A Highly Sensitive and Ultra Low-Power Forward Body Biasing Circuit to Overcome Severe PVT Variations and Extreme Voltage Scaling

Mohsen Radfar*†, Kriyang Shah and Jack Singh

Department of Electronic Engineering, La Trobe University, Melbourne, Victoria 3086, Australia

I. SUMMARY

Dynamic voltage scaling is one of the most popular methods used to reduce energy consumption in today's digital electronic systems. However, addressing process, voltage and temperature variations at subthreshold voltages has become an inevitable procedure. Using a variation sensitive and ultra low-power design, this paper proposed a novel technique capable of sensing and responding to process, voltage and temperature variations as well as dynamic voltage scaling by providing an appropriate forward body bias so that energy delay product of whole system was improved. Theoretical analysis for process variation probability, confirmed by post-layout HSPICE simulations for an 8-bit pipelined Kogge-Stone adder, showed that the circuit performance was enhanced in severe variations and extreme voltage scaling situation. For this adder, for example, assuming a voltage scaling form 0.8V to 0.3V and temperature changes of -15°C to 75°C, the proposed technique brought about a 7x less delay variation while energy delay product improved by 23% compared to a zero body biased adder.

KEY WORDS: dynamic voltage scaling, subthreshold design, forward body biasing, process, temperature and voltage variations

1. INTRODUCTION

With introduction of 65nm technologies, the semiconductor industry started to experience that although more transistors can fit on a die, because of severe static power restrictions this cannot be achieved. Leakage power continued to increase and as a result the supply voltage could not be scaled anymore and therefore switching energy dissipation could not be impeded.

Since then researchers have sought many techniques to overcome this problem. Dynamic voltage scaling (DVS) is an efficient method used to decrease energy consumption. DVS has been utilised in implementation of some low-power processor designs [1]. DVS method scales voltage in accordance with actual delay (or throughput) of the system so that the required performance is met, while fixed voltage systems are designed based on worst case delays in which performance suffers the most. DVS causes a significant energy saving because minimum possible voltage is chosen to meet the required performance. Although it is powerful in controlling dynamic energy, DVS cannot deal with the static energy which is a rapidly growing problem in short-channel devices and dominant in low activity systems. However, simultaneous static and dynamic energy management made viable through dynamic voltage and threshold scaling which tunes both supply and body bias voltages at the same time. In fact, minimum total energy for any required performance can be gained by this technique in digital systems fabricated on a 0.1µm technology or lower

* Correspondence to: Mohsen Radfar, Department of Electronic Engineering, La Trobe University, Melbourne, Victoria 3086, Australia

† E-mail: m.radfar@latrobe.edu.au

[2-3]. For low power high speed applications, this technique has been implemented successfully [3].

Minimum energy can be typically achieved when V_{DD} scales down to the subthreshold region [1]. Subthreshold systems have been proven to be fully functional below 200 mV [4]. Nevertheless, there are still some important challenges for subthreshold designers and the most significant one is variations. Process, voltage and temperature (PVT) have exponential correlations with the subthreshold current and tiny PVT variations, as a result, have a huge impact on performance and energy consumption. This impact still urges designers to employ adaptive body-bias techniques despite the fact that technology scaling counteracts the body biasing effect (with the increase of dopants in the below 100nm device channels to cause stronger inversion). On the other hand, as devices work using the subthreshold current in the subthreshold region, delay worsens substantially in this region too. These challenges need to be addressed to make the widespread exploitation of subthreshold design possible [5].

Process variations is usually divided to inter-die and intra-die variations [6]. Inter-die and intra-die variations can also occur at the same time. Moreover, intra-die variations are classified into systematic and random. In systematic variations, devices that are spatially correlated (e.g. are close together) experience the same effects while random variations can occur accidentally even for spatially correlated devices. Systematic intra-die (as well as inter-die) variations can be addressed by adaptive body-bias techniques [7].

At superthreshold voltages (nominal voltages), body biasing is a well familiar technique to designers for adjusting delay and leakage to overcome inter-die PVT variations. Although efficient in the superthreshold region, the impact of body biasing is especially sensed at the subthreshold voltages because of the exponential increase in the sensitivity of devices to the threshold voltage. For example in a typical 90nm technology, if threshold voltage is changed by 50mV at a 1V supply voltage, delay varies 13% whereas it results in a 55% delay increase at a 0.45V supply voltage [8].

Moreover noise margin susceptibility is also a major challenge at subthreshold voltages. As threshold voltage (V_{TH}) is managed by an independent doping process, PMOS/NMOS V_{TH} can be different considerably. This, for example, can result in an insufficient high output voltage at the fast NMOS slow PMOS corner (in which NMOS devices are much leakier than PMOS ones) or an insufficient low output voltage at the fast PMOS slow NMOS corner. Consequently, at process corners, not only noise margins can be violated, but also either rising or falling time is extremely long which in return results in increased timing breaches.

In a superthreshold circuit, the ratio between the PMOS and NMOS transistors, called the β-ratio or P/N ratio, is usually adjusted so that the noise margin of the gate is maximised. This optimal ratio in superthreshold region, however, is not exactly alike the optimal ratio in subthreshold region. If designed to only maximise the superthreshold noise margin, this P/N ratio can result in a high skew between PMOS and NMOS devices in subthreshold region.

As PMOS and NMOS can be controlled independently using body biasing techniques, this opens many opportunities for designers to optimally tune the β-ratio and preventing V_{TH} mismatch problems. For example, authors in [9] [10] used V_{TH} balancing schemes which enabled them to implement a supply voltage scaling from superthreshold to subthreshold voltages. Their body biasing technique adapts P/N-ratio dynamically while voltage scaling.

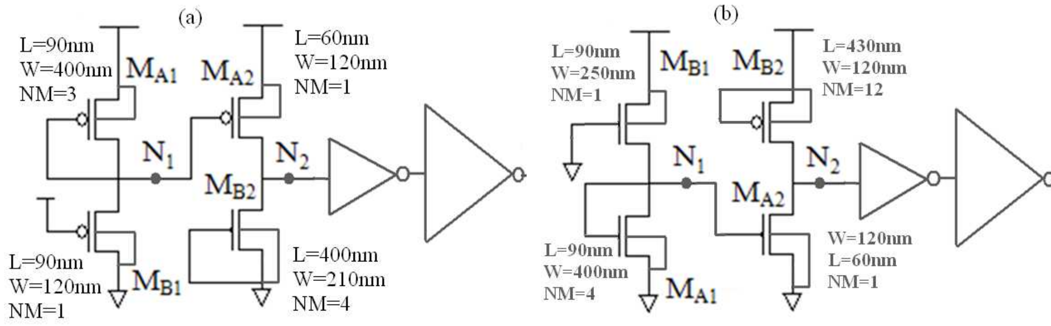

Fig. 1. Body Bias generators for a) PMOS network and b) NMOS network

Authors in [11] [12] also studied the capability of adaptive body-bias techniques to address these variations and implemented a sub-threshold processor to show its effectiveness. They also proved that a body bias which optimises P/N ratio for noise margin also minimises energy per instruction [13].

In this paper, the challenge of effective mitigation of PVT variations is addressed. An extreme process variation sensitive and ultra-low power (SULP) forward body bias (FBB) circuit is proposed capable of addressing PVT variations while V_{DD} scaling happens by tuning P to N ratios in different process corners. This technique help improve the system performance by:

- Applying FBB to both NMOS and PMOS networks at lower voltages and/or lower temperatures
- Applying appropriate FBB to slower devices depending on the process variation

On the other hand, this technique enormously improves delay variation and hence performance yield. In addition, compared to a zero body biased (ZBB) system, this technique also improves the energy delay product (EDP) of the system to which the FBB is applied. Simplicity of the design and also its low power operation incurs low energy and area overhead to the static CMOS system this FBB has been applied to. Results prove that the exponential sensitivity of devices to variations in subthreshold voltages can still be exploited to an extent that can compensate the diminishing FBB effectiveness caused by technology scaling.

The rest of this paper describes how SULP FBB circuit can achieve this by firstly mathematically analysing how circuit works and why it maintains the EDP of the system in section 2, and then, in section 3, providing the HSPICE simulation results obtained from post-layout Monte Carlo runs to back the theoretical findings. Some well known previous works are also explored for comparison purposes.

2. THEORETICAL ANALYSIS OF THE PROPOSED CIRCUIT

2.1. *Proposed Forward Body Bias Circuit*

Fig. 1 shows the schematic of the proposed SULP FBB generator. This circuit is used to apply FBB to the MOS devices so that the EDP of system is improved and PVT variations are addressed while system is working in subthreshold voltages.

Here, focus is on PBB and as NMOS network body bias (NBB) generator works the same way PMOS network body bias (PBB) generator does, discussions and results are valid for NBB too.

This technique comprises two main stages plus process variation independent buffers at the output (which are basically two inverters with large length and width sized devices).

First stage (voltage reference generator) provides M_{A2} with a process/temperature independent V_{gs} proportional to V_{DD}, that is, as V_{DD} increases, V_{gs} also increases correspondingly. In second stage, with a feature size, M_{A2} becomes a PVT dependent device, and with a large channel length and width, M_{B2} becomes PVT invariant. Sizes can be tuned to determine when, in the buffered output, V_{N2} should cause an FBB (or a V_{SS} voltage applied to PMOS network and/or a V_{DD} voltage applied to NMOS network, here T=25°C and V_{DD}=0.4V lead to an FBB). Increasing supply voltage and subsequently V_{gs} of M_{A2}, raises V_{N2} which in turn cancels FBB at superthreshold domain. As M_{B2} and first stage are process/temperature variation independent, a fast PMOS network or a high temperature at a subthreshold domain leads to a leaky and hence strong M_{A2} which withdraws the FBB as it is no longer needed.

Exact sizing of buffers can be achieved by both simulations and formulations (through equation (4)). As channel length has a major role in handling process variations [14-15], M_{B2} is also sized to have a small threshold voltage deviation when facing process variations (and as will be seen to satisfy equation (10)). Multipliers are used to make transistors M_{A1} and M_{B2} leaky and therefore strong in subthreshold voltages (obviously making a leakier PMOS, with a lower mobility than an NMOS, needs more multiplications). For example, a strong M_{A1} in subthreshold will reduce V_{gs} of M_{A2} which in turn leads to a weak M_{A2} and enables now a stronger M_{B2} to be able to reduce V_{N2} and causes an FBB in subthreshold domain.

Considering the fact that all devices of these two stages are always kept off, the large channel length in M_{B2} together with large number of multipliers will have no effect on the performance of the overall circuit and only bring about an area overhead whose cost will be discussed.

If a device is operating in subthreshold region, its drain source current can be modelled by the well known equation [16]:

$$I_{DS} = I_0 e^{\frac{V_{GS}-V_{th}+\gamma V_{DS}}{m\,v_T}} \left(1 - e^{-\frac{V_{DS}}{v_T}}\right) \quad where$$
$$I_0 = \mu \frac{W}{L} \sqrt{\frac{q\epsilon_{si} NDEP}{2\Phi_s}}\, v_T^2 \tag{1}$$

where V_{th} signifies the threshold voltage, γ the body bias coefficient and m the slope factor of transistor. For I_0, μ is the mobility, W the width, L the length, Φ_s the surface potential, v_T the thermal voltage ($=k_BT/q$), $NDEP$ the doping concentration in the channel and ε_{si} the permittivity of silicon for this device.

Transistors in first and second stages of these body bias generators are always kept off and hence the current passing through them is a subthreshold current. Focusing on the first stage, which is a reference voltage generator [17] and using (1), following equation equals the subthreshold currents formulae of M_{A1} and M_{B1} to work out the voltage reference V_{N1} [16].

$$I_{0B1}e^{\frac{(V_{N1}-V_{DD})-V_{thB1}+\gamma V_{N1}}{m_{B1}v_T}}\left(1-e^{-\frac{V_{N1}}{v_T}}\right) = I_{0A1}\mathrm{e}^{\frac{V_{DD}-V_{N1}-V_{thA1}}{m_{A1}v_T}}(1-\mathrm{e}^{-\frac{V_{DD}-V_{N1}}{v_T}}) \quad (2)$$

γ is only applicable to M_{B1}, as it is the only transistor with the body forwarded to ground.

If both M_{A1} and M_{B1} are set to be PMOS transistors with just different widths and multipliers, then (2) is reduced to:

$$W_B e^{\frac{(V_{N1}-V_{DD})+\gamma V_{N1}}{m_p v_T}}\left(1-e^{-\frac{V_{N1}}{v_T}}\right) = mul_A W_A e^{\frac{V_{DD}-V_{N1}}{m_p v_T}}\left(1-e^{-\frac{V_{DD}-V_{N1}}{v_T}}\right) \quad (3)$$

in which mul_A is number of multipliers in device M_{A1}. This shows that process variations, mainly affecting channel length and V_{th}, do not control the reference voltage V_{N1}. This reference voltage can also be designed to be temperature variation independent as stated in [17]. Considering authors' suggestion in [17] to use near-zero threshold voltage MOSFETs for M_{B1} to keep it in the weak inversion mode at negative V_{gs} (in here equal to $V_{N1} - V_{DD}$), an FBB has been applied to this transistor to reduce its threshold voltage as much as possible using a nominal threshold voltage MOSFET device instead of a low-threshold voltage device.

The main well-known drawback of using low-threshold devices (or a multi-V_{th} process) is the incurred overhead in process costs caused by extra masks and steps. Besides, two nominal threshold voltage MOSFETs minimise the process variation effect on reference voltage. Two different threshold voltage MOSFETs have different processes which can result in uncertainty in the output reference voltage. Although applying the FBB eliminates this cost and uncertainty, a PMOS (NMOS) device biased to V_{SS} (V_{DD}) needs a separate n-well (deep n-well) which leads to about 6x area overhead in this body bias generator cell (Note that body biasing needs triple-well process, which is optional for 90nm CMOS process but required for 65nm and lower processes). Nevertheless, this cost is reduced as the system size is raised because the current needed for body biasing is very insignificant compared to the current drained from power supply and hence this generator can provide body bias for the whole design and therefore is employed only once. For example, in the examined 8-bit Kogge-Stone adder, the area overhead has been 14% compared to a ZBB implementation. In an implementation of a 16-bit precision fixed point radix 2 discrete Fourier transform (DFT) core with transform size 1024 and pipelined with 16 complex words per cycle, the FBB circuit overhead reduces to as little as ~0.001%.

As process variations affect many parameters such as channel length and device threshold, it is necessary to inspect how final buffered output is going to overcome these variations. It is usually assumed that all process variations can be abstracted in V_{th} variation [18]. By considering this fact and through sizing, the buffer stages of PBB circuit (two inverters that follow the second stage) can be designed so that:

$$VBSP = \begin{cases} 0 & \frac{V_{DD}}{2} \le VN2 < V_{DD} \ (or\ VTHP < V_{thp0}) \\ V_{DD} & 0 < VN2 \le \frac{V_{DD}}{2} \ (or\ VTHP \ge V_{thp0}) \end{cases} \quad (4)$$

where *VBSP* is a random variable representing V_{BSP} that is the body bias voltage applied to PMOS network (potential difference between V_{DD} and final buffered output).V_{thp0} is a specific threshold voltage of M_{A2} in which V_{N2} is equal to $V_{DD}/2$. *VN2* and *VTHP* also represent random variables for variable V_{N2} (output voltage of the second stage) and variable V_{thp} (threshold voltage of M_{A2}), respectively. Note that larger buffers can be used to form larger fanouts, if required, but it is always the first two buffers that determine the final output voltage. This is because, V_{N2} is converted to V_{SS} or V_{DD} through first two buffers and therefore the final output will be determined before being amplified any further and besides even huge circuits' body biases can be driven by first two buffers if designed properly because a small proportion of current is needed compared to power supplying. So far just *VTHP* has a known probability density function (PDF) as variations in V_{thp} presumably follow a Normal Distribution.

It is obvious that V_{BSP} has a PDF as follows:

$$f_{VBSP}(V_{BSP}) = \begin{cases} p & V_{BSP} = V_{DD} \\ 0 & V_{BSP} = 0 \end{cases} \tag{5}$$

where p is the probability of $V_{BSP}=V_{DD}$ or $V_{N2}\leq V_{DD}/2$. If defined $VBSP=V_{DD}X_B$, it can be seen that X_B has a Bernoulli distribution. Hence:

$$\mu_{VBSP} = V_{DD}\mu_{X_B} = V_{DD}p = V_{DD}P_{VN2}\left(VN2 \leq \frac{V_{DD}}{2}\right) \tag{6}$$

The first stage reference voltage (V_{N1}) is designed to be very close to V_{DD} so that if applied to PMOS transistor M_{A2}, it keeps this transistor off but at the same time increases its subthreshold leakage current needed for creating the reference voltage. A similar equation to (3) can again be formed for the second stage:

$$\begin{aligned} &I_{0p}e^{\frac{V_{DD}-V_{N1}-V_{thp}}{m_p v_T}}\left(1-e^{-\frac{V_{DD}-V_{N2}}{v_T}}\right) \\ &= mul_n\, I_{0n}e^{\frac{-V_{thn}}{m_n v_T}}\left(1-e^{\frac{-V_{N2}}{v_T}}\right) \end{aligned} \tag{7}$$

V_{thp0} can be found by letting $V_{N2} \leq \frac{V_{DD}}{2}$ in (7):

$$\begin{aligned} &mul_n\, I_{0n}e^{\frac{-V_{thn}}{m_n v_T}}\left(1-e^{-\frac{\frac{V_{DD}}{2}}{v_T}}\right) \\ &\geq mul_n\, I_{0n}e^{\frac{-V_{thn}}{m_n v_T}}\left(1-e^{\frac{-V_{N2}}{v_T}}\right) \\ &= I_{0p}e^{\frac{V_{DD}-V_{N1}-V_{thp}}{m_p v_T}}\left(1-e^{-\frac{V_{DD}-V_{N2}}{v_T}}\right) \\ &\geq I_{0p}e^{\frac{V_{DD}-V_{N1}-V_{thp}}{m_p v_T}}\left(1-e^{-\frac{V_{DD}-\frac{V_{DD}}{2}}{v_T}}\right) \end{aligned} \tag{8}$$

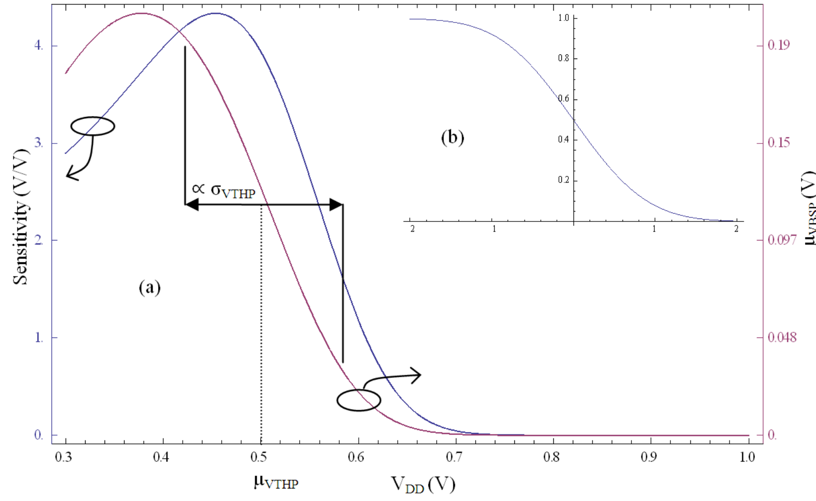

Fig. 2.a) μ_{VBSP} and its sensitivity to μ_{VTHP} variations (at μ_{VTHP}~0.5V) for m_p=1.7, m_n=1.48, v_T=0.026V, σ_{VTHP}=0.04V, μ_{VTHP}=0.5V, μ_{VTHN}=0.45V, $T=25^oC$, and multipliers and sizes of Fig. 1. b) ½ *Erfc(x)*

Solving with respect to V_{thp}, (8) summarises in

$$V_{thp} \geq V_{DD} - V_{N1} + \frac{V_{thn}\, m_p}{m_n} + m_p v_T \ln\left(\frac{I_{0p}}{I_{0n}\, mul_n}\right) \tag{9}$$

Sized to its feature size, M_{A2} is prone to process variations while M_{B2} is sized large enough (especially in channel length) to be resistant to these variations [14-15]–here standard deviation of V_{THN} decreased to 0.01V while V_{THP} standard deviation is 0.04V. By exploiting this fact and defining

$$V_{thp0} = V_{DD} - V_{N1} + \frac{\mu_{VTHN} m_p}{m_n} + m_p v_T \ln\left(\frac{I_{0p}}{I_{0n}\, mul_n}\right) \tag{10}$$

as the PDF of *VTHP* is already known, then the expected value of *VBSP* can be simply found by:

$$\begin{aligned} \mu_{VBSP} &= V_{DD} P_{VOUT}\left(VN2 \leq \frac{V_{DD}}{2}\right) \\ &= V_{DD} P_{VTHP}\left(VTHP \geq V_{thp0}\right) \\ &= \frac{V_{DD}}{2} Erfc\left(\frac{V_{thp0} - \mu_{VTHP}}{\sqrt{2\sigma_{VTHP}^2}}\right) \end{aligned} \tag{11}$$

in which *Erfc(x)* is the complementary error function (see Fig. 2.b).

For NBB circuit, similarly, V_{thn0} can be defined $V_{N1} + \frac{\mu_{VTHP} m_n}{m_p} + m_n v_T\, ln\left(\frac{I_{0n}}{I_{0p}\, mul_p}\right)$ and therefore:

$$\mu_{VBSN} = \frac{V_{DD}}{2} Erfc\left(\frac{V_{thn0} - \mu_{VTHN}}{\sqrt{2\sigma_{VTHN}^2}}\right) \tag{12}$$

To explain how this circuit addresses PVT variations, (11) shows that V_{BSP} follows

variations occurring in transistor M_{A2}. For example, if M_{A2} is a fast (leaky) transistor, which makes μ_{VTHP} lower than its nominal value of V_{THP}, then μ_{VBSP} reduces which means it becomes less likelier for FBB to be applied in these circumstances (Fig. 2.b). This simple example, in fact, illustrates how this circuit copes with process variations. Fig. 2.a shows that moving toward superthreshold voltages reduces μ_{VBSP} to zero leading to once again FBB cancellation. In subthreshold voltages, however, mean of V_{BSP} for PMOS network tends toward V_{DD} (and thus as V_{DD} reduces, mean value of V_{BSP} also decreases). This also describes how SULP FBB handles the voltage variations as well as voltage scaling. And finally, because V_{N1} in (10) is temperature independent, it can be seen that rising temperature increases V_{thp0} and thus reduce μ_{VTHP} (Fig. 2.b). This means in higher temperatures and subthreshold voltages, where transistors become faster, SULP FBB generator starts cancelling FBB to prevent more leakage energy consumption.

FBB cancellation in higher voltages also suggests that, when using this technique, system should be placed in standby mode at higher voltages which, as a result, incurs no FBB and thus no leakage (neither by subthreshold leakage because of subthreshold voltage application nor by means of FBB application). It should be noted that SULP FBB behaviour, shown in Fig. 2, is dependent on V_{thp0}, hence it can be altered by tuning parameters in (10) (for example, in case of experiencing very slow devices, a higher FBB mean values might be needed which can be achieved by adding $mult_n$ or number of multipliers in M_{B2}, which decreases V_{thp0}. This way, SULP FBB circuit is tuned so that *x*-axis interval of Fig. 2.a ranges from nearthreshold (instead of subthreshold) voltages to superthreshold ones). This means SULP FBB can be exploited in variety of systems. Furthermore, when number of multipliers is set to create very asymmetrical and uneven shapes (e.g. a gate with very big NMOS devices and very small PMOS ones), this can be mitigated by skewed buffers.

Another useful observation from (11) is that if variance of process variations (σ^2_{VTHP}) increases, which means for example if system is implemented by narrower channel technologies, then the chance of FBB being applied in higher voltages increases too, which is a reasonable action to take.

Sensitivity of this circuit to process variations is defined by the rate of V_{N2} variation when μ_{VTHP} varies [19] and can be calculated by differentiating (11) with respect to μ_{VTHP} which results in:

$$\frac{\partial \mu_{VBSP}}{\partial \mu_{VTHP}} = \frac{V_{DD}}{\sqrt{2\pi\sigma^2_{VTHP}}} e^{\left(\frac{V_{thp0}-\mu_{VTHP}}{2\sigma^2_{VTHP}}\right)^2} \tag{13}$$

Equation (13) has been sketched for μ_{VTHP}≈0.5V in Fig. 2.a which maximises to ~4.3V/V on V_{DD}≈0.45V. This sensitivity is directly proportional to threshold voltage. For example, at μ_{VTHP}=0.6V, maximum sensitivity of ~6V/V is obtained at V_{DD}≈0.64. Once again, sensitivity to process variations is cancelled at superthreshold voltages. Sensitivity to process variations at nearthreshold voltages in SULP FBB circuit equals to the latest sensitivity that has so far been achieved (for superthreshold applications) in [19]. Likewise [19], SULP FBB is also a bias free circuit as its first stage provides a reference voltage which is process variation independent.

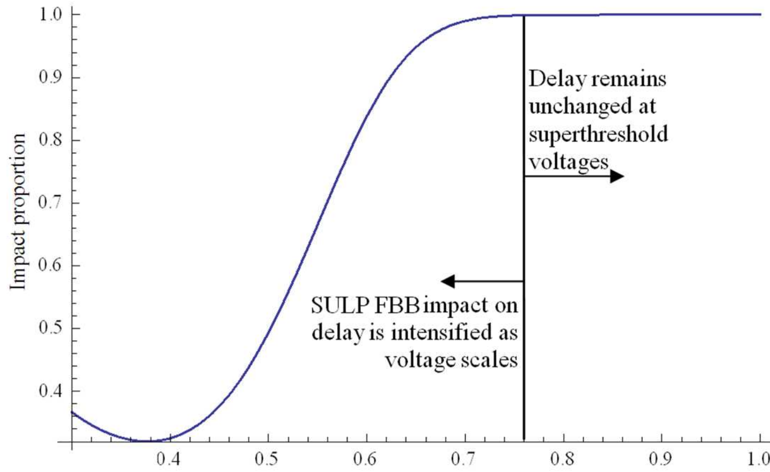

Fig. 3. $f_{Dp1}(V_{thp0})$ for m_p=1.7, m_n=1.48, v_T=0.026V, σ_{VTHP}=0.04V, μ_{VTHP}=0.518V, μ_{VTHN}=0.45V and multipliers and sizes of Fig. 1.

2.2.Effect of SULP FBB circuit on energy and delay

This sub-section focuses on application of this FBB to a typical inverter in a data-path. The goal of this section is to find the effect of FBB on delay and energy in terms of improvement or degradations and not calculating their exact amount. As FBB is cancelled at superthreshold voltages, it can be assumed that no improvement or degradations happens while operating in superthreshold region. Therefore, results of the following equations are valid for superthreshold voltages as long as there is no impact on energy and delay at superthreshold voltages, which will be proved true later on. When output of an inverter in subthreshold region switches from 0 to 1, PMOS is the device which is playing the main role in the delay of the gate, therefore it can be written [18]:

$$t_d = \frac{\frac{1}{2}\eta C_S V_{DD}}{I_{on_p}} \quad where \ I_{on_p} = I_{0p} e^{\frac{V_{DD}-V_{thp}+\gamma V_{BSP}}{m_p v_T}} \left(1 - e^{-\frac{V_{DD}}{v_T}}\right) \tag{14}$$

where η is the delay factor of the inverter non-step input, C_S the inverter switching load capacitance and I_{on_p} is the current passing through PMOS when it is switched on.

A mathematical method for analysis of the energy and EDP of the proposed circuit has been developed and introduced in the Appendix which can further be used for analysis of any other circuits under PVT variations.

The mean value of delay can be worked out as follows: (for details of derivations refer to (19) and (20) in subsection B of Appendix in which, by assuming a Normal PDF for V_{thp}, equation (14) is used to calculate the mean value):

$$E_{T_d}(t_d) = \mu_{T_{dZBB}} . f_{Dp1}(V_{thp0}) \tag{15}$$

where μ_{TdZBB} is the mean value of the delay of the inverter when ZBB is applied and $f_{Dp1}(V_{thp0})$ is the impact of SULP FBB on the inverter *mean* delay (refer to (18) subsection A of Appendix for its definition). Fig. 3 shows how $f_{Dp1}(V_{thp0})$ influences on delay mean

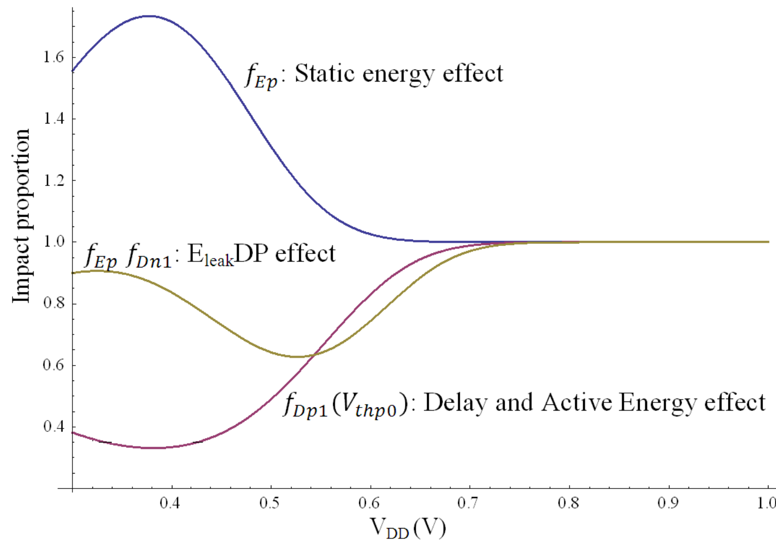

Fig. 4. Effect of SULP FBB on energy and delay of the examined inverter for m_p=1.7, m_n=1.48, v_T=0.026V, σ_{VTHN}=0.034V, μ_{VTHP}=0.518V, μ_{VTHN}=0.493V and multipliers and sizes of Fig. 1.

value.

As Fig. 3 implies, in subthreshold voltages when V_{DD} and hence V_{thp0} are small enough, delay reduces. As V_{DD} rises, impact of FBB is dropped and, as a result, delay in higher voltages is no longer affected by this technique.

The rest of this subsection explores how EDP alters after SULP FBB technique is put into practice.

Before determining EDP equations, let's formulate the equation of leakage current for a stable inverter gate (an inverter that is not switching). If it is assumed that PMOS device is off and NMOS on, then PMOS will have a leaking current with following current:

$$I_{leak_p} \approx I_{0p} e^{\frac{-V_{thp}+\gamma V_{BSP}}{m_p v_T}} \; (assuming\; 1 - e^{-\frac{V_{DD}}{v_T}} \approx 1) \tag{16}$$

where V_{BSP} is the body bias applied to PMOS network.

In subsection C of Appendix, (16) is used for leakage energy calculation and $\frac{1}{2}\alpha C_S V_{DD}^2$ for active energy, and with t_d, using (14), altogether results in EDP to be formulated thoroughly. Following equation shows the impact of SULP FBB on EDP for an inverter in a data-path (for details refer to subsection C of Appendix). It has been assumed in (17) that the expected value of EDP when ZBB is applied equals A+B+C+D+E (see (23) in subsection C of Appendix for extended version) where B and C are related to leakage energy-delay product and D and E are related to active energy-delay product:

$$\begin{aligned} & E_{VTHN,VTHP}(EDP_{SULP}) \\ & = A + B\, f_{Ep}(V_{thp0}) f_{Dn1}(V_{thn0}) \\ & + C f_{En}(V_{thn0})\, f_{Dp1}(V_{thp0}) + D f_{Dp1}(V_{thp0}) \\ & + E f_{Dn1}(V_{thn0}) \end{aligned} \tag{17}$$

Fig. 4 shows how $f_{Ep}(V_{thp0})\, f_{Dn1}(V_{thn0})$ in (17) changes leakage energy-delay product as

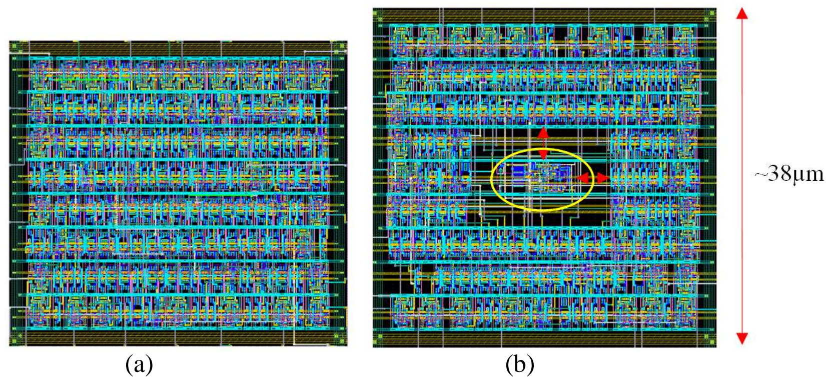

Fig. 5. Layout drawing for the 8-bit Kogge-Stone adder with a) ZBB design and b) SULP FBB generator located inside the yellow ellipse.

V_{DD} changes. In fact, $f_{Ep}(V_{thp0})$ in (17) shows the impact of SULP FBB on static energy, which as expected, causes static energy to rise at subthreshold voltages (Fig. 4). It should be noticed that $f_{Dp1}(V_{thp0})$ for active energy acts the same as Fig. 3. Fig. 4 also shows that the overall EDP remains unchanged (or even reduces) as V_{DD} scales to subthreshold voltages that makes this technique not only beneficial in terms of performance increase but also conservative regarding the higher consumed energy. Moreover, EDP stays unchanged in superthreshold domain, too, because of FBB withdrawal and the near zero energy overhead of SULP FBB circuit. These results are proved by simulations as discussed in next section.

3. Simulation results and discussions

In this section, previously obtained equations are examined through simulations to verify their predictions and later on some prominent previous researches are investigated to validate the novelty and illuminate the contribution made in this work.

3.1.Simulations

A pipelined 8-bit Kogge-Stone adder was chosen to verify the SULP FBB technique. All simulations were performed using Low Power 65nm TSMC technology model. 1000 Monte Carlo (MC) runs were executed on this adder to simulate the process variations. In addition, 100 MC post-layout simulations were also performed to verify the results. For this purpose, a Standard Cell library was created with PBB and NBB connections. Then Design Compiler (Synopsys synthesis tool) and IC Compiler (Synopsys place and route tool) were utilised to create the final GDSII file. The temperature sensitive RC extraction was performed on this file to obtain the final post-layout HSPICE code.

Because one main duty of the SULP FBB technique is to address inter-die process variations by applying the appropriate body bias to the whole system, the foundry provided global process variation model was exploited to produce the most accurate possible results. This model contained all foundry proven Gaussian distributions which lead to the most practical and realistic Monte Carlo simulation results. Temperature and voltage were also swept to resemble both temperature variations and V_{DD} scaling. In this study, temperature varies from -15 to 75 (90°C variation) and if higher or lower working temperatures are required then SULP FBB circuit has to be simply tuned to support this.

Fig. 5 shows the layouts of the above mentioned adder in two different schemes. Fig. 5.a

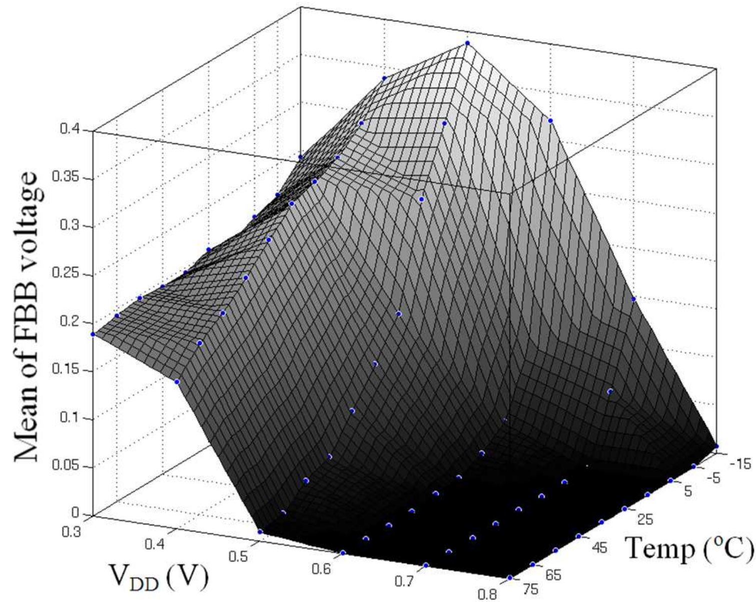

Fig. 7. Simulation results for mean of PBB circuit output in different temperatures and voltages

has a ZBB implementation while Fig. 5.b has the SULP FBB generator in the heart of system in which red arrows indicate the required space between two deep n-well layers (FBB generator has its own deep n-well layer for the V_{DD} biased NMOS while the rest of NMOS devices in the chip are biased to the generated FBB).

Fig. 6 exhibits the mean value of FBB voltage applied to PMOS network. Referring to Fig. 2.a, which was sketched for T=25°C, it can be seen that Fig. 6 actually follows the prediction made by (11) showing its usefulness as a model for studying SULP FBB behaviour.

Adder delay was also recorded before (ZBB) and after (SULP FBB) body bias application during MC runs and temperature/voltage sweeps. Fig. 7 is the outcome that depicts the impact of the proposed technique on the delay of the examined adder. This also acknowledges Fig. 3 that predicted the same behaviour by mathematical analysis.

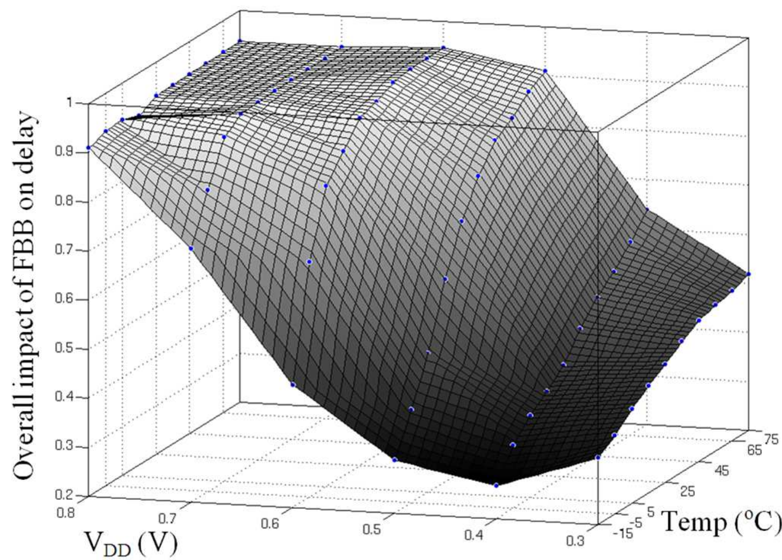

Fig. 6. Impact on ZBB Mean Delay after introduction of the SULP FBB technique.

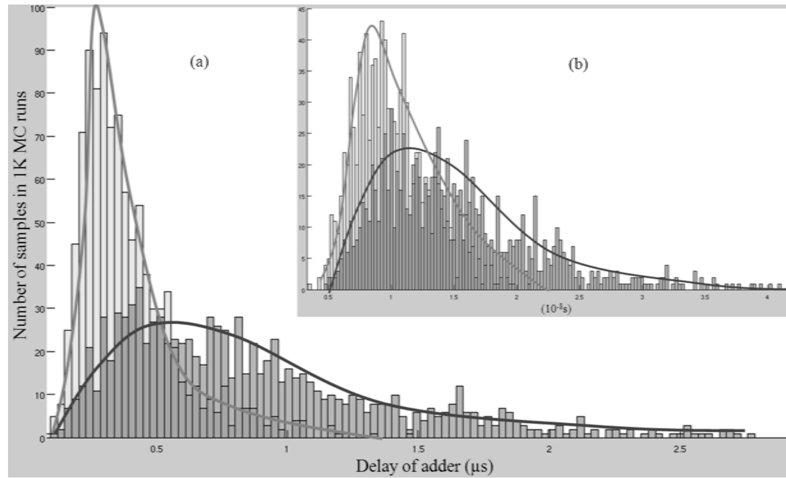

Fig. 9. Variations in Delay resulting from 1K Monte Carlo simulations for SULP FBB and ZBB cases for T=25°C and a) V_{DD}=0.3V b) V_{DD}=0.5V. Lighter bars represent SULP FBB and darker ones signify ZBB case.

MC simulations on pre and post-layout HSPICE codes proved negligible difference too, in terms of circuit delay and energy consumption. For higher accuracy gained by higher number of MC iterations, Fig. 8 shows the distribution of delays in the adder resultant from pre-layout 1K MC runs for the ZBB and the SULP FBB cases. If sketched with respect to t_d (delay of the examined inverter in section 2), the PDF of the analysed inverter showed in (21) (refer to subsection B) would be again similar to Fig. 8. Equation (21) also specifies that a Lognormal Distribution has to be expected for delay as it is obvious in Fig. 8.a. Fig. 8.b also shows that, as (22) stated, increasing voltage makes the SULP FBB ineffective and therefore impact on variance gradually starts to subside. This, in return, makes SULP FBB similar to ZBB case at higher voltages.

Fig. 9.a examines the energy overhead of the SULP FBB circuit. In average, body bias

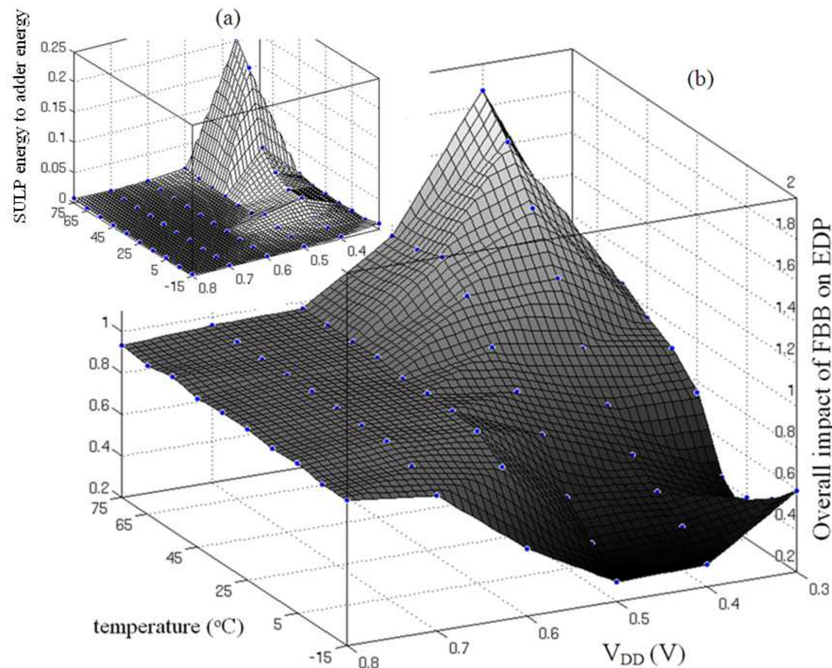

Fig. 8. a) SULP Body bias generator's total energy to adder's total energy. b) Overall EDP effect for SULP FBB technique with respect to ZBB case.

generator has about 1% energy overhead across the temperature and voltage range for this adder and with more complicated systems this overhead (and especially area overhead of 14%) will become negligible. And finally, Fig. 9.b shows the impact of SULP FBB on EDP of the adder in which the overall (static and dynamic as well as adder and SULP FBB circuit) energy consumption has been measured. Comparing Fig. 9 and Fig. 4 it can be seen that Fig. 9 in 25°C reacts almost the same as Fig. 4 to voltage scaling. This is as a result of the effect of leakage energy increase at subthreshold voltages which is severer than what active energy reduction can compensate. Fig. 9 suggests that although this technique in average improves the EDP, subthreshold voltages together with high temperatures should not dominate the typical working condition of the system in which this technique resides.

At high frequencies, which can only be achieved at superthreshold domain, a rise in temperature is expected which has no effect on EDP as SULP FBB is not exerted at this domain. On the other hand, with larger circuits, and therefore more transistors, consumed energy and hence temperature may undergo a rise depending on voltage domain. However, because of the low frequency and low power supply voltage at subthreshold domain, dynamic power dramatically falls, and leakage energy will become the only concern. If the huge number of transistors in the design cannot be evaded, this concern can be settled by remaining at near-threshold voltages [13]. Nevertheless, at subthreshold supply voltages, high leakage is usually an issue regarding the energy consumed not the heat generated, while at superthreshold domain, switching energy at very high frequencies can bring about a considerable temperature rise. As a result at subthreshold domains, unless there is an ambient temperature present (or a packaging or an environment with low heat transfer), a chip with SULP FBB technique (or without it) will not by itself generate that amount of heat leading to 45°C temperature or over. Thus as mentioned before, the ambient temperature beyond 45°C should be avoided as a long term practice, while at subthreshold domain, or else the static energy consumption will dominate and the SULP FBB benefit will not be justified anymore.

Table I compares the result of post-layout simulations for ZBB and SULP FBB cases measured across 0.3V to 0.8V (raising voltage above 0.8V renders no consequences in terms of SULP FBB functionality as FBB will be withdrawn above 0.8V in this design and accordingly all achieved results will be identical). Improvements are calculated based on the ratio of the measured parameter after SULP FBB implementation to before SULP FBB implementation (or ZBB). Delay is measured for the critical path (average of rise and fall times), and so is frequency (maximum of fall and rise times reversed), energy using the integral of power over the period of simulation, and delay variations (measured in pre-layout simulations) based on 1K Monte Carlo runs. Improvement of delay variations by 7x to 9x, depending on temperature interval and when EDP is reduced, indicates the power of this technique in addressing PVT variations as well as voltage scaling in a very efficient and well adaptive way.

It has to be mentioned that applying absolute FBB in all conditions, no matter what PVT situation the circuit is in, brings about a very leaky system which delivers 7x less delay compared to SULP FBB technique while consuming 100x more energy! It has to be noticed too that (at superthreshold voltages when performance is high enough to be sacrificed for reducing the leakage power dissipation) reverse body bias can be applied (instead of ZBB) by simply changing the buffering stages to voltage converters.

3.2. Comparison with previous works

As it was shown, SULP FBB generator had many advantages such as:

- High sensitivity to process variations at subthreshold voltages
- Low energy consumption
- Ability to cancel FBB in superthreshold voltages to prevent system from consuming more power when it is not necessary
- Sensitivity to PVT variations only in subthreshold voltages

Authors in [5] [20], as was discussed, studied adjusting the P/N ratio for optimal noise margin which can also result in minimum energy per instruction. However, in high temperatures and subthreshold voltages, transistors become so leaky that it is very damaging, in terms of total energy consumption and not energy per instruction, to provide the optimal noise margin P/N body bias. For example, a simple simulation on an inverter, with PMOS device sized twice the NMOS device, V_{DD}=0.4V and T=75°C, shows that an optimal FBB applied on PMOS improves noise margin by only ~1% while increases static current by 10%. Furthermore, delay is very sensitive to body biasing, at subthreshold voltages, which helps prevent timing violations at extreme voltage scaling. This, therefore, means that a non-optimal FBB is needed to overcome extreme voltage scaling situations. For instance, in the same inverter with V_{DD}=0.3V and T=25°C, when the non-optimal FBB of 0V is applied to PMOS, delay is improved by 2.23x while noise margin worsens by 1.17x when compared to the optimal FBB. On the other hand, noise margin for data-path is not of a primary concern at superthreshold voltages which means forward body biasing can be cancelled to prevent any extra energy consumption regardless of optimal P/N ratio. Unlike [5], SULP FBB is cancelled at high temperatures, FF corners, superthreshold voltages and is fully applied when at subthreshold voltages.

The problem of [9] is, though, there is always an FBB in subthreshold for either NMOS or PMOS network which makes the circuit leaky when at TT, FF, or SS corners. Moreover their circuit uses large switches to disconnect the circuit from delivering FBB on superthreshold voltages which means both controlling and area, not to mention energy, overhead. In addition to being low power, SULP FBB technique also does not need any controller as PVT variations plus voltage scaling are automatically sensed.

Because of complexity, the body biasing circuit used in [10] incurs energy overhead on the chip and needs to be disabled on superthreshold while the proposed body biasing circuit works in subthreshold region (except the buffer parts which only consume static power) for any V_{DD} and hence imposes almost no power overhead as Fig. 9 states.

There are techniques that manage V_{TH} and V_{DD} at the same time to minimise the energy consumed by means of algorithms that identify the appropriate V_{DD} and V_{TH} depending on work load. Sometimes V_{TH} adaptive biasing is used to gain the minimum energy as it can aid to reduce leakage in lower supply voltages (but variations are forgotten) [3] and sometimes it is utilised to compensate PVT variations while scaling is missing [21]. There are cases that claim to handle both at the same time but they use very complicated architectures only suitable for high performance GIPS processors [22].

As mentioned earlier, random intra-die variations do not have spatial correlation. Hence it cannot be addressed by SULP FBB unless a calibration method is exploited [19]. These random variations, however, are damaging in SRAM rather than data-path architectures [19].

4. CONCLUSION

As discussed, when energy consumption is a concern, voltage scaling toward subthreshold voltages has become a well-known technique which brings many challenges with for designers. Addressing PVT variations, as an important challenge, has been experienced through body bias application techniques in many previous works. Here, however, a very high resistance to PVT variations could be gained in subthreshold voltages using an ultra-sensitive novel body biasing technique that automatically withdraws the body bias in superthreshold voltages. By applying FBB when necessary, this technique reduced timing violations by means of ~7x less delay variation compared to a ZBB case across all PVT variations examined in here while decreasing EDP. The low cost implementation of body biasing circuit using nominal voltage threshold MOSFETS and its ultra low power consumption offset its area overhead which will be negligible in bigger circuits anyway. Results proved the proposed FBB effectiveness for subthreshold voltages despite the decline in impact of FBB in short-channel devices.

II. APPENDIX

A. *Definitions:*

Some definitions are brought for simplicity:

$$\begin{aligned} f_{Dx\alpha}(V_{thx0}) &= \frac{1}{2}\left(1 - Erf\left(\frac{\alpha\sigma_{VTHX}}{\sqrt{2}m_x v_T} + \frac{\mu_{VTHX} - V_{thx0}}{\sqrt{2\,\sigma_{VTHX}^2}}\right)\right) + \frac{e^{\frac{-\alpha\gamma V_{DD}}{m_x v_T}}}{2}\left(1 + Erf\left(\frac{\alpha\sigma_{VTHX}}{\sqrt{2}m_x v_T} + \frac{\mu_{VTHX} - V_{thx0}}{\sqrt{2\,\sigma_{VTHX}^2}}\right)\right) \\ f_{Ex}(V_{thx0}) &= \frac{1}{2}\left(1 + Erf\left(\frac{\sigma_{VTHX}}{\sqrt{2}m_x v_T} - \frac{\mu_{VTHX} - V_{thx0}}{\sqrt{2\,\sigma_{VTHX}^2}}\right)\right) + \frac{e^{\frac{\gamma V_{DD}}{m_x v_T}}}{2}\left(1 - Erf\left(\frac{\sigma_{VTHX}}{\sqrt{2}m_x v_T} - \frac{\mu_{VTHX} - V_{thx0}}{\sqrt{2\,\sigma_{VTHX}^2}}\right)\right) \end{aligned} \tag{18}$$

Note that $f_{Dx\alpha}$ and f_{Ex} are equal to 1 at superthreshold voltages.

B. *Mean Value and Variance of Delay in a typical inverter:*

Equation (14) can be rewritten as:

$$t_d = D_0 e^{\frac{V_{thp} - \gamma V_{BSP}}{m_p v_T}} \quad where\ D_0 = \frac{\frac{1}{2}\eta C_s V_{DD}}{I_{0p} e^{\frac{V_{DD}}{m_p v_T}}\left(1 - e^{-\frac{V_{DD}}{v_T}}\right)} \tag{19}$$

Using (19), the expected value of the random variable T_d can be sought as following:

$$
\begin{aligned}
E_{T_d}(t_d) &= D_0 \int_{-\infty}^{\infty} e^{\frac{V_{thp}-\gamma V_{BSP}}{m_p v_T}} f_{VTHP}(V_{thp}) dV_{thp} \\
&= \frac{D_0}{\sqrt{2\pi\sigma_{VTHP}^2}} \left(\int_{-\infty}^{V_{thp0}} e^{\frac{V_{thp}}{m_p v_T}} e^{-\frac{(V_{thp}-\mu_{VTHP})^2}{2\sigma_{VTHP}^2}} dV_{thp} \right. \\
&\quad \left. + e^{\frac{-\gamma V_{DD}}{m_p v_T}} \int_{V_{thp0}}^{\infty} e^{\frac{V_{thp}}{m_p v_T}} e^{-\frac{(V_{thp}-\mu_{VTHP})^2}{2\sigma_{VTHP}^2}} dV_{thp} \right) \\
&= D_0 e^{\frac{\mu_{VTHP}}{m_p v_T}+\frac{\sigma_{VTHP}^2}{2m_p^2 v_T^2}} \left(\frac{1}{2}\left(1 - Erf\left(\frac{\sigma_{VTHP}}{\sqrt{2}m_p v_T} + \frac{\mu_{VTHP} - V_{thp0}}{\sqrt{2\,\sigma_{VTHP}^2}}\right)\right) \right. \\
&\quad \left. + \frac{e^{\frac{-\gamma V_{DD}}{m_p v_T}}}{2} \left(1 + Erf\left(\frac{\sigma_{VTHP}}{\sqrt{2}m_p v_T} + \frac{\mu_{VTHP} - V_{thp0}}{\sqrt{2\,\sigma_{VTHP}^2}}\right)\right)\right) \\
&= \mu_{T_{dZBB}} . f_{Dp1}(V_{thp0})
\end{aligned}
\tag{20}
$$

By means of (19) and with some mathematical derivations, PDF of T_d can be realised too:

$$
\begin{aligned}
PDF(t_d) &= \left(\frac{m_p v_T}{\sqrt{2\pi\sigma_{VTHP}^2}\ t_d} \right) . \\
&\left(e^{-\frac{\left(\lambda V_{DD}+m_p v_T ln\left(\frac{t_d}{D_0}\right)-\mu_{VTHP}\right)^2}{2\sigma_{VTHP}^2}} p + e^{-\frac{\left(m_p v_T ln\left(\frac{t_d}{D_0}\right)-\mu_{VTHP}\right)^2}{2\sigma_{VTHP}^2}} (1-p) \right) \\
PDF(t_{dZBB}) &= \left(\frac{m_p v_T}{\sqrt{2\pi\sigma_{VTHP}^2}\ t_d} \right) e^{-\frac{\left(m_p v_T ln\left(\frac{t_d}{D_0}\right)-\mu_{VTHP}\right)^2}{2\sigma_{VTHP}^2}}
\end{aligned}
\tag{21}
$$

where as shown in (11) $p = \text{Erfc}\left(\frac{V_{thp0}-\mu_{VTHP}}{\sqrt{2\sigma_{VTHP}^2}}\right)$.

Variance of delay can be derived using the following equations:

$$
\begin{aligned}
E_{T_d}(t_d^2) &= D_0^2 \int_{-\infty}^{\infty} e^{2\frac{V_{thp}-\gamma V_{BS}}{m_p v_T}} f_{VTHP}(V_{thp}) dV_{thp} \\
&= \frac{D_0^2}{\sqrt{2\pi\sigma_{VTHP}^2}} \left(\int_{-\infty}^{V_{thp0}} e^{\frac{2V_{thp}}{m_p v_T}} e^{-\frac{(V_{thp}-\mu_{VTHP})^2}{2\sigma_{VTHP}^2}} dV_{thp} \right. \\
&\quad \left. + e^{\frac{-2\gamma V_{DD}}{m_p v_T}} \int_{V_{thp0}}^{\infty} e^{\frac{2V_{thp}}{m_p v_T}} e^{-\frac{(V_{thp}-\mu_{VTHP})^2}{2\sigma_{VTHP}^2}} dV_{thp} \right) \\
&= D_0^2 e^{2\left(\frac{\mu_{VTHP}}{m_p v_T}+\frac{\sigma_{VTHP}^2}{m_p^2 v_T^2}\right)} f_{Dp2}(V_{thp0}) = \mu_{T_{dZBB}^2} . f_{Dp2}(V_{thp0}) \quad hence \quad \sigma_{Td}^2 = \\
&\quad \mu_{T_{dZBB}^2} . f_{Dp2}(V_{thp0}) - \mu_{T_{dZBB}}^2 . f_{Dp1}^2(V_{thp0})
\end{aligned}
\tag{22}
$$

C. Energy Delay Product in a typical inverter:

Following derivations find EDP of an inverter in execution stage of the adder in which it is assumed that the probabilities of a gate input being 0 or 1 are equal to ½. Leakage energy is calculated using (16) and active energy using the well-known expression $\frac{1}{2}\alpha C_s V_{DD}^2$ and t_d using (14):

$$EDP = (E_{act} + E_{leak}).t_d = \left(\frac{1}{2}\alpha C_s V_{DD}^2 + V_{DD}\frac{1}{2}\left(I_{leak_p} + I_{leak_n}\right)t_{total}\right).\eta C_s V_{DD}\left(\frac{1}{I_{on_p}} + \frac{1}{I_{on_n}}\right)$$
$$= \left(\frac{1}{2}\alpha C_s V_{DD}^2 + V_{DD}\frac{1}{2}\left(I_{0p}e^{\frac{-V_{thp}+\gamma V_{BSP}}{m_p v_T}} + I_{0n}e^{\frac{-V_{thn}+\gamma V_{BSN}}{m_n v_T}}\right)t_{total}\right).\eta C_s V_{DD}.\left(\frac{1}{I_{0p}}e^{\frac{-V_{DD}+V_{thp}-\gamma V_{BSp}}{m_p v_T}} + \frac{1}{I_{0n}}e^{\frac{-V_{DD}+V_{thn}-\gamma V_{BSn}}{m_n v_T}}\right)$$

First active energy effect is taken into account:

$$E_{act}DP = \left(\frac{\eta}{2}\alpha C_s^2 V_{DD}^3\right)\left(\frac{1}{I_{0p}}e^{\frac{-V_{DD}+V_{thp}-\gamma V_{BSp}}{m_p v_T}} + \frac{1}{I_{0n}}e^{\frac{-V_{DD}+V_{thn}-\gamma V_{BSn}}{m_n v_T}}\right)E_{VTHN,VTHP}(E_{act}DP)$$
$$= \left(\frac{\eta}{2}\alpha C_s^2 V_{DD}^3\right).\left(\frac{1}{I_{0p}}e^{\frac{-V_{DD}}{m_p v_T}}E_{VTHP}\left(e^{\frac{V_{thp}-\gamma V_{BSp}}{m_p v_T}}\right) + \frac{1}{I_{0n}}e^{\frac{-V_{DD}}{m_n v_T}}E_{VTHN}\left(\frac{1}{I_{0n}}e^{\frac{V_{thn}-\gamma V_{BSn}}{m_n v_T}}\right)\right)$$
$$= \left(\frac{\eta}{2}\alpha C_s^2 V_{DD}^3\right).$$
$$\left(\begin{array}{c}\frac{1}{I_{0p}}e^{-\frac{V_{DD}}{m_p v_T}}e^{\frac{\mu_{VTHP}}{m_p v_T}+\frac{\sigma_{VTHP}^2}{2m_p^2 v_T^2}}.\\ \left\{\frac{e^{\frac{-\gamma V_{DD}}{m_p v_T}}}{2}\left(1+Erf\left(\frac{\sigma_{VTHP}}{\sqrt{2}m_p v_T}+\frac{\mu_{VTHP}-V_{thp0}}{\sqrt{2\,\sigma_{VTHP}^2}}\right)\right)+\frac{1}{2}\left(1-Erf\left(\frac{\sigma_{VTHP}}{\sqrt{2}m_p v_T}+\frac{\mu_{VTHP}-V_{thp0}}{\sqrt{2\,\sigma_{VTHP}^2}}\right)\right)\right\}\\ +\\ \frac{1}{I_{0n}}e^{-\frac{V_{DD}}{m_n v_T}}e^{\frac{\mu_{VTHN}}{m_n v_T}+\frac{\sigma_{VTHN}^2}{2m_n^2 v_T^2}}.\\ \left\{\frac{e^{\frac{-\gamma V_{DD}}{m_n v_T}}}{2}\left(1+Erf\left(\frac{\sigma_{VTHN}}{\sqrt{2}m_n v_T}+\frac{\mu_{VTHN}-V_{thn0}}{\sqrt{2\,\sigma_{VTHN}^2}}\right)\right)+\frac{1}{2}\left(1-Erf\left(\frac{\sigma_{VTHN}}{\sqrt{2}m_n v_T}+\frac{\mu_{VTHN}-V_{thn0}}{\sqrt{2\,\sigma_{VTHN}^2}}\right)\right)\right\}\end{array}\right)$$
$$= \left(\frac{\eta}{2}\alpha C_s^2 V_{DD}^3\right).\left(\frac{1}{I_{0p}}e^{-\frac{V_{DD}}{m_p v_T}}e^{\frac{\mu_{VTHP}}{m_p v_T}+\frac{\sigma_{VTHP}^2}{2m_p^2 v_T^2}}f_{Dp1}(V_{thp0}) + \frac{1}{I_{0n}}e^{-\frac{V_{DD}}{m_n v_T}}e^{\frac{\mu_{VTHN}}{m_n v_T}+\frac{\sigma_{VTHN}^2}{2m_n^2 v_T^2}}f_{Dn1}(V_{thn0})\right)$$

Now using the same method, effect of SULP FBB on leakage energy is calculated:

$$E_{leak}DP = \frac{1}{2}\eta C_s V_{DD}^2.\left(I_{0p}e^{\frac{-V_{thp}+\gamma V_{BSP}}{m_p v_T}} + I_{0n}e^{\frac{-V_{thn}+\gamma V_{BSN}}{m_n v_T}}\right)t_{total}.\left(\frac{1}{I_{0p}}e^{\frac{-V_{DD}+V_{thp}-\gamma V_{BSP}}{m_p v_T}} + \frac{1}{I_{0n}}e^{\frac{-V_{DD}+V_{thn}-\gamma V_{BSN}}{m_n v_T}}\right)$$

$$= \frac{1}{2}\eta C_s V_{DD}^2 t_{total}\left(e^{\frac{-V_{DD}}{m_p v_T}} + e^{\frac{-V_{DD}}{m_n v_T}} + \frac{I_{0p}}{I_{0n}} e^{\frac{-V_{thp}+\gamma V_{BSP}}{m_p v_T}+\frac{-V_{DD}+V_{thn}-\gamma V_{BSN}}{m_n v_T}} + \frac{I_{0n}}{I_{0p}} e^{\frac{-V_{thn}+\gamma V_{BSN}}{m_n v_T}+\frac{-V_{DD}+V_{thp}-\gamma V_{BSP}}{m_p v_T}}\right)$$

As VTHN and VTHP are presumably independent random variables then it can be written:

$$E_{VTHN,VTHP}(E_{leak}DP) = \frac{1}{2}\eta C_s V_{DD}^2 t_{total}\left(e^{\frac{-V_{DD}}{m_p v_T}} + e^{\frac{-V_{DD}}{m_n v_T}} + \frac{I_{0p} e^{\frac{-V_{DD}}{m_n v_T}}}{I_{0n}} E_{VTHP}\left(e^{\frac{-V_{thp}+\gamma V_{BSP}}{m_p v_T}}\right) E_{VTHN}\left(e^{\frac{V_{thn}-\gamma V_{BSN}}{m_n v_T}}\right) + \frac{I_{0n} e^{\frac{-V_{DD}}{m_n v_T}}}{I_{0p}} E_{VTHN}\left(e^{\frac{-V_{thn}+\gamma V_{BSN}}{m_n v_T}}\right) E_{VTHP}\left(e^{\frac{V_{thp}-\gamma V_{BSP}}{m_p v_T}}\right)\right)$$

$$= \frac{1}{2}\eta C_s V_{DD}^2 t_{total}. \left(e^{\frac{-V_{DD}}{m_p v_T}} + e^{\frac{-V_{DD}}{m_n v_T}} + \frac{I_{0p} e^{\frac{-V_{DD}}{m_n v_T}}}{I_{0n}} e^{\frac{-\mu_{VTHP}}{m_p v_T}+\frac{\sigma_{VTHP}^2}{2m_p^2 v_T^2}} e^{\frac{\mu_{VTHN}}{m_n v_T}+\frac{\sigma_{VTHN}^2}{2m_n^2 v_T^2}}. \left\{\frac{e^{\frac{\gamma V_{DD}}{m_p v_T}}}{2}\left(1 - Erf\left(\frac{\sigma_{VTHP}}{\sqrt{2}m_p v_T} - \frac{\mu_{VTHP} - V_{thp0}}{\sqrt{2\,\sigma_{VTHP}^2}}\right)\right) + \frac{1}{2}\left(1 + Erf\left(\frac{\sigma_{VTHP}}{\sqrt{2}m_p v_T} - \frac{\mu_{VTHP} - V_{thp0}}{\sqrt{2\,\sigma_{VTHP}^2}}\right)\right)\right\}. \left\{\frac{1}{2}\left(1 - Erf\left(\frac{\sigma_{VTHN}}{\sqrt{2}m_n v_T} + \frac{\mu_{VTHN} - V_{thn0}}{\sqrt{2\,\sigma_{VTHN}^2}}\right)\right) + \frac{e^{\frac{-\gamma V_{DD}}{m_n v_T}}}{2}\left(1 + Erf\left(\frac{\sigma_{VTHN}}{\sqrt{2}m_n v_T} + \frac{\mu_{VTHN} - V_{thn0}}{\sqrt{2\,\sigma_{VTHN}^2}}\right)\right)\right\} + \frac{I_{0n} e^{\frac{-V_{DD}}{m_n v_T}}}{I_{0p}} e^{\frac{-\mu_{VTHN}}{m_n v_T}+\frac{\sigma_{VTHN}^2}{2m_n^2 v_T^2}} e^{\frac{\mu_{VTHP}}{m_p v_T}+\frac{\sigma_{VTHP}^2}{2m_p^2 v_T^2}}. \left\{\frac{e^{\frac{\gamma V_{DD}}{m_n v_T}}}{2}\left(1 - Erf\left(\frac{\sigma_{VTHN}}{\sqrt{2}m_n v_T} - \frac{\mu_{VTHN} - V_{thn0}}{\sqrt{2\,\sigma_{VTHN}^2}}\right)\right) + \frac{1}{2}\left(1 + Erf\left(\frac{\sigma_{VTHN}}{\sqrt{2}m_n v_T} - \frac{\mu_{VTHN} - V_{thn0}}{\sqrt{2\,\sigma_{VTHN}^2}}\right)\right)\right\}. \left\{\frac{1}{2}\left(1 - Erf\left(\frac{\sigma_{VTHP}}{\sqrt{2}m_p v_T} + \frac{\mu_{VTHP} - V_{thp0}}{\sqrt{2\,\sigma_{VTHP}^2}}\right)\right) + \frac{e^{\frac{-\gamma V_{DD}}{m_p v_T}}}{2}\left(1 + Erf\left(\frac{\sigma_{VTHP}}{\sqrt{2}m_p v_T} + \frac{\mu_{VTHP} - V_{thp0}}{\sqrt{2\,\sigma_{VTHP}^2}}\right)\right)\right\}\right)$$

$$= \frac{1}{2}\eta C_s V_{DD}^2 t_{total}. \left(e^{\frac{-V_{DD}}{m_p v_T}} + e^{\frac{-V_{DD}}{m_n v_T}} + f_{Ep}(V_{thp0})\, f_{Dn1}(V_{thn0}). \frac{I_{0p}}{I_{0n}} e^{\frac{-V_{DD}}{m_n v_T}} e^{\frac{-\mu_{VTHP}}{m_p v_T}+\frac{\sigma_{VTHP}^2}{2m_p^2 v_T^2}+\frac{\mu_{VTHN}}{m_n v_T}+\frac{\sigma_{VTHN}^2}{2m_n^2 v_T^2}} + f_{En}(V_{thn0})\, f_{Dp1}(V_{thp0}). \frac{I_{0n}}{I_{0p}} e^{\frac{-V_{DD}}{m_n v_T}} e^{\frac{-\mu_{VTHN}}{m_n v_T}+\frac{\sigma_{VTHN}^2}{2m_n^2 v_T^2}+\frac{\mu_{VTHP}}{m_p v_T}+\frac{\sigma_{VTHP}^2}{2m_p^2 v_T^2}}\right)$$

And finally combining these two results, overall impact of SULP FBB application on EDP, on a typical inverter can be extracted:

$$
\begin{aligned}
&E_{VTHN,VTHP}(EDP) \\
&= \frac{1}{2}\eta C_s V_{DD}^2 t_{total} \cdot \left(\begin{array}{l} e^{\frac{-V_{DD}}{m_p v_T}} + e^{\frac{-V_{DD}}{m_n v_T}} \\ + f_{Ep}(V_{thp0})\, f_{Dn1}(V_{thn0}) . \frac{I_{0p}}{I_{0n}} e^{\frac{-V_{DD}}{m_n v_T}} e^{\frac{-\mu_{VTHP}}{m_p v_T} + \frac{\sigma_{VTHP}^2}{2m_p^2 v_T^2} + \frac{\mu_{VTHN}}{m_n v_T} + \frac{\sigma_{VTHN}^2}{2m_n^2 v_T^2}} \\ + f_{En}(V_{thn0})\, f_{Dp1}(V_{thp0}) . \frac{I_{0n}}{I_{0p}} e^{\frac{-V_{DD}}{m_n v_T}} e^{\frac{-\mu_{VTHN}}{m_n v_T} + \frac{\sigma_{VTHN}^2}{2m_n^2 v_T^2} + \frac{\mu_{VTHP}}{m_p v_T} + \frac{\sigma_{VTHP}^2}{2m_p^2 v_T^2}} \end{array} \right) \\
&+ \left(\frac{\eta}{2}\alpha C_s^2\, V_{DD}^3\right) . \left(\frac{1}{I_{0p}} e^{-\frac{V_{DD}}{m_p v_T}} e^{\frac{\mu_{VTHP}}{m_p v_T} + \frac{\sigma_{VTHP}^2}{2m_p^2 v_T^2}} f_{Dp1}(V_{thp0}) \right. \\
&\left. + \frac{1}{I_{0n}} e^{-\frac{V_{DD}}{m_n v_T}} e^{\frac{\mu_{VTHN}}{m_n v_T} + \frac{\sigma_{VTHN}^2}{2m_n^2 v_T^2}} f_{Dn1}(V_{thn0}) \right)
\end{aligned} \tag{23}
$$

One has to notice that in (23), total is constant and it is not the time per instruction in which case it changes when body bias changes. In fact energy per instruction does not change in subthreshold voltages when V_{TH} is changed by body bias as showed in [12].

III. ACKNOWLEDGMENT

The authors would like to thank Dr. David Fitrio at Centre for Technology Infusion, La Trobe University, Melbourne, Australia, for his valuable comments and recommendations.

IV. REFERENCES

[1] B. Zhai, *et al.*, "Theoretical and practical limits of dynamic voltage scaling," presented at the Proceedings of the 41st annual Design Automation Conference, San Diego, CA, USA, 2004, DOI: 10.1145/996566.996798

[2] K. Nose and T. Sakurai, "Optimization of VDD and VTH for low-power and high speed applications," presented at the Proceedings of the 2000 Asia and South Pacific Design Automation Conference, Yokohama, Japan, 2000, DOI: 10.1145/368434.368755

[3] L. Yan, *et al.*, "Joint dynamic voltage scaling and adaptive body biasing for heterogeneous distributed real-time embedded systems," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on,* vol. 24, pp. 1030-1041, 2005,

[4] J. T. Kao, *et al.*, "A 175-mV multiply-accumulate unit using an adaptive supply voltage and body bias architecture," *Solid-State Circuits, IEEE Journal of,* vol. 37, pp. 1545-1554, 2002,

[5] S. Hanson, *et al.*, "Exploring variability and performance in a sub-200-mV processor," *Solid-State Circuits, IEEE Journal of,* vol. 43, pp. 881-891, 2008,

[6] K. A. Bowman, *et al.*, "Impact of die-to-die and within-die parameter fluctuations on the maximum clock frequency distribution for gigascale integration," *Solid-State Circuits, IEEE Journal of,* vol. 37, pp. 183-190, 2002,

[7] J. Tschanz, *et al.*, "Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage," in *Solid-State Circuits Conference, 2002. Digest of Technical Papers. ISSCC. 2002 IEEE International*, 2002, pp. 422-478 vol.1,

[8] J. Rabaey, *Low power design essentials*: Springer Verlag, 2009,

[9] Y. Pu, *et al.*, "An Ultra-Low-Energy Multi-Standard JPEG Co-Processor in 65 nm CMOS With Sub/Near Threshold Supply Voltage," *Solid-State Circuits, IEEE Journal of,* vol. 45, pp. 668-680, 2010,

[10] M. E. Hwang and K. Roy, "ABRM: Adaptive $ beta $-Ratio Modulation for Process-Tolerant Ultradynamic Voltage Scaling," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on,* vol. 18, pp. 281-290, 2010,

[11] S. Hanson, *et al.*, "Ultralow-voltage, minimum-energy CMOS," *IBM journal of research and development,* vol. 50, pp. 469-490, 2006,

[12] S. Hanson, *et al.*, "Performance and Variability Optimization Strategies in a Sub-200mV, 3.5pJ/inst, 11nW Subthreshold Processor," in *VLSI Circuits, 2007 IEEE Symposium on*, 2007, pp. 152-153,

[13] R. G. Dreslinski, *et al.*, "Near-Threshold Computing: Reclaiming Moore's Law Through Energy Efficient Integrated Circuits," *Proceedings of the IEEE,* vol. 98, pp. 253-266, 2010, DOI: 10.1109/JPROC.2009.2034764

[14] A. Srivastava, *et al.*, "Modeling and analysis of leakage power considering within-die process variations," presented at the Proceedings of the 2002 international symposium on Low power electronics and design, Monterey, California, USA, 2002, DOI: 10.1145/566408.566426

[15] S. Narendra, *et al.*, "Full-chip sub-threshold leakage power prediction model for sub-0.18µm CMOS," presented at the Proceedings of the 2002 international symposium on Low power electronics and design, Monterey, California, USA, 2002, DOI: 10.1145/566408.566415

[16] W. Liu, *et al.*, "BSIM4. 6.4 MOSFET Model,"

[17] M. SEOK, *et al.*, "REFERENCE VOLTAGE GENERATOR HAVING A TWO TRANSISTOR DESIGN," ed: WO Patent WO/2010/151,754, 2010

[18] B. Zhai, *et al.*, "Analysis and mitigation of variability in subthreshold design," presented at the Proceedings of the 2005 international symposium on Low power electronics and design, San Diego, CA, USA, 2005, DOI: 10.1145/1077603.1077610

[19] M. Meterelliyoz, *et al.*, "Characterization of Random Process Variations Using Ultralow-Power, High-Sensitivity, Bias-Free Sub-Threshold Process Sensor," *Circuits and Systems I: Regular Papers, IEEE Transactions on,* vol. 57, pp. 1838-1847, 2010,

[20] M. E. Hwang and K. Roy, "ABRM: Adaptive β-Ratio Modulation for Process-Tolerant Ultradynamic Voltage Scaling," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on,* vol. 18, pp. 281-290, 2010,

[21] S. Narendra, *et al.*, "Forward body bias for microprocessors in 130-nm technology generation and beyond," *Solid-State Circuits, IEEE Journal of,* vol. 38, pp. 696-701, 2003,

[22] M. Miyazaki, *et al.*, "A 1.2-GIPS/W microprocessor using speed-adaptive threshold-voltage CMOS with forward bias," *Solid-State Circuits, IEEE Journal of,* vol. 37, pp. 210-217, 2002,

Table I
SULP FBB and ZBB Simulation results comparison

SULP Compared to ZBB	-15°C to 75°C	-15°C to 45°C
EDP improvement	23%	42%
Delay reduction	63%	75%
Frequency increase	45%	54%
Energy increase	33%	23%
Delay Variation reduction	730%	900%