

Intel® performance hybrid architecture & software optimizations

Part One: Introduction to performance hybrid architecture for 12th Generation Intel Core processors

This document is the first in a series of whitepapers that will serve to provide an overview of performance hybrid architecture.

This first whitepaper introduces audiences to the performance and efficiency improvements that come from implementing a performance hybrid architecture solution.

In summary, performance hybrid architecture enables optimal performance on multi-threaded, limited threading, and power-constrained workloads, allowing end-users with varied data processing needs to experience faster computing speeds and a higher level of focused computing power.

Authors

Nikhil Rukmabhatla

Software Enabling and
Optimization Engineer

Rajshree Chabukswar

Senior Principal Engineer

Sneha Gohad

Software Enabling and
Optimization Engineer

Michael Chynoweth

Intel Fellow

Table of Contents

Introduction to performance hybrid architecture.....	1
Intel performance hybrid architecture.....	2
Motivation	2
Architecture	2
ISA Support.....	3
OS Support	4
Intel Thread Director	4
Conclusion.....	4
References.....	5

Introduction to performance hybrid architecture

In every industry and every part of the world, the ability to collect and process data is becoming an increasingly important factor in business success. As data processing needs grow, the need to regulate how that data is collected, processed, managed, and stored increases as well.

For developers, the serious questions start to pile up fast.

How do I...

- Better handle all of my workloads by utilizing all of the compute capabilities and cores?
- Manage the use of more efficient cores for tasks that need less performance but better energy efficiency?
- Ensure that each independent workload is directed to the right processor?
- Get the necessary power to the right cores at the right time?
- Tap into performance cores when needed and prevent them from being used for less important tasks?
- Scale my performance across multiple cores properly to account for the increased workload?
- Enable my application for new performance hybrid architecture without rewriting the code?

**The answer to all of these questions is
Intel performance hybrid architecture.**

A first of its kind in computing innovation, performance hybrid architecture was developed to meet the most demanding workload requirements by bringing maximum computing power to the areas where it is needed most.

Research shows that this is a solution required by real-world workloads. A recent Intel study, which examined the performance of various workloads from multiple segments by using an increasing number of CPU cores, produced the following results:

- A majority of the workloads do not scale beyond 4 cores (many of these limited threading workloads closely resemble actual user experience workloads).
- A minority of the workloads can scale to 8 cores but do not scale any further.
- An even smaller minority of workloads can scale higher than 10 cores and continue to scale with core count.

The results of this study highlighted the fact that the majority of client applications would benefit from better scalability to 8 cores or more. To better serve this market segment, Intel has designed a System on Chip (SoC) architecture where larger cores are unleashed to go after single-threaded and limited threading scenarios, while the efficient multi-threaded cores can help extend scalability performance over prior generations. The result of this effort is the development and introduction of Intel performance hybrid architecture.

Intel performance hybrid architecture

Motivation

Intel's contributions to computing technology go way beyond the hardware and software options that we bring to market. Our goal is to assist developers, making it easier for them to write software targeting our platforms to benefit their customers and elevate their experience. One of the ways that we achieve this is by keeping a consistent focus on making sure that our hardware and software offerings work in harmony.

Intel performance hybrid architecture is the ultimate result of this focus.

Intel performance hybrid architecture helps enable optimal performance on multi-threaded and limited threading workloads. It also enables power-constrained workloads when there may be a need to divert power to other compute hardware. Developers can harness this technology to help achieve the best possible performance on available hardware helping users to experience faster computing speeds on any software stack including gaming, content creation, productivity, or browsing.

Architecture

The new 12th Generation Intel Core processor SoC incorporates performance hybrid architecture, merging two existing architectures: Performance-cores (P-cores) and Efficiency-cores (E-cores). Both facets complement each other to deliver the best possible experience to customers and end-users. Single and multiple-threaded operations are performed simultaneously, despite the constraints dictated by the power capacities and other equipment limitations. Intel Thread Director—a new hardware feature on our 12th Generation Intel Core processors—has been added to assist the OS scheduler to allow the system to make more intelligent and data processing decisions regarding thread scheduling.

By having both P-cores and E-cores, performance hybrid architecture can generate a more efficient distribution of core usage depending on the application. This happens because P-cores increase performance to handle complex (typically workloads with limited threading) workloads. E-cores meanwhile focus on multi-threaded throughput and power-limited scenarios.

As shown in the graph below, a P-core provides more performance compared to an E-core at higher power envelopes. At lower power envelopes, E-cores have better PnP characteristics than P-cores.

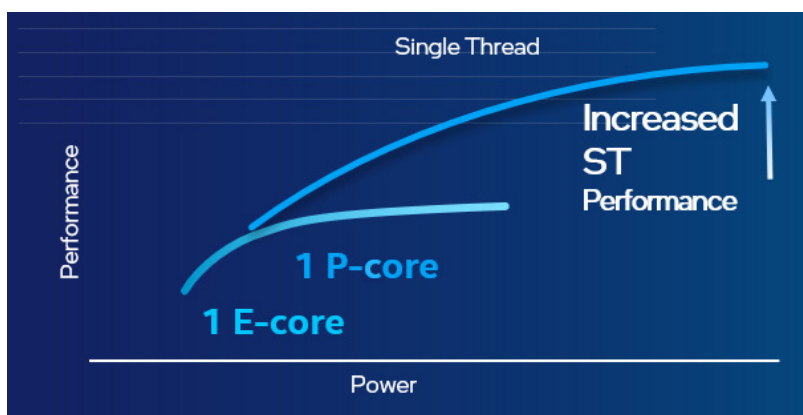


Figure 1. Single-threaded PnP characteristics on P-core and E-core

The multi-threaded performance throughput of an E-core module is better than a P-core at all power envelopes. At a similar power envelope, the E-core module outperforms SMT-enabled P-core and throughput can be increased significantly as shown in the graph below.

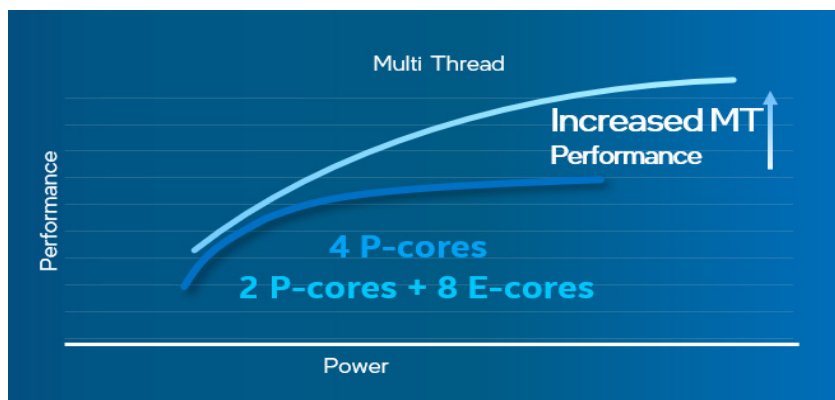


Figure 2. Multi-threaded PnP characteristics on P-cores and E-cores

For workloads and configurations visit www.intel.com/PerformanceIndex. Results may vary.

For these reasons, the P-cores are preferred for the priority tasks and limited threading applications, while the E-cores are available to perform in power-limited scenarios and/or background applications that can meet their Quality of Service (QoS) requirements on that performance. By blending the two in our performance hybrid architecture, the Intel Thread Director can assist the scheduler in handling complex, multi-tiered workloads. These workloads can be directed by the Intel Thread Director to the appropriate and most efficient core.

To summarize, performance hybrid architecture has P-cores and E-cores working together. This creates an exciting opportunity because this solution can work well on single-threaded or partially threaded applications as well as multi-threaded applications. This is how Intel achieves generation-over-generation performance: efficient hybrid thread scheduling through the Intel Thread Director, IPC improvements through uArch, and the related OS optimizations.

Instruction Set Architecture (ISA) Support

As a new architecture emerges, the support for Instruction Set Architecture (ISA) has been adapted to the new working parameters, along with the most recent updates and changes to the design added to the system. Our 12th Generation Intel Core processors have a common subset of ISA for both P-cores and E-cores. AVX512 is not supported on P-cores to ensure that we have symmetric ISA available across E-cores and P-cores on the platform.

Our 12th Generation Intel Core processors have enhanced support for a variety of instructions over previous generations of Intel CPUs. One of the most recent examples is porting AVX512 Vector Neural Network (VNNI) instructions to AVX to allow our 12th Generation Intel Core processors to efficiently run and scale this ISA across all available E-cores and P-cores.

In addition to the AVX support, optimized wait instructions (UMWAIT, TPAUSE) support is also added to our 12th Generation Intel Core processors ISAs via already enabled threading libraries. These ISAs can be utilized to create lightweight spins that have lower latency compared to traditional busy-spins utilizing Sleep(0), SwitchToThread, or sched_yield calls. Also, these spins have a preference for running on the E-cores to ensure they are not taking up higher-demand compute resources.

The table below shows a list of ISA support on the most recent 12th Generation Intel Core processor architecture.

Feature	Alder Lake Support
AVX512	Not supported. VNNI for Machine Learning on AVX on both P-core and E-core is available.
VNNI	VNNI backported to AVX
AES512/PCLMULQDQ512 (Encryption Perf)	vAES/vCLMUL in AVX2 with slight perf delta
Remote Action Request (RAR) (perf request)	Not supported
FP16 Data Type (For AI) on AVX512	Disable & default back to FP32
PMON, Debug (PEBS, PT)	Profiling tools cover performance monitoring differences
IPI virtualization	Disabled in E-core
TSX	Disabled in P-core
Lightweight spins (UMWAIT, TPAUSE)	Enabled through OS, Intel® Thread Director and threading libraries

Table 1. ISA availability on the new 12th Generation Intel Core processors with performance hybrid architecture

OS Support

On our 12th Generation Intel Core processors, unbalanced thread execution has been carefully reviewed to try and fix corner cases that could arise and compromise performance.

To address these implications, Intel, Operating System (OS) vendors, Independent Software Vendors and other tools developers have collaborated to accomplish the following:

- Refine OS scheduling with specific optimizations and the addition of Intel Thread Director feedback in scheduling decisions.
- Prepare threading libraries to ensure the optimal scheduling of threads on hybrid platforms.

Two main characteristics help the OS scheduling enhancements determine the thread QoS and guide scheduling: whether or not the application is a foreground or background application, and whether or not the application has an impact on the customer or user experience. This is to reduce the burden of enabling these applications through optimal thread scheduling which delivers the best performance and throughput. The expectation is that, commonly, software developers will not need to make major changes since OS optimizations are readily available out of the box.

Intel Thread Director

One of the key ways we worked to make performance hybrid architecture successful is through the new Intel Thread Director from the 12th Generation Intel Core processors. With Intel Thread Director, we can monitor and build intelligence directly into the core. The Intel Thread Director accomplishes three primary tasks:

- Monitors the runtime instruction mix of each thread as well as the state of each core, with nanosecond precision.
- Provides runtime feedback to the OS to make optimal scheduling decisions for any workload or workflow.
- Dynamically adapts guide based on the thermal design point, operating conditions, and power settings—without any user input.
- *For workloads and configurations visit www.intel.com/PerformanceIndex. Results may vary.*

Normally, the OS would be making these decisions based on static, pre-programmed threading with core assignments. To keep this overhead from going on the software, we wanted to develop a hardware solution that would assist the OS achieve optimal runtime scheduling by the Intel Thread Director giving the OS more information by monitoring instruction mix, the current state of each core, and the relevant microarchitecture telemetry at much more granular level than would be available via instrumentation

As we will discuss here and will be discussed more in future papers, this opens up a wide variety of opportunities for more directed and intelligent performance. We did this because we wanted to eliminate the need for software developers to have to rewrite their existing code while removing overhead from software scheduling tasks.

We already had a performance monitoring unit (PMU) that provides some of the best hardware telemetry in the industry. By developing a hardware solution like our 12th Generation Intel Core processors and pairing them with the Intel Thread Director, we allow this new solution to still access the PMU and provide the OS with that much-needed telemetry data. Intel Thread Director can communicate directly to Windows OS scheduler, providing “hints” to the scheduler about what task should be handled by which core. By making foreground and background tasks more efficient, Intel Thread Director can give hints to Windows OS scheduler to see that each task gets assigned to the proper core. As a result of these functionality and performance enhancements, tech news outlets are already seeing gains in Intel systems.

Conclusion

We believe that our performance hybrid architecture solution can help developers meet the new performant needs of their end-users through its unique architecture.

We believe this because, at Intel, we put a great deal of effort into continuously expanding our knowledge base regarding not only architecture but in the tools and techniques available to put them to better use. The creation of optimizations for our performance hybrid architecture has already begun and will not only expand as demand grows but be used to create a new baseline in building the future of this technology.

As a part of this effort, Intel optimized multiple frameworks and libraries for performance hybrid architecture, and enabled profiling tools for performance analysis. We will cover those topics in our second and third white papers:

- Developing for Intel performance hybrid architecture
- Debugging for Intel performance hybrid architecture

To read more about maximizing the use of hybrid chips, you can read over our press release titled “Directing Traffic to Maximize the Use of Hybrid Chips” [here](#) or search our site at Intel.com, keywords “Hybrid Chips”.

Disclaimer

Notice: This document contains information on products in the design phase of development. The information here is subject to change without notice. Do not finalize a design with this information.

References

Intel®. "Directing Traffic to Maximize the Use of Hybrid Chips." newsroom.intel.com. September 16, 2021.

Intel®. "Alder Lake and Intel® Thread Director- Architecture Day 2021". [YouTube.com](https://www.youtube.com). August 19, 2021.

Funk, Ben. "Windows 11 Seems to Give Intel® Hybrid CPU Architectures A Notable Performance Boost". [HotHardware.com](https://www.hothardware.com). June 18, 2021.



Performance varies by use, configuration and other factors. Learn more at www.intel.com/performance/index.

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. See backup for configuration details. No product or component can be absolutely secure.

All product plans and roadmaps are subject to change without notice.

Code names are used by Intel to identify products, technologies, or services that are in development and not publicly available. These are not "commercial" names and not intended to function as trademarks.

Intel disclaims all express and implied warranties, including without limitation, the implied warranties of merchantability, fitness for a particular purpose, and non-infringement, as well as any warranty arising from course of performance, course of dealing, or usage in trade.

Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.

No product or component can be absolutely secure. Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

Your costs and results may vary. Intel technologies may require enabled hardware, software or service activation.

© Intel Corporation.

Intel®, the Intel® logo, and other Intel marks are trademarks of Intel® Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.