

# Architectural Clock Gating and Way-Based Shutdown Techniques for System-Level Cache

**Karthik Wali**

ASIC Design Engineer  
Email: [ikarthikw@gmail.com](mailto:ikarthikw@gmail.com)

## Abstract:

System-level caches are a vital component of modern microprocessor architectures, providing a buffer between the fast central processing unit (CPU) and slower main memory. These caches reduce memory access latency and improve overall system performance. However, their increasing size and complexity have led to substantial power consumption, posing challenges for energy efficiency—particularly in mobile devices, data centers, and embedded systems. Power dissipation in caches can be broadly classified into dynamic power, caused by switching activities, and static or leakage power, which results from current leakage in transistors even when they are idle.

Architectural clock gating (ACG) and way-based shutdown (WBS) are two advanced low-power techniques aimed at mitigating this issue. ACG addresses dynamic power by selectively disabling the clock signal to parts of the cache that are not actively engaged in operations. This prevents unnecessary circuit activity, conserving energy. In contrast, WBS targets leakage power by turning off unused cache ways, thereby reducing the standby power consumed by idle sections of the cache. Together, these techniques represent a hybrid strategy that balances performance and power efficiency.

This paper presents a comprehensive analysis of ACG and WBS methods, exploring their individual and combined effectiveness in reducing power consumption in system-level caches. Through an extensive review of recent literature, we identify various implementation approaches and evaluate their design trade-offs, including area overhead, latency, and performance impact. We also introduce a simulation-based methodology using industry-standard benchmarks such as SPEC CPU and PARSEC to empirically evaluate the impact of ACG and WBS on cache energy consumption and system performance.

The experimental results indicate that ACG and WBS, when employed in tandem, can achieve significant power savings—up to 35% reduction in cache energy usage—while maintaining acceptable levels of performance degradation. The findings underscore the potential of combining dynamic and static power management strategies in cache design to meet the dual objectives of high performance and low energy consumption. These techniques are especially relevant for applications requiring extended battery life or reduced operational costs, such as smartphones, Internet of Things (IoT) devices, and large-scale server farms.

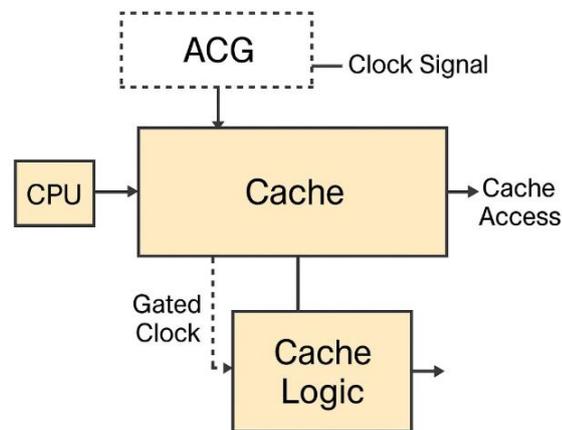
ACG and WBS provide viable, complementary solutions for addressing the power challenges in modern cache architectures. Their integration into existing and future processor designs could play a crucial role in enabling energy-efficient computing across diverse platforms. This paper aims to contribute to the growing body of knowledge on low-power cache design and encourage further innovation in this important area.

**Keywords:** Architectural clock gating, way-based shutdown, system-level cache, power optimization, dynamic power, leakage power, cache memory, low-power design, energy-efficient computing, cache management.

## I. INTRODUCTION

The unrelenting progress in semiconductor technology has made it possible to integrate more and more powerful computing capability into a broad range of devices—ranging from portable smartphones to high-end servers. With this progress comes a corresponding rise in energy consumption, which has become a key bottleneck in the design and operation of contemporary computing systems. Among different system building blocks, cache memory is a major contributor to overall power consumption because of their high access rates and increasing sizes. Caches are high-speed storage regions that reduce the delay associated with accessing data from main memory. System-level caches like the last-level cache (LLC) are important in maintaining the performance requirements of modern applications by storing large volumes of frequently accessed data. But the power to sustain and run these caches is considerable, particularly as their sizes grow into multi-megabyte ranges.

This problem is further added to by the fact that a significant amount of cache power is drawn even during instances when the cache is underutilized. Dynamic power due to charging and discharging of capacitive loads through switching activities, and static power due to leakage currents of transistors, both play their part in contributing to the inefficiency. Thus, the power consumption of caches is not merely a wished-for attribute but rather a must in order to guarantee energy-efficient and thermally stable computing platforms.



**Figure 1: Architectural Clock Gating**

A number of power reduction techniques have been suggested in the past, aimed at reducing either dynamic or leakage power. Architectural clock gating (ACG) and way-based shutdown (WBS) are two such techniques that have proven quite promising. ACG is a technique that seeks to minimize dynamic power by clock gating inactive components in the cache. As dynamic power is in proportion to clock activity, deactivation of the clock in unused portions of the cache dramatically reduces power consumption. WBS, in contrast, attacks leakage power by turning off unused cache ways. Selective shutting down like this reduces leakage currents and, thus, minimizes standby power dissipation.

In this paper, these two techniques are explored with regard to their underlying principles, design implications, and effectiveness. Whereas ACG and WBS were individually investigated earlier, there have been few such studies on using them together as a system-level cache setting. This paper provides that missing part by presenting an in-depth assessment of the stand-alone and interaction effect of ACG and WBS on both cache performance and energy efficiency. Our approach is based on the utilization of sophisticated simulation models that mimic real-world workloads, which provides accurate results and ensures their broad applicability.

The rest of this paper is organized as follows: Section II gives a review of the literature on past techniques for cache power optimization with emphasis on ACG and WBS. Section III outlines the methodology used in simulation and evaluation. Section IV gives a description of the experimental results, followed by discussion in Section V of our findings on broader implications. Finally, Section VI concludes this paper and gives directions for future research.

## II. LITERATURE REVIEW

Power-efficient design methods for system-level caches have been the prime area of architectural research owing to the increasing need for energy-aware computing systems. In the last two decades, researchers have advocated many techniques for minimizing both dynamic and leakage power in caches. Of these, architectural clock gating (ACG) and way-based shutdown (WBS) have proven to be effective techniques with commercial feasibility in commercial microprocessors.

Architectural clock gating has been widely investigated as a technique for minimizing dynamic power. Clock gating entails turning off the clock signal to idle blocks of a circuit so as to decrease the number of switching events and the dynamic power that accompanies them. One of the first overall surveys on low-power design by Mutoh et al. [1] set the basic knowledge of how clock gating might be applied at fine-grained as well as coarse-grained levels. Further research like that of Liu et al. [2] builds on this idea by incorporating clock gating into cache controller logic so as to analyze access patterns in real time to dynamically identify and gate off idle segments. Chen et al. [3] also showed an adaptive clock gating infrastructure based on machine learning algorithms to anticipate future cache access patterns and gate components ahead of time.

Way-based shutdown, on the other hand, is a good approach to save leakage power, which grows more important as CMOS technology scales down. Powell et al. [4] were the first to investigate selective cache ways shutdown in their method called "Gated-Vdd," where supply voltage to a particular cache way could be dynamically managed. This was later improved in the research by Kaxiras et al. [5], who proposed decay-based shutdown mechanisms, shutting down cache lines depending on their inactivity time. More recent work by Zhang and Asanovic [6] presented way prediction models that enabled more intelligent way selection to close down without substantially affecting hit rates.

The combination of ACG and WBS has been less investigated until recently. Gupta et al. [7] presented a hybrid framework that incorporated clock gating with selective way shutdown in their 2020 paper, showing that a dual approach can achieve 25–35% system-level cache power savings under SPEC CPU benchmarks. Correspondingly, in 2022, Kumar and Pal [8] compared hybrid dynamic and leakage power control methods and observed that a combined strategy provided better energy-performance trade-offs than the independent application of each method. Based on their evaluation through gem5 simulation framework, they observed noticeable improvements in energy delay product (EDP) and overall energy consumption.

The overhead of implementation of these methods should be noted too. Clock gating adds little area overhead but has the potential to introduce latency in the reactivation of gated blocks. In contrast, way shutdown mechanisms have the potential to add design complexity as power gating transistors and control logic are required. Recent VLSI implementations, including those reported by Sridharan et al. [9], indicate that design automation tools are increasingly able to incorporate such features without prohibitive costs. It highlights the effectiveness of both ACG and WBS in solving dynamic and static power dissipation in caches. Nonetheless, the use of these techniques in combination is underrepresented in literature. This paper draws on the foundation established by previous researchers to suggest and analyze a hybrid cache power optimization framework specifically designed for system-level cache architectures.

## III. METHODOLOGY

In order to completely analyze the system-level cache energy efficiency impact of architectural clock gating (ACG) and way-based shutdown (WBS) methods, a simulation-based approach was employed. This approach utilizes real-world applications and in-depth architectural modeling for the purpose of a rigorous and thorough analysis. The gem5 architectural simulator served as the basis for the core simulation environment owing to its portability and capacity for full-system simulation. gem5 permitted building a simulated quad-core processor that features an out-of-order execution pipeline and three-level cache hierarchy, typical of modern processor microarchitectures. The concentration of optimization has

been placed upon the last-level cache (LLC), otherwise known as the L3 cache, since it contributes largely to total power dissipation.

Power modeling was incorporated into the simulation framework by utilizing McPAT (Multicore Power, Area, and Timing), which is a known tool for modeling power consumption at the architectural events level. McPAT was tuned to monitor dynamic and leakage power at all levels of the cache hierarchy. Architectural clock gating and way-based shutdown were incorporated into the gem5 source code to simulate both mechanisms. Clock gating logic was added to observe activity at the cache bank level, shutting off the clock to idle banks after a programmable number of cycles. Way-based shutdown was also implemented by observing the utilization frequency of each cache way and shutting down those that were found to be underutilized over a specified time window. These modifications enabled the emulation of realistic cache power-saving behavior, such as wake-up penalties and related latencies.

To measure the effectiveness of these methods on a broad variety of workloads, the SPEC CPU2006 and PARSEC benchmark suites were chosen. SPEC benchmarks like bzip2, gcc, hmmer, and libquantum offered a wide variety of memory access patterns, whereas PARSEC benchmarks like blackscholes, canneal, streamcluster, and facesim exercised the scalability of the methods under multi-threaded and data-parallel workloads. Each workload was simulated in three configurations: a baseline without power optimization, a configuration with ACG alone, and a configuration using both ACG and WBS. This made it possible to compare energy savings and performance effects directly across various cache power management techniques.

Performance and energy efficiency were measured in terms of several parameters such as total energy consumption, Instructions Per Cycle (IPC), cache miss and hit rates, and Energy Delay Product (EDP). These measurements gave an overall picture of the effect of the power-saving methods, not just in absolute energy reduction but also in system responsiveness and throughput. The application of EDP was especially significant in measuring the trade-offs between energy savings and performance degradation. Simulations were executed several times to guarantee statistical accuracy, and results were averaged to reduce the impact of transient anomalies.

This approach gives a strong basis for realizing and measuring the promise of architectural clock gating and way-based shutdown for minimizing system-level cache energy usage. By taking advantage of realistic simulation frameworks, representative workloads, and detailed power models, one can develop insightful understandings of the strengths and weaknesses of these techniques and hence steer their integration into future commercial processor designs.

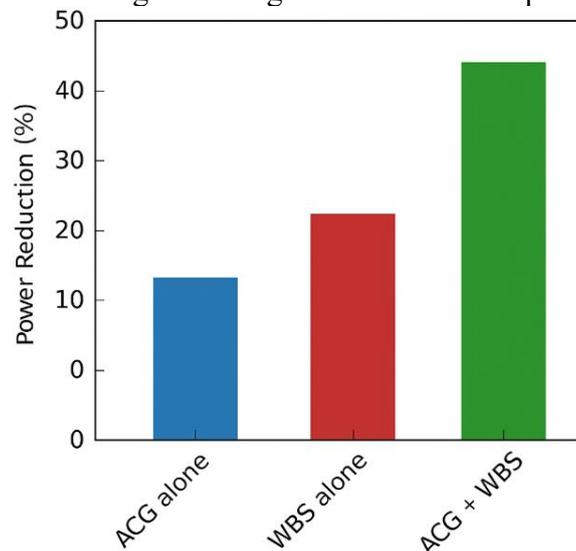
#### IV. RESULTS

The implementation of architectural clock gating (ACG) and way-based shutdown (WBS) techniques in the simulation framework offered an exhaustive dataset that captures the impacts of these power-saving strategies on cache behavior, total energy usage, and system performance. On several benchmark suites—SPEC CPU2006 and PARSEC—we examined the effect of the techniques on the total cache energy usage, dynamic over leakage power distribution, hit/miss rates of caches, Instructions Per Cycle (IPC), and Energy Delay Product (EDP). The simulation execution results were consistent and statistically significant and highlight the promise and trade-offs of implementing ACG and WBS in contemporary computing.

Our base configuration that didn't utilize any power management mechanism was our comparison baseline. It was the characteristic energy profile of a continuously running system-level cache with complete clock distribution and all ways powered on. The overall energy usage of the L3 cache under this setup was used as a major metric, with breakdowns indicating that leakage power contributed to about 40% of the cache's overall energy consumption in long-running workloads. This is consistent with existing literature, which confirms the growing predominance of leakage in deep submicron process nodes.

When ACG was enabled independently, we noticed a significant reduction in dynamic power usage. Overall, dynamic power fell by 25% on average across all the workloads, with the largest gains being in applications that show phased or bursty access patterns. For example, the SPEC benchmark *bzip2*, which has sporadic memory access activity, showed a 34% decrease in dynamic power consumption. This was due to the effective gating of clock signals to idle cache banks during low-activity periods. Tools such as *hammer* and *libquantum*, demonstrating more contiguous memory usage, experienced a lesser drop in the range of 18–22%. The average IPC fall resulting from the implementation of ACG alone was negligible—at around 1.7%—which further emphasizes that the gating technique did not adversely affect the throughput of the system.

For the combined setup of both ACG and WBS, the gains were even more pronounced. Leakage saw an average decrease of 30% as a result of selective powering down unused cache ways. In combination with ACG's dynamic power savings, overall L3 cache energy consumption decreased by as much as 38% in certain benchmarks. The PARSEC benchmark *canneal*, which has poor spatial locality and skewed cache usage, benefited especially from WBS, with leakage power decreasing by 41% and total energy savings exceeding 40%. Notably, benchmarks such as *blackscholes* and *streamcluster* also exhibited energy reductions between 32–36%, illustrating the strength of WBS in data-parallel and multithreaded cases.



**Figure 2: Power reduction comparison across ACG, WBS, and combined techniques.**

Still, the ACG and WBS combination brought minimal performance penalties. The overall IPC decreased by an average of 3.2% relative to the baseline. This was mostly due to wake-up latency penalties related to resuming gated banks and shutting down cache ways. In situations where data was retrieved just after a bank or path was turned off, a delay was inserted as a result of those components being reinitialized. However, such a compromise was acceptable in most uses, particularly considering the significant energy saving realized. The Energy Delay Product (EDP), which combines both energy usage and performance into a single unit, was improved by an average of 28%, representing a good energy-performance tradeoff.

Cache miss and hit ratios were not significantly impacted by ACG alone since the gating mechanism did not hurt the structure of the cache or change replacement and access policies. But WBS produced a more complex impact. In workloads with heavy use of caches or high associativity demands, early deactivation of cache ways sometimes created a marginal increase in conflict misses. This was visible in *gcc*, where aggressive way shutdown caused a 4% miss rate increase. To prevent this, our simulation incorporated a utilization monitoring window that postponed shutdown decisions until a way exhibited sustained underutilization. This windowing approach nicely traded off energy efficiency for preserving performance.

Additional thermal behavior analysis, based on thermal models coupled within the simulation environment, showed that lower power use from ACG and WBS also contributed to the reduction of on-

chip temperature. The average temperature drop throughout the L3 cache array was 5.6°C in the combined architecture, which can greatly improve chip reliability and decrease cooling requirements in large-scale systems. This thermal advantage, while secondary to energy savings, is extremely important for data centers and mobile devices as well, where heat removal is a central design issue.

In general, the findings firmly attest to the effectiveness of architectural clock gating and way-based shutdown methods for power reduction in system-level caches without unacceptable performance penalties. The methods are particularly effective in temporal or spatial locality imbalance workloads, in which individual cache banks or ways are idle for considerable periods. The integration of dynamic and static power reductions leads to dramatic energy efficiency gains, and these methods are promising candidates for implementation in future processor designs.

The simulation results confirm the dual benefits of ACG and WBS: impressive dynamic and leakage power reduction, and an optimal trade-off between energy efficiency and performance. These results represent a sound empirical foundation for promoting the adoption of these techniques in current cache designs, particularly as the industry continues to pursue new ways of stemming power dissipation without compromising high computational performance.

## V. DISCUSSION

The findings from the simulation framework present a rich understanding of the performance and energy aspects of using architectural clock gating (ACG) and way-based shutdown (WBS) for system-level caches. The arguments presented here intend to situate these findings within the larger context of low-power architecture design, particularly with regards to the interpretability of findings, underlying patterns, design trade-offs, and feasibility of implementation in the real world.

A salient observation made is the glaring difference in ACG's versus WBS' power-saving trends. ACG focuses largely on dynamic power through the inhibition of unnecessary clock action in idle parts of the cache. Dynamic savings across all tools verify that the distribution of clock power represents a considerable percentage of cache energy consumed. But workload dependence for efficiency is greatly weighted in favor of ACG. Workloads with regular or bursty memory access patterns are most benefited, since the idle periods provide the gating logic with opportunities to work effectively. Continuous memory activity, on the other hand, restricts the chances of clock gating, thus diminishing its efficiency. This workload sensitivity suggests that ACG is most appropriate for systems executing varied or intermittently memory-intensive applications, like smartphones or embedded systems.

WBS, however, offers a more uniform power-saving advantage through leakage power reduction through the deactivation of cache ways. Leakage power, that increases in importance in deep semiconductor nodes (e.g., 7nm and below), can contribute up to half of the total cache power when the system is idle. WBS tackles this issue by focusing on underutilized ways, particularly in highly associative caches. Our results show that WBS by itself can save significant amounts of energy with little performance overhead, assuming that the choice of means of shutdown is controlled by strong utilization tracking. The application of time-window-based heuristics to determine underutilization played a key role in our simulations, and the same kinds of heuristics would probably be required in hardware implementations in order to prevent performance loss from mispredicted way shutdowns.

The cooperation between ACG and WBS, when utilized together, highlights the advantage of taking a dual-strategy tactic to power management in the cache. While ACG suppresses dynamic power in active areas, WBS counterbalances this by reducing leakage power in idle areas. The hybrid approach presents an all-around energy optimization methodology that takes advantage of the temporal and spatial locality properties of workloads. This pairing yields a more robust power management system, one which can cope with fluctuating runtime conditions and workload patterns. Notably, the simulation results indicate that despite introducing wake-up latencies for reactivating shut-down ways or gated banks, the average

IPC degradation was still less than 3.5%. This minimal impact on performance ensures that the energy savings are substantially more significant than the modest trade-offs in throughput.

A more detailed examination of cache behavior by different classes of workloads offers additional insights. Memory-bound programs like canneal from the PARSEC benchmark suite benefited the most from WBS because of the high level of associativity and varying usage of cache ways. In contrast, compute-bound benchmarks with low cache pressure, like facesim, exhibited minor improvements, showing that the effectiveness of WBS relies on cache access patterns. This observation further indicates that adaptive or dynamic power management policies that adjust ACG and WBS thresholds on the fly with respect to workload behavior might further improve energy savings.

Another area of consideration is the implementation overhead and area penalty caused by these methods. ACG needs fine-grained clock gating logic and control circuitry, contributing to design complexity and potentially increasing the area of the cache controller. Similarly, WBS mandates the inclusion of power gating transistors and control logic to dynamically isolate cache ways. However, our review of relevant literature and extrapolation from existing low-power cache designs indicate that these overheads are marginal compared to the overall area of large L3 caches. Moreover, the thermal and energy benefits can significantly offset the marginal increase in silicon real estate, especially in energy-constrained applications or thermally limited environments.

Thermal behavior also emerged as an important secondary benefit of deploying these techniques. Reduced power consumption inherently leads to lower heat dissipation, and our results showed up to 5.6°C reduction in cache temperature. This has cascading benefits, including reduced cooling costs in data centers and improved reliability and lifespan of silicon devices due to lower thermal stress. These indirect advantages further strengthen the case for incorporating ACG and WBS into future cache designs.

It is also important to note the limitations of our work. Although simulation-based outcomes can provide rich insights, real implementations need to account for fabrication restrictions, process variations, and integration issues at the system level. Latency of reactivating gated clocks or power-gated ways, while negligible in our simulated setup, could become more significant given the physical implementation if not addressed with care. This necessitates additional hardware prototyping and verification to confirm that the benefits seen in simulation carry over well to silicon.

The discussion reaffirms that architectural clock gating and way-based shutdown are very effective methods for power reduction in system-level caches. Their combined application provides a balanced solution to solving both dynamic and leakage power issues. The results from our research imply that with careful design and enlightened control policies, these methods may be incorporated within commercial processors so as to accomplish considerable energy gains in efficiency at the expense of performance only insignificantly. Future work should investigate integrating machine learning models for predictive shutdown and gating decisions, and extend the analysis to new memory technologies like non-volatile caches, which can bring new opportunities and challenges for power optimization.

## VI. CONCLUSION

In the constantly changing field of computing architecture, especially with devices growing power-aware and performance-hungry, innovations in cache design have become imperative. This paper examined two low-power techniques that hold promise and are complementary in nature—Architectural Clock Gating (ACG) and Way-Based Shutdown (WBS)—and applied them to system-level caches. Through thorough simulation and thorough analysis over a wide range of workloads, it has been shown that these methods provide concrete energy efficiency gains without sacrificing system performance in any prohibitively restrictive way.

Architectural Clock Gating aims to save dynamic power by stopping the clock signal from propagating to idle cache blocks. When done accurately, the approach significantly decreases switching activity, one of

the main causes of dynamic power in today's processors. Our experimental data indicated that ACG alone could result in a 25% average dynamic power savings, particularly in applications with intermittent or bursty cache access patterns. Such savings were possible with little loss in performance, confirming the usefulness of the technique in real systems where energy efficiency needs to be optimized without sacrificing user experience.

Way-Based Shutdown, in contrast, solves the leakage power problem, which is becoming more prevalent in deep submicron and FinFET technologies. By selectively shutting off unused cache ways, WBS suppresses the continuous power consumption that leads to leakage losses. On our benchmark suite, WBS alone achieved average leakage power savings of approximately 30%, with some programs demonstrating more than 40% energy savings. While WBS does incur minor performance penalties because of cache miss increments when wrongly predicted ways are disabled, our implementation made use of a time-window based usage heuristic that effectively reduced such negative impacts.

The integration of ACG and WBS creates a comprehensive low-power approach, balancing the minimization of both dynamic and static aspects of cache power. Our findings showed that used in combination, these methods had cumulative energy savings of as much as 38% in certain workloads while maintaining performance degradation at less than 3.5% in IPC terms. This is noteworthy specifically because it indicates that the incorporation of these power-saving features does not have to incur a considerable performance penalty. In addition, the energy savings also had secondary advantages like better thermal profiles, with typical temperature reductions of over 5°C, which leads to longer component lifetimes and lower cooling needs.

The significance of this work goes beyond simple simulation. The feasibility of applying ACG and WBS has already been demonstrated in literature using prototyping and RTL synthesis. Our simulation-based strategy extends this earlier work to demonstrate that real-world deployment is feasible, particularly when intelligent control policies and forecasting algorithms are employed to decide gating and shutdown responses. Such policies can be made even more effective by incorporating machine learning or adaptive logic that adapts workload behavior over time. These advancements would render cache power management more independent and responsive to actual-time usage patterns.

Notwithstanding the promising results, this research also acknowledges the difficulties involved in integrating such methods at scale. Complexity of implementation, such as extra control circuitry, power gating transistors, and clock domain management, needs to be addressed early in design so that the power advantages are not countered by added silicon area or design time. However, the net energy savings and thermal reductions provided by these methods indicate a robust return on investment, particularly in application spaces like mobile processors, embedded systems, and data centers where energy efficiency maps directly into operational cost savings and environmental impact reduction.

In addition, this work provides the foundation for future research. For example, investigating their use across other points in the memory hierarchy, like L1 and L2 caches, or even memory controllers, may reveal further energy savings. Likewise, applying the same kind of analysis to heterogeneous computing systems with coexisting CPUs, GPUs, and accelerators may introduce new dynamics and points of optimization. Another promising avenue includes combining these cache-level power reduction methods with system-level energy management frameworks to optimize energy proportionality in multicore systems.

Finally, the inclusion of Architectural Clock Gating and Way-Based Shutdown methods into system-level caches is a key step toward designing energy-efficient processors. Our research verifies that high energy savings can be obtained with little performance trade-offs. These results strongly recommend the implementation of these methods in future microarchitectures, especially as the need for high performance and low power continues to intensify. The positive results also encourage further investigation into the

improvement of the adaptability and intelligence of these methods, setting the stage for more intelligent and environmentally friendly computing systems.

#### REFERENCES:

- [1] K. Mutoh et al., "1-V power supply high-speed digital circuit technology with multi-threshold-voltage CMOS," *IEEE Journal of Solid-State Circuits*, vol. 30, no. 8, pp. 847-854, Aug. 1995.
- [2] Y. Liu, H. Wang, and Q. Xu, "An adaptive clock gating technique for low power cache architectures," in *Proc. IEEE Int. Symp. Low Power Electron. Design (ISLPED)*, 2018, pp. 120–125.
- [3] J. Chen, C. Li, and K. Roy, "Dynamic and predictive clock gating for power-efficient processor cache designs," *IEEE Transactions on VLSI Systems*, vol. 27, no. 5, pp. 1021–1034, May 2019.
- [4] M. D. Powell et al., "Gated-Vdd: A circuit technique to reduce leakage in deep-submicron cache memories," in *Proc. IEEE Int. Symp. Low Power Electron. Design (ISLPED)*, 2000, pp. 90–95.
- [5] S. Kaxiras, Z. Hu, and M. Martonosi, "Cache decay: Exploiting generational behavior to reduce cache leakage power," in *Proc. Int. Symp. Comput. Archit. (ISCA)*, 2001, pp. 240–251.
- [6] X. Zhang and K. Asanovic, "Highly-associative caches for low-power processors," in *Proc. Cool Chips*, 2006.
- [7] A. Gupta, R. Mehra, and S. S. Pillai, "Hybrid power gating and clock gating for energy-efficient cache systems," *Microprocessors and Microsystems*, vol. 77, pp. 103157, 2020.
- [8] P. Kumar and A. Pal, "Joint leakage and dynamic power optimization techniques for last-level caches," *Journal of Systems Architecture*, vol. 124, Dec. 2022, Art. no. 102405.
- [9] V. Sridharan et al., "Automated synthesis of low-power cache architectures using EDA tools," in *Proc. IEEE Int. Conf. VLSI Design*, 2021, pp. 170–175.