HPC APPLICATION SUPPORT FOR GPU COMPUTING

Michael Feldman   **Christopher G. Willard, Ph.D.**

Addison Snell

Sponsored research report

November 2015

EXECUTIVE SUMMARY

Since 2007, when GPU computing was introduced to High Performance Computing (HPC) users, the technology has advanced rapidly. Demand for computational GPUs has been driven largely by HPC practitioners looking for greater performance and improved energy efficiencies. The former has been a fundamental driver in the HPC market since its inception, while the latter is a more recent demand that has come about as a result of the breakdown of Dennard scaling. Thanks to GPUs' throughput-optimized design and considerable floating point capability, GPU computing has been demonstrated to provide substantial speed improvements over unoptimized x86 architectures, in the range of 2x to 10x, to HPC application codes in which vector or SIMD parallelism is applicable. In these cases, GPU acceleration also generally exhibits strong performance per watt characteristics. GPU technology has found acceptance in the HPC market, with nearly half of all new HPC systems deploying some level of the technology.[1]

The growth of GPU adoption for HPC has been driven almost entirely by NVIDIA, which has invested heavily in building a robust software ecosystem to support its hardware. Specifically, the company has developed a set of parallel programming APIs, libraries, and associated software development tools to support application development on its CUDA (Compute Unified Device Architecture) GPU platform. NVIDIA has also established over 20 GPU centers of excellence around the world, as well as a global network of GPU research and educations centers encompassing hundreds of academic institutions. Those efforts have supported the acceleration of software tool development for general-purpose GPU programming, in addition to aiding the porting of open-source HPC codes to GPUs. The company has also collaborated with ISVs to help accelerate commercial HPC applications and libraries. Today, hundreds of such codes are now available, promising speed-ups of anywhere from 10 to 20 percent up to more than 1,000-fold. For the most commonly accelerated codes, performance increases are usually in the 2x to 10x range.

In this report, Intersect360 Research has identified the commonly used HPC applications that currently have GPU acceleration incorporated. According to the latest HPC User Site Census data and additional research, of the 50 most popular application packages mentioned by HPC users, 34 offer GPU support (including two under current development), including 9 of the top 10. We believe GPU computing has reached a tipping point in the HPC market that will encourage continued increased in application optimization.

---

[1] HPC User Site Census: Processors. Intersect360 Research, October 2015.

TABLE OF CONTENTS

Table 1. GPU-Accelerated Applications

Software Supplier - Package Name	Mentions	Rank	GPU Accelerated
Gaussian – Gaussian	118	1	Under Development
ANSYS – Fluent	114	2	Accelerated
Gromacs.org – GROMACS	93	3	Accelerated
Dassault Systemes - SIMULIA Abaqus	91	4	Accelerated
U of Illinois, UC – NAMD	75	5	Accelerated
NCAR – WRF	72	6	Accelerated
U of Vienna – VASP	68	7	Accelerated[2]
OpenFoam Foundation – OpenFOAM	54	8	Accelerated
LSTC - LS-DYNA	53	9	Accelerated
AmberMD.org – AMBER	50	10	Accelerated
NCBI – BLAST	46	11	No
PNNL – NWChem	35	12	Accelerated
Iowa State University – GAMESS	30	13	Under Development
Quantum-espresso.org - Quantum ESPRESSO	30	14	Accelerated
Sandia Nat Lab – LAMMPS	28	15	Accelerated
ANSYS – CFX	17	16	No
Exelis – IDL	16	17	Accelerated
MSC Software – NASTRAN	16	18	Accelerated
CD-adapco - Star-CD	15	19	No
Schrodinger – Schrodinger	14	20	Accelerated
NCAR – CCSM	14	21	No
COMSOL – COMSOL	14	22	No
ANSYS - ANSYS Mechancial	12	23	Accelerated
Charmm.org – CHARMM	11	24	Accelerated
CD-adapco - Star-CCM+	10	25	No
CP2K.org - CP2k	10	26	Accelerated
Thermo Scientific – Sequest	9	27	No
Kitware – ParaView	9	28	Accelerated
USQCD.org – MILC	8	29	Accelerated
CERN – CMSSW	8	30	Accelerated
Scripps Research Institute – AutoDock	8	31	No
Open Source – MrBayes	8	32	No
Q-Chem - Q-Chem	8	33	Accelerated
LLNL – VisIT	8	34	Accelerated
Terraspark - Insight Earth	7	35	Accelerated
Open Source – Galaxy	7	36	No
Illumina – Casava	7	37	No
CMOP – SELFE	7	38	No
Tecplot, Inc. – Tecplot	6	39	Accelerated
SCM – ADF	6	40	Accelerated
USQCD.org – CHROMA	6	41	Accelerated
Dassault Systemes - Accelrys Materials Studio	6	42	No
Abinit.org – ABINIT	6	43	Accelerated
NASA – Overflow	6	44	No
CIG - Specfem3d	6	45	Accelerated
Kitware – VTK	6	46	Accelerated
Metacomp - CFD++	6	47	No
Open Source – Nbody	6	48	Accelerated
Altair Engineering – HyperWorks	6	49	Accelerated
U of Massachusetts – FVCOM	5	50	No

---

[2] GPU implementation of VASP is beta complete; formal release expected soon.

RESEARCH AND ANALYSIS

High Performance Computing (HPC) spans a range of application domains in both the technical and business realms. For the purposes of this report, they are organized into major application domains that reflect specific user communities. GPU-accelerated applications are present in the majority of these domains, including chemical research, fluid dynamics analysis, structural analysis, environmental modeling, geophysics, visualization/image processing, and physics. This paper describes how these accelerated codes fit into the application domains and their significance to their respective user communities.

According to the latest HPC User Site Census data and additional research, of the 50 most popular application packages mentioned by HPC users, 34 offer GPU support (including two currently under development), including 9 of the top 10. The complete distribution is given above in Table 1.

Chemical Research

Chemical research employing HPC systems is used to understand chemical compounds and processes, to design new substances based on this understanding, and to organize this information for future research and product development. This segment includes such applications as computational chemistry, molecular modeling, and software used to analyze mass spectrometry data.

Approximately one-fifth of all application mentions in the Site Census surveys conducted by Intersect360 Research are in the chemical research category. The popularity of this software is being driven by the intense interest in biomolecular research and in developing new compounds and materials for industrial use, spanning multiple economic sectors. Products based on better solvents, increased battery performance, advanced construction materials, and more efficient fuels are just a few examples of commercial applications that can be derived from these packages.

As is illustrated in Table 2, GPU support for the top chemical research codes is quite deep, thanks in large part to a preponderance of open-source codes in this application domain and the presence of active developer communities. As a result, 11 of the top 14 applications are currently available with GPU acceleration, with two applications in the list under development.

Table 2. GPU-Accelerated Applications: Chemical Research

Software Supplier - Package Name	Mentions	Rank	GPU Accelerated
Gaussian – Gaussian	118	1	Under Development
Gromacs.org – GROMACS	93	3	Accelerated
U of Illinois, UC – NAMD	75	5	Accelerated
U of Vienna – VASP	68	7	Accelerated[3]
AmberMD.org - AMBER	50	10	Accelerated
PNNL - NWChem	35	12	Accelerated
Iowa State University - GAMESS	30	13	Under Development
Sandia Nat Lab - LAMMPS	28	15	Accelerated
Schrodinger - Schrodinger	14	20	Accelerated
Charmm.org - CHARMM	11	24	Accelerated
CP2K.org - CP2k	10	26	Accelerated
Q-Chem - Q-Chem	8	33	Accelerated
SCM - ADF	6	40	Accelerated
Dassault Systemes - Accelrys Materials Studio	6	42	No

---

[3] GPU implementation of VASP is beta complete; formal release expected soon.

For example, GROMACS, a popular open-source molecular dynamics code, aimed primarily at simulating biological molecules, enjoys typical speed-ups of three to five times when CUDA-based GPU acceleration is in effect, according to the GROMACS web site.[4] NAMD is a molecular dynamics package that is used especially for modeling large structures such as viruses and ribosomes; its GPU-accelerated version was used to reveal the capsid structure of HIV-1. AMBER, another popular molecular modeling code, offers a CUDA implementation that is used in hundreds of labs.[5]

Fluid Dynamics Analysis

Solutions based on computational fluid dynamics (CFD) appear across an exceptionally broad range of applications, such as: aircraft design, internal combustion engine design, analysis of airflow around buildings, heart pump design, modeling of protein transport within cells, and so on. Similar to structural analysis, it is an area largely dominated by commercial codes, but with some originally coming from NASA development and open-source efforts. The emergence of OpenFOAM as a top application in this category is indicative of a trend toward increased use of open-source codes.

As is evident from Table 3, the two most popular CFD applications, ANSYS Fluent and OpenFOAM, are now available with GPU support. ANSYS Fluent, the top-ranked commercial code, and the second-most popular HPC application overall, has been shown to deliver better performance when NVIDIA's GPU acceleration library for Fluent is invoked. For example, adding GPU acceleration in a Formula 1 car simulation decreased the time-to-result by a factor of 2.1, according to ANSYS.[6] Similarly, OpenFOAM, the leading open-source CFD package, offers a CUDA GPU-accelerated linear solver that can deliver a 2x or better performance boost on applications.[7] This reflects similar speed-ups attained by CFD researchers with GPU implementations of Navier-Stokes models and Lattice Boltzman methods.[8]

Table 3. GPU-Accelerated Applications: Fluid Dynamics Analysis

Software Supplier - Package Name	Mentions	Rank	GPU Accelerated
ANSYS - Fluent	114	2	Accelerated
OpenFoam Foundation - OpenFOAM	54	8	Accelerated
ANSYS - CFX	17	16	No
CD-adapco - Star-CD	15	19	No
CD-adapco - Star-CCM+	10	25	No
Tecplot, Inc. - Tecplot	6	39	Accelerated
NASA - Overflow	6	44	No
Metacomp - CFD++	6	47	No

Structural Analysis

Structural analysis includes applications used to analyze a variety of structures, including explicit and implicit finite element analysis (FEA), an important computational tool used for engineering analysis. It is primarily used to determine loads and stresses on structures.

---

4 http://www.gromacs.org/GPU_acceleration
5 http://ambermd.org/gpus/
6 http://www.ansys.com/staticassets/ANSYS/staticassets/resourcelibrary/article/Accelerationg-ANSYS-Fluent-Simulations-with-NVIDIA-GPUs-AA-V9-I1.pdf
7 http://www.fluidyna.com/content/culises
8 http://www.nvidia.com/object/computational_fluid_dynamics.html

Table 4 delineates the top structural analysis packages that have GPU support incorporated, in this case, five out of the top six packages, including the three most popular ones in this domain: Abaqus, LS-DYNA, and NASTRAN.

Table 4. GPU-Accelerated Applications: Structural Analysis

Software Supplier - Package Name	Mentions	Rank	GPU Accelerated
Dassault Systemes - SIMULIA Abaqus	91	4	Accelerated
LSTC - LS-DYNA	53	9	Accelerated
MSC Software - NASTRAN	16	18	Accelerated
COMSOL - COMSOL	14	22	No
ANSYS - ANSYS Mechancial	12	23	Accelerated
Altair Engineering - HyperWorks	6	49	Accelerated

The most frequent mentioned structural analysis code, Abaqus, delivers a 1.3x to 1.45x speed-up with its GPU-accelerated solver, according to Dassaul Systemes.[9] For LS-DYNA, GPU acceleration has proved useful for certain types of applications, in particular, solid element modeling, where a 3x speed-up was demonstrated with 2011 NVIDIA Tesla technology.[10] NASTRAN, a popular FEA code, has realized a 1.4x to 2.0x speed-up employing two GPUs on top of an 8-core CPU base, again, employing older 2012-era NVIDIA GPU technology, according to MSC Software.[11]

Environmental Modeling

Environmental modeling includes applications used to simulate weather, climate, air quality, and ocean systems. The potential use of weather forecasting ranges from saving lives and property with advanced severe weather alerts, to predicting wildfires, to managing the power grid by predicting energy demands, to estimating demand for products at your local supermarket based on weather forecasts. Add in climate modeling research for predicting global climate change and hydrodynamic modeling to predict water movement, and the impact of environmental modeling becomes obvious. Almost all these codes have been developed from non-commercial sources and approximately three-quarters of them are open-source.

Table 5 shows the presence of GPU support in the most popular environmental modeling packages. Two of the top packages are supported, including the most widely mentioned, Weather Research and Forecasting (WRF). According to work done at the National Oceanic and Atmospheric Administration (NOAA), a single GPU, in this case the NVIDIA Tesla K40, enabled a two-fold performance speed-up of the WRF WSM5 microphysics model compared to execution on a two-socket Xeon "Ivy-Bridge" platform.[12]

Table 5. GPU-Accelerated Applications: Environmental Modeling

Software Supplier - Package Name	Mentions	Rank	GPU Accelerated
NCAR - WRF	72	6	Accelerated
NCAR - CCSM	14	21	No
Terraspark - Insight Earth	7	35	Accelerated
CMOP - SELFE	7	38	No
U of Massachusetts - FVCOM	5	50	No

---

[9] http://www.3ds.com/fileadmin/PRODUCTS/SIMULIA/PDF/rums/uk-rum-13-abaqus-613-isight-5.8.pdf

[10] http://www.dynalook.com/8th-european-ls-dyna-conference/session-6/Session6_Paper2.pdf

[11] http://www.mscsoftware.com/Submitted-Content/Resources/nvidia_ps_ltr_w.pdf

[12] http://on-demand.gputechconf.com/supercomputing/2013/presentation/SC3133-CUDA-Weather-Research-Forecasting-Model.pdf

Geophysics

Geophysics includes applications used to understand and predict the earth's characteristics. The primary subcategories within geophysics are seismic analysis, which is used to identify potential natural resources (i.e., natural gas and petroleum fields) beneath the earth's surface , and reservoir modeling, which is also used in oil and gas exploration to map the characteristics of the reservoirs.

Table 6 shows the top geophysics applications in the Site Census, which both have GPU acceleration support. There are no dominant geophysics packages, because of the predominance of in-house applications in this segment; the two listed are the only ones that garnered enough mentions in the latest Intersect360 Research survey to make it to the top 50 applications list.

Table 6. GPU-Accelerated Applications: Geophysics

Software Supplier - Package Name	Mentions	Rank	GPU Accelerated
Terraspark - Insight Earth	7	35	Accelerated
CIG - Specfem3d	6	45	Accelerated

According to geoscience company CGG, an NVIDIA GPU will speed up Terraspark's Insight Earth seismic interpretation code by a factor of 50, compared to an 8-core CPU-based system, with scalable performance possible as more cards are added.[13] Another popular seismic code, Specfem3d, running on a system based on dual-socket, 16-core x86 nodes, realized a performance increase of up to 3.8x when an NVIDIA GPU was added.

Visualization/Image Analysis

Visualization is fundamental to most HPC workflows, allowing researchers and engineers to see the results of their simulations in graphic form. There are a large number of visualization packages that users cited in the latest Intersect360 Research survey. Of these, four made it into our top 50 of HPC applications. Visualization software is roughly evenly split between commercial and non-commercial applications.

As can be seen in Table 7, all the top visualization packages are GPU accelerated, which is hardly a surprise considering that graphics processors were originally designed for image and video processing. Pixel analysis algorithms in particular are extremely well-suited to the GPU architecture.

Table 7. GPU-Accelerated Applications: Visualization/Image Analysis

Software Supplier - Package Name	Mentions	Rank	GPU Accelerated
Exelis - IDL	16	17	Accelerated
Kitware - ParaView	9	28	Accelerated
LLNL - VisIT	8	34	Accelerated
Kitware - VTK	6	46	Accelerated

Exelis's IDL, the most popular commercial package, has been accelerated with a CUDA-based implementation and, with additional optimizations, can provide a 50x to 100x reduction in the time required to process large imagery, according to the Exelis web site.[14] ParaView, VisIT, and VTK, all popular open-source visualization packages, have also incorporated CUDA support.

---

[13] http://www.cgg.com/InsightEarth/
[14] http://www.exelisvis.com/Home/NewsUpdates/TabId/170/ArtMID/735/ArticleID/13966/Using-GPUs-to-Accelerate-Orthorectification-Atmospheric-Correction-and-Transformations-for-Big-Data-Recorded.aspx

Physics

Physics includes applications are used to model the fundamental principles involved in the behavior of matter and energy. Work in the field ranges from cosmological models of the evolution of the universe to quantum mechanics simulations of matter at the subatomic level. Physicists were among the first users of HPC and continue to have some of the most demanding computational requirements. As might be expected from a research-based domain, the vast majority of these packages are provided by non-commercial organizations as open-source codes. Table 8 shows the pervasiveness of GPU acceleration in the top physics codes.

Table 8. GPU-Accelerated Applications: Physics

Software Supplier - Package Name	Mentions	Rank	GPU Accelerated
Quantum-espresso.org - Quantum ESPRESSO	30	14	Accelerated
USQCD.org - MILC	8	29	Accelerated
CERN - CMSSW	8	30	Accelerated
USQCD.org - CHROMA	6	41	Accelerated
Abinit.org - ABINIT	6	43	Accelerated
Open Source - Nbody	6	48	Accelerated

Quantum ESPRESSO, an open-source code for electronic-structure calculations and nanoscale materials modeling, has been accelerated with GPU support, offering a 4x to 8x speed-up over an 8-core CPU implementation, according to the web site.[15] Another popular open-source physics code, CHROMA, is able to reduce simulation times by a factor of five when GPU acceleration is used.[16]

Biosciences

Bioscience applications are used to reveal the nature of life and help design medical treatments. They are also used to understand and enhance agricultural production. This segment includes such applications as genomics, proteomics, and drug discovery, among others. We see much of this segment being driven by the genomics revolution, which uses bioinformatics software to analyze the function of genes as it relates to cellular function. This information led to a flowering of applications in biology, pharmacology, and biochemistry/bioengineering. Note that many of the molecular dynamics codes, which were placed in the Chemical Research category, have bioscience applications as well.

As Table 9 shows, there are no mainline GPU ports of the top bioscience codes. However, there are successful GPU-accelerated versions for BLAST, AutoDock, and MyBayes in the research community. They have not been integrated into the official code distributions maintained by the software suppliers.

Table 9. GPU-Accelerated Applications: Biosciences

Software Supplier - Package Name	Mentions	Rank	GPU Accelerated
NCBI - BLAST	46	11	No
Thermo Scientific - Sequest	9	27	No
Scripps Research Institute - AutoDock	8	31	No
Open Source - MrBayes	8	32	No
Open Source - Galaxy	7	36	No
Illumina - Casava	7	37	No

---

[15] http://www.quantum-espresso.org/wp-content/uploads/QE-dev-meeting-2013/heterogeneous_Spiga.pdf
[16] http://arxiv.org/pdf/1408.5925.pdf

INTERSECT360 RESEARCH ANALYSIS

As is evident from the number of GPU-accelerated software packages available across application domains, the use of this accelerator technology has become well established in the HPC user community. In nearly every domain, GPU-acceleration is now available on one or more of the most commonly used application packages. More significantly, in concert with the expanded use of GPUs reported by HPC users in our latest Site Census surveys, GPU usage is moving from a test-and-evaluation phase into production work.

In certain areas such as chemical research, physics, structural analysis, and visualization, the availability of GPU acceleration is nearly ubiquitous. In other areas, like biosciences and environmental modeling, penetration remains incomplete. But even in environmental modeling, the recent announcement by the Swiss National Supercomputing Centre that it would be running its production weather models on GPUs reflects the extent to which this technology can succeed there.

The growth of GPU adoption for HPC has been driven almost entirely by NVIDIA, which has invested heavily in building a robust software ecosystem to support its hardware. Specifically, the company has developed a set of parallel programming APIs, libraries, and associated software development tools to support application development on its CUDA (Compute Unified Device Architecture) GPU platform. Since its introduction, CUDA has been available on platforms running any NVIDIA GPU and was downloadable for free by anyone interested in experimenting with using GPUs as application accelerators. This combination of free software running on existing hardware effectively bootstrapped the accelerator market and positioned NVIDIA as the market leader. Subsequently the company now holds over three quarters of mentions of accelerator implementations in our survey research.

While NVIDIA's efforts have been centered on CUDA-based software tools and middleware, implementations also include support of OpenCL, an open standard framework for developing parallel applications across a variety of processor architectures, and OpenACC, a set of standard compiler directives for high-level languages that can be used for both x86 CPUs and accelerators. NVIDIA has also established over 20 GPU centers of excellence around the world, as well as a global network of GPU research and educations centers encompassing hundreds of academic institutions. Those efforts have supported the acceleration of software tool development for general-purpose GPU programming, in addition to aiding the porting of open-source HPC codes to GPUs. The company has also collaborated with ISVs to help accelerate commercial HPC applications and libraries. Today, hundreds of such codes are now available, promising speed-ups of anywhere from 10 to 20 percent up to more than 1,000-fold. For the most commonly accelerated codes, performance increases are usually in the 2x to 10x range.

The movement from experimentation into production is an outcome of the ecosystem support from NVIDIA and the availability of the application packages described in this report. At this point, we believe that GPU computing has reached a tipping point, which will encourage GPU support to be incorporated even more broadly into applications that can make use of the architecture's highly parallel features. In fact, this process is well underway. In addition to the 34 top GPU-enabled applications identified here, NVIDIA cites hundreds more on its website. We expect this expansion to continue as the GPU computing user base grows and becomes more comfortable with the technology.