Notices and Disclaimers

Performance varies by use, configuration and other factors. Learn more at www.Intel.com/PerformanceIndex.

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.

Intel contributes to the development of benchmarks by participating in, sponsoring, and/or contributing technical support to various benchmarking groups, including the BenchmarkXPRT Development Community administered by Principled Technologies.

Your costs and results may vary.

Intel technologies may require enabled hardware, software or service activation.

Some results may have been estimated or simulated.

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

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

Statements in this document that refer to future plans or expectations are forward-looking statements. These statements are based on current expectations and involve many risks and uncertainties that could cause actual results to differ materially from those expressed or implied in such statements. For more information on the factors that could cause actual results to differ materially, see our most recent earnings release and SEC filings at www.intc.com.

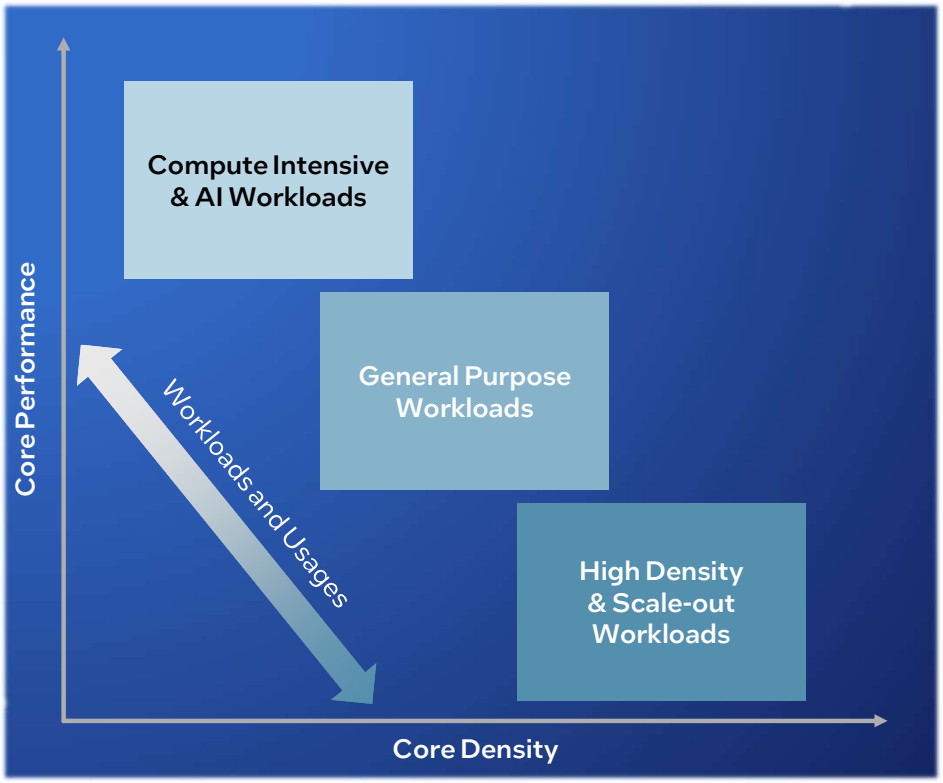
Hot Chips 2023

The Next Generation of High Performance, Energy-Efficient Computing: Intel® Xeon® Processors Built on Efficient-Core

Don Soltis
Senior Principal Engineer, Xeon with Efficient-Core Architect, Intel

Stephen Robinson
Intel Fellow, Efficiency Core Architect

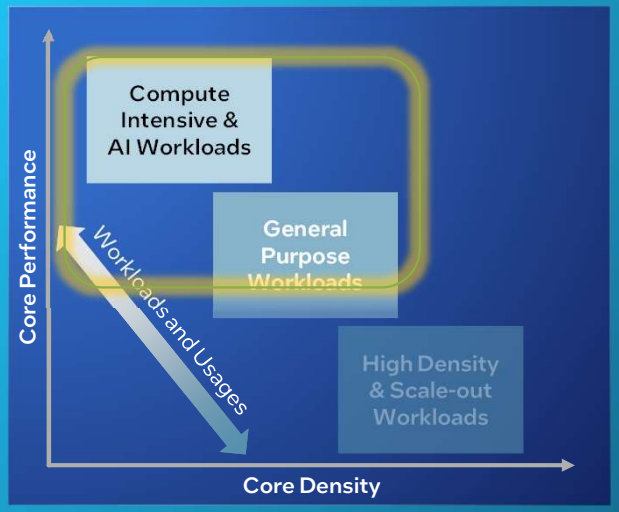
Data Center Requirements Are Expanding



- Continuing demand to increase core performance, power efficiency and throughput performance
- Deliver the best performance per Watt at the desired core performance
- Workload requirements place trade-offs between core performance & core density
- Expanding deployment models demanding increased power, IO & memory bandwidth

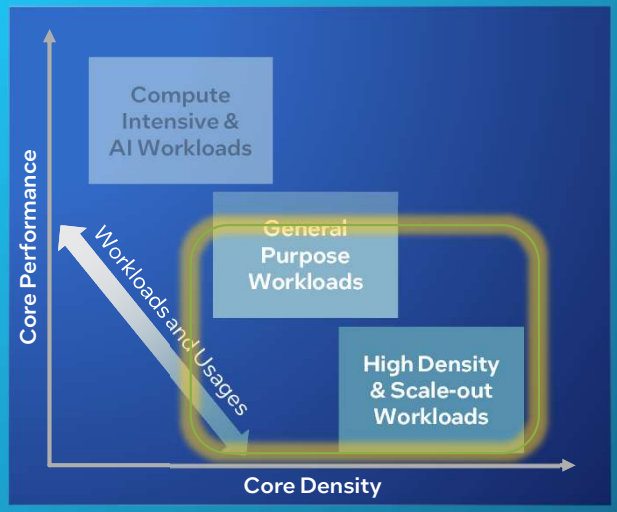
Expanded Xeon Product Portfolio with Optimized Processors

P-core



Optimized for Performance
in Compute Intensive & AI Workloads

E-core

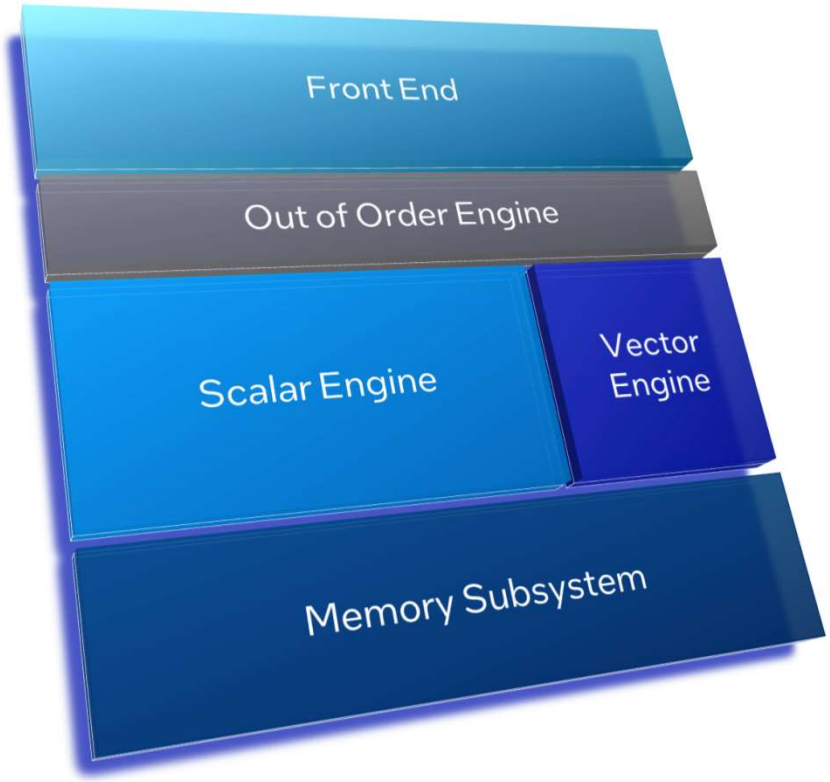


Optimized for Efficiency
in High Density & Scale-out Workloads

intel XEON

Common Platform Foundation
& Shared Software Stack

Intel's Newest Data Center Efficiency Optimized E-Core Microarchitecture



Designed for scalable throughput performance

Optimized for power and density efficient throughput with:

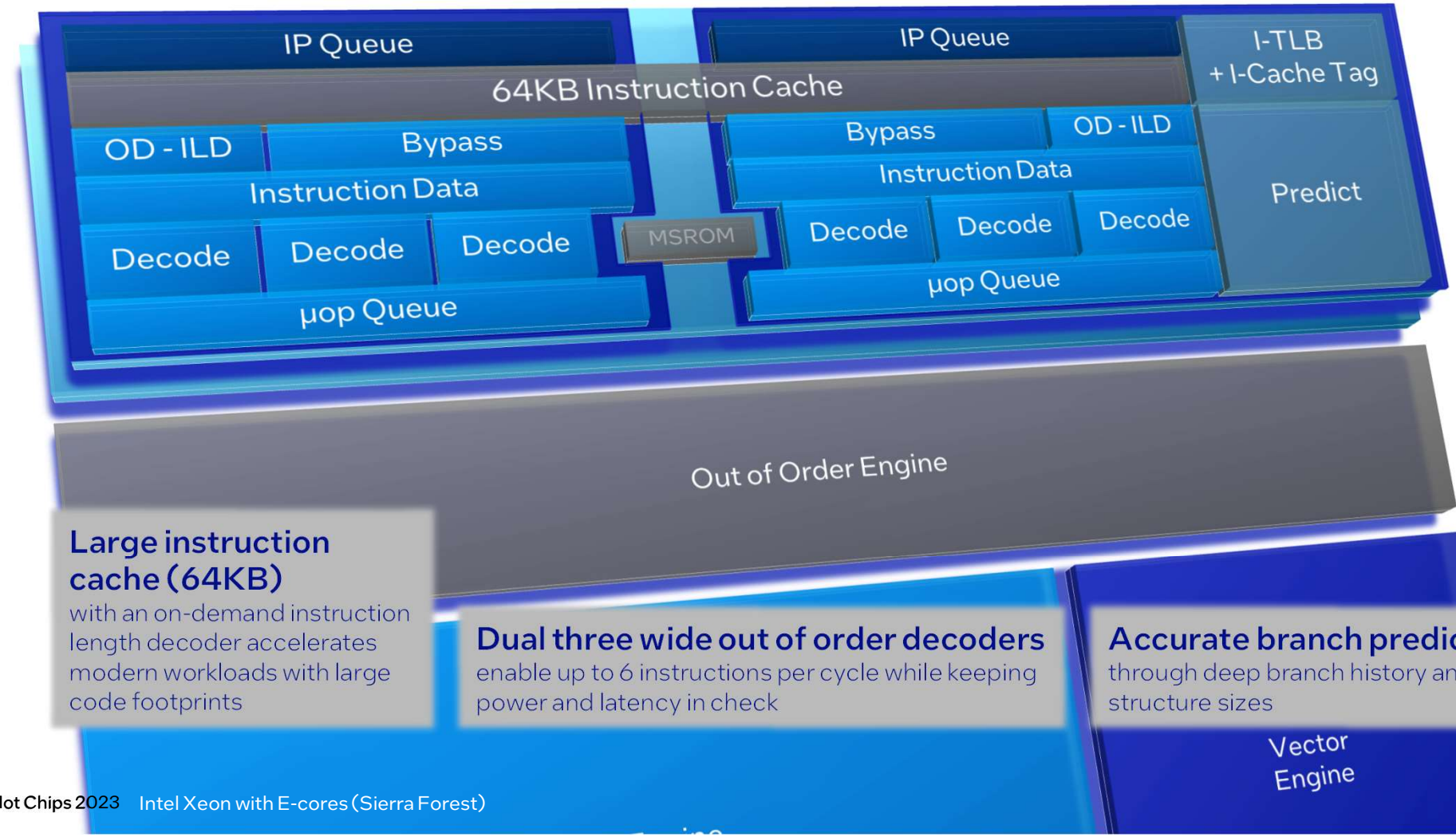
Deep Front-End with on-demand length decode

Wide Back-End with many execution ports

Optimized Design for latest transistor technologies

Data Center Ready

Core Instruction Control

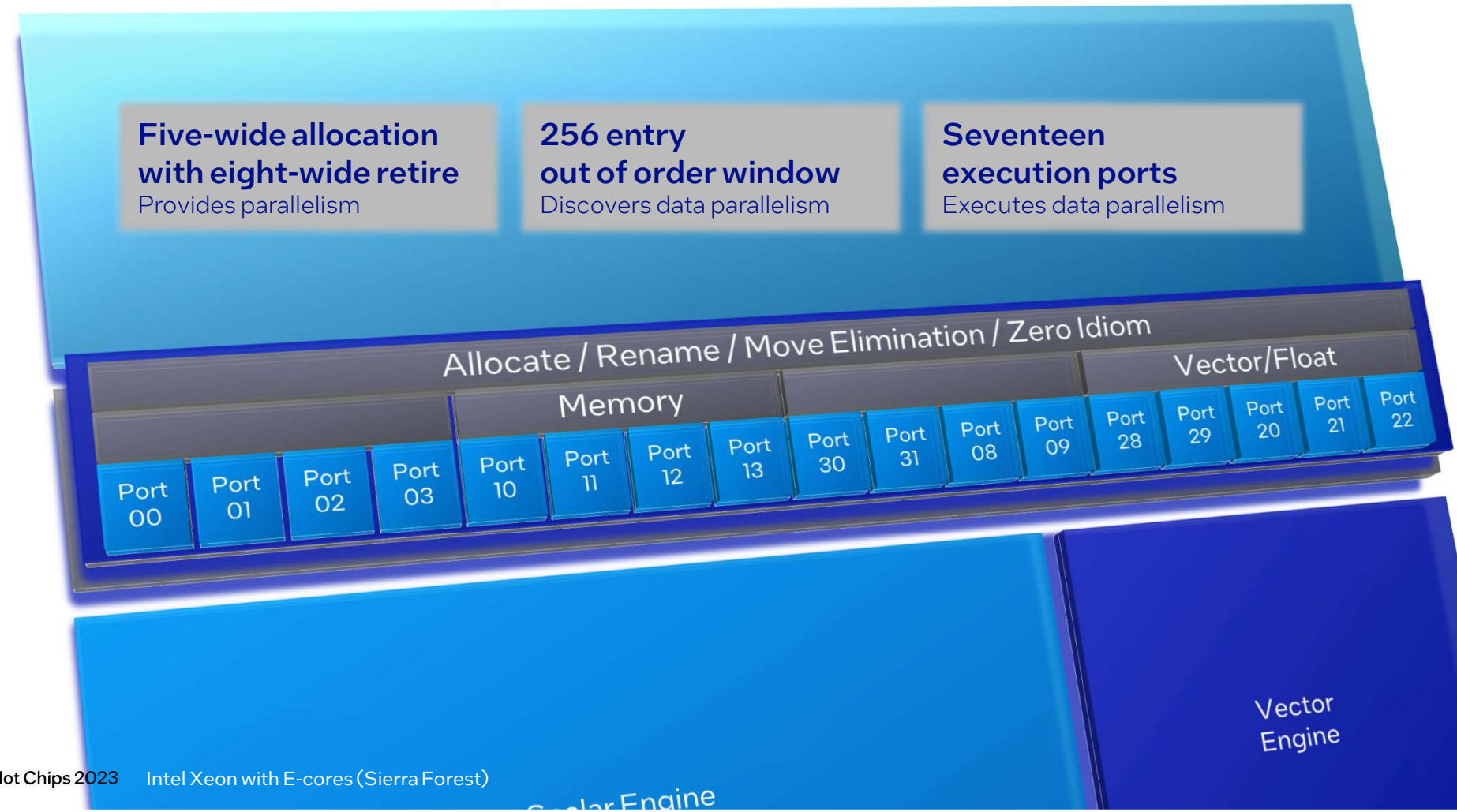


Large instruction cache (64KB)
with an on-demand instruction length decoder accelerates modern workloads with large code footprints

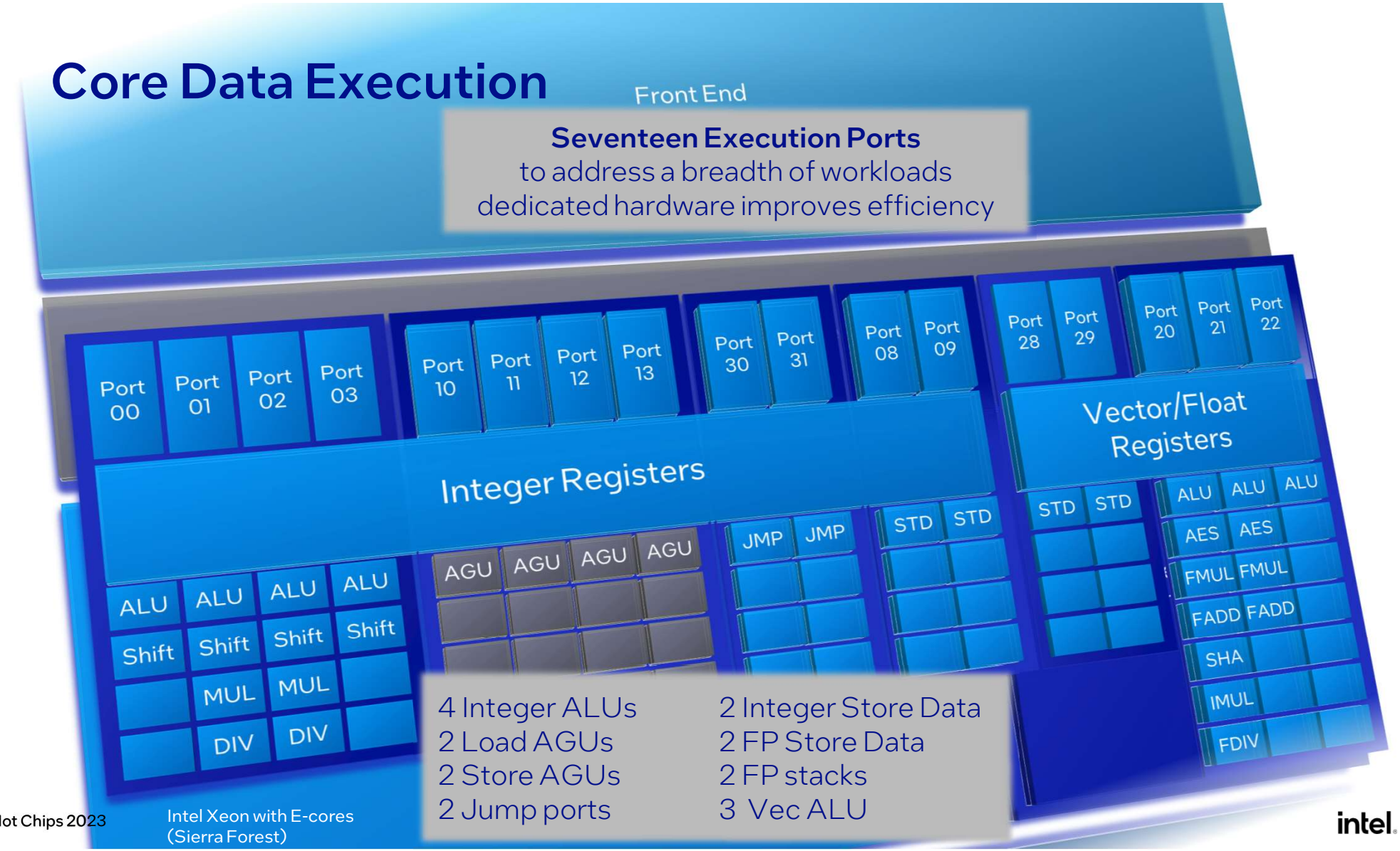
Dual three wide out of order decoders
enable up to 6 instructions per cycle while keeping power and latency in check

Accurate branch prediction
through deep branch history and large structure sizes

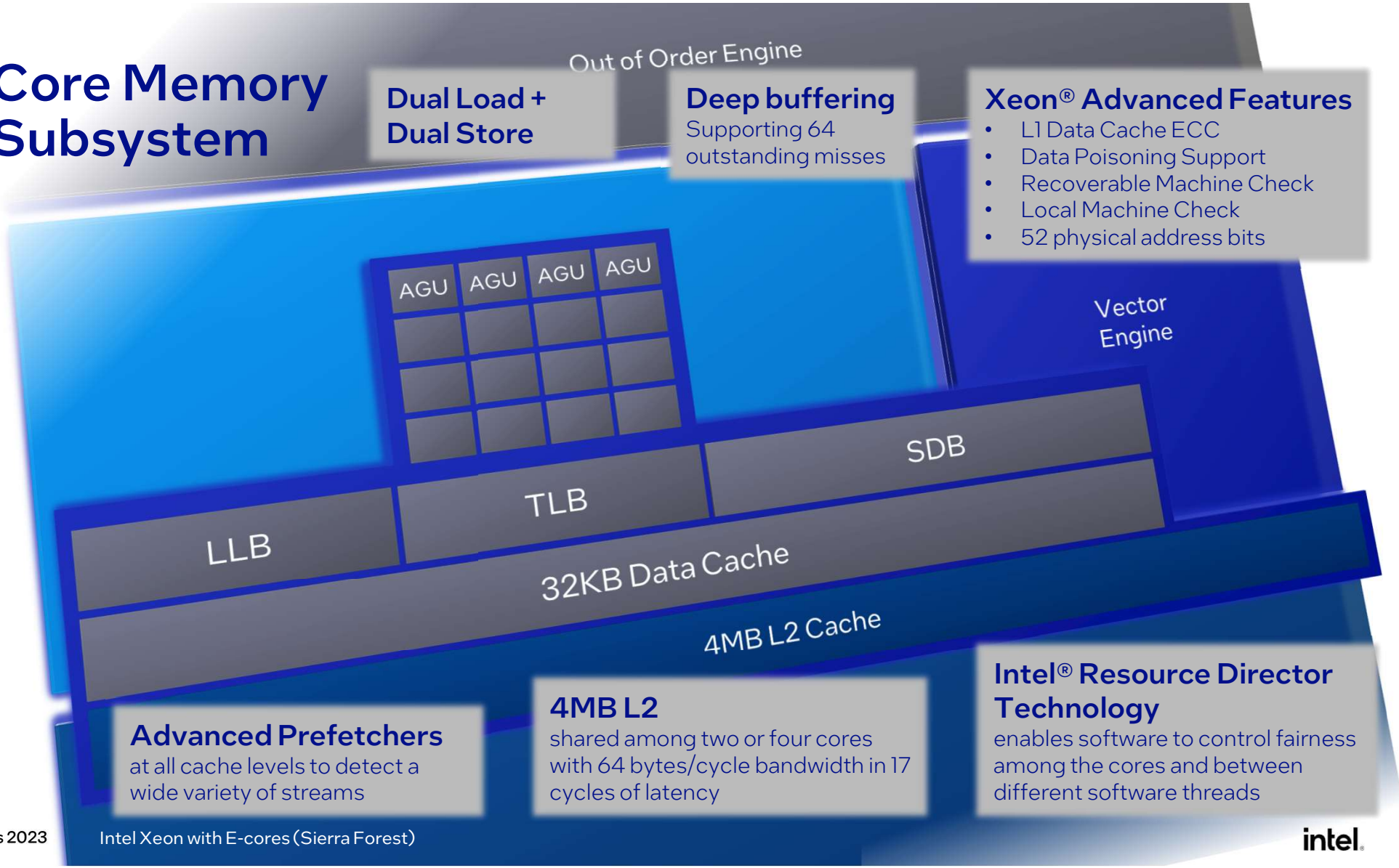
Core Execution



Core Data Execution



Core Memory Subsystem



Dual Load + Dual Store

Deep buffering
Supporting 64 outstanding misses

Xeon® Advanced Features
- L1 Data Cache ECC
- Data Poisoning Support
- Recoverable Machine Check
- Local Machine Check
- 52 physical address bits

Advanced Prefetchers
at all cache levels to detect a wide variety of streams

4MB L2
shared among two or four cores with 64 bytes/cycle bandwidth in 17 cycles of latency

Intel® Resource Director Technology
enables software to control fairness among the cores and between different software threads

Modern Instruction Set

Security Features

Intel® Control-Flow Enforcement Technology
(Intel® CET) designed to improve defense in depth

Intel® VT-rp
(Virtualization Technology redirect protection) Supported

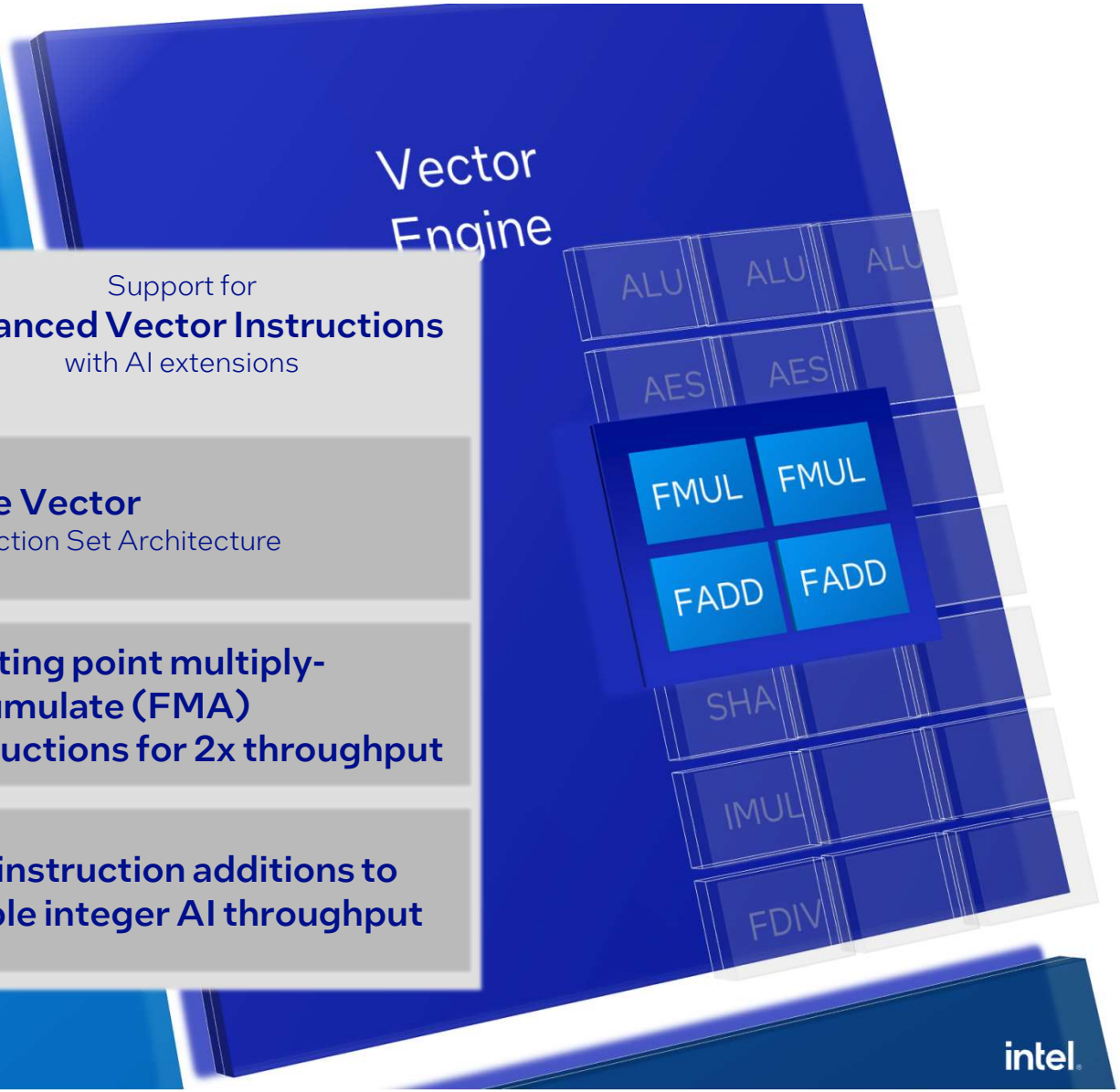
Advanced speculative execution validation methodology

Support for
Advanced Vector Instructions
with AI extensions

Wide Vector
Instruction Set Architecture

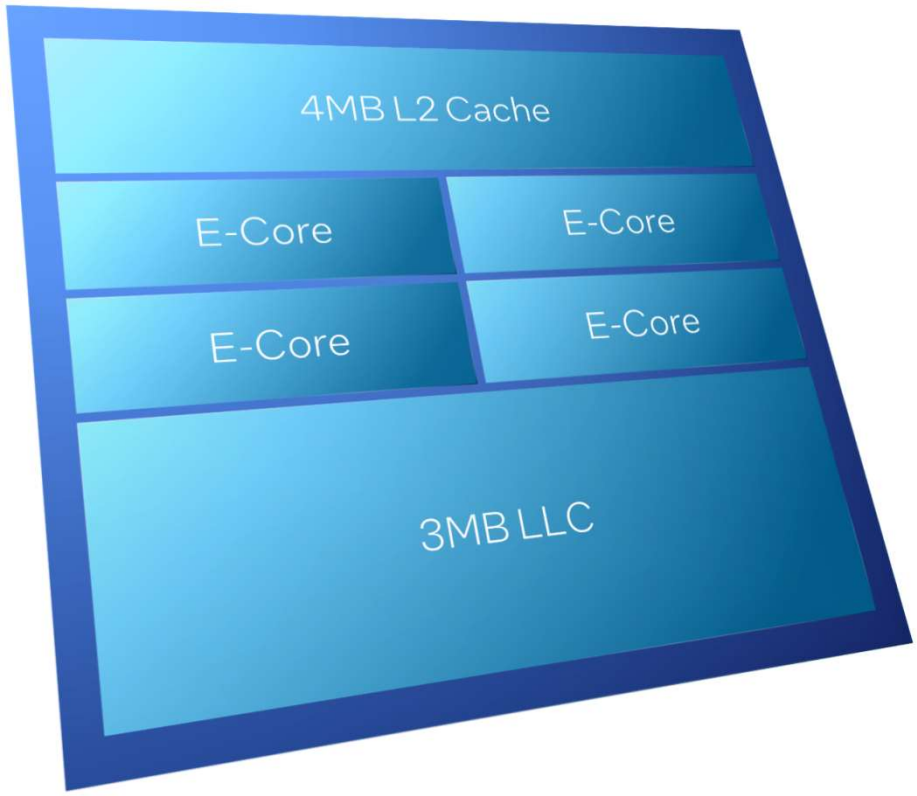
Floating point multiply-accumulate (FMA) instructions for 2x throughput

Key instruction additions to enable integer AI throughput



intel

Core Tile



2 or 4 cores per module
Shared L2 cache
Shared frequency and voltage domain
Shared mesh fabric interface

Each core is single threaded
Providing performance isolation

LLC slice shared among all cores in socket
High bandwidth pipeline per slice

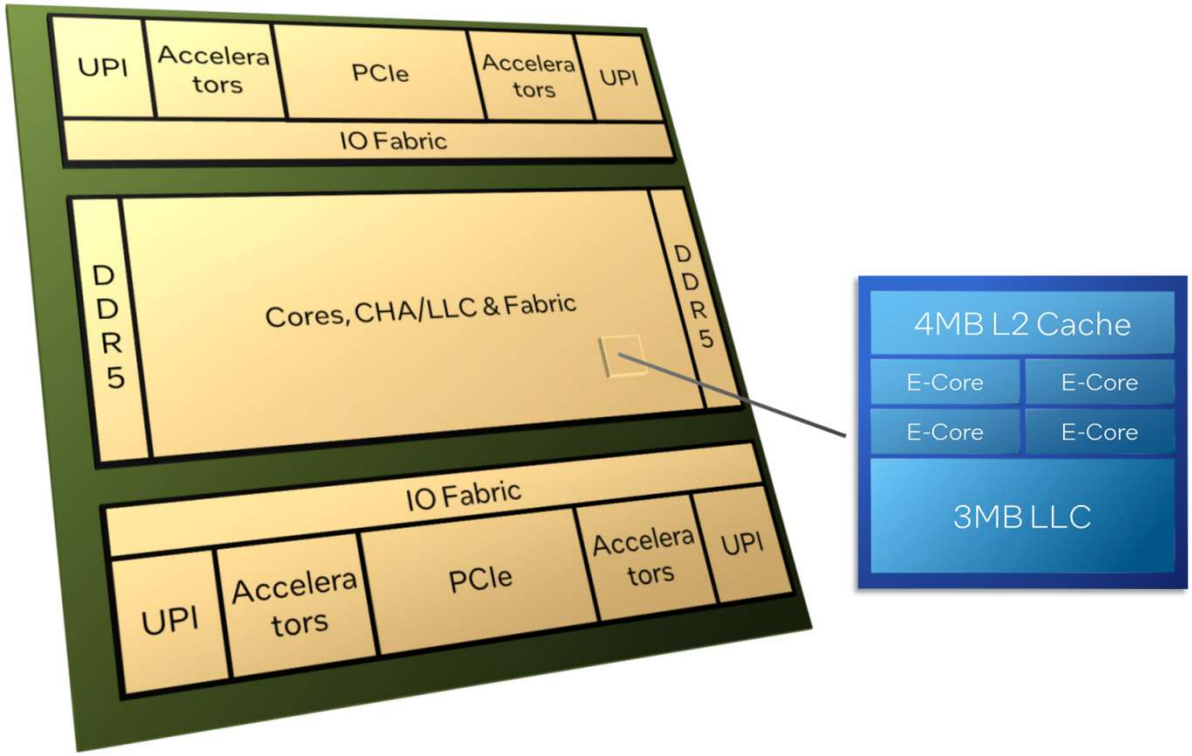
Package View

I/O Chiplet
Common with Xeon P-core (Granite Rapids)
PCIe gen 5, CXL 2.0, UPI coherency links
Crypto, compression, data streaming accelerators
Self-boot

Compute Chiplet
Mesh of core tiles
Single domain, shared Last-Level cache
DDR5 6400 interface

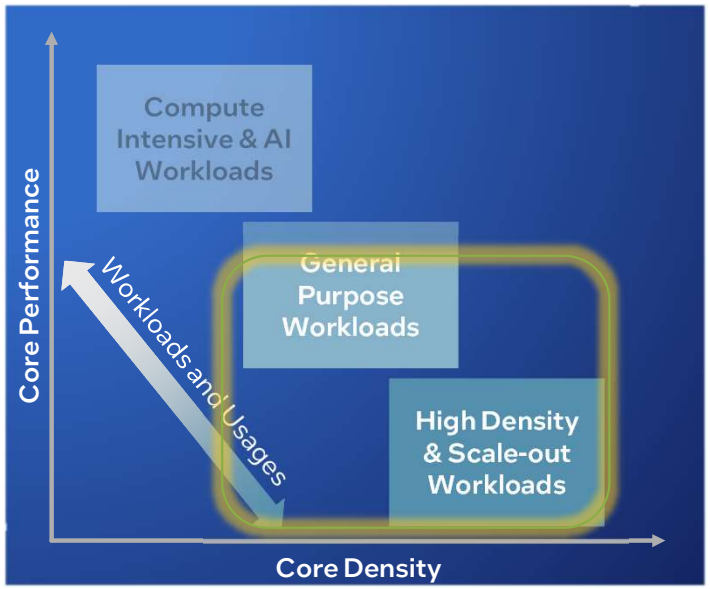
Security Isolation[1]
Intel® Trusted Domain Extensions (Intel TDX)
Intel® Software Guard Extensions (Intel SGX)

Focus on Throughput Performance
Up to 288 cores (2S)
205W and higher power per socket



1 Common capabilities across both Intel Xeon with P-cores and Intel Xeon with E-cores

Harness the Efficiency of Xeon with E-cores



Improve OpEx and CapEx
Increased performance per Watt at increased core density

Augments Xeon breadth of coverage
Same hardware, software and firmware

Vector and AI Instruction Support
Including support for FP16, BF16 and INT8 data types

Focus on throughput, density and efficiency
Optimize across full range of utilization

intel XEON

Highly Scalable Architecture To Address the Throughput Efficiency Needs For the Next Decade of Compute

Thank You

References

TDX https://www.intel.com/content/www/us/en/developer/articles/technical/intel-trust-domain-extensions.html

VT-rp https://www.intel.com/content/dam/www/central-libraries/us/en/documents/intel-virtualization-technologies-white-paper.pdf

CET https://www.intel.com/content/www/us/en/developer/articles/technical/technical-look-control-flow-enforcement-technology.html

VNNI https://community.intel.com/t5/Blogs/Tech-Innovation/Artificial-Intelligence-AI/Deep-Learning-Performance-Boost-by-Intel-VNNI/post/1335670

CXL https://www.computeexpresslink.org/

New instruction set extensions
https://www.intel.com/content/www/us/en/developer/articles/technical/intel-sdm.html#inpage-nav-4


intel