caches on modern microprocessors are protected by ECC [1]–[3]. If a memory structure is protected with ECC, a piece of data is encoded first and the entire codeword including the ECC check bits are written into the memory array. When the codeword is loaded from memory, it has to be decoded and corrected if errors are detected to obtain the original data.

Data comparison circuit is usually in the critical path of a pipeline stage because the result of the comparison determines the flow of the succeeding operations. When the memory array is protected by ECC, it exacerbates the criticality because of the added latency due to ECC logic. In the cache tag match example, the cache tag directory must be accessed first. After the tag information is retrieved, it must go through

Manuscript received April 28, 2011; revised August 04, 2011; accepted September 13, 2011. Date of publication October 17, 2011; date of current version July 27, 2012.

S.-L. Lu and W. Wu are with the Circuit and System Research, Intel Labs, Hillsboro, OR 97124 USA (e-mail: shih-lien.l.lu@intel.com; wei.a.wu@intel.com).

D. Somasekhar was with the Intel Labs, Hillsboro, OR 97124 USA. He is now with Global Foundries, Sunnyvale, CA 94085 USA. (e-mail: dinesh.somasekhar@globalfoundries.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVLSI.2011.2169094

1063-8210/$26.00 © 2011 IEEE