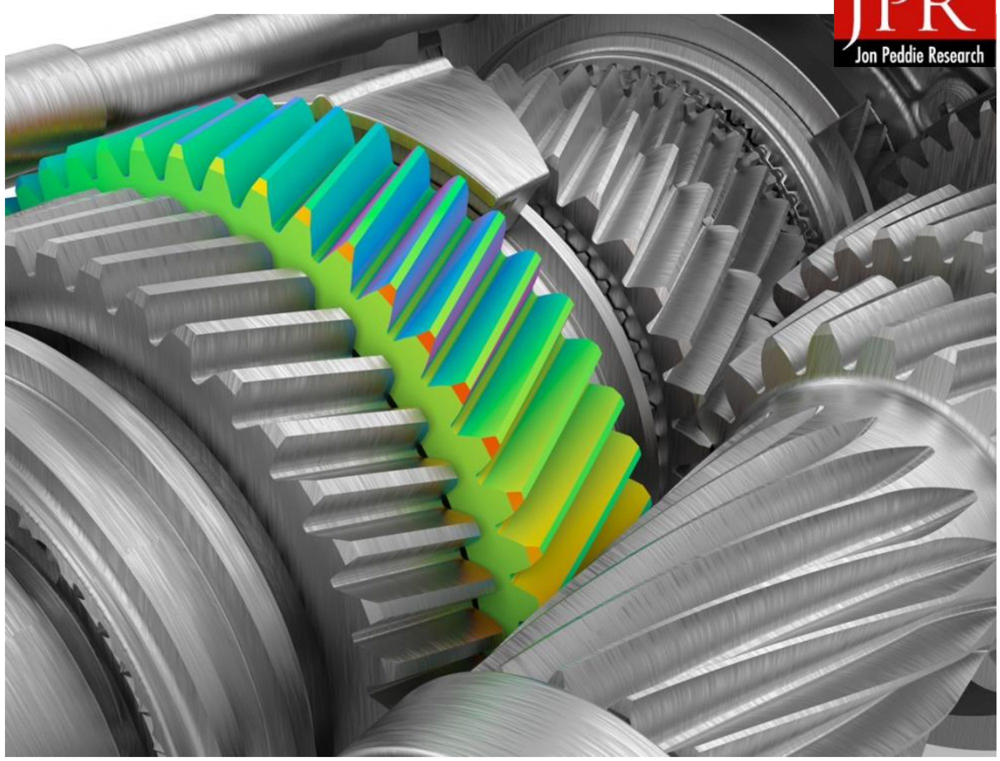
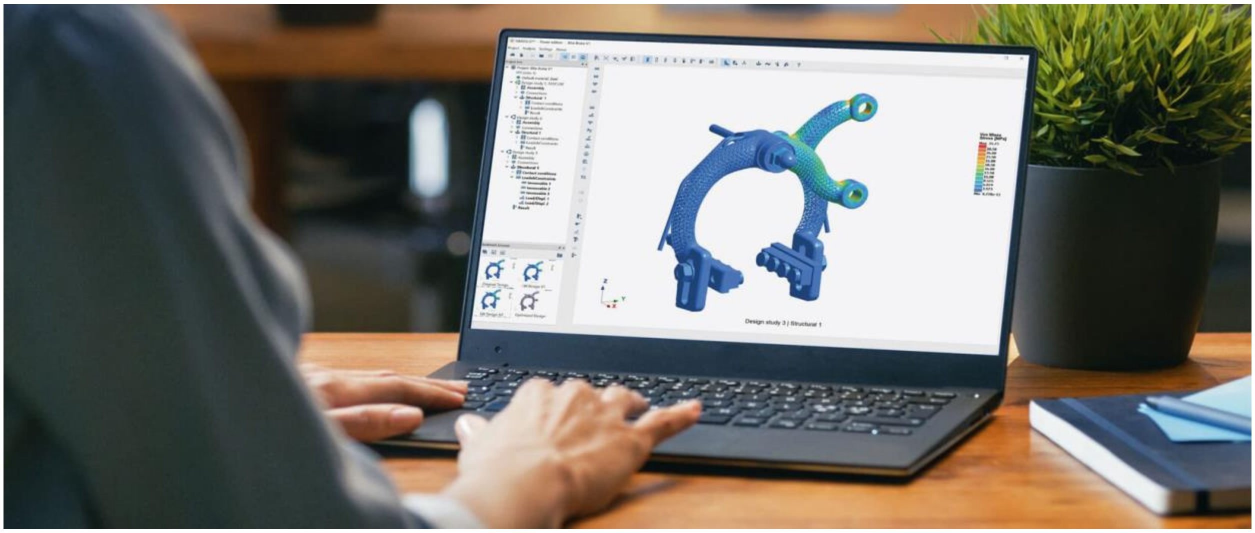
ACCELERATING AND ADVANCING CAE

GPUs change the equation of workstation CAE

By Kathleen Maher, Analyst, Jon Peddie Research



An eBook prepared by Jon Peddie Research for Nvidia



Workstations are running ever more demanding simulations as GPU acceleration becomes accessible. (Source: Altair)

Introduction

Computer-Aided Engineering (CAE) has gone through several significant phases from its origins in the ’50s as a technology developed by mathematicians, scientists, and engineers to improve design for the aeronautics and automotive industries. A resource-intensive technology, CAE has been a challenge waiting for a solution. CAE performance scales with every turn of the semiconductor process crank driving Moore’s Law, every advance in memory, and the multiplying parallel processing capabilities of CPUs and GPUs. Today, CAE is responding to advances in semiconductors and software that are fundamentally changing the landscape.

The major engineering simulation software providers including Altair, Ansys, Autodesk, Dassault Systèmes (Simulia), Hexagon MSC, and Siemens have relied on technology developed when CPUs were the primary engine driving computations. Originally models were simplified to accommodate the computing systems, but engineers wanted more realism and complexity. As a result, the compute time for these larger and more complex problems could take several hours, even days and weeks, of processing before getting results. This delayed the interpretation of the results, and as a result, helped define the way CAE was used. For example, CAE was often used after a product was manufactured to forensically determine the reasons for failure. However, the real value of CAE comes with its use much earlier in the design process. CAE is better used to help define designs rather than diagnose failure.

The arrival of graphics processing units (GPUs) in 2000 set the stage for a major transition in CAE as has happened in so many fields that benefit from parallel processing. The major jobs of CAE are massively parallel processes. CAE evaluates a model by creating a mesh of nodes on the model and then applies forces and conditions to the nodes to evaluate suitability of the design for its use. The denser the mesh, the more reliable the simulation. That's an obvious application for GPUs, and software vendors and hardware developers have recognized that early in the history of GPUs. Simulation approaches designed to run on CPUs have had to make allowances for the CPU's capabilities. They've had to reduce the size of the model, simplify the design, and manage the mesh size, and as a result, evaluate an entity that might be substantially different than the real thing being analyzed. GPUs have the advantage of having many more processing units on one chip than CPUs. In a comparison of processor to processor, GPU processors are much less expensive than CPU processors. So, if GPU processors are less expensive than CPUs, and GPUs are better suited to CAE workloads, then why haven't all software programs switched to GPUs?

The challenge is that GPUs and CPUs work in different ways, requiring specific programming methods for each. CAE is based on complex applications developed decades ago. Adapting those programs for GPUs is not an easy prospect, but programming tools such as Nvidia's CUDA and OpenCL have arrived to make it easier for developers to take advantage of GPU acceleration.

Transforming CAD

We have seen two major trends in CAE that began to take shape at the beginning of the 21st century. In the first decade of 2000, computer-aided design (CAD) software vendors began acquiring CAE technology to add to the design pipeline. The idea was to integrate CAE capabilities into CAD and automate the processes for designers. That work is ongoing, and so are the acquisitions. At the same time, the CAE community has also recognized the need for advanced CAE requiring specialized skill and resources. The CAE software vendors are investing in GPU acceleration and advancing high-performance computing (HPC).

Likewise, the semiconductor companies have been investing in tools to create on-ramps to their hardware for software developers. Intel and AMD are looking at opportunities for their CPUs and GPUs. Nvidia stands out with its early commitment and investment into CUDA in 2006. CUDA is a library of specialized code for GPU acceleration of applications. As part of that work, Nvidia has been working with CAE developers to create tools tailored to common tasks for simulation-analysis-visualization. The company is unique in its singular focus on the GPU, which has driven its entry into specialized tool development. In 2023, that focus is due to expand with the introduction of Nvidia's Grace CPU, which will leverage Arm processors as well as CPUs.

Between the CPU and the GPU, the paths of CAE diverge. The analysis tools integrated into CAD processes have a higher level of automation and are designed for simpler problems; they allow for the limitations of CPUs and have helped grow the market for CAE tools by making visualization and simulation tools accessible to designers and engineers. Those earlier simulation tools relied on CPUs because the CAD programs were written for the CPU. Today the CAD vendors are finding ways to take advantage of GPUs just like their counterparts in the simulation industry. The CAD vendors are building plug-ins and add-ons to enable GPU acceleration. In addition, they're adding cloud resources to provide advanced capabilities not available to the CPU-based core technology of their CAD flagships. These advances give workstation applications including analysis and simulation room to grow.

To cite one notable example, Dassault Systèmes has built their 3DExperience platform to unify Catia and SolidWorks design platforms. Design engineers working on the 3DExperience platform can perform structural analysis using Abaqus directly on their design and take advantage of the program's support for GPU acceleration without having to import the CAD model, perform meshing, run the analysis, and then update the design in a separate tool. The cloud-based CAD tools Autodesk Fusion 360 and PTC's Onshape can also take advantage of GPUs in the cloud for simulation extensions to the programs.

Although the opportunities for CAE applications working in mainstream design apps are growing, there will be a concurrent demand for high-end analysis requiring specialized practitioners. The ability for designers to perform simple analysis as they work clears the way for advanced simulation to further evolve for high-performance systems, be they workstations, HPC servers, or cloud-based systems.

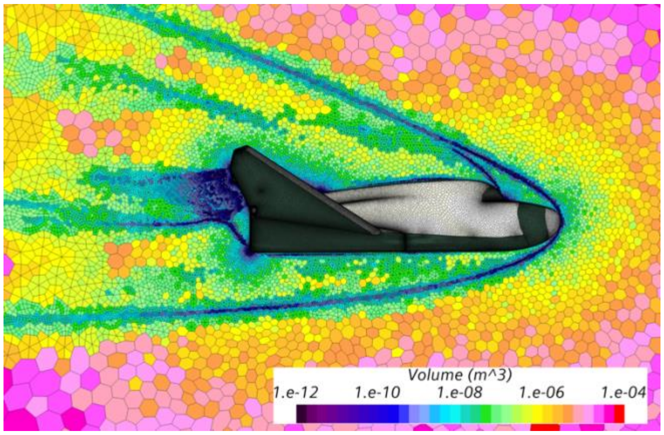
GPU tools evolve to accelerate every step

Traditionally, designers and engineers export a simplified model of the area they're interested in. They defeature the model as much as possible and add a mesh as a framework for the analysis. They also define materials and apply loads and constraints. These steps are part of the preprocessing stage of CAE.

Traditionally, the process was handled by experts, and it is often a time-consuming process. Simulation tools integrated in CAD programs often automate the process on simplified geometry. In more advanced implementations, the process may be a combination of manual and automated processes. The drive is for increased automation on workstations to avoid the challenges that lie in wait for inexperienced users and even for specialists who might miss an area that needs special attention.

There is also growing interest in the use of Machine Learning (ML) and Artificial Intelligence (AI) for preprocessing and identifying areas of concern in models.

When ready, the customized geometry created in preprocessing is sent to the solver.



CFD study on the space shuttle using Simcenter STAR-CCM+. Notice that automatic adaptive mesh refinement has been applied resulting in denser mesh on the area of interest around the model to better capture shockwaves, airflow, temperature, force, etc. (Source: Siemens)

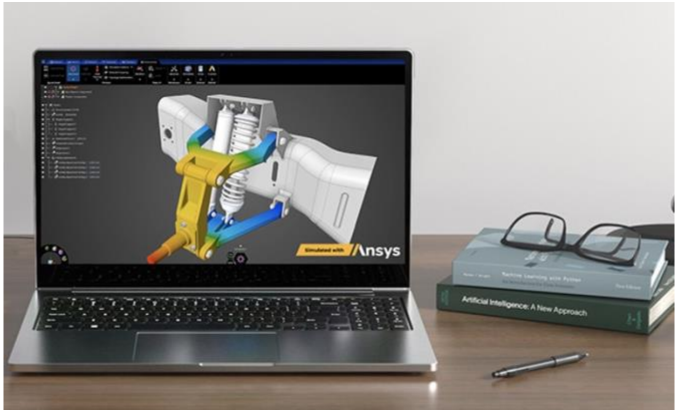
Once the mesh is defined, next comes the heavy lifting: solving. The algorithms for computational fluid dynamics (CFD), finite element analysis (FEA), and computational electromagnetic (CEM) simulations are run, and depending on the problem, can require some time—minutes, hours, or even days—depending on the application and workload.

Finding the best route to the GPU

Since 2014, every major CAE vendor has taken advantage of GPU acceleration in some way. A great deal of work has been done on the solver side of the equation because there are big gains to be realized. It's been estimated that linear solvers represent 50–70% of the work of CAE. For some software vendors, it has represented low hanging fruit in some cases because the process maps easily to GPUs. In other cases, the code must be rewritten.

For example, the AMG (Algebraic Multi-Grid) algorithm is a solver central to all CFD software tools. Nvidia and Ansys worked together to parallelize AMG and create AmgX libraries to make use of GPUs, and then further develop CUDA tools for CFD.

That work has led Ansys' development of a new product, Discovery, which is transforming simulation workflows in mainstream CAD tools, allowing iterative design exploration in fluid, thermal, structural, and modal applications.



Ansys Discovery was built for the GPU from the ground up.

Ansys fellow and vice president of research, Dipankar Choudhury, says the development of Discovery represents a break for Ansys from its traditional products, which were traditionally developed for the end of the design cycle. He says, “Discovery really pushes its way upstream in the design process. So, using Discovery technology, you can actually evaluate candidate designs, and you can also evaluate design concepts early on.”

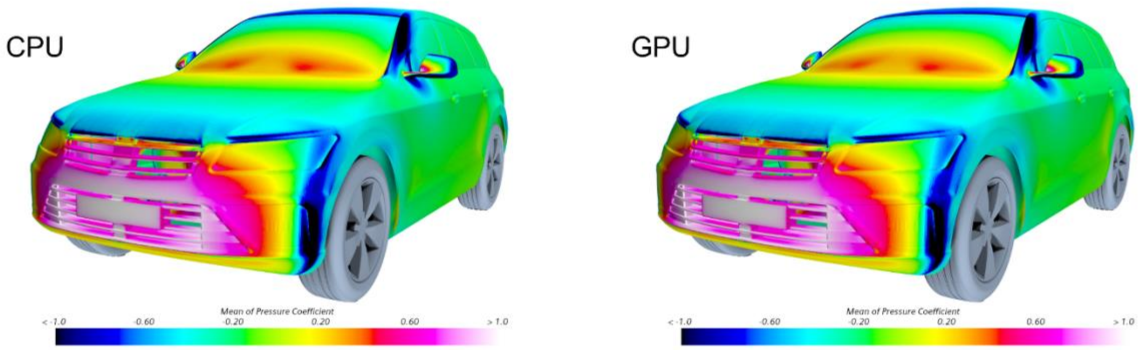
For Discovery, Ansys took the opportunity to design from the ground up for GPUs, rather than port a very large application like their Fluent CFD tool. “We don’t do porting, says Choudhury” The challenge simulation companies are dealing with as they make the transition to GPU acceleration is that the parts of the code that are written for the CPU may slow down performance overall. “So, we very consciously decided not to take that approach and write code software from the ground up on GPUs.”

Following up on the development of Discovery, PTC and Ansys have collaborated on an integrated version of Ansys Discovery, Creo Simulation Live, which enables designers to iteratively design and analyze.

Siemens didn’t rush into the GPU market but waited until they felt the investment would benefit their users. Their approach to moving to GPUs was based on the assumption that their customers would be able to seamlessly move from CPU to GPU and vice versa without any change to their simulation environment and results. With C++ tooling maturity and the performance improvements that came with the arrival of Nvidia’s Volta architecture, the time felt right.

Siemens’s CFD software Simcenter STAR-CCM+ is a large codebase but the majority of the changes were handled on the internal framework level. The developers at Siemens modified the framework to support a single codebase that compiles for both CPUs and GPUs architectures uniformly with the internal details abstracted out. To kick start this effort and deliver their first GPU version in Simcenter STAR-CCM+ 2022.1 by early 2022, they relied on AmgX from NVIDIA. For this version, they focused on vehicle external aerodynamics applications because that work requires less porting of physics models and their associated frameworks, but comes with a large compute overhead, making parallelization a necessity and GPU acceleration very attractive. Subsequently, Siemens engineers are heavily investing in porting all physics, solvers and associated software parts that can benefit from GPU in the coming years.

The Simcenter STAR-CCM+ lead product manager for Physics Stamatina Petropoulou says that “at Siemens they believe that heterogeneous architectures are the future. In search for the best cost-to-solution for any given simulation project, CFD engineers will want to switch from CPU to GPU, from on-premises to cloud or even leverage the best of all worlds in one single simulation run.” She pointed out that “for the Simcenter STAR-CCM+ customers that have spent years validating the software consistency of results, the ability to seamlessly switch between CPU to GPU is key!” so they made sure that their customers can use the same code and get equivalent results on CPU or on GPU (see image).



Simcenter STAR-CCM+ calculation of the mean of pressure coefficient compared between (left) CPU- and (right) GPU-based runs. Vehicle: Aero SUV, FKFS Stuttgart https://www.ecara.org/drivaer-1

In the case of Hexagon, the company decided to write their new product MSC Apex Generative Design for the GPU from the start, says Hugues Jeancolas, Hexagon VP of product management at MSC Software. He says the result is an application that ran with an Nvidia GPU using calculations that previously required expensive clusters to run. The CUDA framework provided an easy entry point for Hexagon’s developers, who were able to start coding right away. The MSC development team used MSC Apex Generative Design, CUDA, CuBLAS, and CuSPARSE to enable GPU acceleration in their generative design application.

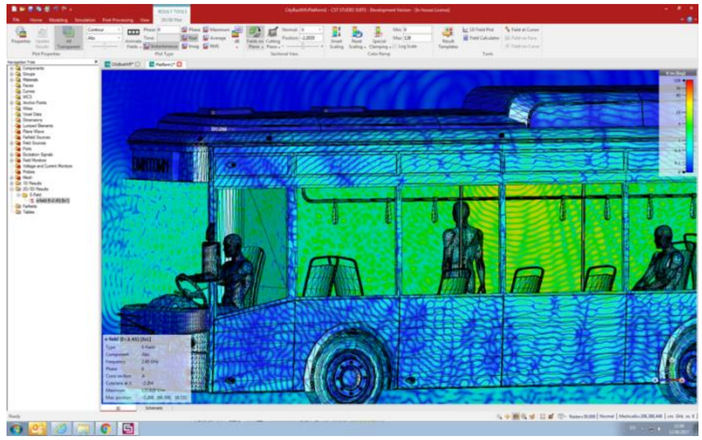
Hexagon built their product MSC Apex Generative Design from the ground up using GPUs. Not only is the finished product faster, but it also combines the functions of design, meshing, and analysis in one product. (Source: Hexagon)

At the end, MSC Apex Generative Design outputs a fully manufacturable model for additive manufacture machines without the need for manual rework. Jeancolas estimates that the computing time for one optimization run of a standard component can be reduced by a factor of five, comparing CPU and GPU on a standard workstation. He says the numbers go up as GPUs are added or more advanced GPUs are used.

Similarly, Hexagon’s engineers have been able to rework their Actran DGM (discontinuous Galerkin method) acoustic application to take advantage of GPUs. Actran DGM is used to predict noise propagation in complex physical conditions. It is widely used in the automotive industry. Jeancolas says that about 10 years ago, in the early days of GPGPU (general-purpose GPU) applications, the Actran DGM team was able to move a large part of their code to the GPU. That code, based on CPUs, was massively parallel and required thousands of cores. By porting it to the GPU, not only did the team make the code much more efficient, but it was much cheaper to run given that CPUs cores are much more expensive than GPU cores.

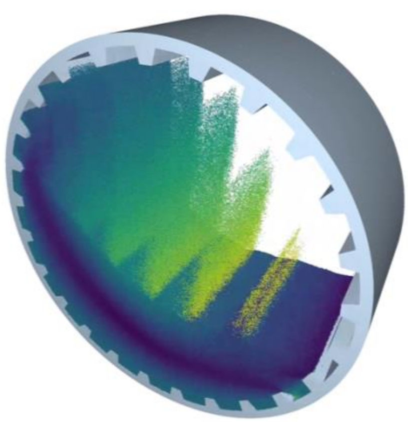
Today, Actran DGM has been optimized to take further advantage of GPU acceleration. Actran DGM’s performance scales with the number of GPUs. Performance also depends on the amount of memory capacity available on the board and the performance of each individual core and the number of cores—conditions that scale with the introduction of new graphics boards and GPUs.

Electromagnetic analysis using Simulia CST Studio Suite is used to evaluate the performance and efficiency of antennas and filters. The technology is used to determine electromagnetic compatibility and interference (EMC/EMI) and gauge the exposure of the human body to EM fields. (Source: Dassault Systèmes Simulia)



In these early days of adapting CAE programs for GPU acceleration, electromagnetic analysis has emerged as an early beneficiary of GPU acceleration. In 2016, Dassault Systèmes acquired CST (Computer Simulation Technology), a German company that specializes in electromagnetic simulation and analysis software to expand its Simulia-brand Multiphysics portfolio.

Peter Hammes, Simulia R&D Electromagnetics Technology senior director at Dassault Systèmes, says CST Studio Suite's technology, based on the finite-difference time-domain simulation algorithm, is well suited for GPU architectures. It also benefits from large GPU memory and memory bandwidth and scales very well from workstation GPUs to datacenter compute GPUs such as Nvidia A100. The CST team credits Nvidia's CUDA libraries with making the development of new projects from the ground up a much easier proposition.



20 million particles in a mill running on EDEM multi-GPU solver and colored according to their velocity. Each particle and their collisions are tracked independently by the EDEM solver, and GPU technology enables both faster computational performance as well as increased scale of problem that can be solved. (Source: Altair EDEM)

Altair acquired EDEM in 2019 and as a result has added discrete element modeling to its large portfolio of solver technology. EDEM can work as a stand-alone tool or be combined with other CAE tools including CFD solvers for simulations related to the behavior of particle-based materials. It is used to model the behavior of coal and other ores, soil, fibers, grains, etc. Working with particle-based simulations is already an inherently massively parallel proposition.

“The actual process of development went quite smoothly, and customers are seeing a big payoff in performance increases—up to 20x versus the CPU-based process,” says Altair SVP CFD/Vice President David Curry. “To be more specific, the addition of a GPU improved the performance of EDEM 20x compared to 12 CPUs going to work on a similar workload.” Altair says their latest EDEM multi-GPU solver enables even larger industrial problems to be solved—with many millions of particles—and offers performance scalability as additional GPU cards are added.

According to Curry, the deployment of large-scale GPU systems has increased enthusiasm for artificial intelligence applications. AI has helped increase the availability of cloud-based GPUs by the major cloud service providers, which in turn has fueled the solid growth of GPU-accelerated tools and applications.

Altair’s engineers started supporting GPUs as the tools for GPGPU development became available. They credit Nvidia’s continuous technology updates, developer tools, and support with making the process easier.

To date, over 120 CAE applications from more than 10 ISVs have been accelerated by GPUs. The results have been impressive, with results being delivered up to 100 times faster depending on the application and workload. Furthermore, as GPUs are added, the performance improvements scale up. There will be considerably more breakthroughs coming as more solvers are ported to GPUs.

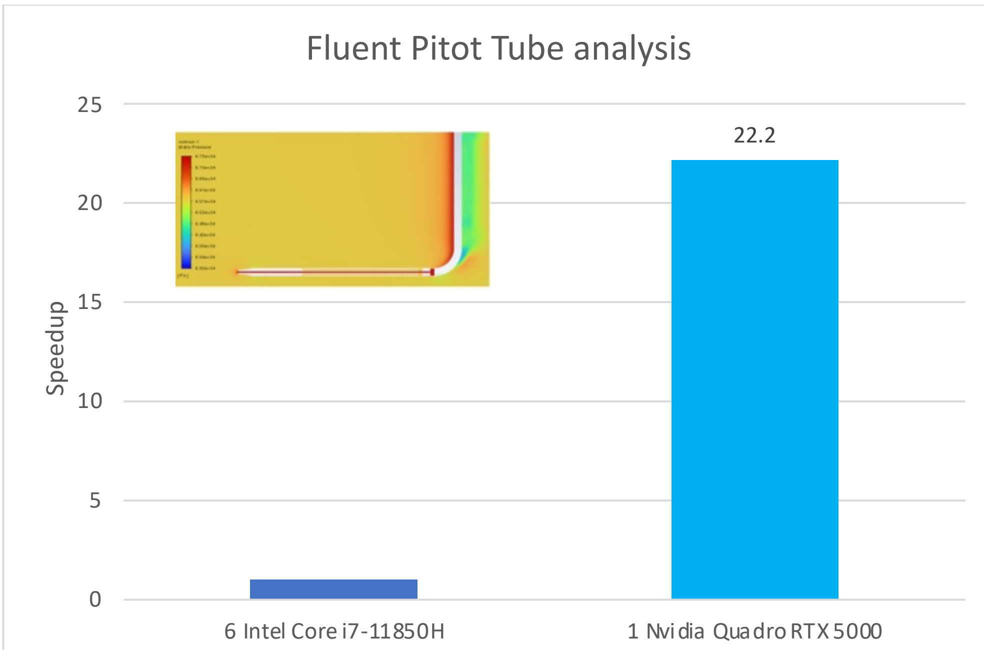
Getting to the customer

The human factor comes into play in the final phases of CAE, which includes visualization, analysis of the results, and reporting. The CAE applications deliver results with visualizations and related data. Analysts evaluate the results and will decide how to proceed—with more analysis or back to the model for adjustments as suggested by the analysis.

Analysis is usually a separate process from design even when the analysis process is initiated from within a CAD model. As we’ve outlined, the geometry is optimized, meshed, analyzed, visualized, and evaluated. The findings are then taken back to adjust the design geometry for more analysis or to move the design forward.

In situations where advanced analysis is used, the process is even more disjointed because the analysis itself may be an external process performed outside the design group and even by a different company. “The State of Simulation, Prototyping, and Validation,” a recent study written by Roopinder Tara of Engineering.com for Dassault Systèmes documented a similar workflow. They surveyed 268 engineers and found that the actual practice of CAE lags the potential of today’s computing machines and software. Physical prototypes are still being made and designs are still being sent to outside experts for analysis, which may take days to turn around.

The Ansys team that developed Discovery believe they're forging a new path for analysis. Justin Hendrickson, Ansys' director of product management, says, "We're a little bit in the mode 'if you build it, they will come.'" Dipankar Choudhury agrees and says they're seeing changes in the way Discovery is being introduced in companies and how it is being taught in schools. Instead of being reserved for research departments and for graduate students, Discovery is making its way directly to end users and taught by early education instructors.

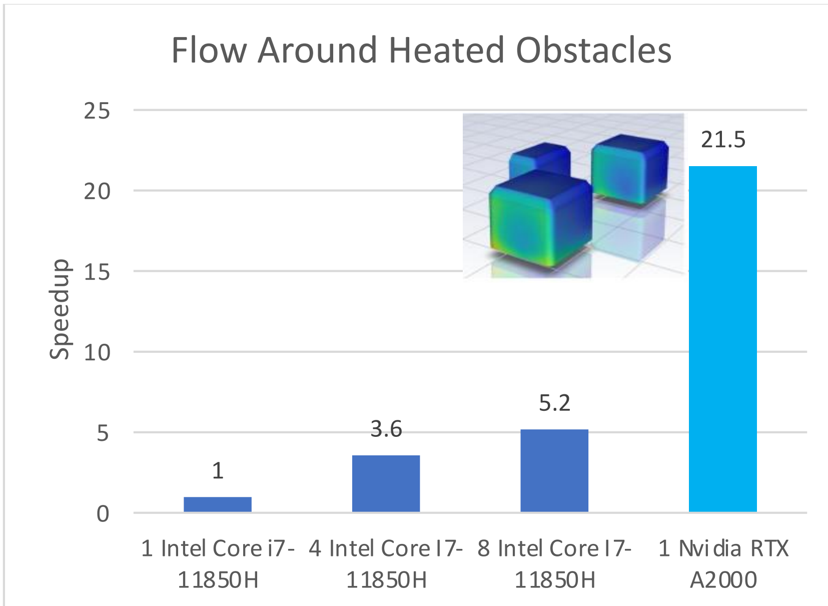


An example of a simulation using Ansys Fluent with a simplified model of a Pitot Tube run on a workstation using one Quadro RTX 5000 GPU compared to six Intel Core i7-11850H CPUs. (Source: Ansys)

The numbers also support GPU acceleration throughout the design process. The work the Discovery team did to create Discovery also informed their work on Fluent. The CFD portion of Discovery also formed the basis for Ansys' Fluent.

Choudhury notes that as GPUs began to come into play for advanced compute processes like CAE, "The traditional computing community always had the assumption that the methods we were choosing were lower in accuracy, or that we were making compromises in the numerical methods to have them run efficiently on GPUs." That was not the case, says Hendrickson. They're seeing the same level of accuracy, "which is amazing and not really expected," he adds. "When we look at some of the benefits of GPUs, one is speed; but what does speed mean?

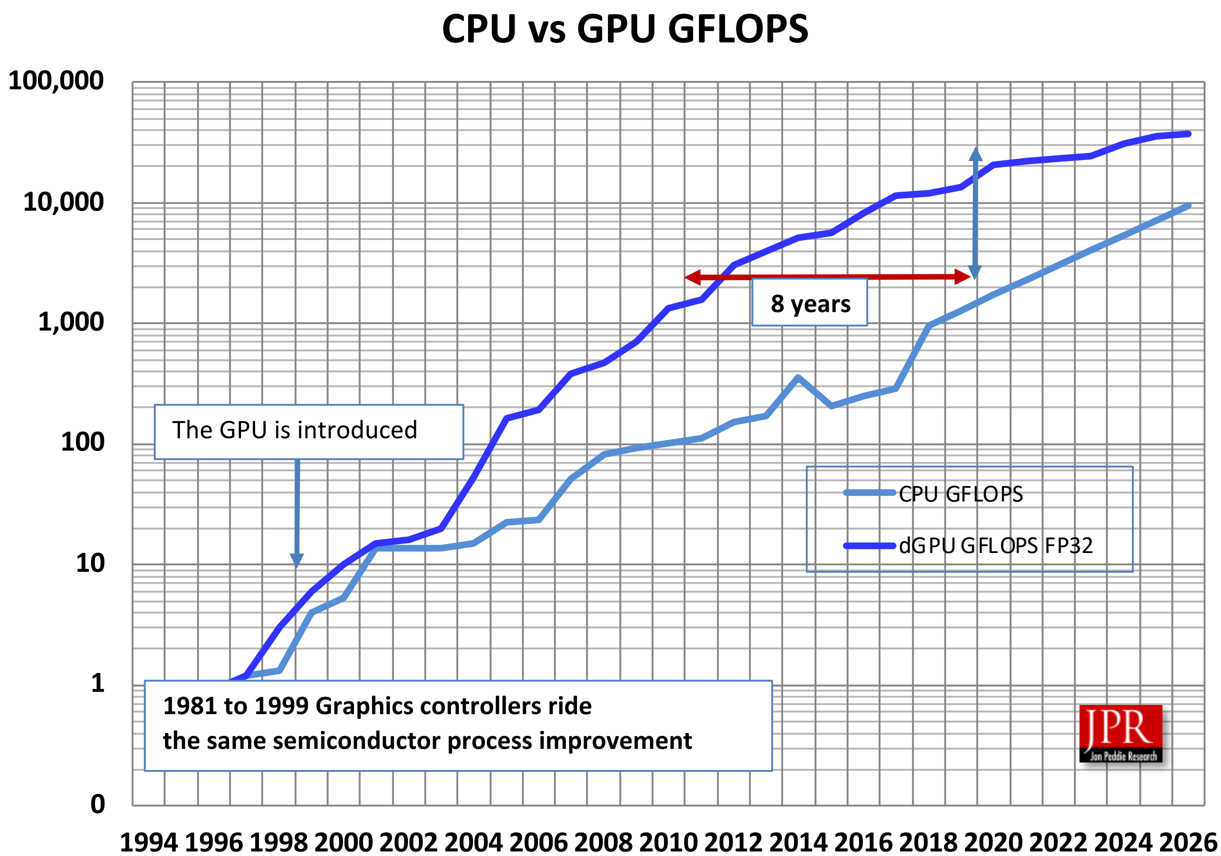
We're really talking about price per dollar, performance per dollar, or performance per watt." He says many of Ansys' customers care about sustainability; they care about minimizing energy, and it's really these two measures where we see 10X and sometimes even larger increases."



As one of the most challenging applications in simulation, CFD is a good showcase for GPU acceleration. Ansys' latest version of Fluent runs natively on GPUs, and the speedup is significant. (Source: Ansys)

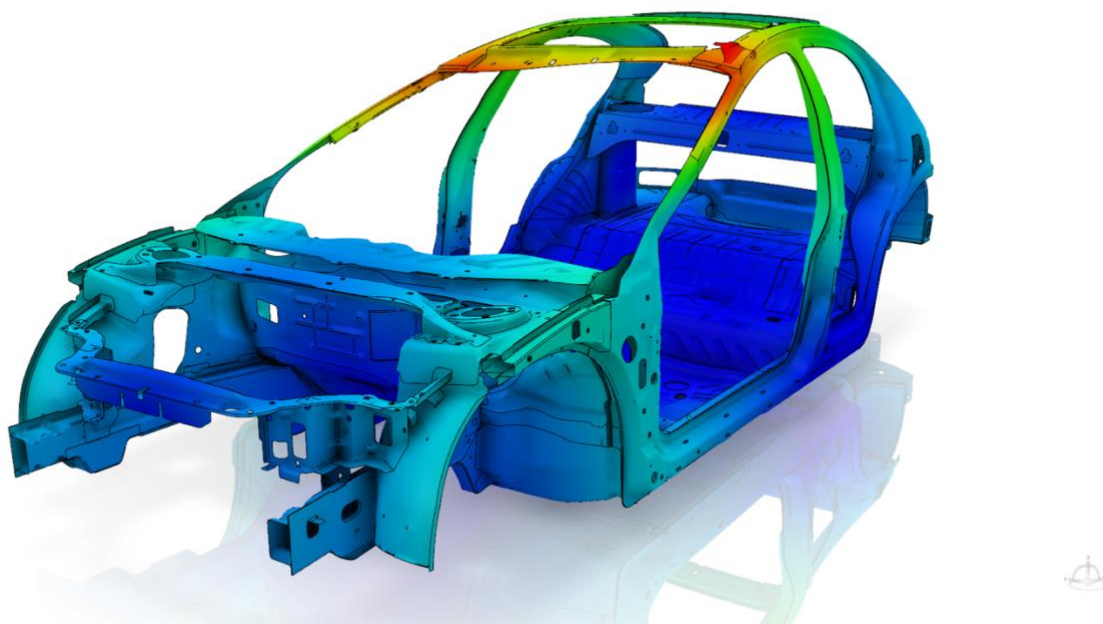
Ansys' scientists have calculated that in some cases, four GPUs can outperform over a thousand CPU cores for certain CAE applications, at 16% of the cost and with one-fourth less power consumption. The company has published a blog that discusses the scalability of GPUs for CFD applications. The blog also highlights the attractive cost differential between CPUs and GPUs. On the desktop, performance in CAE performance can scale with each new generation of processor as new semiconductor manufacturing processes enables more processors per core.

Nvidia GPUs—measured in cores or floating-point operations per second (FLOPS)—outpace the CPU. Mesh processing and solver programs are ideally suited for parallel processors because of the redundancy of the algorithm and the quantity of data.



Comparison of computing power of a CPU and a GPU. The GPU offers an eight-year lead over CPUs in performance. (The gap between CPUs and GPUs narrows in the future due to the surge in CPU cores enabled by new architectures. But, because CPUs suffer greater thermal issues as cores increase, GPUs will always have a processing edge.)

Dassault Systèmes chief architect for Simulia Matt Dunbar says that Simulia Abaqus, an FEA tool primarily for structural simulation, has benefitted from the fast evolution of GPU processors. In the past, he says, "anytime you added compute power to the code, people just made their problems bigger." That trend is no longer a hard and fast rule as customers welcome the ability to perform more iterations as well as tackle challenging problems. He notes that modern workstations can have a lot of memory, and a lot of cores. Now, he says, "adding a GPU or two GPUs into a workstation really turbocharges the workstation, it extends the scope of the workstation, and doesn't force you to go to an HPC." GPU acceleration is beneficial for several Abaqus features, including the use of the AMS Eigensolver, which offers a performance boost for natural frequency extraction in full vehicle models that have a very large number of nodes. GPU compute resources are also beneficial for Abaqus Steady State Dynamics and computationally intense Direct Sparse Solver operations.



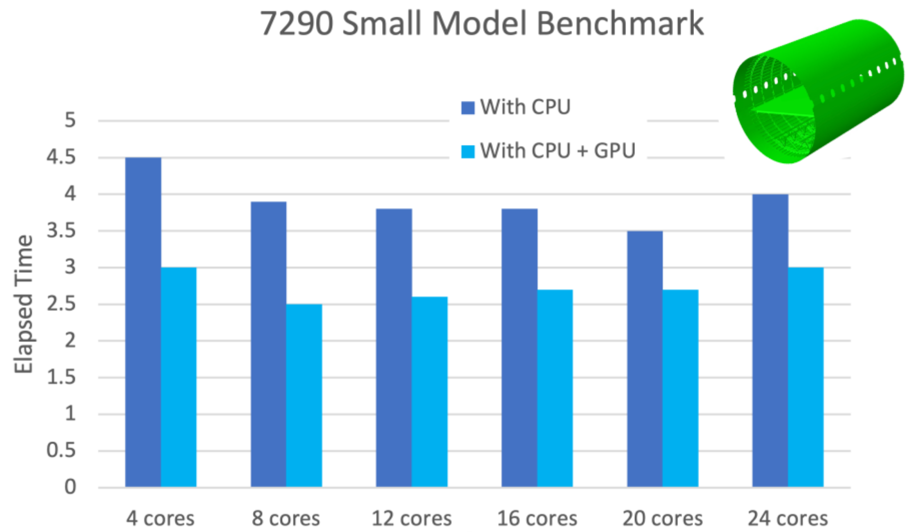
Abaqus has several features that benefit from GPU acceleration. For instance, Abaqus gets a performance boost when the AMS Eigensolver comes into play for full vehicle models with a large number of nodes. (Source: Dassault Systèmes)

As workstations become more capable of advanced compute processes, their role in the design workflow is changing. Advanced capabilities are also changing the role of the people working on those computers who are now about to do more with the machines on their desk.

How much more?

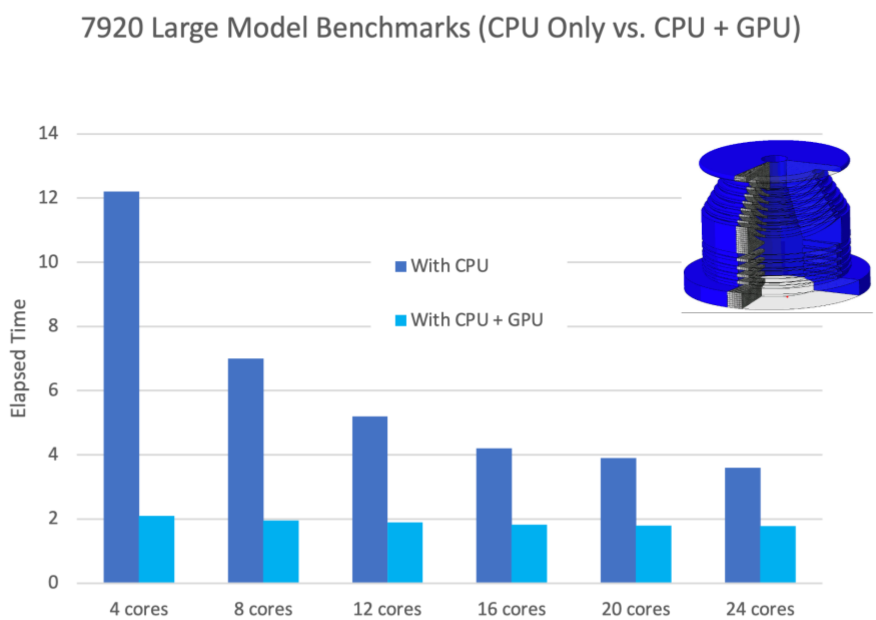
Clearly, GPUs speed up CAE processes compared to similar applications being run on CPUs. Simply, there are more GPU processors in a single chip than there are processors in a CPU semiconductor—it's an order of magnitude but it's not a pure numbers game. The application must be optimized for the GPU, the hardware platform needs to be optimized for simulation, and every problem has its own demands. As mentioned throughout this article, some applications such as electromagnetic simulation, particles, and acoustical simulation lend themselves well to GPUs. Also, GPUs love large problems, which is why we're seeing Ansys and Siemens coming to market with revamped CFD applications. The larger the problem, the more efficient the simulation.

To gauge improvements on small and large model simulations, Dassault Systèmes worked with Dell and *Digital Engineering Magazine* to test the 2021 Dell Precision 7920 Tower with dual Intel Xeon Gold 6146 3.2 GHz GPUs, and Nvidia's Quadro GV100.



Dell and Dassault Systèmes sponsored productivity studies of Simulia Abaqus software running on Dell Precision workstations. The studies ran and compared simulations using CPUs only and CPUs plus a GPU. Adding a GPU resulted significant performance boosts. (Source: Dassault Systèmes)

The cost benefit is obvious. GPU processors are cheaper than CPUs. That goes for GPUs in the cloud, in a cluster, on the desktop.



In this case, Abaqus, using a more complex simulation approach, shows a larger gain with the addition of a GPU and with CPUs. (Source: Dassault Systèmes)

In addition to fast CPUs and fast GPUs, CAE workloads benefit from large caches and wide bandwidths. It's important that memory not be a bottleneck, so in many cases CAE workflows are being moved to the graphics board, which is why we're seeing professional graphics boards arrive with plenty of high-speed memory.

Ansys has said in their blog that the benefits of putting GPUs to work on CAE workloads are:

- Increased performance.
- Reduced hardware costs.
- Reduced power consumption.

The companies profiled in this eBook have also demonstrated that the use of GPUs has, in many cases, enabled workstations to take over the role formerly played by HPC clusters. In an upcoming eBook, we will also demonstrate how HPCs are stepping up to new challenges thanks to the performance boost brought by GPUs.

Conclusion

If it were not obvious already, the experience of the developers we've talked to clearly demonstrates that GPUs offer a performance edge and cost advantage over CPUs. The wide and growing availability of GPUs optimized for rendering, CAE, AI/ML, video editing, and gaming ensures that almost every capable workstation system is equipped with a GPU, and usually a very capable GPU. In addition, the GPU's role in off-loading the CPU in these resource-hungry applications like simulation ensures the entire system is more efficient.

For all these reasons, GPU acceleration, once a nice-to-have component for design and engineering, is becoming a linchpin for companies interested in advancing their practices. The CAD industry is changing rapidly as the idea of digital twins captures the imagination of the industry. We expect to see the technology improvements enable faster, cheaper iterations on the desktop and more sophisticated analysis (and bigger geometries) taking advantage of powerful HPC appliances. The sky's not the limit and neither is the sea; whole digital worlds are being built in the very near future, and they are being tested, analyzed, and measured by GPUs.

For information on the tools available for GPU acceleration in CAE, see Nvidia's expanding list of tools in its catalog. GPU acceleration is allowing more iterations for faster error correction and better analysis, and as advances continue, analysts can tackle bigger problems.