

Francisco Paulo de Aboim

A Finite Element Implementation Using GPUs and OpenCL

**Dissertação de Mestrado**

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Civil da PUC-Rio como requisito parcial para obtenção de título de Mestre em Engenharia Civil.

Orientador: Luiz Fernando Campos Ramos Martha

Rio de Janeiro, Março de 2013

|  |
| --- |
| Francisco Paulo de Aboim  A Finite Element Implementation Using GPUs and OpenCL  Tese apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Engenharia Civil do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.  Prof. Luiz Fernando C. R. Martha  Orientador  Departamento de Engenharia Civil - PUC-Rio  Prof. José Eugênio Leal  Coordenador Setorial  do Centro Técnico Científico - PUC-Rio  Rio de Janeiro, 29 de março de 2013. |

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização do autor, do orientador e da universidade.

Francisco Paulo de Aboim

Graduou-se em Engenharia Civil, pela PUC-Rio, em fevereiro de 2010. Como projeto final de graduação desenvolveu com o prof. Luiz Fernando Martha o Misulatool, ferramenta educativa para cálculo barras em mísula. Atuou de 2010 a 2013 como estagiário e bolsista no TecGraf no grupo Mvgeo deselvolvendo software para geração de modelos de elementos finitos para análise geomecânica.

Agradecimentos

Agradeço, em primeiro lugar, aos meus pais pelo incondicional apoio em todas as decisões da minha vida. A todos os professores da PUC-Rio que contribuíram na minha formação e fizeram possível com que chegasse a este ponto. Aos meus orientadores que me guiaram e tiveram inabalável paciência e por terem acrediato em mim. A Rita de Cássia e a Deane Roehl pela compreensão e auxílio durante minha pós-graduação. Aos meus colegas de trabalho e amigos do TecGraf por tudo que aprendi e pelas discussões que ajudaram a me tornar um programador melhor.

Abstract

Aboim, Francisco Paulo de; Martha, Luiz Fernando; Sotelino, Elisa. **A Finite Element Implementation with OpenCL**. Rio de Janeiro, 2012. Master's Dissertation - Department of Civil Engineering, Pontifícia Universidade Católica do Rio de Janeiro.

The paper investigates an implementation of the finite element method using the OpenCL (Open Computing Language). GPU computing is becoming increasingly popular because of its ability to efficiently perform highly parallel computations, such as vector and matrix operations. Current GPUs have approximately 100 times more potential computing power than a CPU but achieving this potential requires an understanding of the hardware and how it affects the programming model. The massive multicore programming model is also especially well-suited for problems of a parallel nature. On the other hand, GPUs are less flexible and require more care and patience in their programming. Furthermore, until recently knowing different programming frameworks was necessary for developing for different hardware vendors. OpenCL was thus made to function as the standard for heterogeneous platforms, using a single framework. Its development is supported by a consortium of major vendors such as Apple, Nvidia, AMD, Intel, ARM, and others. The finite element method is the most widely used method to solve complex computational mechanics and engineering problems. It involves matrix and vector operations in many of its phases and therefore is an excellent candidate for GPU computing. In this investigation, the use of OpenCL is explored on the different stages of the method. Both code implementation and its optimization are discussed.

**Keywords:** GPU computing, OpenCL, High Performance Computing, Parallel Finite Element Analysis

Sumário

[1 Introduction 8](#_Toc353234172)

[2 Historical Background 14](#_Toc353234173)

[2.1 The CPU Architectures and the Road to Massively Multicore 14](#_Toc353234174)

[2.2 Vector Processors 17](#_Toc353234175)

[2.2.1 The Development of Vector Processors 17](#_Toc353234176)

[2.2.2 SSE and CPU Instruction Sets 19](#_Toc353234177)

[2.3 GPU Architecture and the Development of General Purpose GPU Computing 20](#_Toc353234178)

[2.3.1 Programmable Shaders 20](#_Toc353234179)

[2.3.2 Programming for the Unified Shader Architecture 23](#_Toc353234180)

[3 OpenCL 25](#_Toc353234181)

[3.1 Introduction 25](#_Toc353234182)

[3.2 Execution Model 25](#_Toc353234183)

[3.3 Memory Model 30](#_Toc353234184)

[4 Sparse Matrix Data Structures 32](#_Toc353234185)

[4.1 Diagonal Format 33](#_Toc353234186)

[4.2 Compressed Sparse Row 33](#_Toc353234187)

[4.3 Ellpack Format 34](#_Toc353234188)

[4.4 Eigen Format 35](#_Toc353234189)

[4.5 Searching index arrays 35](#_Toc353234190)

[4.5.1 Linear Search 36](#_Toc353234191)

[4.5.2 Binary Search 36](#_Toc353234192)

[4.5.3 Linear Search with SSE 37](#_Toc353234193)

[4.5.4 Performance 38](#_Toc353234194)

[5 Implementation of The Finite Element Method 39](#_Toc353234195)

[5.1 Determining the Structure’s Stiffness Matrix 41](#_Toc353234196)

[5.2 Parallel implementation using OpenMP 45](#_Toc353234197)

[5.3 Parallel CPU implementation using Graph Coloring 45](#_Toc353234198)

[5.4 Parallel implementation in OpenCL 46](#_Toc353234199)

[5.5 Setting Boundary Conditions 49](#_Toc353234200)

[6 Solving for Displacements 50](#_Toc353234201)

[6.1 Naïve Implementation 50](#_Toc353234202)

[6.2 Introducing Vectorization Primitives 51](#_Toc353234203)

[6.3 Utilizing Local Memory 52](#_Toc353234204)

[6.4 Utilizing Tiling/Blocking 53](#_Toc353234205)

[6.5 Summing it all up 54](#_Toc353234206)

[7 Results 56](#_Toc353234207)

[8 Conclusion 105](#_Toc353234208)

[9 References 108](#_Toc353234209)

List of Figures

[Figure 1 - Evolution of Heat Generation 9](#_Toc353234210)

[Figure 2 - Gate Leakage (KIM, N.S. et al.) 10](#_Toc353234211)

[Figure 3 - Evolution of CPU Cache (tomshardware.com) 17](#_Toc353234212)

[Figure 4 - AMD Cypress SIMD engine (old nomenclature) 21](#_Toc353234213)

[Figure 5 - *ATI* 5870 Architecture (AMD Parallel Processing Guide) 24](#_Toc353234214)

[Figure 6 - Compute Device Structure (Copyright AMD) 26](#_Toc353234215)

[Figure 7 - Memory Pipelining 29](#_Toc353234216)

[Figure 8 - Latency Hiding 29](#_Toc353234217)

[Figure 9 - OpenCL Memory Architecture (Khronos Group) 30](#_Toc353234218)

[Figure 10 – Example two dimensional mesh 39](#_Toc353234219)

[Figure 11 - Example mesh analysis results 40](#_Toc353234220)

[Figure 12 - Three dimensional mesh 41](#_Toc353234221)

[Figure 13 - Stiffness Matrix Assembly Flowchart 44](#_Toc353234222)

[Figure 14 - Avoiding Race Conditions through Coloring 49](#_Toc353234223)

1. Introduction

The motivations of researching new programming models targeted at heterogeneous multicore systems are inherently tied to the motivations that led to the development of multicore systems. Moore's law is well known inside and outside the computing community for having been able to predict with remarkable accuracy back in 1965, in the celebrated and often cited article "Cramming more components onto integrated circuits" (G. Moore, 1965), that transistors on integrated circuits would double roughly every two years. This claim has gained such notoriety because, regardless of silicon price fluctuations, technological breakthroughs and investment in research and development, it has held true to this day. Not only did Moore predict this trend, but he also was able to envision what would be one of the greatest barriers processor manufacturers would face to keep up with the trend and postpone diminishing returns: heat generation. In the same article Moore poses: "Will it be possible to remove the heat generated by tens of thousands of components in a single silicon chip?”, a question that lingered for decades while the heat generated by these processing components increased.

While the growing problem was addressed by cooling solutions which got progressively more complex and expensive, the inherent problem behind heat generation remained. In a few decades passive heat sinks evolved to active heat sinks (including fans to increase the heat exchange), copper heat tubes and a variety of fan designs to optimize air flow (Figure 1 - Evolution of Heat Generation).

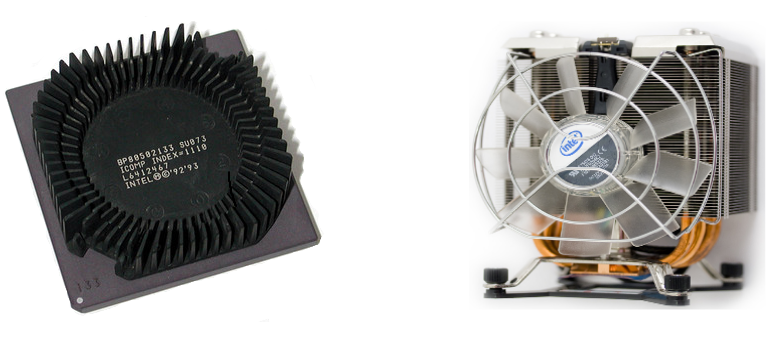


Figure 1 - Evolution of Heat Generation

We have now reached a point where current processor gates are around 20 nm wide (As in Intel's 22nm Ivy Bridge) or about 74 atoms of silicon wide (with prototypes being even smaller). When dealing with these reduced dimensions, transistor gates have become increasingly susceptible to an effect defined in quantum physics as quantum tunneling. Quantum tunneling tells us that the smaller the barrier is, the greater the probability that an electron will be found on the other side of the barrier. Recently gates have gotten so small that this probability has risen dramatically and the loss of energy due to electrons tunneling through the insulator has become a barrier to further miniaturization, since energy loss translates into an even higher production of heat. The loss of energy due to these effects is commonly named *leakage*. During decades, the processor industry was able to increase processing power by miniaturization and increasing clock frequency.

During the last years, however, due to Leakage becoming an increasingly important factor. The industry has hit a frequency wall where further gains in processing power by increasing clock speeds are leading to a higher marginal cost in terms of power consumption [Figure 2 - Gate Leakage (KIM, N.S. et al.)]. To deal with this, many techniques have been proposed, all of which lead to an increase in complexity, be it in the processor itself, compilers or shifted towards the programmer. The era of easy gains in algorithm speed coming from advances in the semiconductor industry with increasing clocks which made even old code run faster has come to an end (The Free Lunch is Over, Herb Sutter, 2005).

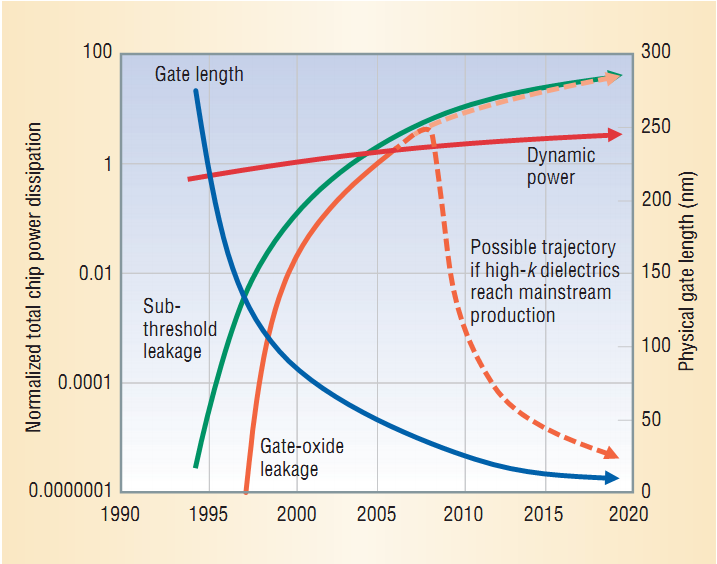


Figure 2 - Gate Leakage (KIM, N.S. et al.)

While the long awaited replacement for the common variants of the MOSFET (metal-oxide semiconductor field-effect transistor) technology do not arrive and contenders such as the graphene transistor, carbon nanotube transistor and other incipient exotic technologies are still being researched - albeit a long way from mass production - other solutions have been proposed. One of the solutions found to overcome these limitations was to simply pack more cores into a single processor, which set the trend of the multicore era. This solution, although obvious, was not implemented at earlier stages because of one problem: almost all production code was designed with the serial programming model in mind and would gain no immediate benefit from the availability of multiple cores.

A more recent market segment which did not require maintaining compatibility with legacy code and which was well positioned to benefit from the multicore paradigm given the parallel nature of most of the computation required, the segment of personal computer graphics hardware or GPUs (Graphics Processing Units), was one of the first to buy into the multicore trend. The fact that initially vendors supplied their own implementations for APIs such as OpenGL also facilitated the transition by making programmers oblivious by the complex nature of the computations being performed behind the scenes. When the capacity to program for these processors was opened to the common programmer, a number of different uses for this massively parallel computational model transpired, creating a new market for graphics card vendors. This market took advantage of the higher throughput these devices offered for parallel computation with the diminished power consumption with respect to traditional processors, an increasingly important factor for the reasons formerly mentioned.

In the burgeoning mobile industry, efficient use of processing power is paramount and heat dissipation is one of the main concerns. Industry players such as ARM, which have for decades researched how to build efficient processors which address these requirements, have gained a significant amount of market share due to the difficulty other vendors in dealing with these setbacks, becoming the largest producer of processors in the world. Unsurprisingly, these same advantages which GPUs possess in relation to CPU (Central Processing Units), the traditional computer processors, come from architectural similarities in the hardware. With the rapid boom of tablet and smartphone production which require more processing power then traditional mobile phones and are required to perform a myriad of uses, which once were restricted to personal computers, a legion of devices with considerable processing power has been made available. Unfortunately, these processors, in accordance with GPUs, require a multicore programming model to be efficiently utilized, further increasing the importance of understanding these programming models in this new era. With the onslaght of mobile personal computing, many people have migrated from desktops to tablets and smartphones, making personal desktop computing a market of shrinking growth. While HPC computing is becoming an increasingly popular tool gaining wider applicability for scientific computing and being a growth market, traditional vendors of processors for the personal computing market are increasingly pushing products with an eye on these facts, putting further pressure on throughput oriented processors and the massively multicore model.

The current state of development of finite element analysis in the context of massive multicore programming can be considered to be in its inital phase. While de development has roots in the vector processor architecture previously described, aiding the discovery of the paths that can lead to higher gains, many algorithms have to be tuned and rethought to perform adequately for this specific architecture. Notable recent developments include the utilization of graph coloring to avoid race conditions and permit parallelization in the assembly stage (Komatitsch et al., 2009) and also methods of assembly in parallel by nonzeros (Cecka et al.).

On the effort of solving the resulting system of linear equations, for finite elements where sparse matrices are ubiquitous, especially considering commercial applications which deal with large models with at least thousands of degrees of freedom, iterative solvers for sparse matrices are the norm. Since iterative solvers do not pertain only to finite element analysis and are found in many other areas, the development of this class of solvers in the field of numerical methods has gained much attention in recent years. Recent developments which deserve special attention and indirectly contribute to speeding up the solution of the stiffness-displacement equation by creating sparse matrix formats tailored to extract the most performance out of sparse matrix-vector multiplication; which is by far the most expensive linear algebra operation involved using iterative solvers. The modern approach has somewhat converged to using a variant of the Ellpack format, initially developed for vector processors which have many similarities to modern GPUs. The more recent versions initially targeted CUDA (Bell, N and Garland, M.) and posteriorly evolving to a more general and flexible approach (Vásquez et al.). Given that most of the computational time in FEM analysis is spend on these two stages, these new approaches have made the utilization multicore for these purpose of analysis progressively interesting.

With all of the above being considered, this is an attempt to make available an overview of how technology has evolved to this point and how the new architectures can be used to develop algorithms that take advantage of the newer paradigms, algorithms and formats developed for the massively multicore era. Taking into account the narrow relation between the novel programming architectures for massive multicore systems and available programming models, for didactic purposes, optimization strategies are offered within the context of finite elements. Finally, an analysis of the potential impact of these changes within the field is discussed. The full sourcecode regarding this thesis is available at http://github.com/fpaboim

1. Historical Background
   1. The CPU Architectures and the Road to Massively Multicore

Such fundamental hardware changes brought by the move to multicore meant that programmers had to redesign their hardware and algorithms with a different mentality. The burden of having increasing computing performance from a piece of code was now partially passed on to the programmer, who could no longer depend on better compilers and faster processors to gain increased performance. Not only would serial code cease to run faster on newer generations of processors, but the code could actually see its performance degrade, since clock rates stagnated or even fell to be able to accommodate the extra heat produced by the addition of more cores. Still, clock speeds and transistor counts are not the only factors which affect processing power, and to keep up with the growing demand for performance every possible avenue was explored.

One of the main factors influencing overall performance is the number of instructions per cycle which translates roughly to the amount of work which is done per clock cycle. Instruction-level parallelism (ILP), which is also known as dynamic or out-of-order execution, is a technology processors use to analyze dependencies between instructions and reorder instruction execution to eliminate pipeline stalls (Akhter, S and Robers, J). This increases the number of instructions executed per clock. Unfortunately, the number of instructions per clock has hit its own wall (Stamatis, Vassiliadis et al.) and has remained fairly constant in the past years. Not being able to achieve more performance out of ILP, once an oasis for squeezing more performance out of processors, has led to focusing efforts in other areas.

Another imperative factor for driving performance is memory speed. While processing power has kept up with Moore's law as described, memory speed has risen at a dramatically slower pace (McKee, Sally A., 2004). This has made the spread between processing speed and memory access speed grow larger with time and no new technology was developed which significantly increased memory access speed in a cost effective manner. Moreover, with the addition of more cores to processors communication between cores has increased compared with previous models. If the architecture uses a common bus for communication between cores, a competition for the same communication channel arises which must be taken into account. Also, *ceteris paribus*, more cores to longer bus wires which require lower clocks, negatively impacting speed. Due to wire delay, which causes costs of communication to increase relative to physical separation, this trend has obviated the necessity of exploring data locality - particularly in more recent years with the rapid increase in numbers of cores. Solutions such as optical on-chip networks have been proposed to address these shortcomings (Psota et al.) but are still far from commercial availability. Because of this, many of today's computational algorithms are not actually limited by processing capability but by memory bandwidth and speed.

Before processors started using cache as means to speedup processing, faster types of memory did exist and were available but at a much higher cost which made it impossible for them to hit mainstream production. The solution Intel first implemented on the i386 chip was to allow an extra space on the motherboard where a small amount of high speed memory could be connected to provide the processor with high speed memory access for whoever would be willing to pay for it. With the advancement of compiler technology to include this new level of available memory, engineers saw that adding a small amount of cache to a processor could greatly increase the processing speed, given that all data was not used equally. This is due to the principle of locality of reference (Denning, P. J., 2005). As an example, while performing a task such as multiplying an n-sized vector by a scalar, the scalar is used n times while the vector is used only once. If the scalar were to be cached, the reading time would be an order of magnitude faster without the need to use high speed memory to store the whole problem's data. One would simply need space for a single float or double, normally a measly 4 or 8 bytes, to achieve outstanding performance gains. In reality, the scalar would probably be placed on a register by an intelligent compiler, but the same principle which makes registers use more efficient applies to cache (although there are subtle difference on how both behave). This was the realization which prompted Intel to incorporate cache as an option on the motherboard, given the high marginal gain of using a small portion of costly high speed memory. This architectural change was such a success that on the next family of processors, the i486, the cache started to be included on the same wafer as the processor.

Thus, in modern processors the traditional architecture with one memory space was replaced by a hierarchy of different memory levels. AMD's Athlon XP and 64 and Intel's Pentium 3 and 4 added one more layer to the memory hierarchy with the addition of a larger but cheaper L2 cache. With the advent of the multicore architectures Intel's Core 2 Duo created an L2 cache which was common to both cores and therefore permitted sharing of cache memory between cores while the L1 cache was faster but core specific(for counters and such) which further impacted programmability. Current processors such as the Phenom X4 and the Intel core i7 have yet another added level of memory with cores having their own L1 and L2 caches, and sharing an L3 cache [Figure 3 - Evolution of CPU Cache (tomshardware.com)].

One unintended effect was that the complexity of the compilers increased dramatically with this new type of memory. Compilers have to decide what to put in the cache, when to put it there - since even changing the order of execution of independent events can impact performance - and where (at what cache level) to write the information. The sharing of information on certain caches also created new possibilities for optimizing code by allowing parallel computing within the cache memory level. While compilers have evolved significantly over the last decades in terms of optimizing code, they are limited in what they know of the task at hand. If one were to have an algorithm with deterministic memory access it would be easy to outperform a compiler by planning ahead and manually decide which variables to store in the different levels of cache and by tweaking memory use to maximize bandwidth by sharing information at high speeds through the common memory cache. In this manner, many linear algebra libraries can achieve performance much higher than would be expected by running a naïve version of the routines and letting the compiler do the work. Nevertheless the growing hierarchical nature of memory increases the possibilities of information storage and, therefore, difficulty for the programmer to define and optimal strategy.

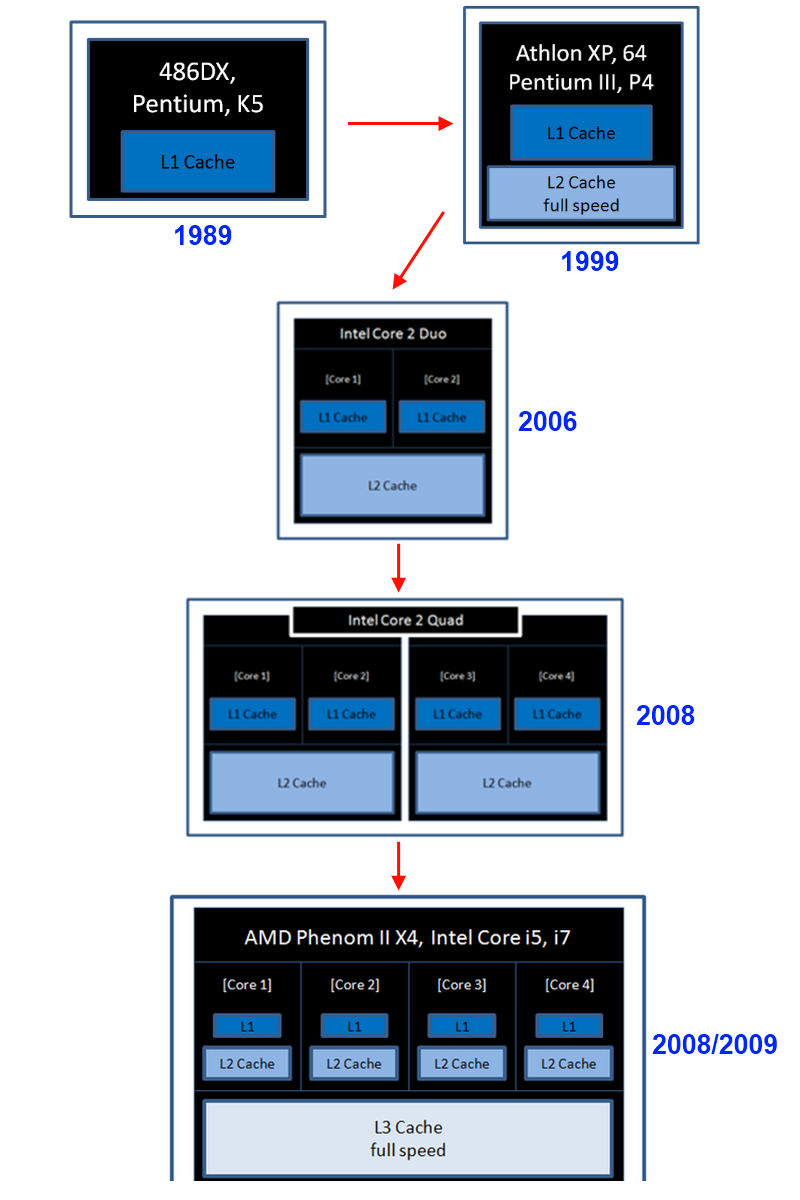


Figure 3 - Evolution of CPU Cache (tomshardware.com)

* 1. Vector Processors
     1. The Development of Vector Processors

The SIMD (Single Instruction, Multiple Data) paradigm as described by Flynn’s taxonomy (Flynn, 1972) describes an execution model where the same instruction is executed on multiple units of data (words, i.e., smallest executable unit of data by the processor). This contrasts with the traditional SISD (Single Instruction, Single Data) paradigm where one instruction is executed on a single word, such as in the old Commodore 64 platform (Table 1 - Flynn's Taxonomy).

|  |  |  |
| --- | --- | --- |
|  | Single Instruction | Multiple Instruction |
| Single Data | SISD | MISD |
| Mulitple Data | SIMD | MIMD |

Table 1 - Flynn's Taxonomy

A vector processor is an implementation of this model where the same instruction is executed for N words, which can also be viewed as executing on a N-dimensional array of words (which gives insight to why the technology was named a "vector processor") following the SIMD model as described above in Flynn’s taxonomy. Immediately, we can identify that the instruction bandwidth will be much smaller since only one instruction read will be required for operating on N words and, therefore, the instruction read overhead will be (1/N)th of that of a scalar processor. Nevertheless, this will only happen if all N processors are actually processing useful information. Since vector processors execute in lock-step, i.e., all individual processors are required to concurrently perform the same operation, if the operation is not data-parallel or barely so, only one or a few of the processors will be used and unused processors simply operate on any data and have the result discarded, resulting in a less than optimal efficiency. It becomes clear that vector processors are, in essence, engineered to solve problems which are predominantly data parallel.

Given that vector architectures are only suited for a specific type of problem, widespread adoption only occurred in the 70’s with the infamous Cray-1 supercomputer. Besides being the first successful supercomputer to employ vector architecture, the Cray-1 was also ahead of time in that it employed individual registers for each vector processor and used a pipeline model[], both of which are ubiquitous in today’s computers. One of the reasons for its success was that, like modern GPU’s, the Cray-1 could deliver a much greater performance compared to other supercomputers of similar price – for problems that were inherently data parallel. In this sense the Cray supercomputer brought the vector processing model to the mainstream, which at the time was limited to few large and expensive computers mainly used for research.

While modern CPUs are not vector processors *per se*, vendors have added instruction sets over the years that make use of vectorization for added performance for specific operations with examples such as Intel's SSE/MMX and 3DNOW!. These processors are, rather, a myriad of processing models developed over decades with different companies (or even divisions within companies) focusing on different processing models based on the target market's necessities. For this reason only antique processors such as the Commodore64 can be considered as an example of SISD, while newer processors can be considered to have SIMD extensions which can be executed by independent threads being, therefore, MIMD (Multiple Instruction, Multiple Data).

* + 1. SSE and CPU Instruction Sets

A characteristic of the x86 architecture is that one of the ways vendors can deliver additional performance is by using wafer space for adding instruction sets. On the other hand, wafer space is limited and there is a tradeoff between adding instructions or cache, for example. Adding instruction sets also increases chip complexity, but when they can be used, large gains in performance can be obtained. While the work can be done on the software side, complex instructions transfer the load to the hardware and usually accomplish the task in a fraction of the time required. Instruction sets exist for a diverse number of applications, from multimedia to 3D rendering (which recently has been dropped given that nowadays it is rare to utilize a CPU to process graphics).

One of the most useful, particularly for scientific computing, is the Streaming SIMD Extensions (SSE) family of instructions which was born to address the shortcomings of the first SIMD instructions proposed by Intel, MMX. It has since then been expanded four times since the inception and is still in the process of being extended in new chips, as with the addition of the new AVX (Advanced Vector Extensions) instruction set. The SSE family, as the name implies, is a SIMD extension which permits the processor to perform as if it were a vector processor, being able to achieve much better performance on parallel operations. This is accomplished by storing data in registers which possess faster memory access and by operating on these registers by using a set of instructions which can each operate on a whole vector of data in one clock cycle. Thus, it is possible to compress into one cycle the computation that would take multiple cycles if done without vector extensions. It is also valid to point out that many times the way an algorithm is programmed makes it difficult to expose the vectorization potential to the compiler, which will result in suboptimal code. To trust the compiler for performance code is tricky and not straightforward, as sometimes even unrolling loops, aligning access and performing separate load/store operations as to expose vectorization to the compiler will not be sufficient to guarantee the code will utilize the appropriate vectorization instructions.

Although considerable gains are possible using extended instruction sets, one counterpoint is that instruction utilization depends not only on the hardware where it is run but is also compiler specific. That means roughly doubling the lines of code if you want software that can be built on a windows compiler and GCC, for example. Using a framework which compiles to multiple platforms such as OpenCL (Open Computing Language) makes the code capable of vectorization and compiler independent. For small functions which get called many times, such as in linear algebra libraries for CPUs, the overhead which OpenCL entails makes the code better suited for utilizing instructions directly. Although the resulting code may not be the most elegant, with conditional compilation to account for the platform and the necessity to port from one compiler to the other, for applications where performance is paramount, it is the best option.

* 1. GPU Architecture and the Development of General Purpose GPU Computing
     1. Programmable Shaders

With the advent of the programmable vertex and pixel shaders, developers were given the flexibility of creating their own code for customizing the rendering process to meet their needs. While most developers used this new technology for its original purpose, i.e. creating custom graphics effects or adding realism, it became clear that while the shader was made to operate on a vertex or pixel, any data could be treated as such and thus the GPU could be programmed to be used for any general purpose computation. This formed the field of GPGPU (General Purpose GPU) computing. With the rapid growth of the processing power of GPUs which quickly surpassed that of CPUs pushed by the highly competitive gaming industry, research began to appear that took advantage of the great and untapped pool of GPU resources.

To process hundreds of vector and matrix required by shader algorithms, state of the art GPUs possess hundreds of “small vector processors”, also called SIMD engines (Figure 4 - AMD Cypres SIMD engine (old nomenclature)), in AMD’s nomenclature, each of which contains a few ALUs (Arithmetic Logic Units) (fig). ALUs can perform single precision or integer operations. To perform double precision operations, some architectures group ALUs together which yielding double precision performance which is inversely proportional to the number of ALUs grouped. This means that even though GPUs, in general, operate with a smaller clock rate with regard to CPUs, current hardware is able to perform *thousands* of arithmetic operations per cycle, and frequently hundreds of special functions (sine, cosine, square root, etc.). Since special function operations, also called transcendental operations, occur less frequently, there is usually only one special function processing element per engine. Another design groups ALUs, potentially sacrificing ALU power to perform a transcendental operation. The loss in processing power is offset by gained wafer which can be used for memory, control or processing. For this reason, algorithms that rely heavily on such operations such as Fourier transforms and random number generators are one of the prime beneficiaries of GPU computing.

It is valid to take a detour here and observe that comparisons regarding "processing power" are used rather loosely and can be taken to mean peak throughput, and this liberty was taken for the sake of comparison. In reality, achieving peak throughput or anything close to it is highly algorithm dependent which makes it difficult to speak of processing power in general terms. Such figures such as gigaflops and teraflops should thus be read as a comparison between the potential of rivaling technologies with the addendum that GPU potential is naturally harder to extract, given the many layers of parallelism present .

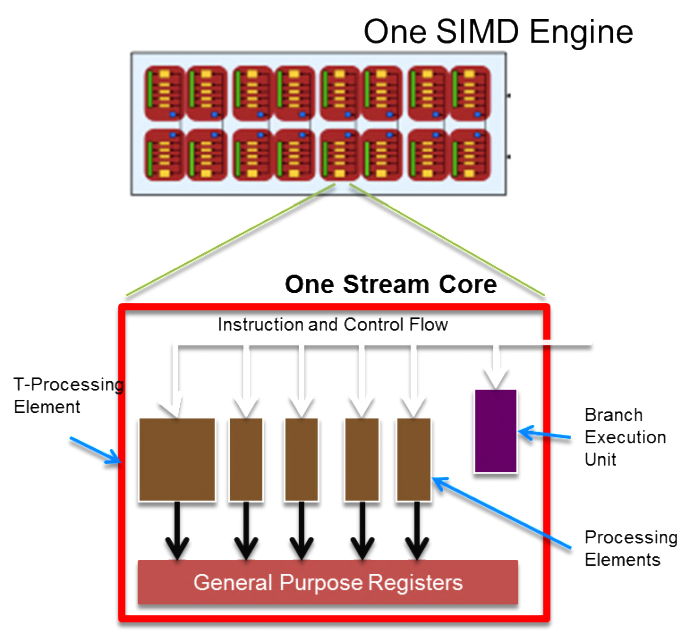


Figure 4 - AMD Cypress SIMD engine (old nomenclature)

With the advent of the first programmable shaders, although the potential gains were high, so was the initial investment in learning to program for general purpose computation on a GPU. Initially only programmable in assembly, GLSL (OpenGL Shading Language) greatly increased the ease of programming shaders by raising the level of abstraction and offering the developer the ability to program his own kernels in C variant. GLSL was also a standard to which each vendor had his own implementation. This meant that programmers did not need to learn vendor specific assembly code to write their own custom shaders. While this was more practical than dealing with instruction language, GLSL was still a language written for shaders and not for performing general purpose computations on a GPU. This introduced some awkwardness in the process of programming because of the necessity of treating any piece of data which was operated upon as a pixel or vertex, making code unreadable to a layman of the world of shader programming. General purpose computation with GLSL left an overarching feeling of being more a "hack" than something usable for serious development in the field of general purpose computing.

Nevertheless, the ease of being able to program a shader in C, which is heavily used by the scientific community and known to be performant for HPC made this new technology gain much wider appeal. Furthermore, for many programmers familiar with the OpenGL API and the intricacies of managing buffers and the interface in general, the transition was facilitated, since the primary focus of the language was to extend OpenGL's capabilities and this was, after all, their targeted audience. A boom in research ensued, which was primarily targeted at computer graphics but extended to other areas as well with the realization that many of the benefits reaped by this specific audience could be used by a wider family of algorithms.

While the hardware continued to evolve, hardware vendors realized that vertex and pixel shaders had many overlapping functions. By joining the two shaders into a unified shader architecture, manufacturers were able to save wafer area and pack more transistors into the chip, while also being able to balance the computational load when one form or the other was being more heavily required - further increasing processing power. Also, this required that a new level of abstraction be introduced, since shader was to be used for vertices and pixels, leading to a further increase in the programmability of the GPU.

Nvidia and ATI, the only two large GPU manufacturers who managed to survive the intense competition, recognized the potential of this new model and developed their own high level languages to facilitate the task of creating a program to run on a shader, commonly known as a kernel. Of the two languages created by Nvidia and ATI (AMD), respectively CUDA and Brook+, CUDA enjoyed the most hype and was embraced as the standard for GPU computing because of a gentler learning curve and higher level of abstraction which made it more accessible, as well as Nvidia's capacity to quickly develop a functional and mature programming language for the GPU. The fallback was that CUDA and Brook+ were still vendor specific languages and therefore allowed no flexibility for porting code to other platforms. Meanwhile, IBM also found its way into the general purpose computing scene by creating the acclaimed CELL BE processor for the Playstation 3. Promptly recognized as a viable alternative to GPUs for massively parallel computationally intensive tasks such as dense solvers (Kurzak et al., 2007), the Playstation 3 spawned the creation of a number of Beowulf type clusters because the low cost per FLOP (Floating Point Operation), and the added benefit of networking and storage in a cheap setup. The cell architecture itself is an interesting compromise between traditional CPUs and the first programmable GPUs. The added flexibility of the CELL BEs multiple cores with DMA (Direct Memory Access) controllers and with a threading model in mind made it a powerful processor for general purpose computing. With it IBM added yet another language to the growing list of languages tailored for the massively multicore era.

* + 1. Programming for the Unified Shader Architecture

OpenCL was therefore created in an industry effort to standardize the programming for these different platforms. While each vendor would have his own implementation, since there was a standard they would have to adhere to, the portability to different platforms would be guaranteed. This was made possible because while the underlying hardware was completely different when dealing with a DSP (digital signal processor), CPU or GPU, all of these have hardware that can benefit from exposing vectorization, memory hierarchies and parallelism with a greater detail and control. In essence, what was needed was an abstraction to deal with programming for multiple cores, possibly hundreds, and to be able to manually manage the different memory hierarchies, mapping the hardware abstraction to a programming model [Figure 3 - *ATI* 5870 Architecture (Copyright AMD)].

OpenCL was conceived with this purpose, and has been constantly growing in adoption, posing itself as a potential new industry standard for heterogeneous programming.

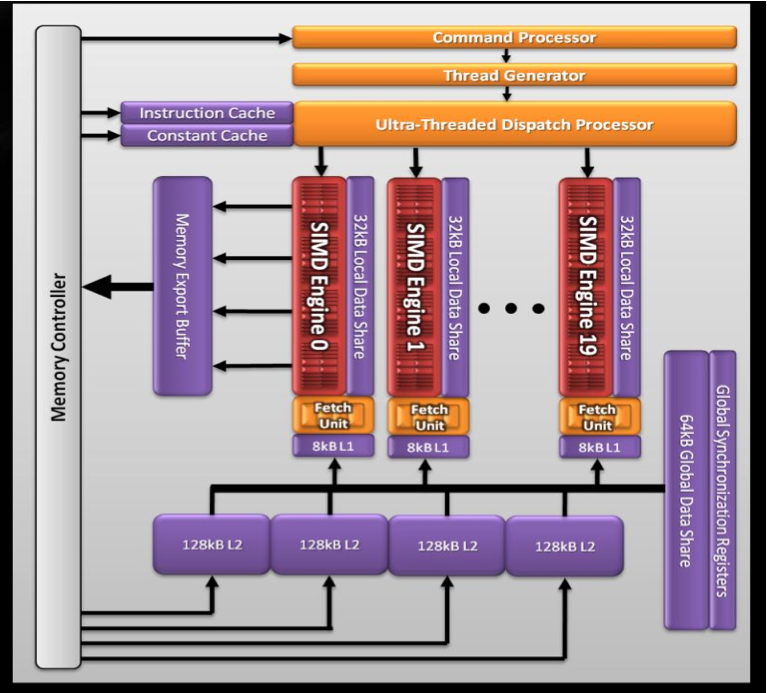


Figure 5 - *ATI* 5870 Architecture (AMD Parallel Processing Guide)

The standard is supported by most of the large processor vendors, including the largest growing market segment of the industry, that of mobile processors for smartphones and tablets which utilize architectures which can greatly benefit from such a programming paradigm. It is also valid to note that by creating a high level language that gives the developer explicit management of the different levels of the memory hierarchy also greatly benefits the CPUs, which have taken a road similar to that of GPUs to be able to keep up with Moore's law: add cores and layers of memory (cache) to increase performance. Nevertheless, the high-bandwidth memory coupled with smaller cache sizes which are present in a number of cores which is orders of magnitude higher make obtaining cache efficiency for a GPU a completely different task (Govindaraju, N. K. and Manocha, D.), not to mention communications directives which permit block memory transfers an synchronization at some level. The fact that OpenCL and other frameworks address this problem is what makes it, in part, possible to extract more performance out of such a wide range of platforms.

1. OpenCL
   1. Introduction

OpenCL is a standard defined by the Khronos Group, which is formed by a consortium of major players in the technological industry including hardware manufacturers, software developers and middleware vendors with the purpose of creating an open and portable API for accessing the capabilities of multicore systems. OpenCL provides a level of abstraction which enables the developer to write code for heterogeneous platforms consisting of CPUs, GPUs, APUs (Accelerated Processing Units), DSPs (Digital Signal Processors) and Accelerators (e.g. IBM Cell BE processor) within one unified programming model. Through the OpenCL programming language it is possible to write code based on the ISO C99 standard (ISO/IEC 9899:1999) extended to write functions, called kernels, which execute on the device. Through the abstraction of the hierarchical memory present in modern GPUs and CPU with multiple levels of cache, OpenCL provides the capability of creating code which will take advantage of greater bandwidth provided by some levels of memory to create optimized code without the need resort to vendor specific instruction language. Although programming in instruction language provides a greater level of control and performance, using OpenCL in to program in C99 like language makes the code comparatively more readable and maintainable, besides being a cross-platform solution. A large number of resources for programming with OpenCL are available online. Thorough references on utilizing the framework, describing the programming language and the functionality available through the API exist, such as: (Benedict R. Gaster et al., Heterogeneous Computing with OpenCL, 2012), and (Aaftab Munshi et al., OpenCL Programming Guide, 2012), which should be consulted for a broader understanding of programming with OpenCL. This chapter will offer only a small overview of the framework necessary for the comprehension of posteriorly proposed optimization strategies.

* 1. Execution Model

OpenCL works on the abstraction of a GPU composed of multiple compute units. The compute units are, in turn, composed of multiple processing elements which work as vector processors as shown in fig. The processing elements contain multiple ALUs and, in some cases, a SFU (special function unit) which some vendors call “cores”. Since the underlying hardware differs from that of a CPU, certain concepts such as that of a CPU “core” do not map directly between both architectures and the term “core” must be taken with a grain of salt. For that reason, the ALUs are called *processing elements* within the context of OpenCL and also on the newer AMD architectures (Southern Islands Devices) where there is one ALU per processing element [Figure 4 - Compute Device Structure (Copyright AMD)]. Switching between different vendor’s documentation and the official OpenCL specification or even between different families of graphics processors can generate some confusion regarding nomenclature. It is therefore advised to focus on one development platform for the purpose of simplicity. After the basics are well ingrained, consulting other vendor’s architectures and documentation becomes easier, and parallels or differences between implementations become clearer.

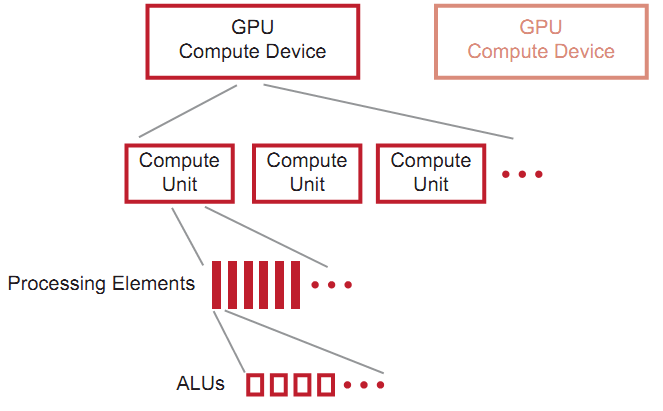


Figure 6 - Compute Device Structure (Copyright AMD)

A kernel, as previously described, is basically a C function with some syntax limitations, particularly with regard to memory management. Examples of immediate differences which can be observed are pointers, which do not work in the same way as C and cannot be dynamically allocated, vector data types which are built in, native parallel processing synchronization directives and qualifiers which suggest to the compiler where to store data (locally, registers, globally, etc.). Alas, aside from a few differences a C programmer can easily transition to writing OpenCL kernels. There are also has many built in functions which, given native hardware support, can be used to perform more complex instructions in a single cycle such as MAD (fused multiply and add), 2D and 3D vector operations such as scaling, addition, dot product, sine, cosine and random number generation, and various others. Many of these functions would require dealing with assembly code to perform which would be compiler specific and, therefore, limit portability or at least require code duplication.

Memory management within the kernel is in large part deferred to the OpenCL library running on the host and cannot be dynamically managed from with a running kernel, which is one of the shortcomings of programming for the GPU. Memory is managed through the creation of buffers which indicate the amount of memory which will be allocated and queues which can be used to schedule the asynchronous or synchronous transfer of data to the GPU. Optionally, the GPU can operate directly on the data stored in RAM, but since transfers between the GPU and host memory are very costly, it is important to analyze if this would is a good strategy (multiple reads from the data, for example, would indicate that copying the data to GPU would be a more suitable choice).

Each compute unit runs an instance of this kernel, called a *work item* which is somewhat analogous to a CPU thread. Within a compute unit, all work items execute the same instruction, similar to vector processors with the exception that they are not all executed in lockstep. The work items are grouped (called a workgroup in OpenCL terminology) and within a group execute in lockstep just as a vector processor but multiple groups are pipelined, which is explained further when addressing wavefronts. In a graphics card, hundreds of work items can be generated and because of the common use with image and graphics processing, these work items are arranged in an N-dimensional grid, called *ND-Range.* Utilizing a 2D range for images, where each work item corresponds to a pixel, and 3D range for a 3D model, where each work item corresponds to a voxel (3-dimensional pixel) to deal with images or 3-dimensional objects makes it easier for the programmer to create algorithms around these objects, dealing with them in a more natural way. Using this indexing has other benefits such as increasing cache hits by exploring data locality for certain algorithms, such as antialiasing and a variety of filters which operate on pixels and their neighbors, besides making the code more elegant and maintainable by providing a standard indexing scheme. For linear algebra operations, utilizing a 1D range suffices, although some matrix algorithms could potentially benefit from exploring data locality from a 2D range, especially within the context of dense matrices.

Work items are divided into groups called *workgroups.* A workgroup has access to common shared memory and synchronization directives which it can use to perform operations such as parallel reductions. The task of determining the size of the workgroups is deferred to the programmer, and the size chosen determines how local much local memory will be reserved for each group since workgroup size is determined by hardware and local memory is determined by hardware. The selection of work group size, therefore, has a large impact on the overall performance of the algorithm. Determining the best workgroup size is still something of an art and oftentimes counterintuitive, since there are always a number of conflicting interests e.g., larger numbers of work groups are generally good for filling wavefronts (pipelined workgroups) which on the other hand implies on less local memory per workgroup to perform parallel reductions or other operations which take advantage or local memory synchronization. Work items are executed in fixed groups which depend on the hardware, called wavefronts. When workgroup sizes are not evenly divisible by the wavefront size, some work items will remain idle while executing a wavefront, being that the wavefront is the minimum unit of execution.

While wavefronts are determined by hardware, grouping is determined by software. For the hardware used in the benchmarks, the wavefront size is 64. This is because there are 320 processing elements (fig) divided into 20 compute units yielding 16 cores per compute unit. Without going into much detail, the VLIW (very long instruction word) architecture permits the compute unit to issues one big instruction for 4 cycles at once. To have the hardware completely utilized it would be necessary to have a group size of, at least, 64 (one work-item per core times cycles necessary) and be a multiple thereof, since any factionary Dnumber would leave a wavefront only partially full (AMD Accelerated Parallel Processing Guide, 2012). Besides having the compiler extract instruction level parallelism through the VLIW architecture (which has been deprecated in more recent hardware), another reason to which also benefits more recent GPU architectures is to diminish the latency in memory access.

Due to characteristics of the GPU model, memory latency is preponderant factor in GPU efficiency. To minimize the latency conundrum, GPUs pipeline work item execution so that while one work item fetches information from memory, other work items are executing. Up to four work items of the same wavefront can be pipelined on the same processing element(in a full wavefront). Performance is therefore dependent on keeping the pipeline full so that latency does not affect performance, a technique called *latency hiding*. An example of the pipelined execution can be seen in fig.

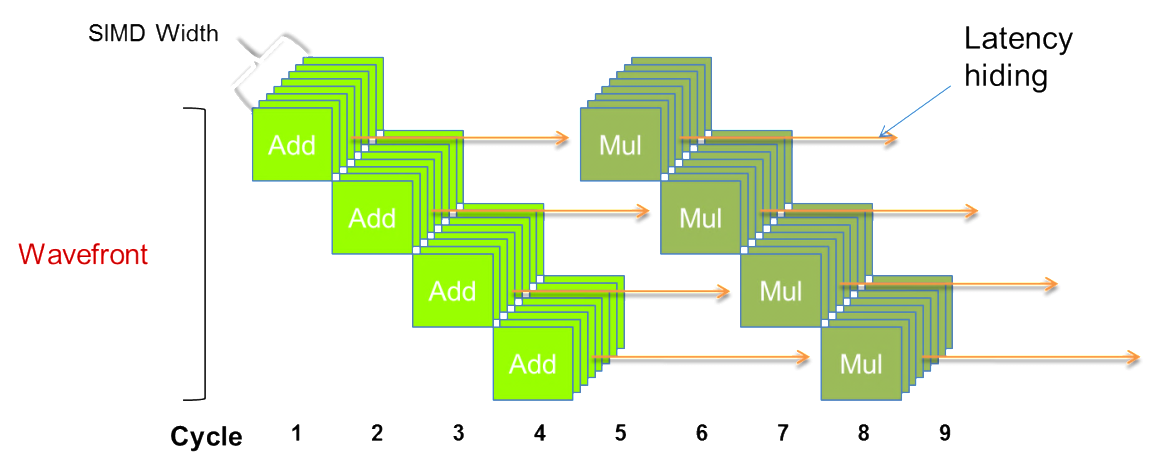


Figure 7 - Memory Pipelining

Figure 8 - Latency Hiding

As a concrete example, one of the GPUs used for the benchmarks, the ATI Radeon™ HD 5870 HD has 20 compute units which each contain 16 stream processing elements. These cores each contain 5 ALUs adding up to a colossal 1600 ALUs. The wavefront size is 64 work items. While there is an immense added complexity, the theoretical peak throughput of this already outdated graphics card is 2.72 teraflops in single precision and 544 gigaflops in double precision (processing elements are linked together to perform double precision arithmetic leading to lower throughput). Nevertheless, as has been noted many times, achieving anything close to maximum throughput requires a special set of problems that can be parallelized into hundreds of different portions, preferably without the need of communication between these portions, and without branching. It is clear that the GPU evolved with a very specific problem in mind. Still, even a fraction of peak performance can provide large gains when compared to CPUs given the raw performance capabilities of modern GPUs.

On more recent models, the choice has been made to sacrifice peak throughput in favor of a model more suitable for general purpose computing. The more recent models offer greater flexibility for wavefront execution and have shifted the job of scheduling wavefronts from the compiler to the hardware simplifying the compiler’s job at the cost of occupying wafer space. On the other hand this allows more flexibility for wavefront executing regarding branching and exploring thread level parallelism. On most SIMD execution models, branching takes a heavy toll since lockstep execution does not allow diversion of execution paths and therefore should be avoided. This deficiency is partially mitigated by the more recent developments at the cost of some computational potential. Also while one of the tested GPUs (Radeon 5870) has only 1/4th of the computational capacity for double precision compared to single precision since 4 ALUs are joined to perform one double precision operation, on most modern GPUs this has been reduced to 2 ALUs, leading to only half the performance penalty.

* 1. Memory Model

OpenCL contains four main memory domains: private, local, global and constant memory (Figure 9 - OpenCL Memory Architecture). Private memory pertains only to a work item and is not visible by other work items. It is the smallest level of granularity at which memory is divided on the GPU, and a rough analog in traditional CPU computation would be a general purpose register (GPR). Local memory is a level above, and can be shared between work items contained in a common workgroup and provides synchronization directives. Here an analog to CPU architecture would be something akin to the L2 cache introduced by Intel in the Core 2 Duo, which is common to both cores. Global memory is visible by every work item and is at the highest level of granularity, being one monolithic chunk of memory every work item has access to but without possibility of synchronization between work items and can be thought of as an analog to traditional RAM. The global memory present in GPUs is actually a special type of RAM where the main difference is the bandwidth, much higher in the GPU than between the CPU and traditional RAM.

As in the different cache hierarchies in a CPU, there is progressively more memory on chip moving up towards global memory on the graphics board, while memory access speed decreases by roughly an order of magnitude for each level (which can be used as a rule of thumb).

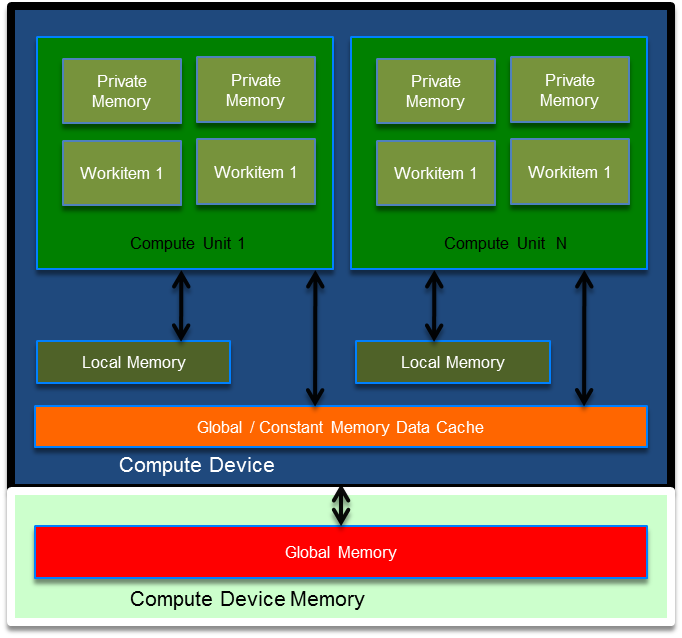


Figure 9 - OpenCL Memory Architecture (Khronos Group)

The many layers the programming model contains, be it in execution or in memory management, are what make programming in OpenCL challenging. Because of this, more variations of how to utilize the different levels of memory are possible adding to the complexity of the program, even more so when considering clusters of computers which add another layer of data locality to juggle. Wrapping one’s head around such concepts takes time to get used to, even for experienced programmers, particularly because the programming model adds many dimensions to be managed - beyond what programmers are generally used to.

1. Sparse Matrix Data Structures

The formats used to store sparse matrices are central to the performance of algorithms which rely heavily on matrix-vector or matrix-matrix operations. Contrary to a dense matrix which stores all the information in a matrix in a contiguous fashion, a sparse matrix storage format usually stores only nonzeros or at least tries to store the smallest number of nonzeros possible. This is necessary because the storage size increases exponentially with the matrix dimension, while for many problems the number of nonzeros increases at a smaller rate (usually linearly). For very large matrices, as generated by the finite element method for structures with many elements, the storage space necessary when using a sparse format is only a small fraction of the space that would be used by a dense matrix. Using sparse matrices enables analysis of structures that otherwise would not be possible in personal computers and graphics card memory, or would be possible with a great performance handicap of being forced to use hard disk space which is an order of magnitude slower than RAM.

The format used to store sparse matrices determines how the data of the matrix will be accessed and stored, i.e. the memory access pattern, which impacts the time to read or write a matrix entry and also the total size required to store the sparse matrix. Most modern formats are relatively cheap to store and with very little variability between format storage sizes to make a significant difference. As observed in some operations in linear algebra sometimes there is some reuse of rows or scalars which makes dividing a matrix into rows or columns, depending on the operation which the format is tailored to, a natural choice. Depending on the way a matrix is divided, indexed, and stored, these decisions can make certain formats better suited to one general way the nonzero elements are disposed (sparsity patterns) in the matrix or to solve one particular problem.

Matrix operations can be shown to take up a considerable amount of the computational time when profiling, and are therefore sections which provide considerable impact upon performance and prime candidates for optimization. Thorough discussions on the difference between data structures and which structures are more adequate for given problems and hardware architectures, such as vector architectures and GPUs have thus been widely discussed and large bodies of comprehensive studies are available (Wafai, M., 2009). Of particular interest are the formats created for vector architectures dating from as early as the 70s, targeting vector supercomputers which share similarities with modern multicore architectures and can offer insight as to where gains can be achieved, given that sparse matrix operations are still relatively inefficient, utilizing about 10-20% of peak performance (Stathis, P. T., 2004). Only a succinct description of each format employed will be provided with brief implementation remarks (Code for all sparse matrix data structures can be found on appendix A)

* 1. Diagonal Format

The diagonal sparse matrix format, also called compressed diagonal, stores sparse matrices as diagonals, which are stored if they contain a nonzero. The main advantage of the diagonal sparse matrix format is that given a regular band, this format can have a very small memory overhead, especially for stiffness matrices which have gone through a simple nodal reordering algorithm. If a matrix were to have a full row, this format would be worse than storing a dense matrix. It enjoyed popularity within the finite element community for structures because of the regular band densely packed about the diagonal makes storing the matrix by diagonals a natural tendency. Unfortunately the small gain in memory consumption is offset by poor performance for the required operations, as the format does not offer an efficient way to access data for operations such as matrix vector multiplications (one must access all diagonals to perform a matrix vector operation), even more so for vector processors, and for random insertion which is important on the assembly stage where format performs as well as the other formats. It requires storing the number of diagonals, a vector with all the stored diagonals, and an n-sized vector to indicate the offset from center diagonal of each diagonal. Positions are defined by convention. The general C data structure would include these fields:

typedef struct DiagonalFormat {

int ndiagonals;

fem\_float\* diagdata;

int\* diagoffset;

} DiagonalFormat;

* 1. Compressed Sparse Row

This particular implementation of the CSR format may have a few differing points from other implementation with regard to memory allocation, namely that each row has its own pointer which adds a memory allocation overhead but at the same time permits independent memory reallocations for each row. In the context of finite elements this decision allows performance gains at the assembly stage when used in conjunction with mesh coloring, since independent memory management of rows to be resized in parallel, avoiding the issues of race conditions other formats have to consider. The strategy used was to create a data structure to hold each compressed row's data as follows:

typedef struct ROWdata {

fem\_float\* values;

int\* col;

int numNZ;

int maxSz; // Used to control reallocations

} ROWdata;

The matrix is then represented by a vector of these structures which represent a compressed row. The advantage of utilizing this format for finite elements is that it allows rows to be reallocated in parallel without needing to move a chunk of memory, which provides more flexible filling, compared other data structures which group all non-zeroes into one vector at the cost of slower deallocation. Also, since dynamically created pointers to pointers are not supported by the GPU, the data is not directly transferrable to the GPU.

* 1. Ellpack Format

While the Ellpack format lost some ground with the demise of most of the vector supercomputers, replaced by clusters containing thousands of regular CPUs which had become a more cost-efficient solution, the advent of GPGPU programming brought back the interest in formats better suited for vector architectures, given the similarities both share. Recently, other formats such as Ellpack-R (Vásquez, F., Fernández, J. J. and Garzón, E. M., 2010) based on the Ellpack format have been devised to take advantage of vector processing and address prerequisites such as data alignment and to overcome some performance problems which can arise due to varying numbers of nonzeros per row. The traditional Ellpack format, besides being especially well-suited for vector architectures (Bell, N and Garland, M.) also performs well in the context of finite element problems in structures, particularly because as was previously mentioned in general the problems have a somewhat regular banding. This makes storing padding for rows less common and the storage scheme more efficient. For problems with more variable row width, formats such as described in the former reference would be a better option, adding the storage of number of nonzeros in each row as additional information. To store a matrix in the Ellpack format the matrix has all the nonzeros shifted left and the matrix is padded to length of the row with maximum number of nonzeros. The entries are then scanned column-wise to a matrix. A matrix of the same size is stored keeping an integer with the column value for the corresponding entry. The row can be inferred since they are scanned sequentially when building the matrix, restarting at the first row after the first column is entirely scanned. The data necessary to store the information is therefore minimally:

typedef struct Ellpack {

fem\_float\* data;

int\* colindex;

int maxrownnz;

} Ellpack;

* 1. Eigen Format

While not a separate format in itself, the Eigen sparse matrix library was added to the sparse matrix interface to have a library that performs well to compare to. Documentation suggests the use of a variant of the CSC (Compressed Sparse Column) format. While our formats were tailored for our problem and our needs, Eigen which is a general format performed well, particularly the dynamic format which has a memory overhead which is a bit larger than the regular format but has better performance regarding random insertion which is where most of the time is spent assembling a stiffness matrix. Another advantage of using the Eigen library is that since it is a header only library, it can be compiled with the code utilizing the same compilation optimization level as the other formats.

* 1. Searching index arrays

Sorted arrays are commonly used to store indices and thus getting or setting information in the sparse matrix passes necessarily through these operations. When assembling the stiffness matrix, entries have to be read and updated multiple times per element and the speed at which memory can be retrieved and stored impacts the performance greatly at this stage. Efficiently searching for the correct index is therefore a very important part in having a sparse structure that performs well. While understanding of Big O notation is necessary to understand the performance of an algorithm, relying solely on the limiting behavior of the algorithm as an optimization choice can yield counter-intuitive results if the nature of the problem is not taken into consideration. For problems with a small number of elements, for example, although considering Big O notation a certain algorithm may yield better performance, setup costs may make the initial overhead trump the gains of a theoretically faster algorithm. As Donald Knuth points out “People often abuse O-notation by assuming that it gives as exact order of growth; they use it as if it specifies a lower bound as well as an upper bound” (Knuth, D; 1997). It is valid to point out that one could theoretically take Big Omega to be able to find a lower bound and from a theoretical perspective select the appropriate algorithm. Nevertheless, other factors such as caching, memory and bandwidth boundaries as well as the distinct performance of multiple memory hierarchies which are paramount for optimization are not captured by these models.

In the context of searching within a row of a matrix generated by the finite element method, row size is a function of mesh structure. The belief that selecting the “best” algorithm according to Big O notation in hope of, at least, building a system that will scale well is in this case erroneous, since the algorithm will scale according to mesh structure and not problem size. Considering that elements connected to many other elements (e.g. “star” structure) are the exception rather than the rule, it would be wiser to program for the most common connectivity of each element. For each element type analysis was performed in order to find the impact of searching algorithms on sparse matrix performance. This analysis does not include OpenCL searching within the GPU, which has its own particularities and is discussed in the appropriate section. Differences between the different searching techniques have been discussed at length at: <http://schani.wordpress.com/2010/04/30/linear-vs-binary-search/>.

* + 1. Linear Search

The traditional linear search algorithm was used as a basis to compare with other approaches used for searching an ordered array. The pseudocode for finding the correct place to insert a value in an array which returns and index is:

**FOR** position = 1 to length of vector

**IF** vector[position] >= value **THEN**

**RETURN** position

**END IF**

**END FOR**

We can then check if the value exists to discover if the value has to be inserted or modified or simply return that it does not exist in the case of a query. For small vectors, or in the case of sparse matrices for narrow bands, benchmarks indicate that contrary to popular belief, linearly searching across the vectors usually is faster than performing a binary search.

* + 1. Binary Search

The implementation for binary search was also used as a basis for posterior comparison with other versions. This implementation uses only one conditional per loop. The pseudocode for the implementation is as follows:

**STORE** variable min set to zero, variable max set to length

**WHILE** (min < max)

**STORE** middle is average of min and max

**IF** (searched value > vector[middle]) **THEN**

**STORE** min is middle + 1

**ELSE**

**STORE** max is middle

**END IF**

**IF** (searched value <= min) **THEN**

**RETURN** min // Return insert position

**ELSE**

**RETURN** -1 // Indicates key was not found

**END IF**

**END WHILE**

**RETURN** min

* + 1. Linear Search with SSE

Using the SSE instruction set it is possible to speed up our linear search algorithm to search more positions per clock cycle. This can be done because utilizing special high speed CPU registers we are able to process multiple operations per clock cycle and, therefore, reduce the number of clock cycles necessary to perform the required operation. One of the downfalls is that aligned memory access is required to transfer the data efficiently to the registers, which impacts the requirements of the underlying data structure. Besides this, since up to 16 comparisons can be fitted into the registers, larger array sizes benefit the most from utilizing SSE linear searches. Second to all this, while the primary objective of writing optimal code is performance, utilizing SSE instructions makes the code harder to read, debug and maintain, especially considering that the code is compiler dependent and thus compiling for GCC would require extending the code with compiler directives. The portability issue could be resolved utilizing OpenCL, but for a lightweight function which is called millions of times, even a very small overhead can be detrimental to performance. The SSE code utilized and compiled with MSVC was the following (description of the code was added as comments):

// Value to be compared is copied into 4 vector positions

\_\_declspec(**align**(16)) int v4[4] = {val, val, val, val};

const \_\_m128i\* v4ptr = (\_\_m128i\*)v4;

// Loads comparison values into registry

\_\_m128i key4 = \_mm\_load\_si128(v4ptr);

int i = 0;

\_\_declspec(**align**(16)) unsigned short res;

// Does search 16 elements at a time

for (i = 0; i <= len; i += 16) {

//

const \_\_m128i\* in0ptr = (\_\_m128i\*)&intvector[i ];

const \_\_m128i\* in1ptr = (\_\_m128i\*)&intvector[i + 4];

const \_\_m128i\* in2ptr = (\_\_m128i\*)&intvector[i + 8];

const \_\_m128i\* in3ptr = (\_\_m128i\*)&intvector[i + 12];

\_\_m128i in0 = \_mm\_load\_si128(in0ptr);

\_\_m128i in1 = \_mm\_load\_si128(in1ptr);

\_\_m128i in2 = \_mm\_load\_si128(in2ptr);

\_\_m128i in3 = \_mm\_load\_si128(in3ptr);

\_\_m128i cmp0 = \_mm\_cmpgt\_epi32(key4, in0);

\_\_m128i cmp1 = \_mm\_cmpgt\_epi32(key4, in1);

\_\_m128i cmp2 = \_mm\_cmpgt\_epi32(key4, in2);

\_\_m128i cmp3 = \_mm\_cmpgt\_epi32(key4, in3);

\_\_m128i pack01 = \_mm\_packs\_epi32(cmp0, cmp1);

\_\_m128i pack23 = \_mm\_packs\_epi32(cmp2, cmp3);

\_\_m128i pack0123 = \_mm\_packs\_epi16(pack01, pack23);

// Result is saved as a mask (fits more comparisons)

res = \_mm\_movemask\_epi8(pack0123);

if (res != 0xffff)

break;

}

int count = 0;

// To find position scans the bits of the resulting mask

if (res) {

unsigned long rb = 0;

\_BitScanForward(&rb, (unsigned long)~res);

count = rb;

}

return i + count;

* + 1. Performance

The best compromise found between overall performance and vector length was with the simple binary search. For analysis utilizing the Q4 and Q8 element where the connectivity is smaller, linearly searching yielded marginally better results. Utilizing SSE made results vary since performance having the searched vector not be a multiple of 16 leads much redundant computation, and being that vector widths for the meshes analyzed do not surpass 64, vectorization could not be adequately utilized to its potential. Nevertheless, forcing vectors to reallocate to multiples of 16 and memory alignment on these vectors could potentially lead to good results if elements with a higher number of nodes were to be utilized.

1. Implementation of The Finite Element Method

To perform analysis utilizing the finite element method, simple models were constructed for didactic purposes utilizing modeling software developed by MVGEO, the modeling and visualization group which builds modeling tools for geomechanical analysis for Petrobras, at TecGraf for generating two and three dimensional meshes. While the modeling tools permit much higher degrees of complexity, the focus of this study was to determine the effect of sparse matrix formats on analysis and, as such, only simple linear elastic analyses were performed.

The two dimensional mesh consists of a square with increasing numbers of subdivisions on each side, which supplied the analysis with meshes of quadratically increasing numbers of elements and, consequently, degrees of freedom. The square was constrained on both vertices of one side with one vertex having a free vertical degree of freedom to allow dislocation, given a poisson ratio between 0 and 1 and a horizontal load was applied to one of the opposing nodes (Figure 9 – Example two dimensional mesh).

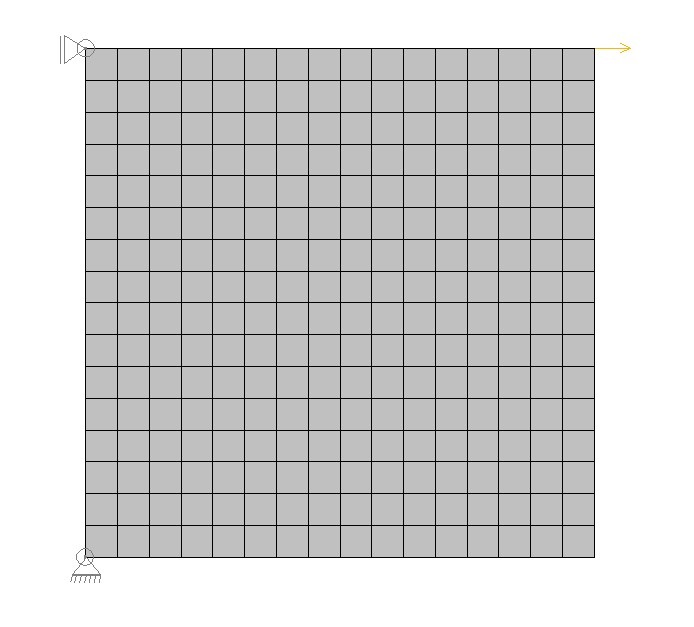


Figure 10 – Example two dimensional mesh

While the model is simple, the analysis provides results which demonstrate phenomena such as tension concentration, rotations due to deformation of the mesh and non-deterministic stress distributions which suffice for the purpose of benchmarking the performance of the analysis with regards to the number of degrees of freedom, the traditional metric of complexity of a finite element problem. To exemplify, results can be seen for a model utilizing Q8 quadratic elements with results taken at gauss points and extrapolated to nodes, with smoothing applied to visualize continuous resulting bands of intervals of equal stress (Figure 10 - Example mesh analysis results).

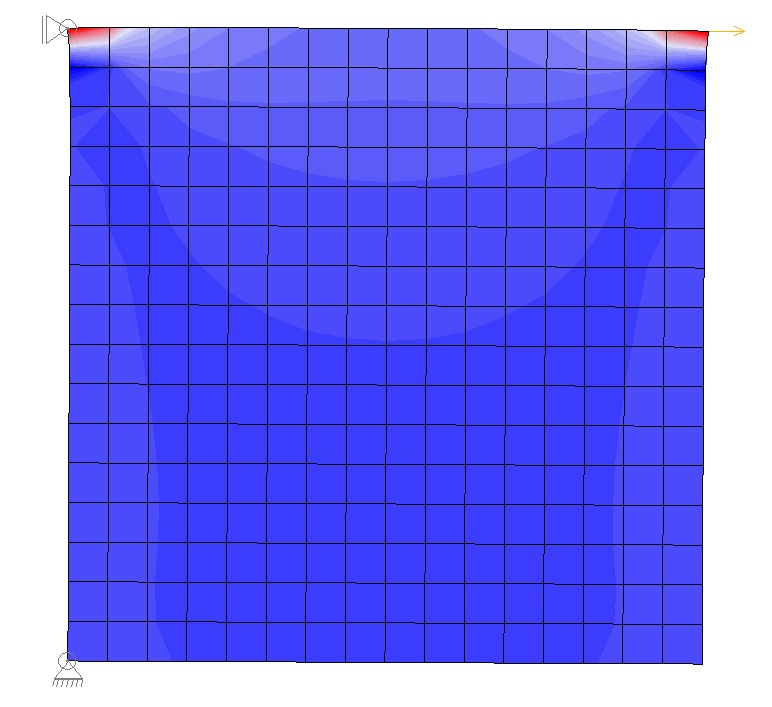


Figure 11 - Example mesh analysis results

Similarly, the three dimensional counterpart mesh was built utilizing a cube with increasing subdivisions on each side (Figure 11 - Three dimensional mesh) making the growth of the number of degrees of freedom with regards to the number of subdivisions increased proportionally to the number of subdivisions on each of the segments cubed. The difference to the two dimensional model is that the cube has the force applied to only one node of the opposing face, resulting in torsional forces on the topmost face which can only be adequately captured by three dimensional finite element analysis. All models utilized the same elasticity modulus, poisson coefficients and dimension, and for benchmarking only the number of subdivisions was changed.

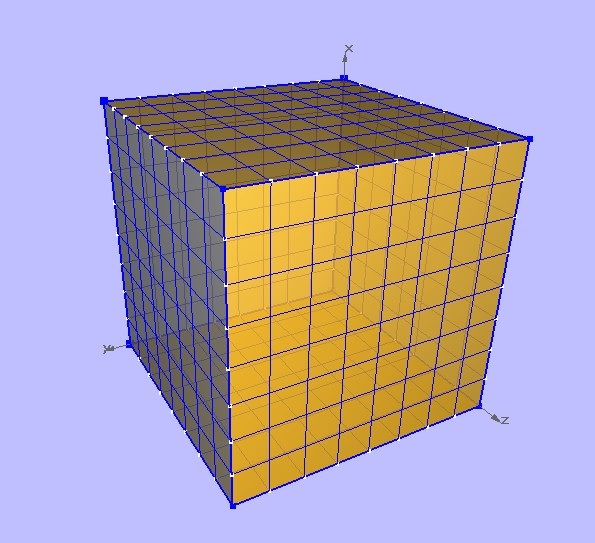


Figure 12 - Three dimensional mesh

The analysis utilized isoparametric finite elements, considered an effective method for most practical analyses (Bathe, K. J) vis. utilizing generalized coordinates and implementation details of constructing interpolation functions and the constitutive matrix were extracted from the former text.

* 1. Determining the Structure’s Stiffness Matrix

Determining the stiffness matrix is one of the main steps in performing analysis by the finite element method. The global stiffness matrix is required to solve the stiffness-displacement equation

and consequently determine the displacements for a given model and it’s boundary conditions. To first determine the global stiffness matrix, we first determine the local stiffness matrices of each element with their corresponding local degrees of freedom in the model and then assemble to a global stiffness matrix which represents all of the model's degrees of freedom. The element stiffness matrix can be evaluated as:

In the two dimensional case:

Where **B** is the strain-displacement transformation matrix and **C** is the constitutive matrix. The work of calculating the integral is facilitated by using normalized (natural) coordinates, i.e. isoparametric elements, since performing Gaussian numerical integration is always done in the same domain. Also, isoparametric elements permit us to interpolate either displacements or coordinates from nodal values (Cook, R.D. et al). To transform from natural coordinates to the global coordinate system, the strain-displacement transformation matrix can be used and is calculated by deriving the elements shape functions, which can be generated systematically using Lagrange interpolation polynomials (Rao, S. S.). The corresponding pseudocode for the algorithm is:

**FOR** element = 1 to number of elements

**STORE** element Coordinates from node coordinates and element connectivity matrix

**FOR** gauss point = 1 to (number of gauss points squared if in 2D / cubed if in 3D)

**STORE** r,s,t natural coordinate values and corresponding weights given the current gauss point

**COMPUTE** derivative of shape functions for current element type and r,s,t position

**COMPUTE** jacobian matrix given the element coordinates and derivative of shape functions

**COMPUTE** determinant of jacobian matrix

**COMPUTE** the inverse of the jacobian matrix given itself and its determinant

**COMPUTE** the derivative of shape function matrix in cartesian coordinates by multiplying the inverse of jacobian matrix by the derivative of shape functions in natural coordinates

**COMPUTE** B matrix by utilizing data in computed derived shape function matrix in natural coordinates

**COMPUTE** B transposed by transposing elements in B

**COMPUTE** CxB matrix by multiplying constitutive matrix C (given by constitutive model adopted) by strain-displacement transformation matrix B

**COMPUTE** temporary stiffness matrix by multiplying prior result, CxB, by B transposed and by the determinant of the jacobian matrix to map to the global coordiantes

**STORE** in local stiffness matrix values calculating by adding the partial results of the numerical integration

**END FOR**

**STORE** in global stiffness matrix values of local stiffness matrix by assembling matrix into corresponding global degrees of freedom

**END FOR**

Since the in assembly process the corresponding degrees of freedom are modified more than once, to parallelize this portion it is necessary to ascertain that concurrent threads are not simultaneously accessing the same element, which could lead to a race condition and yield erroneous results, e.g. two threads read the same value, increment the value by the amount that will be assembled, and then write to the same location irrespective of the work the other thread has done. To sidestep this problem it is possible to tell the compiler through OpenMP directives that a certain portion of the code should not be parallelized as covered in the OpenMP implementation section. Another possibility that maintains parallelism would be to calculate where these race conditions would arise, i.e. where elements have corresponding degrees of freedom, and “color” these elements with different colors. Therefore, while elements of different colors must be assembled serially, elements of matching color can be assembled without need for worrying about race conditions. Yet another option which is not covered in this text but which has been extensively discussed and which also yields good results would be to parallelize in the domain of the nonzeros (Cecka et al.). The different paths possible which were implemented are all shown in the following workflow Figure 12 - Stiffness Matrix Assembly Flowchart.

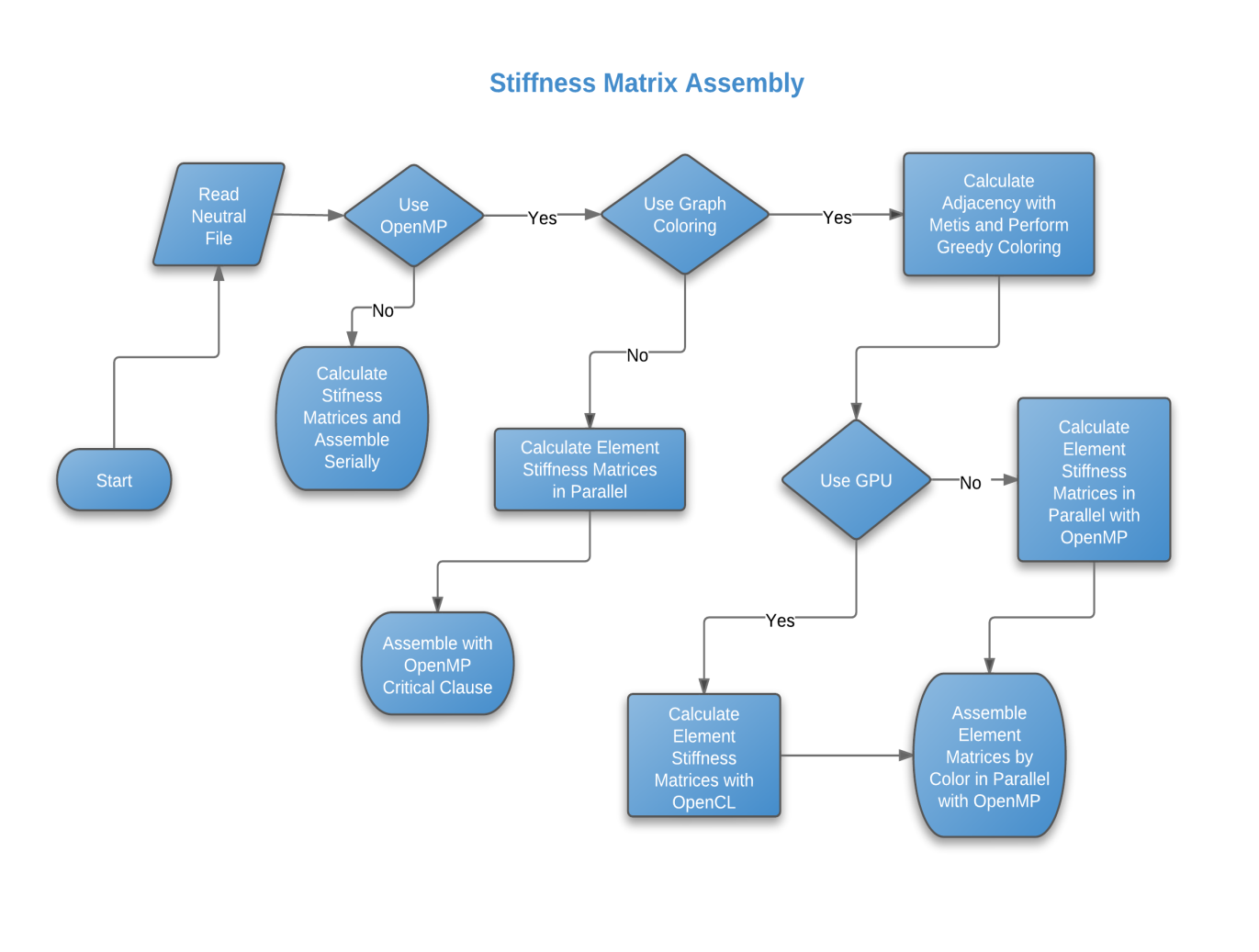


Figure 13 - Stiffness Matrix Assembly Flowchart

* 1. Parallel implementation using OpenMP

By utilizing OpenMP it is possible to obtain a considerable speedup in multicore systems without having to refactor a great quantity of code when using C or C++. It enables us to delegate and distribute work to different processors, while having control over locks, atomic decisions, load distributions and other features of parallel processing while maintaining a simple API.

One straightforward way to parallelize our code would be to compute the local stiffness matrices in parallel and then assemble to the global stiffness matrix. Care must be taken that race conditions do not affect the result, for while calculating the local stiffness matrices are independent calculations, when assembling the results race conditions may appear of two processors modify the same global degree of freedom at the same time. It is therefore necessary to utilize the CRITICAL construct from OpenMP to assert that only one thread performs this operation at a time. The resulting pseudocode would therefore be:

#pragma omp parallel for

**FOR** element = 1 to number of elements

**STORE** element Coordinates from node coordinates and element connectivity matrix

**FOR** gauss point = 1 to number of gauss points squared if in 2D, cubed if in 3D

Calculate local stiffness matrix

**END FOR**

#pragma omp critical

**STORE** in global stiffness matrix values of local stiffness matrix by assembling matrix into corresponding global degrees of freedom

**END FOR**

* 1. Parallel CPU implementation using Graph Coloring

By utilizing METIS, a third party library, to obtain a k=1 k-way graph partitioning (KARYPIS, G. and KUMAR, V., 1998), commonly referred simply as graph coloring (also known as map coloring where only directly adjacent elements cannot share the same color) of the elements of the mesh it is possible to eliminate race conditions at the assembly stage enabling parallelization of this stage of the finite element calculation process (Komatitsch et al., 2009). This is executed by passing metis the elements connectivity and receiving back the partitioned mesh, after which the proper mesh coloring can be carried out without much computational expense utilizing a simple greedy coloring algorithm (Markall, G., 2010). Metis then returns a colored graph which can be used to perform a quick greedy coloring and for a vector with all the elements for a given color. It is shown in the previous reference that there exists a coloring where the whole mesh can be colored in parallel in no more than n + 1 colors, where n is the largest connectivity of all elements. This means that the mesh can be colored in no more than n + 1 steps, where each step may perform its calculations in parallel. Meshes with elements connected to many other elements are also a problem for other balancing methods and are usually discouraged during modeling. Nevertheless the presence of a few of these will not sufficiently impact the performance gain since for each of the added steps of an added connectivity there are usually hundreds of elements to be executed in each step which dilutes the added time of having to force a few sequential steps. All elements of a given color are executed in parallel for a given step, since elements that were given the same color have no degrees of freedom in common. This way it is possible to perform parallel loops over the elements without having to worry about race conditions in accessing the same element, as would happen with a dense matrix.

With sparse matrices certain particularities arise with regard to memory management since adding an entry sometimes requires reallocating an array which holds all nonzero values and therefore requires that the operation be performed by a single thread. This is a bottleneck of many sparse implementations. The compressed sparse row implementation, as described in the sparse matrix formats section, permits parallel access to rows and with graph coloring we are guaranteed that no two shared degrees of freedom are accessed concurrently. This way we can modify values and reallocate sparse compressed row matrices without having to worry about race conditions. The pseudocode for the parallelized version using graph coloring is as follows:

**FOR** color = 1 to number of colors

**#pragma omp parallel for**

**FOR** auxelem = 1 to number of elements with this color

**STORE** find current element based on elements for current color and auxelem index

**STORE** element Coordinates from node coordinates and element connectivity matrix

**FOR** gauss point = 1 to number of gauss points squared if in 2D, cubed if in 3D

**COMPUTE** local stiffness matrix

**END FOR**

**STORE** in global stiffness matrix values of local stiffness matrix by assembling matrix into corresponding global degrees of freedom

**END FOR**

**END FOR**

* 1. Parallel implementation in OpenCL

The parallelization in OpenCL was made calculating all of the element’s stiffness matrices in parallel on the GPU and assembling element matrices onto the structures global stiffness matrix on the host device utilizing the CPU with OpenMP and graph coloring. On section 5.3 relative to the parallel CPU strategy with graph coloring implemented an outer loop on the elements which could be assembled independently by color and proceeded to calculate the element stiffness matrices and assemble them, while other colors were performing the same action. On the other hand, the GPU code implemented with OpenCL passed back a vector with all of the element stiffness matrices calculated utilizing parallelization and vectorization to calculate this local matrix, and subsequently performed the assembly in with graph coloring with OpenMP. Being that both the GPU and CPU implementations allow the assembly of elements that do not share any degrees of freedom, it follows that within a same color group no two elements will access the same row at the same row. This allows the possibility of assembling the elements in parallel, being that race conditions will not arise. This implementation can be scaled to larger problem sizes or distributed systems easily by assigning the calculation of local stiffness matrices to different GPUs and performing the assembly of sub-blocks of local matrices by the host, also permitting the utilization of both devices. Also, the implementation is independent of sparse matrix structure utilized, since the GPU only deals with the calculation of the dense element stiffness matrices, being adaptable to any structure but also being highly dependent on the sparse format's random insertion performance. Furthermore, for a given color multiple hosts can perform assembly in parallel and while GPUs are busy, given that proper care is taken to avoid race conditions regarding memory allocation (providing critical sections or preallocating matrices by counting nonzeros which can be done by third party libraries such as METIS itself).

One complication which arises is regarding memory management. To allow dynamic management of the stiffness matrices sizes, data structures must be reallocated when the number of nonzeros grow. Reallocation depends on the format chosen, with allocated data growing at varying rates which depend on the cost of reallocation for each format. The growth factors utilized were chosen to amortize this cost, whenever possible, without having much impact on the total size of the matrix and the implementation can be observed directly in the source code. Due to graph coloring, independent access to rows by each thread is guaranteed, but this is not sufficient to avoid problems coming from concurrent access. Since memory reallocation is done dynamically, concurrent memory allocation no only changes the underlying size of the data structures but frequently must reorder indexing and other format specific variables in use by threads that are concurrently in flight. This does not limit the whole operation of assembly useless when utilizing dynamic memory management, but requires extra care. The obvious approach would be to provide a barrier requiring all threads to be at the end of the function which inserts/adds nonzeros to the stiffness matrix before proceeding with memory reallocation, which would then be concluded by a single thread which would pass a message to the other threads indicating it was taking care of reallocation in case any other thread also discovered its own row needed reallocation. The trouble is that the synchronization primitives (barrier) and clauses necessary to apply this strategy cannot be applied within an OpenMP worksharing construct (Do/For Loops, Single and Sections directives) as defined in the OpenMP 2.0 specifications (OpenMP Review Board, OpenMP C and C++ Application Program Interface). Thus, the strategy adopted to avoid such problems was to create a Boolean value to indicate the matrix is close to being reallocated, and checking this variable on the outer assembly loop for each node to guarantee that if the matrix has a chance of being reallocated the assembly the node in the element must be done using a critical clause, analogous to the way described in the previous section. Given that for a three dimensional model there are three degrees of freedom to be assembled on the global stiffness matrix, two for a two dimensional model, we can assert that at most for each node there will be three insertions per row. Since the distance to reallocation for most formats depends on the numbers of nonzeros in a row or column, given the dimension of the model we can conclude that a safe distance to reallocation would be 4 nonzeros for a 3D model and 3 nonzeros for a 2D model. A flowchart summarizing the adopted strategy can be observed in Figure 10 - Avoiding Race Conditions through Coloring.

In this manner, using graph coloring permitted greater gains than possible using a naïve approach where all assembly should be performed with an OpenMP critical clause, to avoid race conditions. The benefit of adopting this strategy on distributed systems or by concurrently performing assembly in parallel while other element stiffness matrices are being calculated is a topic for further research. Different assembly parallelization implementations exist such as by row or by non-zero (Cecka, C. et al, 2011) as the domain of parallelization, which were very well documented and discussed.

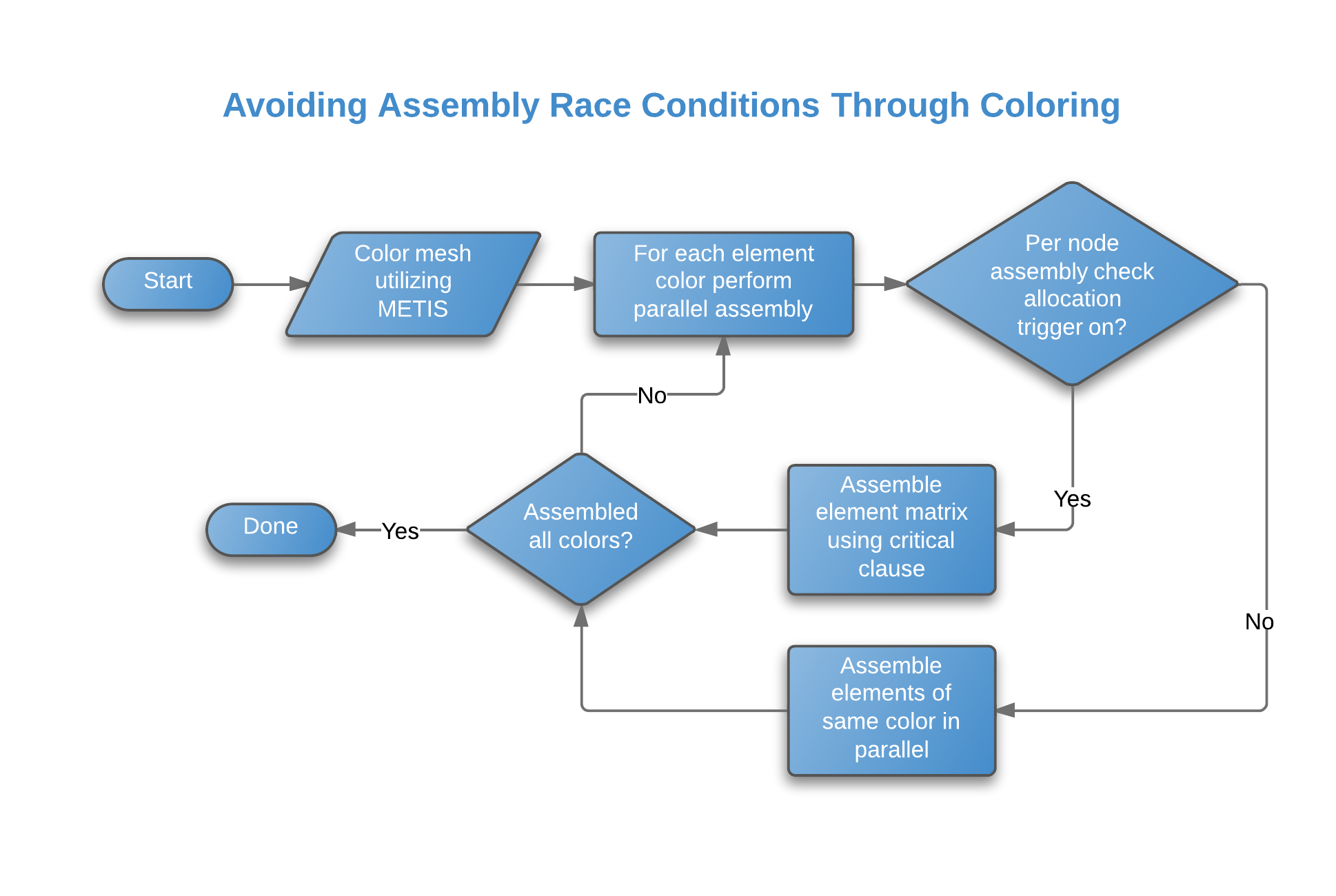


Figure 14 - Avoiding Race Conditions through Coloring

* 1. Setting Boundary Conditions

Given the focus on performance and optimization, boundary conditions were enforced using the penalty method and heuristics such as the “Square Root Rule” (Felippa, C., 1998) on the stiffness matrix before solution of the stiffness-displacement equation. While it is clearly not the best option in terms of numerical accuracy, this step does significantly affect the overall running time, and thus is not a primary concern in this context and the source provides the necessary abstractions for easily incorporating other methods in the future.

1. Solving for Displacements

The global stiffness matrix resulting from assembling the system of the previous chapter is a positive definite matrix. Between the solvers used for positive definite matrices, the conjugate gradient method is one of the most widely employed. Even current state of the art solvers such as Multigrid Solvers (MG) typically use conjugate gradient methods in solving the coarser grid stage (Geveler et al., 2011). Iterative solvers such as the conjugate gradient method are especially suited for solving large linear systems of sparse matrices, where traditional factorization methods would destroy matrix sparsity. The foremost reference regarding the conjugate gradient method used for this section is the seminal, aptly named paper "An Introduction to the Conjugate Gradient Method Without Agonizing Pain" (Shewchuck, 1994) which is highly recommended for a more detailed explanation of the development of the conjugate gradient method. Most material exposed in this section pertains to particularities of implementation regarding the finite element method.

Many implementations of the CG (Conjugate Gradient) method are available, and the implementation of the algorithm is relatively straightforward, especially considering the vast available literature on the subject. Being that in a traditional finite element CPU computation, the largest part of the computational time is spent on the solver which has been shown to be the object of considerable gains utilizing GPGPU computing (Helfenstein, R. and Koko, J., 2012) which makes it a prime target for exemplifying some OpenCL optimization techniques. These techniques can be used in a variety of different situations when programming for GPUs and also targeting vector extensions of CPUs, increasing the processing efficiency.

* 1. Naïve Implementation

Within the CG method, the operation which, by far, consumes most of the attention of the processor is the matrix-vector multiplication. This operation is the basis of almost all iterative solvers, particularly Krylov subspace methods. For this reason there is a considerable amount of academic research devoted to this simple linear algebra operation, making it an important benchmark for determining performance of different sparse matrix formats. Given an Ellpack format as described in section 4.3, A naïve implementation of the matrix vector multiplication operation using OpenCL would be:

uint gid = get\_global\_id(0);

if (gid < matDim) {

uint nnz = rowNnz[gid];

float sum = 0;

for (int i = 0; i < nnz; i++) {

int index = i \* matDim + gid;

int col = colIdx[index];

float aval = matData[index];

float xval = vector\_x[col];

sum += aval \* xval;

}

vector\_y[gid] = sum;

}

This implementation does not take into consideration local memory or grouping. The approach displayed simply assigns one thread to each row and calculates the dot product between the vector and a matrix row using a private register as an accumulator. It is valid to point out that although no workgroup indication appears in the kernel, being that no local memory was used nor were workgroup specific calculations were executed; when executing, a workgroup size must be defined and different sizes can still influence performance. That occurs because the workgroup size defines the number of wavefronts to execute which in turn affects memory allocation and warp execution, as per sections 3.3 and 3.2.

* 1. Introducing Vectorization Primitives

Another technique used to extract parallelism out of the GPU is to expose vectorization by using built in vector data types. This allows the compiler to more easily identify opportunities for using vectorization capabilities, oftentimes performing multiple arithmetic operations in one cycle. Code readability does suffer by treating variables as vectors which in practice do not possess a vector representation:

uint gid = get\_global\_id(0);

if (gid < matDim) {

uint nnz = rowNnz[gid];

float4 sum = 0;

for (int i = 0; i < nnz; i += 4) {

int4 index, col;

float4 aval, xval;

index.s0 = i \* matDim + gid;

index.s1 = (i + 1) \* matDim + gid;

index.s2 = (i + 2) \* matDim + gid;

index.s3 = (i + 3) \* matDim + gid;

col.s0 = colIdx[index.s0];

col.s1 = colIdx[index.s1];

col.s2 = colIdx[index.s2];

col.s3 = colIdx[index.s3];

aval.s0 = matData[index.s0];

aval.s1 = matData[index.s1];

aval.s2 = matData[index.s2];

aval.s3 = matData[index.s3];

xval.s0 = vector\_x[col.s0];

xval.s1 = vector\_x[col.s1];

xval.s2 = vector\_x[col.s2];

xval.s3 = vector\_x[col.s3];

sum.s0 += aval.s0 \* xval.s0;

sum.s1 += aval.s1 \* xval.s1;

sum.s2 += aval.s2 \* xval.s2;

sum.s3 += aval.s3 \* xval.s3;

}

vector\_y[gid] = sum.s0 + sum.s1 + sum.s2 + sum.s3;

}

* 1. Utilizing Local Memory

One possibility of taking advantage of local data and synchronization primitives within a local workgroup is to assign a workgroup to a row. This way, each workgroup is responsible for a dot product operation and this operation can utilize a parallel reduction. While it seems this approach would be faster, the large number of work items per workgroup means that each work item will also do less ALU operations, which can negatively affect performance. Additionally, although fewer cycles are necessary in a parallel reduction, memory latency is larger than the naïve case which uses registers and the potential performance gain also depends the number of nonzeros per column, i.e. the vector length:

uint grpid = get\_group\_id(0);

uint locid = get\_local\_id(0);

uint loclen = get\_local\_size(0);

uint numblocks = ELLwidth / loclen;

uint blockstride = loclen \* matDim;

uint slaboffset = grpid + (locid \* matDim);

if (grpid < matDim) {

auxShared[locid] = 0;

for (int i = 0; i < numblocks; i++) {

// vertical slab + row + x\_pos

int index = (i \* blockstride) + slaboffset;

int col = colIdx[index];

float aval = matData[index];

float xval = vector\_x[col];

float val = aval \* xval;

auxShared[locid] += val;

}

barrier(CLK\_LOCAL\_MEM\_FENCE);

if (get\_local\_id(0) == 0) {

float sum = 0;

for (int i = 0; i < loclen; i++) {

sum += auxShared[i];

}

vector\_y[grpid] = sum;

}

}

* 1. Utilizing Tiling/Blocking

While sharing inside a row did not work, another alternative which commonly yields good results is to divide the problem into a tiled format. By dividing the matrix into a slab of rows and looping over these groups of rows by a fixed width, it is possible to compute the matrix vector product by visualizing these blocks as tiles. This approach increases the number of operations performed by a workgroup increasing the percentage of time the ALU is busy, an important indicator of performance, besides permitting parallel reductions:

uint gloidy = get\_global\_id(1);

uint locidx = get\_local\_id(0);

uint locidy = get\_local\_id(1);

uint loclenx = get\_local\_size(0);

// Zero out local memory

auxShared[loclenx \* locidy + locidx] = 0;

uint slabsize = matDim \* loclenx;

uint NumBlocksX = ELLwidth / loclenx;

float sum = 0;

// Loops the vertical "slab" over ELLwidth

for (int i = 0; i < NumBlocksX; ++i) {

int index = i \* slabsize + gloidy + (matDim \* locidx);

int col = colIdx[index];

float aval = matData[index];

float xval = vector\_x[col];

sum += aval \* xval;

}

auxShared[locidy \* loclenx + locidx] = sum;

// Only one thread per row reduces

if (locidx == 0) {

barrier(CLK\_LOCAL\_MEM\_FENCE);

sum = 0;

uint localbase = locidy \* loclenx;

for (int i = 0; i < loclenx; ++i) {

sum += auxShared[localbase + i];

}

barrier(CLK\_LOCAL\_MEM\_FENCE);

vector\_y[gloidy] = sum;

}

* 1. Summing it all up

Applying the same techniques of using vector data types it is also possible to increase the efficiency of our tile based approach. Using the built in vector data types it is possible to speed up both the vector-row multiplications as the parallel reductions performed on local memory:

uint gloidy = get\_global\_id(1);

uint locidx = get\_local\_id(0);

uint locidy = get\_local\_id(1);

uint loclenx = get\_local\_size(0);

// Zero out local memory

auxShared[loclenx \* locidy + locidx] = 0;

uint slabsize = matDim \* loclenx;

uint NumBlocksX = ELLwidth / loclenx;

// Loops the vertical "slab" over ELLwidth

float4 sum4 = 0;

for (int i = 0; i < NumBlocksX; i += 4) {

int4 index, col;

float4 aval, xval;

index.x = i \* slabsize + gloidy + (matDim \* locidx);

index.y = (i + 1) \* slabsize + gloidy + (matDim \* locidx);

index.z = (i + 2) \* slabsize + gloidy + (matDim \* locidx);

index.w = (i + 3) \* slabsize + gloidy + (matDim \* locidx);

col.x = colIdx[index.x];

col.y = colIdx[index.y];

col.z = colIdx[index.z];

col.w = colIdx[index.w];

aval.x = matData[index.x];

aval.y = matData[index.y];

aval.z = matData[index.z];

aval.w = matData[index.w];

xval.x = vector\_x[col.x];

xval.y = vector\_x[col.y];

xval.z = vector\_x[col.z];

xval.w = vector\_x[col.w];

sum4 += aval \* xval;

}

auxShared[locidy \* loclenx + locidx] = sum4.x + sum4.y + sum4.z + sum4.w;

// Only one thread per row reduces

if (locidx == 0) {

barrier(CLK\_LOCAL\_MEM\_FENCE);

sum4.x = 0; sum4.y = 0; sum4.z = 0; sum4.w = 0;

uint localbase = locidy \* loclenx;

for (int i = 0; i < loclenx; i += 4) {

sum4.x += auxShared[localbase + i];

sum4.y += auxShared[localbase + i + 1];

sum4.z += auxShared[localbase + i + 2];

sum4.w += auxShared[localbase + i + 3];

}

barrier(CLK\_LOCAL\_MEM\_FENCE);

vector\_y[gloidy] = sum4.x + sum4.y + sum4.z + sum4.w;

}

1. Results

The following assembly speedup graphs displayed consider as a baseline the CPU implementation utilizing one thread without coloring so that the extra overhead from coloring is not included in the baseline of the results. This gives a rough comparison of the speedup that could be achieved compared to a naïve single threaded approach. Results are presented for each format type, for 2 and 3 gauss points utilized for numerical integration comparing versions with and without graph coloring on the same page. The raw data for all computations can be found with the source code on http://github.com/fpaboim. For parallelizing the global stiffness calculation using the GPU, the GPU handled calculating local element stiffness matrices while the task of performing parallel assembly utilizing graph coloring was performed by OpenMP with or without graph coloring as per section 5.3. The test systems utilized for conducting the benchmark consist of System1: Intel Core i7 CPU Q720@1.6Ghz, quad core with 8 hardware threads; 6Gb RAM; AMD Radeon 5870 Mobility GPU and System2: AMD Phenom II X4 945@3.01Ghz 4Gb RAM; AMD Radeon 5870 HD GPU.

**System1:**

**System2:**

1. Conclusion

Comparing the gains in performance with the time dedicated to optimizing code, OpenMP offers much easier gains with far fewer modifications and a greater ease maintaining. Conversely, in cases where performance is truly paramount, with OpenCL it is possible to extract the most out of the hardware utilizing vectorization primitives at a lower level of abstraction. Although with OpenCL it is possible to write portable code that will run on a variety of different platforms, some tuning is still required given that many options on what strategy to use for parallelizing depend on the underlying hardware. Still, the extra work can be mitigated by introducing some form of auto tuning or heuristics, and is much less work then rewriting for a different framework.

Much of the discussion of the validity of using GPUs for finite element analysis stem from having a problem which is not a perfect fit for the original hardware as other problems, and hence with less potential gains. The current trend towards GPUs with a greater focus on general purpose computing makes newer hardware better suited for these problems, making the switch even more appealing in the long run which is exemplified by better support for double precision arithmetic which achieves only one half of single precision peak performance in newer GPUs and also the lessening of performance penalties for branching in kernels. Moreover, there are still many experiments as to which sparse formats or strategies adopted to perform linear algebra operations on graphics processors and accelerators can produce the highest gains, and thus the potential gains from developing new procedures are higher for these novel architectures.

Since the assembly of the global stiffness matrix, also in this stage called the scattering of element stiffness matrices, is a memory intensive operation with few arithmetic operations, adding more threads quickly makes the problem memory bound and diminishes gains without a parallelization strategy such as graph coloring. Thus coloring, as displayed in the results for every element type tested, can greatly contribute to accelerate this stage of analysis. If more memory operations were carried out while the GPU was still calculating part of the stiffness matrix, further gains would be potentially possible at this stage. This would also be a solution that would permit scaling to other GPUs and calculating matrices larger then GPU memory, which would permit scaling clusters as described on section 5.4. Coloring is also shown to have a negligible computational cost for even the smaller problem sizes and this cost decreases with problem size. Given the formats benchmarked, the Eigen sparse matrix benefited the most from the adopted assembly approach, which is to be expected as one of the format’s main points is exactly the random insertion performance. Still, the speedup results must be taken with a grain of salt as the large gains are due, in part, to the poor baseline performance and not only to the large gains obtained with multithreading and vectorization. It is also apt to note that the some of the performance benefits from fast random insertion probably comes at the cost of larger sparse matrix sizes, which can be observed in the raw data present in Appendix A.

On the other hand the Ellpack format is the one who would potentially benefit the most from solving utilizing the conjugate gradient method which depends mainly on matrix-vector multiplication. Although the format’s main concern is matrix-vector multiplication performance, considerable speedups were obtained utilizing the parallel assembly method described on section 5.3 as were with the Compressed Sparse Row format. Given the importance of assembly time increases as less time is spent on the solving portion of the finite element analysis, optimizing not only for matrix-vector multiplication but also for assembly becomes increasingly important and thus such assembly strategies are expected to gain emphasis.

One interesting conclusion which can be drawn from these results is that 3D elements, with more degrees of freedom and higher connectivity are able to attain a greater speedup. This happens because the number of degrees of freedom of the elements used in the mesh directly influences the matrix’s band size. Clearly, elements with more degrees of freedom have more connections to other elements originating the larger band sizes, which in turn provide the GPU with more data on which to compute. The greater gains in elements with more degrees of freedom make the decision to use 3D, or even more complex elements within 3D, more attractive.

Using more gauss points is another factor which correlates directly to a greater speedup in GPUs for the same reasons. Both of these factors increase arithmetic intensity when calculating element stiffness matrices on the GPU by increasing the computation required on this device, which increases the ratio of ALU operations to memory operations - an especially important indicator for memory bound problems such as ours. This increases the efficiency of the processors utilization since the processor will have more to compute while, otherwise, it would be waiting for memory operations to be concluded. Consequently, the marginal cost of using more complex elements, i.e. elements with more degrees of freedom or more Gaussian interpolation points, on GPUs is lower when compared to CPUs. Although there are variables which factor into element selection for modeling, this decrease in computational cost is expected to give rise to a larger utilization of elements with more degrees of freedom when utilizing GPUs or Accelerators. Also the greater increase in numerical accuracy obtained by utilizing a larger number of gauss points which also has a smaller cost when compared to CPUs is expected to obtain greater importance, albeit not without further considerations such as phenomena such as overall “stiffening” of the models integrated with 3 gauss points relative to 2 gauss integration points.

1. References

AKHTER, SHAMEEM; ROBERS, JASON. **Multi-Core Programming**, Intel Press, 2006

Advanced Micro Devices, Inc. **AMD Accelerated Parallel Processing Guide**, rev. 2.3, Sunnyvale, CA. 2012

AMDAHL, G. M. **Validity of the single-processor approach to achieving large scale computing capabilities.** AFIPS Conference Proceedings, vol. 30, Atlantic City, NJ, April, p. 483-485, 1967

BATHE, K. J. **Finite element procedures in engineering analysis.** Prentice-Hall, Inc. ISBN 0-13-301458-4, 1996

BELL, N. and GARLAND, M. **Efﬁcient sparse matrix-vector**

**multiplication on CUDA**. NVIDIA Technical Report NVR-2008-004, NVIDIA Corporation, Dec. 2008

CECKA, C.; LEW, A. J.; DARVE, E. **Assembly of finite element methods on graphics processors**, International Journal for Numerical Methods in Engineering, vol. 85, Issue 5, p. 640–669, 4 February 2011

COOK, R. D.; MALKUS, D. S.; PLESHA, M. E.; WITT, R. J. **Concepts and Applications of Finite Element Analysis.** 4th Edition, Wiley. ISBN **0471356050**, 1996

DENNING, P. J. **The locality principle**, Communications of the ACM, v.48 n.7, July 2005

FELIPPA, C. **Introduction to Finite Element Methods**, online course material (ASEN 5007) , available at: <http://www.colorado.edu/engineering/CAS/courses.d/IFEM.d/Home.html>,1998

FLYNN, M. **Some Computer Organizations and Their Effectiveness**, IEEE Trans. Comput. C-21: p. 948, 1972

GASTER, B. et al. **Heterogeneous Computing with OpenCL,** 1st. ed. Morgan Kaufmann. ISBN 978-0-12-387766-6, 2012

GEVELER, M. et al. **Efficient Finite Element Geometric Multigrid Solvers for Unstructured Grids on GPUs,** Proceedings of the The Second International Conference on Parallel, Distributed, Grid and Cloud Computing for Engineering (PARENG 2011), Ajaccio, Corsica, France, 2011

GOVINDARAJU, N. K; MANOCHA, D. **Cache-efficient numerical algorithms using graphics hardware,** Parallel Computing, v.33 n.10-11, p.663-684, November, 2007

HELFENSTEIN, R. and KOKO, J. **Parallel preconditioned conjugate gradient algorithm on GPU**, J. Comput. Appl. Math. 236, No. 15, p. 3584-3590, ISSN 0377-0427, 2012

KARYPIS, G.; KUMAR, V. **Multilevel k-way Partitioning Scheme for Irregular Graphs CUDA,** J. Parallel Distrib. Comput. 48(1): p.96-129, 1998

KIM, N. S. **Leakage current: Moore's law meets static power**, IEEE Comput. Vol. 36, 2003

KOMATITSCH, D.; MICHÉA, D; ERLEBACHER, G. **Porting a high-order finite-element earthquake modeling application to NVIDIA graphics cards using CUDA,** Journal of Parallel and Distributed Computing, Volume 69, Issue 5, p.451-460, May, 2009

KURZAK, J; BUTTARI, ALFREDO; DONGARRA, JACK. **Solving Systems of Linear Equations on the CELL Processor Using Cholesky Factorization,** IEEE Trans. Parallel Distrib. Syst., 19(9), p. 1175-1186, 2008

MARKALL, G. **Making Fast FEM Solvers, Faster,** Imperial College, London, 2010

MCKEE, SALLY A. **Reflections on the memory wall,** Proceedings of the 1st conference on Computing frontiers, p. 162 , 2004.

MOORE, G. **Cramming more components onto integrated circuits,** Electronics, Volume 38, Number 8, April 19, p. 114-117, 1965

MUNSHI, A. et al. **OpenCL Programming Guide,** 1st. ed. São Paulo: Addison-Wesley. ISBN-10 0-321-74964-2, 2012.

OpenMP Architecture Review Board. **OpenMP C and C++ Application Program Interface,** version 2.0, March 2002

KNUTH, D. E. **The art of computer programming,** volume 1 (3rd ed.): Addison-Wesley Longman Publishing Co., Inc., Boston, MA, 1997

PSOTA , J. et al. **ATAC: A Manycore Processor with On-Chip Optical Network,** CSAIL Technical Reports, MIT-CSAIL-TR-2009-018, 2009

SHEWCHUCK, J. R. **An Introduction to the Conjugate Gradient Method Without the Agonizing Pain,** Technical Report, Carnegie Mellon Univ., Pittsburgh, PA, USA, 1994

STATHIS, P. T. **Sparse Matrix Vector Processing Formats,** PhD Dissertation, University of Delft, 2004

SUTTER, H. **The Free Lunch Is Over,** Dr. Dobb’s Journal, 30(3), p. 114-117, 1965.

VÁSQUEZ, F; FERNÁNDEZ, J. J.; GARZÓN, E. M. **A new approach for sparse matrix vector product on NVIDIA GPUs,** Concurrency and Computation: Practice and Experience, Vol. 23, Issue 8, p. 815-826, June 2011

WAFAI, M. **Sparse Matrix-Vector Multiplications on Graphics Processors,** Master Thesis, University of Stuttgart, 2009