Možnosti užití GPU s programy Ansys

TechSoft

Michal Moštěk, TechSoft Engineering
mostek@techsoft-eng.cz

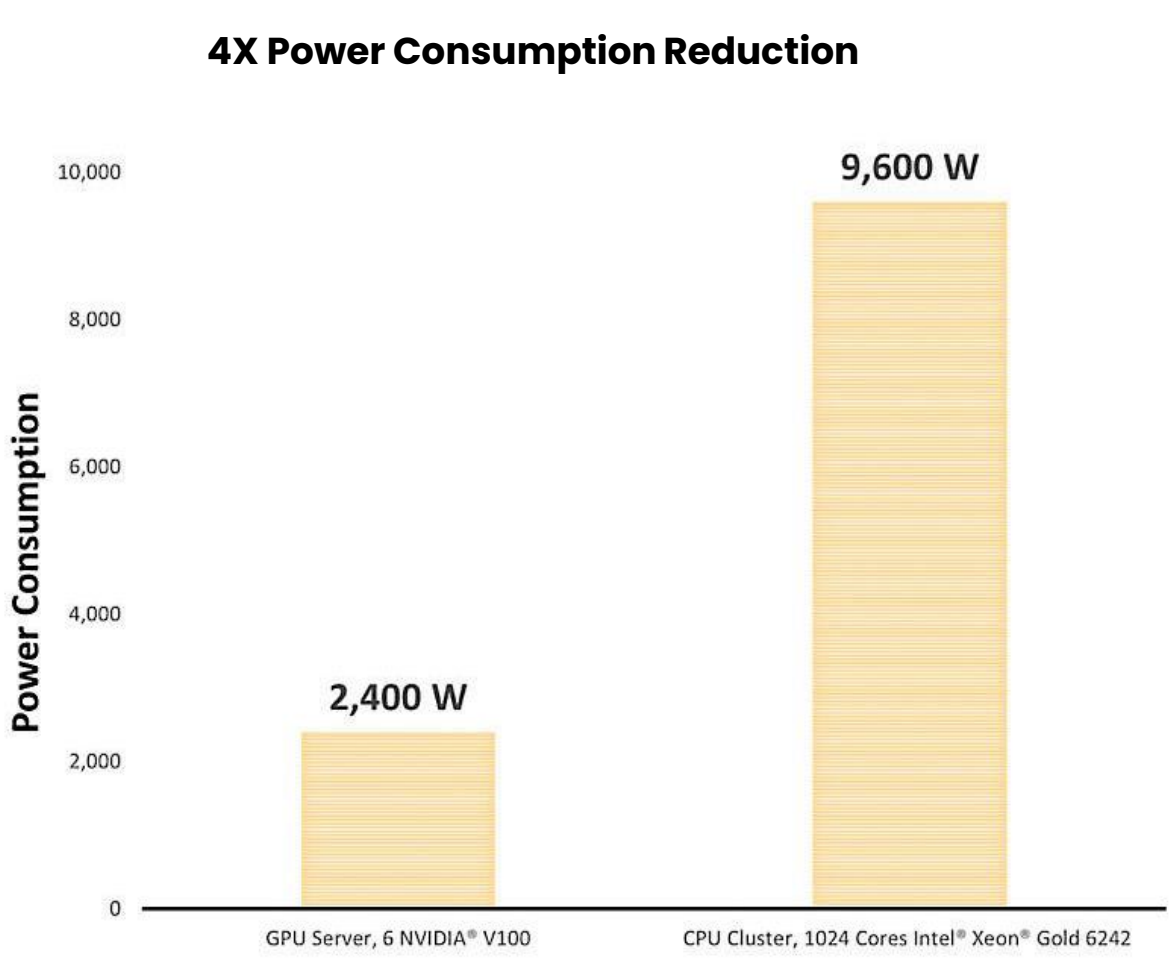
HPC infrastructure and data is constantly evolving…

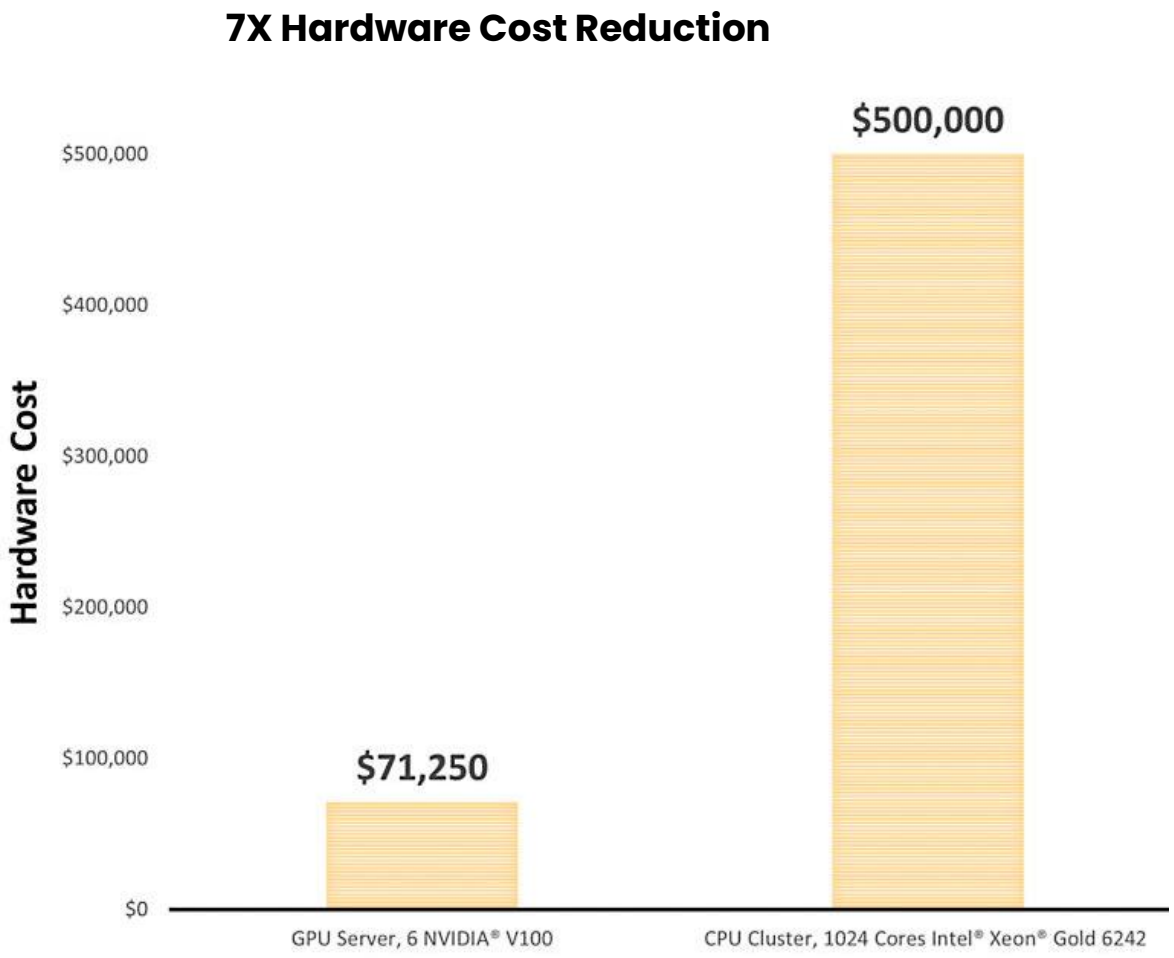
Towards centralized and consolidated resources, with users remote, global, and mobile

Driven by:

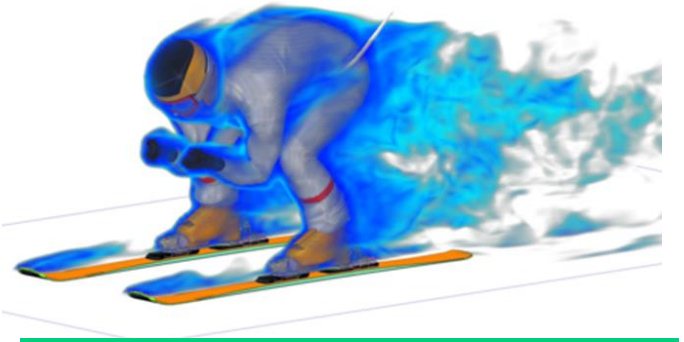
- **Demand for throughput and higher fidelity simulation results**
- **Operational efficiency and asset optimization**
- **More workloads, more design points and parameters, V&V, and uncertainty quantification**
- **Need for collaboration and data sharing across geographically distributed users**
- **Increasing focus on simulation IP protection and process traceability**

Native GPU Benefits Go Beyond Fast Turnaround

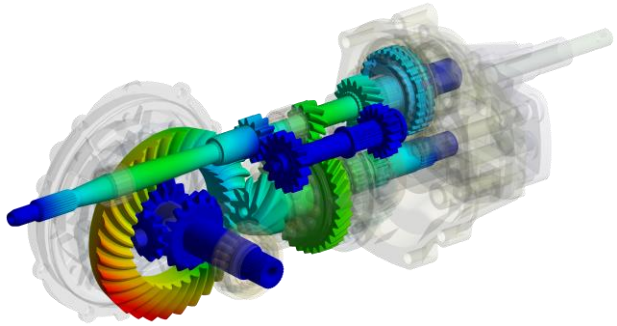




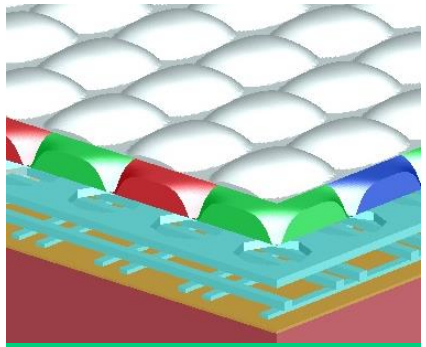
Ansys products exploit parallelism in all forms
– Enabling efficient many-core computations



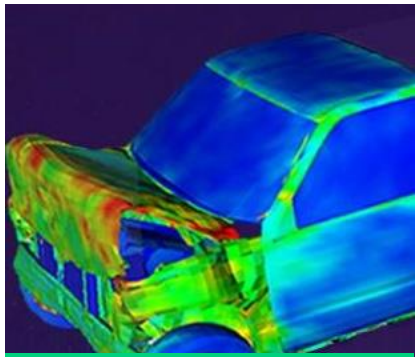
Fluent



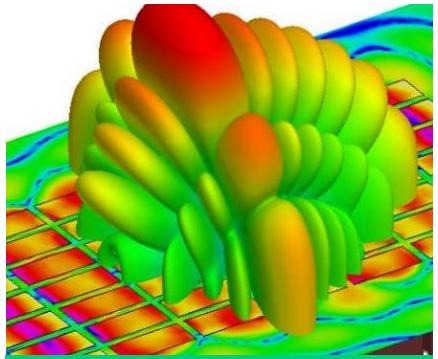
Mechanical



Lumerical FDTD



LS-Dyna



HFSS/SI-Wave

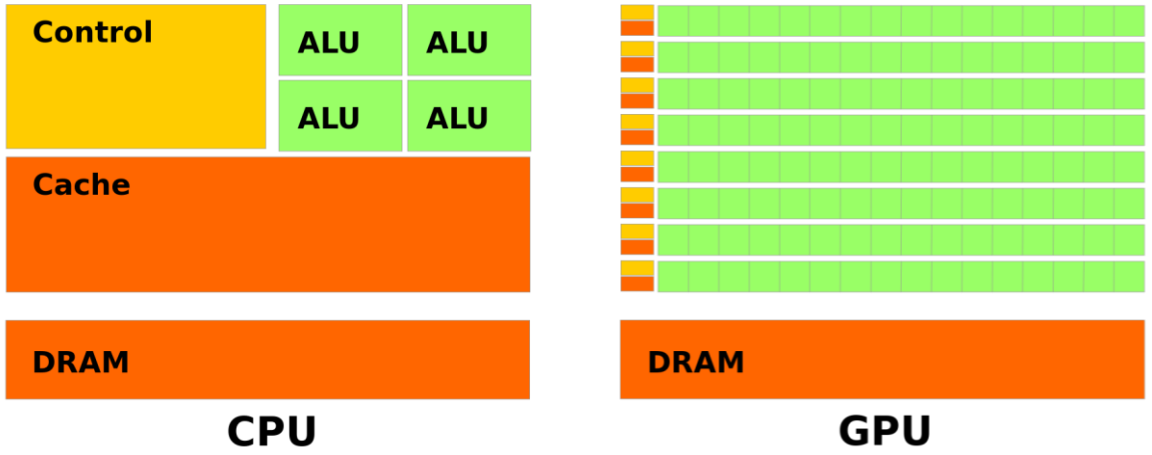
4 Layers of HPC

Parallel Parametric Analysis	HPC Job Scheduler
Domain Decomposition	MPI – NUMA & Distrib. Mem. Cluster
Parallel Tasks/Loops	Multicore / Multithread
SIMD Vector Processing GPU Parallelism	Intra-Core / GPU

Harnessing GPUs in Currently Released Ansys Software

- **Ansys Fluent, SPEOS, Discovery Explore** solvers can run end-to-end **natively** on GPUs
  - Within HFSS, the **SBR+ solver** runs entirely on GPUs for real-time radar simulations
  - **Lumerical's FDTD HF** solver also well-suited for efficient GPU computation

- Ansys Flagship Products such as **Mechanical, HFSS, Maxwell** use GPUs as accelerators in **"offload"** mode.
  - GPUs have large bandwidth and computational throughput
  - Good for isolated, expensive computations
  - Solver modules like linear algebra, ray tracing, radiation models can run efficiently on GPUs

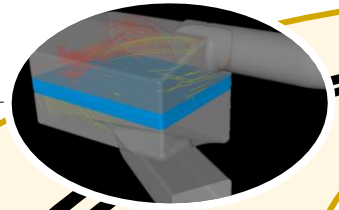


Multi-GPU Solver in Fluent: Release Timeline

UTILIZE THE POWER OF MULTIPLE GPUS TO ACCELERATE YOUR CFD SIMULATIONS

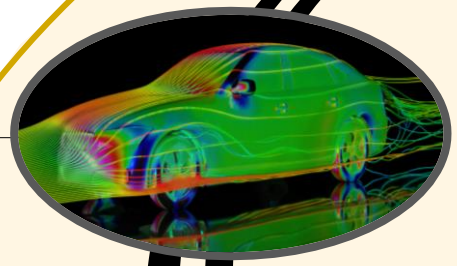
Beta support for :
- Single/multi-GPU (shared / distributed memory)
- Steady & transient simulations
- Incompressible & subsonic compressible flows
- All mesh types
- Ideal Gas and Materials with variable properties
- Turbulence: standard k-e, SST, GEKO, RKE, SBES
- Solid conduction and CHT
- Porous media
- Windows and Linux

2022/R1

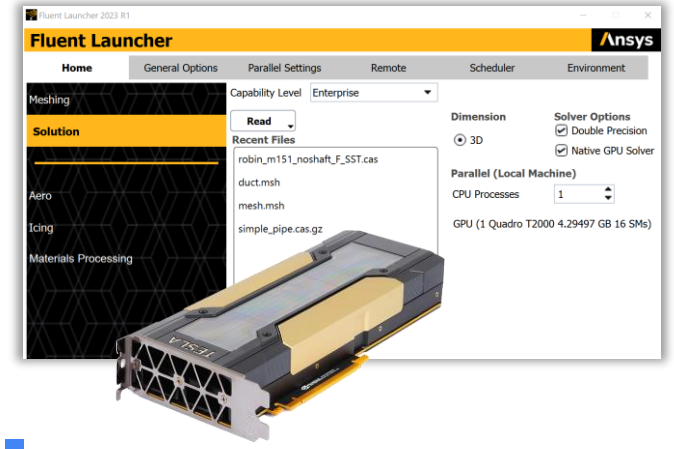


Beta support for :
- Transient scale-resolving simulations
- Non-conformal mesh interfaces
- Moving walls & Moving Reference Frame

2022/R2



2023/R1

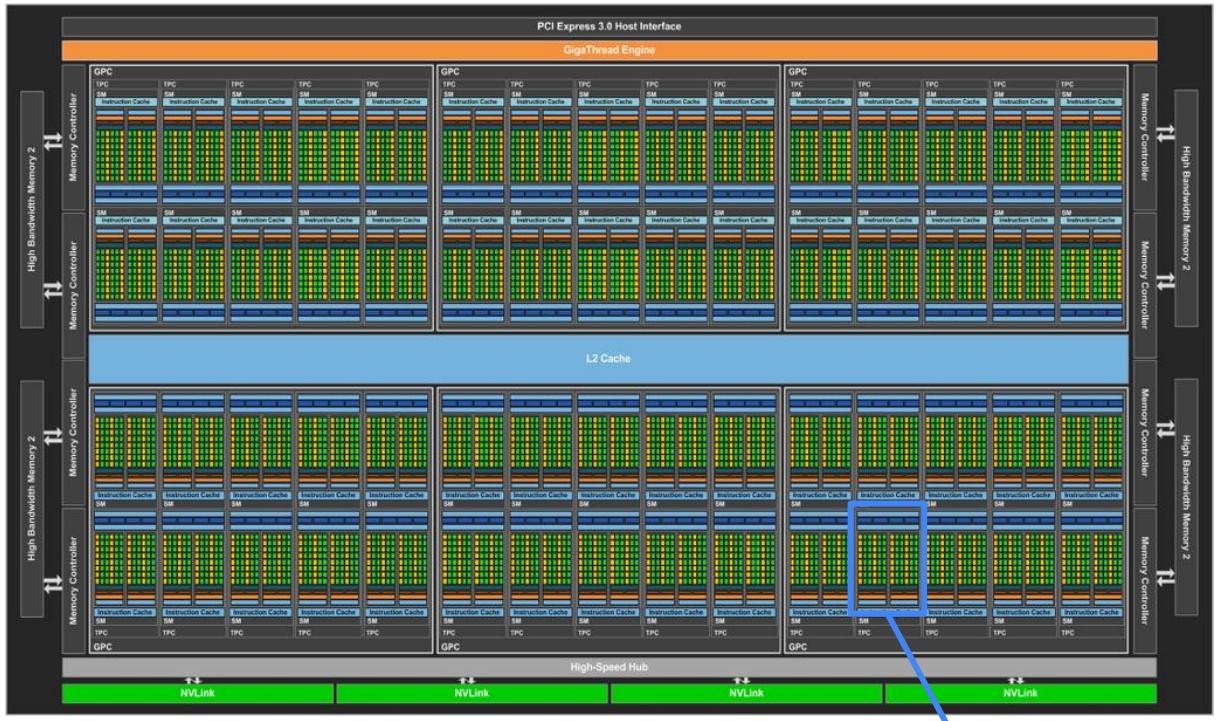


All previous features released, along with :
- Various numerics improvements
- Enhanced robustness and accuracy for RANS
- Species transport
- Enhanced LES numerics
- Enhanced monitors: Uniformity, HTC, Porous force, Moments, Volume averaging
- Velocity directions for pressure inlet , Porous jump, Intake fan, Outlet vent , Wall thickness
- Enhanced Handling of unsupported features
- Profiles: Primitive variables for boundary
- Setup time reduction & simplified UI
- Enhanced Materials: Piecewise-linear, Polynomial, Piecewise-Polynomial , Viscosity: Sutherland
- Initialization workflow enhancement
- Improved UX and launcher enhancement

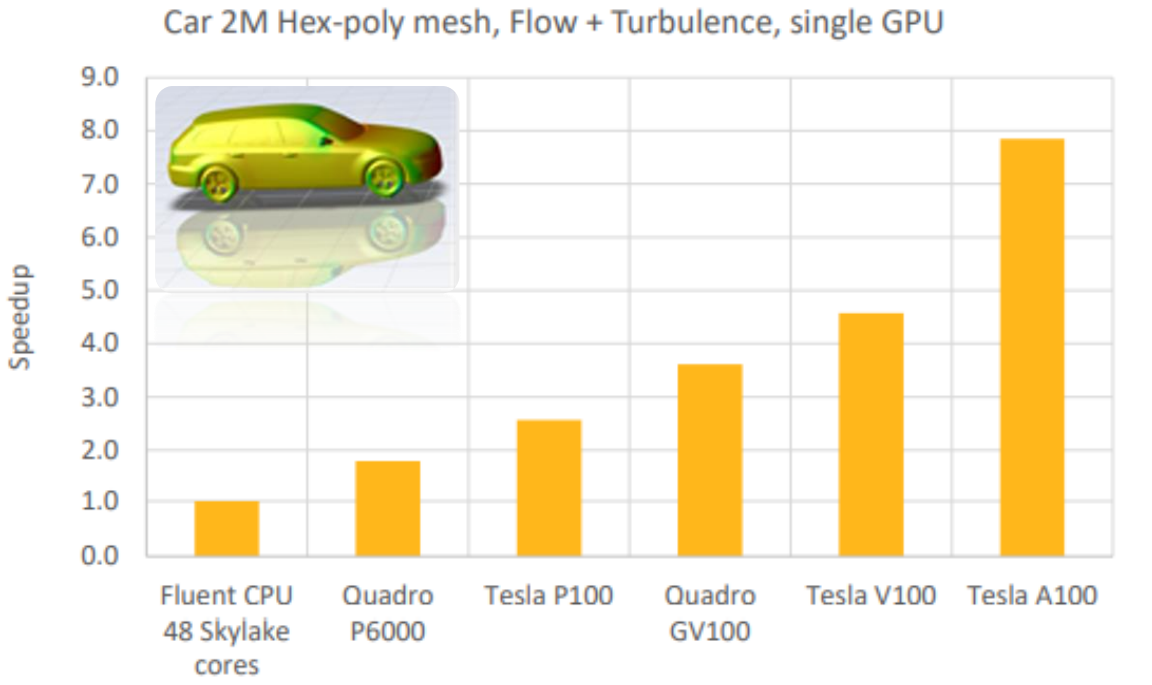
Beta support for :
- Sliding meshes
- Non-stiff reacting flows
- Compressible flows

What are Streaming Multiprocessors (SMs)?

- GPU cards are made up of many SMs, and each SM contains many CUDA cores
- More powerful GPU cards typically contain more SMs



Tesla P100 Layout
56 SMs, 3584 CUDA cores



# SMs	28	56	80	80	108
# CUDA	3840	3584	5120	5120	6912

Licensing

CFD Pro

Entry level physics capabilities

- Steady state flow
- Transient flow
- Heat transfer
- CHT
- Porous media
- Massless/inert particle tracking
- Multi-stream component mixing
- Rotating reference frames
- 2-equation RANS turbulence
- Compressible and incompressible
- Non-Newtonian fluids
- Watertight Meshing workflow
- Materials Processing – entry level extrusion, co-extrusion, fiber spinning, blow molding and thermoforming

CFD Premium

Full range of physics models

- All physics models in the Fluent CPU solvers
- All physics models in the Materials Processing workspace
- Full Fluent Meshing capabilities, including Fault Tolerant Meshing workflow

CFD Enterprise

GPU Solver and additional modules

In addition to CFD Premium features:

- Fluent GPU solver
- Fluent Aero workspace
- Fluent Icing workspace

- Multi-GPU Solver included in Ansys CFD Enterprise
- 40 SMs (streaming multiprocessors) included out-the-box
- Additional SMs enabled with HPC licensing
  - 1 HPC Task = 1 SM
  - HPC Packs scale as usual (e.g., 3 Packs enables 128 additional SMs)

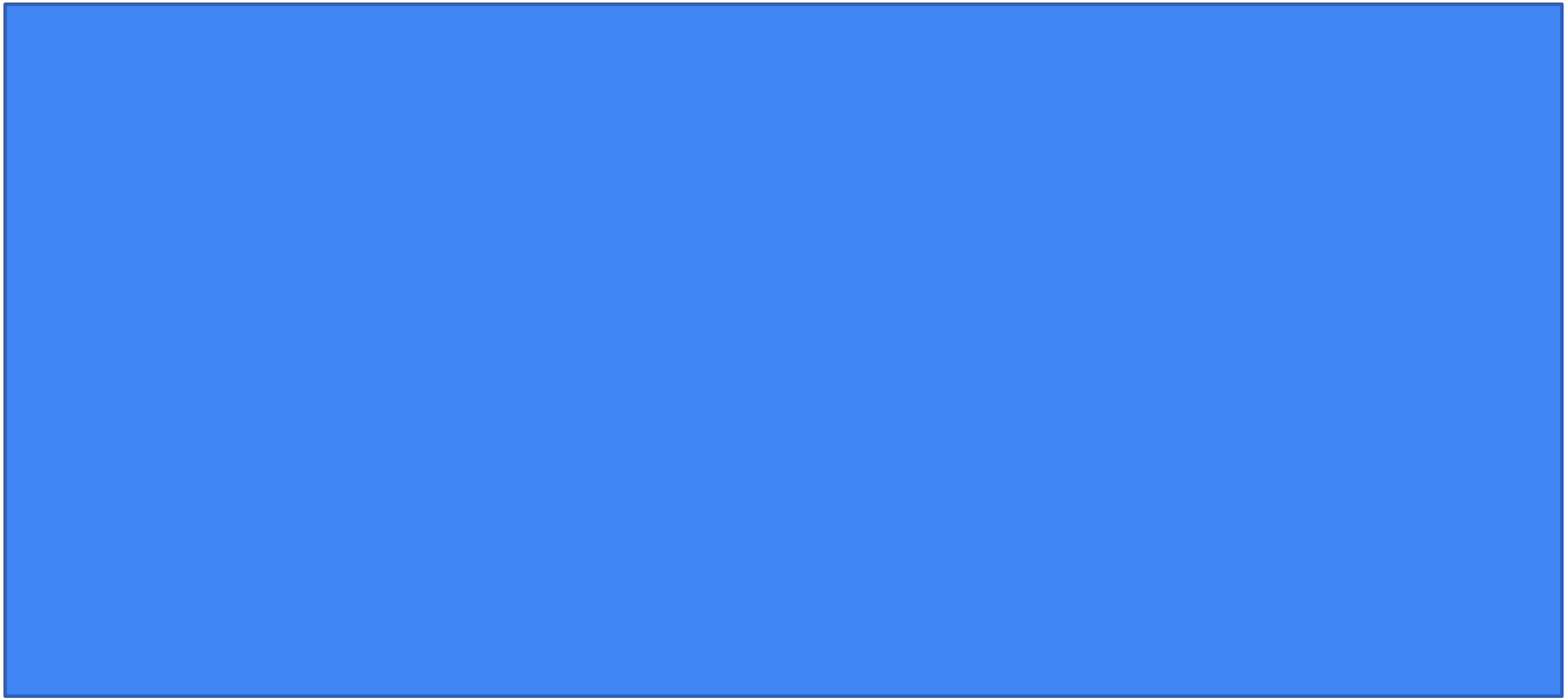
# SM's	HPC Workgroups	HPC Packs
1 – 40	0	0
41 – 48	1 – 8	1
49 – 72	9 – 32	2
73 – 168	33 – 128	3
169 – 552	129 – 512	4
553 – 2088	513 – 2048	5

HPC Requirements for Common GPUs

GPU Card	Description	#SMs	RAM (GB)	~CPU Core Equivalent	Approx. max problem size* (per GPU)	#HPC Packs
RTX A2000 Mobile	Released 2021. Standard laptop GPU.	20	4	6 – 12	3M cells	0
Quadro P2000 Mobile	Released 2017. Standard laptop/desktop GPU	6	4	10 – 15	3M cells	0
Quadro T2000	Released 2019. Standard laptop GPU	16	4	10 – 15	3M cells	0
Quadro P6000	Released 2016. Older high-end workstation GPU	30	24	50 – 90	20M cells	0
Quadro RTX 4000	Released 2018. Typical workstation GPU	36	8	60 – 100	7M cells	0
Tesla P100	Released 2016. Older server GPU	56	16	100 – 130	13M cells	2
Quadro RTX 6000	Released 2018. High-end workstation GPU	72	24	120 – 200	20M cells	2
Tesla V100	Released 2017. Previous gen server GPU	80	32	150 – 220	27M cells	3
Quadro RTX A5000	Released 2021. Top-end workstation GPU	64	24	130 – 220	20M cells	2
Tesla A100	Released 2020. Top-end server GPU	108	40, 80	200 – 400	33M, 67M cells	3
Tesla H100	Released 2023. Next-gen server GPU	132	80	Not yet tested	67M cells	3
RTX 6000 Ada	Released 2023. High-end workstation GPU	142	48	Not yet tested	40M cells	3

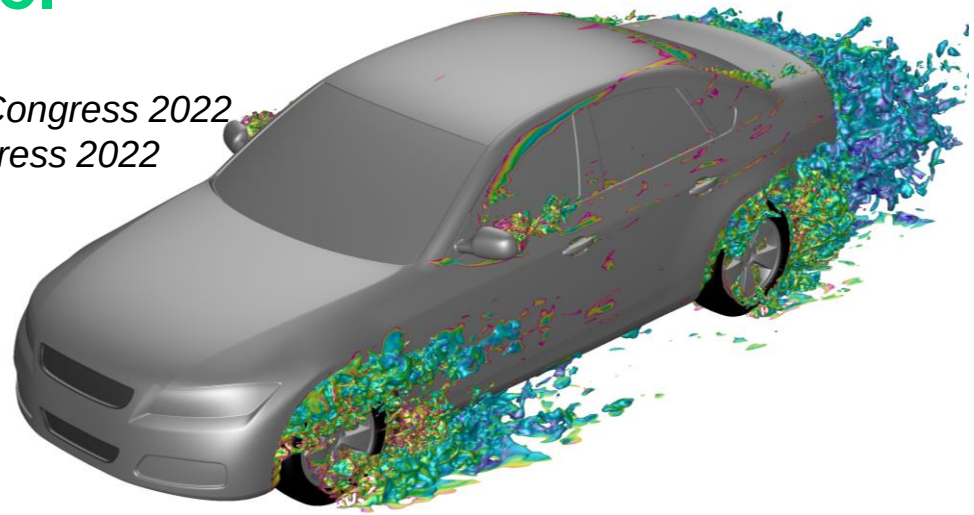
** Assuming 1.2GB RAM per millions cells. Actual RAM requirements will be case specific and will depend on the mesh type, physics solved, single vs double precision and other factors. Larger mesh sizes can be solved using multiple GPUs.*

Ansys Fluent GPU Solver Roadmap

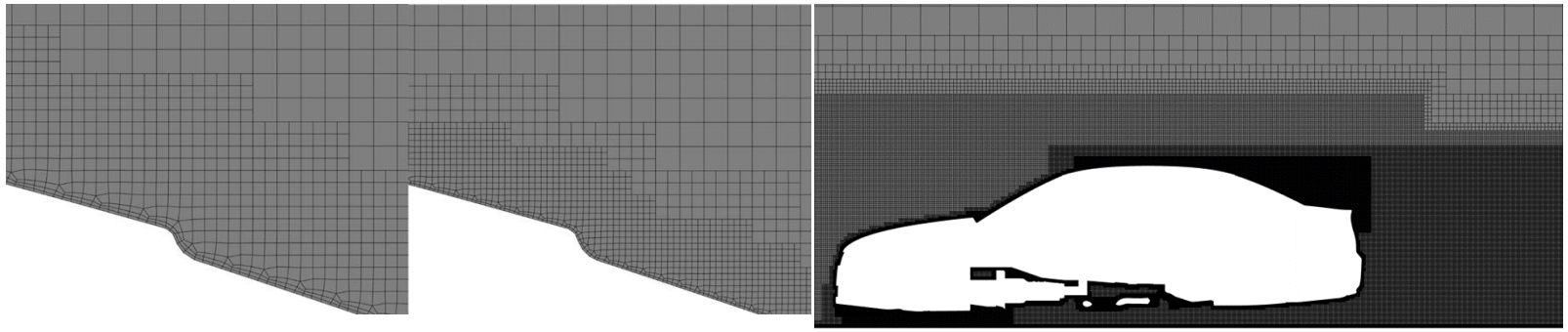


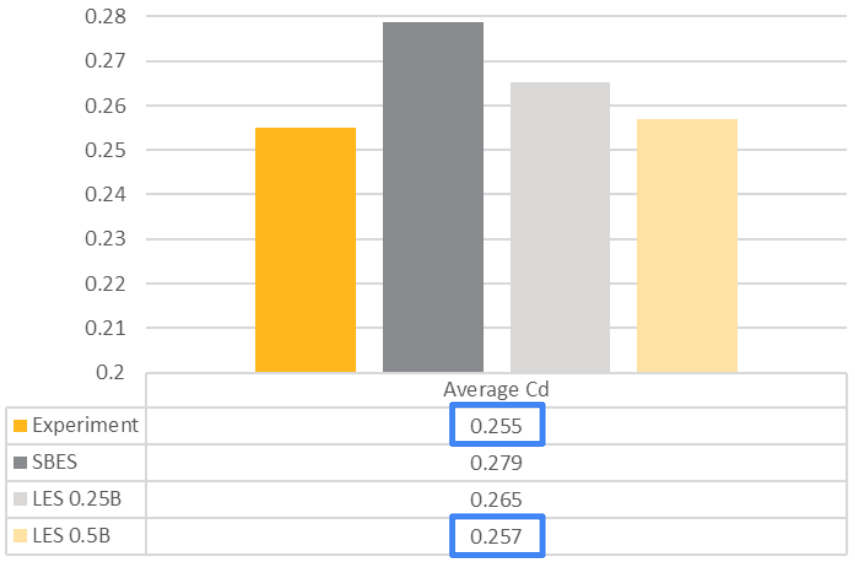
New LES Numerics + RO Mesh on DrivAer Model

Winstanley et al., SIMVEC Congress 2022
Hüppe et al., SIMVEC Congress 2022



- **DrivAer Notchback – front wheel deflector comparison**
  - AutoCFD2 0.244 Billion Octree CPU SBES
  - 0.25 Billion Octree GPU/CPU LES
  - 0.5 Billion Octree GPU/CPU LES
- **Comparing WFLES with SBES and Experiment**



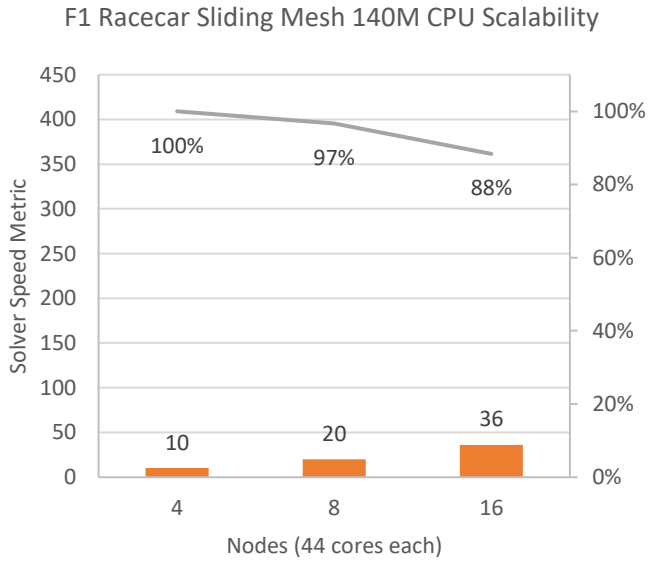


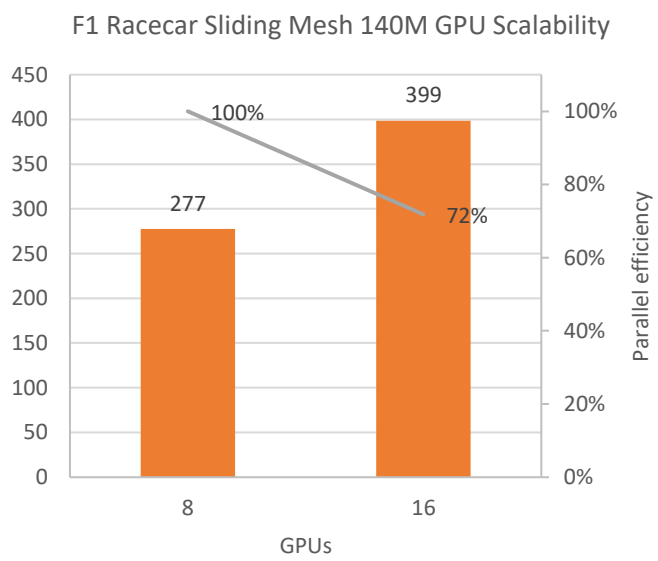
C_D error for new LES on 0.5B cells is 0.78%

Sliding Mesh (beta) – F1 Racecar

F1 Racecar 140M cells

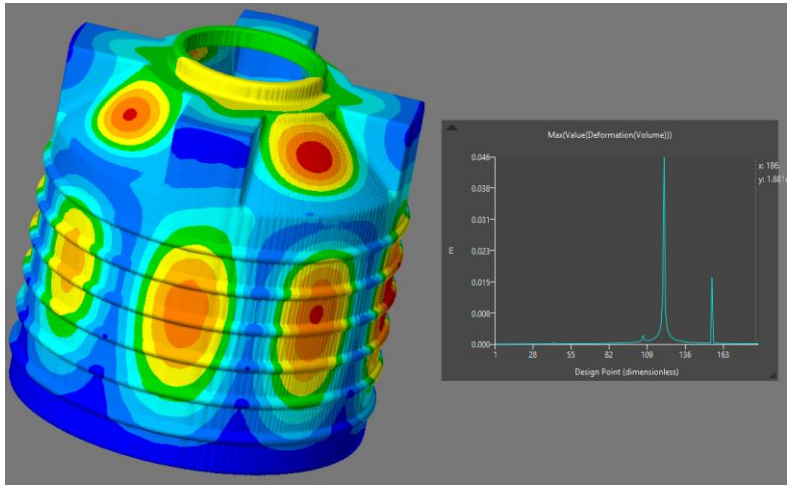
- SST k-omega, transient
- **Computing resources**
  - Intel Platinum 8168 dual-socket cluster
    - 44 cores per node
  - Nvidia A100 40GB 8-GPUs per node
- **1 GPU delivers the performance of 16 CPU nodes (704 cores)**



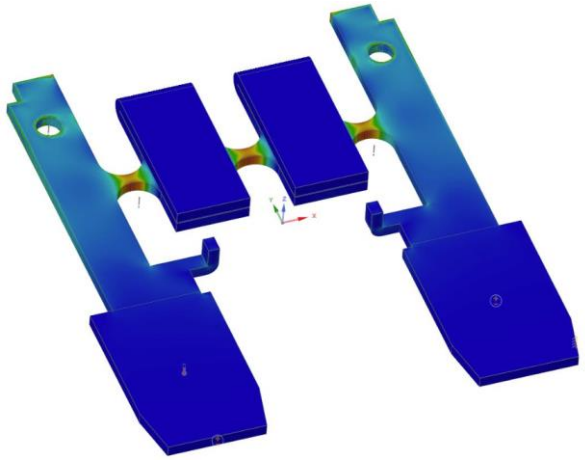


1000 timesteps in 40 minutes on 8 GPUs

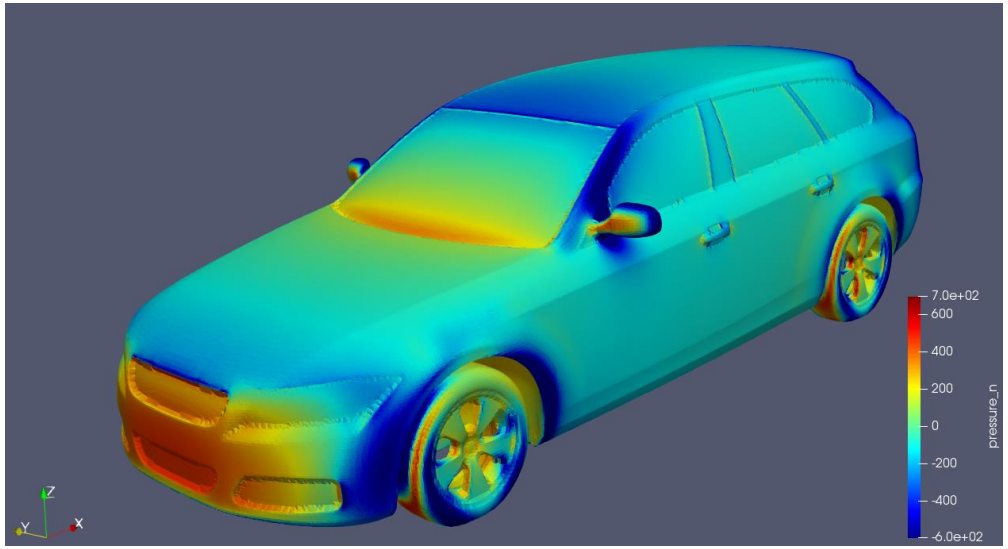
Ansys Discovery: Example of a Fully Native GPU Solver Family



Harmonic Modal Superposition



Electro-Thermal Stress



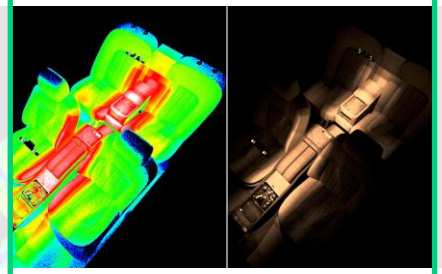
Fluid Mechanics

- End-to-end solution runs **natively** on GPUs
- Structural, Modal, Thermal, Electro-Thermal, Fluids, Modal
- Topology optimization
- Rapid simulation process: geometry, meshing, solution, results analysis and visualization with a high degree of automation
- Supporting near-instantaneous concept exploration and preliminary design

Ansys Speos – Example of Progressive Development: HPC/GPU

2001

- Monte-Carlo Raytrace
- CPU
- Local Workstation
- Full Physics
- Days to compute



2006

- Monte-Carlo Raytrace
- CPU
- **Distributed**
- Full Physics
- Days to compute

2008

- Deterministic Raytrace
- **GPU**
- Local Workstation
- Specular path only
- Semi-Realtime

2012

- Hybrid Raytrace/Raster
- **Multi-GPU**
- **Distributed** on Local Network
- Specular + First Bounce
- Realtime

2015

- Monte-Carlo Raytrace
- CPU
- **HPC/Cloud compatible**
- Full Physics
- Hours to Compute

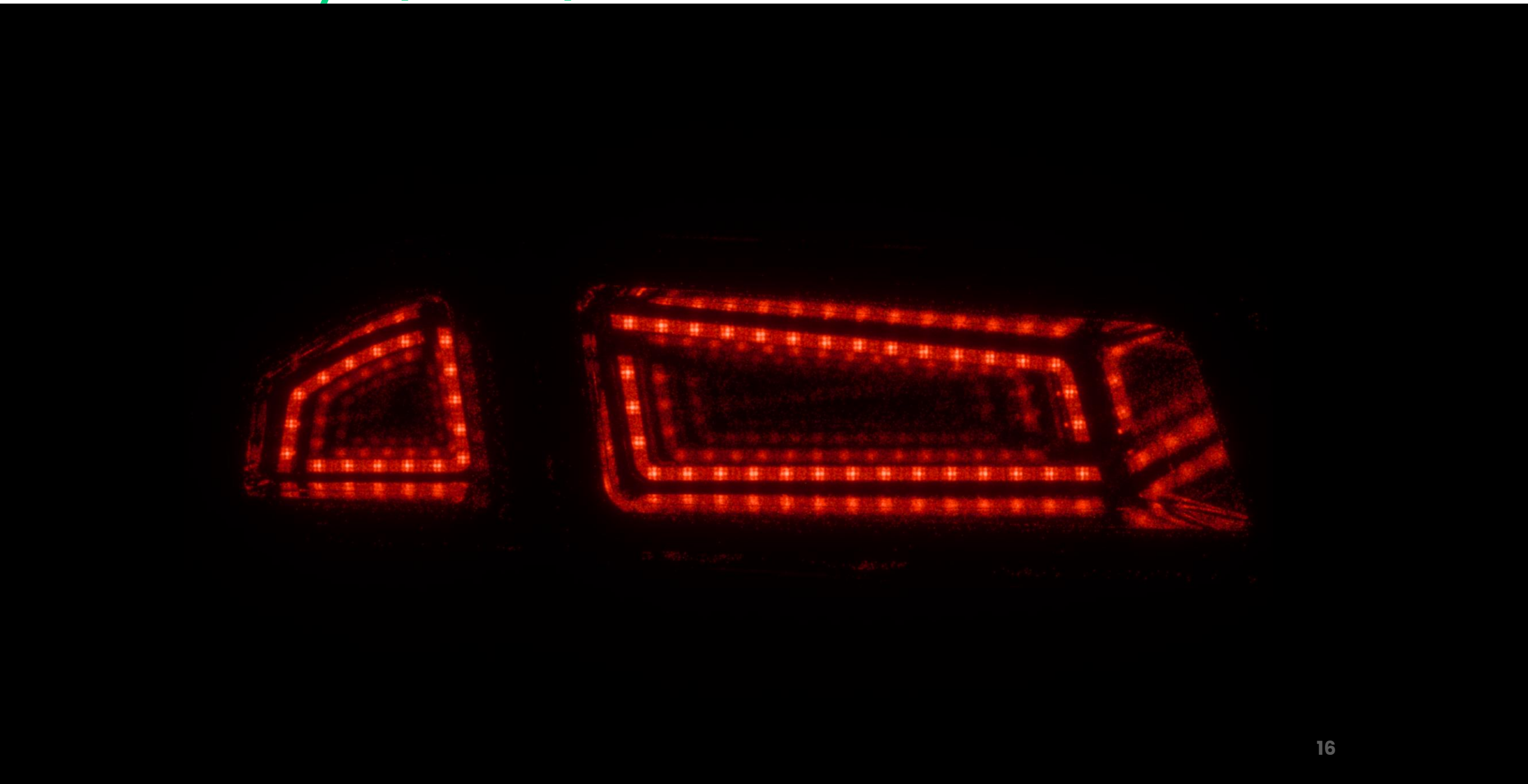
2018

- Monte-Carlo Raytrace
- **Multi-GPU**
- Local
- Full Physics
- Interactive - Seconds to Compute

GPU enabled by using HPC Workgroup 32 license

CPU 24 cores, 2.5e9 rays (8 hours 8 min)

GPU, 2.5e9 rays (20 min)



Ansys HPC product licenses enable GPU Computing
- Ansys HPC, HPC Pack and HPC Workgroup

Fluids* / Structural products

1 GPU requires 1 HPC task as long as **GPUs ≤ CPU cores**

Examples:

- 2 HPC licenses enable up to 3 CPU cores + **3 GPUs** through the available 6 HPC tasks
- 1 HPC Pack enables up to 6 CPU cores + **6 GPUs** through the available 12 HPC tasks
- 2 HPC Packs enable up to 18 CPU cores + **18 GPUs** through the available 36 HPC tasks

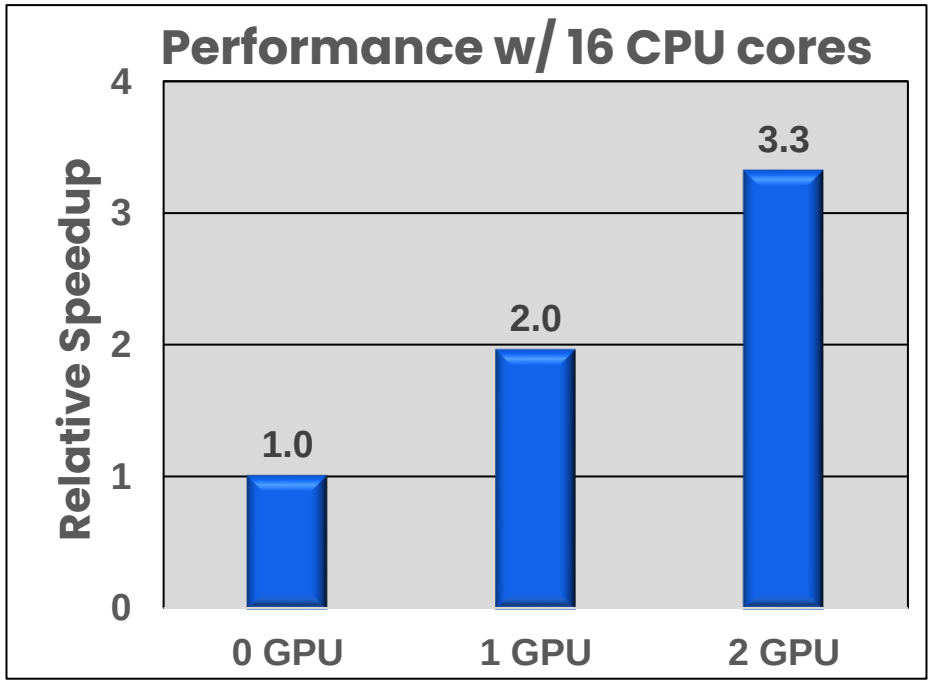
Electronics products

1 GPU unlocked by every 8 HPC tasks

- 4 HPC licenses enable **1 GPU** through the available 8 HPC tasks
- 1 HPC Pack enables up to 12 CPU cores + **1 GPUs** through the available 12 HPC tasks
- 2 HPC Packs enable up to 36 CPU cores + **4 GPUs** through the available 36 HPC tasks

Ansys Mechanical with GPU Acceleration: Offload Mode

- **Hardware Support**
  - NVIDIA GPUs and AMD Instinct GPUs are supported
  - Windows and Linux are supported for NVIDIA GPUs
  - Linux is supported for AMD Instinct GPUs
  - Single or multiple cards are supported
- **Software Support**
  - **Sparse (direct) and PCG (iterative) solvers are both GPU enabled, including most eigensolvers**
  - **This breadth of solver algorithm GPU acceleration is unique in the industry**
  - Distributed and Shared memory parallel are supported
- **Relies on offload model (CPU still required)**
  - Certain key computations in linear equation solvers are offloaded to the GPU for acceleration



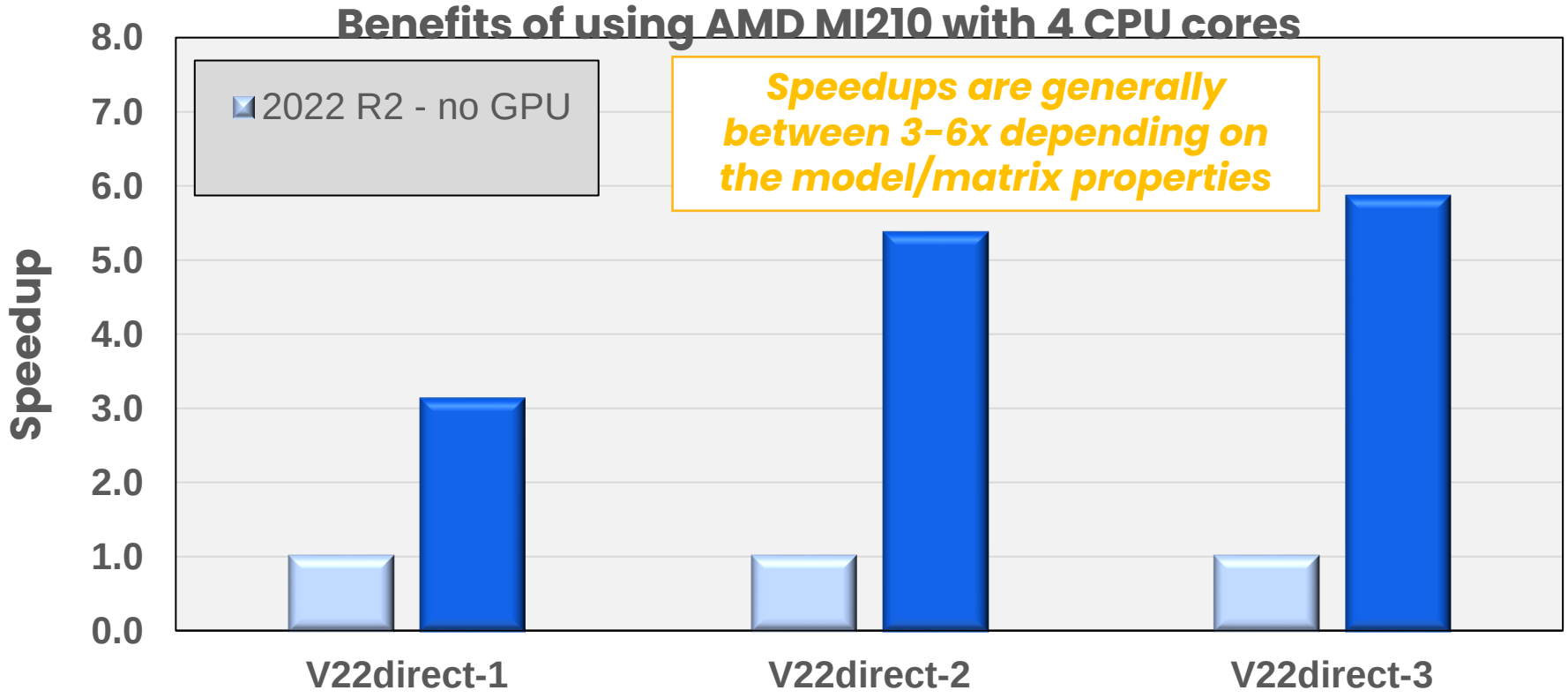
Total Simulation Time Speedup

Ansys Mechanical with GPU Acceleration: Offload Mode

Significantly faster performance for direct solver benchmarks on AMD Instinct GPU cards

Measuring entire matrix factorization time (not just the calculations that are accelerated on the GPU)

Linux cluster; each compute node contains 2 AMD EPYC 7643 processors (48 cores), 512GB RAM, SSD, 1 AMD Instinct MI210 PCIe card, RHEL 8.5



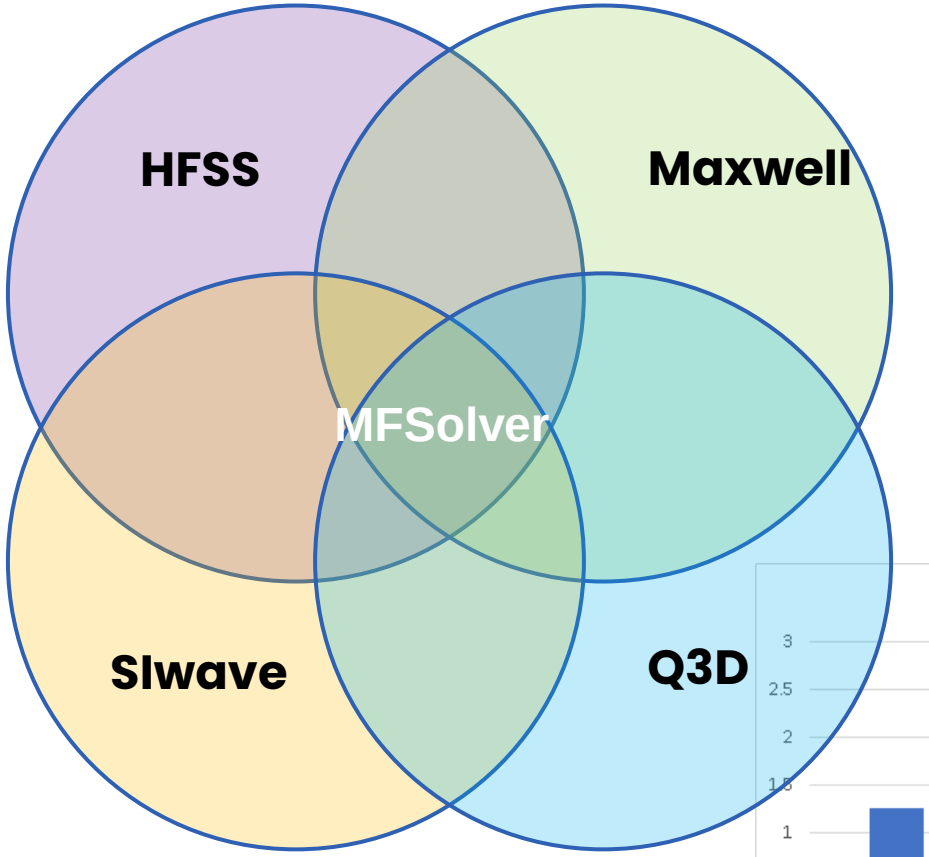
Ansys HFSS and Maxwell: GPU Acceleration in Offload Mode

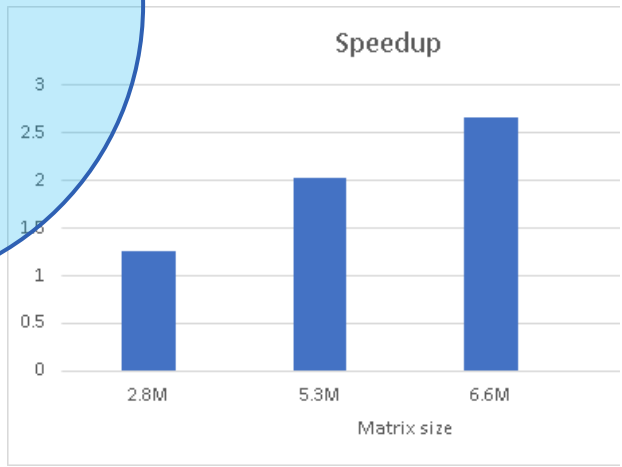
MFSolver

- High performance in-house 64-bit direct solver for sparse linear systems.
- Efficient multifrontal implementation with supernodes.
- Exploits parallelism through SMP, distributed memory, GPU.

Other GPU Functionality for EM:

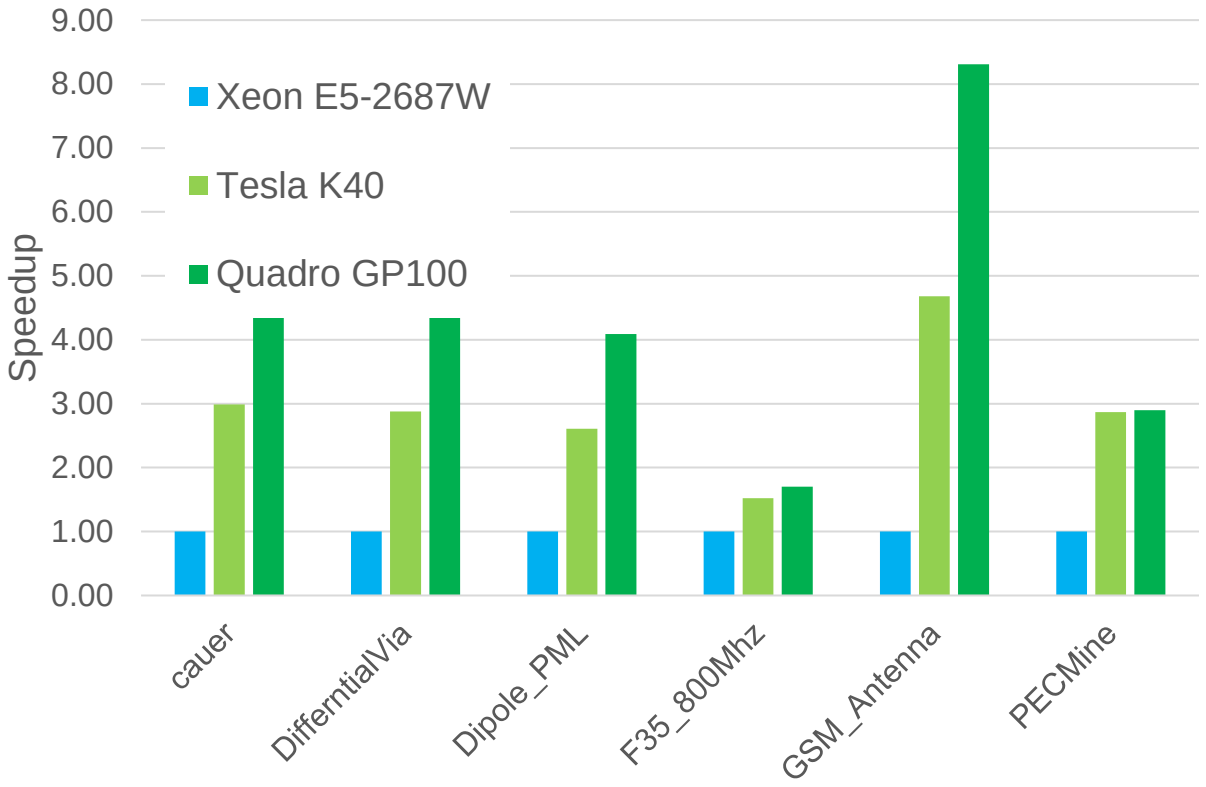
- HFSS transient DG-TD engine
- SBR+ ray tracing engine



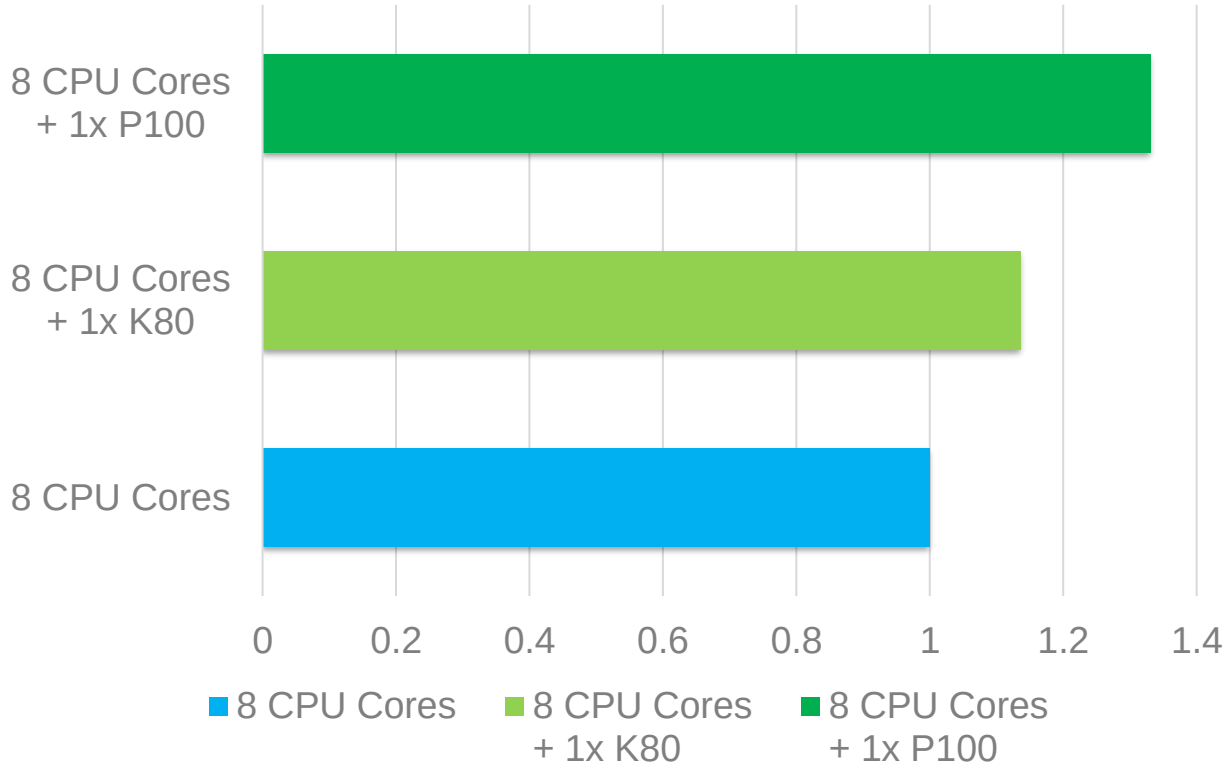


Ansys HFSS and Maxwell: GPU Acceleration in Offload Mode

Comparison of Ansys HFSS Transient entire simulation solution time (not just the calculations that are accelerated on the GPU)



Comparison of Ansys Maxwell relative speedups compared to base of 8 CPU cores only



Concluding Remarks

- Ansys software design and implementation aims to increase customer productivity by harnessing available HPC resources efficiently end-to-end in the simulation workflow
- Large dedicated Ansys HPC development teams and strategic partnerships with hardware partners and national labs are critical to success as the field evolves rapidly
- Methods are evolving rapidly to keep ahead of new hardware as released for consumer HPC, particularly GPU-centric computing

Jsme připraveni Vám pomoci vyřešit Vaše úlohy