22FDX Ultra Low Power Design with Dynamic Body Biasing and Liberty Variation Format

Suman Dwivedi
Design Methodology-CAD
GLOBALFOUNDRIES
Bangalore,India
Email: suman.dwivedi@globalfoundries.com

Ulrich Hensel,Chenbo Liu
Design Methodology-CAD
GLOBALFOUNDRIES
Dresden, Saxony, Germany
Email: ulrich.hensel@globalfoundries.com, chenbo.liu@globalfoundries.com

RamyaSrinivasan,Siddharth Sawant,Haritez Narisetty
Design Methodology-CAD
GLOBALFOUNDRIES
Santa Clara,California,USA
Email: ramya.srinivasan@globalfoundries.com, siddharth.sawant@globalfoundries.com , haritez.narisetty@globalfoundries.com

***Abstract*—In this paper, we are demonstrating how 22FDX dynamic body bias enables low voltage design operating at nominal voltages of 0.4V, 0.5V, and 0.65V using the ULP_LV platform. Basic concepts of dynamic body bias for low voltage design are introduced. The necessity of Liberty Variation Format (LVF) deployment due to high local variations at worst corners under low operating voltages is discussed. The paper is detailing how LVF is used through the tool chain from library characterization to timing sign-off and concludes with example block level implementation results**

Keywords— Body Biasing, LVF-liberty variation format, Sensitivity based analysis

I. INTRODUCTION

Integrated circuit designs today require smallest possible power consumption to support mobile applications and to enable the IOT (Internet of Things) market. However, this cannot be achieved anymore at reasonable cost by moving to even smaller technology nodes. Just shrinking planar technologies has started to lead to increased number of required masks because double patterning for multiple routing has become a requirement. For FinFET technologies, which could enable smaller designs at lower power consumptions, the more complex manufacturing process compared to planar technologies, comes with higher cost as well. This drives the motivation to stay with a planar technology and reduce the power consumption by other means. As a result low power features, which use different kinds of voltage level adaptation, have become more and more popular, and the scope varies from supply voltage to bias voltage variations and from fixed power supply levels to tunable power supply levels. All these features can be used separately or in combinations to the full extent by GLOBALFOUNDRIES' 22FDX technology, which is an FDSOI- (Fully Depleted Silicon on Insulator) technology and this comes with a cost close to a 28nm planar technology but with a smaller area than 28nm.

II. MOTIVATION

GLOBALFOUNDRIES 22FDX provides technology features that allow designs to operate at very low supply voltage down to 0.4V, which is ideal for IOT devices. Most prominently, 22FDX enables body biasing as a means to effectively change the threshold voltage (VT) dynamically after fabrication. The challenge is that at low operating voltages, local delay variations at worst case corners are extremely high. Advanced On-Chip Variation (AOCV) based methodologies with local uncertainty margins lead to area bloat or non-converging designs. Deployment of Liberty Variation Format (LVF) through the tool chain is a must to close timing.

III. THE FDSOI TECHNOLOGY

As experienced over the last years, just continuing to shrink a transistor size does not always result in the desired improvements for a given product in performance, area and/or power at reasonable cost - hence a change in architecture is required. However, unlike other technologies FDSOI does not change the construction of the transistor substantially. The innovation in FDSOI is rather based on adding a thin insulation layer (referred to as buried oxide) just below the transistor channel, and in eliminating the need to add doping to the channel, thus making it fully depleted. The comparison of a bulk technology transistor to an FDSOI transistor is shown in Figure 1:

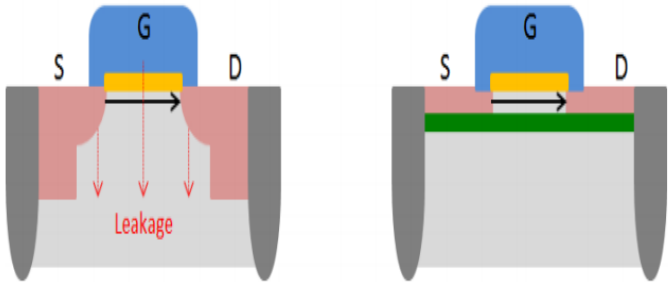

Figure1. Comparison of a Transistor in Bulk Technology and FDSOI

One advantage of the FDSOI transistor structure is that effective use of body biasing can be made which means applying a voltage to the substrate to improve the transistor performance. While in bulk technologies the ability to use body biasing is limited due to increased parasitic leakage current effects, the FDSOI transistor with its insulating layer below the channel allows for higher body biasing. This means that higher transistor performance can be achieved while still keeping leakage current low. This effect is shown in Figure 2:

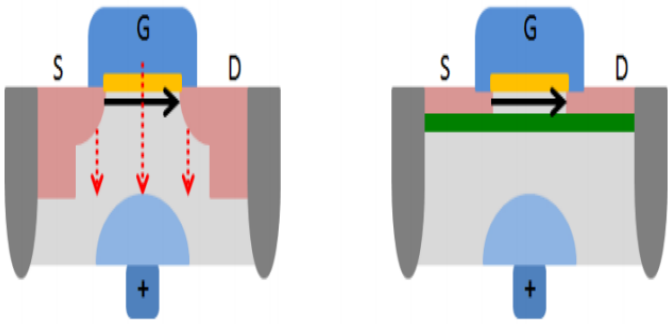

Figure2. Effects of Body Biasing in Bulk Transistor and FDSOI Transistor

Off-State leakage current is the current that leaks through the transistors even when they are turned off and has three components in CMOS circuits:

- Subthreshold leakage current - which is a leakage current across the device when the transistor is operating in weak inversion region. Weak inversion region means the gate voltage (Vg) is lesser than the threshold voltage.
- Gate Leakage, is due to the tunneling phenomena where the electrons (holes) from the bulk and source/drain overlap region tunnel through the gate oxide barrier into the gate. Gate leakage is also called as the gate oxide tunneling current.
- PN-junction reverse-bias current, which is due to the reverse biased PN-junctions between drain and source to well/substrate and this reverse bias constitutes to leakage current. The leakage current in FDSOI technology is substantially reduced because of the thin insulation layer (referred to as buried oxide). The buried oxide layer cuts off the pn reverse bias junctions and hence eliminate the pn junction reverse bias current. In Figure 1, the black arrow represents the subthreshold leakage current and the red dotted arrow represents the gate leakage and PN junction reverse bias currents. Another advantage of FDSOI technology is that variation of transistor behavior across a die caused by different amounts of doping at different locations is reduced. This means that for example timing variation across a die can play a less important role, hence allowing the use of less pessimistic timing constraints for design closure and sign-off STA.

IV. PROPOSED TECHNIQUE

22FDX Dynamic Body-Biasing

In FDSOI technology the voltages applied to the transistor gate and to the substrate can be independently adjusted. This allows changing the behavior of the same transistor from a high performance to a low power transistor, even during run-time of the chip in case an adaptive bias voltage generator is used. Two modes of operation can usually be looked at:

- Forward body biasing (FBB), for which a positive substrate voltage is applied to the NMOS devices and a negative substrate voltage is applied to the PMOS devices, hence lowering the effective VT of these devices, which puts them into a high-performance mode. The FBB case is using a flipped well approach as shown in Figure 3 with a NMOS transistor over a N-well.
- Reverse body biasing (RBB), for which a negative substrate voltage is applied to the NMOS devices and a positive substrate voltage is applied to the PMOS devices, hence raising the effective VT of these devices, which puts them into a low-leakage or even stand-by mode. The RBB case is using a “bulk like” well assignment with an NMOS transistor over a P-well. For completeness also a so called zero-bias mode could be listed, for which the substrate voltages for NMOS devices as well as PMOS devices are 0V or GND. A connectivity example for FBB is show in Figure 3:

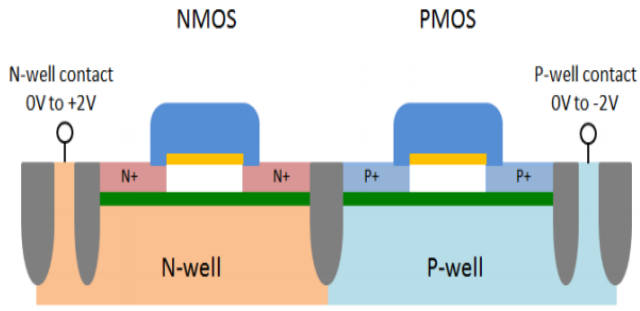

Figure3. Connectivity Example for Forward Body Biasing with Flipped Well

Dynamic Threshold Voltage Shift through Forward Body Biasing Playing with Body Biasing voltage gives designers a new tool to optimize their design to achieve timing closure. Threshold voltage is reduced by applying FBB to LVT and SLVT devices. The device is then able to work at higher frequency at the cost of higher leakage. On the other hand, by applying RBB to RVT and HVT devices, VT is increased and the devices can work at lower

frequency with lower leakage. However, this is out of the scope of this discussion.

By reducing the effective VT, FBB is increasing the overdrive – the delta between supply and threshold voltage – and, thus, the performance of the transistor. Increasing the supply voltage is an alternative way to increase the overdrive. This traditional approach is commonly applied in Dynamic Frequency Voltage Scaling (DVFS), supply voltage based corner trim or adaptive voltage scaling (AVS) techniques also for Bulk and FinFet technologies.

Now with FDX, the designer can optimize power consumption by choosing how to increase the overdrive. FBB increases the performance at the cost of increased leakage; supply voltage increase pushes the performance at the cost of increased dynamic power – as the dynamic power will rise in quadratic relationship to the supply voltage increase. Of course, also a combination of supply voltage increase and FBB is possible – this enables a wider range of dynamically changing the overdrive and thus dynamically tuning the performance of the design.

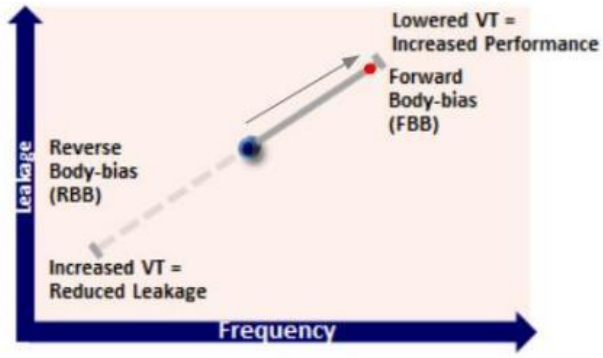

Figure4: VT Manipulation's Impact on Frequency and Leakage

Need to review the two major design approaches to make use of dynamic FBB:

1. Dynamic Voltage Body Bias Frequency Scaling (DVBFS) is akin to the well-known Dynamic Voltage Frequency Scaling (DVFS). As described above, the designer can utilize increasing the supply voltage or the body bias voltage dynamically in order to achieve a higher frequency in a specific functional mode – often called boost mode – of the design. This paper focuses on low voltage design and its impact to variability modeling with LVF, so DVBFS is not further discussed here.
2. Delay variation compensation applies FBB in order to reduce the delay variation distribution. That is, slow delay conditions can be selectively accelerated by applying FBB. Delay variation compensation is also known as **corner trim**. Delay variation compensation using FBB (and not supply voltage scaling) is a key prerequisite to enable the design at ultra-low supply voltages such as 0.4V or 0.5V. This paper focuses on delay compensation or corner trim for the purpose of low voltage design.

Trim - Delay Compensation through Forward Body Biasing

The essence of delay compensation through FBB is to lower the VT to speed up the design selectively. The delay distribution diagram in Figure 5 shows the principle of operation for 2 different trim targets. The blue distribution of the curve shows a normal delay distribution, the red curve depicts how the delay distribution would change by applying the red dotted bias voltage to the delay in blue distribution. Apparently, this scenario regulates for a trim target of the original typical delay , that FBB is larger for slow conditions and is reduced until it reaches 0 for typical unchanged. The green curve shows a different trim target (Super Trim).This time, start with a higher FBB for the slow delay condition and ramp FBB down towards the fast delay condition.

So even typical and faster delays get acceleration by FBB, only the fastest stay at default body biasing delay.

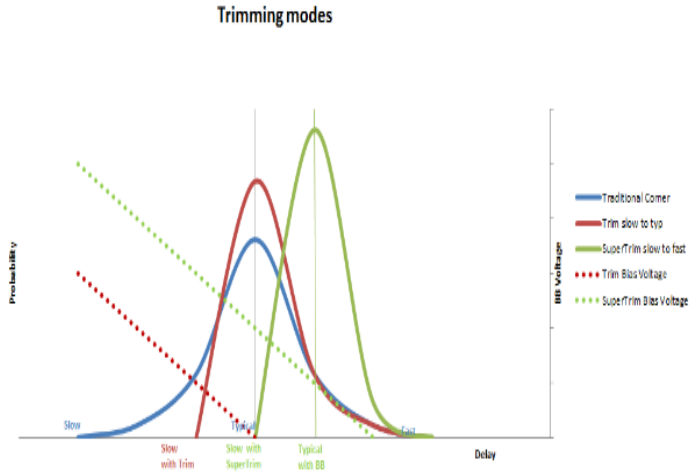

Figure5. Different Trimming Target Moving Delay Distribution Diagram

FBB Delay compensation works in principle through the following algorithms :

1. Measure die delay
2. Adjust Body Biasing voltage to regulate die delay to a trim target.
3. Effective squeeze delay distribution
4. Design with trim corners

LVF as Must-have Characterization to Cope with High Local Variability with increasing complexity and operating

voltages getting lower and closer to threshold voltages of devices, variation and variation modeling become important aspects of timing closure. The variation modeling solutions that exist at higher nodes don't accurately model the variation as we move towards lower tech nodes.

AOCV vs. LVF

AOCV:

- Only one single input transition/output load
- Model based on depth and distance
- Derate only for delay
- Need to add uncertainty margins

LVF:

- Variation data for each slew/load combination
- Sigma for each timing arc
- Model variation for delay, setup and hold constraints, and transition

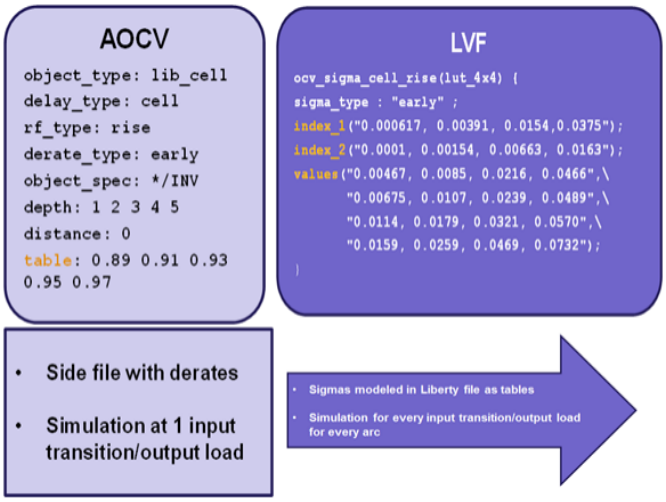

Figure 6 AOCV vs. LVF

In addition, LVF is not only modelling delay, but also modelling variation for setup and hold constraints, as well as transition. LVF based variation modeling for setup/hold constraints removes the necessity to model setup/hold variation as part of (global) clock uncertainty constraints. To summarize, LVF provides a fine grain representation of timing variation.

- The constraints vary based on the slew on both Data_pin and CLK_pin, or input slew and load for delay and slew.
- Characterization of delay,slew and constraint sigma for every input transition, output load and timing arc.
- Early and late sigma is modeled separately. Each statistical parameter from variation mode is modeled independently.

There are two methods available in industry tool for LVF characterization.

1. Monte-Carlo Method: Typically, Monte Carlo (MC) analysis is used for statistical modeling, and is considered as golden reference method in terms of accuracy

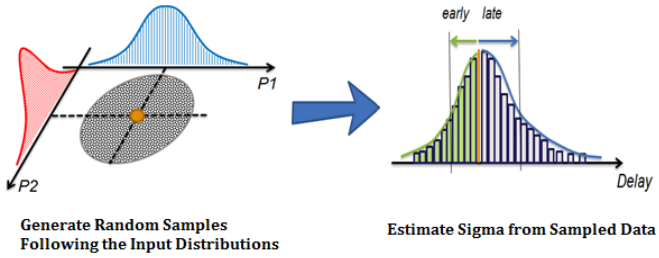

Figure 7: Monte Carlo -Method

Based on the sample size set in the library characterization tool MC LVF flow, the tool generates random samples for the statistical variation parameters in the spice model and estimates early and late sigma from the sampled data. MC method is the most accurate method for statistical modeling however it comes with a cost of really high runtime.

2. Sensitivity Based Analysis: Due to the exhaustive number of simulations in LVF, MC is not the best optimized solution in terms of library characterization time. Industry tools have an alternative method named **Sensitivity Based Analysis (SBA)** that solves this problem. This method is recommended in production environment.

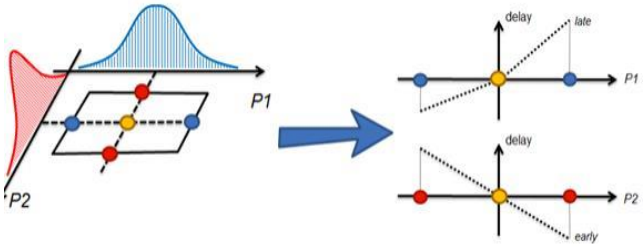

Figure 8: Sensitivity Based Analysis

SBA uses statistical parameters from transistor spice model, samples them at specific points, and computes sensitivity to each parameter to come up with variation numbers. Compared to MC, it runs less number of samples and also uses other optimizations like netlist pruning, binning, and screening and reduction factor, to reduce the number of simulations. Hence, the runtime for characterization is reduced.

Best Practices in Digital Flow with LVF:

No More Setup and Hold Uncertainty and Max Transition

- Using a fully LVF characterized library including variation tables for setup and hold constraints in the .lib allows the removal of global clock uncertainty margins except for the modeling of external PLL jitter that is still needed for setup uncertainty. The timing analysis will accurately model the delay variation along the cells

of the logic paths all the way to the setup and hold constraints at the receiving sequential element and also take slew/load or slew/slew influences to the variation into account. The designer is just left with defining the desired local variation deviation – the tool defaults assume a 3-sigma here.

- For early design stages before fully synthesized and routed clocks (pre CTS) , we recommend to add a setup uncertainty for the unknown clock tree implementation. This is a standard approach also used in AOCV based methodologies. This uncertainty needs to be removed post CTS to avoid over design
- Throughout the flow, hold uncertainty shall be set to zero.
- Another benefit brought in by LVF analysis is that slew dependency of variation is accurately modeled even for very large slews. This of course depends on the characterization range of the available library.
- For accurate variation calculation in timing analysis, the timing engine calculates the actual transitions and transition variations and derives delay and setup/hold constraint variations from that.
- This is in contrast to the AOCV methodology, where setup and hold uncertainty margins are applied globally and only are valid for all paths if a global max transition constraint is not violated.
- Such a global max transition constraint is not needed anymore for an LVF based timing signoff flow.
- This leaves new options to the designer during implementation, for instance to relax transition constraints for slow paths or paths that are not that prone to cross-talk.
- Of course, designers should still care for max transition in general because large slews are bad for design convergence, make designs prone to cross-talk and thus may cause bad quality of results metrics.

V. BENCHMARK RESULT

CASE 1: AOCV VS. LVF AT LOW VOLTAGE

Table 1: Detail of Case 1

Case 1	**SIMD Engine**
Target Frequency	1.25Ghz at Vnom-10% = 0.72V
Max Tran	100/200ps
Uncertainty for hold (in case of AOCV equivalent to 3 sigma)	55ps
Uncertainty for setup (in case of AOCV equivalent to 1 sigma)	18.25ps
Uncertainty for setup and hold (in case of LVF equivalent to 3 sigma)	3 - sigma (default)
PLL Jitter used for both AOCV and LVF	Not Applied Jitter

Table 2: LVF vs AOCV tool run time

HH:MM:SS	**Syn**	**init_design**	**place_opt**	**clock_opt**	**route_auto**
AOCV	1:38:26	0:08:38	0:31:12	0:18:31	1:10:59
LVF	1:20:58	0:08:30	0:32:28	0:26:23	0:40:52

The run times do not show a significant difference until after clock_opt_cts. Starting with clock_opt_opto the AOCV run has to resolve a higher number of timing violations than the LVF run and therefore the runtime for AOCV increases.

Table 3: QOR metrics at chip finish in P&R Tool

	Setup WNS	**Setup TNS**	**Setup NVP**	**Hold WNS**	**Hold TNS**	**Hold NVP**
AOCV	-21.5	-450	89	-69.7	-9050	546
LVF	-4.38	-17.04	12	0.98	0	0

Table 4: QOR metrics at chip finish for area in P&R Tool

	Cell Area	**Comb Area**	**NonComb Area**	**Buf Area**	**Inv Area**
AOCV	69708	47630	21906	14801	3007
LVF	56545	34638	22078	303	2666
% diff in area	18.88	27.27	-0.78	97.95	11.31

Table 3 and table 4 show that the AOCV run has to spend a higher effort to close on HOLD because AOCV with flat clock uncertainty has a very pessimistic modelling of 3-sigma hold variation. This results in longer run time but also in larger design area. The overhead is primarily a much larger buffer area because of HOLD fixing.

At very similar constraints (even tighter setup at 3sigma for LVF), the LVF run converges much faster. HOLD is not an issue because of the fine granular modelling of hold constraints within the cells in the liberty file. The STA reports shows a similar result on hold metrics

Table 5: Hold QOR in STA: AOCV vs. LVF

Hold QOR in Primetime	**AOCV (GBA**	**AOCV (PBA)**	**LVF (GBA)**	**LVF (PBA)**
WNS(in ps)	-58.4142	-53.2	0	0
TNS(in ps)	-7506.5	-7076.76	0	0
NVP	510	371	0	0
ECO Cells		1016	0	0

CASE 2: AOCV VS. LVF AT ULTRA LOW VOLTAGE

Table 6: Detail of Case2

Case 2	**SIMD Engine**
Target Frequency	100MHz at SSG, Vnom-10% = 0.36V, -40C
Max Tran	300/600ps
Uncertainty for hold (in case of AOCV equivalent to 3 sigma)	309ps
Uncertainty for setup (in case of AOCV equivalent to 1 sigma)	338ps
Uncertainty for setup and hold (in case of LVF equivalent to 3 sigma)	3 - sigma (default)
PLL Jitter used for both AOCV and LVF	30ps

Table 7: AOCV vs. LVF

	1-sigma uncertainty/AOCV	3-sigma LVF
Sequential Area (squmm)	24474.828	24474.378
Combo Area (squmm)	61343.493	41168.799
Total Power FFG,0.44V, 125C (mW)	7.95	6.38
Leakage Power FFG,0.44V, 125C (mW)	5.79	4.65
Total Area	85818.321	65643.177
Leakage Reduction		19%
Area Reduction		23.5%

At Vnom = 0.4V, the design did not converge with AOCV and a global clock uncertainty of 3-sigma. Therefore, as a reference point, the design was closed at a clock uncertainty of 1-sigma. Still, hold closure was difficult and result in a large number of hold buffers and area increase.
The LVF based implementation was converging at smaller area and power at 3-sigma default settings.
As Table 7 shows, the LVF flow provides, at 3-sigma settings, an area reduction of 23.5% and a leakage reduction of 19.7%, compared to AOCV flow.

VI. CONCLUSIONS

GLOBALFOUNRDRIES 22FDX technology enables ultra-low-voltage design utilizing dynamic forward body bias to compensate delay variation. Libraries characterized with LVF and an implementation flow with statistical timing analysis exploiting accurate per cell, slew, load modelling of variation are a crucial prerequisite to close designs at low operating voltages. Industry characterization, P&R and STA tools support LVF implementation. The paper shows that LVF is enabled throughout the flow and ready for production design.

VII. REFERENCES

[1] R. Srinivasan et al., "Silicon Trimming using body-bias interpolation for GLOBALFOUNDRIES22FDX Technology", SNUG Silicon Valley 2017 K.

[2] H. Narisetty, R. Srinivasan, "22FDX IC Compiler II Body-Bias Interpolation and Correlation to PrimeTime", SNUG Silicon Valley 2018.

[3] GLOBALFOUNDRIES Design Methodology Team, "Liberty Variation Format (LVF) Usage in EDA Tools", Version 5.0, March 15th, 2018.