

View Point



Accelerating Compute Intensive Applications with GPGPU

Dr. Santonu Sarkar and Dr. Sudeep Mallick

Large enterprises like Healthcare and life Sciences, e-Governance, Finance, Medical Sciences, and others; have mission-critical IT systems to solve complex computational problems. A paradigm shift is happening — computing is shifting from the CPU towards the GPU, the parallel co-processor, at a fractional cost and with amazing throughput. If a legacy application has to shift to such a parallel computing platform, the opportunities for parallelism need to be identified and re-engineered accordingly. In this point-of-view we present our understanding of the landscape, and our solution approach using General-purpose computing on graphics processing units (GPGPU).

Introduction to GPGPU Based Computing

Parallel computing aims to increase the throughput and performance of computations through exploitation of inherent parallelism existing in the computation logic. Recently it has been found that one can effectively utilize modern General Purpose Computing using Graphics Processors (GPGPU) like the ones manufactured by NVIDIA, to compute massively data parallel, numerical tasks and achieve a manifold improvement of speed. The secret behind such an astonishing improvement of throughput lies in the architecture of GPU processor. The large number of cores on a single GPU along with its suitably designed memory and thread hierarchy enables manifold speed up of data parallel tasks having relatively high compute to communication ratio. What's more surprising is that GPU processor, once utilized only for graphics processing and visualization, can perform high performance computing (HPC) tasks much faster than a multi-core Central Processing Unit (CPU), at a fraction of today's operational and platform infrastructure cost.

Large corporations have observed this phenomenon with avid interest and started using GPU to execute some of their core domain specific algorithms. Monte-Carlo simulation in financial application or seismic data processing in oil exploration, to name a few, have shown manifold improvement in computation using GPGPU compared to a multi-core system. With this success, there is a renewed interest in the scientific community to invent new and more accurate algorithms to replace hitherto approximate algorithms, traditionally used for HPC. Infosys, with its rich research and practical experience in HPC, recognizes that the potential impact of exploiting GPGPU in non-graphic, data-parallel applications is far reaching in the world of HPC.

GPGPU Opportunities for Infosys - An Early Analysis

Infosys has analyzed the evolution of GPGPU and its growing usage patterns. An overall market survey supported by external research reports, shows that GPGPU is applicable to a number of industry verticals to speed-up various number-crunching algorithms. A sample of such applications that is amenable to GPGPU is provided in the table 1.

Infosys' Value Proposition in HPC

Infosys has executed numerous HPC projects for its Fortune 500 clients in various industry verticals. It plans to leverage its earlier experience in HPC and parallel computing space, in-house developed accelerators and its expertise in NVIDIA powered CUDA based GPGPU platform to address HPC needs of its Fortune 500 clients in various industry verticals such as energy and utilities, retail, manufacturing, banking and capital markets, life sciences to name a few.

SETLabs, which is Infosys R&D division, aims to develop tools and methodologies to facilitate development and deployment of high throughput applications on NVIDIA

Domain	Areas of application	Sample nature of computation
Life sciences	Molecular dynamics, simulation of proteins, protein sequencing	Smith Waterman sequence alignment
Manufacturing	Fluid dynamics, 3D, CAD/CAM, electronics design simulations, parallel flow simulation	Non linear complementarity problems, finite element analysis, Lattice-Boltzmann Model, sparse matrix solvers, combinatorial graph algorithms
Energy, oil and gas	Stock market simulations, normalization of market data, options pricing, risk analysis, etc	Monte Carlo Sim, FFT, stochastic models, stochastic optimization
Software engineering and IT	Power system simulation, 3D simulation of seismic data sets	
BI	Power system simulation, 3D simulation of seismic data sets	

Table 1: Sample Applications

GPUs leveraging its existing expertise and IPs in Grid Computing and Parallel Computing. Infosys delivery team can take the solution and offerings to market to its existing customer base in various industry verticals.

Infosys plans to apply a combination of its project execution and delivery skills with the results of its state-of-the-art R&D projects to deliver value to its customers in terms smooth and predictable application analysis, migration, and deployment for parallel computing.

Infosys plans to exploit its access to its customers' existing application architectures to proactively conduct analysis and make value propositions in the parallel computing space for high throughput LOB applications.

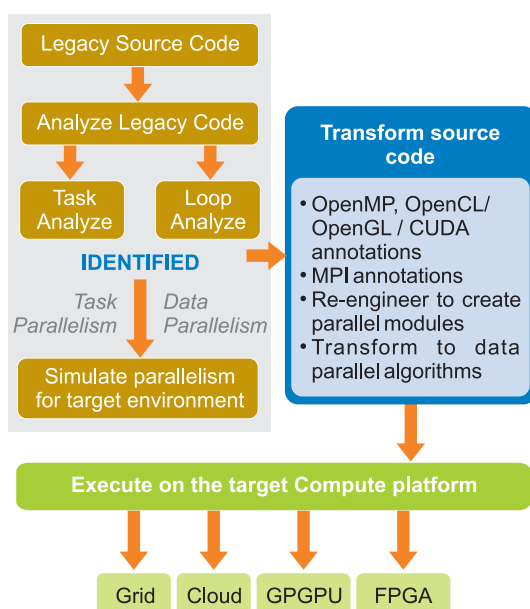
Infosys has a presence in various geographical regions across the globe and its affiliate companies in Australia, China, and Sweden. Its India Business Unit looks into business from Indian companies as well as Government organizations and bodies. Hence Infosys is well placed at the moment to look at both government and non-government sectors.

HPC Research at SETLabs

SETLabs has done elaborate work in the area of Grid Computing, Parallel Computing, Performance Engineering and Cloud Computing resulting into development of numerous world class Patent pending IPs which have been leveraged in many existing client engagements resulting into case studies and testimonials. The Lab is manned by a team of knowledgeable software engineers, computer science graduates, post graduates and Ph.D.s who have regularly published in IEEE/ ACM conferences/journals and various industry trade journals and conferences in the area of Grid computing, Parallel computing. This work would serve as the launching platform for it to take up GPU technology which is basically one form of parallel computing.

Infosys plans to leverage its IPs for re-engineering and on boarding HPC applications on various target parallel computing architectures keeping in view various architectural trade-offs such QoS, ROI, TCO, etc. It plans to offer manifold performance gains to existing legacy applications of its customers by conducting efficient analysis of the legacy application to identify parallelization opportunities, estimate performance gains expected by on boarding / migrating legacy applications to parallel computing architectures. Infosys plans to apply a combination of its project execution and delivery skills with the results of its aggressive R&D projects to deliver value to its customers in terms smooth and predictable application analysis, migration, and deployment for parallel computing. Infosys plans to exploit its access to its customers' existing application architectures to proactively conduct analysis and make value propositions in the parallel computing space for business applications having HPC requirements.

Research in Application Re-engineering for HPC



One of key challenges in on boarding legacy code onto parallel computing platforms is to be able to quickly identify computing hot spots and opportunities for parallelism in the legacy code so that it could be re-engineered accordingly. Infosys' patent pending IP Legacy Code Analyzer (LCA) framework (see the Figure above) enables identification of parallelism opportunities inherent in the legacy code base. It aids decision making in situations where legacy code base has to be re-factored and re-engineered by helping identify execution hot spots, task parallelism opportunities for parallel computing platforms such as clusters and grids. Research efforts are underway at SETLabs to extend the LCA Framework to cater to coarse and fine grained data parallelism identification requirements targeting various languages such as Java, C, C++, C#. With the help of LCA manual code analysis time is reduced manifold times enhancing productivity of HPC legacy re-engineering projects.



For more information, contact askus@infosys.com

About Infosys

Many of the world's most successful organizations rely on Infosys to deliver measurable business value. Infosys provides business consulting, technology, engineering and outsourcing services to help clients in over 30 countries build tomorrow's enterprise.

For more information about Infosys (NASDAQ:INFY), visit www.infosys.com.