HPCSIM25 - Frontiers of High-Performance & Cloud Computing in Modeling and Simulation

2025 September 11th-12th Padova, Italy

DAY 1 - 11 September

9:00
Matthew Szyndel | SLB
Keynote speech
Hardware agnostic reservoir simulation: enabling our clients to solve fluid flow in porous media efficiently and flexibly
The simulation of fluid flow in porous rock has been of interest to the energy sector since computers were first able to solve a useful problem within a reasonable timescale. As computing performance has increased, so too has the spatial and physical fidelity of the models used. While 40 years ago the focus of these simulations was exclusively on the efficient extraction of hydrocarbons, today we also model carbon storage, lithium extraction, geothermal energy extraction and hydrogen storage.
The Intersect™ reservoir simulator is the flagship reservoir simulator of SLB. Customer data is proprietary and general; we cannot make assumptions about the grid we are simulating at the outset as it may not be shared with us and can take any form we allow. Furthermore, to be competitive, the Intersect simulator needs to make effective use of the high-performance computing hardware that our customers have available and to offer an effective software as a service solution. Upon initial release in 2010, the Intersect simulator was designed from the ground up to be a scalable distributed parallel solution with no underlying assumptions on the grid topology and characteristics. Since that release we have extended the Intersect simulator to use hybrid shared/distributed memory parallelism (2016), and to run as a full graphics processing unit (GPU) simulator (2020), should that be the best option for the customer. Multi-GPU simulation is now our development focus. Our architecture also enables us to experiment quickly and efficiently with alternative accelerator choices.
In this talk we will outline the algorithms and architecture which underpin Intersect. We will discuss the similarities and differences between central processing unit and graphics processing unit implementations and how these implementations are kept aligned. We will present performance results for a range of hardware options and discuss scalability and future options for development.

10:00
Massimo Bernaschi | Institute for Applied Computing, National Research Council of Italy (IAC-CNR)
Performance and Energy Efficiency of BootCMatchGX
BootCMatchGX is a library designed to exploit to their best high-performance and energy-efficient Graphics Processing Unit (GPU) accelerators in solving sparse linear systems with dimensions that far exceed the memory capacity of a single node.

In this talk, we present our implementation design patterns and solutions to problems that could limit the scalability. We complement performance results with energy profiles that address the sustainability requirements of modern HPC platforms.

10:30
Gianluca Carnielli | Mathworks and Imperial College
Enabling High-Performance Scientific Computing with MATLAB and Simulink
Discover how MATLAB and Simulink enable scientists and engineers to harness the power of high-performance computing for scientific discovery and engineering innovation.
This presentation showcases scalable workflows for parallel computing, GPU acceleration, and integration with HPC clusters and cloud platforms. Learn how to accelerate simulations, process large datasets, and deploy complex models efficiently bridging the gap between desktop development and large-scale execution.

11:00
Coffee Break

11:30
Stefano Regazzo | University of Padova
URANOS-2.0: Exascale-Ready CFD computations for Europe's Leading GPU Architectures
URANOS-2.0 represents a significant advancement in high-performance CFD, delivering a robust, open-source, and vendor-neutral solver for compressible wall-bounded flows. Optimized for Europe's leading GPU architectures, it ensures exceptional portability through a hybrid OpenACC/CUDA implementation and demonstrates outstanding scalability up to 1024 GPUs.
This contribution presents a comprehensive performance assessment of URANOS-2.0 and highlights its capability to handle DNS, WRLES, and WMLES with exascale efficiency. Thanks to extensive kernel refactoring, the solver achieves a twofold speedup over its previous version, establishing a new standard for high-fidelity turbulence modeling in aerospace and energy engineering.

12:00
Matthieu Robeyns | CNRS
A parallel approach for solving the Green's function based on the nested-dissection and selected-inverse methods
This talk deals with the parallelization of the solution of the retarded Green's function which is a critical and computationally expensive step in quantum transport applications based on the NEGF method.
Our approach takes advantage of the sparsity of the involved operators and reformulates this operation in the form of a selected-inverse algorithm. We evaluate the use of matrix permutation techniques based on the nested-dissection method to increase parallelism and analyze the associated added computational load resulting from fill-in in order to achieve a favorable trade-off between operational complexity and parallelism.
We finally discuss how the proposed algorithm can be parallelized using a task-based parallel programming model.

12:30
Lunch

14:00
Mario Tacconi | ENI
Composable Hybrid Computing: Integrating Microservices with HPC
As scientific and industrial computing demands evolve, the convergence of artificial intelligence (AI), quantum computing (QC), and high-performance computing (HPC) is becoming essential for enabling next-generation workflows. This project investigates the feasibility of a composable hybrid computing architecture that could integrate AI, QC, and other emerging paradigms, such as neuromorphic processors, into traditional HPC environments. These components, designed as service-oriented architectures (SOA), aim to create a flexible, scalable, and modular ecosystem where AI-driven insights, quantum acceleration, and brain-inspired computing can be dynamically integrated to enhance simulation-based models and solve classically intractable problems.
At the core of this integration is a Kubernetes (K8s) cluster embedded within the HPC infrastructure. This cluster orchestrates microservices that expose the heterogeneous computing backends as computing services, enabling seamless communication between loosely coupled AI/QC/neuromorphic components and tightly coupled simulation workloads. By leveraging containerization, service discovery, and hardware abstraction, the system supports dynamic resource allocation, fault tolerance, and rapid deployment of diverse computational models alongside legacy HPC applications.
The hybrid architecture is designed to harness the unique strengths of each computing paradigm: HPC for large-scale numerical simulations, AI for pattern recognition and surrogate modeling, QC for solving specific classes of problems with exponential speedups, and neuromorphic systems for energy-efficient cognitive tasks. By combining these capabilities, the system enables new approaches to grand challenge problems in fields such as climate and weather modeling, nuclear fusion, materials discovery and complex systems optimization.
This project aims to demonstrate the feasibility and transformative potential of such a hybrid paradigm through a series of domain-specific use cases. It also addresses key challenges including workload orchestration, data interoperability, and infrastructure integration, laying the foundation for a new approach to intelligent, heterogeneous, and composable scientific computing.

14:30
Omar Bettinotti | Dassault Systemes
AI-Powered 3D Surrogate Physics Models: Learning from Historical Data
Estimating 3D physics fields from virtual twin simulations is critical since the very early phases in the design of industrial products. However, exploring multiple design options in such phases can be computationally expensive, especially when using high-fidelity models. Surrogate modeling provides faster alternatives for early decision-making. Machine Learning (ML) techniques have emerged as non-intrusive techniques, performing well in terms of computing times and accuracy for a wide range of design exploration. Within the ML subset, traditional statistical learning techniques interpolate scalar quantities over a design space with acceptable accuracy, whereas more advanced Deep Learning (DL) techniques based on Neural Networks are better positioned to recover the full history of transient quantities or even full 3D fields.

Starting from parametrized design and fully labeled dataset in design exploration, DL techniques have been returning astonishing results. On the other hand, the complex iterative workflow from design to modeling and simulation of complex products and assemblies often makes it impractical and challenging to consider and keep track of parameter dimensions. Therefore, a new paradigm shift is required in order to tackle unlabeled and historical datasets in input. For this accomplishment, self-supervised learning approaches can map geometry to latent representations, addressing the feasibility limits of parametrized surrogate modeling methods.
With this presentation the Authors will share their prototyping efforts and progress for a variety of test cases, ranging from quasi-statics to transient dynamics and multi-physics applications.

15:00
Simone Bna | Cineca
SPUMA: a minimally-invasive GPU porting of OpenFOAM
High Performance Computing (HPC) on hybrid clusters represents a significant opportunity for Computational Fluid Dynamics (CFD) when modern accelerators are utilized effectively. However, despite the widespread adoption of GPUs, programmability remains a challenge, especially in open-source contexts. In this talk, we present SPUMA, a full GPU porting of OpenFOAM targeting NVIDIA and AMD GPUs.
The implementation strategy, based on a portable programming model and on the adoption of a memory pool manager that leverages the unified memory feature of modern GPUs, requires minimal changes in the core of the OpenFOAM source code. Numerical tests on the two pre-exascale clusters in Europe, LUMI and LEONARDO, hosting AMD MI250X and NVIDIA A100 GPUs respectively, will be presented.

15:30
Coffee Break

16:00
Eric Pascolo | Cineca
Funding and Infrastructure for Innovation: AI Factory, HPC, and European Opportunities for Enterprises and Academia
This talk will explore the evolving landscape of European funding programs that support innovation across both enterprises and academia, with a focus on initiatives such as AI Factory. It will highlight how these opportunities intersect with engineering and artificial intelligence, and how they can be leveraged to accelerate research and development.
The session will also present upcoming infrastructure developments at Cineca, including advancements in high-performance computing (HPC), the integration of AI Factory resources, and the emergence of quantum computing capabilities—positioning Italy at the forefront of European digital innovation.

16:30
Ruggero Poletto | CFD FEA Service
Scalability comparison between Calculix and CodeAster

DAY 2 - 12 September

9:00
Stefano Zampini | KAUST
Keynote speech
Device-accelerated solvers with PETSc
Despite the revolutionary impact of GPUs on various computational tasks, challenges still emerge when trying to efficiently utilize these architectures for the solution of large nonlinear systems of partial differential equations, highlighting the need for tailored algorithms.
In this talk, I will discuss recent developments towards exploiting GPUs with the Portable and Extensible Toolkit for Scientific Computing (PETSc), an award-winning software for the numerical solution of large systems of nonlinear and time-dependent partial differential equations.
The focus will center on diverse components, encompassing Krylov methods, direct solvers, preconditioners, multi-GPU communications, distributed sparse matrix-matrix multiplications, and quasi-Newton methods. The objective is to furnish the audience with valuable insights into attainable and realistic performance expectations.

10:00
Gianluigi Rozza | Sissa
Surrogate Modelling by Reduced Order Methods and Machine Learning to enhance Scientific Computing: focus on CFD
Partial differential equations (PDEs) are invaluable tools for modeling complex physical phenomena. However, only a limited number of PDEs can be solved analytically, leaving the majority of them requiring computationally expensive numerical approximations. To address this challenge, reduced order models (ROMs) have emerged as a promising field in computational sciences, offering efficient computational tools for real-time simulations. In recent years, deep learning techniques have played a pivotal role in advancing efficient ROM methods with exceptional generalization capabilities and reduced computational costs [1, 2, 3], especially for parametric settings and turbulent flows.
In this talk we explore how classical ROM techniques can be elevated through the integration of some deep learning models. We will introduce hybrid approaches, which consider both physics-based and purely data-driven techniques [4, 5, 6, 7], as well as aggregated ones, where the model is built as a combination of different pre-trained models [8]. Examples will deal with parametric flows in presence of compressibility and turbulence, as well as industrial applications.
References
[1] Benner, P., Grivet-Talocia, S., Quarteroni, A., Rozza, G., Schilders, W., & Silveira, L. M. (2021). System-and data-driven methods and algorithms (p. 378). De Gruyter.
[2] Benner, P., Schilders, W., Grivet-Talocia, S., Quarteroni, A., Rozza, G., & Miguel Silveira, L. (2020). Model Order Reduction: Volume 2: Snapshot-Based Methods and Algorithms (p. 348). De Gruyter.
[3] Benner, P., Schilders, W., Grivet-Talocia, S., Quarteroni, A., Rozza, G., & Miguel Silveira, L. (2020). Model order reduction: Volume 3: Applications (p. 466). De Gruyter.
[4] Ivagnes, A., Stabile, G., Mola, A., Iliescu, T., & Rozza, G. (2023). Hybrid data-driven closure strategies for reduced order modeling. Applied Mathematics and Computation, 448, 127920
[5] Ivagnes, A., Stabile, G., & Rozza, G. (2024). Parametric Intrusive Reduced Order Models enhanced with Machine Learning Correction Terms. arXiv preprint arXiv:2406.04169.
[6] Zancanaro, M., Mrosek, M., Stabile, G., Othmer, C., & Rozza, G. (2021). Hybrid neural network reduced order modelling for turbulent flows with geometric parameters. Fluids, 6(8), 296.

[7] Stabile, G., Zancanaro, M., & Rozza, G. (2020). Efficient geometrical parametrization for finite-volume-based reduced order methods. International journal for numerical methods in engineering, 121(12), 2655-2682.
[8] Ivagnes, A., Tonicello, N., Cinnella, P., & Rozza, G. (2024). Enhancing non-intrusive reduced-order models with space-dependent aggregation methods. Acta Mechanica, 1-30.

10:30
Xavier Trias | Polytechnic University of Catalonia
On the evolution of Poisson solvers for extreme-scale CFD simulations
This work addresses a fundamental question: is the numerical complexity of solving Poisson's equation increasing or decreasing for very large DNS and LES simulations of incompressible flows? As Reynolds numbers grow, simulations require both finer meshes and smaller time steps. There are two competing effects influencing the convergence of Poisson solvers: namely, the former increases the condition number of the discrete Poisson's equation whereas the latter leads to better initial guesses. Physical and numerical arguments are combined to derive power-law scalings at very high Reynolds numbers.
A theoretical convergence analysis of both Jacobi and multigrid solvers defines a two-dimensional phase space, partitioned into two regions depending on whether the number of solver iterations decreases or increases with the Reynolds number.
These insights can help to define a road-map for the development of Poisson solvers, providing guidance on how CFD codes may need to evolve for future HPC architectures approaching the zetta (and eventually yotta) scale.

11:00
Coffee Break

11:30
Simone Montangero | University of Padova
TBD

12:00
Carlos Roig | CIMNE
SYCL based acceleration of a Structured CFD Solver
We present a novel structured mesh solver tailored for large scale incompressible external flow simulations, specifically designed for extreme-scale computing.
The method leverages a minimal data structure that eliminates the need to store the mesh explicitly in memory and uses SYCL to bring the possibility of targeting different accelerators. These features are combined with an implicit representation of the analysed geometries via distance functions (i.e., level set) and a boundary condition imposition done with the Shifted Boundary Method (SBM). Thanks to these, the solver is able to handle complex domains while keeping the benefits of a structured mesh representation. Besides, a fractional step numerical scheme is employed aiming to maximize computational efficiency. Specifically, velocity prediction is performed explicitly using a 4th order Runge-Kutta method and pressure correction is solved by solving a Laplacian system using Conjugate Gradients, both highly parallelizable.

12:30
Lunch

14:00
Alfonso Amendola | ENI
High Performance Computing for the Energy Transition: challenges and opportunities
Energy sector has traditionally used high performance computing modelling techniques in key areas such as seismic imaging, geological modelling, and reservoir flow simulation.
In this talk, we highlight new areas of application of high-performance computing in Eni, with practical examples drawn from advanced material simulation and wind resource modelling and optimization, showing that these applications might enhance competitiveness and sustainability of the Energy sector.

14:30
T.B.D. | Saipem
TBD

15:00
Fabio Villa | ESSS
Reducing time to solution in ANSYS: from GPU acceleration to AI solution
Reducing time to solution in ANSYS is critical for accelerating engineering workflows and enhancing productivity, especially in computational fluid dynamic. In this presentation we will explores the advancements solutions proposed by ANSYS.
From the traditional HPC solution to GPU-accelerated. GPU acceleration enhances simulation performance, reduces time to solution, and enables more complex or higher-fidelity analyses. Also the new AI solution introduces a paradigm shift by enabling predictive modeling and surrogate simulations. By integrating machine learning techniques, ANSYS workflows can bypass iterative solvers for certain tasks, achieving near-instantaneous results with acceptable accuracy. This evolution not only reduces computational load but also opens new possibilities for real-time design iteration and optimization. The transition from GPU to AI signifies a transformative step in simulation efficiency.

15:30
Coffee Break

16:00
Sebastian Gries | Frauhnofer SCAI
Grey-Box Approach to Linear Solvers: Tailored Strategies in Different Practical Applications
While it appears to be a simple linear algebra operation of inverting a matrix, the linear solver part typically accounts for the largest portion of runtime in a numerical simulation. Which methods to choose to design a robust and efficient solution strategy depends on the type of application that you are considering. With further emerging physical complexity in the simulation models, it is hardly possible to simply utilize a standard text-book method if optimal performance and robustness were expected.
With a Grey-Box Approach, we refer to identifying a suitable linear solution method for a particular type of application.
While this requires some initial design-work, it can be used robustly and efficiently in a black-box manner with different simulations of that kind afterwards.

We will discuss this approach with practical examples from different industrial simulation applications. The range of solver strategies will consider application-specific algebraic-multigrid approaches. But we will also demonstrate how adjusted incomplete-factorization-type methods can be successfully applied in applications that do not (yet) require the strength of multigrid.
In all cases, after finding the basic strategy, a Machine-Learning-based parameter optimization can be applied to realize a case-based performance optimization.

16:30
Joel Guerrero | Leonardo
How HPC and physical modeling enables industrial innovation