

# Using Statistical Static Timing Sign-off Methodology for Voltage Binning in 45 nm

Xiaoyue Wang (xxwang@ca.ibm.com), Eric A Foreman (eforeman@us.ibm.com), Peter A Habitz (habitz@us.ibm.com)

IBM Systems and Technology Group, Essex Junction, VT

## I. INTRODUCTION

As performance requirements continue to increase, technology challenges also rise. In the 45 nm technology node, leakage power has become more problematic. In order to address this issue, a selective voltage binning (SVB) technique has been developed. Chips can be binned at slow silicon process and operated at high voltage, and binned at fast silicon process and operated at lower voltage. Thus maximum power of a chip across full process is minimized while same chip performance is maintained in all bins.

This binning process can be modified for any number of bins, and the cut point of where voltage versus process changes can be unique. Figure 1 shows an SVB implementation flow, and Figure 2 shows a 16-bin SVB solution. In order to make this possible, every chip manufactured receives an Electronic Chip Identification (ECID) based on its performance screen ring oscillator (PSRO). By running a chip from the fast bin at lower voltage and running a chip from the slow bin at high voltage, the maximum power on the chip at the fast process is reduced - up to 17% as shown in an example in Figure 3, while the required performance of the chip at the slow process is maintained.

Static timing analysis coverage must match the binning plan. This can be achieved by performing many deterministic timing runs at multiple process-voltage-temperature (PVT) corners, which is burdensome due to an excess number of timing corners, and could be pessimistic. Our SVB timing methodology, with the use of statistical static timing analysis (SSTA), provides an ideal solution for that. In this paper we will discuss enablement of SVB with SSTA.

## II. SOURCES OF VARIATION

SSTA propagates delay distributions when performing timing analysis. The distributions of timing slacks can be expressed as a canonical model [1]:

$$S = a_0 + \sum_{i=1}^n a_i \Delta X_i + a_{n+1} \Delta R \quad (1)$$

where  $a_0$  is the mean of the distribution,  $a_i$  is the sensitivity and  $\Delta X_i$  is the deviation from nominal value for the “i”th source of variation, and  $a_{n+1}$  and  $\Delta R$  represent random

uncertainty.

For SVB timing, at least two sources of variation are required: silicon process and voltage. With these two sources of variation, timing closure can be performed only on the PVT points that the chip will operate at. Since SSTA is used, the extra number of timing runs that SVB requires can be merged into a single timing run.

Voltage is treated as a source of variation during SVB timing, but must use a small range to reduce error due to non-linearity. When voltage varies within a large range, at least two SSTA runs will need to be employed, which one is run at the highest voltage and another is run at the lowest voltage. The voltage variation modeled by a source of variation in SSTA will now be a small value around the voltage extremes.

## III. SVB TIMING METHODOLOGY

When generating timing reports, the canonical model in Eq. (1) can be projected to different corners. The worst slack canonical equation will be in the form:

$$Slack = a_0 - 3 \left[ \sum_{i=1}^m |a_i| \right] - 3 \left[ \sum_{j=m+1}^n (a_j)^2 + (a_{n+1})^2 \right]^{1/2} \quad (2)$$

where  $i$  is the index of sources of variation that are worst-cased by projecting to unique -3-sigma corner,  $j$  is the index of sources of variation that are statistically combined by the function of Root Sum of Squares (RSS), then projected to unique -3-sigma corner [2]. This equation can be used to report slacks only needed for timing closure at selected PVT corners:

$$Slack = a_0 + \left[ \sum_{i=1}^m P_i a_i \right] - 3 \left[ \sum_{j=m+1}^n (a_j)^2 + (a_{n+1})^2 \right]^{1/2} \quad (3)$$

where the selected sources of variation are projected to their unique  $P_i$ -sigma corner. For a 2-bin SVB solution, the report should include slacks only at low voltage / fast process and high voltage / slow process of two bins. Timing closure is now only needed on the process sub-space. Figure 4 shows an example of the process subspace. Two SSTA runs cover full space of Voltage / Silicon process which is a dash-dotted

rectangle. The subspace needed for SVB timing closure is the highlighted region, which has 8 VDD / process corners. Considering other parameters such as temperature, the total number of PVT corners will be doubled at least.

This 2-bin SVB timing methodology can be extended straight forward to N bins in which the number of PVT corners increase significantly and timing closure by many deterministic timing runs becomes too difficult to be completed. On other hand, trying to bound all PVT corners (dash-dotted rectangle in Figure 4) either through non-SVB SSTA or deterministic timing runs, is too pessimistic and not necessary, especially when the chip timing is dominated by voltage. Figure 5 shows a timing slack comparison between SVB timing and non-SVB timing from a dirty part whose timing has not been closed yet. Every tick mark represents a unique timing slack. The diagonal lines are used for comparison. We can see that SVB timing reduces slack up to 900ps, and so design turn-around time (TAT) is significantly improved.

#### IV. CONCLUSION

In order to address power problems in 45 nm, designs can be binned and operated based upon optimizing performance and voltage. Designs that are binned can reduce leakage power by reducing the timing closure subspace.

This paper focused on enabling an SVB SSTA sign off methodology for 45 nm. The key idea of SVB timing methodology is to do timing analysis for the full process space in all dimensions but only close the chip timing anywhere in the subset of the full process space. We showed comparative timing data to validate our methodology. From 45 nm onwards, more and more chips will be in the SVB paradigm, therefore this paper points out a practical solution to a very important problem.

#### REFERENCES

- [1] Visweswariah, C., et al, "First-order incremental block-based statistical timing analysis," IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems, vol. 25, no. 10, pp. 2170–2180, 10/2006.
- [2] Foreman, E.A., Marshall, L.B., "Timing Closure in 65-nanometer ASICs Using Statistical Static Timing Analysis Design Methodology", Design Automation Conference (DAC), San Francisco, CA, 07/2009.

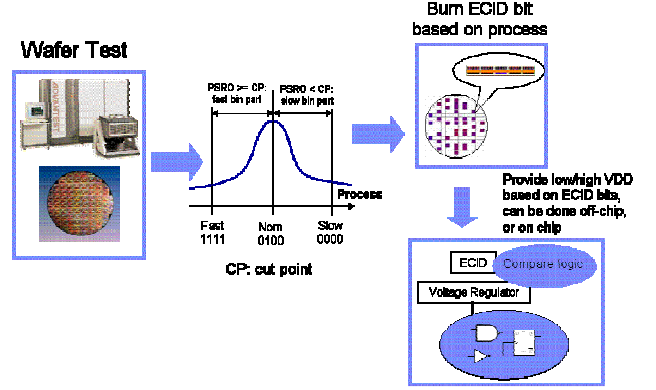


Figure 1: SVB implementation flow

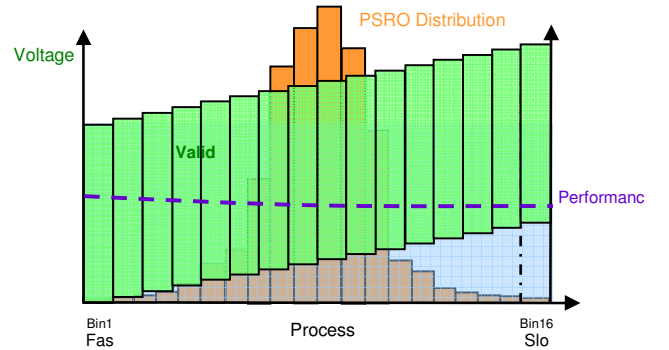


Figure 2: A 16 bin SVB solution. Chip is binned based on PSRO measurement and operated at bin-specific voltage. Chip performance is the same or much close in each bin.

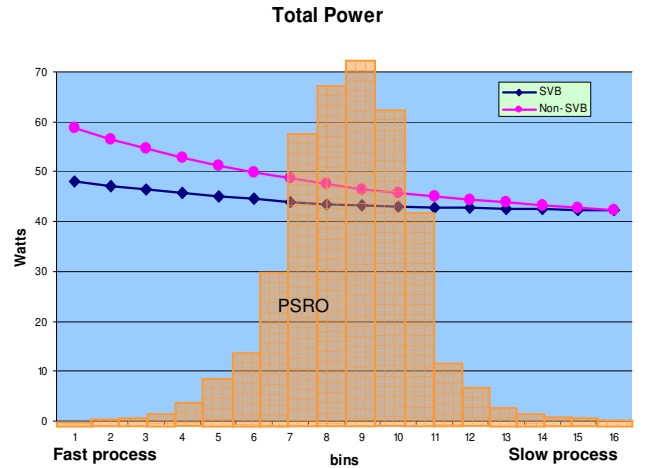


Figure 3: An example of chip power comparison between 16 bin SVB and non\_SVB. Up to 17% total power is saved at fast end of process by SVB.

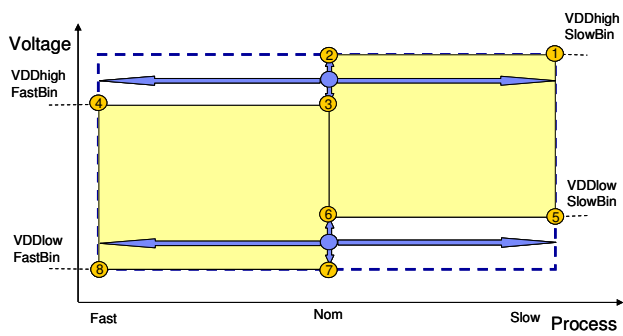


Figure 4: Two SSTA cover full space of Voltage/Silicon process variation. SVB timing is represented by the highlighted subspace.

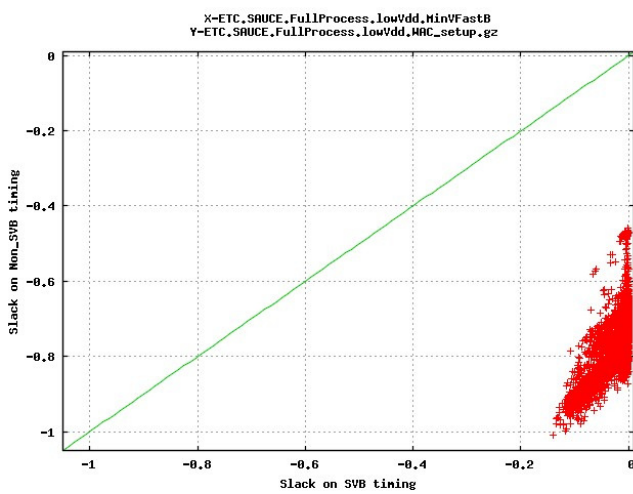


Figure 5: Timing slack comparison between SVB timing and Non-SVB timing. The x-axis represents slacks from a timing run with SVB. The y-axis represents slacks from a timing run without SVB. The green diagonal line represents same value between two runs. Look how much closer to zero it is with SVB!