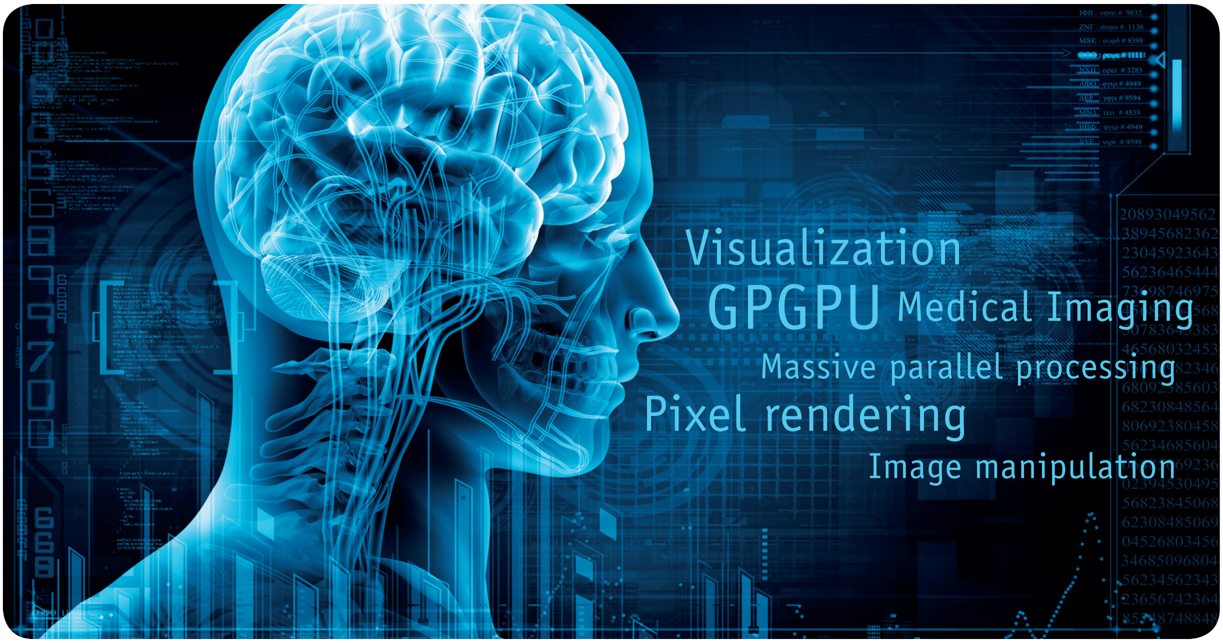
kontron

» Whitepaper «



Frameworks optimize analysis and visualization of medical imaging

Developing backend systems for medical imaging

New frameworks optimize analysis and visualization of medical imaging

Developing backend systems for medical imaging

Each advancement in medical imaging enables significant improvements in the quality and efficiency of diagnostics and therapy. The challenge: Each advancement requires significant computing performance gains to allow medical image data to be measured and analyzed. And once this massive amount of digital data has been created, it needs to be visualized too – including even stereoscopic 3D in real-time. With CUDA from NVIDIA, OpenCL from Khronos Group and Xeon Phi™ from Intel, several alternative massive parallel processing technologies are available for these tasks – all with their own benefits and restrictions. So which technology is right for data processing on medical backend systems? And how can this technology be combined most efficiently with different processor boards, operating systems and graphic algorithms? This whitepaper helps engineers understand the major differences in technology and identify the most appropriate framework for building their professional, medical-grade image processing systems.

CONTENT:

Frameworks optimize analysis and visualization of medical imaging 1

Abstract 2

Frameworks optimize analysis and visualization of medical imaging 2

Seeing means understanding 3

Processing massive image data 3

Benefits of parallel processing 3

Parallel processing technologies 3

CUDA – rich ecosystem and vendor-specific 3

OpenCL – the Open Source programming standard for GPU and CPU 4

From GPU to GPGPU 4

Intel® Xeon Phi™ 4

Parallel processing benchmarks 5

Embedded parallel processing configurations 6

Purchasing individual systems as finished components 6

Original Design & Manufacturing Services 6

Interface-free development and manufacturing expertise 7

High-tech center for ODMs 7

About Kontron 8

Seeing means understanding

Each advancement in medical imaging enables significant improvements in the quality and efficiency of diagnostics and therapy. This is because seeing is key to understanding. Sharper and higher image resolutions of organs, muscles, veins and bones enable more precise insights into pathological structures. The more precise the contours, colors and structures are visualized, the better physicians can differentiate between benign and malignant mutations. Improved visualization also helps to reduce patients' stress by enabling more precise and more conservative surgery. Augmented surgery entailing, for example, stereoscopic 3D in real-time, can give surgeons a realistic look into the blood vessels of a liver. A further benefit is that improved imaging will also help computer-aided diagnostics in identifying disease symptoms in an increasingly automated way – similar to face and fingerprint recognition technologies. A multitude of other very promising opportunities lies in innovations for improving medical imaging technologies. Many of them can lead to a reduction in healthcare costs. All of them, however, are ultimately designed to help people live more healthy lives.

Processing massive image data

As more precise medical imagery continues to become available, automatically an increase occurs in the mass of raw data produced as well as more demanding analysis. As simple mathematics demonstrate: If you have a picture increasing to the double in both dimensions, the number of pixels increase by a factor of four. If this quadrupled image data is shifted from 2D to 3D, each image layer adds the same amount of data. In the case of a video stream, the data rate will be multiplied by factor 50 to 60. But is this really as much as it sounds? Consider the saga of the grains of rice on a chess board, which demonstrates exponential growth: When starting with one grain of rice on the first square on a chess board and doubling the number on each following square, the total number exceeds 18 billion billion grains - or 18.446 * 1018 = 264 -1. Each of these multiplications means an exponentially larger amount of data. In the case of HD or even 4K resolutions, up to billions or even quadrillions of pixels have to be processed per minute. Each single pixel needs to be computed, analyzed and visualized in medical backend systems (systems that collect and process image raw data from frontend systems' sensors or pre-processors). So providing massive data processing performance is absolutely key to enabling further innovations in the medical imaging industry.

Figure 1: As an award for defeating the king in a game of chess, a sage asked for a grain of rice on the first square of the board and thereafter for the amount to double for every square. The king accepted, expecting this to be a small request. He soon found out that the amount of grains equals 210 billion tons.

Benefits of parallel processing

Parallel processing is one of the major key technologies which enable the rapid processing of massive amounts of data. In parallel processing, the core speed of single processors does not need to be increased to process more data within a certain time frame. When using parallel processing, existing processing frequencies can be simply multiplied by each added core. Theoretically, this setup is generally capable of supporting even the highest demands by adding the right number of cores. One example is an early field deployment of a 3D ultrasound CT breast imaging system a decade ago, that was based on Kontron single board computers (SBCs). With parallel processing on seven SBCs it was possible to achieve a substantial performance increase: The analysis time to process 60 GB of raw data was reduced from several hours to less than 60 minutes. This was revolutionary in those days. But nearly 10 years later, the amount of data and the number of cores have significantly increased. Today, engineers can utilize parallel processing technologies with more than 3000 processing units in parallel – a solid basis to improve medical imaging systems which have recently been deployed in the field.

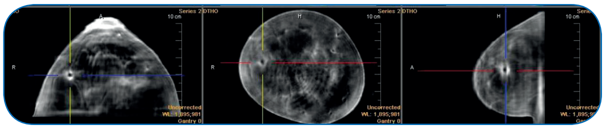


Figure 2: A 10 year old digital ultrasound breast scan detects a water-filled cyst from a solid mass. Higher resolutions in real time are very beneficial for more precise surgery.

Parallel processing technologies

Currently, three major parallel processing technologies are available: CUDA from NVIDIA, OpenCL by the Khronos Group and the Intel® Xeon Phi™ platform. How do they differ? And what are their benefits and restrictions? Let's start with CUDA technology from NVIDIA, which was first to market:

CUDA – rich ecosystem and vendor-specific

CUDA was invented by NVIDIA to use graphics cards for General Purpose Computing on Graphics Processing Units (GPGPU). NVIDIA introduced CUDA at the end of 2006. CUDA stands for Compute Unified Device Architecture. This architecture uses either NVIDIA Tesla co-processors or NVIDIA GPUs to run symmetric multi-processing applications for massively parallel workloads. Currently, up to eight CUDA processors can be used in one system. CUDA requires a central host processor, usually an x86 processor, to delegate the tasks to the CUDA capable devices.

Due to the fact that CUDA has the longest presence in the market, it offers a broad spectrum of toolkits, accelerated libraries, compiler directives and extensions to industry-standard programming languages. The CUDA Toolkit from NVIDIA provides, for example, a comprehensive environment for C/C++ developers. The support for further industry-standard programming languages includes extensions for Fortran, Python as well as Microsoft's new functional language F# for the .Net framework.

The fact that CUDA is supported by a host of GPU-accelerated libraries is an advantage. The NVIDIA CUDA page currently lists 20 libraries for various tasks, ranging from video encoding across Fast Fourier transformation to Open Source libraries like OpenCV for computer vision. Further third party tools include solutions for debugging, cluster management and performance analysis.

Currently, three different product ranges are available for CUDA-based parallel processing:

» The Tesla coprocessors - dedicated to pure parallel processing, i.e. they lack video output and cannot drive a display.
» The Quadro workstation GPUs - can be used to drive either several monitors in parallel or run CUDA code.
» The GeForce GPUs - dedicated to the consumer and gaming PC sector. Although they also support GPGPU applications they are not recommended for the medical sector.

From GPU to GPGPU

Traditionally, graphics cards had the task of visualizing computer data on a monitor. Ever since the introduction of computer graphics, developers have aimed to mimic the real world as best as possible. To visualize images, graphics processor had to execute uniform tasks on every pixel. In order to increase the quality, resolution, and frame rate, more and more processing units were integrated into the GPU to handle as many pixels as possible in parallel. With increased 3D capabilities, these processing units transformed from dedicated hardware accelerators into programmable processing units. This was when the term and usage model of General Purpose Computing on Graphics Processing Units, or GPGPU for short, was initiated. As a GPGPU, the GPU can now also turn raw data generated by the sensors and frontend systems of medical imaging applications into usable visualizations.

OpenCL – the Open Source programming standard for GPU and CPU

OpenCL is an open standard for parallel programming of heterogeneous systems. It was introduced to the market in August 2009 by Apple. Today, OpenCL is hosted by the Khronos Group, a non-profit industry consortium. Leading vendors such as AMD, ARM, Intel® and NVIDIA are members, amongst others. This broad support provides a reliable basis to make OpenCL a long-lasting, open industry standard.

Just like CUDA, OpenCL requires a host processor. Unlike CUDA, however, OpenCL can also be run on the host. Another benefit of OpenCL is that the code can be executed not only on GPGPUs, but also on CPUs, APUs (that contain both a multi-core x86 CPU and a GPU) as well as FPGAs and other processor architectures. To simplify programming, OpenCL presents developers an abstract, hierarchical platform model, so they do not have to care about the underlying hardware. This makes the code highly portable from one platform to another. Additionally, it helps programmers preserve their source code, if the hardware is changed or updated.

Furthermore, OpenCL supports both data and task parallelism, which offer distinct benefits for CPUs and GPUs. Task parallelism, the simultaneous execution of many different functions on multiple cores across the same or different datasets, is a perfect match for multi-core CPUs. Data parallelism is also well suited for multi-core GPUs. It enables simultaneous execution of the same function on multiple cores across the elements of a dataset.

OpenCL is supported by many different hardware platforms, including:

» 3rd gen Intel® Core™ processors and following generations
» AMD FirePro graphics as well as AMD APUs and discrete processors
» NVIDIA Tesla coprocessors and Quadro graphics
» Intel® Xeon Phi™ coprocessors

This makes OpenCL a very suitable programming standard for the entire range of potential parallel processing tasks. A configuration dedicated for x86 code is the Intel® Xeon® Phi coprocessor.

Intel® Xeon Phi™

Intel® Xeon Phi™ coprocessors are based on the Intel® Many Integrated Core Architecture (Intel® MIC Architecture) and combine tens of Intel® CPU cores onto a single chip. Like CUDA and OpenCL, Intel® Xeon Phi™ also requires a host processor. The current generation of Intel® Xeon Phi™ Coprocessors features up to 61 x86 cores on a single die. With hardware-based hyper threading, up to 244 parallel threads are possible. The coprocessor runs a Linux operating system, supports x86 memory order model and IEEE 754 floating-point arithmetic. It is capable to execute applications written in industry-standard programming languages such as Fortran, C, and C++. Developers can utilize the broad ecosystem of programming languages, models and software which already exist for x86 processors, thanks to the compatibility of the Xeon Phi™ coprocessor with the Intel® Xeon® processor. Applications that run on one of the two processor families will run on the other as well. The benefit is that developers can reuse existing code and maintain a common code base using familiar tools and methods. To reach full throughput on the Coprocessor, existing code will need to be tuned and recompiled. But as with so many parallel executing cores, this is self explanatory. The Intel® Xeon Phi™ also supports OpenCL 1.2. So does the Intel® Xeon Phi™ perhaps also compete with GPGPUs? Ultimately, this depends on the applications, algorithms and codes developers prefer. From a basic technological point of view, it also depends on the raw performance these solutions provide. So the next step is to look at the available processing cores and their computing performance.

Parallel processing benchmarks

A bunch of professional, embedded parallel processing silicon from AMD, Intel® and NVIDIA can be leveraged by OEMs to accelerate medical imaging applications. A very basic but useful starting point for a performance comparison is the floating point operation values with single or double precision. They range from several hundred Gflops up to the factor 10 with the recent benchmark design of 5196 Gflops. This is an enormous bandwidth which can address different application areas from low up to high-end. All listed silicon is available on PCI Express expansion cards. Only the solutions from the silicon vendors are listed here. During evaluation and purchasing processes, it is important not to forget third party solutions.

Vendor	Model	Cores	Support for	Memory GDDR5	TDP [Watt]	Peak single precision [Tflops]	Peak double precision [Tflops]	Gflops SP/W	Gflops DP/W
NVIDIA	TESLA K20-K40	2496 - 2880	CUDA	5GB - 12GB	225 - 235	3.52 – 4.29	1.17 – 1.43	15.6 - 18.3	5.2 - 6.1
NVIDIA	TESLA K10	2x 1536	CUDA	2x 4GB	250	4.58	0.19	23.1	0.8
NVIDIA	Quadro K4000 - K6000	768 – 2880	CUDA and OpenCL[2]	3GB - 12GB	80 – 225	1.25 - 5.20	≤ 1.73	15.6 - 23.1	≤ 7.7
Intel	Xeon Phi™ 5120D -7120D	240 – 244[1]	OpenCL 1.2	8GB - 16GB	245 – 270	2.02 – 2.41	1.01 – 1.21	8.3-8.9	4.1 - 4.5
AMD	FirePro W5000-W9000	768 – 2048	OpenCL 1.2	2GB - 6GB	75 - 274	1.30 – 4.00	0.08 – 1.00	14.6 - 17.3	1.1 - 3.6
AMD	Embedded Radeon E8860	640	OpenCL 1.2	2GB	37	0.77	na	20.8	n.a

1) = WITH HYPERTHREADING

2) = OPENCL SUPPORTED, BUT VERSION NOT SPECIFIED

And for OEMs wanting to design compact mobile devices, i.e. handheld ultrasound tablets, solutions are also available. The solutions listed below are available on MXM boards, a form factor originating from the laptop and ultrabook segment.

Vendor	Model	Cores	Support for	Memory GDDR5	TDP [Watt]	Peak single precision [Tflops]	Peak double precision [Tflops]	Gflops SP/W	Gflops DP/W
NVIDIA	Quadro K3100M-K5100M	768 - 1536	CUDA	4GB - 8GB	75 - 100	1.05 – 2.35	Not specified	14.0 - 23.5	Not specified
AMD	Embedded Radeon E8860	640	OpenCL 1.2	2GB	37	0.77	Not specified	20.8	Not specified

Embedded parallel processing configurations

All these three parallel processing technologies are very attractive for improving medical image processing. Consequently, many medical engineers would opt to always work with the latest launched products of their preferred technology. But all the technologies also present several restrictions. One of the most limiting factors is the fact that the design, certification and life cycle of the systems are much longer than the availability of the latest GPUs or coprocessors. Every six months a new and more compelling GPU becomes available. The older versions become obsolete, because the race for more GPU performance is more intense than the one for CPU performance. This race will go on until one day virtual reality graphics will not be identifiable as such anymore, until then we will continue to see these significant performance gains.

Another crucial argument is that the quality of the silicon for medical applications needs to be higher compared to solutions for the consumer and desktop gaming market. In medical technology, the GPUs need to be more reliable and more energy-efficient and with lower tolerances in TDP. Consequently, either professional or embedded versions of these parallel processing technologies are recommended.

All the silicon vendors, as well as several embedded boards and modules vendors, provide corresponding GPUs or Xeon Phi™ designs which can be purchased off-the-shelf. The same applies to corresponding embedded CPU boards. Engineers can – more or less easily – configure and assemble a system, e.g. based on ATX-compliant form factors for 19″ rack mounts. Quite a few components can be made available without any custom design.

The availability of COTS products does not free OEM engineers from the task of having to find and configure an optimized system setup with compatible components, perfectly in tune with one another. Besides selecting the right components, other important factors to achieve an optimal configuration are the specific settings of the operating system with the corresponding configuration and drivers with appropriate settings. On the path to an optimally configured platform, numerous potential pitfalls lurk along the way. One missing tick in the submenu during configuration can quickly lead to a system not working as required. A lack of know-how can result in nerve-wracking trial & error processes, costing time and money. As the number of system components required for an individual configuration increases, the potential number of configuration errors increases. To add to this, the sheer number of potential suppliers makes system configuration even more complex. The effort and the possible consequential problems are usually out of proportion to any additional costs which might occur with the purchase of an application-ready, pre-configured system. This especially holds true if the system is to go into production and is being used on a global scale.

However, are all these efforts in configuring components really the core competence of engineers and will this work result in added value for the customer? Do medical engineers really need to work on computer hardware configurations and parameterizations while, parallel to this, they are involved in the challenging race of the latest algorithms and codecs for image processing to constantly improve medical imaging? Definitely not. But they demand easily deployable frameworks that cover all the required aspects to deliver application-ready, custom-specific medical imaging backend systems.
Such a medical imaging framework has to fulfill several requirements. Some of the most determining factors are:

1. Long-term availability (sometimes well over a decade)
2. Modularity
3. Specific requirements for the chosen imaging technologies must be addressed
4. Standard form factor basis for electronics
5. Scalability over several performance classes and processor generations
6. EN60601-compliance for medical environments
7. Application-ready availability, including drivers and OS support
8. Standardization of programming interfaces (EAPI) to simplify migration

Vendors can make such a framework available off-the-shelf. But for many medical device manufacturers, purchasing standard groups of components plus integration off-the-shelf does not suffice.

Purchasing individual systems as finished components

Medical device manufacturers would rather receive complete systems which they can deploy in their equipment as ready-made components. In order to provide this, a manufacturer of medical computers has to be able to develop and produce systems which are tailored to meet the individual demands of the application. The manufacturer has to master the whole process from the board development to the manufacturing of customized medical computers and ideally offer a wide range of Original Design & Manufacturing (ODM) services. This entails – in addition to the systems themselves and their assembly - the management of supply chain and lifecycle. With this synergy, the medical device manufacturer can fully concentrate on his actual core competences and purchase the medical computer as a just-in-time, ready-made component.

Original Design & Manufacturing Services

Kontron offers a range of ODM services and has, for many years, carried out these services as a reliable partner for leading global medical device manufacturers. Kontron's systems and components for medical devices are designed to meet individual customer requirements through flexibility in manufacturing and engineering. In particular, Kontron's products fulfill the customer's need for easy certifiability and long-term availability. They are also designed to meet specific environmental requirements, contributing to long-term stability and reliability. Continuous delivery quality, revision control (down to chip level) as well as optimized supply chain and EN60601-compliant development are all facets of Kontron's service which have solidified Kontron's good reputation in the market as a medical device manufacturer. Kontron has over 30 years of experience in supply chain management and has installed and established the processes and QM systems to meet the stringent demands of the medical market.

Interface-free development and manufacturing expertise

The interaction of design and manufacturing expertise makes system responsibility almost free of interfaces. Original Design & Manufacturing Services are therefore a guarantee for high product and service quality. The logistics system, used by Kontron for years, for manufacturing medical computers ensures the company's consistently high delivery reliability and excellent delivery quality. Recently, Kontron once again received a Supplier Performance Award, which distinguishes the company for its long-standing delivery performance and high product and service quality. With 99.98 percent the delivery performance was rated exceptionally high.

Sources:
http://www.mevis.fraunhofer.de/en/news/press-release/article/tablet-pc-unterstuetzt-leberchirurgen-neue-app-von-fraunhofer-mevis-erstmals-in-deutschem-op-getest.html

http://developer.download.nvidia.com/shaderlibrary/docs/GPU-Img-Proc-Intro.pdf

Link to OpenCV OCL: http://docs.opencv.org/modules/ocl/doc/introduction.html

High-tech center for ODMs

For Kontron though, there will be no resting on these laurels. As a German company with German engineering practice Kontron will expand the Augsburg, Germany, site into a high-tech center. And as a global company with a worldwide presence, Kontron supports customers locally with local engineering and manufacturing services following the same engineering standards. This will further increase the attractiveness of this internationally recognized high-tech company. Medical OEMs will profit from these investments and enjoy even closer and more sustainable interaction between the two pillars of ODM, i.e., 'design' and 'manufacturing'. Ultimately, this will further increase the quality of medical computing solutions for OEMs and secure sustainability.

About Kontron

Kontron is a global leader in embedded computing technology. With more than 40% of its employees in research and development, Kontron creates many of the standards that drive the world's embedded computing platforms. Kontron's product longevity, local engineering and support, and value-added services, helps create a sustainable and viable embedded solution for OEMs and system integrators.

Kontron works closely with its customers on their embedded application-ready platforms and custom solutions, enabling them to focus on their core competencies. The result is an accelerated time-to-market, reduced total-cost-of-ownership and an improved overall application with leading-edge, highly-reliable embedded technology.

Kontron is listed on the German TecDAX stock exchanges under the symbol "KBC". For more information, please visit: **www.kontron.com**

CORPORATE OFFICES

Europe, Middle East & Africa

Oskar-von-Miller-Str. 1
85386 Eching / Munich
Germany

Tel.: +49 (0) 8165 / 77 777
Fax: +49 (0) 8165 / 77 219
info@kontron.com

North America

14118 Stowe Drive
Poway, CA 92064-7147
USA

Tel.: +1 888 294 4558
Fax: +1 858 677 0898
info@us.kontron.com

Asia Pacific

17 Building,Block #1, ABP.
188 Southern West 4th Ring Road
Beijing 100070, P.R.China

Tel.: +86 10 63751188
Fax: +86 10 83682438
info@kontron.cn

Whitepaper – Medical Imaging #29072014WMH
All data is for information purposes only and not guaranteed for legal purposes. Subject to change without notice. Information in this whitepaper has been carefully checked and is believed to be accurate; however, no responsibilty is assumed for inaccurancies. All brand or product names are trademarks or registered trademarks of their respective owners.