NAFEMS World Congress 2025 - Accelerating Scientific Workflows with Domain-Specific Hardware: GPUs, Arm Chips and Beyond

Dr Sam Zakrzewski PhD
Rescale (Denmark)

Romain Klein
Rescale (France)

Abstract

The demand for high-performance computing (HPC) is expanding as scientific and engineering challenges grow increasingly complex. Traditional CPU-based architectures, while versatile, often struggle to efficiently handle specialised workloads such as machine learning, computational fluid dynamics (CFD), and molecular dynamics. To address this, the integration of domain-specific hardware accelerators like NVIDIA GPUs and Arm chips has emerged as a game-changer, enabling unparalleled performance and efficiency for targeted applications.

This paper delves into the role of domain-specific hardware accelerators in revolutionising scientific workflows. We focus on how specialised architectures, available on Rescale's intelligent cloud HPC platform, empower researchers and engineers to leverage cutting-edge hardware tailored to their workloads. By combining GPUs for compute-intensive tasks with aarch64 architectures for energy-efficient operations, users can achieve optimal performance while addressing cost and sustainability goals.

A critical aspect of this discussion involves the optimisation of workflows for hybrid and heterogeneous computing environments. Integrating domain-specific accelerators requires not only hardware availability but also seamless software orchestration to manage data flows, scheduling, and execution. Rescale's platform addresses these challenges by offering a unified environment where users can dynamically select the most suitable hardware configurations based on workload requirements. This flexibility is particularly impactful for industries ranging from aerospace to pharmaceuticals, where precision and efficiency are paramount.

The paper also explores practical use cases, such as leveraging GPUs for AI-driven simulation post-processing or Arm-based chips for low-power, high-throughput scenarios. These examples illustrate how hardware accelerators can drastically reduce time-to-solution and cost, enabling organisations to push the boundaries of innovation. Additionally, we will discuss the importance of workload profiling and benchmarking to ensure optimal hardware utilisation, drawing on real-world insights from Rescale's platform.

Attendees will gain an understanding of the technical and operational considerations involved in adopting domain-specific hardware, from software compatibility to deployment strategies in cloud-

based environments. We will highlight how Rescale simplifies these complexities, allowing users to focus on their core scientific and engineering objectives.

Through this exploration of domain-specific accelerators, we aim to demonstrate how specialised architectures are not just advancing performance but also reshaping how organisations approach HPC, paving the way for new breakthroughs in research and industry.

1. Introduction

High-performance computing is the backbone of modern engineering and scientific research. As the complexity of simulations grows, so does the demand for efficient and scalable compute resources. Traditional CPU-based systems, governed by Moore's Law, have struggled to meet this demand. Specialized architectures, such as GPUs and domain-specific accelerators, offer a solution, delivering exponential performance improvements [1]. Cloud-based platforms have emerged as a pivotal enabler, providing engineers with seamless access to these technologies while optimizing costs and sustainability.

Over the past decades, engineering simulations have undergone a profound transformation. From initial wireframe models to sophisticated, multi-physics simulations, the computational demands have increased manifold. The limitations of traditional HPC infrastructures have spurred the need for innovative solutions. Today, engineers require platforms that not only deliver raw computational power but also adapt to specific workloads and reduce environmental impacts.

Rescale addresses these needs by offering a unified platform that integrates state-of-the-art hardware and intelligent software. By leveraging cloud-based infrastructure, Rescale eliminates the bottlenecks associated with static, on-premises systems. Engineers can dynamically scale resources to match the complexity and urgency of their projects, ensuring that insights are derived faster and more efficiently. This capability is particularly critical in industries such as aerospace, automotive, and life sciences, where time-to-market and precision are paramount.

The integration of AI and machine learning has redefined how simulations are performed. AI-assisted workflows not only enhance accuracy but also automate repetitive tasks, freeing engineers to focus on innovation. Rescale's platform supports these advancements, offering pre-configured environments for deploying AI-driven simulations. This symbiosis of advanced hardware and intelligent algorithms marks a new era in digital engineering.

2. Trends in the Industry

Sustainability and Energy Efficiency

Emerging architectures promise up to five times better energy and performance efficiencies compared to traditional on-premises solutions [2]. This shift aligns with global sustainability goals, encouraging organizations to adopt cloud-based HPC resources that are not only faster but also greener. Companies now prioritize reducing their carbon footprints while optimizing computational workloads, and cloud-based platforms play a crucial role in achieving these objectives.

Collaborative Research and Data Sharing

Approximately 70% of organizations globally engage in collaborative R&D, often bridging academia, suppliers, and service providers. This collaboration ensures a steady exchange of knowledge and innovation across sectors. Platforms like Rescale facilitate seamless data sharing and collaborative workflows, such as multiple users accessing simulation output stored in secure, centralized location, enabling geographically dispersed teams to work together efficiently. The rise of interdisciplinary projects further underscores the importance of robust data-sharing mechanisms, which enhance the development of next-generation products and solutions.

Accelerated Computing and AI-Driven R&D

The performance of GPUs, RISC architectures, and similar accelerators is advancing at a rate double that of traditional x86 CPUs. These technologies, coupled with AI and machine learning, have transformed research and development processes across industries. AI-driven R&D enables faster prototyping, predictive modeling, and optimization, cutting development cycles by up to 50%. This progress not only accelerates time-to-market but also enhances product quality and operational efficiency, ensuring companies stay ahead in competitive markets .

3. Specialized Architectures in Engineering Simulations

Advantages of Domain-Specific Accelerators

Domain-specific hardware accelerators, including GPUs, FPGAs, TPUs, and quantum computing platforms, offer tailored solutions for various simulation needs. For instance, GPU-based solutions excel in parallelized workloads such as computational fluid dynamics (CFD), while high-memory-bandwidth architectures are ideal for finite element analysis (FEA). These accelerators not only enhance computational speeds but also reduce operational costs by optimizing resource allocation. By matching hardware capabilities to specific tasks, organizations can achieve significant gains in productivity and performance [1,2,3].

Case Studies

- **ANSYS Fluent**: Benchmarks show significant performance gains using Nvidia GPUs, enabling faster external flow simulations with millions of cells. This improvement has allowed teams to conduct more iterations in less time, enhancing design accuracy.

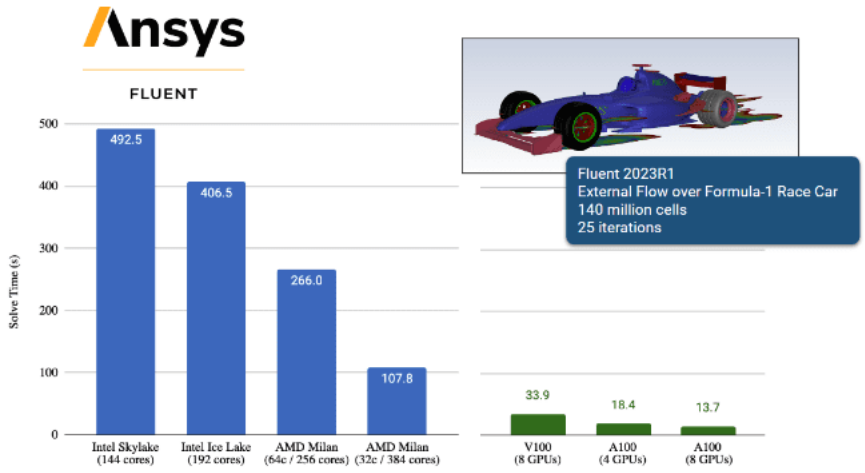

Figure 1 Comparison of CPU and GPUs performance on ANSYS Fluent [5]

The workflow is as follows:

- **Data Preparation:** Preprocesses the geometry and mesh data, then upload it to Rescale's cloud storage.
- **Hardware Choice:** Select a Rescale coretype equipped with Nvidia A100 GPUs, known for their high performance in parallel computing tasks.
- **Simulation Execution:** The CFD solver code utilizes CUDA for parallel processing. The ANSYS Fluent solver is configured to take advantage of the GPU's capabilities.
- **Optimization and Scaling:** Using Rescale's Performance Profile tools, identify performance bottlenecks and optimize the workflow. Additional options to scale the simulation across multiple GPUs to further reduce computation time for a cost-efficient solution

- **OpenFOAM**: Rescale supports aarch64 and x86 architectures, offering flexibility in deployment and optimized performance for specific workloads [3]. OpenFOAM users benefit from Rescale's ability to provide tailored computational environments, streamlining their simulation processes and boosting overall efficiency.

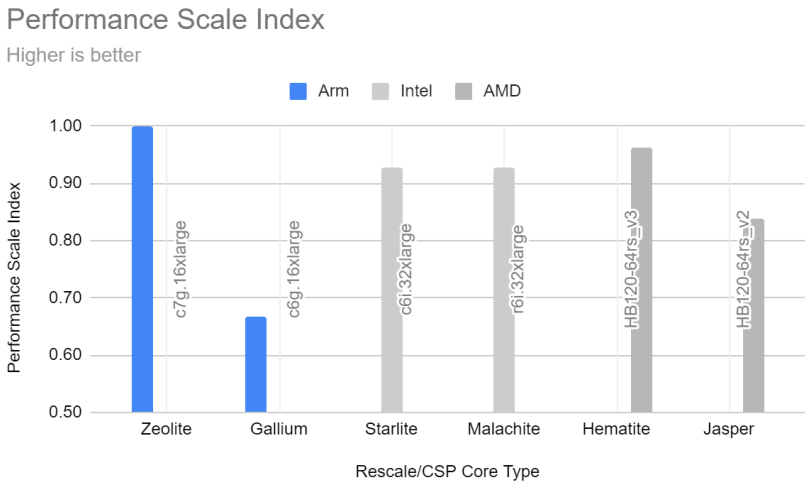

Figure 2 Comparison of aarch64 and x64 architectures on OpenFoam [6]

Here the data shows a favourable comparison between Arm architecture (Zeolite AWS Graviton 3) and x86 based cloud systems

Benchmarks

Hardware benchmarks highlight the performance of the latest accelerators, such as Nvidia's A100 GPUs and Arm's Graviton processors. These benchmarks validate the cloud's ability to deliver best-in-class performance and cost efficiency. For example, simulations conducted on the A100 GPUs demonstrated up to a 4x [4] improvement in processing times, underscoring the advantages of using specialized architectures for intensive engineering tasks.

4. AI-Driven Engineering

Impact of AI/ML

AI/ML has emerged as a transformative force in engineering. From training neural networks on physics data to deploying pre-trained models for predictive simulations, AI accelerates insights and optimizes workflows. These capabilities allow engineers to identify potential issues early in the design process, reducing costly revisions and ensuring products meet stringent quality standards. By integrating AI into simulation workflows, Rescale empowers organizations to transition from reactive to proactive engineering methodologies.

AI Physics Breakthroughs

AI enables computational reducibility, unlocking accurate predictive physics that traditional methods cannot achieve. Pre-trained physics models and accelerated computing architectures further enhance these capabilities, reducing runtime and computational costs. Recent breakthroughs include the development of neural networks capable of solving complex fluid dynamics problems in real time, offering engineers unprecedented speed and precision in their analyses.

Design Optimization

AI integrates seamlessly into simulation workflows, automating tasks such as drag coefficient calculations and design validation. Engineers can rapidly evaluate multiple designs, shortening the design cycle and improving product quality. Additionally, AI-driven optimization tools can explore vast design spaces, identifying configurations that maximize performance while minimizing material usage and costs. This holistic approach to design ensures that products are not only high-performing but also sustainable.

5. Digital Transformation and Ecosystem Integration

Elastic Cloud Workstations

Rescale's Elastic Cloud Workstations offer a powerful, web-based platform for simulation workflows. Engineers can access full application catalogs, deploy secure virtualization solutions, and collaborate with colleagues in real time. These workstations provide the flexibility to pre-process, solve, and post-process simulations from any location, ensuring productivity remains uninterrupted. The integration of cutting-edge hardware and intuitive software enables seamless transitions between different phases of simulation, enhancing overall efficiency.

End-to-End Workflows

The platform completes the digital thread by integrating with tools for product lifecycle management (PLM), workflow orchestration, and schedulers. This integration ensures a cohesive ecosystem for engineering simulations, from data preprocessing to post-processing. Rescale's support for diverse applications and hardware configurations allows teams to customize their workflows, aligning them with specific project requirements [5]. This adaptability is critical in dynamic industries where innovation and agility are key to success.

6. Rescale's Role in HPC

Platform Capabilities

Rescale offers a cloud-based HPC platform that integrates seamlessly with the latest hardware technologies, enabling organizations to harness diverse compute resources on-demand [7]. Engineers can pre-process, solve, and post-process simulations using a web-based GUI, fostering accessibility and collaboration. This approach eliminates the logistical and financial constraints of maintaining on-premises HPC infrastructure, providing engineers with flexibility and scalability. The platform's intuitive interface and performance insights allow users to make data-driven decisions, optimizing resources for maximum efficiency.

Strategic Partnerships

Rescale's collaborations with industry leaders like Nvidia [10] and Arm [9] bring cutting-edge hardware to its users. Nvidia's contributions, including integration with the NGC Catalog and DGX Cloud, enable efficient AI and simulation workloads. Similarly, Arm-powered instances optimize chip design and verification processes, showcasing the versatility of domain-specific hardware. These partnerships ensure that Rescale's platform remains at the forefront of technological advancements, offering users unparalleled performance and cost-efficiency.

7. Conclusion and Future Directions

Cloud-based platforms represent a paradigm shift in HPC for engineering simulations. By enabling access to domain-specific accelerators and AI-driven workflows, They empower organizations to achieve faster, more sustainable, and cost-effective results. As the adoption of specialized architectures continues to grow, these versatile power houses will remain at the forefront, driving innovation and shaping the future of engineering and scientific research. Future developments may include deeper integration with quantum computing platforms, expanded AI capabilities, and enhanced support for interdisciplinary collaborations, ensuring that Rescale continues to meet the evolving needs of its users.

8. References

[1] Nvidia. "Heeding Huang's Law: Video Shows How Engineers Keep the Speedups Coming." https://blogs.nvidia.com/blog/huangs-law-dally-hot-chips/

[2] ANSYS. "Volvo Cars Leverages Ansys and NVIDIA GPUs to Accelerate CFD Simulations by 2.5x for the EX90 Electric Vehicle" https://www.ansys.com/news-center/press-releases/3-18-25-ansys-volvo-accelerate-cfd-simulation-with-nvidia-gpu

[3] Arm. "Ansys Fluent and Ansys LS-DYNA demonstrates leading performance on Arm" https://community.arm.com/arm-community-blogs/b/servers-and-cloud-computing-blog/posts/ansys-fluent-ansys-ls-dyna-on-arm

[4] ANSYS. "Unleashing the Full Power of GPUs for Ansys Fluent, Part 2" https://www.ansys.com/blog/unleashing-the-full-power-of-gpus-for-ansys-fluent-part-2

[5] Rescale. “Leveraging Specialized Architectures with Domain-Specific Hardware Accelerators: Nvidia GPUs and Arm Chips on Rescale.” https://rescale.com/blog/leveraging-specialized-architectures-with-domain-specific-hardware-accelerators-nvidia-gpus-and-arm-chips-on-rescale/

[6] FOAM Iberia 2023. “Leveraging Arm Architecture And Rescale Cloud Hpc Platform For Enhanced Openfoam Performance: A Comparative Analysis” https://2023.foam-iberia.eu/program/ Session T3

[7] Rescale. “Cloud Computing Infrastructure for HPC.” https://rescale.com/platform/hpc-ai-architectures/

[8] Rescale. “GPU Center of Excellence.” https://rescale.com/solutions/by-use-case/gpu-coe/

[9] Arm. "Arm HPC Ecosystem." https://www.arm.com/solutions/high-performance-computing

[10] Nvidia. "NVIDIA GPU Solutions for HPC." https://www.nvidia.com/en-us/data-center/gpu-acceleratedapplications/high-performance-computing/