# Exploring the Potential of ARM Processors: Evaluating LS-DYNA Performance for Cloud-Based High-Performance Computing

Eric Day

Ansys LST

## 1 Abstract

In the realm of high-performance computing (HPC), x86_64 architecture has traditionally dominated, driven by its robust performance and extensive software support. However, recent benchmarks indicate the emerging viability of ARM processors for compute-intensive workloads, particularly when running LS-DYNA software. This study explores the performance of LS-DYNA on ARM-based chips, specifically evaluating its effectiveness on Amazon Graviton in the HPC cloud environment and Apple M, Cavium ThunderX2, Ampere Altra, Fujitsu A64FX and Amazon Graviton in standalone computing. Power efficiency, high throughput, cost-effectiveness, and scalability position ARM processors as compelling options for cloud-based LS-DYNA computations.

## 2 Introduction

In the ever-evolving landscape of computing, processor architectures play a crucial role in determining the performance and efficiency of various computing systems. This paper delves into the exploration of ARM processors and their potential in cloud-based high-performance computing (HPC) environments. We aim to evaluate the advantages and drawbacks of ARM-based systems, comparing them to the more dominant x86_64 architecture found in traditional datacenters.

### 2.1 The Clash of Architectures: CISC vs. RISC

Complex instruction set computer (CISC) processors, exemplified by the x86_64 architecture, boast an extensive instruction set, enabling them to perform complex tasks like multiplication or memory manipulation in a single cycle [1]. Consequently, CISC processors have become pivotal in PCs and servers, providing substantial computing power. On the other hand, reduced instruction set computer (RISC) processors, such as those based on ARM (ARM64) architecture, embrace a minimalist approach, with only base-level instructions [1]. Although RISC instructions might require multiple cycles to accomplish what a CISC instruction does in one, the simplicity of RISC architecture results in lower power consumption. It is vital to examine how these fundamental differences in architecture impact performance and efficiency in cloud based HPC scenarios.

### 2.2 The Dominance of x86_64 in Datacenters

The remarkable raw performance and computational power of x86_64 processors have positioned them as dominant players in the desktop and server markets. Their longstanding presence has also fostered the development of well-established optimizations, mature compilers, and a vast library of scientific software tailored specifically for x86 architecture. This robust software ecosystem makes x86_64 a reliable choice for applications heavily reliant on single and multi-threaded performance or those requiring specific software optimizations. However, as we shall discuss, this very focus on performance has led to some trade-offs in terms of energy efficiency and resource utilization.

### 2.3 Evolving CPU Design Philosophy: From Clock Rates to Parallelism

The pursuit of increased performance by chip OEMs saw a phase of escalating CPU clock rates. However, physical limitations, including excessive heat generation and power consumption, forced a shift towards enhancing performance through parallelism, incorporating multi-core designs. Modern CPUs now feature multiple cores operating at lower clock speeds, leading to more efficient performance in multi-threaded applications. In recent generations, x86_64 processors have seen an influx of cores, such as 60 cores in Intel Sapphire Rapids and 96 cores in AMD Genoa CPUs. Nevertheless, this exponential increase in core counts may not always translate to commensurate gains in communication and efficiency, potentially making x86_64 architecture less suitable for HPC in power-hungry datacenter environments.

### 2.4 The Rise of ARM Processors in HPC and Datacenters

ARM (ARM64) architecture brings several compelling advantages to the table. Its renowned power efficiency is particularly attractive for HPC clusters that consume massive amounts of power. The simplified RISC design requires fewer transistors, leading to more power-efficient and compact packages. This efficiency is well-aligned with the ongoing global efforts to reduce datacenter energy consumption. Notably, companies like NVIDIA with their ARM-based Grace CPU, Fujitsu with their A64FX CPU, Ampere with their Altra CPU, and Amazon with their Graviton CPU have strategically designed ARM-based processors for high-performance computing, offering competitive performance per watt. Embracing ARM processors in datacenters could lead to lower energy costs, improved power density, and a smaller datacenter footprint.

### 2.5 Emphasizing Parallel Processing and High Throughput

ARM processors shine in parallel processing. The power conscious core design allows for the deployment of many cores without excessive heat generation. This characteristic of ARM mitigates thermal-induced performance degradation that can occur with extensive core usage. ARM's advanced SIMD instructions and often excellent memory bandwidth also contribute to the efficient parallel processing of large datasets. Moreover, ARM-based architectures can leverage specialized accelerators, such as GPUs or FPGAs, to further enhance performance in specific tasks, making them a scalable choice for HPC workloads.

### 2.6 Cost-Effectiveness of ARM-based Systems

ARM processors present an advantage in terms of cost-effectiveness. Their simpler architecture and licensing model makes them less expensive to produce compared to x86 processors. Additionally, the reduced chip area and power consumption results in higher energy density, ultimately leading to potential cost savings in datacenter operations. The combination of enhanced performance per watt and performance per datacenter area can contribute to more economical HPC deployments.

### 2.7 Acknowledging the Dominance of x86_64 and Emerging Sentiment around ARM

Despite the advantages of ARM processors, the advanced software ecosystem and performance optimization of x86_64 processors have helped maintain their prominence in datacenter deployments. However, there is a growing sentiment favoring ARM for the reasons outlined above. As ARM processor performance continues to improve, its presence in datacenters is becoming increasingly relevant and compelling.

### 2.8 Sneak Peek

This paper endeavors to explore the potential of ARM processors in cloud-based high-performance computing, assessing their performance, scaling, efficiency, and cost-effectiveness relative to the dominant x86_64 architecture. As HPC deployments strive to maximize performance per watt and reduce energy consumption, ARM processors hold promise in offering power-efficient solutions without compromising computational capabilities. Embracing the ARM architecture may mark a transformative step towards achieving more sustainable and efficient datacenter operations. However, it is essential to recognize the continued strength of x86_64 systems, and we shall delve into further detail to compare these architectures in the following sections.

## 3 Assessed ARM Processors

This section presents the specifications of the ARM processors used in this paper's LS-DYNA benchmarks. Each processor outlined below is founded on the ARMv8 architecture. Table 1 displays the specifications of the CPU models benchmarked.

| Processor | Year Released | Cores, Frequency | Architecture, Microarchitecture | SIMD | Tech | Memory | Memory BW |
|---|---|---|---|---|---|---|---|
| Cavium ThunderX2 CN9975 | 2018 | 28 @2.4GHz | ARMv8.1-a, Vulcan | 128bit Neon | 16nm | 8xDDR4 -2666 | 170GB/s |
| Fujitsu A64FX | 2019 | 48 @2.0GHz | Armv8.2-a, A64FX | SVE | 7nm | | 1,024GB/s |
| Ampere Altra Q64-30 | 2021 | 64 @3.0GHz | ARMv8.2-a, Neoverse-N1 | 2x128bit Neon | 7nm | 8xDDR4 -3200 | 204GB/s |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Apple M1 Max | 2021 | 8 @3.2GHz + 2 efficiency | ARM 8.5-a, Firestorm | 128bit Neon | 5nm | 16xLP DDR5-6400 | 408GB/s |
| AWS Graviton2 | 2020 | 64 @2.5GHz | ARMv8.2-a, Neoverse-N1 | 2x128bit Neon | 7nm | 8xDDR4 -3200 | 204GB/s |
| AWS Graviton3 | 2022 | 64 @2.6GHz | ARMv8.4-a, Neoverse-V1 | 2x128bit Neon and 2xSVE | 5nm | 8xDDR5 | 300GB/s |

*Table 1: ARM64 Processor Specifications*

### 3.1 Cavium ThunderX2

The ThunderX2 that was benchmarked is based on the Vulcan architecture. It is characterized by two sockets with 28 cores and 112 threads operating at up to 2.4 GHz [10]. Blending power efficiency and performance, this model employs 16nm technology, incorporates eight DDR4 memory channels, and integrates the 128-bit Neon extension, an advanced SIMD ISA by ARM [10]. Introduced by Cavium in 2018, the ThunderX2 aimed to make strides in the server market.

### 3.2 Fujitsu A64FX

The FX700, introduced in 2019, harnesses the Fujitsu A64FX chip based on the ARMv8.2-a architecture [13]. Crafted with 7nm technology, it boasts 48 cores clocked at 2.0 GHz, Scalable Vector Extension (SVE) SIMD, and a staggering 1TBps bandwidth [13], focusing on optimizing parrallel processing and efficient vector processing. The Tofu interconnect technology enhances communication between nodes [13], making it a contender for high-performance computing applications.

### 3.3 Ampere Altra

The Ampere Altra Q64-30, assessed through an Azure Standard_D64pls_v5 VM, emerged in 2021. The chip was based on TSMC's 7nm platform and featured 64 Neoverse-N1 cores, reaching 3.0 GHz [11]. Prioritizing scalability and performance, it wields eight DDR4 memory channels per socket, facilitating robust data throughput [11]. This Azure VM series with Ampere Altra stands out for its cost-effectiveness within the general-purpose Azure Virtual Machine portfolio.

### 3.4 Apple M1 Max

The Apple M1 Max, based on ARMv8.5 architecture, houses a 10-core CPU with 8 Firestorm performance cores clocked at up to 3.2 GHz and 2 efficiency Icestorm cores [12]. Unveiled in 2021 and fabricated on the TSMC 5nm platform, it flaunts a 408 GBps bandwidth courtesy of its sixteen DDR5 memory channels [12].

### 3.5 AWS Graviton

Amazon Graviton chips, designed by AWS, are optimized to provide excellent price-performance ratios for cloud workloads [3]. The first-generation Amazon Graviton chip, introduced in 2018, featured 16 Cortex A72 physical cores running at 2.3GHz [2]. While not exceptionally powerful, it marked the initial step in advancing the Graviton platform.

### 3.6 AWS Graviton2

Graviton2, launched in December 2019, boasted 64 Neoverse N1 physical cores operating at 2.5GHz [2]. Like Ampere Altra, there is an entire physical core for each vCPU. This chip was equipped with large L1 and L2 caches for each virtual CPU, along with the Neoverse N1 mesh interconnect, enabling low-latency and bandwidth-efficient performance [7]. Graviton2 also incorporated 2x 128-bit Neon, further enhancing its capabilities [2]. As a result, Graviton2-based instances exhibited up to 40% better price-performance compared to fifth-generation Amazon instances [7].

### 3.7 AWS Graviton3

While ARM's Neoverse N-series focuses on power efficiency and performance per unit area, Neoverse V-series prioritizes maximum performance even at the expense of power and space. At equivalent frequencies, the V1 platform demonstrates a 50% increase in instructions per cycle (IPC) compared to the N1 platform [4]. Amazon Graviton3 aims to push ARM server performance by leveraging the V1 platform.

Launched in 2022, Graviton3 employs 64 Neoverse V1 cores running at 2.6GHz and includes 4x 128-bit Neon vectors and 2xSVE 256-bit [2], providing 25% better compute performance over Graviton2 [6].

The increased IPC is largely due to the increased number of transistors, growing from 30 billion in Graviton2 to 55 billion in Graviton3 [5]. Graviton3 also features DDR5 memory, providing 50% more bandwidth over DDR4 on Graviton2 [2]. The 5nm design looks to offset the additional power consumption drawn from the larger V1 core and added transistors over the 7nm Graviton2.

### 3.8 AWS Graviton3E

For high-performance computing (HPC) applications, AWS Graviton3E processors offer even higher vector-instruction performance than AWS Graviton3 processors, achieving up to a 35% improvement [6]. These chips are utilized in instances like C7gn and Hpc7g, offering up to 200Gbps of network bandwidth [6], catering to HPC workloads with exceptional performance benefits.

## 4 Neon

The Neon benchmark model features a Dodge Neon frontal crash with a rigid wall. The termination time was set to 30ms. Neon has 500,000 elements. The modest size of the model makes it well-suited for benchmarking standalone computers. Table 2 displays the machine specifications used to benchmark each processor as well as the armflang mcpu flag used in compiling the LS-DYNA binary.

| Processor | Machine | Cores, Frequency | Sockets | Architecture, Microarchitecture | Recommended mcpu flag |
|---|---|---|---|---|---|
| Cavium ThunderX2 CN9975 | on-prem | 28 @2.4GHz | 2 | ARMv8.1-a, Vulcan | thunderx2t99 |
| Fujitsu A64FX | FX700 | 48 @2.0GHz | 1 | Armv8.2-a, A64FX | a64fx, neoverse-512tvb |
| Ampere Altra Q64-30 | Azure D64pls_v5 | 64 @3.0GHz | 1 | ARMv8.2-a, Neoverse-N1 | neoverse-n1 |
| Apple M1 Max | MacBook Pro | 8 @3.2GHz + 2 efficiency | 1 | ARM 8.5-a, Firestorm | neoverse-n1 |
| AWS Graviton2 | AWS c6gn.16xlarge | 64 @2.5GHz | 1 | ARMv8.2-a, Neoverse-N1 | neoverse-n1 |
| AWS Graviton3 | AWS c7gn.16xlarge | 64 @2.6GHz | 1 | ARMv8.4-a, Neo-verse-V1 | neoverse-512tvb |
| Intel 8375C Ice Lake | AWS c6i.32xlarge | 32 @2.9GHz | 2 | x86_64, Ice Lake | |
| AMD 7R13 Milan | AWS c6a.48xlarge | 48 @2.65GHz | 2 | x86_64, Zen3 | |
| AMD 9654 Genoa | on-prem | 96 @2.4GHz | 2 | x86_64, Zen 4 | |

*Table 2: Machine Specifications*

### 4.1 Performance and Scaling – Neon

To create a more balanced comparison between different machine sizes, each Neon benchmark utilized cores from a single socket and NUMA node. This approach aimed to create a fair evaluation between larger and smaller systems. The ARM processors were pitted against Intel Ice Lake, AMD Milan, and AMD Genoa processors, with testing conducted across 1, 2, 4, and 8 cores. For these assessments, LS-DYNA MPP single precision from the LS-DYNA development source was employed, utilizing Open MPI 4.x. LS-DYNA for ARM processors was compiled using gcc and armflang22.0.2, employing the recommended mcpu flag tailored to the specific chip. Meanwhile, LS-DYNA for x86_64 processors was compiled using gcc and ifort190 with AVX2 instruction set. Performance below is based on total elapsed time.

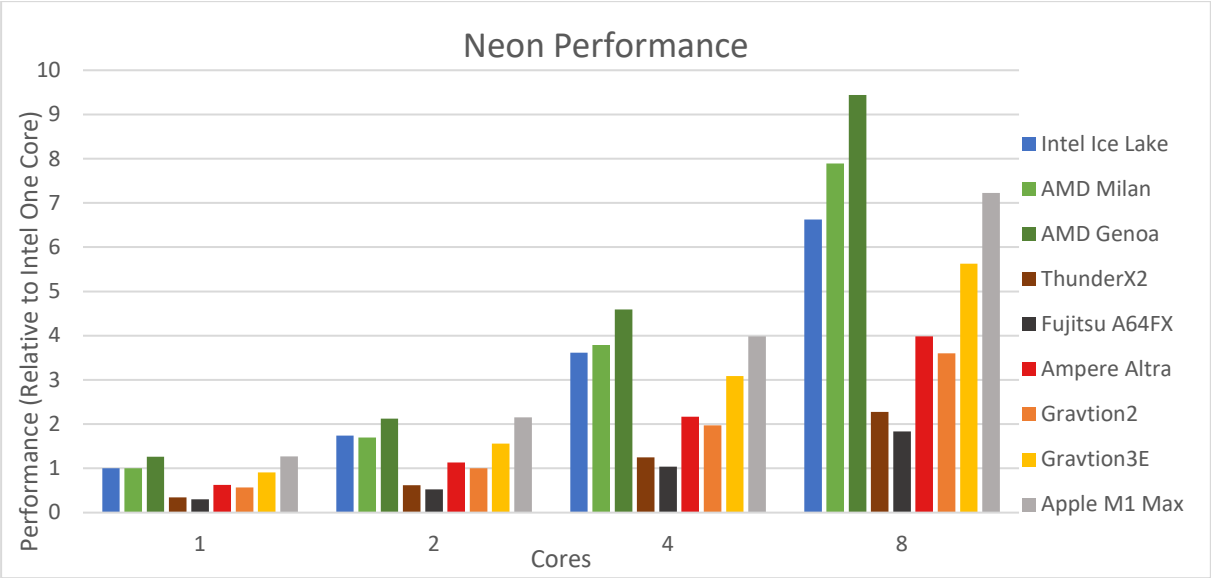


*Fig.1: LS-DYNA Trunk Standalone Computer Relative Performance – 500,000 Elements*

While older-generation ARM processors, ThunderX2 and A64FX, might not match the compute performance of contemporary x86_64 processors, the Apple M1 Max demonstrates remarkable resilience at both 1 and 2 cores, even outdueling the top offering from AMD. Graviton3E, in comparison, exhibits respectable performance when pitted against Intel Ice Lake. Moving to 8 cores, Apple M1 Max maintains its strong performance, continuing to beat out Intel, but succumbing to the superior scaling of AMD in this configuration.


*Fig.2: LS-DYNA Trunk Standalone Computer Speedup – 500,000 Elements*

Graviton2 and Ampere Altra 1st generation Q64-30 are remarkably similar chips. They share the same N1 microarchitecture, 7nm technology, quantity of cores, 2x128 Neon, RAM, memory bandwidth, and system level cache. Ampere, however, has a larger clock at 3.0GHz, compared to 2.5GHz in Graviton2. The identical memory characteristics are evident in the near identical scaling.

AMD demonstrates the most pronounced scaling, likely benefiting from the ample L3 cache of Milan and Genoa, which likely fits all of Neon. Despite Apple M1 Max boasting one of the highest memory bandwidths among the tested processors, its speedup falls behind. From testing, Neoverse-N1 was the

best performing armflang mcpu or march flag for Apple, but still may not be the optimal choice for M1 Max. Submitting the LS-DYNA jobs on Apple via a Linux emulator, among other factors, may have also contributed to the suboptimal scaling.

The Neon benchmark reveals that contemporary ARM processors, particularly Apple M1 Max and Graviton3E, deliver competitive LS-DYNA performance on standalone systems. Notably, Apple achieves exceptional single and double-core performance. However, ARM's computational prowess may still lag behind the latest 4th generation EPYC and Xeon processors. Subsequent benchmarks will delve into the assessment of a more substantial model, full CPU utilization, and HPC performance of ARM processors.

## 5 ODB-10M

ODB-10M features a refined Ford Taurus model crash with an LSTC shell ODB barrier. ODB-10M has 10 million elements. The termination time was set to 50ms.

### 5.1 Performance and Scaling – ODB-10M

This study focuses on the HPC capabilities of ARM processors on cloud. Graviton2, Graviton3E, Intel Ice Lake, and AMD Milan are evaluated up to 8 nodes on AWS. One node of Ampere Altra was available for testing on Azure. LS-DYNA R12.1 MPP single precision with Open MPI 4.x was employed. Multi-threading was disabled on x86_64 instances. Elastic Fabric Adapter (EFA) network interface was utilized on Amazon instances. Figure 3 displays the elapsed time on clusters composed of each of the listed instances.

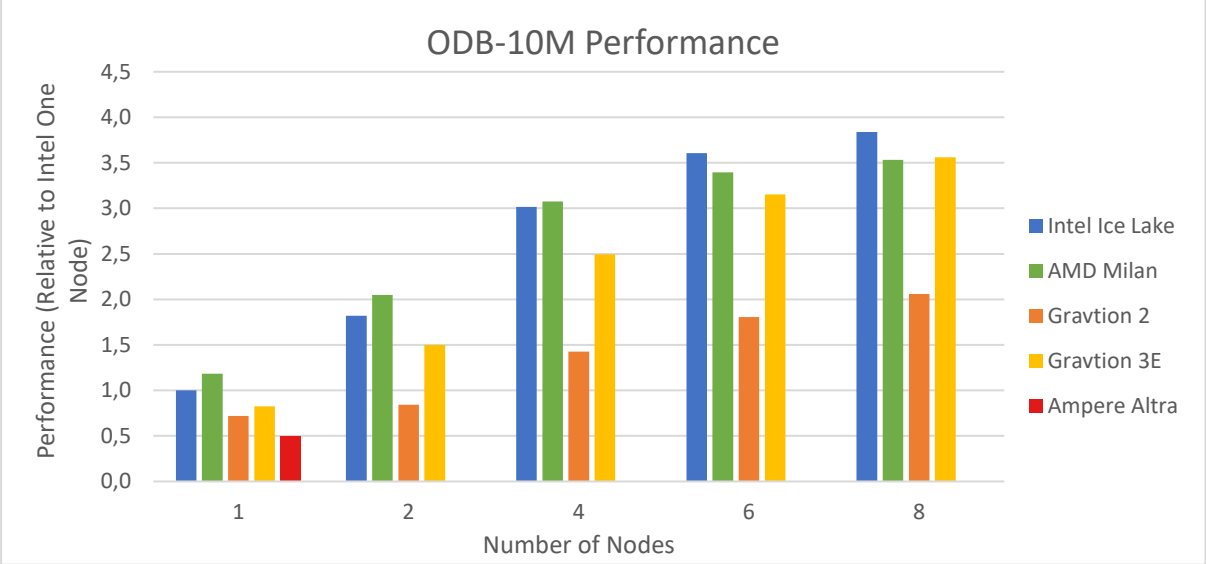


*Fig.3: LS-DYNA R12.1 Multi-Node Relative Performance – 10 million Elements*

Despite its increased clock speed, Ampere's single-node performance fell behind that of Graviton2 by 27%.

Graviton3E showed a fairly minor 12% improvement in single-node performance over Graviton2. Nonetheless, the enhanced vectorization of Graviton3E may have been compounded across supplemental nodes. Graviton3E averaged 43% greater performance than Graviton2 in clusters sized 2-8 nodes.

AMD's single-node performance outshined the rest, surpassing Graviton3E by 43%. However, as the cluster size expanded, AMD's edge began to dull. At 8 nodes, Graviton3E performance equaled that of AMD's. The core count advantage in the AMD instance was evident at lower node counts, but its scaling is hindered by limited memory bandwidth per core and network bandwidth per core, which can lead to delayed memory access and communication between nodes. This scaling challenge is pronounced in the speedup, as illustrated in Figure 4.

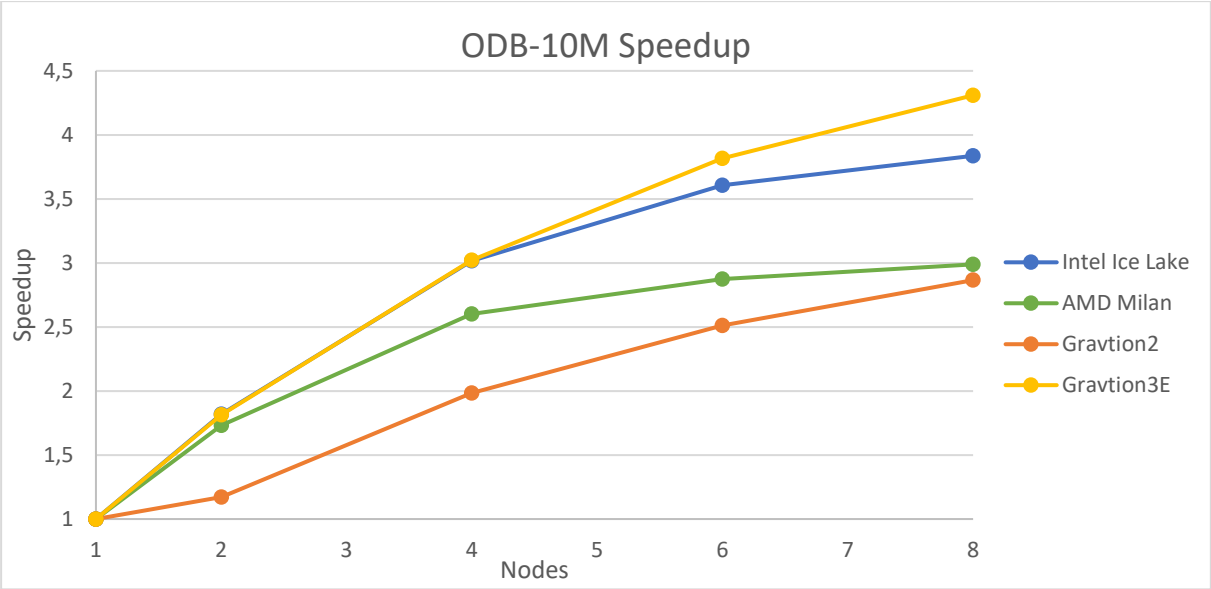


*Fig.4: LS-DYNA R12.1 Speedup – 10 million Elements*

Graviton3E demonstrates great speedup. The 50% increase in memory bandwidth from Graviton2 to Graviton3E is vividly displayed in the Figure.

Both AMD and Intel instances could also be experiencing latency due to their multi-socket configurations. LS-DYNA may need to access NUMA memory on these machines, whereas all memory is local on each Graviton node with its single NUMA domain. Latency from remote memory access could be compounded across multiple nodes of the x86_64 clusters. However, it's important to note that the benchmarked Graviton instances feature an impressive network bandwidth of 200 Gbps, in contrast to x86_64 instances which are limited to 50 Gbps. This discrepancy in network bandwidth could potentially lead to slower remote access between nodes.

### 5.2 Power Consumption – ODB-10M

The Thermal Design Power (TDP) is the maximum power that a chip is designed to dissipate under full load. It is often considered an approximation of the power consumption when the CPU is running at or near 100% usage. Notably, the TDP for Intel Ice Lake 8375C stands at 300W, while AMD EPYC 7R13's TDP is noted at 225W [3]. Both AMD and Intel instances adopt dual-socket configurations. Ampere Altra's Q64-30 TDP is revealed as 180W [14].

Regrettably, Amazon has not disclosed the TDP for their Graviton processors, but we can make an informed estimate. As previously mentioned, Graviton2 and Ampere Altra 1st generation 'QuickSilver' chips are remarkably similar. Although the Q64-30 benchmarked in this paper operates at a higher frequency, Ampere chips in the same lineup have more comparable frequencies to Graviton2. For instance, the Ampere Q64-24, clocking at 2.4GHz, carries a TDP of 95W, while the Ampere Q64-26, with a 2.6GHz frequency, carries a TDP of 125W [14]. Given the Graviton2's 2.5GHz frequency, a reasonable assumption places its TDP around 110W. Frumusanu from AnandTech estimates Graviton2's TDP is between 110-130W [15]. We can use the conservative figure of 130W.

For Graviton3, we need to make another informed estimate. ARM states that the V1 platform's power efficiency ranges between 0.7x to 1x that of N1 [4], where power efficiency is the quotient of the performance increase and the power increase. Given the increased IPC of 1.5x and the slight increase in clock speed, assuming Graviton2 TDP is 130W, I arrive at an estimate of 210W-295W for Graviton3E TDP. We can use the conservative figure of 295W.

Figure 5 showcases the comparative performance, estimated power consumption, and performance per watt of the instances, both in single node and eight node clusters. The metrics from each cluster are normalized to Intel Ice Lake single node.

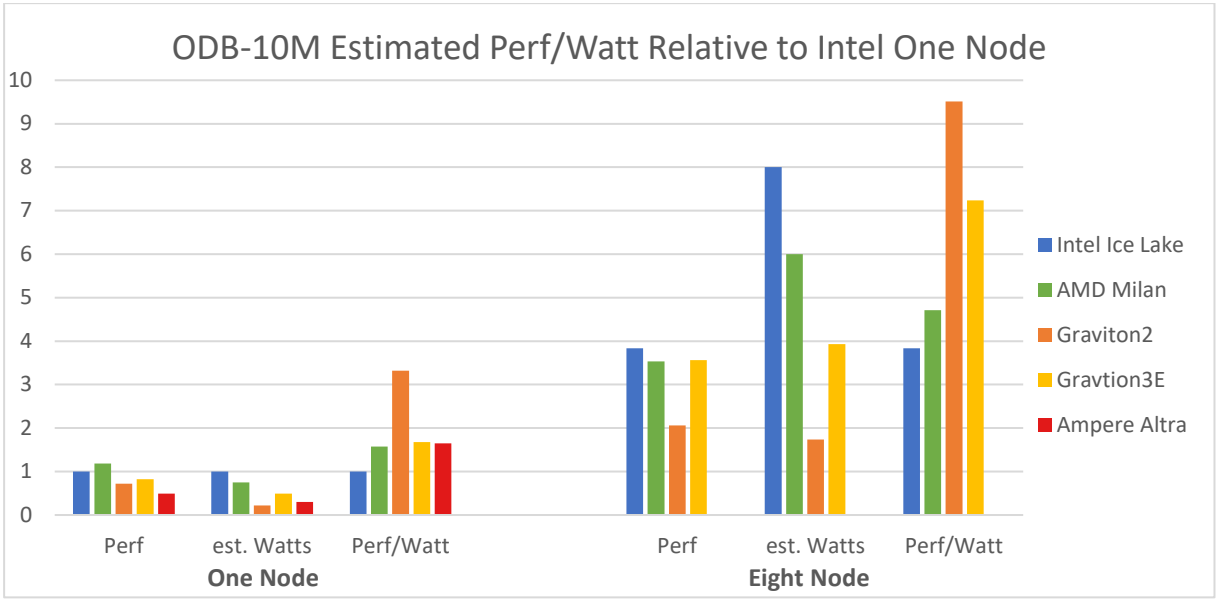


*Fig.5: LS-DYNA R12.1 Performance per Watt – 10 million Elements*

The reduced clock speed of AMD Milan could contribute to its lower power consumption in comparison to Intel's offerings. Nevertheless, the performance per watt of both x86_64 CPUs falls behind that of Graviton2 with its remarkable efficiency. Graviton2 sets an impressive single node benchmark, with a performance per watt 3.3 times greater than that of Intel, while Graviton3E's performance per watt is 1.7x that of Intel. Ampere Altra and Graviton3E, although using different ARM microarchitectures, arrive near the same performance per watt in the single node cluster scenario.

As the cluster expands, the enhanced IPC and scalability of Graviton3E greatly benefits its performance in an attempt to outweigh the favorable power consumption of Graviton2. However, in the context of the eight-node cluster, Graviton2 continues to assert its dominance, showcasing a performance per watt 2.5 times higher than that of Intel. Meanwhile, Graviton3E demonstrates its strength with a performance per watt 1.9 times greater than Intel's.

Graviton3E's respectable performance and efficiency matched with its impressive scaling allows its performance per watt advantage to continually grow over x86_64 processors as the cluster increases in size. Graviton2's shear efficiency makes its performance per watt difficult to match at any scale.

### 5.3 Price Performance – ODB-10M

Datacenters are increasingly prioritizing performance per watt to not only reduce their environmental impact but to optimize energy costs. ARM chips hold a price advantage over x86_64 in this regard due to their simplified architecture and licensing model, resulting in more cost-effective production and operation. These efficiencies in datacenter operations can be translated into cost savings for customers running workloads. For instance, at the time of testing, the on-demand Linux pricing for Intel C6i.32xlarge and AMD C6a.48xlarge instances stands at $5.44/hr and $7.344/hr, while the ARM-based counterparts, C6gn.16xlarge and C7gn.16xlarge, offer more economical rates at $2.77/hr and $3.99/hr, respectively.

Figure 6 presents the normalized price-performance, or performance per cost, for each cluster running ODB-10M, normalized against Intel's single node price-performance. The price referenced is the cost of running the entire cluster for the duration of the job.

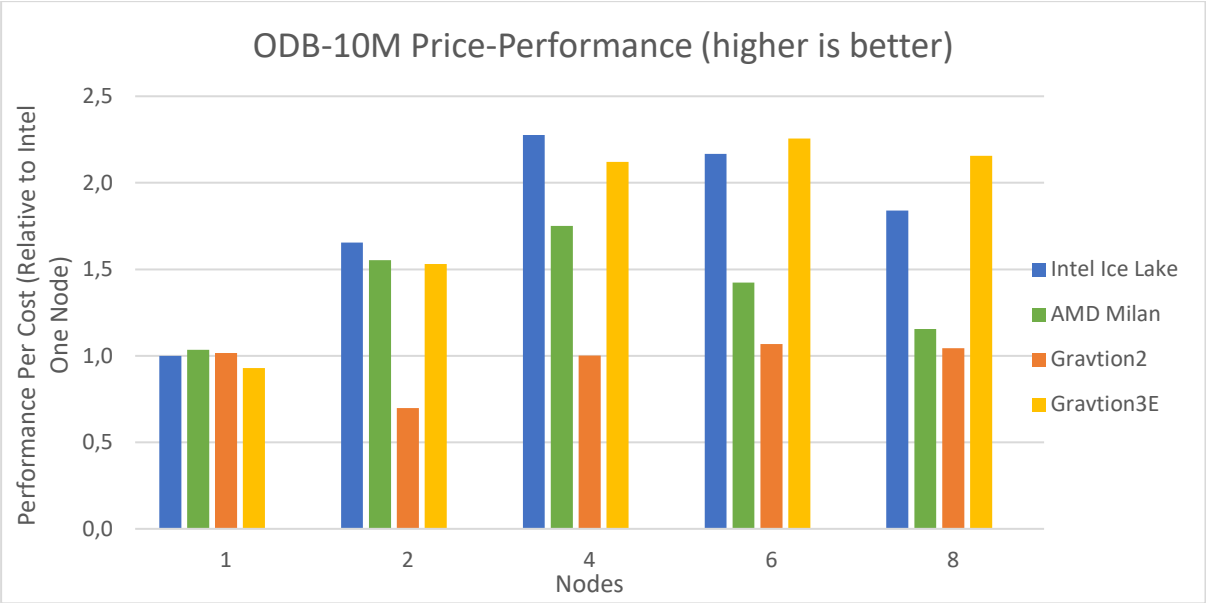


*Fig.6: LS-DYNA R12.1 Price Performance – 10 million Elements*

A high price-performance instance offers the greatest computational capacity for every dollar invested. This can lead to increased productivity, with the potential to achieve more job iterations per day within budget constraints.

At one node, the price-performance is fairly equal across the board. It is a different story for larger sized clusters.

Intel achieves the highest LS-DYNA price-performance at four nodes, while Graviton3E also shines at six nodes. Accordingly, both x86_64 instances hit their price-performance peak at four nodes, whereas ARM64 instances hit their stride at six nodes.

While the ARM N1 platform prioritizes area and power efficiency, the V1 platform emphasizes computational prowess. Amazon's focus for Graviton2 therefore centers on performance per watt, while Graviton3's centers on enhancing price-performance for compute-intensive workloads [6]. The scalability and attractive pricing of Graviton3E positions it as a compelling choice for larger cluster configurations.

## 6 Future ARM Innovations

The paper's evaluation of ARM processor performance shows great promise, yet the realm of ARM innovations holds even more exciting prospects on the horizon.

One notable development is Ampere's introduction of the AmpereOne CPU, which capitalizes on the ARMv8.6 architecture and boasts an impressive 192 cores [16]. AmpereOne surpasses AMD and Intel's 4th generation processors by 2.9x and 4.3x, respectively, in terms of VMs per rack [16]. This leap in compute density could directly translate to more environmentally friendly datacenters.

Equally captivating is Nvidia's ARM64 Grace CPU, built on the ARM Neoverse V2 platform. Boasting an impressive 144 cores, Grace Hopper with its 1 TBps bandwidth, offers a CPU+GPU coherent memory model for AI and HPC applications [8]. The Grace CPU has already found deployment in Isambard 3, establishing itself as one of the world's most energy-efficient non-accelerated supercomputers [9].

The momentum behind ARM processors ensures that we can confidently anticipate the emergence of Amazon Graviton4 and Apple M3 processors. Rumored to leverage TSMC's cutting-edge 3nm technology, these processors promise substantial power savings and heightened efficiency.

As the software ecosystem continues to evolve in tandem with chip technology advancements, further speedup and efficiency gains are within reach. Notably, OpenMPI 5.0 is poised to optimize parallel performance for ARM architectures. Chip OEMs are diligently expanding their software libraries to align with the unique capabilities of their ARM processors. Furthermore, tuning of LS-DYNA for specific ARM hardware is already underway. Collectively, these developments paint a promising future characterized by elevated performance, efficiency, and compatibility for ARM processors and LS-DYNA.

## 7 Summary

In the evolving landscape of datacenters, the balance between cost considerations and environmental concerns is paramount. ARM architecture presents a promising alternative to the longstanding x86_64 dominance in HPC deployment. While RISC processors historically posed limitations on compute-based performance, rapid architectural advancements, an expanding software ecosystem, and chips tailored for HPC, are progressively bolstering ARM's viability in datacenters. LS-DYNA benchmarks showcase ARM64's competitive price-performance, outstanding performance-per-watt, and remarkable scalability with Amazon Graviton. As ARM processors continue to advance and gather momentum, the potential for a substantial shift towards more streamlined and eco-friendly datacenter operations becomes increasingly tangible.

## 8 Literature


[1] “ARM vs. x86: What’s the difference?”, Red Hat, 2022. Available online: https://www.redhat.com/en/topics/linux/ARM-vs-x86

[2] “AWS Technical Guide”, aws-graviton-getting-started, 2023. Available online: https://github.com/aws/aws-graviton-getting-started/blob/main/README.md

[3] “AMD EPYC 7R13 vs Intel Xeon Platinum 8375C”, GadgetVersus. Available online: https://gadgetversus.com/processor/amd-epyc-7r13-vs-intel-xeon-platinum-8375c/

[4] Frumusanu, A: “ARM Announces Neoverse V1, N2 Platforms & CPU’s, CMN-700 Mesh: More Performance, More Cores, More Flexibility”, Anandtech, 2021. Available online: https://www.anandtech.com/show/16640/ARM-announces-neoverse-v1-n2-platforms-cpus-cmn700-mesh

[5] Saidi, A: “Deep Dive Into AWS Graviton3 and Amazon EC2 C7g Instances”, AWS re:Invent 2021. Available online: https://www.youtube.com/watch?v=WDKwwFQKfSI

[6] “Amazon EC2 C7g Instances”, Amazon, 2023. Available online: https://aws.amazon.com/ec2/instance-types/c7g/

[7] Raman, S: “Deep Dive On AWS Graviton2 Processor-powered EC2 Instances”, AWS re:Invent 2020. Available online: https://www.youtube.com/watch?v=NLysl0QvqXU

[8] “NVIDIA Grace CPU” , NVIDIA, 2023. Available online: https://www.nvidia.com/en-us/data-center/grace-cpu/

[9] “NVIDIA Grace Drives Wave of New Energy-Efficient ARM Supercomputers”, NVIDIA, 2023. Available online: https://nvidianews.nvidia.com/news/nvidia-grace-drives-wave-of-new-energy-efficient-ARM-supercomputers

[10] “ThunderX2 – Cavium”, WikiChip, 2019. Available online: https://en.wikichip.org/wiki/cavium/thunderx2

[11] “Ampere Altra Family Product Brief”, Ampere, 2023. Available online: https://amperecomputing.com/briefs/ampere-altra-family-product-brief

[12] Frumusanu, A: “Apple’s M1 Pro, M1 Max SoCs Investigated: New Perfromance and Efficiency Heights”, Anandtech, 2021. Available online: https://www.anandtech.com/show/17024/apple-m1-max-performance-review

[13] “PRIMEHPC Specifications”, Fujitsu. Available online: https://www.fujitsu.com/global/products/computing/servers/supercomputer/specifications/

[14] Cutress, I: “Ampere Altra 1P Server Pictured: GIGABYTE’s 2U with 80 Arm N1 Cores, PCIe 4.0 and CCIX”, Anandtech, 2020. Available online: https://www.anandtech.com/show/15949/ampere-altra-1p-server-pictured-gigabytes-2u-with-80-ARM-n1-cores-pcie-40-and-ccix

[15] Frumusanu, A: “Hot Chips 202: Marvell Details ThunderX3 CPUs – Up to 60 cores Per Due, 96 Dual-Die in 2021” Anandtech, 2020. Available online: https://www.anandtech.com/show/15995/hot-chips-2020-marvell-details-thunderx3

[16] Shilov, A: “Smpere Unveils 192-Core CPU, Controversial Benchmarks”, Tom’s Hardware, 2023. Available Online: https://www.tomshardware.com/news/ampere-unveils-192-core-cpu