

Statistical Timing Modeling Based on a Lognormal Distribution Model for Near-Threshold Circuit Optimization*

Jun SHIOMI^{†a)}, Student Member, Tohru ISHIHARA^{†b)}, Member, and Hidetoshi ONODERA^{†c)}, Fellow

SUMMARY Near-threshold computing has emerged as one of the most promising solutions for enabling highly energy efficient and high performance computation of microprocessors. This paper proposes architecture-level statistical static timing analysis (SSTA) models for the near-threshold voltage computing where the path delay distribution is approximated as a lognormal distribution. First, we prove several important theorems that help consider architectural design strategies for high performance and energy efficient near-threshold computing. After that, we show the numerical experiments with Monte Carlo simulations using a commercial 28 nm process technology model and demonstrate that the properties presented in the theorems hold for the practical near-threshold logic circuits.

key words: near-threshold computing, statistical static timing analysis (SSTA)

1. Introduction

The internet of things (IoT), or it is sometimes called machine-to-machine communication (M2M), has emerged as a new concept in which billions of computers and sensor devices are interconnected each other, enabling an autonomous exchange of information. It requires SoCs on the wireless sensor devices to handle much bigger and far more complex multimedia data than ever before. Both high computational-power and high energy-efficiency are thus extremely crucial to the digital processors embedded in these SoCs. Traditionally, the wireless sensor devices employ sub-threshold logic circuits that operate with the power supply voltage less than the transistors' threshold voltage. They will be suitable for specific applications which do not need high performance, but require extremely low power consumptions. For the IoT applications, however, the computational-power of the sub-threshold logic circuits is no longer satisfiable. As a solution to this issue, a concept of near-threshold computing has emerged [2]. It scales the supply voltage to the near-threshold region, which brings not only quadratic dynamic energy savings, but also super-linearly reduced leakage energy, with keeping the performance degradation of the circuit to a minimum.

This paper, for the first time, shows architectural-level models for Within Die (WID) variability in the near-

threshold voltage (denoted *NTV* in the following) operation. In the last fifteen years, techniques for statistical static timing analysis (SSTA) for the super-threshold voltage (denoted *STV* in the following) operation have been widely studied. They assume that the delay distributions of the circuits follow the Gaussian distribution. This makes it easier to perform the SUM and the MAX operations for estimating the timing yield of the targeting circuit. Recent literature [3] revealed that the delay distributions in *NTV* operation do not follow the Gaussian distribution, but follow lognormal distribution. Based on this fact, this paper proves several interesting theorems that help consider architectural design strategies for high performance and energy efficient near-threshold computing. Since decisions taken at the architectural-level have a significant impact on both performance and the energy consumption, the architectural-level model that gives insight into the impact of near-threshold timing variation on both performance and the energy consumption is crucial to the processor design.

This paper is an extension of our previous work [1]. This paper newly introduces several properties which enhance theorems presented in [1]. Discussing the impact of gate sizing on the performance improvement in near-threshold operation and showing an application of the theorems are also key enhancements of this paper. The minimum size transistor suffers from drastic performance degradation when the supply voltage is scaled down to near-threshold region. This paper analytically shows that gate upsizing exponentially improves the performance of a transistor in *NTV* operation while it just linearly improves the performance of a transistor in *STV* operation.

The rest of this paper is organized in the following way. Section 2 describes related work and contributions of this work. Section 3 explains fundamental characteristics of near-threshold voltage operation. Several important theorems and their proofs are presented in Sect. 4. Section 5 shows experimental results which demonstrate that the theorems proved in Sect. 4 hold for the actual *NTV* operation. Section 6 discusses how the theorems and corollaries presented in this paper can be effectively exploited in an architectural design phase. Section 7 concludes this paper.

2. Related Work and Contributions of This Work

2.1 Related Work

Early studies that related variability to architecture were by

Manuscript received September 17, 2014.

Manuscript revised January 12, 2015.

[†]The authors are with the Department of Communications and Computer Engineering, Kyoto University, Kyoto-shi, 606-8501 Japan.

*This work was partly presented in the conference [1].

a) E-mail: shiomi-jun@vlsi.kuee.kyoto-u.ac.jp

b) E-mail: ishihara@i.kyoto-u.ac.jp

c) E-mail: onodera@i.kyoto-u.ac.jp

DOI: 10.1587/transfun.E98.A.1455

Bowman et al. [4]–[6], which presented a statistical predictive model for the distribution of the maximum operating frequency (Fmax) for a chip in the presence of process variations. The model provides insight into the impact of different components of variations on the distribution of Fmax. The within-die (WID) delay distribution depends on the total number of independent critical paths (N_{cp}) for the entire chip. For a larger number of critical paths, the mean value of the maximum critical path delay increases. As the number of critical paths increases, the probability that one of them will be strongly affected by process variations becomes higher, and therefore, the mean of critical path delay also increases. On the other hand, the standard deviation (or delay spread) decreases with larger N_{cp} . Another factor that affects the delay distribution is the logic depth per critical path. Random WID variations have an *averaging effect* on the overall critical path distribution, which reduces the relative delay variation $\frac{\sigma}{\mu}$ [5]. If the mean and the standard deviation of a single gate are μ and σ , respectively, the standard deviation of a path comprising the serially connected identical n gates is $\sqrt{n}\sigma$. Since the mean delay of the path is $n\mu$, the relative delay variation is proportional to $\frac{1}{\sqrt{n}}$ [6].

The Non-Gaussian delay distributions as well as the correlations among delays make statistical timing analysis more complicated. Several previous techniques present efficient statistical timing analysis approaches which can accurately predict Non-Gaussian delay distributions from realistic nonlinear gate and interconnect delay models [7]–[9]. All of those techniques are aiming at accurately reflecting the Non-Gaussianity when performing the two atomic operations of SSTA, SUM and MAX. None of them explicitly discusses the averaging effect or the effect of parallel critical paths on lognormal delay distributions. Unlike those previous work, this paper proves several important theorems that provide architectural insight into the impact of timing variation in NTV operation on the performance of a target circuit.

A statistical leakage power distribution is also known as a Non-Gaussian distribution. Chang et al. [10] showed a statistical prediction method for leakage power distribution. They approximate each leakage component as a spatially correlated lognormal distribution and calculate the total leakage by summing up all the leakage components. In their work, they assume that the sum of lognormal random variables (RVs) is also approximated as a lognormal RV. We use this approximation for deriving theorems in this paper. Recent literature [3] showed that the delay distributions in NTV operation can be approximated as lognormal distribution. However, none of them discusses the averaging effect on lognormal delay distributions. Unlike those previous work, this paper numerically shows that a stronger averaging effect is observed in NTV operation than that observed in the STV operation. More specifically, this paper, for the first time, shows that the distance between the median and the specific high- σ worst case of n -stage logic path delay distribution in NTV operation is approximately proportional to values ranging from $0.3\sqrt{n}$ to $0.7\sqrt{n}$ which are much less

than \sqrt{n} if the delays of individual logic gates are all lognormal independent RVs. Note that the distance between the median and the $k\sigma$ worst case of n -stage logic path delay in the STV operation is proportional to \sqrt{n} if the delays of individual gates are all normal independent RVs [6].

2.2 Contributions of This Work

The contributions of this work are summarized below.

- We give conditions for logic paths operated with NTV, with which the delay distributions of the paths can be exactly fit to a lognormal distribution. We also show that the conditions given here are practical in NTV operation. Although actual delay distributions in NTV operation may not exactly follow the lognormal distribution basically, it can be closely fit to a lognormal distribution if we limit the ranges of several parameter values. These conditions are summarized in Lemma1.
- We give an analytical model which provides μ and σ for a lognormal delay distribution of an n -stage logic path in NTV operation. The properties of the model are summarized in Theorem1. With this model, we can numerically see how the path delay distributions in NTV operation change as the logic depth n of the path increases or decreases.
- We analytically show the *averaging effect* in NTV operation. This property is summarized in Corollary1. This demonstrates that the averaging effect observed in NTV operation is stronger than that observed in STV operation.
- It is well known that the maximum operating frequency (Fmax) of a circuit needs to be lowered as the number of critical paths (N_{cp}) increases to maintain the same timing yield [5]. This paper, for the first time, shows that the speed of the Fmax degradation in NTV operation is exponential to that in STV operation. However, the degradation speed is still slow even in NTV operation. These properties are summarized in Theorem2 and Corollary2.
- We analytically show that gate upsizing brings exponential performance improvement in NTV operation compared with that in STV operation. This property is summarized in Corollary3.

3. Preliminaries

3.1 Statistical Static Timing Analysis

Statistical Static Timing Analysis (SSTA) is a popular solution for timing analysis of digital circuits. A typical approach of the SSTA is to perform SUM and MAX operations for the delay probability density functions (PDFs) of individual gates through the circuit and estimate the timing yield of the circuit. Although the techniques of SSTA for the circuits operated with STV have been well studied over the last 15 years, those for NTV operation have not

been sufficiently investigated. In this paper, we focus on the architectural-level statistical timing analysis for the circuits operated with NTV.

For discussing the statistical timing, the concept of *timing yield* is indispensable. The timing yield is defined as the probability that the critical path delay of the circuit is no more than a specific delay x_k considering variation. If the x_k corresponds to a delay in the $k\sigma$ worst case condition, the timing yield for the x_k in a normal distribution can be calculated using the cumulative distribution function (CDF) of the circuit delay in the following way:

$$\int_{-\infty}^{x_k} f(x)dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^k \exp\left(-\frac{x^2}{2}\right)dx = \Phi(k), \quad (1)$$

where $\Phi(x)$ is a CDF of a standard normal distribution and $\exp(x) = e^x$.

3.2 Properties of Lognormal Distribution

It is well known that if an RV X has a normal distribution $N(\mu, \sigma^2)$, its PDF $f(x)$ is represented as

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \quad (2)$$

Suppose we have an RV Y which is represented as $Y = \exp(X)$. Then Y is a lognormal distribution function $LN(\mu, \sigma^2)$. Its PDF $g(y)$ is represented as

$$g(y) = \frac{1}{y\sqrt{2\pi}\sigma} \exp\left(-\frac{(\ln y - \mu)^2}{2\sigma^2}\right). \quad (3)$$

Note that μ and σ^2 in (3) do not correspond to a mean and a variance of Y , respectively. They correspond to a mean and a variance of $\ln(Y)$. The shape of $g(y)$ is asymmetric unlike a normal distribution. The CDF of a lognormally distributed RV Y can be formulated using $\Phi(x)$ in (1) as

$$(\text{CDF of } Y) = \Phi\left(\frac{\ln y - \mu}{\sigma}\right). \quad (4)$$

3.3 Delay Distribution for NTV Operation

Alpha power law model (Sakurai et al. [11]) is commonly used for representing MOSFET's ON current. This model is accurate for the operating voltage which is sufficiently higher than the threshold voltage. We refer to this operating voltage as *STV* (*super-threshold voltage*). Recently, Keller et al. [3] pointed out that NTV operation needs another model for accurately representing the transistor ON current in the near-threshold region as follows:

$$I_{\text{on}} = I_0 k_0 \exp\left(k_1 \frac{V_{\text{DT}}}{n_i \phi_t} + k_2 \left(\frac{V_{\text{DT}}}{n_i \phi_t}\right)^2\right), \quad (5)$$

where $V_{\text{DT}} = V_{\text{DD}} - V_{\text{th}}$, k_0 , k_1 and k_2 are fitting coefficients. ϕ_t is the thermal voltage and n_i is an ideal factor of MOSFET. With a simple linear RC-delay model, the delay of a

gate can be approximated as:

$$t_{\text{pd}} = k_f C_{\text{load}} \frac{V_{\text{DD}}}{I_{\text{on}}} = \alpha \exp\left(\frac{\Delta V_{\text{th}}}{V_0} - k_2 \left(\frac{\Delta V_{\text{th}}}{n_i \phi_t}\right)^2\right), \quad (6)$$

where $\Delta V_{\text{th}} = V_{\text{th}} - V_{\text{th0}}$ which represents a threshold voltage variation and α is a constant, and V_0 is $\left(\frac{k_1}{n_i \phi_t} + 2k_2 \frac{V_{\text{DD}} - V_{\text{th0}}}{(n_i \phi_t)^2}\right)^{-1}$.

Lemma1

If $k_2 \left(\frac{\Delta V_{\text{th}}}{n_i \phi_t}\right)^2$ is sufficiently smaller than $\frac{\Delta V_{\text{th}}}{V_0}$ meaning that $k_2 \left(\frac{\Delta V_{\text{th}}}{n_i \phi_t}\right)^2$ can be ignored in comparison with $\frac{\Delta V_{\text{th}}}{V_0}$, t_{pd} can be exactly fit to a lognormal distribution function.

(pf.) From the assumption, $k_2 \left(\frac{\Delta V_{\text{th}}}{n_i \phi_t}\right)^2$ in (6) can be ignored in comparison with $\frac{\Delta V_{\text{th}}}{V_0}$. Therefore, we obtain

$$t_{\text{pd}} = D \sim \alpha \exp\left(\frac{\Delta V_{\text{th}}}{V_0}\right) = D_0 \exp\left(\frac{V_{\text{th}}}{V_0}\right), \quad (7)$$

where we rename t_{pd} to D , and $D_0 = \alpha \exp\left(-\frac{V_{\text{th0}}}{V_0}\right)$ for simplicity. This work assumes that the threshold voltage (V_{th}) is only a variable that represents all impacts of device parameter variations on delay variation. It also assumes that the V_{th} is a normally distributed RV with mean μ' and variance σ'^2 (i.e., $N(\mu', \sigma'^2)$). The PDF $f(v)$ of the V_{th} is given as

$$f(v)dv = \frac{1}{\sqrt{2\pi}\sigma'} \exp\left(-\frac{(v-\mu')^2}{2\sigma'^2}\right)dv. \quad (8)$$

Since $dV_{\text{th}} = V_0 \frac{dD}{D}$, delay's PDF $g(d)$ is obtained as

$$g(d)dd = \frac{V_0}{d\sqrt{2\pi}\sigma'} \exp\left(-\frac{\left(\ln\left(\frac{d}{D_0}\right) - \frac{\mu'}{V_0}\right)^2}{2\left(\frac{\sigma'}{V_0}\right)^2}\right)dd, \quad (9)$$

which means that a variable $\frac{D}{D_0}$ has a lognormal distribution $LN\left(\frac{\mu'}{V_0}, \left(\frac{\sigma'}{V_0}\right)^2\right)$. (q.e.d.)

If we rename $\frac{\mu'}{V_0}$ to μ and $\frac{\sigma'}{V_0}$ to σ , the variable t_{pd} can be represented as a lognormal distribution $LN(\mu, \sigma^2)$.

In (6), ϕ_t and n_i are the thermal voltage and an ideal factor of MOSFET, respectively and typical values for them are 26 mV and within a range of 1.3 to 1.7, respectively in a room temperature. Figure 1 shows an approximation error introduced by ignoring $k_2 \left(\frac{\Delta V_{\text{th}}}{n_i \phi_t}\right)^2$ in (6). The approximation error is superlinear to ΔV_{th} . However, based on a model fitting result for our target process technology, if $\Delta V_{\text{th}} = 50$ mV, an approximation error for t_{pd} introduced by ignoring $k_2 \left(\frac{\Delta V_{\text{th}}}{n_i \phi_t}\right)^2$ in (6) is less than 7% which is reasonably small. If we use the values reported in [3] for the parameters k_1 and k_2 , the approximated error is less than 8%. Note that the error depends on ΔV_{th} (i.e., threshold voltage variation). If ΔV_{th} increases, the error also increases. Therefore, the assumption stated in Lemma1 is feasible for a reasonable ΔV_{th} such as $\Delta V_{\text{th}} < 50$ mV.

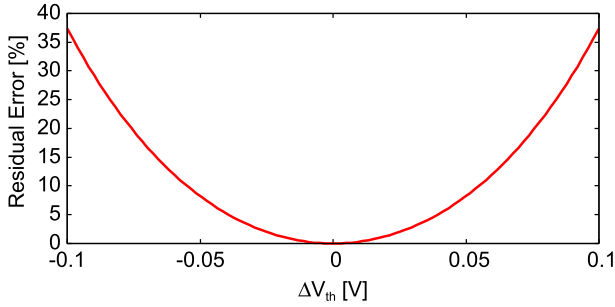


Fig. 1 An approximation error for t_{pd} .

4. Near-Threshold Statistical Timing Model

It is common that the delay distributions of the circuits follow the Gaussian distribution in STV operation, which makes easy to perform two atomic operations, SUM and MAX operations. In this section, we analytically derive SUM and MAX operations for the lognormal delay distribution, which can more accurately represent the delay in NTV operation than the Gaussian delay distribution. We then, derive several interesting properties from these two SUM and MAX operations.

4.1 SUM Operation for Lognormal Distributions

Generally, the distribution of the sum of lognormal RVs does not have the closed form. However, the sum of lognormal RVs can be reasonably approximated as a lognormal RV. Let L be the sum of n correlated lognormal RVs (L_1, L_2, \dots, L_n),

$$L = \sum_{i=1}^n L_i = \sum_{i=1}^n \exp(X_i) \sim \exp(Z), \quad (10)$$

where X_i and Z are normally distributed correlated RVs as follows:

$$X_i \sim N(\mu_i, \sigma_i^2), \quad Z \sim N(\mu_Z, \sigma_Z^2). \quad (11)$$

The correlation coefficient of X_i and X_j is defined as

$$r_{ij} = \frac{E[(X_i - \mu_i)(X_j - \mu_j)]}{\sigma_i \sigma_j}. \quad (12)$$

Note that if X_i and X_j are independent, then $r_{ij} = 0$.

As a simple approximation method, this work uses Wilkinson's method [12], which fits only the first two moments of the lognormal's sum to those of another lognormal distribution function. The first two moments, u_1 and u_2 , are as follows:

$$u_1 = E[L] = E[\exp(Z)] = \sum_{i=1}^n \exp\left(\mu_i + \frac{\sigma_i^2}{2}\right) \quad (13)$$

$$\begin{aligned} u_2 &= E[L^2] = E[\exp(2Z)] \\ &= \sum_{i=1}^n \exp(2\mu_i + 2\sigma_i^2) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \{\exp(\mu_i + \mu_j) \end{aligned}$$

$$\cdot \exp\left(\frac{1}{2}(\sigma_i^2 + \sigma_j^2 + 2r_{ij}\sigma_i\sigma_j)\right)\}. \quad (14)$$

From (13) and (14), μ_Z and σ_Z are derived as

$$\mu_Z = 2 \ln u_1 - \frac{1}{2} \ln u_2 \quad (15)$$

$$\sigma_Z^2 = \ln u_2 - 2 \ln u_1. \quad (16)$$

Note that μ_Z and σ_Z correspond to the μ and σ of L defined in (10), respectively. From (15) and (16), the following interesting theorems can be derived.

Theorem1

Let L_1, L_2, \dots, L_n be n identical lognormally distributed RVs $LN(\mu, \sigma^2)$ and $\sigma \ll 1$ meaning that σ 's quadratic terms are negligible compared with the other terms, then μ_Z and σ_Z can be represented as $\mu + \ln n$ and $\frac{\sigma}{\sqrt{n}} \sqrt{1 + \frac{2}{n} \sum_i \sum_j r_{ij}}$, respectively.

(pf.) μ_Z and σ_Z can be derived from (15) and (16) as follows:

$$\begin{aligned} \mu_Z &= \mu + \ln n + \frac{\sigma^2}{2} \\ &\quad - \frac{1}{2} \ln \left\{ \frac{\exp(\sigma^2) + \frac{2}{n} \sum_i \sum_j \exp(r_{ij}\sigma^2)}{n} \right\} \end{aligned} \quad (17)$$

$$\sigma_Z = \sqrt{\ln \left(\frac{\exp(\sigma^2) - 1}{n} + \frac{\frac{2}{n} \sum_i \sum_j \exp(r_{ij}\sigma^2) + 1}{n} \right)}. \quad (18)$$

Since $\sigma \ll 1$, μ_Z and σ_Z can be represented as follows by ignoring the σ 's quadratic terms:

$$\mu_Z \sim \mu + \ln n \quad (19)$$

$$\sigma_Z \sim \frac{\sigma}{\sqrt{n}} \sqrt{1 + \frac{2}{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^n r_{ij}}. \quad (20)$$

(q.e.d.)

Theorem1 indicates that the median $\exp(\mu)$ is proportional to the number of RVs summed up together (n):

$$(\text{Median of } L) \sim \exp(\mu + \ln n) = n \exp(\mu). \quad (21)$$

Since r_{ij} ranges from 0 to 1, σ_Z ranges from $\frac{\sigma}{\sqrt{n}}$ to σ , which indicates that the parameter σ decreases as the number of RVs summed up together (n) increases. Specifically, if X_i and X_j are mutually independent (i.e., $r_{ij} = 0$), then $\sigma_Z = \frac{\sigma}{\sqrt{n}}$. In this case the L 's $k\sigma$ worst case can be expressed as follows:

$$\begin{aligned} (k\sigma \text{ worst case of } L) &\sim \exp\left(\mu + \ln n + \frac{k\sigma}{\sqrt{n}}\right) \\ &= n \exp\left(\mu + \frac{k\sigma}{\sqrt{n}}\right). \end{aligned} \quad (22)$$

Based on a model fitting result for our target process

technology, if $\sigma = 0.2$, $\mu = -20$ and $r_{ij} = 0$, approximation errors for μ_Z and σ_Z introduced by ignoring the second order of σ in (17) and (18) are less than 0.1% and 1%, respectively, for $n = 4$. Therefore, the assumption stated in Theorem1 (i.e., the second order of σ is negligible) is feasible.

4.2 Averaging Effect of Paths in NTV Operation

In super-threshold voltage (STV) operation where the operating voltage is sufficiently higher than the threshold voltage, random WID variations have an *averaging effect*, which reduces the relative delay variation $\frac{\sigma}{\mu}$ of a path as the number of logic gates chained in series increases.

If the path consists of n identical gates chained in series, the variance of the path delay is $n\sigma^2$. Since the mean delay of the path is $n\mu$, the relative delay variation is proportional to $\frac{1}{\sqrt{n}}$ [6]. Therefore, the relative delay variation decreases as n increases. In near-threshold voltage (NTV) operation where the operating voltage is close to the threshold voltage, a stronger averaging effect is observed than that observed in STV operation. This means that the relative delay variation of an n -stage path is proportional to a value which is less than $\frac{1}{\sqrt{n}}$.

Corollary1 (Averaging Effect)

For a given RV L which is the sum of n independent lognormal RVs, let $x_{k\sigma,n}$ be the L 's $k\sigma$ worst case as shown in Fig. 2 and let *variation* v_n be defined as $x_{k\sigma,n} - x_{0\sigma,n}$ which indicates the difference between the $k\sigma$ worst case and the median of L . Then the ratio of v_n to v_1 , which we refer to as *averaging effect ratio* is represented as:

$$(\text{averaging effect ratio}) = \frac{v_n}{v_1} = n \cdot \frac{\exp\left(\frac{k\sigma}{\sqrt{n}}\right) - 1}{\exp(k\sigma) - 1}. \quad (23)$$

(pf.) Immediate from (21) and (22). (q.e.d.)

The averaging effect ratio represents the magnitude of the delay variation. If the averaging effect is stronger, the magnitude of the delay variation gets smaller. Therefore, if the averaging effect is strong, the averaging effect ratio is small. Since, in a normal distribution, the median is equal to the mean and the distance between the median and the $k\sigma$ worst case is $k\sigma$, the averaging effect ratio in the normal distribution is \sqrt{n} . Corollary1 shows that the averaging effect

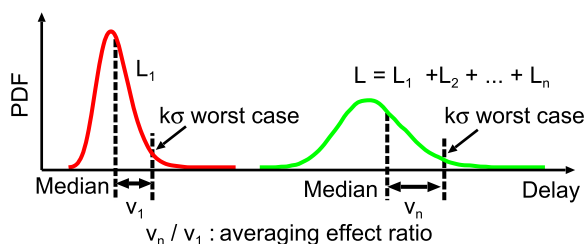


Fig. 2 Definition of averaging effect ratio.

ratio in NTV operation is smaller than that in STV operation (i.e., \sqrt{n}). In an actual MOS circuit, the median corresponds to the delay in a TT condition where both the threshold voltages of pMOS and nMOS are typical values. Note that the σ in (23) is corresponding to $\frac{\sigma'}{V_0}$ where the σ' corresponds to a standard deviation of V_{th} if we consider a delay variation of a circuit using this model. Typical values for the σ in the latest process technologies are ranging from 0.2 to 0.6. For example, the averaging effect ratio of a 16-stage logic path in a circuit operated with NTV is 1.62 if $\sigma = 0.4$ and $k = 5$ while that in a circuit operated with STV is 4. This intuitively means that the performance degradation along with an increase of the logic depth in NTV operation is slower than that in STV operation.

4.3 Delay Distribution of Parallel Paths in NTV Operation

The impact of the number of independent parallel paths (N_{cp} , denoted p for simplicity) on the maximum critical path delay is obtained by performing a MAX operation for the paths. It is well known that the maximum operating frequency (Fmax) of a sequential circuit needs to be lowered as the number of critical paths increases to maintain the same timing yield [5]. In this paper, we refer to the speed of the Fmax degradation as an *Fmax degradation speed*. This key word is used in the following theorem.

Theorem2

Suppose we have a lognormally distributed RV L_{log} and a normally distributed RV L_{norm} . Let $\max_i(L)$ be defined as a result of the MAX operation for i identical independent RVs L . Then the Fmax degradation speed along with the increase of the number of RVs p for $\max_p(L_{log})$ is exponential to that for $\max_p(L_{norm})$.

(pf.) Let $f(x)$ be a PDF for either normally or lognormally distributed RVs. The MAX of p parallelly connected $f(x)$ is

$$G(x) = \left(\int_{-\infty}^x f(x) dx \right)^p = \left(1 - \int_x^{\infty} f(x) dx \right)^p. \quad (24)$$

For a large x in (24), which means $\int_x^{\infty} f(x) dx$ is small, $G(x)$ can be approximated as

$$G(x) \sim 1 - p \int_x^{\infty} f(x) dx = 1 - p + pF(x), \quad (25)$$

where $F(x)$ is a CDF of $f(x)$.

Let y_0 and d be defined as a timing yield and a target delay which satisfies the timing yield y_0 , respectively. To keep this timing yield y_0 constant, the value of (25) should be kept at y_0 for different p values. If $f(x)$ is a lognormal distribution function $LN(\mu, \sigma^2)$, with (4), the target delay (d_{log}) for satisfying the yield y_0 is

$$d_{log} = \exp\left(\mu + \sigma\Phi^{-1}\left(\frac{y_0 - 1 + p}{p}\right)\right). \quad (26)$$

In a similar way, if $f(x)$ is a normal distribution function

$N(\mu', \sigma'^2)$, the target delay (d_{norm}) for satisfying y_0 is

$$d_{\text{norm}} = \mu' + \sigma' \Phi^{-1} \left(\frac{y_0 - 1 + p}{p} \right). \quad (27)$$

Since the Fmax is inversely proportional to d , it is immediate from (26) and (27) that the Fmax degradation speed along with the increase of p for $\max_p(L_{\log})$ is exponential to that for $\max_p(L_{\text{norm}})$. (q.e.d.)

Although d_{\log} is exponential to d_{norm} as presented in (26) and (27), the degradation speed of d_{\log} along with the increase of p is very slow. This property is summarized in the following corollary:

Corollary2

The degradation of d_{\log} is sublinear to the increase of p . (pf.) The CDF of Gaussian distribution, $\Phi(x)$ is represented by an *error function* ($\text{erf}(x)$):

$$\Phi(x) = \frac{1}{2} \left(1 + \text{erf} \left(\frac{x}{\sqrt{2}} \right) \right). \quad (28)$$

In [13], $\text{erfc}(x) (= 1 - \text{erf}(x))$ is approximated as the sum of two exponential terms for a positive number x :

$$\text{erfc}(x) \sim \frac{1}{6} \exp(-x^2) + \frac{1}{2} \exp(-4x^2/3). \quad (29)$$

In this paper, for simplicity, we further approximate (29) by ignoring a non-dominant term for a large x :

$$\text{erfc}(x) \sim \frac{1}{6} \exp(-x^2). \quad (30)$$

The approximation (30) provides an inverse function of $\Phi(x)$ in a closed form:

$$\Phi^{-1}(x) \sim \sqrt{-2 \ln(12(1-x))}. \quad (31)$$

From (26) and (31), the degradation of d_{\log} can be expressed by an elementary function as follows:

$$d_{\log} = \exp \left[\mu + \sigma \sqrt{2} \sqrt{\ln \left(\frac{p}{12(1-y_0)} \right)} \right]. \quad (32)$$

Since $\exp(\sqrt{\ln p})$ is sublinear to p^a for any positive number a [†], the degradation of d_{\log} is sublinear to the increase of p . (q.e.d.)

Based on a model fitting result for our target process technology, the approximation errors for a single stage buffer delay d_{\log} introduced by the approximation (30) are less than 2.7% and 0.2% for 3σ timing yield when $p = 1$ and $p = 256$, respectively. Hence the approximation (30) is feasible.

4.4 Impact of Gate Sizing on NTV Operation

A smaller transistor gate width causes a larger threshold

[†]Since $\lim_{p \rightarrow \infty} \exp(\sqrt{\ln p})/p^a = 0$ for $a > 0$, $\exp(\sqrt{\ln p})$ is sublinear to p^a .

voltage variation, resulting in the larger high- σ worst case delay. According to Pelgrom's model [14], the standard deviation of threshold voltage (V_{th}), $\sigma_{V_{\text{th}}}$ is represented as:

$$\sigma_{V_{\text{th}}} = \frac{A_{\text{vt}}}{\sqrt{WL}}, \quad (33)$$

where A_{vt} , W and L are a *Pelgrom coefficient*, transistor gate width and length, respectively. Let us define the term of *gate sizing effect on the worst case delay* as a ratio of the reduction in the worst case delay to the increase in gate size. Then we obtain the following corollary by (7) and (33).

Corollary3

The gate sizing effect on the worst case delay in NTV operation is approximately exponential to that in STV operation if load capacitance of the gate is proportional to its gate width. (pf.) Let us suppose a normally distributed RV $V_{\text{th}} \sim N(V_{\text{th}0}, \sigma_{V_{\text{th}}}^2)$, then, from (4), (7) and (33), the $k\sigma$ worst case delay of a single gate in NTV operation D_{NTV} is as follows:

$$D_{\text{NTV}} = D'_0 \exp \left(\frac{\beta \sigma_{V_{\text{th}}} k}{\sqrt{WL}} \right), \quad (34)$$

where k is a parameter which determines the target timing yield of the gate. D'_0 and β are constants. It is well known that the distribution of single gate delay D_{STV} in STV operation has the Gaussian distribution if V_{th} is a normally distributed RV. From (33), the $k\sigma$ worst case delay of a single gate in STV operation D_{STV} can be represented:

$$D_{\text{STV}} = \beta'_1 V_{\text{th}0} + \beta'_2 k \sigma_{V_{\text{th}}} + \text{const.}, \quad (35)$$

where β'_1 and β'_2 are constants. From (34) and (35), it is immediate that the gate sizing effect on the worst case delay in NTV operation is exponential to that in STV operation. (q.e.d.)

This corollary indicates that gate sizing for a logic gate operated with NTV *exponentially* affects the performance of the logic gates.

5. Validation with a Commercial Process Technology Model

This section validates the properties described in the theorems and corollaries by comparing them with circuit simulation results obtained using a commercial 28 nm process technology. We perform transistor-level circuit simulation [15] using a foundry provided Monte Carlo simulation package. We use a 1.0 V and a 0.4 V as the STV and the NTV, respectively. We examined the delay of circuits having different logic depths and different numbers of parallel paths. L_i ($i = 1, 2, \dots, k$) shown in Fig. 3 is a buffer of i th logic stage in a circuit. All the buffers have an identical lognormal distribution $LN(\mu, \sigma^2)$ if they are operated with the NTV. The delay value D_k is obtained from a delay of an intermediate part of a sufficiently long buffer chain as shown in Fig. 3.

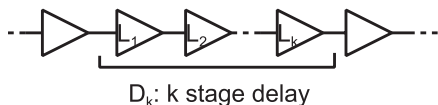


Fig. 3 Buffer chain example where all buffers have the same fan-out.

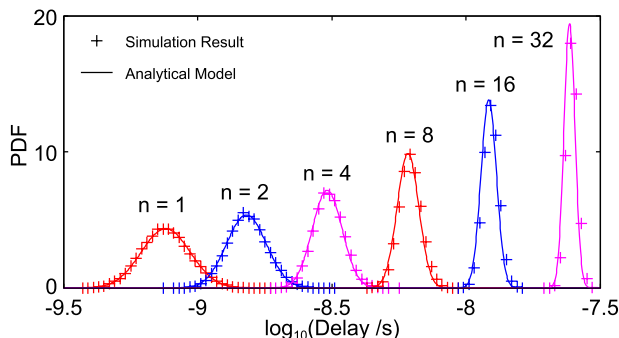


Fig. 4 Buffer chain simulation result. $\mu = -21$, $\sigma = 0.21$ and $r = 0.32$.

5.1 Delay Distribution of a Buffer Chain

Figure 4 shows delay distributions of a buffer chain having different logic depths shown as n . Through an input slew, buffer delays are mutually correlated. If the delay of a buffer is large, the output transition time (i.e., input slew of the next buffer) of the buffer is also large, which results in an increase of an input slew rate in the next buffer. This causes an increase in the propagation delay of a buffer in the next stage. This is the mechanism of the correlation between the two consecutive buffers. For reflecting this correlation in our analytical model, we assume that the adjoined buffers are correlated from each other as follows:

$$r_{ij} = \begin{cases} 1 & (i = j) \\ r & (i = j \pm 1) \\ 0 & (\text{otherwise}). \end{cases} \quad (36)$$

Solid lines in Fig. 4 are obtained using a model derived from Theorem1 by fitting the parameter values μ , σ and r . Dots show Monte Carlo simulation results. As can be seen from Fig. 4, the delay distributions estimated using our analytical model are well matched with the results of Monte Carlo simulation using a commercial process technology model. It demonstrates that the delay distributions of actual multiple stage buffer chains closely fit the lognormal distribution.

5.2 Validation of Averaging Effect

We examine the averaging effect ratio in the 4σ worst case of the n -stage buffer chain for various logic depths n . Note that in order to obtain a 3σ timing yield per chip, we have to obtain more than the 3σ timing yield per critical path. Hence, as a representative value, we use a 4σ to evaluate the worst case delay per critical path, which corresponds to a 3σ timing yield per chip and there are about 40 independent critical paths in it. Figure 5 shows the results. The

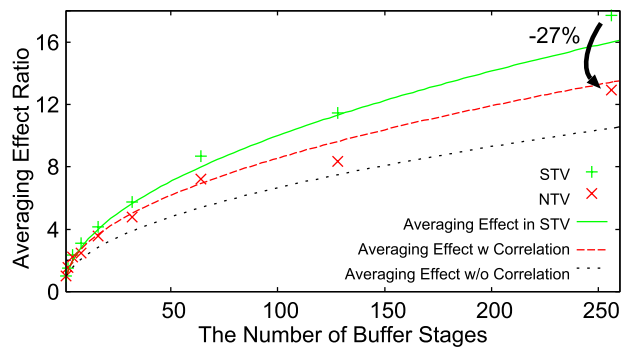


Fig. 5 Averaging effect ratio for buffer chains.

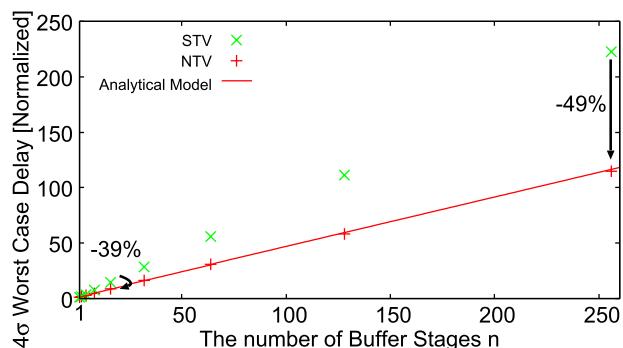


Fig. 6 Logic depth vs. the 4σ worst case delay.

dots show the results of Monte Carlo simulation. The solid line labeled “Averaging Effect in STV” shows the averaging effect ratio for *normally distributed independent* RVs. The graph labeled “Averaging Effect w/o Correlation” shows the averaging effect ratio for *lognormally distributed independent* RVs. The graph for lognormal RVs can be obtained using (23). For $n = 256$, a 27% smaller averaging effect ratio (i.e., stronger averaging effect) can be observed in NTV operation compared with that in STV operation. Since each logic gate has a correlation, the averaging effect ratio gets larger.

5.3 Validation of Fmax Degradation Speed

Since Fmax is an inverse of the worst case delay, we evaluate the worst case delay for analyzing the Fmax degradation speed. First we examine the speed of the worst case delay degradation in a buffer chain along with the increase of a logic depth. We suppose that the buffer chain represents a critical path in a processor. Figure 6 shows the 4σ worst case delay for different logic depths. The number of parallel paths is one in this case. Dots show the results of Monte Carlo simulation and the solid line shows the results obtained with an analytical model presented in Theorem1. Each dot represents the 4σ worst case delay normalized by the delay of a single stage buffer operated with a corresponding supply voltage. The results show that the degradation speeds for the worst case delay of 16-stage and 256-stage buffer chains in NTV operation are 39% and 49% smaller

than those in STV operation, respectively. This demonstrates that the averaging effect in NTV operation is stronger than that in STV operation as described in Theorem1.

Let us consider to reduce the number of pipeline stages of a processor from 4 to 2. If the number of pipeline stages is halved, the logic depth in a critical path is roughly doubled. Therefore, if the logic depth of the 4-stage processor is 32, that of the 2-stage processor is 64. According to Fig. 6, when the logic depth is doubled from 32 to 64, the 4σ worst case delay becomes 91% larger in NTV operation while it becomes 98% larger in STV operation. Therefore, the performance degradation of a processor incurred by reducing the number of its pipeline stages in NTV operation is smaller than that incurred in STV operation.

Next we evaluate the degradation speed of the worst case delay in parallel buffer chains in order to see the performance of a processor with many parallel critical paths. We use a circuit where the 8-stage buffer chains are connected in parallel as shown in Fig. 7. We consider the parallel paths as a representative of a chip for the evaluation. If the chip requires a 3σ timing yield, the parallel paths also require the 3σ timing yield. Therefore, we evaluate the 3σ worst case delay of the parallel paths in this experiment. The overall delay in this circuit is obtained by performing MAX operation for all critical paths. Figure 8 shows the worst case delays for different numbers of critical paths N_{cp} . Dots show the results of Monte Carlo simulation and the solid line shows the results obtained with an analytical model presented in Theorem2 and Corollary2. Each dot represents the 3σ worst case delay normalized by the delay of a single-path buffer chain operated with a corresponding supply voltage.

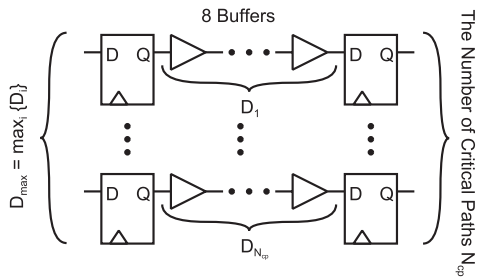


Fig. 7 Parallelism of 8-stage-buffer chains.

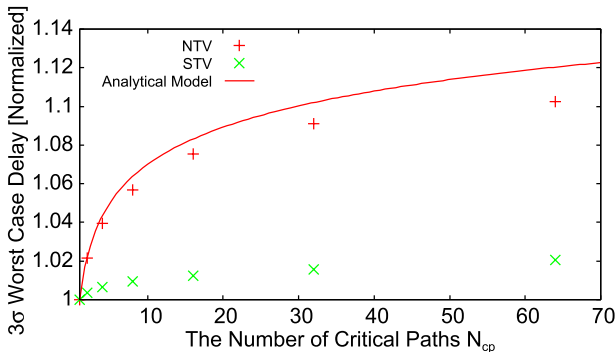


Fig. 8 The number of critical paths N_{cp} vs. the 3σ worst case delay.

The results demonstrate that the speed of the worst case delay degradation for parallel critical paths against an increase of N_{cp} in NTV operation is faster than that in STV operation. However, the 3σ worst case delay in NTV operation is sublinear to the number of parallel critical paths, and the degradation speed is less than 11% both in NTV and STV cases, which indicates the increase of the number of parallel paths has weak impact on the delay degradation. There are about two percent of errors between simulation results and the analytical model. The error comes from the approximation (31) which is used for calculating the worst case delay of parallel chains.

5.4 Validation of Gate Sizing Effect

We evaluate the 4σ worst case delays in single-stage buffers (i.e., D_1 in Fig. 3) for different gate sizes. Figure 9 shows PDFs of different sized single-stage buffers in NTV and STV operations, respectively. “X” represents the drive strength of the gate, which is proportional to the gate width. The figure shows that PDFs of buffers with a large X have a smaller relative variation although their median does not move widely. This is because the typical value of transistor’s ON current and load capacitance are roughly proportional to the gate width, and the delay variation is strongly dependent on the gate width according to Pelgrom’s model (33). Between NTV operation and STV operation, for the same gate size, the relative variation (i.e., σ/μ) in NTV operation is larger than that in STV operation. Figure 10 shows the 4σ

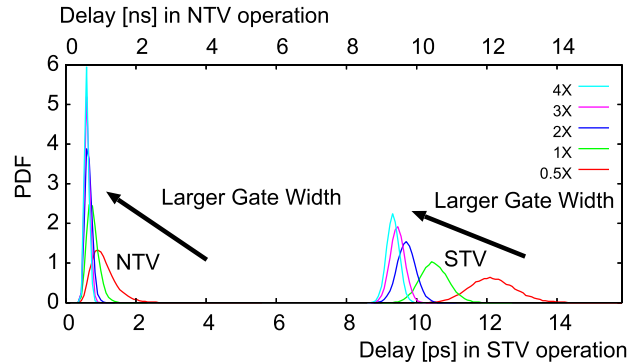


Fig. 9 Delay distributions for different gate sizes.

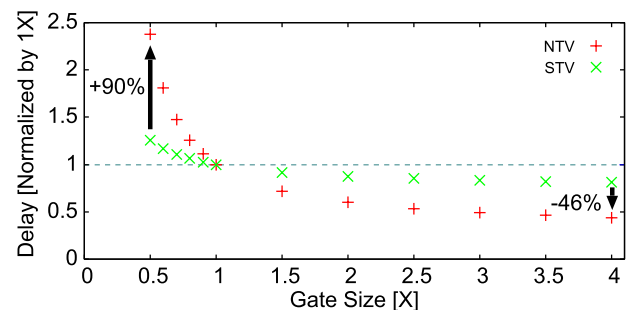


Fig. 10 Buffer size X vs. the 4σ worst case delay.

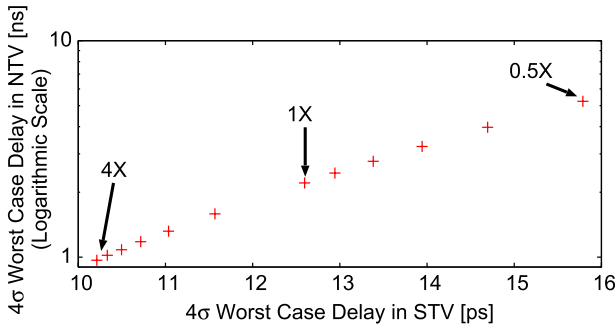


Fig. 11 The NTV 4σ worst case delay vs. the STV 4σ worst case delay.

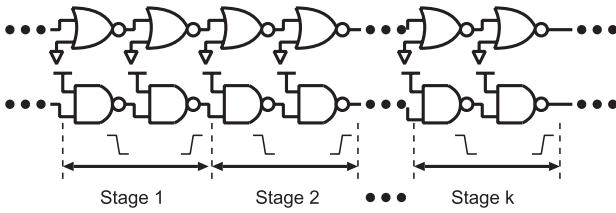


Fig. 12 Test circuit structure for NAND2 and NOR2.

worst case delay of single buffers with different gate sizes. Each delay is normalized by the delay with $X = 1$. If we look at the 4σ worst case delay, a 90% delay degradation for a 0.5X buffer and a 46% delay improvement for a 4X buffer can be observed in NTV operation compared with those in STV operation, respectively. Figure 11 shows single buffer delays for different gate sizes. The vertical axis shows the NTV operation delay in a logarithmic scale, and the horizontal axis shows the STV operation delay in a linear scale. Since dots roughly line up in a straight line, Corollary 3 holds even for the commercial process technology model. Main reasons for the nonlinearity are that a buffer’s load capacitance is not exactly proportional to the gate width and that the ON current is not linear to the gate width due to the narrow channel effect.

5.5 Validation of Properties in Other Logic Cells

We validate that the properties derived for the buffer chain hold for 2-input NAND (NAND2 in short) or 2-input NOR (NOR2 in short) with stacked transistors. In order to evaluate rise and fall delays simultaneously, we evaluate the propagation delay of two serially connected logic gates as the delay of NAND2/NOR2 as shown in Fig. 12.

Figure 13 shows averaging effect ratios for NAND2 and NOR2 chains. The line labeled “Averaging Effect in STV” shows the averaging effect ratio for normally distributed independent RVs (i.e., \sqrt{n}). The other lines show the results obtained with the analytical model presented in Theorem 1. Although the worst case delay of the chains increases because of stacked transistors in NAND2/NOR2, we confirmed that the stronger averaging effect in NTV operation is still remained.

Figure 14 shows the 4σ worst case delay for different

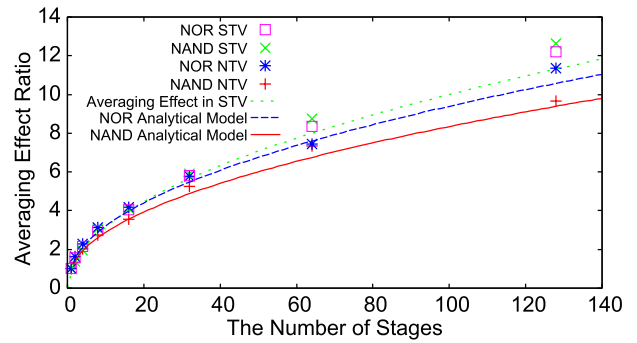


Fig. 13 Averaging effect ratio for NAND2/NOR2 chains.

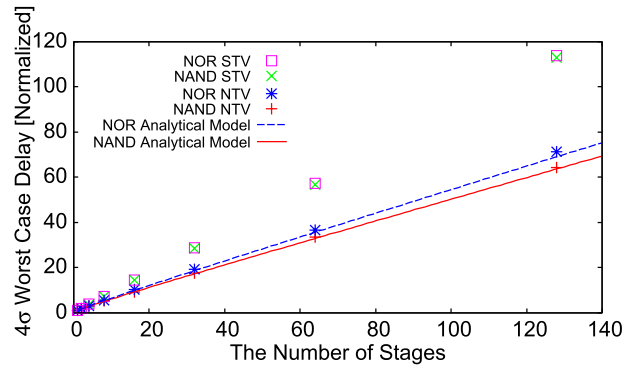


Fig. 14 Logic depth vs. the 4σ worst case delay for NAND2/NOR2 chains.

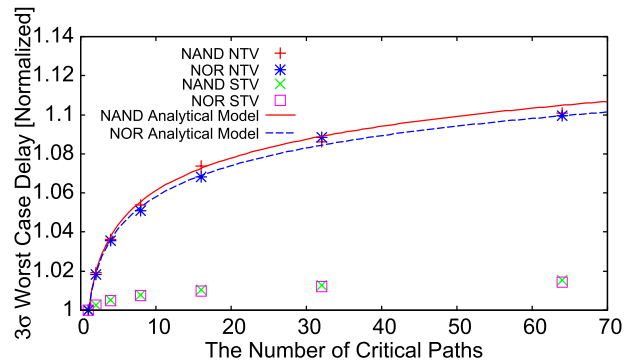


Fig. 15 The number of critical paths N_{cp} vs. the 3σ worst case delay for NAND2/NOR2 chains.

logic depths for NAND2/NOR2 chains. The lines show the results obtained with the analytical model presented in Theorem 1. The results show that the averaging effect ratio of the NAND2/NOR2 chains in NTV operation is always smaller than that in STV operation. Figure 14 shows the 4σ worst case delays with the different number of stages normalized by the delay of the single stage chain operated with the corresponding supply voltage. The normalized worst case delay in NTV operation is smaller than that in STV operation. This demonstrates that there is a stronger averaging effect in NTV operation than that in STV operation.

Figure 15 shows the 3σ worst case delays for different numbers of critical paths N_{cp} . Each dot represents the 3σ

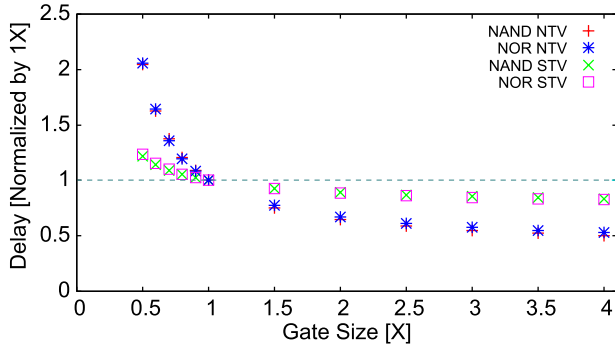


Fig. 16 Gate size X vs. the 4σ worst case delay for NAND2/NOR2 chains.

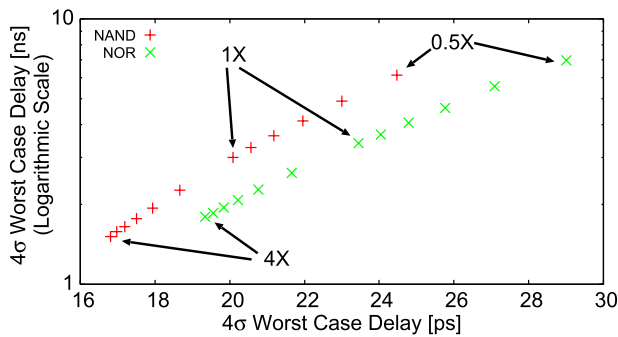


Fig. 17 The NTV 4σ worst case delay vs. the STV 4σ worst case delay for NAND2/NOR2 chains.

worst case delay normalized by the delay of a single-path NAND2/NOR2 chain operated with a corresponding supply voltage. The lines show the results obtained with the analytical model presented in Theorem2 and Corollary2. Like buffer chains, the 3σ worst case delay in NTV operation is sublinear to the number of parallel critical paths as described in Corollary2, and the increase of the delay over the delay of the single path is less than 10% both in NTV and STV cases.

Figure 16 shows the 4σ worst case delay of single stage NAND2/NOR2 gates with different gate sizes. Each delay is normalized by the delay with $X = 1$. This result demonstrates the sensitivity of the normalized 4σ worst case delay to gate size X is also larger in NTV operation than that in STV operation for NAND2/NOR2 chains.

Figure 17 shows delays of single stage NAND2/NOR2 chains for different gate sizes. The vertical axis shows the delay for the NTV operation in a logarithmic scale, and the horizontal axis shows the delay for the STV operation in a linear scale. Since dots roughly line up in a straight line, Corollary3 holds even for the logic gates with stacked transistors.

6. Summary of Properties for NTV Operation

Table 1 summarizes the theorems and corollaries presented in previous sections. In this section, we show how these theorems and corollaries are mutually dependent and dis-

Table 1 Summary of properties. C: Corollary. L: Lemma. T: Theorem. p : degree of parallelism. N : logic depth. W : gate width. L : gate length.

Property	STV	NTV
Delay distribution	Gaussian	Lognormal (L1)
Sum of distributions	Gaussian	Lognormal (T1)
Averaging effect	\sqrt{N}	Stronger than STV (C1)
Fmax degradation speed	Sublinear to p (C2)	Exponential to STV (T2) but still sublinear to p (C2)
Gate sizing effect	$1/\sqrt{WL}$	Exponential to STV (C3)

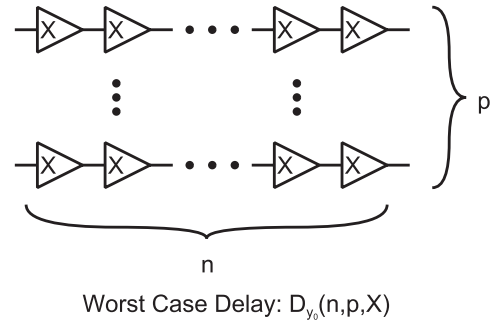


Fig. 18 p -parallel n -stage buffer chains where all buffers in chains have the same gate size X .

cuss how we can effectively exploit these properties in an architectural design stage.

Let us consider p -parallel n -stage buffer chains where individual buffers have the same gate size X from each other as shown in Fig. 18. We suppose to keep the timing yield y_0 constant. A buffer whose drive strength $X = 1$ is supposed to have a delay distribution $LN(\mu, \sigma^2)$ in NTV operation. Under these assumptions, we find the optimal set of n , p , and X , which minimizes the energy consumption of the circuit with keeping the timing yield y_0 constant.

From (34) in Corollary3, we can derive each buffer in chains has a delay distribution $LN(\mu, (\sigma/\sqrt{X})^2)$. Then, from Theorem1, an n -stage chain composed of X -drive buffers has the following delay distribution:

$$LN\left(\mu + \ln n, \frac{\sigma}{\sqrt{nX}} \sqrt{1 + \frac{2}{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^n r_{ij}}\right). \quad (37)$$

Finally, from Theorem2 and Corollary2, the worst case delay of the circuit $D_{y_0}(n, p, X)$ can be expressed as:

$$D_{y_0}(n, p, X) = n \exp\left(\mu + \frac{\sigma}{\sqrt{nX}} \sqrt{1 + \frac{2}{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^n r_{ij}}\right) \cdot \sqrt{2} \sqrt{\ln \frac{p}{12(1-y_0)}}. \quad (38)$$

For better understanding, we consider (38) in two conditions according to the degree of n as follows:

When n is sufficiently large

The factor $\frac{1}{\sqrt{nX}} \sqrt{1 + \frac{2}{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^n r_{ij}} \sqrt{2} \sqrt{\ln \frac{p}{12(1-y_0)}}$ in

(38) is sufficiently small. This factor corresponds to the averaging effect of the buffer chains presented in Corollary1. In this condition, the sensitivity of p and X to $D_{y_0}(n, p, X)$ is very small, which means that we can take the same design strategy for the optimization of NTV circuit as taken in STV circuit design. For example, in case of $n = 64$, $p = 256$ and $y_0 = \Phi(4)$, doubling the value of X from 1 to 2 reduces the worst case delay of the circuit only by 5.0%. Similarly, doubling the value of p from 256 to 512 with $X = 1$ increases the worst case delay by 0.4%. These trends are quite similar to those in STV operation.

When n is small

The factor $\frac{1}{\sqrt{nx}} \sqrt{1 + \frac{2}{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^n r_{ij}} \sqrt{2} \sqrt{\ln \frac{p}{12(1-y_0)}}$ in (38) is relatively large in comparison with μ . Hence, the sensitivity of p and X to $D_{y_0}(n, p, X)$ is relatively large in this case. From Corollary2 and Corollary3, we can conclude that for short critical paths, tuning X has a stronger impact on the performance than tuning the degree of parallelism in NTV operation. Similar trend is also observed in STV operation but it is weaker than NTV operation. For example, in case of $n = 8$, $p = 256$ and $y_0 = \Phi(4)$, doubling the value of X from 1 to 2 reduces the worst case delay of the circuit by 13% in NTV operation while the delay is reduced by 2% in STV operation. Similarly, doubling the value of p from 256 to 512 with $X = 1$ increases the worst case delay by 1.2% in NTV operation while the delay increases by 0.16% in STV operation. Both gate sizing effect and parallelization effect in NTV operation are much stronger than those observed in STV operation.

7. Conclusion

Near-threshold computing is one of the most promising approaches for achieving high performance and energy efficient computation of microprocessors. In this paper, we prove several theorems that help consider architectural design strategies for NTV operation where the logic gate delay has a lognormal distribution. Corollary1 shows that near-threshold voltage (NTV) operation has a stronger averaging effect than the super-threshold voltage (STV) operation where the standard deviation is proportional to the square root of the number of chained gates in series. With Monte Carlo simulation using a commercial 28 nm process technology model, we show that the averaging effect of a buffer chain in the NTV operation is 27% stronger than that in the STV operation. For example, according to the discussion in Sect. 5.3, if we increase the logic depth of a path from 32 to 64, the 4σ worst case delay becomes 91% larger in NTV operation while it becomes 98% larger in STV operation. Theorem2 shows that the maximum operating frequency (Fmax) of a circuit operated with the NTV is more widely degraded than the same circuit operated with the STV when the number of critical paths increases. However, Corollary2 shows that the Fmax degradation speed along with the increase of the number of parallel paths (N_{cp}) is negligibly slow both

for NTV and STV operations. This means that the impact of the parallelization on the Fmax degradation is negligible even in NTV operation. Corollary3 shows the gate upsizing exponentially improves the worst case delay in NTV operation compared with the delay improvement in STV operation. For example, according to the discussion at the end of Sect. 6, if we double the size of gates in a path from 1X to 2X, the 4σ worst case delay is reduced by 13% in NTV operation while it is reduced by 2% in STV operation. The theorems and corollaries presented in this paper intuitively imply that the performance degradation of a processor incurred by reducing the number of its pipeline stages in NTV operation is smaller than that incurred in STV operation. If we use a deeply pipelined processor in NTV operation, gate upsizing is an effective way to improve the throughput of the processor since it has an exponential impact on the worst case delay in NTV operation.

Our future work will be devoted to develop optimization methods for near-threshold circuit design by exploiting the properties proved in this paper.

Acknowledgment

This work has been partly supported by KAKENHI Grant-in-Aid for Scientific Research B-25280014 and B-26280013. This work is also supported by VLSI Design and Education Center (VDEC), the University of Tokyo in collaboration with Cadence Design Systems, Inc. and Synopsys, Inc.

References

- [1] J. Shiomi, T. Ishihara, and H. Onodera, "Microarchitectural-level statistical timing models for near-threshold circuit design," *The 20th Asia and South Pacific Design Automation Conference*, pp.87–93, Jan. 2015.
- [2] S. Jain, S. Khare, S. Yada, V. Ambili, P. Salihundam, S. Ramani, S. Muthukumar, M. Srinivasan, A. Kumar, S.K. Gb, R. Ramanarayanan, V. Erraguntla, J. Howard, S. Vangal, S. Digha, G. Ruhl, P. Aseron, H. Wilson, N. Borkar, V. De, and S. Borkar, "A 280 mV-to-1.2 V Wide-Operating-Range IA-32 Processor in 32 nm CMOS," *IEEE International Solid-State Circuits Conference*, pp.66–68, Feb. 2012.
- [3] S. Keller, D.M. Harris, and A.J. Martin, "A compact transregional model for digital CMOS circuits operating near threshold," *IEEE Trans. VLSI Syst.*, vol.22, no.10, pp.2041–2053, Oct. 2014.
- [4] K.A. Bowman, S.G. Duvall, and J.D. Meindl, "Impact of die-to-die and within-die parameter fluctuations on the maximum clock frequency distribution for gigascale integration," *IEEE J. Solid-State Circuits*, vol.37, no.2, pp.183–190, Feb. 2002.
- [5] D. Marculescu and E. Talpes, "Energy Awareness and Uncertainty in Microarchitecture-Level Design," *IEEE Micro*, vol.25, no.5, pp.64–76, Sept. 2005.
- [6] M. Eisele, J. Berthold, D. Schmitt-Landsiedel, and R. Mahnkopf, "The impact of intra-die device parameter variations on path delays and on the design for yield of low voltage digital circuits," *Proc. 1996 International Symposium on Low Power Electronics and Design*, pp.237–242, Aug. 1996.
- [7] Y. Zhan, A.J. Strojwas, X. Li, L.T. Pileggi, D. Newmark, and M. Sharma, "Correlation-aware statistical timing analysis with non-Gaussian delay distributions," *Proc. Design Automation Conference*, pp.77–82, June 2005.

- [8] K. Chopra, B. Zhai, D. Blaauw, and D. Sylvester, "A new statistical max operation for propagating skewness in statistical timing analysis," Proc. 2006 IEEE/ACM International Conference on Computer-Aided Design, pp.237–243, Nov. 2006.
- [9] J. Singh and S. Sapatnekar, "Statistical timing analysis with correlated non-Gaussian parameters using independent component analysis," 2006 43rd ACM/IEEE Design Automation Conference, pp.155–160, 2006.
- [10] H. Chang and S.S. Sapatnekar, "Prediction of leakage power under process uncertainties," ACM Trans. Des. Autom. Electron. Syst., vol.12, no.2, April 2007.
- [11] T. Sakurai and A.R. Newton, "Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas," IEEE J. Solid-State, vol.25, no.2, pp.584–594, April 1990.
- [12] A.A. Abu-Dayya and N.C. Beaulieu, "Comparison of methods of computing correlated lognormal sum distributions and outages for digital wireless applications," Proc. IEEE Vehicular Technology Conference (VTC), vol.1, pp.175–179, June 1994.
- [13] M. Chiani, D. Dardari, and M.K. Simon, "New exponential bounds and approximations for the computation of error probability in fading channels," IEEE Trans. Wireless Commun., vol.2, no.4, pp.840–845, July 2003.
- [14] M.J.M. Pelgrom, A.C.J. Duinmaijer, and A.P.G. Welbers, "Matching properties of MOS transistors," IEEE J. Solid-State Circuits, vol.24, no.5, pp.1433–1439, Oct. 1989.
- [15] I. Synopsys, HSPICE User's Manual: Simulation and Analysis, 2010.



Hidetoshi Onodera received the B.E., M.E., and Dr. Eng. degrees in Electronic Engineering from Kyoto University, Kyoto, Japan, in 1978, 1980, and 1984, respectively. He joined the Department of Electronics, Kyoto University, in 1983, and currently a Professor in the Department of Communications and Computer Engineering, Graduate School of Informatics, Kyoto University. His research interests include design technologies for Digital, Analog, and RF LSIs, with particular emphasis on low-power design, design for manufacturability, and design for dependability. Dr. Onodera served as the Program Chair and General Chair of ICCAD and ASP-DAC. He was the Chairman of the IPSJ SIG-SLDM (System LSI Design Methodology), the IEICE Technical Group on VLSI Design Technologies, the IEEE SSCS Kansai Chapter, and the IEEE CASS Kansai Chapter. He served as the Editor-in-Chief of IEICE Transactions on Electronics and IPSJ Transactions on System LSI Design Methodology.



Jun Shiomi received the B.E. degree in Electronic Engineering from Kyoto University, Kyoto, Japan in 2014. He is currently working toward the M.E. degree at the Communications and Computer Engineering from Kyoto University. His research interest includes computer-aided design for low power and low voltage system-on-chips.



Tohru Ishihara received his B.E., M.E., and Dr.E. degrees in computer science from Kyushu University in 1995, 1997 and 2000 respectively. From 1997 to 2000, he was a Research Fellow of the Japan Society for the Promotion of Science. For the next three years he worked as a Research associate in VLSI Design and Education Center, the University of Tokyo. From 2003 to 2005, he was with Fujitsu Laboratories of America, Inc. as a member of research staff. From 2005 to 2011, he was with System LSI Research Center,

Kyushu University as an Associate Professor. In 2011, he joined Kyoto University, where he is currently an Associate Professor in the Department of Communications and Computer Engineering. His research interests include low power design and methodologies for embedded systems. He served as an editorial board member of Journal of Low Power Electronics and an executive committee member of DATE conference from 2009. He also served in technical program committees of a number of IEEE/ACM conferences including DATE, CODES+ISSS, and ISLPED.