Abstract
Drastic yield reduction at sub/nearthreshold voltage domains, caused by the severe process, voltage, and temperature (PVT) variations in this region, is a challenging characteristic of recent nanometre sensory chips. Using a variation sensitive and ultra-low-power design, this paper proposes a novel technique capable of sensing and responding to PVT variations by providing an appropriate forward body bias (FBB) so that the delay variations and timing yield of the whole system as well as energy-delay product (EDP) are improved. Theoretical analysis for the error probability, confirmed by post-layout HSPICE simulations for an 8-bit Kogge-Stone adder and also two large Fast Fourier Transform (FFT) processors, shows considerable improvements in severe PVT variations and extreme voltage scaling. For this adder, for example, the proposed technique can reduce error rate from 50% to 1% at 0.4V. In another implementation, in average ~7x delay variation and ~4x EDP improvement is gained after this technique is applied to an iterative 1024pt, radix 4, complex FFT while working in sub/nearthreshold voltage region of 0.3V to 0.6V. Also, pipelined version of the FFT consumed only 412pJ/FFT at 0.4V while processing 125K FFT/sec.

Keywords: timing yield, subthreshold design, forward body biasing, process, temperature and voltage variations

1. Introduction
With the introduction of 65nm technologies, reliability of circuits started to challenge the transistor scaling which is an essential trend for continuation of performance, area and energy improvements in silicon chips fabricated for sensory systems.

Increasing defects in fabrication process as a result of aggressive transistor scaling is the source of this unreliability which reduces manufacturing yield. As it will be seen, the final yield for ultra-low power circuits depends on reliability (or resilience to temperature variation and voltage noise) and satisfying the performance and energy goals which are all correlated, and if not met, the specific die will be discarded. These challenges are pronounced even more seriously at the subthreshold voltages, from which low energy wireless applications benefit the most by trying to minimize the power use for a given performance requirement [1]. As PVT variations rise exponentially with the voltage scaling, this results in a dramatic uncertainty and still urges designers to employ adaptive body-bias (ABB) techniques [2] 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).

Although efficient in the superthreshold region, the impact of the 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[3].

As mentioned, susceptibility to noise or voltage variations is a major source of performance failure at subthreshold voltages. Because threshold voltage ($V_{TH}$) is managed by an independent doping process, $V_{TH}$ in PMOS and NMOS devices 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 a fast PMOS slow NMOS corner. Consequently, not only noise margins can be violated at process corners, but also either rising or falling time is

1 Corresponding author. Tel: +61394793210, E-mail: m.radfar@latrobe.edu.au.
extremely long which in return results in increased timing breaches.

As PMOS and NMOS transistors 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 [4] 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.

Authors in [5] also studied the capability of ABB techniques to address these variations and implemented a sub-threshold processor to show its effectiveness. They also proved that a body bias which optimises β-ratio for noise margin also minimises energy per instruction [6].

In this paper, the purpose is addressing the challenge of timing yield reduction at subthreshold voltages. An extreme process variation sensitive and ultra-low power (SULP) FBB circuit is proposed capable of addressing PVT variations, while $V_{DD}$ is scaling, by tuning β-ratio in different process corners. This technique helps improve the system performance and hence timing yield by applying FBB to NMOS and/or PMOS networks at lower voltages and/or lower temperatures and applying appropriate FBB to slower devices depending on the process variation.

In addition to sensing the PVT variations, the proposed technique is at the same time capable of reacting to the voltage level under which the circuit is working. This means that if a circuit has a low voltage level or any other performance restricting conditions (due to PVT variations), the proposed technique detects it and helps the system cope with the PVT variations which are more pronounced at low voltages. The originality of this technique is in its capability to handle harsh temperature and process variations while addressing aggressive voltage scaling that is unprecedented in the literature.

On the other hand, this extra help is withdrawn when the system is working in high voltage levels or any other high performance circumstances. This results in considerable reduction of functionality failures at slow conditions and no energy overhead at fast situations which adds to the technique’s originality as well. This technique also leads to a significant improvement in production yield as a result of its adaptive approach towards PVT variations as well as voltage reductions.

Another factor important to timing yield is delay variation which is enormously improved by this technique as it will be seen. Simplicity of the design and also its low power operation incurs low energy and negligible 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. By scaling voltage form 0.8V to 0.3V and temperature changes of -15°C to 75°C, error probability in the simulated 8-bit Kogge-Stone adder was decreased from 50% to 1% at 0.4V as a result of this technique.

The outcome of mixed signal simulations on a 1024 point, radix 4, 32x32bit complex input iterative FFT processor showed not only seven times improvement in delay variations, but EDP was also reduced to around 4 times, after this technique was applied, which showed the real benefit of the technique lying in large scale circuits.

Finally, a pipelined version of the FFT consumed 412pJ/FFT which was ~43 times less than the latest sub/nearthreshold FFT processor while being only ~1.9 times slower (with 125K FFT/sec throughput). SULP FBB also resulted in ~250% delay improvement in this pipelined FFT processor, with respect to a ZBB FFT processor, with 34% energy overhead.

The demonstrated mathematical platform also helps designers understand and adjust the parameters and factors which have the highest and lowest sensitivity in the output of a particular sub/nearthreshold circuit. They can also analyse and predict, with a high accuracy, what effects a technique can have on a design under test, and optimise the approach mathematically.

The rest of this paper describes how SULP FBB circuit can achieve this by firstly mathematically analysing how circuit works and why it reduces error rate 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.

2. Objective, Functionality and Theoretical Analysis of the Proposed Circuit

2.1 Objective

Fig. 1 shows the schematic of the proposed SULP FBB generator introduced in [7] which was used to apply FBB to the MOS devices so that the EDP of system was improved and process, voltage and temperature (PVT) variations were addressed while system was working in subthreshold voltages. The main duty of the SULP FBB technique was to address inter-die process variations by applying the appropriate FBB to the whole system.

It should be noted that inter-die and intra-die process variations are considered independent parameters and their effects are usually represented by different uncorrelated random variables [8]. As the SULP FBB technique has been characterised and confirmed, in this paper, for tackling inter-die process (as well as temperature and voltage) variations, the final results on the
improvement of performance, power and reliability will be independent of intra-die process variations. That is, final results will even enhance, should this (or any other) technique be used for handling intra-die and random process variations. In a 22nm technology, for example, the issue of inter-die variation will still be present, even worse than before, and the proposed technique will still be applicable for such technologies. Nevertheless, a 22nm circuit with the proposed technique applied to, will still benefit from other techniques to compensate intra-die and random variations which may not be addressed by the proposed circuit.

It is worth mentioning that this technique has a very small area overhead by having the total area of 168.7 \( \mu \text{m}^2 \). Besides, this small circuit is capable of providing independent PMOS and NMOS body biases, based on PVT variations, and it is designed for addressing process variations independent of being inter-die or intra-die. Therefore, SULP FBB generator can be readily adapted and exploited for dealing with the both variation types. This can be realised by implementing the circuit in various regions of the chip that have high correlations between process variation parameters. In fact the area overhead of this circuit is 500 times less than the latest body bias circuit in the literature which claims to be the smallest at the time [9] and therefore highly suitable for addressing intra-die variations, as well.

The above mentioned total area of 168.7 \( \mu \text{m}^2 \) includes the isolated deep n-well layer and the required deep n-well to deep n-well gaps which is sometimes not reported or considered in body bias generators [10]. Furthermore, this paper’s focus on the inter-die process variations implies that the proposed method should be designed for and applied at the chip level. On the other hand, techniques aiming at fine-grained body biasing, such as [11], are implemented to tackle process variation or subthreshold leakage at a more detailed/clustered level instead of a global/block level, hence, they are more appropriate for addressing intra-die variations or for post-silicon tuning purposes.

Here, however, this technique is mathematically analysed and new and novel aspects and applications in reliability and error rate reduction are further revealed and discussed [12]. Equations are, mainly, for the PMOS network body bias (PBB) generator but discussions and results are valid for the NMOS network body bias (NBB) generator too. First a quick background is given in subsection A, and then the error rate is worked out in subsection B.

### 2.2 Functionality of the Proposed Forward Body Bias Circuit

Two transistors in the first stage form a voltage reference generator adapted from [13]. However, this adaptation is not exactly what the authors designed which was a very largely sized (L=60\( \mu \text{m} \), \( W_{A1} = 1.5 \mu \text{m}, W_{A2} = 3.3 \mu \text{m} \)) supply voltage insensitive generator. In addition to having different (and rather opposite) sizings, the voltage dependency is crucial in this paper’s voltage reference generator.

Transistor \( M_{B1} \) is off but at the same time forward body biased. This, as a result, causes a leakage current passing through its channel, and creates a very small current sink from \( M_{A1} \). On the other hand, \( M_{A1} \) is always saturated and with this configuration, (gate connected to drain), it acts as a diode and results in a voltage drop. Fig. 2 depicts the reference voltage output (\( V_{N1} \) for NBB or \( V_{DD}-V_{N1} \) for PBB) in different corners and temperatures. In subthreshold voltages, it can be seen that the reference voltage is independent of temperature and process variations. When approaching superthreshold voltage, reference voltage is increased rapidly, which leads to leakier \( M_{A2} \) and the effect of PVT or voltage scaling is cancelled on \( M_{A2} \).

This stage, therefore, provides \( M_{A2} \) with a process/temperature independent \( V_{GS} \) proportional to \( V_{DD} \) at subthreshold voltages; 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. Increasing supply voltage and subsequently \( V_{GS} \) of \( M_{A2} \), raises \( V_{N2} \) which in return 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.

It is usually assumed that all process variations can be abstracted in \( V_{\text{th}} \) variation \([14]\). Assuming \( V_{\text{thp}0} \) is a specific threshold voltage of \( M_{A2} \) in which \( V_{N2} \) is equal to \( V_{DD}/2 \) and \( VN2 \) and \( V_{THP} \) represent random variables for variable \( V_{N2} \) (output voltage of the second stage) and variable \( V_{\text{thp}} \) (threshold voltage of \( M_{A2} \)), respectively, then \( V_{THP} \) has a known probability density function (PDF) as variations in \( V_{\text{thp}} \) presumably follow a Normal Distribution. \( V_{\text{thp}0} \) can be found by letting \( VN2 \leq VDD/2 \) which results in (please refer to equation (A.1) for definitions):

\[
V_{thp0} = V_{DD} - VN1 + \frac{\mu_{VTHN} m_p}{m_n} + m_p v_T \ln \left( \frac{I_{op}}{I_{on} \mu n} \right)
\]  

(2.1)

As mentioned before, by sizing 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 \([15]\)—here standard deviation (\( \sigma \)) of \( V_{THN} \) decreased to 0.01V while \( V_{THP} \) standard deviation is 0.04V. It should be mentioned that despite the common practice of using \( \sigma/\mu \) as the variability criterion, here only \( \sigma \) is used when comparing deviations or variations differences before and after SULP FBB application. The reason of popularity of \( \sigma/\mu \) lies in the fact that it enables comparison of the variations of two variables with considerably different mean values (\( \mu \)) like delay of a circuit in subthreshold and superthreshold voltages. But here, as all circumstances under which SULP FBB and ZBB circuits work (except the body biasing) are the same, therefore two being compared variables do not have significant mean value changes. As the PDF of \( V_{THP} \) is already known, then the probability of \( V_{BSP} = V_{DD} \) or \( VN2 \leq V_{DD}/2 \) can be found by:

\[
P \left( VN2 \leq V_{DD}/2 \right) = P \left( V_{THP} \geq V_{thp0} \right) = \frac{1}{2} \text{Erfc} \left( \frac{V_{thp0} - \mu_{VTHP}}{\sqrt{2} \sigma_{VTHP}} \right)
\]  

(2.2)

in which \( \text{Erfc}(x) \) is the complementary error function.

2.3 Theoretical Analysis of the Effect of SULP FBB circuit on error rate

While \([7]\) showed the improved EDP resulted from SULP FBB technique, here the main focus is to show how this technique was designed to take advantage of PVT variations to improve the performance at a reasonable energy cost in order to significantly reduce the error rate. Like the assumption made in \([7]\), 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 region too. When output of an inverter in the 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 [14]:

\[ t_d = D_0 e^{\frac{V_{thp} - \gamma V_{BSP}}{n_p V_T}} \]

where \( D_0 = \frac{1}{2} \eta C V_{DD} \)

\[ = \frac{I_{DD}}{I_{p}e^{\frac{V_{DD}}{n_p V_T}} \left(1 - e^{-\frac{V_{DD}}{V_T}}\right)} \]

(2.3)

Timing yield is modelled as the probability of an erroneous event in a data-path [16] and can be approximated by the probability of error in an inverter which has to have a delay of less than \( t_0 \) to meet the required speed. Given this assumption and by defining \( T_d \) as the random variable of \( t_d \), the probability of error or simply error rate can be acquired by the following equation and using the law of total probability:

\[ P(T_d > t_0) = P(D_0 e^{\frac{V_{THP} - \gamma V_{BSP}}{n_p V_T}} > t_0) = P(V_{THP} > m_p n_p \ln \left(\frac{t_0}{D_0}\right) + \gamma V_{BSP}) \]

\[ = P(B | V_{THP} > V_{thp0}) P(V_{THP} > V_{thp0}) + P(B | V_{THP} \leq V_{thp0}) P(V_{THP} \leq V_{thp0}) \]

where \( B = m_p n_p \ln \left(\frac{t_0}{D_0}\right) + \gamma V_{BSP} \)

(2.4)

Using above equation and a few mathematical derivations, final probability of error (or timing yield) is worked out in (A.3) for the zero body-bias (ZBB) circuit and in (A.4) for the SULP FBB case (refer to the Appendix for detailed equations).

Equations (A.3) and (A.4) have been plotted in Fig. 3.a and Fig. 3.b, respectively, by substituting coefficients with numbers extracted by simulations and curve fitting procedure (values have also been brought into Fig. 3). Fig. 3 shows that higher supply voltages or looser (higher) delay constraints \(^1\) lead to zero probability of error signifying a data-path free of functional error. On the contrary, very tight (small) delay constraints or very small supply voltages lead to probability of error equal to one which indicates total functional failure of the data-path.

It can be observed that there is a concavity in the error probability of SULP FBB inverter shown in Fig. 3.b, when in subthreshold voltage domain. This leads to less area in the probability equal to one or total failure plain (marked with the upper red trapezoids) and more area in the probability equal to zero or error free plain (marked with the lower blue trapezoid), when SULP FBB technique is applied. This also means that tighter (smaller) delay constraints can now be tolerated by an inverter with SULP FBB applied to it with respect to a ZBB inverter.

If the error probability of ZBB inverter is subtracted from the error probability of SULP FBB inverter (using expression \( P(T_{dZBB} > t_0) - P(T_{d} > t_0) \) or equation (A.3) – equation (A.4) or Fig. 3.a – Fig. 3.b) then the outcome has the form of Fig. 4.

---

\(^1\) Delay constraints, in here, refer to the time needed for a combinational circuit to meet the required setup and hold times

![Fig. 3. Probability of error for \( m_p=1.7, m_n=1.48, n_p=0.026V, \sigma_{VTBI}=0.04V, \mu_{VTBI}=0.5V, \mu_{VTBN}=0.45V, \eta=2.1, \text{ and } C_s=1pF \) in a) ZBB inverter and b) SULP FBB inverter](http://www.sciencedirect.com/science/article/pii/S0026271414003400)
Fig. 4 shows the improvement (reduction) in the probability of error in a ZBB inverter after the SULP FBB technique is applied. As it can be seen in Fig. 4, improvement in the probability of error in the superthreshold voltage region tends to zero. That is, both ZBB and SULP FBB inverters have an error free functionality in a superthreshold voltage. This is again due to the FBB cancellation in this region.

In the subthreshold voltage region, however, when $t_0$ or the delay constraint on the inverter’s output is very tight and close to zero (presented in Fig. 3), the probability of delay being larger than the expected constraint, for both ZBB and SULP FBB designs, are the same and equal to one, which results in zero probability improvement. This is true for very loose delay constraints too as both ZBB and SULP FBB inverters can meet the constraint and having the same probability of error equal to zero results in zero improvement too.

But delay constraints are determined by clock frequency and are set to practical limits which are neither too tight damaging the production yield nor too loose not meeting the required performance, but they are set to the tightest error free condition. For example, in Fig. 3, this delay constraint (or clock frequency for a data-path) is just set to the points where the surface is about to rise from zero to one. The extended area of the lower blue trapezoid in the Fig. 3.b is, in fact, highlighting the obtained room for the clock frequency improvement.

On the other hand, Fig. 4 also suggests the same fact that improvements in the error rate is possible if SULP FBB applied as the values in Fig. 4 are always bigger than zero, indicating values in Fig. 3.a are always bigger than Fig. 3.b (here maximum error rate reduction is 0.35 for a theoretical inverter). Although delay variation analysis for a single inverter does not represent the exact effect of variations on a data-path comprised of many various cells but, as showed in [14], a path made up from different serried cells will improve variation effects which means the approximation of improvement will be even more experienced in a data-path as simulations prove later.

### 3. Simulation Results and Discussions

In this section, as well as investigating how the proposed technique reacts to PVT changes, previously obtained equations are examined through simulations to verify their predictions. It should be noted that, as it will be mentioned later, a low power Standard Cell Library was designed for examining the SULP FBB generator. The cells of this library were characterised across all corners and necessary temperature and voltage ranges by applying a wide variety of input signals, depending on being a combinational or sequential cell, in order to obtain all the timing and power information. This information were required for the synthesis and place & route tools to be able to allocate the timing budgets to the clock skew and jitter and to satisfy the timing/power constraints.

#### 3.1 Simulations

##### 3.1.1 An 8-bit Kogge-Stone adder

An 8-bit Kogge-Stone adder was chosen to verify the SULP FBB technique. The adder’s verification was performed by feeding all the possible input vectors in to the circuit to simply add the two 8-bit numbers and by confirming the output. The critical path was also identified for the further stochastic timing verification and delay measurement. 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 for each supply voltage ranging from 0.3V to 0.8V. Foundry provided global process variation
model was exploited to produce the most accurate possible results in simulating inter-die process variations. This model contained all foundry proven Gaussian distributions which led to the most practical and realistic Monte Carlo simulation results possible. Temperature and voltage were also swept to resemble both temperature variations and V_{DD} scaling. In this study, temperature varies from -15°C to 75°C (90°C variation) and if higher or lower working temperatures is required then SULP FBB circuit has to be simply tuned to support this.

Fig. 5 shows a range of fast to slow processes in a scatter graph of 1K MC runs for the data-path spread by each run’s effect on the delay of PMOS and NMOS devices of a typical ZBB inverter. It has to be pointed out that both axes and black lines are in logarithmic scale. Black lines demonstrate the normalised delay reduction in data-path at a specific process corner. When both PMOS and NMOS networks are slow (blue dots), SULP FBB automatically applies FBB to both networks whereas in SF or FS corners, FBB is only applied to NMOS or PMOS networks, respectively.

In an FF corner (orange diamonds), neither NMOS nor PMOS networks are forward body biased hence resulting in no delay improvement. Black lines, indicating delay reduction, only exist when SULP FBB is applied to either or both networks. However, black lines stretch exponentially when PMOS network is fast showing that the nature of PMOS devices being slower than NMOS counterparts has a great impact in subthreshold region to an extent that even forward body biased PMOS network is not as fast as forward body biased NMOS network. This imbalance can be addressed by libraries which are designed for both sub and superthreshold regions [17] and can improve the achieved results in this study even further.
In Fig. 5, when temperature (left to right)/voltage (up to down) increases, SULP FBB generator reduces number of FBB applications as higher temperature leads to higher subthreshold leakage and faster devices in subthreshold domain and higher supply voltage leads to stronger and hence faster devices. This way by sensing the temperature and voltage, SULP FBB technique avoids applying FBB to devices which are fast enough due to process, temperature or voltage causes.

Simulations for the adder also verify the predictions of equations (A.3) and (A.4) abstracted in Table I and Fig. 6. Fig. 6 is a combination of three figures Fig. 3.a, Fig. 3.b and Fig. 4 but showing simulation results instead theoretical predictions. To be exact, Fig. 3.a is represented in Fig. 6 by the adder’s simulation results for the probability of error when ZBB is applied, Fig. 3.b by the simulation results for the probability of error when SULP FBB is applied, and Fig. 4 by the solid lines indicating probability improvement.

As suggested by Fig. 4, there is a maximum improvement curve for Kogge-Stone adder data-path which, as explained, is expected to be higher (here 0.52) than what was predicted by the theoretical inverter. As this adder clearly has higher delay than the analysed single inverter in the previous section, its delay constraints are also higher than the inverter.

<table>
<thead>
<tr>
<th>Design objective→</th>
<th>1% error rate</th>
<th>5% error rate</th>
<th>maximum improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voltage↓</td>
<td>error rate reduction/delay constraint relaxation(x)</td>
<td>error rate (%)</td>
<td></td>
</tr>
<tr>
<td>0.3V</td>
<td>0.296/3.3113</td>
<td>0.422/2.8184</td>
<td>0.528/2.8184 13.8</td>
</tr>
<tr>
<td>0.4V</td>
<td>0.495/3.3884</td>
<td>0.571/2.9512</td>
<td>0.589/2.6303 13.9</td>
</tr>
<tr>
<td>0.5V</td>
<td>0.256/1.9953</td>
<td>0.335/1.7378</td>
<td>0.394/1.5849 18.8</td>
</tr>
<tr>
<td>0.6V</td>
<td>0.012/1.0471</td>
<td>0.033/1.0965</td>
<td>0.048/1.0233 26.3</td>
</tr>
</tbody>
</table>

It can be seen in both Table I and Fig. 6, although error rate reduction is maximum along this curve, the SULP FBB error rate is not satisfactory even for subthreshold voltage domain.

For example, Table I shows the error probability of 0.138 for V_{DD} of 0.3V when the maximum error rate reduction is sought which is 0.528. In addition, Table I shows that, if SULP FBB is applied, ~2.8x delay constraint relaxation can also be achieved, compared to a ZBB data-path with the same error rate of 0.138. But choice of maximum error rate improvement leads to error rate of 13.8% or higher (across different voltages) in the data-path which is very yield damaging and in the superthreshold region this choice has no benefit in terms of the error rate reduction.

![Fig. 6 Probability of error for SULP FBB and ZBB data-path and the maximum gained improvement for 1K MC simulations at 25°C](http://www.sciencedirect.com/science/article/pii/S0026271414003400)

It is error rate that is often set as a design goal for which case Table I shows examinations of two error rate constraints 1% and
5%. Clearly, as limit on final error rate (probability of error) in the SULP FBB data-path is tightened, error rate reduction declines as well due to distancing from the maximum reduction point which is experienced at higher error rates. Despite this decline, the error reduction is still significant even for the expected error rate of 1%. And with this error rate, when SULP FBB is applied, the delay constraint can be relaxed more, compared to the case of maximum error rate reduction. This is as a result of the SULP FBB design feature which 1) not only improves the performance of data-path, by providing FBB when required, 2) but also reduces variations in data-path delay.

The former feature can be observed in Fig. 6 in which, as delay constraint reduces (tightens), probability of error raises to one in ZBB earlier than SULP FBB does. This demonstrates SULP FBB supremacy in performance because it now enables the data-path to perform faster as delay constraints can be tougher. And the latter is apparent in the slope of error rate curve in Fig. 6 when rising up from zero to one with steeper curves belonging to SULP FBB due to its lower delay variation. As a result, when error rate is reduced to lower percentages (like 1% here), a ZBB data-path will fail more and more as extreme corner variations play an important role in this case and satisfying such extreme process corners will need looser delay constraints and hence SULP FBB will be more beneficial in this situation as Table I and Fig. 6 both show.

Another significant problem in the subthreshold region is severe temperature dependency. As blue asterisks in Fig. 7 show for a ZBB data-path, when supply voltage approaches to subthreshold voltage domain, variance in the delays leading to 10% error rate or less, increases across different temperatures. As temperature or voltage increases, applied delay constraint can be dropped dramatically (z axis is in logarithmic scale) which suggests a decline in the rate of FBB application in higher temperatures or voltages if an FBB were to be used. This decline has been taken into account in designing the SULP FBB [7].

Red points, signifying SULP FBB impact in Fig. 7, show that although delay constraints can be reduced hugely after SULP FBB application and especially in subthreshold domain, the best temperature independent voltage at sub/nearthreshold domain occurs at 0.5V. This point has the highest temperature yield or the smallest red ellipse in Fig. 7 (except superthreshold voltages of 0.7V and 0.8V) which indicates how important the priorities (timing yield or temperature yield) are while tuning the SULP FBB generator. If higher timing yields at subthreshold voltages are more important than temperature yield (which was in this study), then mean of FBB voltage has to be higher at subthreshold region with respect to nearthreshold domain [7]. For example, Fig. 8, in which error rate improvement curve in Fig. 6 has been extended to different temperatures, demonstrates that higher error rate improvement (which leads to higher timing yield) is achieved at subthreshold domain, with less regard to temperature, but at nearthreshold domain this improvement will be largely temperature dependant as timing yield is less significant in higher voltages in this design.

3.1.2 A 1024pt, radix 4, 32x32bit complex FFT

In this subsection, the proposed technique is examined in a very larger scale by applying it to two extensive Fast Fourier Transform (FFT) processors. Authors in [18] have created a framework by which the iterative and streaming concepts in
datapath can be represented. Using this framework, called Spiral, two FFT structures were generated, both 1024 point, radix 4, 32x32bit complex input: 1) an iterative version with 64 64x64bit SRAMs and 48 32x1024bit ROMs which takes 241 cycles to perform an FFT 2) a pipelined version with Flip-Flops which takes 32 cycles per FFT. First version is simulated comprehensively to examine the SULP FBB effect. However, only corner analysis were performed as Monte Carlo simulations were impossible given the scale of FFTs. In this case, stimuli was created using the hardware description language Verilog for the functionality verifications and the timing verification tool, Synopsys Design Compiler, was used for the critical path identification needed in the subsequent delay analyses.

Fig. 9 shows the layout of the iterative FFT. Two voltage domains of 1.2V and variable voltages (0.3V-1.2V) plus different partitions have been highlighted. A low power Standard Cell Library was also created in two versions of high voltage (no body bias) and variable voltage (body biased). The required level shifter, capable of shifting voltages down from 0.3 to 1.2V, was adapted from [19]. In this layout the SULP FBB generator is 70,000 smaller than the entire chip which brings about almost no area overhead. As the current sink from the FBB generator is negligible compared with the current of main power sources, the FBB generators and the connected straps can be designed to be small in the area despite sourcing huge amount of cells. Simulations on this enormous circuit with ~880,000 standard cells were performed using Synopsys Discovery AMS (Analogue Mixed Signal) suite mixing VCS-MX and CustomSim-XA.

Fig. 9. Layout of the iterative FFT; red area (middle part) is 1.2V domain and the rest variable voltage domain; green area (bordering accumulation) shows adders/multipliers and the blue area (between two previous ones) is the permutation part in the FFT
Fig. 10 shows the outcome of simulations (with and without SULP FBB application) in a TT corner and 25°C temperature across different voltages. It is clear from figure that, in this corner and temperature, SULP FBB is applied only at voltages under 0.5V.

As the memory and its related low-to-high level shifters and buffers (inverters) all are placed in a $V_{DDH}$ (1.2V) voltage domain, Fig. 10 depicts four types of energy consumptions. Black lines (marked ►) show the energy per FFT consumed in $V_{DDH}$ domain except memory consumption, brown lines (marked ●) show switching energy of $V_{DD}$ domain energy (subthreshold domain switching energy per FFT) and magenta lines (marked ◄) show total energy per FFT. Blue lines (marked ▲) show frequency in the FFT. As shown in [6], energy per instruction does not change in subthreshold voltages, when $V_{TH}$ is changed by body biasing, and subthreshold switching energy in Fig. 10 illustrates this fact. This is because the same amount of energy is needed to switch a gate in both scenarios no matter how fast gates switch.

Although leakage current is higher after SULP FBB application, the frequency is higher too, which in turn increases throughput and reduces working times resulting in reduced total leakage energy and compensating part of the increased leakage current. Moreover, overall delay is reduced by ~3.6x (see Table II) when FFT is in subthreshold voltage domain and SULP FBB is applied. However, consumed power remains the same for $V_{DDH}$ domain (as voltage and therefore currents remain the same) at subthreshold voltages, and together with delay reduction, the SULP FBB application in fact decreases the $V_{DDH}$ domain energy and in return total consumed energy.

Table II. Improvements after SULP FBB technique is applied for TT corner and 25°C temperature.

<table>
<thead>
<tr>
<th>voltage</th>
<th>Delay improvement</th>
<th>energy reduction per FFT</th>
<th>EDP improvement</th>
<th>$V_{DDH}$ energy reduction per FFT</th>
<th>$V_{DD}$ energy increase per FFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>3.45</td>
<td>1.32</td>
<td>4.55</td>
<td>1.70</td>
<td>1.29</td>
</tr>
<tr>
<td>0.4</td>
<td>3.72</td>
<td>1.25</td>
<td>4.64</td>
<td>5.82</td>
<td>1.07</td>
</tr>
<tr>
<td>average</td>
<td>3.58x</td>
<td>1.29x</td>
<td>4.60x</td>
<td>3.76x</td>
<td>18%</td>
</tr>
</tbody>
</table>

It is clear that, at subthreshold voltages, $V_{DDH}$ domain has comparable power consumption to $V_{DD}$ domain as $V_{DD}$ domain consumes less switching power and $V_{DDH}$ domain also needs to wait most of the time for the subthreshold part which results in huge leakage power consumption as well (especially in memory). This difference in energy consumption has been brought to attention in Table III.
Approaching to higher voltages, $V_{DD}$ domain takes more part in consuming the overall energy. It should be noted that the use of an industrial memory in these simulations gave rise to high total energy consumption as whole $V_{DDH}$ domain had to remain supplied with 1.2V power supply disrespect of $V_{DD}$ domain voltage. To give an idea how much this high voltage memory (SRAM, ROM and level shifters and buffers) contributed to total power consumption, a magenta spot can be seen on the Fig. 10 showing total consumed energy at 0.4V (for both ZBB and SULP FBB techniques) if a subthreshold memory (like what proposed in [20] and [21]) were to be used which results in 7x less energy consumption making total energy ~36nJ per FFT at ~6MHz (~25K FFT/sec). In this case, consumed $V_{DDH}$ energy in Fig. 10 is eliminated and now the only energy cost, as a result of SULP FBB application, would be the $V_{DD}$ leakage energy which is 0.5nJ per FFT and negligible compared to the total consumed energy (and hence not distinguishable in Fig. 10 for these two techniques). Therefore, EDP is improved by the amount of delay improvement which is ~3.6x rather than ~4.6x of the industrial memory case (see Table II).

Fig. 11 shows the difference between delays of FF and SS corners signifying ±3σ variation and Table III compares these values and shows delay variation improvements across different voltages after SULP FBB is applied. Comparing Fig. 11 and Fig. 10 shows that FBB is applied in an SS corner earlier than in a TT corner and is not applied in an FF corner at all. This is because SULP FBB generator senses variations in process as well as voltage and applies FBB depending on how slow the PMOS and/or NMOS networks are. As supply voltage scales down, delay variations increase; but with SULP FBB application, in average ~7x improvements occurs in delay variations compared to ZBB case. But this variation improvement deteriorates as $V_{DD}$ scales more towards subthreshold voltages leaving the maximum improvement (in yield) at 0.6V. An SULP FBB design with different supply voltages (in the buffer stages) from the system’s supply voltage can solve this problem by applying higher levels of forward body bias with respect to system’s supply voltage. Applying higher FBB voltages, however, leads to increased energy overhead.

Although SULP FBB, in a bigger scale such as the simulated FFT, once again helps improve DEP and delay variation (timing yield), designing an FFT with two separate voltage domains for the memory and the logic turns out very energy consuming even with a subthreshold memory. Using high voltage memory together with low voltage processors has been very popular [22-24] since memories usually start to fail at subthreshold voltages while logic part is still functional. However, in this particular application, that is FFT implementation, another option is using flip-flops or latches instead of internal memory which also makes pipelining possible (by means of register files) and eliminates the use of level-shifters and the high voltage domain needed.

Table III. Variation improvement and VDD domain energy portion when SULP FBB is applied.

<table>
<thead>
<tr>
<th>voltage (V)</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>variation reduction (SS to FF)</td>
<td>4.98x</td>
<td>6.63x</td>
<td>8.22x</td>
<td>8.43x</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$V_{DD}$ domain energy to total energy (%)</td>
<td>3.15</td>
<td>5.96</td>
<td>8.05</td>
<td>11.15</td>
<td>14.89</td>
<td>18.79</td>
</tr>
</tbody>
</table>

Fig. 11. Maximum delay in FFT for temperature 25°C and SS and FF corners on two techniques of SULP FBB and ZBB.

(http://www.sciencedirect.com/science/article/pii/S0026271414003400)
for memory blocks [25].

By exploiting this idea and for the sake of comparison with other works, a pipelined 1024 point, radix 4, complex 32x32b input FFT processor was implemented. Using SULP FBB on this FFT resulted in ~250% delay improvement with 34% energy overhead with respect to the ZBB FFT. This FFT consumed 412pJ/FFT which is ~43x less than the FFT in [25] while it is ~1.9x slower than it (with 125K FFT/sec throughput).

4. Conclusion

As a result of high PVT variations, production yield drops dramatically at subthreshold voltages especially in recent nanometre transistors. On the other hand, voltage scaling is an inevitable technique in the today’s silicon industry and cannot be evaded in many ultra-low power designs. Here, the effect of proposed FBB on timing yield and delay variations was investigated. It was concluded from results that the proposed technique was able to improve the timing yield from 50% down to 1% at 0.4V for the examined adder. Adder’s EDP was also reduced to around 4 times, as a result of SULP FBB technique.

In a bigger scale of a powerful FFT, a ~7x reduction in delay variations was observed. On the other hand, negligible area and energy overhead of this technique was another important factor which is promising in subthreshold designing. And finally, a pipelined version of FFT consumed ~43x less energy/FFT than the latest low-power FFT in the literature while being ~1.9x slower.

Appendix

If a device is operating in subthreshold region, its drain source current can be modelled by the well-known equation [26]:

\[ I_{DS} = I_0 e^{-\frac{V_{DS}}{m \cdot V_T}} \left( 1 - e^{-\frac{V_{DS}}{V_T}} \right) \quad \text{where} \quad I_0 = \mu \frac{W}{L} \sqrt{\frac{q \varepsilon_{si} NDEP}{2 \Phi_s}} v_T^2 \]  

(A.1)

where \( V_{th} \) signifies the threshold voltage, \( \gamma \) the body bias coefficient and \( m \) the slope factor of transistor. For \( I_0, \mu \) is the mobility, \( W \) the width, \( L \) the length, \( \Phi_s \) the surface potential, \( v_T \) the thermal voltage (=\( k_B T/q \)), \( NDEP \) the doping concentration in the channel and \( \varepsilon_{si} \) the permittivity of silicon for this device.

Using equations (2.2), (2.3), (2.4), and (2.5) it can be concluded that:

\[ P(T_d > t_0) = P\left(V_{THP} > m_p v_T \ln \left( \frac{t_0}{D_0} \right) + \gamma V_{DD} \right) P\left(V_{THP} > V_{thp0} \right) + P\left(V_{THP} > m_p v_T \ln \left( \frac{t_0}{D_0} \right) \right) P\left(V_{THP} \leq V_{thp0} \right) \]

\[ = \frac{1}{2} \text{Erfc} \left( \frac{m_p v_T \ln \left( \frac{t_0}{D_0} \right) + \gamma V_{DD} - \mu_{V_{THP}}}{\sqrt{2 \sigma_{V_{THP}}^2}} \right) + \frac{1}{2} \text{Erfc} \left( \frac{V_{thp0} - \mu_{V_{THP}}}{\sqrt{2 \sigma_{V_{THP}}^2}} \right) \]

\[ + \frac{1}{2} \text{Erfc} \left( \frac{m_p v_T \ln \left( \frac{t_0}{D_0} \right) - \mu_{V_{THP}}}{\sqrt{2 \sigma_{V_{THP}}^2}} \right) \left( 1 - \frac{1}{2} \text{Erfc} \left( \frac{V_{thp0} - \mu_{V_{THP}}}{\sqrt{2 \sigma_{V_{THP}}^2}} \right) \right) \]

(A.2)

A ZBB inverter with \( T_{dZBB} \) as the random variable for output delay has no FBB and hence a zero VBSP and therefore will have the error rate of:

\[ P(T_{dZBB} > t_0) = \frac{1}{2} \text{Erfc} \left( \frac{m_p v_T \ln \left( \frac{t_0}{D_0} \right) - \mu_{V_{THP}}}{\sqrt{2 \sigma_{V_{THP}}^2}} \right) \]

(A.3)

As \( 1 - e^{\frac{-V_{DD}}{v_T}} \approx 1 \) for \( V_{DD} > 0.3V, D_0 \) can be substituted in above equations resulting in:

(http://www.sciencedirect.com/science/article/pii/S0026271414003400)
\[
P(T_{d_{ZBB}} > t_0) = \frac{1}{2} \text{Erfc} \left( \frac{m_D v_T \ln(2) t_0 I_{op}}{\eta C_s V_{DD}} \right) + \frac{V_{DD} - \mu_{VT_{HP}}}{\sqrt{2} \sigma_{VT_{HP}}} \end{align*}
\]

\[
P(T_d > t_0) = \frac{1}{2} \text{Erfc} \left( \frac{m_D v_T \ln(2) t_0 I_{op}}{\eta C_s V_{DD}} \right) + \frac{V_{DD} - \mu_{VT_{HP}}}{\sqrt{2} \sigma_{VT_{HP}}} \end{align*}
\]

References


