Sub-threshold Library Design

***Abstract*—Sub-threshold operation is promising to achieve energy minimization when high performance is not required. The device sizing in sub-threshold region is different from other regions due to significantly different IV characteristics and impact of parasitic effects. Operation in the subthreshold or weak-inversion region exploits the parasitic subthreshold leakage current, and uses it as its primary operation current. These currents are much weaker than standard strong-inversion currents, and so the time needed for charging or discharging capacitive nodes is longer, limiting the operation frequency of the circuit. Various effects in subthreshold region and some of the power reduction techniques have been discussed in the paper.**

***Index Terms*—sub-threshold, energy, weak-inversion, leakage-current, strong-inversion**

I. Introduction

With the increase in demand for power-sensitive devices, conventional methods for low power design may not be sufficient in many applications where there is a requirement for low power consumption with a medium frequency of operation. To handle this deficit, the design of digital sub-threshold logic has been proposed. Operation of devices in the sub-threshold region leads to significantly lower dynamic and leakage power consumption and increased transconductance with a tradeoff to the performance of the device. Designing devices for operation in the subthreshold region is advantageous for portable devices which need low power dissipation. Various power reduction techniques are discussed which further reduce the dynamic and static power in the subthreshold region. Although it restricts speed, the reduced voltage of operation results in greater energy efficiency. Less power consumption results in lighter circuitry, fewer or lighter cooling modules, and the power source's expected lifetime are also increased.

The operation in the subthreshold region poses effects such as inverse narrow width effect and reverse short channel effect which results in a non-proportionate relation between the current and width of the device. This affects the way sizing needs to be done to achieve the desired current and performance. For applications operating at low power and low frequency, the subthreshold design presents good gains in PPA however subthreshold logic design is highly prone to PVT variations and noise. Managing effects of the subthreshold region and its variations is a challenge but these are mitigated with design techniques which thus makes devices operating in subthreshold logic more efficient.

II. Short-Channel Effects on MOSFET Subthreshold Swing

Usually, the introduction of short channel effects is associated with a deterioration in MOSFET properties. However, it is shown here that at least one measure of MOSFET performance, the subthreshold swing, may actually improve once short channel effects start to have an influence. If the gate and drain share the depletion zone charge, the impact can be explained. As a significant punch-through current begins to flow, S increases once more in the extreme short channel regime.

Electric field lines from a short-channel MOSFET's drain extend far into the region under the gate, lowering the effective depletion charge controlled by the gate. Because a given change in Vti results in a larger change in surface potential in a short-channel device than in a long-channel device, S in the short-channel MOSFET is reduced.

In very short MOSFETs, subthreshold swing S may actually decrease as L decreases before catastrophically increasing as punch-through current begins to flow. The effect is most noticeable in devices with deep source/drain junctions, lightly doped substrates, and heavy threshold adjust implants operating with a high drain bias. This dependence on device geometry and bias can be explained in terms of charge sharing in the depletion region between the gate and drain.

III. A Parametric Short-Channel MOS Transistor Model for Subthreshold and Strong Inversion Current

A parametric model for MOS transistors with short-channel capabilities is presented. It includes subthreshold and strong inversion regions, with a smooth transition between them. Mobility reduction, carrier velocity saturation, body effect, source-drain resistance, drain-induced barrier lowering, and channel length modulation are all included in the model. The model simulates the critical analogue circuit characteristics of current, transconductance, and output conductance with high accuracy. For accurate simulation of short channel MOS transistors, models for subthreshold current, channel length modulation, carrier velocity saturation, mobility reduction, and body effect must be included.

The model presented in this paper incorporates all of these critical features. Furthermore, the presented model predicts output conductance and transconductance accurately in both the subthreshold and strong inversion regions of operation.

The two primary operating regions of the MOS transistor in which current flows from the source to the drain are the weak inversion or subthreshold region and the strong inversion region. When the gate voltage V.s is less than the threshold

voltage VT, the subthreshold region occurs, and when the gate voltage is greater than the threshold voltage, the strong inversion region occurs. The diffusion current is the sub threshold current, and the drift current is the strong inversion current. The transition region is the area immediately adjacent to the strong inversion and subthreshold regions.

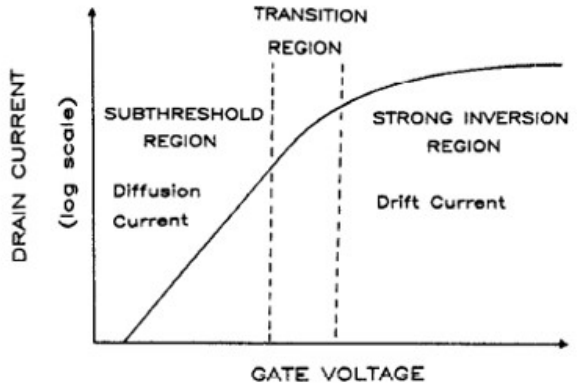

Fig. 1. Different regions based on Gate voltage

The drain voltage has an effect due to two phenomena: channel length modulation and drain-induced barrier lowering. The channel length modulation effect is caused by a shortening of the transistor's effective channel length as the drain voltage increases due to an increase in the drain depletion region. The drain-induced barrier lowering effect occurs when the channel becomes sufficiently short that the drain voltage directly influences the surface potential of the transistor. The channel length modulation effect exists across all channel lengths, but the drain induced barrier lowering effect exists only at short channel lengths.

$$I_0 = \frac{\mu W C_{\mathrm{OX}}}{LA} V_t^2 \exp\left[\frac{A(V_{GS} - V_{T0})}{V_t}\right]\left[1 - \exp\left(\frac{-V_{DS}}{V_t}\right)\right]$$

where

$$\frac{1}{A} = \frac{C_{\mathrm{OX}} + C_D + C_{FS}}{C_{\mathrm{OX}}}$$

$$C_{FS} = qN_{FS}$$

and

$$C_D = \frac{K_1 C_{\mathrm{OX}}}{2\sqrt{\phi_S - V_{BS}}}.$$

K_1 is the body effect coefficient which can be related to the effective doping N_{eff} as

$$K_1 = \sqrt{\frac{2\epsilon_{\mathrm{Si}} q N_{\mathrm{eff}}}{C_{\mathrm{OX}}^2}}$$

In strong inversion, the MOS transistor current is a drift current that is influenced by a number of important effects such as mobility reduction, body effect, channel length modulation, carrier velocity saturation, and source-drain resistance. An MOS model has been developed that incorporates these effects in order to accurately model the current, transconductance, and output conductance for both the triode and saturation regions of strong inversion operation.

A. Body Effect and Mobility Reduction

The transverse field influences the mobility of the carriers, resulting in reduced mobility. The expression that represents mobility reduction is

$$\mu = \frac{\mu_0}{1 + \theta(V_{GS} - V_T)}$$

where μo is the mobility of the threshold voltage and 0 is the mobility reduction factor The body effect is the effect of the substrate voltage on the current of a MOS transistor. The body effect has three primary effects on the current characteristics of a transistor. First, the threshold voltage varies as a function of the substrate voltage. Second, the value of the current at a given operating voltage decreases when compared to no body effect considerations. Finally, the pinch-off voltage V is decreased. The pinch-off voltage is the drain voltage at which the transistor transitions from the triode to the saturation region of operation for a given gate voltage.

B. Channel Length Modulation

By modulating the channel length and moving the threshold voltage by the drain voltage, a MOS transistor's output conductance in the saturation region is produced. As was already said, the threshold voltage is shifted by drain-induced barrier lowering. Channel length modulation for drift current occurs when the drain voltage rises because the pinch-off point shifts further from the drain. The pinch-off point is the distance at which the silicon-oxide interface's potential is no longer sufficient to strongly flip the semiconductor.

$$I_D = I_{DS} \frac{L}{L - L_D}.$$

The value of the channel shortening L can be calculated by considering the MOS structure. The electric field that is perpendicular to the channel at the pinch-off point P is known as the pinch-off field or EG. When the channel length shortening L is calculated, it is discovered that the electric field at point P is the pinch-off field. A is the length of the field line from the drain to point P, where A=LF. The path from the drain to point P is not perpendicular to the surface, which is explained by geometry element F. The value of L can be obtained by applying Poisson's equation along the path from the drain to point P.

$$V_{DS} - V_{DSS} = E_G \lambda + \frac{q N_{\mathrm{eff}} \lambda^2}{2\epsilon_{\mathrm{Si}}}$$

Assuming the pinch-off voltage is the voltage at point P, and finding Ld results in:

$$L_D = \frac{K_2}{F}\left[\sqrt{V_{DS} - V_{DSS} + \left(\frac{E_G K_2}{2}\right)^2} - \frac{E_G K_2}{2}\right]$$

C. Carrier Velocity Saturation

Carrier velocity saturation is a phenomenon that occurs at very large electric fields where the speed of the carriers no longer increases proportionally to the electric field. This phenomenon is critical for short-channel MOS transistors. The general velocity saturation equation is:

$$v = \frac{\mu_0 E_y}{\left(1 + \left(\frac{E_y}{E_c}\right)^{\alpha}\right)^{1/\alpha}}$$

where

$$E_C = \frac{v_L}{\mu_0}$$

and Vl is the saturation rate of the carrier velocity. According to the equation, Fig shows a plot of carrier velocity versus electric field for the values of a = 1 and a = 2. The value of a has been examined in several studies, with a = 2 being the most frequently used value for the silicon surface inversion layer.

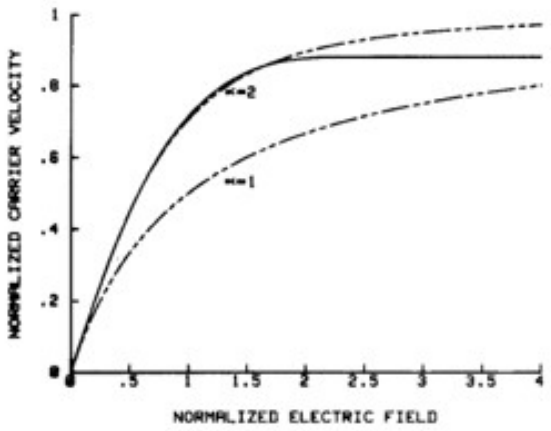

Fig. 2. Normalized carrier velocity Vs Normalized E.F.

D. Source – Drain Resistance

The dc MOS transistor model is incomplete without a source and drain resistance. This is particularly important for the proper fitting of the output conductance in the triode region. The source and drain resistors cause a shift in the location where the output conductance curves in the triode area extrapolate to an output conductance of zero. This intersection with zero output conductance substantially facilitates the extraction of the source and drain resistance for experimental transistors.

IV. Power Reduction Technique

This section focuses on power-saving strategies, and here we will see that they work at both the device and transistor levels. We can always make decisions that will ultimately have a big impact on the power dissipation while designing a circuit, even though they may function at much higher design levels, like the level of architecture. This is not to argue that power considerations should only be made after the circuit has transistors; rather, we can do so even while designing circuits at a very high level, such as the behavioral level. We will have more control over the power dissipation once the final circuit is set up if we adhere to or make changes to the tactics we outline. The methods we will cover here are mostly for lowering dynamic power and lowering static power. The rate at which output changes, or how frequently it is charging or discharging, is measured by the activity factor-alpha. The load capacitance is measured by C, the power supply voltage is measured by VDD, and the clock frequency is measured by f. Now since you're talking about reducing, any one of these four characteristics can be reduced, and these tactics can also entail reducing all four of them.

We'll see how strategies like clock gating and sleep mode can be used to minimize the activity factor. Clock gating refers to the ability to prevent the clock from entering a circuit when it isn't necessary. If the clock can be turned off, all activities will cease, the alpha will automatically decrease, and the value of alpha will be correct. In a similar way, we can strive to lower C by using small transistors, short cables, smaller fan outs, and other ways. Reducing supply voltage is a relatively common method, and since VDD appears to be square, it seems to be especially important. Unfortunately, there is a trade-off since as we lower the supply voltage, latency also goes up. Finally, we can naturally reduce the frequency as much as we can without significantly affecting the performance. It's possible that our circuit has a power-saving mode that automatically reduces the frequency.

We know that static power is spent whenever the output switches, causing the pull-ups and pull-downs to temporarily conduct. To reduce static power, we can use a ratioed circuit or a mix of several other ways. The output voltage is often drawn from a circuit that resembles a voltage divider, which is typically what we refer to as a ratioed circuit. A ratioed circuit is a conventional CMOS style of the circuit where there is a pull-up network and a pull-down network, Since, there will be multiple gates connected in series whenever we have a ratioed circuit, the likelihood of these switching currents flowing will decrease on their own. And we can employ low threshold devices selectively; for example, if we lower the transistor's threshold voltage, the static power is also reduced. And of course, static power AL includes leakage power leakage reduction technique AL includes several ways like stacked devices, which refers to a series of transistors, body bias, where we change the substrate bias voltage, and reduce the temperature operation, several of which exist because they directly affect the leakage current.

A. Clock Gating

Clock gating is one way to reduce alpha. It is the presumption that signal transition activity, or alpha, has a major effect on a circuit's power dissipation. Clock gating is essentially what it entails. We see that the clock is a signal that serves

as the brain of the entire system. The clock is what starts the various circuit operations, and whenever the clock transitions take place, all of the circuit's components operate in synchronism. The fundamental idea behind clock gating is that if there is a section of a circuit that is not being used at the moment, why not totally turn off the clock from that section so that at least for that section, signal switching activity will stop and power consumption would reduce. We do however remember one thing that if we work as a test engineer and your primary responsibility is testing, we can advise against clock gating. If we employ clock gating, testing may become more difficult; because if the error in that gate prevents the clock from ever reaching there, we won't be able to test that sub-circuit. Clock gating is viewed as a very simple way to drastically reduce power, though, from the perspective of low power. Another compromise is that we disable the clock signal from the functional modules that are not being used. The clock signal may incur an additional delay or skew as a result of the additional hardware that may be needed for this.

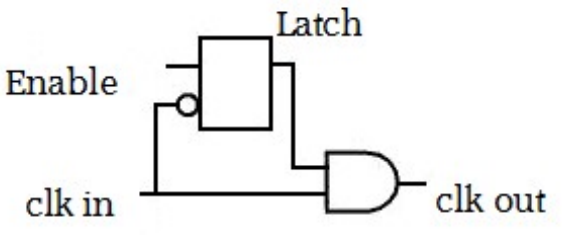

Fig. 3. Latch-based clock gating technique

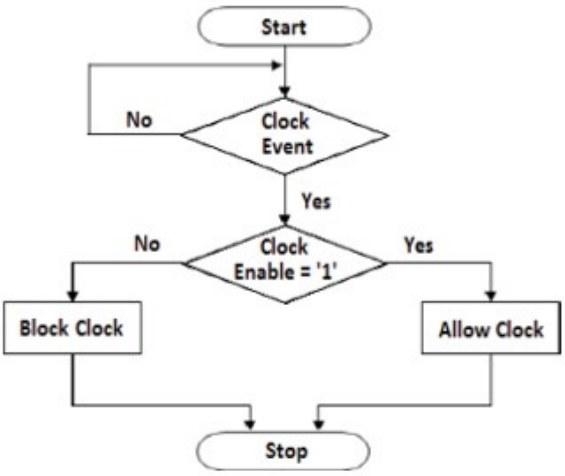

Fig. 4. Flow chart for clock gating

Flip-flop-based gating techniques are typically not preferred due to their excessive switching activity and poor effectiveness in contrast to other techniques. Additionally, when the capacitance's charging and discharging characteristics are constantly altering, it consumes more energy. The required amount of capacitance also increases. The gate-based clock gating consumes more dynamic power as a result of glitches and increased switching activity. The gated approach reduces the delay mismatch and the majority of the sleep duration. The gated method, however, takes up less room than the flip-flop-based clock gating method. That's why we use latch-based clock gating comprised of a special cell ICG (integrated clock gate) in the library to prevent the chance of glitches to a greater extent.

Latch based CG	Gate based CG	FF based CG
Glitches are avoided	Affected by glitches	Affected by glitches
Long sleep period	Sleep period avoided	Long sleep period
Mismatch of delay occurs	No mismatch of delay	Delay occurs
Low dynamic power	High dynamic power	High dynamic power
Less switching activity	High switching activity	High switching activity
Performance is good	Moderate performance	Performance is bad
Less area is occupied	Less area is occupied	Large area is used

Fig. 5. Comparison of various clock gating technique

The gate-based clock gating is preferable since it consumes less energy and space. However, due to its great performance, flawless operation, and tiny footprint, latch-based clock gating is favored. Flip-flops are also not very common due to their subpar performance, excessive power consumption, and switching activity.

B. Voltage Scaling

As we want low power dissipation and we know that P= CV2f and we see that power is directly proportional to the square of the voltage so we are reducing the voltage in subthreshold we using voltage around 0.4V. In fig 6, when S=1 that means input is high so NMOS should turn on and PMOS will be off so M2 is in an active state and M1 is in the cut-off state so a path will form from C1 to the ground through M2 which discharges the capacitor C1 and M3 transistor is a PMOS when voltage Vp at C1 is low M3 will turn on. As M3 will turn on a path from Vdd to C2 will form the C2 is fully charged to Vdd. When M2 and M1 transistors are low then the current in M3 is determined by the voltage difference in PMOS transistors. The Self Stabilization path helps in charge sharing between C1 and C2 capacitors. X, Vtp, and Vdd share a relation which is: X is less than or equal to (Vdd – Vtp4). When the clock makes a transition from high to low M1 becomes active and thus X is dropped to a value very lower than Vdd. Hence, the main purpose of the Voltage Scaling technique is to reduce dynamic power consumption.

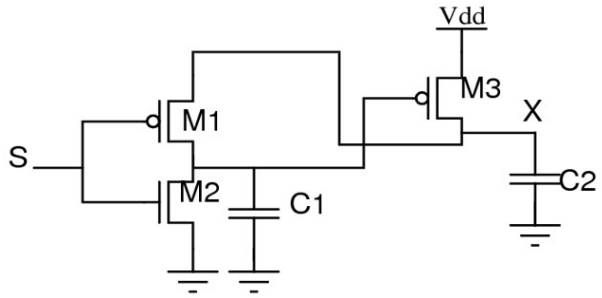

Fig. 6. Voltage scaling technique

V. Techniques to reduce static power dissipation

Static Power is due to leakage current through devices when they are turned off or on. Mainly the leakage current consists of Reverse biased pn junction current, Sub-threshold leakage, Gate induced drain leakage (GIDL), Punch through and Gate tunneling. To decrease static power dissipation, many circuit and device level approaches have been developed. Some of them are transistor stacking, dual-threshold partitioning, variable thresholds, Power Gating using sleep transistors, etc.

A. Forced Stack

The alternative method by which we are able to operate the circuit at lower power is a forced stack technique in this we just break the transistor into two transistors half of the size of the actual one and this is called as stack effect. When implemented in the circuit, these two sizes are halved. When the transistor is in the off state, the transistor is reverse biased Between the two, culminating in a subliminal leakage current. This technique generally reduces power consumption when the circuit is in standby mode that is if it does not work.
Power dissipation is a growingly significant problem as we mode down to lower technology nodes. Earlier, Dynamic power dissipation had a significant role in overall power consumption and there are various techniques to overcome this. As technology is moving towards smaller size and faster operation, leakage power has increased substantially. It is observed that Dynamic power dissipation has been almost constant going down the flow whereas Static dissipation has increased three to four times every generation. As die sizes and integration increase, static power will make up an increasingly large portion of the overall power since leakage current emanates from every transistor that is turned on.

B. Transistor stacking

It is a technique that is typically used in active mode. By keeping a few transistors in series, leakage power can be reduced. In a typical CMOS circuit, stacking means connecting transistors in a series fashion from Vdd or output to ground. When transistors are connected in series and when transistors are off then their effective leakage resistance becomes higher than a single transistor. This leads to a higher leakage current. But when these transistors are connected in parallel then, the leakage current will find parallel paths and the effective leakage current will increase. Whereas, if we have several of these transistors in series which is called stacking, then the overall resistance will be some of the off resistances of all these transistors, which will result in less leakage current to flow. This type of transistor stacking helps in reducing the overall flow of leakage current through the stack. Forced stacking can be used for circuits without a stacking structure. A single transistor of width W can be replaced by two transistors of width W/2 each in forced stacking, as seen in fig. Due to the simultaneous off-switching of two transistors, the leakage current is decreased.

Fig. 7. Forced stacking

C. Dual Threshold Partitioning (DTMOS)

It is a very popular technique used in fabrication units. By changing the doping levels of the different regions during fabrication, the threshold voltage of the transistors can be changed. The threshold voltages of the transistors have a direct impact on the speed of the devices and the circuit netlist consists of a large number of transistors, some of these transistors we classify as low threshold, and some of the transistors as high threshold. Low threshold voltages run faster, but they incur more leakage current. So, there is a trade-off. Circuit partitioning can be done on the basis of critical paths and leakage current. For the circuits which do not need high speed, High threshold transistors can be set which will also have low leakage current. Whereas, for the circuits which are timing critical and need high speed, Low threshold transistors can be used. In this way, partitioning can be done by analyzing critical and delay requirements of the circuit, so that circuit works in an efficient way and yet consumes less power.

D. Variable Threshold Transistors (VTMOS)

It is a technique in which transistors threshold voltages are adjusted dynamically from their final level to many levels in active mode. This technique is also known as Standby Power Reduction (SPR). By changing the substrate bias voltage, the threshold voltage can be changed. This can be achieved by connecting two voltage sources to the substrate, one for pmos and the other for nmos. The approach is to increase the threshold voltage in standby mode by connecting the substrate to a voltage lower than the ground for nmos. Similarly, by connecting to a higher voltage than ground for pmos. The drawback of this technique is that we need the extra power supply for this operation which will increase the cost.

E. Power gating (MTCMOS)

It is a general technique to decrease static current when in the sleep state by turning off the power supply to the sleeping blocks. There are VDD and VSS supplies, in addition to them, an auxiliary voltage at a lower level of VDV and VSSV is added. The circuit can be set to sleep mode by activating these transistors by setting the SL to 1. In this mode, transistors are switched off and thus these transistors are now powered by lower voltage (power down mode). Hence, power consumption is reduced. To eliminate the issue of data loss in the case of storage elements like FFs, this technique is used by not doing power off but instead just lowering the voltage level. This technique, if used in a proper way, has been demonstrated that we can reduce power by as much as 1000 times but it requires a lot of analysis, circuit level analysis, and calculation to find out which blocks are needed to implement or incorporate these sleep modes.

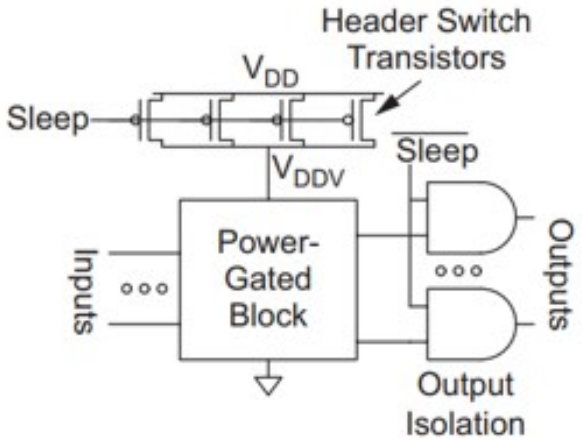

Fig. 8. Power Gating method-1

VI. Inverse Narrow Width Effect (INWE)

The sizing of the transistors in the subthreshold region is different from the normal sizing due to various effects like inverse narrow width effect and reverse short channel effect

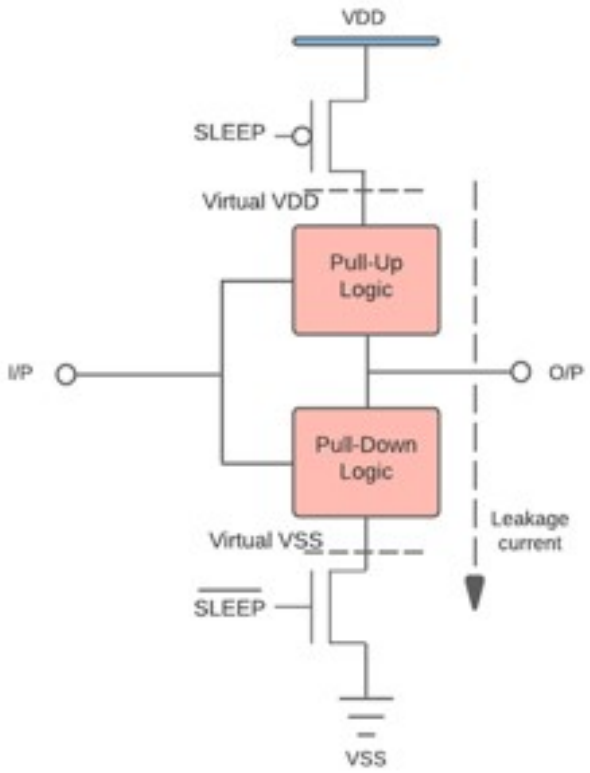

Fig. 9. Power Gating method-2

which is observed in the subthreshold region. The effect of INWE is not significant in the superthreshold region however here it plays a major role.

The threshold voltage of transistors that have narrow width rises quickly as the width of the transistor is increased due to the inverse narrow width effect. Because of the increase in threshold voltage, there is a non-proportional relation between the drain current and transistor width when the width is raised from the minimum width in a given technology node and thus the current rises considerably slower than what is anticipated. Due to the fact that the sub-threshold current is related to the threshold voltage in an exponential manner, this impact becomes more important at a lower value of operating voltages and has a prominent effect in the sub-threshold operation.

A parasitic transistor that is present at the corners during the process of the shallow trench is responsible for the inverse narrow width effect. A sidewall that is orthogonal to the channel is created as a result of STI which creates an abrupt end. Due to fringing which results from the overlapping of the polysilicon gate and the perpendicular arrangement, the electric field strength at the corners of the channel formed is significantly high. Increased channel control along with deep depletion zones leads to a decrease in the threshold voltage of the device. When the width is narrower, the fringing fields in the corners take up more control thus impacting the threshold voltage. The parasitic transistor gets turned on at a lower value of voltages than the main channel due to the geometric impact

of the shallow trench corner.
A rise in the ratio of on to off current is also a result of an increase in the depletion depth at the corners of the channel. The parasitic transistor adds significantly to the performance of the transistor as the transistor width shrinks. Because of this, the threshold voltage drops for a narrower transistor having narrower widths. The threshold voltage rises as we start increasing the width and after a certain value of width it gets constant.
The current of a transistor has quadratic relation with its threshold voltage and therefore, the current is seldom affected by the relatively tiny fluctuation in the value of the threshold voltage. The current and threshold voltage is related exponentially in the subthreshold domain. Due to the wide range of values of threshold voltage because of INWE, the current-width relationship is highly unanticipated. The raised threshold voltage causes the current to rise considerably slower than what is anticipated as we start increasing the width of the transistor from the minimum width. At lower nodes and narrower widths, the change in values of threshold voltage gives the non-proportional relationship between current and width.

VII. Device Sizing in the subthreshold region

Generally, while sizing a device, it is typically assumed that the width of the MOSFET has no effect on the transistor's threshold voltage and as a result, the increase in the value of current is proportionate to the way we increase the device width. However, due to the INWE phenomenon in the subthreshold region, the width has an effect on the threshold voltage of the transistor, and the current is exponentially related to the threshold voltage. In order to attain the required proportionate value of the drain current, a large increase in the value of width is required. This results in a penalty on PPA leading to large delay, area, and power overhead.
In the region of narrower widths, with an increment in width of transistor threshold voltage rises. As a result, the width of the transistor is increased from the minimum width, the drain current increases comparatively at a slower rate than anticipated because of the increment in threshold voltage simultaneously. Thus, in order to reach the required current, a further increase of width is required. Appropriate sizing of a device becomes less predictable because of the non-proportional relation between transistor current and its width. Therefore, determining the correct transistor size necessitates lengthy simulations complicating the design.

A. Sizing of the devices in an INWE-aware manner

To reduce the impact of INWE in the subthreshold region on the drain current, the technique of dividing the transistor into multiple fingers with a minimum width is adopted which helps in improving the PPA. The gain from this approach is that the width and threshold voltage of the transistor stays constant at the minimum width of the given technology node ensuring that an increase in drain current directly corresponds to the increase in transistor width. Approximately, the current changes in a linear manner as the width of the transistor is increased. Additionally, the current is maximized when the design is done with minimum width as the threshold voltage is lowest at the minimum width.
Thus, this method allows for a significant reduction in width as compared to normal sizing while retaining the current. The transistors scaled using this approach have substantially less width than those made using the standard way while still supplying the same amount of current with a significant decrease in area and power. The desired width can be made as = minimum width + no of fingers or multipliers.

B. Impact on the layout

Having the transistor divided into numerous minimum-width transistor fingers or multipliers has an impact on the layout space. Due to design rule checks between different layers, source and drain are distant in finger-based layout thus increasing the area. When multipliers are used, the arrangement is more compact however the multipliers must be spaced apart. All such factors lead to an increment in the area while arranging the transistor and by the usage of minimum width finger the amount of increment in current is significant, also it helps in saving the area.

VIII. Effect of variations in the subthreshold region

During the process of chip fabrication, there are process variations around the desired value of the design parameter that needs to be taken into consideration. These process variations and fluctuations include inter-die (one die to another) and intra-die (within a die) variations. Inter die variations occur due to property of equipment and temperature of operation, whereas intra-die variations originate from factors such as dopant concentration, channel length variations, oxide thickness, doping concentration, and transistor width, these variations have become more important as they directly impact the power and frequency.
Parameter variations have led to increase in the number of leakage currents, therefore leakage currents are necessary to take into consideration as they drive the circuit. To understand its significance, for example, a 3-fold significant difference in sub-threshold leakage current can be created with a 10 percent increase in channel length, similarly, the same variation in gate oxide thickness can cause a 15 times difference in current.

A. Random Dopant Fluctuations (RDF)

Subthreshold designs are sensitive to process variations due dependence of subthreshold current with variations in Vth. Random dopant fluctuations are the main component of variations in subthreshold design. Subthreshold current is exponentially impacted by any variations in Vth and Leff. At lower operating voltage, DIBL and punch through are reduced. Therefore, the variations in Vth that were occurring because of channel length variations are diminished significantly. Therefore variations in Ion(on current) current due to DIBL are nearly constant only but Ion variations due to random doping

fluctuation become dominant as Vdd approaches near to Vth. Random doping fluctuations variations can be addressed by device sizing and using larger logic depths because the increase in W(size) and n(no of inverters) can reduce the number of variations.
Random variations at the subthreshold region is given as,
sigma(Vth) is proportional to (sqrt(W L))-1
where the Vth standard deviation is roughly So, increasing the width and length of the device is a solution (countermeasure) to this problem. The second method includes where a device is stacked in parallel with a duplicate, this helps in t reduction of the variability of a single device as compared to a composite device.

B. Physical Gate Length (Lg)

Since the source and drain regions are close, so the depletion regions of the source and drain penetrate considerably into the channel and hence the channel length decreases, and the threshold voltage also decreases. Consequently, this leads to a reduction in total charge in the channel in comparison to the long channel device. Therefore, less charge needs to be inverted by the applied gate voltage to achieve vth. This reduction in channel length leads to an increase in subthreshold current and hence increases the amount of power consumption.

C. Oxide Thickness

Any variation in oxide thickness impacts the capacitance, which directly impacts the Vth and current. As the reduction in gate oxide thickness exponentially increases the gate leakage current. Therefore, crucial analysis needs to be done for the oxide thickness, to meet the desired device delay and leakage values.

D. Junction Depth

Shallow junction depth helps in reducing the short channel effects and avoids the penetration of depletion width into the channel, this leads to an increase in the amount of sheet resistance and also makes the fabrication process complex. Therefore, it is desirable to keep the parasitic resistance low to achieve sufficient current drive capability.

E. Channel Upsizing

Since subthreshold logic faces a lot of challenges, such as subthreshold swing, DIBL, and Energy gain hence to address optimal performance, channel upsizing is a good approach. Thus, increase in channel length removes the DIBL effect, reducing the leakage current and further increasing the threshold voltage. Hence by relation sigma(Vth) is proportional to (sqrt(W)) L-1, the increase in channel length decreases the variation in threshold voltage and hence makes the device more robust.

IX. Sequential Circuits

Characterization of Flip flop for low power application operating in subthreshold region:
With the advent of technology scaling, transistor sizes scale down, and there is an increase in the power dissipation of digital circuits due to leakage and increased density. The subthreshold operation has become an efficient approach to decreasing both the static and dynamic power dissipation of a circuit. Specifically, we are discussing in this section behavior and operation of flip flop design in the subthreshold region. We are extending our studies to different flip flop design architectures, their compatibility in subthreshold operation, and a kind of comparative analysis in terms of PPA. Our insight is more over basic parameters of flip-flop like Clock-q delay, max operating frequency, setup and holds time, and metastability.
An optimal sequential design should have low power dissipation, offer no significant delay and inherit a valid stable output at all times. While designing the flip flop, there must be consideration of trade-offs between speed, power, and basic operational requirements. While analyzing flip for the subthreshold region, the speed is its foremost considerable parameter. As we are reducing the supply below the threshold it certainly hampers its operating frequency which can be measured by the total delay required for a valid output is available after the input has been stabilized.
tdelay = tsetup + tclk-q delay
where tsetup is the time taken for the input to stabilize in the flip-flop and tclk-q is the time taken from the clock goes high to a valid output Q is available. From this conclusion, we get the following equation for the maximum delay of a flip-flop:
tdelay = max ((tsetup + tctq),HL ,(tsetup + tctq),LH))
where HL is high-to-low transition and LH is low to high transition. Another factor is power dissipation. The dynamic power dissipation of a flip-flop is input data-pattern dependent and is directly proportional to the switching activity alpha. When an alternate sequence of input is provided in that case high power dissipation occurs i.e., alpha =1, while there will be minimum power dissipation in case of a constant high (1) or low (0) input signal (alpha = 0). For the characterization of flip-flop cells performance in subthreshold operation, we are giving a few widely used and explained different flip-flop cells.

A. Transmission-gate Master-Slave flip-flop

The TGMS flip-flop is made up of two transmission gate-based latches operating on complementary clocks. This flipflop may be sensitive to clock-skew of its two complementary clock-phases.

B. C2MOS flip-flop

This flip-flop in comparison to the TGMS is insensitive to overlap clocks. It consists of clocked 4T and basic 2T inverters and relies on two clock phases.

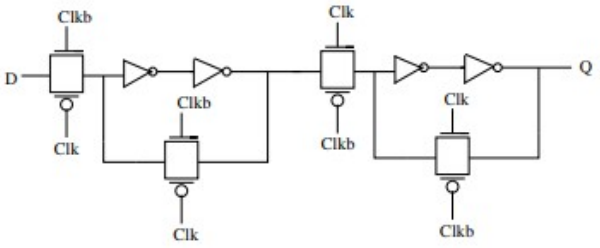

Fig. 10. TG based D Flip-Flop

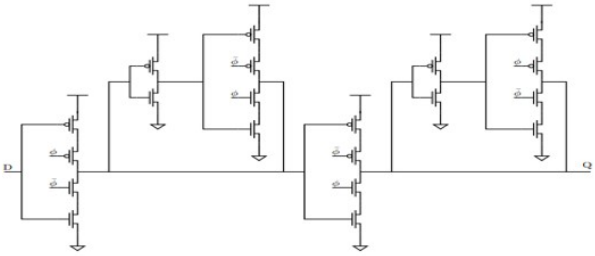

Fig. 11. C2MOS Flip-Flop

C. D flip-flop using Pass-Gates

Normally, transmission gates are used to achieve symmetrical transmission of logic high and logic low signals which is more advantageous in terms of transmission at both levels. While the pass gate design takes advantage of the area and single clock phase. advantages of using single transistors as pass gates rather than 2T transmission gates; the trade-off is signal integrity loss.

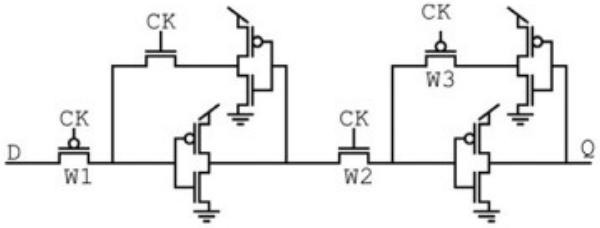

Fig. 12. D Flip-Flop using Pass-Gates

D. PowerPC 603 flip-flop

This Kind of flip-flop works in subthreshold design region operations for the case of static; it is a combination of the TGMS and C2MOS flip-flops, using clocked inverters instead of feedback transmission gates. It was earlier used in the PowerPC 603 microprocessor data-path from where it gets its name.

Regarding PDP, the PowerPC 603 DFF stands out as the finest architecture. The data from the publication to which we have linked confirm these conclusions, but it is also helpful to analyze the DFF results in order to better understand the pass gate flip-flop, which outperforms the PowerPC 603 design

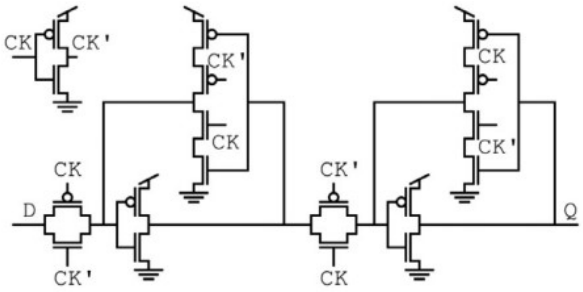

Fig. 13. PowerPC 603 flip-flop

overall on the basis of power delay product, setup and hold variations, leakage power, and energy/cycle when vdd¡ 400 mV except for maximum clock speed and energy per cycle outside of the subthreshold supply voltage region.

X. Conclusion

The paper discusses about various aspects that comes into picture when we are designing a sub-threshold library, for which we have a limit for operating voltage, i.e., around 0.4 V. Here we have discussed about several new effects like INWE, how to size our devices in the sub-threshold region, effects of variations in the sub-threshold region. We also talk about the design of sequential circuits like D flip-flop, seen different architectures. We also saw different techniques for controlling dynamic and static power dissipation.

In future, one can start designing the components for our sub-threshold library with proper defined values for different parameters of the design, by the learning from this paper.

XI. References

[1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18]

References

[1] S. Amarchinta, H. Kanitkar, D. Kudithipudi, Robust and high performance subthreshold standard cell design, in: 2009 52nd IEEE International Midwest Symposium on Circuits and Systems, IEEE, Cancun, Mexico, 2009, pp. 1183–1186. doi:10.1109/MWSCAS.2009.5235946.
URL http://ieeexplore.ieee.org/document/5235946/

[2] P. Arun, S. Ramasamy, A low-power dual threshold voltage-voltage scaling technique for domino logic circuits, in: 2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12), IEEE, Coimbatore, 2012, pp. 1–6. doi:10.1109/ICCCNT.2012.6395896.
URL http://ieeexplore.ieee.org/document/6395896/

[3] S. Bhutada, A. Asati, A. Dubey, Design of ultra low power flip flops in sub-threshold region for bio-medical application in 45nm, 32nm and 22nm technologies, in: 2015 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), IEEE, Coimbatore, India, 2015, pp. 1–5. doi:10.1109/ICECCT.2015.7226146.
URL http://ieeexplore.ieee.org/document/7226146/

[4] Bo Zhai, S. Pant, L. Nazhandali, S. Hanson, J. Olson, A. Reeves, M. Minuth, R. Helfand, T. Austin, D. Sylvester, D. Blaauw, Energy-Efficient Subthreshold Processor Design, IEEE Transactions on Very Large Scale Integration (VLSI) Systems 17 (8) (2009) 1127–1137. doi:10.1109/TVLSI.2008.2007564.
URL http://ieeexplore.ieee.org/document/4895693/

[5] M. Horowitz, T. Indermaur, R. Gonzalez, Low-power digital design, in: Proceedings of 1994 IEEE Symposium on Low Power Electronics, IEEE, San Diego, CA, USA, 1994, pp. 8–11. doi:10.1109/LPE.1994.573184.
URL https://ieeexplore.ieee.org/document/573184

[6] R. Jaramillo-Ramirez, J. Jaffari, M. Anis, Variability-aware design of subthreshold devices, in: 2008 IEEE International Symposium on Circuits and Systems, IEEE, Seattle, WA, USA, 2008, pp. 1196–1199. doi:10.1109/ISCAS.2008.4541638.
URL http://ieeexplore.ieee.org/document/4541638/

[7] D. J. Judy, V. S. K. Bhaaskaran, Review and Analysis of the Impacts and Effects on Low Power VLSI Circuits Operating in Subthreshold Regime, International Journal of Engineering and Technology 5 (5) (2013) 14.

[8] E. Låte, A. A. Vatanjou, T. Ytterdal, S. Aunet, Extended Comparative Analysis of Flip-Flop Architectures for Subthreshold Applications in 28 nm FD-SOI, Microprocessors and Microsystems 48 (2017) 11–20. doi:10.1016/j.micpro.2016.07.016.
URL https://linkinghub.elsevier.com/retrieve/pii/S0141933116301004

[9] B. B. Liu, Standard cell library design for sub-threshold operationPublisher: Technische Universiteit Eindhoven (2014). doi:10.6100/IR782367.
URL https://research.tue.nl/en/publications/standard-cell-library-design-for-subthreshold-operation(11c482b9-9e49-4bfa-9584-efe9fa4fb707).html

[10] J. Morris, P. Prabhat, J. Myers, A. Yakovlev, Unconventional Layout Techniques for a High Performance, Low Variability Subthreshold Standard Cell Library, in: 2017 IEEE Computer Society Annual Symposium on VLSI (ISVLSI), IEEE, Bochum, Germany, 2017, pp. 19–24. doi:10.1109/ISVLSI.2017.14.
URL http://ieeexplore.ieee.org/document/7987489/

[11] M. Nabavi, F. Ramezankhani, M. Shams, Optimum pMOS-to-nMOS Width Ratio for Efficient Subthreshold CMOS Circuits, IEEE Transactions on Electron Devices 63 (3) (2016) 916–924. doi:10.1109/TED.2016.2517446.
URL http://ieeexplore.ieee.org/document/7400979/

[12] Nam Sung Kim, T. Austin, D. Blaauw, T. Mudge, K. Flautner, Jie S. Hu, M. Irwin, M. Kandemir, V. Narayanan, Leakage current: Moore's law meets static power, Computer 36 (12) (2003) 68–75. doi:10.1109/MC.2003.1250885.
URL http://ieeexplore.ieee.org/document/1250885/

[13] M. Priya, K. Baskaran, D. Krishnaveni, Leakage Power Reduction Techniques in Deep Submicron Technologies for VLSI Applications, Procedia Engineering 30 (2012) 1163–1170. doi:10.1016/j.proeng.2012.01.976.
URL https://linkinghub.elsevier.com/retrieve/pii/S1877705812009861

[14] J. F. Ryan, J. Wang, B. H. Calhoun, Analyzing and modeling process balance for sub-threshold circuit design, in: Proceedings of the 17th great lakes symposium on Great lakes symposium on VLSI - GLSVLSI '07, ACM Press, Stresa-Lago Maggiore, Italy, 2007, p. 275. doi:10.1145/1228784.1228853.
URL http://portal.acm.org/citation.cfm?doid=1228784.1228853

[15] M. Sharma, K. G. Sharma, T. Sharma, B. P. Singh, N. Arora, SET D-flip flop design for portable applications, in: India International Conference on Power Electronics 2010 (IICPE2010), IEEE, New Delhi, India, 2011, pp. 1–5. doi:10.1109/IICPE.2011.5728081.
URL http://ieeexplore.ieee.org/document/5728081/

[16] C. Wang, J. Zhou, X. Liu, A. M. Annamalai, M. Je, A sub-threshold to super-threshold Level Conversion Flip Flop for sub/near-threshold dual-supply operation, in: 2014 IEEE Asian Solid-State Circuits Conference (A-SSCC), IEEE, KaoHsiung, Taiwan, 2014, pp. 41–44. doi:10.1109/ASSCC.2014.7008855.
URL http://ieeexplore.ieee.org/document/7008855/

[17] B. Zhai, S. Hanson, D. Blaauw, D. Sylvester, Analysis and Mitigation of Variability in Subthreshold Design 6.

[18] J. Zhou, S. Jayapal, B. Busze, L. Huang, J. Stuyt, A 40 nm Dual-Width Standard Cell Library for Near/Sub-Threshold Operation, IEEE Transactions on Circuits and Systems I: Regular Papers 59 (11) (2012) 2569–2577. doi:10.1109/TCSI.2012.2190674.
URL http://ieeexplore.ieee.org/document/6183496/