

# How Low Can You Go? Pushing the Limits of Transistors

## Deep Low Voltage Enablement of Embedded Memories and Logic Libraries to Achieve Extreme Low Power

### Authors

**Josefina Hobbs**

Sr. Product Marketing  
Manager, Logic Libraries,  
Synopsys

**Anuj Pant**

Sr. Technical Marketing  
Manager, Embedded  
Memories, Synopsys

### Introduction

Rising demand for cutting-edge mobile, IoT, and wearable devices, along with high compute demands for AI and 5G/6G communications, has driven the need for lower power systems-on-chip (SoCs). This is not only a concern for a device's power consumption when active (dynamic power), but also when the device is not active (leakage power). This highly competitive industry provides significant rewards for being the first to achieve best-in-class power efficiency in these markets. And of course, all of this must be achieved without impacting performance or area. Power, performance, and area (PPA) are the critical metrics for today's advanced semiconductor SoCs.

Synopsys Foundation IP Memory Compilers and Logic Libraries enable SoC designers to achieve the best possible PPA, getting the maximum possible performance out of their designs while enabling them at the lowest possible operating voltages (near threshold values of transistors), thus significantly reducing overall power consumption. The result is longer battery life and higher Performance Per Watt.

In this paper we will discuss:

- Deep low voltage requirements (0.4v typical and below) for mobile, IoT, high performance compute (HPC), automotive, and crypto applications
- Various techniques adopted by SoC designers to trade-off PPA, including improvements on existing assist techniques for memory compilers
- Architectural and characterization enhancements to support lower voltages for logic libraries
- How Synopsys Memory Compilers and Logic Libraries have been enhanced to support deep low voltages to save power, while still achieving optimal performance and area and maintaining high reliability

### Applications Requiring Deep Low Voltages

**Mobile:** For mobile application designs, dynamic power reduction is essential for long battery life. Mobile applications typically don't require the entire SoC to operate at maximum performance the entire time. Some parts of the mobile SoC do not necessarily need the maximum speed capabilities at all times and can therefore be "down-shifted" to a lower clock speed, which saves dynamic power. Those lower clock speeds can generally run at lower operating voltages and reducing the voltage will save even more dynamic power. Some parts are not at all active and can therefore be "turned off" when not needed. Designers employ various advanced low power techniques such as dynamic voltage and frequency scaling (DVFS), as well as power shutdown, to reduce dynamic and leakage power, respectively.

**Bluetooth and IoT** devices are even more reliant on battery life, where they are often expected to last weeks or even months between charges. The longer the battery lasts, the better the device is deemed. If one looks at the very basic operation of AI devices (smart watches) and other voice-controlled devices, most of the processing is done locally. The processors that do these calculations must necessarily be extremely low power.

**HPC** applications are generally found in a server farm, where hundreds of thousands of servers are operating simultaneously. Battery life is not an issue since the servers are plugged into power; however, heat dissipation is a major concern and cooling the servers down is a major expense. Another significant problem is overall power consumption since these hundreds of servers can consume hundreds of Mega Watts. The SoC designer can reduce heating and power consumption concerns by applying low voltage techniques to their design blocks, thereby reducing power. Here again, DVFS techniques allow servers to scale both operational voltage and frequency when full computational effort is not required.

**Automotive:** Many automotive applications such as digital cockpit, ADAS, parking sensor, and even the car itself if it's electric, run on battery-powered devices. Furthermore, with new high-resolution driver displays, infotainment systems, and cutting-edge ADAS features, they all must operate at higher performance than before. However, for electric cars these features need to consume less power to get more miles out of a single battery charge. After all, everyone is well-aware of the general consumer anxiety about the range of electric vehicles. One other very important concern in Automotive is reliability, and power consumption is a direct factor in reliability.

**Crypto:** The power demand for Crypto SoCs is quite high, due to massively parallel operations across all blocks with 100% activity 24/7/365. However, excessive high-power demands will reduce the profit margin for Currency Mining, potentially eliminating financial viability of a project. If the price of mining is too high, users will not select that machine for mining. Therefore, it is essential to keep power consumption under control to reduce the electricity bill.

## Embedded Memories - Techniques Adopted by SoC designers to Trade-off PPA

First, let's look at memory compilers. There are various techniques employed by SoC designers for the embedded memories to be able to operate at lower voltages to save power. The **assist techniques** are the most common. **Splitting the supply voltages** for the bitcell and periphery is another technique, where the array operates at a higher voltage and the periphery operates at a lower voltage. However, depending on how assists and other power management techniques are implemented in the SoC, the PPA and reliability can be impacted. Synopsys Memory Compilers have been enhanced to achieve optimal PPA while still maintaining the system's high reliability.

Careful co-optimization between technology and the design of memory assist circuits is required to deliver dense, low-power memory operation at low voltages. There are various read and write assist techniques that can ensure readability and writability in the bitcell at lower voltages. Negative Bit Line (NBL), Dual-supply, Word Line (WL) lowering, write column voltage lowering, and Read-Modify-Write offer large stability improvement at a reasonable area overhead.

### Enhanced Write Assist Techniques to Further Save Power and Area

When implementing low voltage design, it is generally necessary to employ a write assist scheme for the memory. Synopsys Memory Compilers use enhanced write assist schemes that are area efficient and provide higher performance and increased reliability along with the capability to operate at a lower Vmin.

Synopsys continuously explores and enhances the embedded memories IP portfolio with different design techniques for our write assist circuitry. For smaller geometry nodes, Synopsys developed a new write assist scheme with minimal area and power penalty as shown in the tables below.

Memory Size	Conventional Assist Scheme Design	New Assist Scheme Design	Difference	% Difference
Small	2,119	1,434	-686	-32%
Wide	19,241	13,018	-6,223	-32%
Big	315,974	283,223	-32,752	-10%
Tall	34,808	31,200	-3,608	-10%

Table 1: Area (um<sup>2</sup>) savings

Memory Size	Conventional Design	New Assist Scheme Design	Difference	% Difference
Small	1.079	0.876	-0.203	-19%
Wide	8.135	6.021	-2.114	-26%
Big	19.941	12.541	-7.4	-37%
Tall	3.328	2.268	-1.059	-32%

Table 2: Dynamic power savings during write operation

Table 1 and Table 2 show the block area and power improvements enabled by Synopsys Embedded Memories versus the results achieved with embedded memories that use conventional write assist schemes. Also, this write assist scheme leverages metal capacitance over device capacitance. This allows SoC designers to save area and improve the reliability of their SoCs.

### Ensuring Performance and Enhanced Reliability at Lower Voltages

Low voltage operation in SRAM architecture faces several challenges due to process variation, bit cell stability, sensing, and much more. Recall the standard butterfly curve showing the voltage transfer characteristics (VTCs) of two inverters. As the voltage is reduced, the SRAM cell starts showing degradation of the read and write butterfly curves. This degradation can cause multiple issues: reads are upset, bitcell does not flip, SER is pronounced, sensing fails, control signals deviate, and the BL signal weakens. Therefore, assist techniques are needed to support the lower extreme low voltages ( $V_{min}$ ) required by cutting-edge low power applications.

Synopsys Foundation IP Team continuously explores the enhancement of Synopsys Embedded Memories with different techniques for assist schemes to further improve the SoC PPA and reliability at lower voltages. Figure 1 shows the gain by 4X in the read current world line under drive (WLUD) going from higher to lower voltages, thereby improving the access time ( $T_{cq}$ ) using the Synopsys proposed WL lowering assist technique compared to conventional WL lowering.

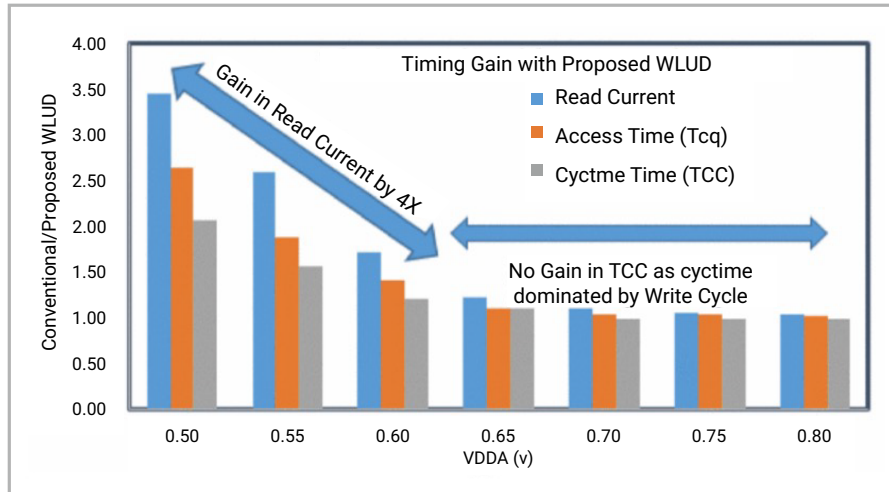


Figure 1: Timing gain with read assist WL lowering

Figure 2 shows the magnitude of the NBL voltage with conventional and Synopsys proposed assist schemes. These results were calculated at SF (Slow Fast) silicon, highest possible voltage,  $-40^{\circ}\text{C}$ , and for maximum load at the bit line. The results show that there is an average 67% improvement for the Synopsys proposed NBL assist technique across 111 6T, 122 6T, 133 6T, and 122 8T bit cells. This improvement in the magnitude of the dip at the bit line leads to significant reliability improvement.

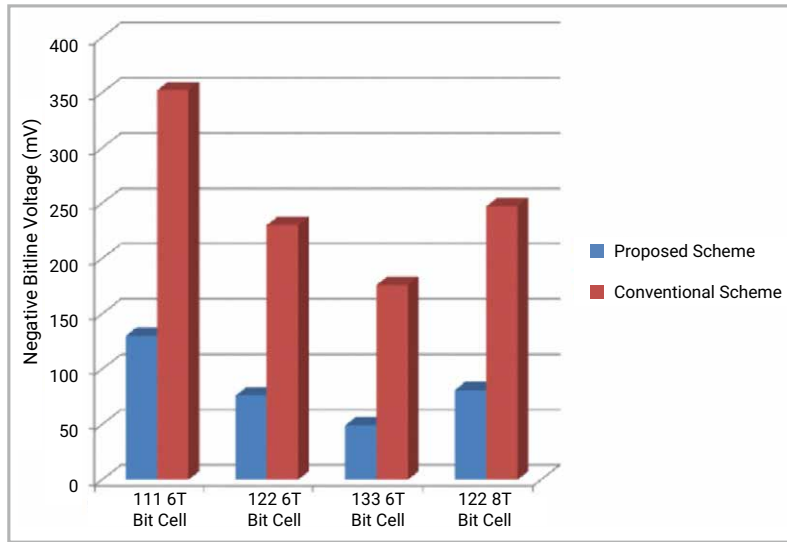


Figure 2: NBL comparison

To calculate the SoC design reliability improvement, Synopsys implemented aging simulations using the foundry MOSRA flow, and the age of the memory bit cell was calculated. The age of the bit cell is the time in which the write time of the bit cell increases by 10% of its typical value. Figure 3 shows the write failure in a 122 bitcell due to aging effects with the conventional scheme. It represents the behavior of internal nodes of 122 6T bit cell during write operation with the conventional write assist scheme and shows that the cell stops flipping (write failure) for an aged cell.

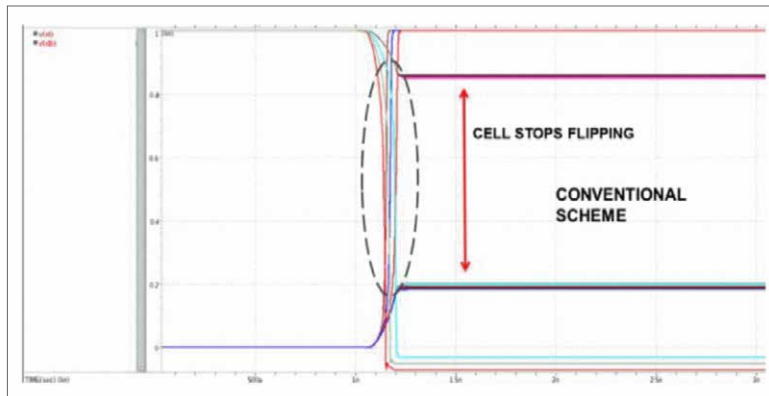


Figure 3. Write failure

Figure 4 shows the improvement in reliability with the Synopsys proposed scheme. It represents the behaviour of internal nodes of 122 6T bit cell during write operation with the proposed write assist scheme, where there is no write failure. This scheme shows better writability into the bit cell, which improves reliability.

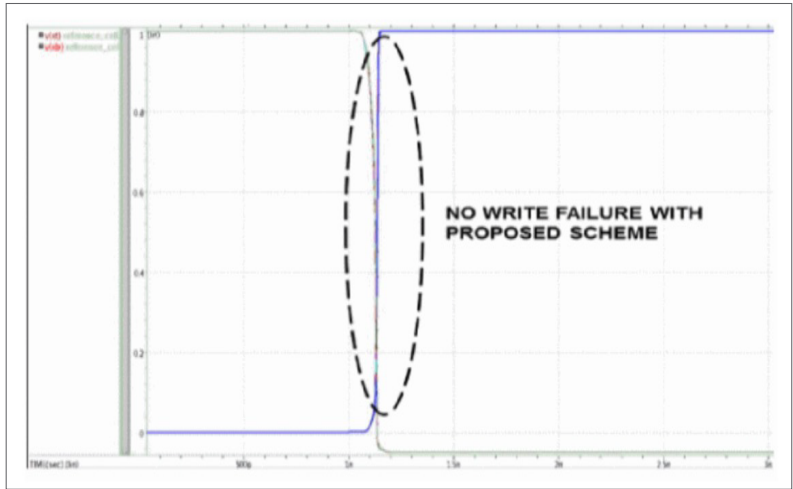


Figure 4. No write failure

Figure 5 compares the reliability simulation (0–10 years) results. It shows that there is a significant improvement in the age of the bit cell.

Bit Cell Type	Age with Conventional Scheme (in years)	Age with Proposed Scheme (in years)
111 6T Cell	1.3	6.2
122 6T Cell	2.7	9
122 8T Cell	2.1	8.2
133 6T Cell	4.1	14.3

Figure 5. Improvement in age of bitcell

To summarize, all the above results show that the Synopsys Embedded Memories with the proposed write and read assist schemes provide a tremendous improvement in both reliability and performance of the SoC. These solutions are essential in addressing the critical challenge of applications that require extreme low voltage operation with the least penalty on device reliability and without impacting transistor performance.

## Logic Libraries - Enabling Deep Low Voltage Operation (0.4v and below)

As mentioned earlier, there are different market segments requiring very low operational voltages of 0.4v and below to reduce dynamic power on their SoCs. Since dynamic power is proportional to the square of the voltage ( $V^2$ ) and frequency ( $f$ ) of an SoC, a designer can realize dynamic power reduction either by reducing the frequency or reducing the voltage. With smaller geometries, it is possible to achieve much better performance even after reducing the voltage.

For Crypto market SoCs, standard cells dominate the design. The power demand for Crypto SoCs is quite high. However, as explained earlier, it is essential to keep power consumption under control to reduce the operational cost. With no need of SRAMs, the design voltage can drop to near threshold, meaning the threshold voltage at which a transistor switch.

Synopsys Logic Libraries support a variety of applications requiring deep low voltage supplies at smaller geometries, with special consideration for two key aspects of deep low voltage development: standard cell architectural optimization and standard cell characterization optimization.

## Architectural Optimization

Standard cell architectural techniques can be employed to reduce both dynamic and leakage power. Synopsys uses stack-based versus stage-based architectural techniques for the optimal topology for deep low voltage operation. These techniques may require adjusting the standard cell area to get better yield. Synopsys also avoids long poly routing to reduce resistance, which will help to improve delay since poly has higher resistance compared to metal. Hspice Monte Carlo simulations are run on every cell in the standard cell library to ensure proper functionality. If a standard cell does not pass the required sigma checks, it will be removed from the library to avoid any potential yield issues.

## Characterization Optimization

Accurate characterization is an essential part of a robust design. One important piece of characterization is modeling process variation across an SoC, referred to as on chip variation (OCV). OCV is usually modeled as a flat derate for 90nm and above. However, using a flat derate can introduce some pessimism on critical timing paths. Starting from 65nm, path depth and distance-based derate was used instead of flat derate, and this was named advanced on chip variation (AOCV). In 28nm, parametric OCV (POCV) was introduced to model single stage delay variability. This results in faster timing closure compared to AOCV because it is less pessimistic.

POCV cell variation information can be represented by either a single cell-based variation coefficient (side file) or the Liberty Variation Format (LVF). The LVF models delay variation for each timing arc at each slew and load combination in the library. Figure 6 summarizes how on-chip variation (OCV) margining approaches have evolved and how Liberty Variation Format (LVF) and moment-based LVF have become mandatory at 7nm and below.

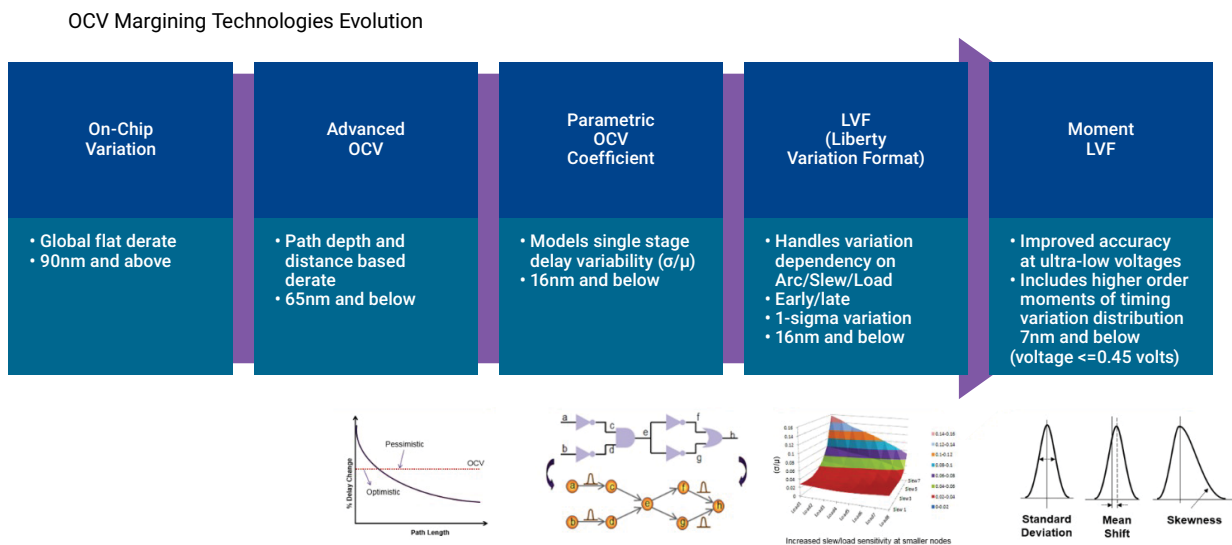


Figure 6: OCM margining technologies evolution

The Synopsys Library Characterization Team uses a machine learning (ML) based Liberty Variation flow. The ML-based LVF flow improves variation accuracy for timing Liberty files. At deep low voltages, the statistical distribution is usually asymmetrical and non-linear, and variation is significant. It is important to model this asymmetrical non-linear distribution behavior as close to linear behavior as possible to get accurate Spice-to-silicon precision. Synopsys follows foundry specific guidelines to accurately model the variation data in the liberty and employs moment-based LVF to capture the more detailed timing variation distribution. Moment-based LVF extension includes mean-shift, standard deviation, and skewness, as shown in Figure 7. To increase accuracy, the ML-based strategy is used to accumulate sufficient data at regions of interest during the characterization of each standard cell.

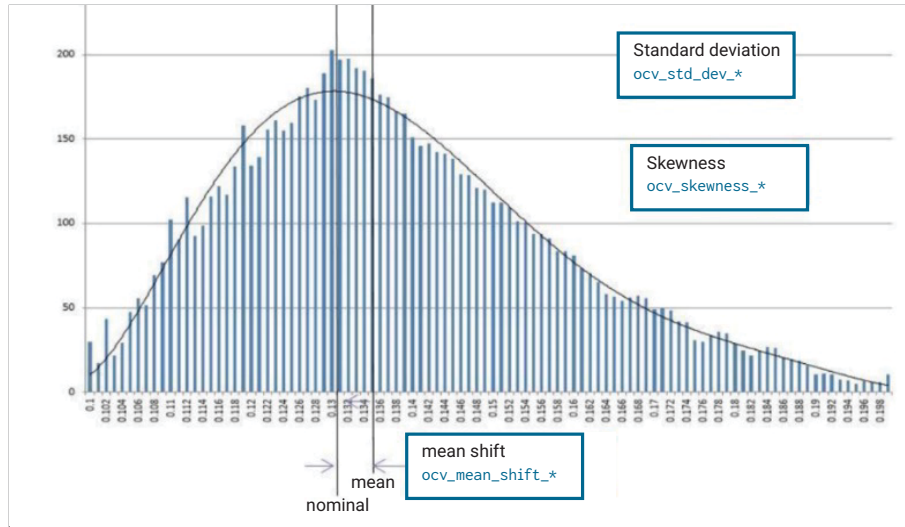


Figure 7. Moment-based LVF

Device parameters are varied during simulation to accurately model setup, hold, delay, and timing parameters. The max slew is bigger with lower voltages, so it takes longer for the Hspice simulation to perform the binary search for setup/hold characterization, which can increase characterization runtime up to 3x for deep low voltage corners. Synopsys level shifters are also designed and recharacterized to support the wider voltage range.

In addition to having a rich and robust standard offering, the Synopsys Logic Library Team is also equipped and experienced to work with customers on specialized custom low voltage operation cells, where optimizations can be made at even the transistor level for this type of operation.

## Putting It All Together at the SoC Level

Now let's look at how to leverage this technology in an SoC. With the availability of Synopsys Logic Libraries and Embedded Memories specifically optimized to perform at low voltages, designers can be confident they will be able to achieve their SoC design goals. Various design level techniques can be employed to achieve deep low voltage operation that pushes the limits of transistors. Some of the techniques are:

- **Voltage reduction**

Crypto customers can use techniques such as voltage reduction to optimize the mining algorithm speed for the lowest voltage. An even more extreme power reduction technique is to remove "non-required" design elements. For example, at the design level, use single level sequential cells and remove scan chains. At the system level, reduce the traffic of external control signals to the lowest frequencies possible or even remove the Power Management System altogether.

- **DVFS**

DVFS techniques allow servers to run at various voltages to support different frequencies. For example, faster processing servers performing compute-heavy operations can run at 0.7v and higher, while the servers that are only doing slow software computations can run at 0.4v and below to reduce power. As there are multiple servers being used in an HPC design, saving just 5% to 10% of power in one server will have a significant impact when applying it to hundreds of servers.

- **Shut down**

Synopsys Memory Compilers offer multiple levels of power management features, including light sleep, deep sleep, shut down and POFF (Periphery OFF) modes to enable array biasing with partial periphery shut down, full periphery shut down with data retention, complete shut down without data retention, and a mode where the periphery supply voltage of the memory can be totally removed externally. An option for dual rail is also available to support DVFS, allowing the periphery to go down to even lower voltages. This provides SoC designers flexibility for how each memory should be controlled within the different power domains on their SoC.

## Summary

Demand for SoCs with extremely low power consumption will always remain high in the semiconductor industry, be it for IoT, mobile, HPC, automotive, crypto, medical devices, or AI. As the technology scales down further and system complexity increases, SoC designers rely on memories and logic libraries that consume very little power. This means more demand for refined memory assist techniques and standard cell architectural and characterization advancements that can support operation at very low voltages while maintaining PPA benefits and increased reliability. [Synopsys silicon proven Embedded Memory Compilers and Logic Libraries](#) offer such capabilities, enabling deep low voltage operation required for these applications.

## References

1. A Temperature Compensated Read Assist for Low Vmin and High Performance High Density 6T SRAM in FinFET Technology - <https://ieeexplore.ieee.org/document/8326969/>
2. Write Assist Scheme to Enhance SRAM Cell Reliability Using Voltage Sensing Technique - <https://ieeexplore.ieee.org/document/7434972>
3. A step towards accurate timing analysis <https://www.edn.com/parametric-on-chip-variation-a-step-towards-accurate-timing-analysis/>
4. Model Variation And Its Impact On Cell Characterization <https://semiengineering.com/model-variation-and-its-impact-on-cell-characterization/>