SYNOPSYS®

WHITE PAPER

PPA(V): Performance-Per-Watt Optimization with Variable Operating Voltage

Author

James Chuang
Sr. Staff Product Marketing Manager, Synopsys

Abstract

Performance-per-watt has emerged as one of the highest priorities in design quality, leading to a shift in technology focus and design power optimization methodologies. Variable operating voltage possess high potential in optimizing performance-per-watt results but requires a signoff accurate and efficient methodology to explore. Synopsys Fusion Design Platform™, uniquely built on a singular RTL-to-GDSII data model, delivers a full-flow voltage optimization and closure methodology to achieve the best performance-per-watt results for the most demanding semiconductor segments.

High-Performance Computing (HPC) Continues to Push Advanced Node PPA Envelope

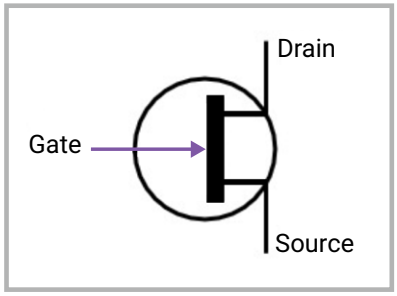
High-Performance Computing (HPC) is one of the fastest-growing design segments in the semiconductor industry, powering a wide span of applications, including cloud data centers, artificial intelligence, mobile computing, autonomous vehicles, and more. While the segment continues to race towards the highest design performance, power consumption can limit design performance in several application areas. For example, energy and cooling costs can directly impact data centers' profitability, and mobile phones must carefully balance performance with battery life.

As a result, performance-per-watt has emerged as one of the highest priorities in HPC design quality, in addition to the established performance, power, and area (PPA) criteria, leading to a shift in technology focus and design power optimization methodologies.

On the other hand, to achieve the best possible performance-per-watt goals, HPC designs are fabricated on state-of-the-art FinFET processes. While the innovative "fins" possess superior control over electron flows, enabling faster switching and lower leakage currents requires more power to switch than an equivalent planer structure. Furthermore, the compute-intensive workloads intended for HPC designs also lead to near-constant switching, resulting in a power profile heavily dominated by dynamic power. The component represents the power consumed when transistors are switching between states.

Emerging Opportunities in Dynamic Power Optimization

The power to complete a switch is consumed by the transistor's gate capacitor (FET). According to the power consumption equation shown in Figure 1, under the same frequency, the power consumed is proportional linearly to the gate capacitance but proportional to the square of the operating voltage.



$$\text{Dynamic Power} = \text{Capacitance}_{\text{Gate}} * \text{Voltage}^2 * \text{frequency}$$

Figure 1: FET Transistor Dynamic Power Equation

Established dynamic power optimization techniques target the reduction of transistor sizes, which directly reduces the gate capacitance. However, lowering the operating voltage possess an even higher potential in lowering dynamic power. As shown in Figure 2, a case study on a 7nm FinFET design shows a mere 5% operating voltage reduction can lead to a 9% dynamic power reduction.

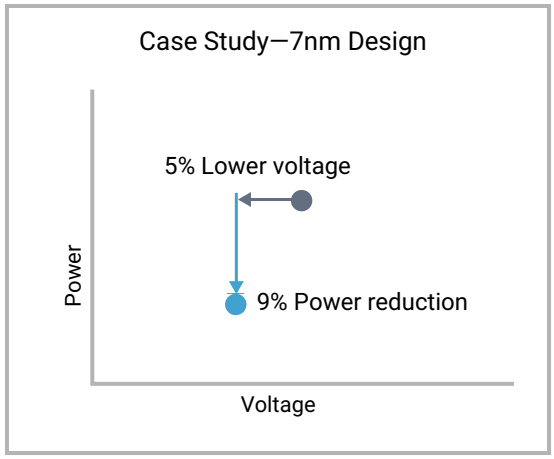


Figure 2: Operating Voltage vs. Dynamic Power Case Study

While a reduction in operating voltage is significantly effective in dynamic power optimization, it has not been a broadly deployed technique in the semiconductor design industry. Next, we will review the challenges that have led to its absence.

Past: Voltage Decisions Are Remote from Design Optimization

In previous semiconductor design environments, operating voltage (Vdd) defines a priori, independent of production design contexts. The process involves transistor device-level analysis from the foundry, accompanied by in-house simulation on a small subset of cells, to identify a reasonable minimum set of operating voltages. The pre-determined operating voltages then drive technology library characterization, design optimization and signoff closure for all designs.

As illustrated in Figure 3, designers optimizing for performance-per-watt will be exploring the solution space within the pre-determined voltages by performing multiple runs with different performance targets.

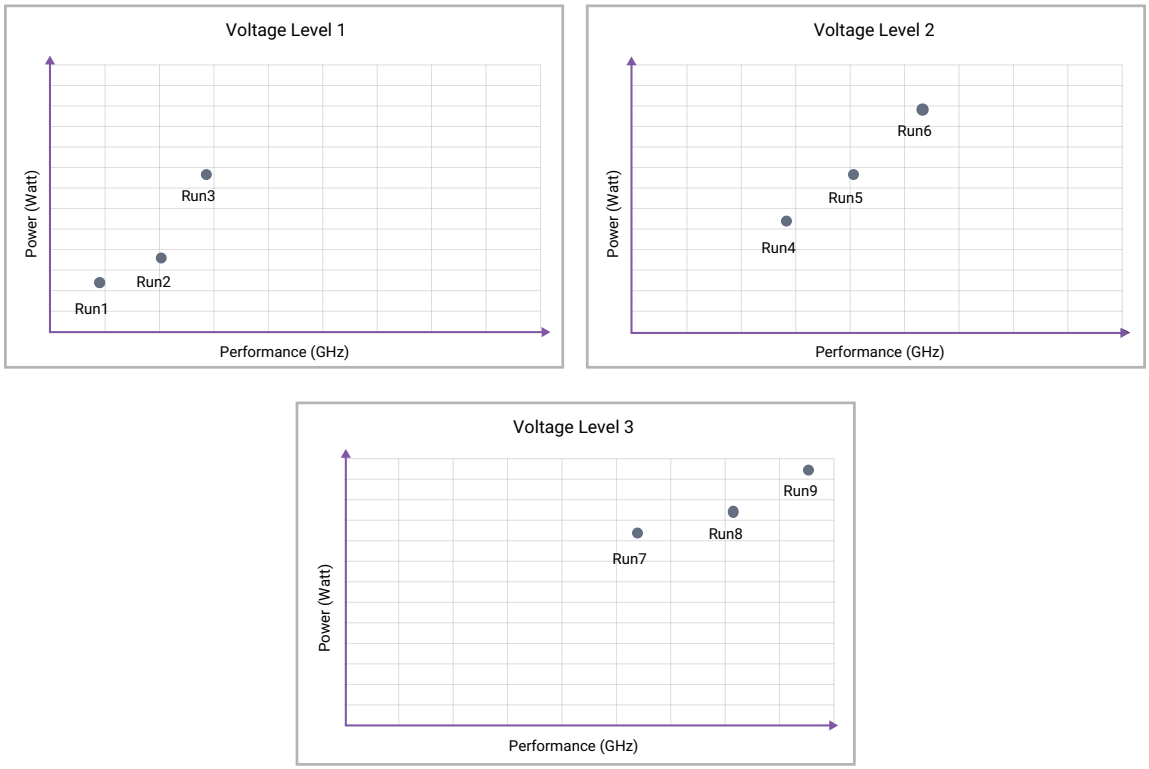


Figure 3: Performance-per-Watt Optimization with Discrete Voltage Levels

Suppose a project met all PPA targets at the pre-defined voltages. In that case, there is no viable path to explore further optimization at a lower operating voltage due to the lack of technology libraries.

As a result, it has become a widespread practice during post-silicon testing to explore lower operating voltages on a silicon testbench, also known as Vmin analysis, and find the lowest operating voltage that a design can continue to operate properly. The result will be used to influence the decision if new libraries are characterized. This feedback loop can take months or multiple design cycles to make a meaningful impact to improve performance-per-watt.

Enabling Free Voltage Exploration in the Design Flow

Non-linearity between voltage levels and timing responses has limited the deployment of linear voltage interpolation to only between two closely spaced libraries at higher, standard voltage levels. In 2017, PrimeTime® timing signoff solution established a foundry-certified advanced voltage scaling technology that enabled accurate analysis at any voltage level within a broad range. As illustrated in Figure 4, signoff-accurate voltage scaling results can be achieved between wide spacing or at lower voltage levels.

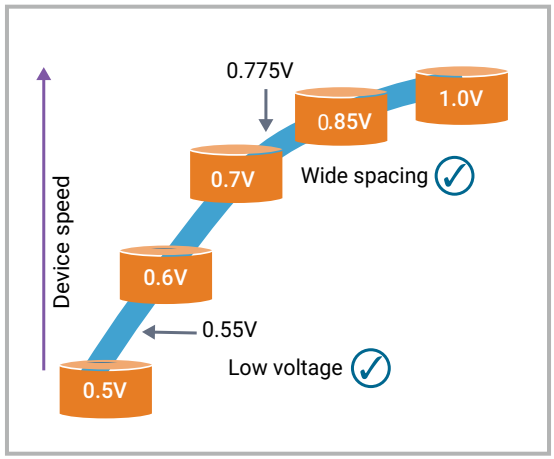


Figure 4: PrimeTime Advanced Voltage Scaling

Designers now had a way to "sweep" the voltage range, trial run the same design at unlimited voltage levels, and eventually, find a voltage sweet-spot for the desired PPA or performance-per-watt targets. As illustrated by Figure 5, while the PrimeTime timing signoff solution proved to be both accurate and effective, the manual sweeping process can be highly time- and resource-consuming. With exploration, runs growing linearly with voltage level candidates.

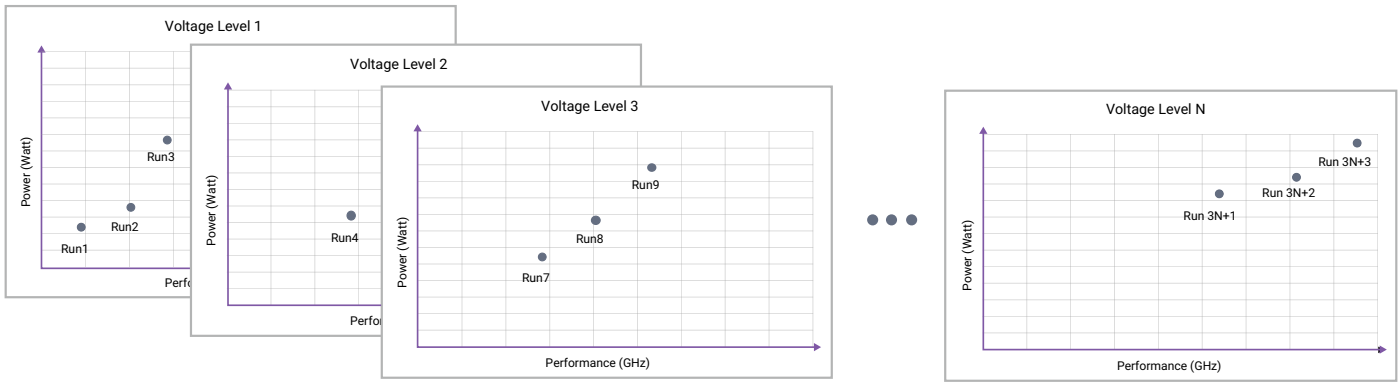


Figure 5: Performance-per-Watt Optimization with Sweeping Voltage Levels

PPA(V) Optimization: Introducing Voltage as a Variable During Design Optimization

Fusion Compiler™ RTL-to-GDSII solution and IC Compiler II™ hyper-convergent place and route solution are the industry's only digital design implementation solutions that deploy Synopsys' most-trusted golden signoff solutions during implementation and PPA optimization. The unique Signoff Fusion technology seamlessly enables the PrimeTime timing signoff solution, PrimePower RTL-to-signoff power analysis, and StarRC™ parasitic extraction signoff analysis engines within the implementation environment for accurate timing, power, and interconnect RC guidance, including PrimeTime timing signoff solution's advanced voltage scaling technology.

Introducing voltage level as a variable during design optimization enables Fusion Compiler RTL-to-GDSII solution and IC Compiler II place and route solution to expand the performance-per-watt solution space. By scaling the operating voltage while pushing the metrics for higher performance, lower power, and smaller area, the optimization engine can natively explore the operational voltage level within a single optimization run without costly external iterations previously required for voltage sweeping.

In a design flow with a fixed frequency target, the variable operating voltage enables further exploration opportunities that can lower total power by directly reducing dynamic power with a lower operating voltage while minimizing leakage and area impact. As shown in Figure 6 below, a 5nm HPC design was able to reduce total power by 26% using a 14% lower operating voltage while meeting the same frequency target:

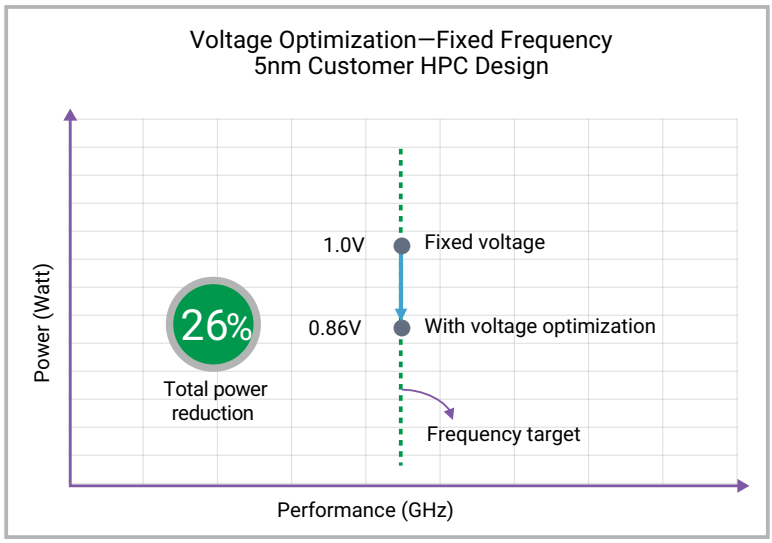


Figure 6: Improved Total Power with Lower Operating Voltage

In a design flow with a fixed total power target, the variable operating voltage can eliminate the frequency and voltage sweeping previously required to find the highest possible frequency. The optimization engine can natively explore the voltage range and find the best possible frequency within a single run, which may not have been achieved otherwise due to time and resource constraints, as illustrated in figure 7:

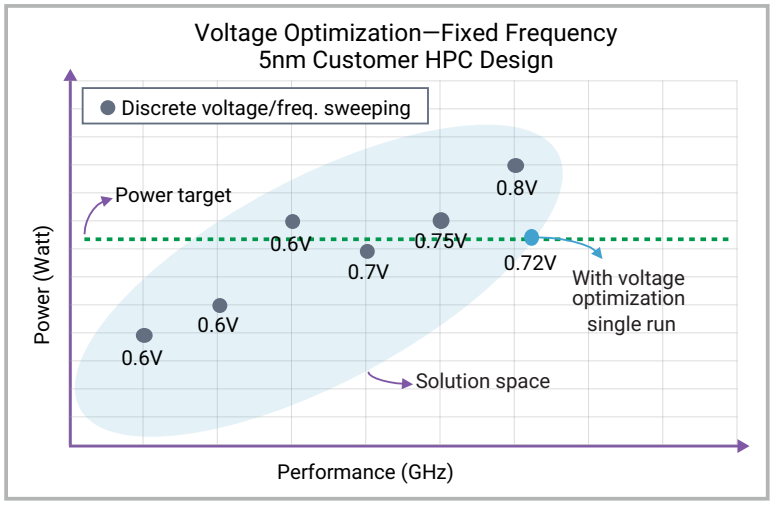


Figure 7: Improved Frequency with Fixed Power Target

PPA(V) Signoff: Introducing Voltage Robustness in Design Analysis and Signoff

Voltage optimization unlocks PPA optimization opportunities, lifts the PPA curve, and improves SoC design performance-per-watt. Its innovative native voltage sweeping extends optimization solution space exploration to achieve the best possible performance-per-watt at lower operating voltages.

While lowering the operating voltage significantly improves performance-per-watt, it also removes the excessive margins unintentionally created by fixed-voltage design flows. Moreover, the higher cell and power density at advanced nodes also requires more sophisticated operating voltage-drop analysis and margining methodologies to prevent voltage drop-related design failures.

PrimeShield™ design robustness solution has expanded on PrimeTime timing signoff solution's core technology and introduced a native voltage analysis feature to address this challenge. In contrast to static timing analysis, which reports critical timing paths according to timing slack, this new analysis reports critical paths according to a new metric: voltage slack. As illustrated in Figure 8, this new metric represents the minimum voltage drop per-cell or per-path for a path to still meet timing requirements.

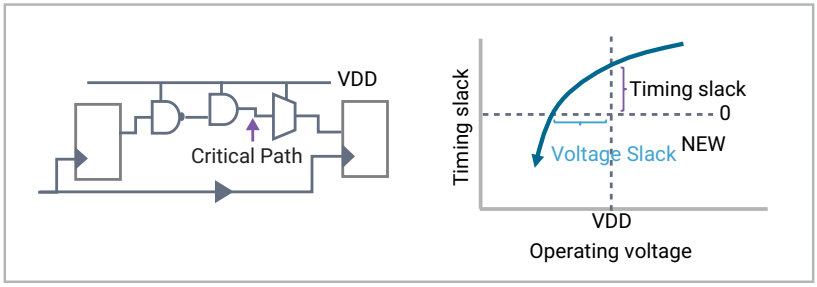


Figure 8: Voltage Slack Reporting for Critical Paths

Timing critical paths typically employ the strongest driving cells to ensure path delays will meet performance requirements. These strong driving cells are usually less sensitive to voltage variation, as illustrated in Figure 9. Simultaneously, less-critical timing paths may employ weaker driving cells that are more sensitive to voltage and fail earlier when the design experiences a drop in operating voltage. These risk-inducing paths are not straightforward to find using static timing analysis methods and usually require extensive voltage sweeping to uncover.

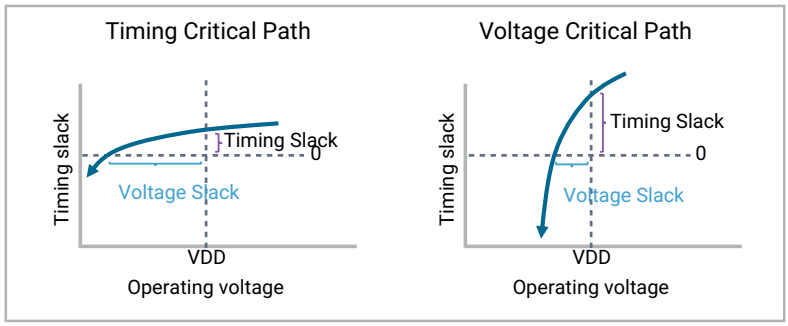
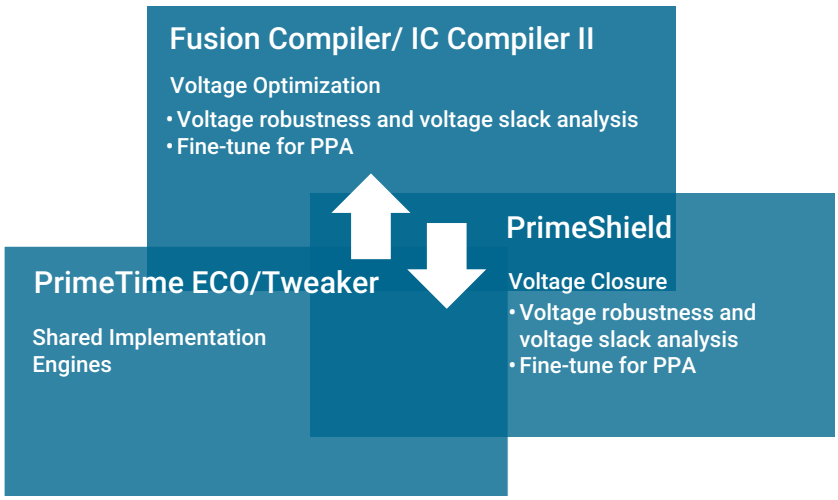


Figure 9: Voltage Slack Analysis on Timing vs. Voltage Critical Path

PrimeShield design robustness solution's voltage slack analysis provides a direct and efficient path to analyze and report such information for further optimization. To facilitate an effective optimization methodology, in addition to voltage slack reporting, PrimeShield design robustness solution's voltage robustness analysis performs bottleneck analysis on voltage critical paths to identify the cells that are most sensitive to voltage variation and impose the highest risk of timing failure.

Synopsys' most popular ECO signoff solutions, PrimeTime ECO design closure signoff and Tweaker ECO signoff, can provide ECO guidance to improve the above metrics. By swapping voltage-sensitive cells with less sensitive counterparts, the ECO changes can improve design robustness against voltage drop, or further fine-tune operating voltage across all signoff timing scenarios.

This technology can also enable methodologies to achieve uniformity in voltage slack and improved voltage margining. By ensuring voltage margining where needed and eliminating any outstanding risk points across the design, designers can avoid applying excessive margining globally while benefiting from the power benefits that lower operating voltages can deliver.



Conclusion

As the semiconductor industry, especially the HPC design segment, continues to push for better performance-per-watt, Fusion Compiler RTL-to-GDSII solution's and IC Compiler II place and route solution's voltage optimization capability, based on Synopsys' golden signoff analysis solutions, provides a differentiated path to efficiently improve advanced node designs' performance-per-watt through introducing operating voltage as a variable during optimization.

PrimeShield design robustness solution's voltage slack analysis is based on the same core foundation. The new analysis metrics enable designers to efficiently pinpoint voltage robustness bottlenecks, drive voltage margining efficiency and uncover opportunities to directly fine-tune operating voltages.

Synopsys Fusion Design Platform, uniquely built on a singular RTL-to-GDSII data model, delivers a full-flow voltage optimization and closure methodology to achieve the best performance-per-watt results for the most demanding semiconductor segments.

09/02/21.CS748550008-PPA(V) wp.

Pub: September 2021