“Новый подход к использованию гибридных технологий»

СТАНАВОВ ПАВЕЛ

НАЦИОНАЛЬНЫЙ СУПЕРКОМПЬЮТЕРНЫЙ ФОРУМ

PERESLAVL ZALESSKIY 2013

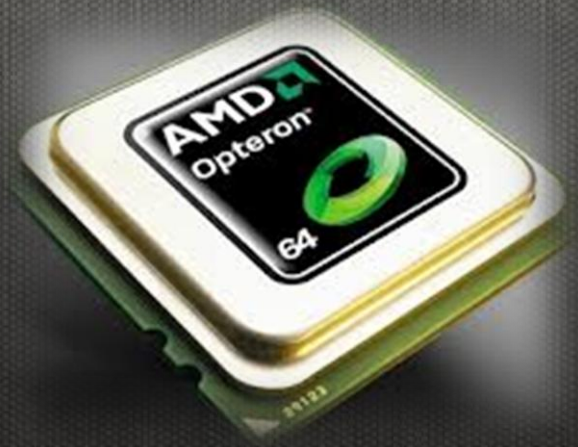
ОБ АРХИТЕКТУРЕ ГЕТЕРОГЕННЫХ СИСТЕМ (HSA)

10 ЛЕТ НАЗАД

AMD

AMD Opteron

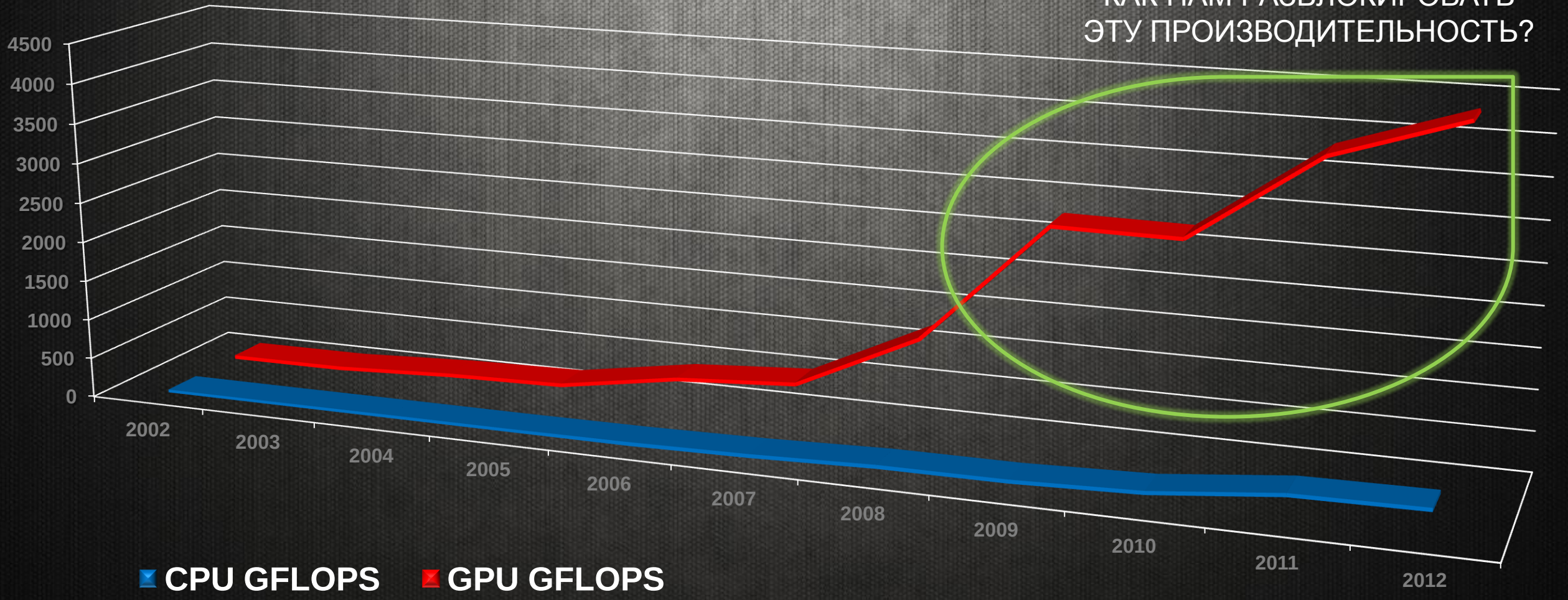
- ***64-битные расширения***
- **HyperTransport**
- ***Контроллер памяти на кристалле***



ВЫЧИСЛИТЕЛЬНАЯ МОЩНОСТЬ GPU ПРЕВЫШАЕТ ВОЗМОЖНОСТИ CPU БОЛЕЕ ЧЕМ В 10 раз

AMD

КАК НАМ РАЗБЛОКИРОВАТЬ ЭТУ ПРОИЗВОДИТЕЛЬНОСТЬ?



- См. слайд 24 для подробной информации

ЧТО ТАКОЕ «АРХИТЕКТУРА ГЕТЕРОГЕННЫХ СИСТЕМ» (HSA)?

Это ***интеллектуальная компьютерная архитектура***, позволяющая центральному, графическому и другим процессорам ***гармонично*** работать вместе на одной кремниевой микросхеме, ***без проблем*** передавая задачи элементам, которые лучше всего для них подходят.

ЭВОЛЮЦИЯ HSA

AMD

Преимущества

Унифицированная энергоэффективность

Увеличенная вычислительная мощность

Упрощенное совместное использование данных

Возможности

Размещение CPU и GPU на одном кристалле

GPU имеет доступ к памяти CPU

Унификация доступа к памяти для CPU и GPU

ЧТО ТАКОЕ hUMA?

гетерогенный
ОДНОРОДНЫЙ
ДОСТУП К
ПАМЯТИ

ЧТО ТАКОЕ UMA?

AMD

UMA — это сокращение от **Uniform Memory Access, или однородный доступ к памяти**

- *Определяет, каким образом процессорные ядра отображаются в системе и как они обращаются к памяти*
- *Все процессорные ядра с UMA в системе делят одно адресное пространство памяти*

Перенос вычислений на GPU с неоднородным доступом к памяти (NUMA)

- *Необходимо, чтобы данные управлялись через множество куч с разными адресными пространствами*
- *Усложняет программирование из-за создания частых копий, синхронизации и преобразования адресов*

HSA обеспечивает однородный доступ к памяти для GPU

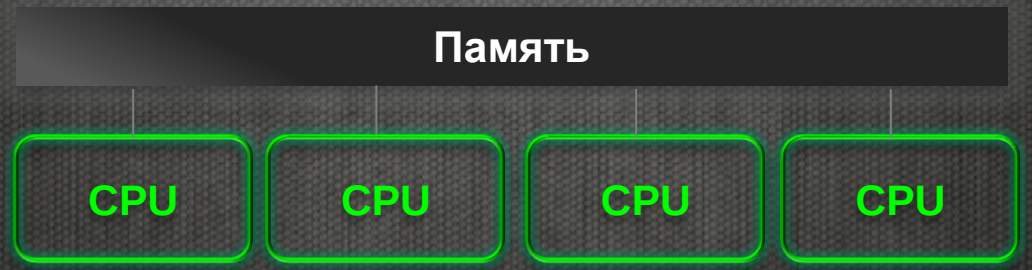
- *Гетерогенные вычисления заменяют вычисления на GPU*

ПРЕДСТАВЛЯЕМ hUMA

AMD

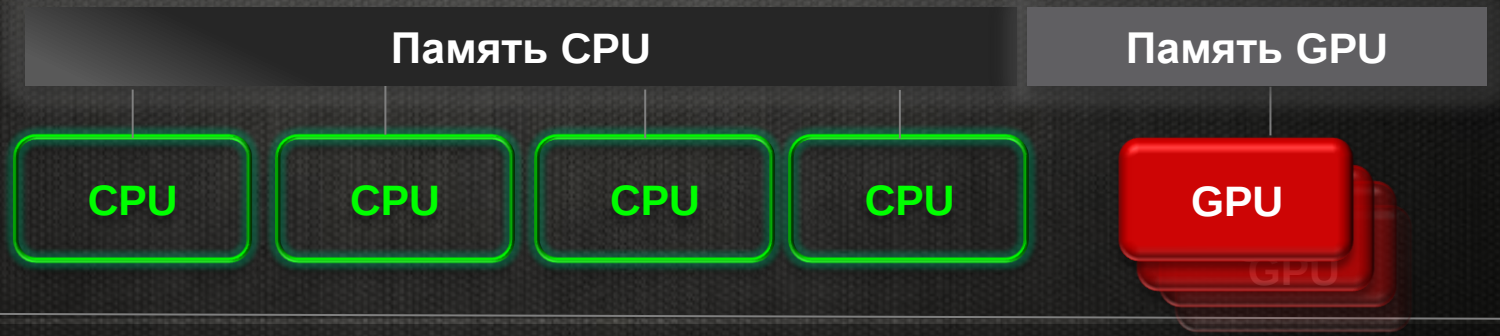
CPU

UMA



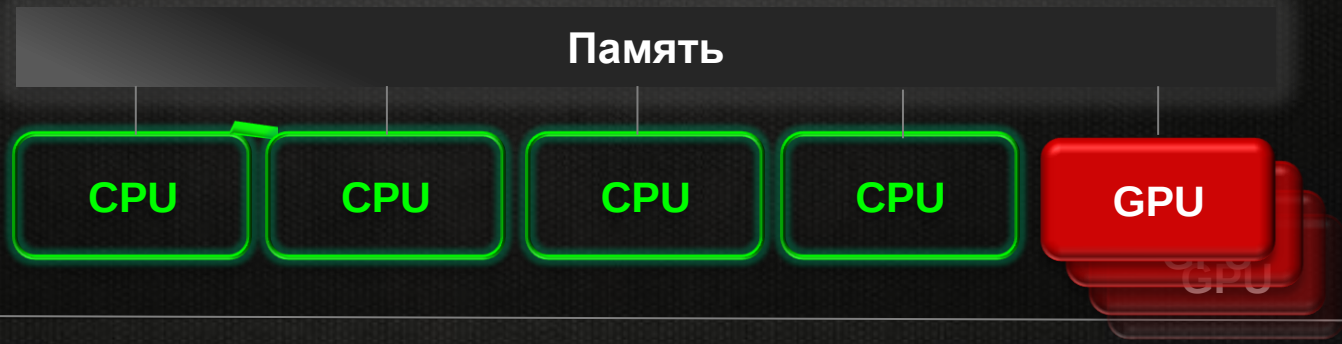
APU

NUMA



APU
с
HSA

hUMA



ГЛАВНЫЕ ОСОБЕННОСТИ hUMA

AMD

ДВУНАПРАВЛЕННАЯ КОГЕРЕНТНАЯ ПАМЯТЬ

Любые изменения, сделанные одним процессорным устройством, будут видимы другими процессорами — GPU или CPU

ПАМЯТЬ СО СТРАНИЧНОЙ ОРГАНИЗАЦИЕЙ

GPU может обрабатывать ошибки страницы памяти и напрямую работать с файлом подкачки

ОБЩЕЕ АДРЕСНОЕ ПРОСТРАНСТВО ПАМЯТИ

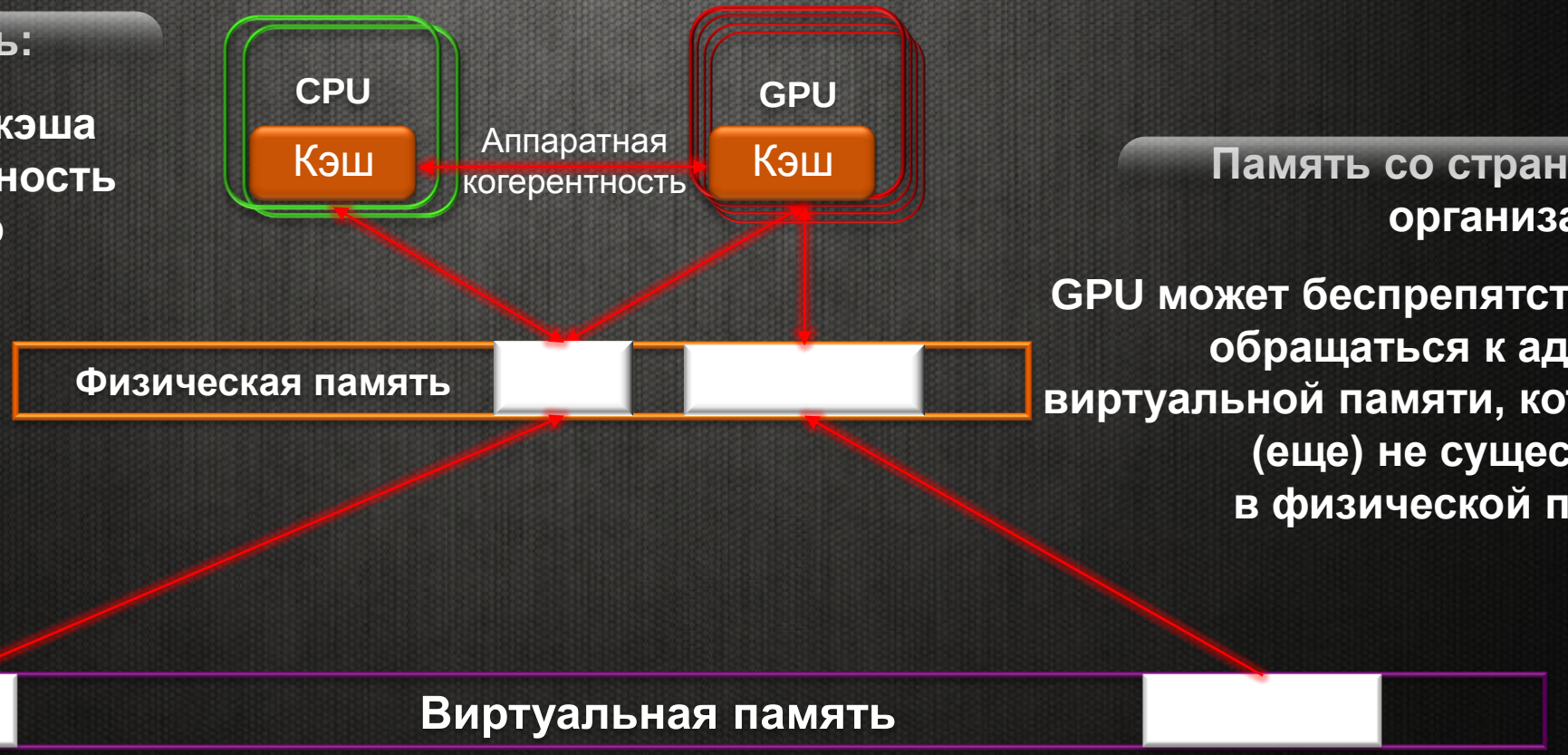
Процессы, выполняемые на CPU и GPU, могут динамически распределять использование памяти из ее общего адресного пространства

ГЛАВНЫЕ ОСОБЕННОСТИ hUMA

AMD

Когерентная память:

Обеспечивает для кэша CPU и GPU возможность видеть актуальную версию данных



Память со страничной организацией:

GPU может беспрепятственно обращаться к адресам виртуальной памяти, которые (еще) не существуют в физической памяти

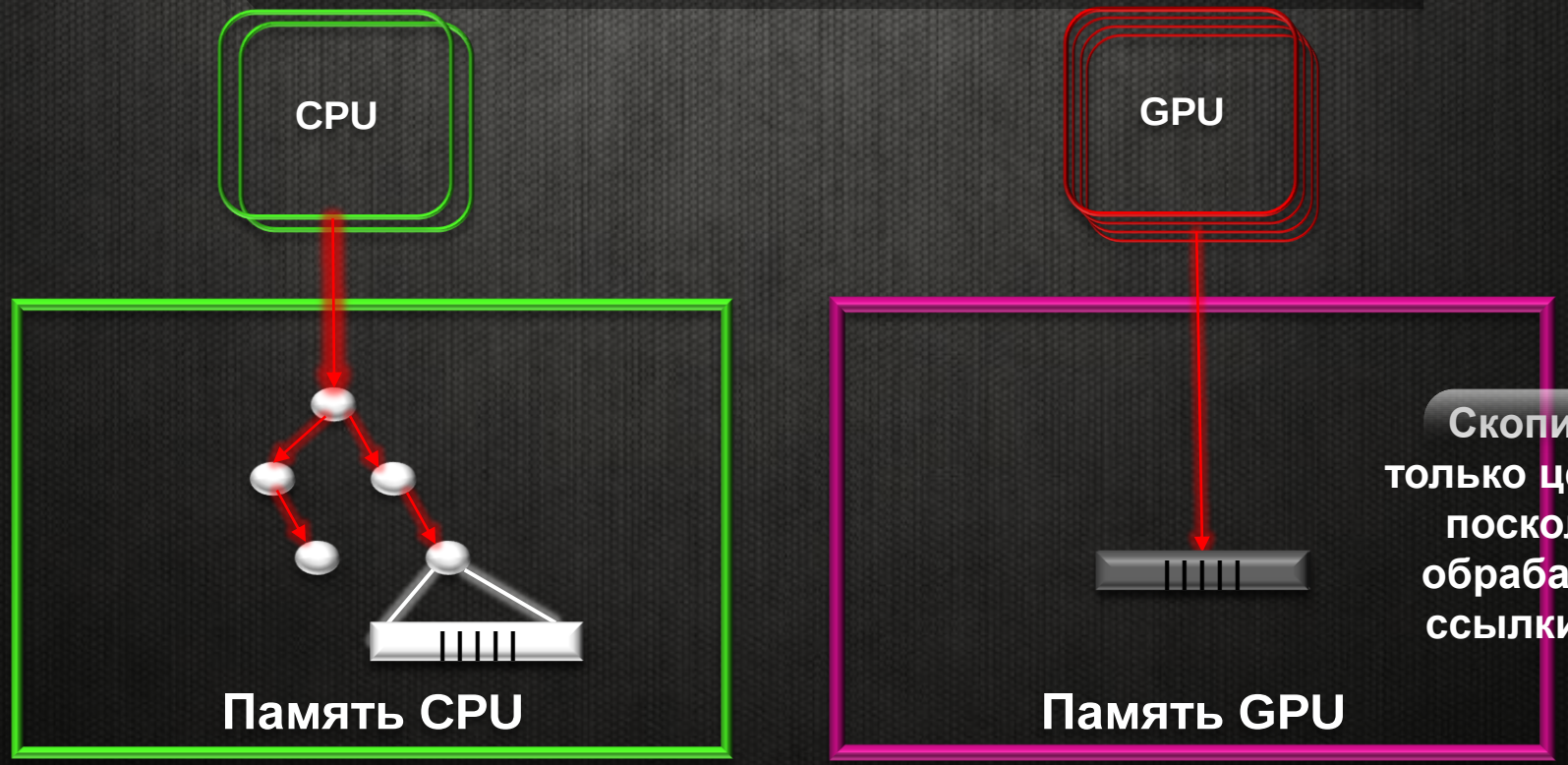
Общее адресное пространство:
CPU и GPU могут обращаться и занимать любую область в адресном пространстве виртуальной памяти

БЕЗ УКАЗАТЕЛЕЙ* И РАЗДЕЛЕНИЯ ДАННЫХ

AMD

Без hUMA:

- CPU явным образом копирует данные в память GPU
- GPU завершает вычисления
- GPU явным образом копирует результат назад в память CPU



Скопирован может быть только целый массив данных, поскольку GPU не может обрабатывать встроенные ссылки структуры данных

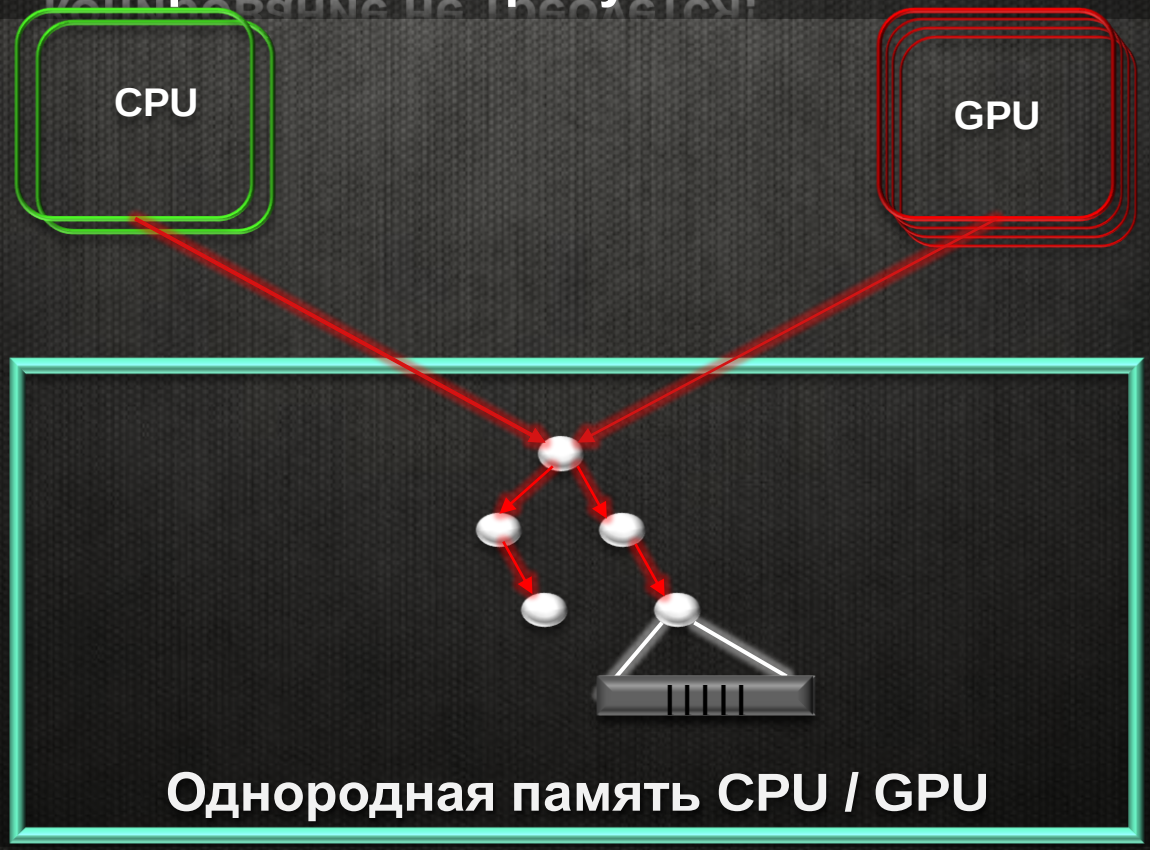
*Указатель — именная переменная, которая содержит адрес памяти. Это делает возможным легко ссылаться на данные или сегменты кода по названию и избавляет разработчика от необходимости знать фактический адрес в памяти. Указателями могут управлять те же самые выражения, используемые для работы с любой другой переменной.

С УКАЗАТЕЛЯМИ* И ОБМЕНОМ ДАННЫМИ

AMD

С hUMA:

- CPU просто передает указатель памяти на GPU
- GPU завершает вычисления
- CPU может напрямую прочитать результат – **копирование не требуется!**



CPU может передавать указатель всей структуре данных, так как GPU теперь может обрабатывать встроенные ссылки

*Указатель — именная переменная, которая содержит адрес памяти. Это делает возможным легко ссылаться на данные или сегменты кода по названию и избавляет разработчика от необходимости знать фактический адрес в памяти. Указателями могут управлять те же самые выражения, используемые для работы с любой другой переменной.

10 ГЛАВНЫХ ПРИЧИН ДЛЯ ВНЕДРЕНИЯ КОГЕРЕНТНОЙ ПАМЯТИ В GPU/APU

1. Программисты могут легко писать код одновременно для CPU и GPU
2. Не нужно использовать специальные интерфейсы программирования приложений (API)
3. Алгоритмы, выполняемые на многоядерных CPU, можно переносить на GPU без повторного написания кода
4. Возможность совместного использования фрагментарных данных, недоступного при обращении к программе через процессор
5. Легче внедрить один раз в аппаратное обеспечение, чем делать это множество раз в различное ПО
6. Легче писать код без ошибок
7. При наличии когерентной памяти уменьшается вероятность появления ошибок в ОС
8. Использование probe-фильтров, предназначенных для снижения контрольного трафика между процессорами, уменьшает энергопотребление системы
9. Общее адресное пространство памяти открывает возможности для централизованного программирования неуправляемого и управляемого кода на гетерогенных платформах
10. Оптимальная архитектура для гетерогенных вычислений на гибридных процессорах и однокристальных системах

ФУНКЦИИ hUMA

- Доступ ко всему пространству памяти 
- Память со страничной организацией 
- Двунаправленная когерентность 
- Быстрый доступ GPU к системной памяти 
- Динамическое распределение памяти 

ПРЕИМУЩЕСТВА hUMA

ПРЕИМУЩЕСТВА HSA

ПРЕИМУЩЕСТВА ОДНОРОДНОЙ ПАМЯТИ ДЛЯ РАЗРАБОТЧИКОВ

AMD

ЛЕГКОСТЬ И ПРОСТОТА ПРОГРАММИРОВАНИЯ

Единая стандартизированная вычислительная среда

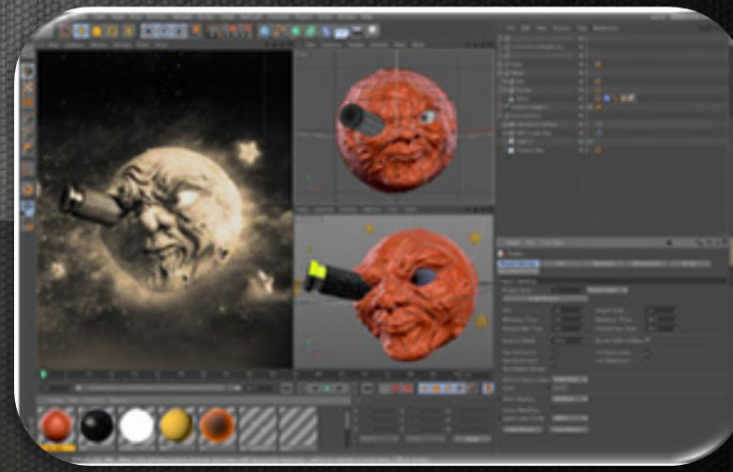
ПОДДЕРЖКА ПОПУЛЯРНЫХ ЯЗЫКОВ ПРОГРАММИРОВАНИЯ

Python, C++, Java

УМЕНЬШЕНИЕ СТОИМОСТИ РАЗРАБОТКИ

Более эффективная архитектура позволяет делать меньшему количеству людей прежний объем работы

ПРЕИМУЩЕСТВА ДЛЯ ПОЛЬЗОВАТЕЛЕЙ



НОВЫЕ ОЩУЩЕНИЯ

Радикально отличающийся пользовательский опыт

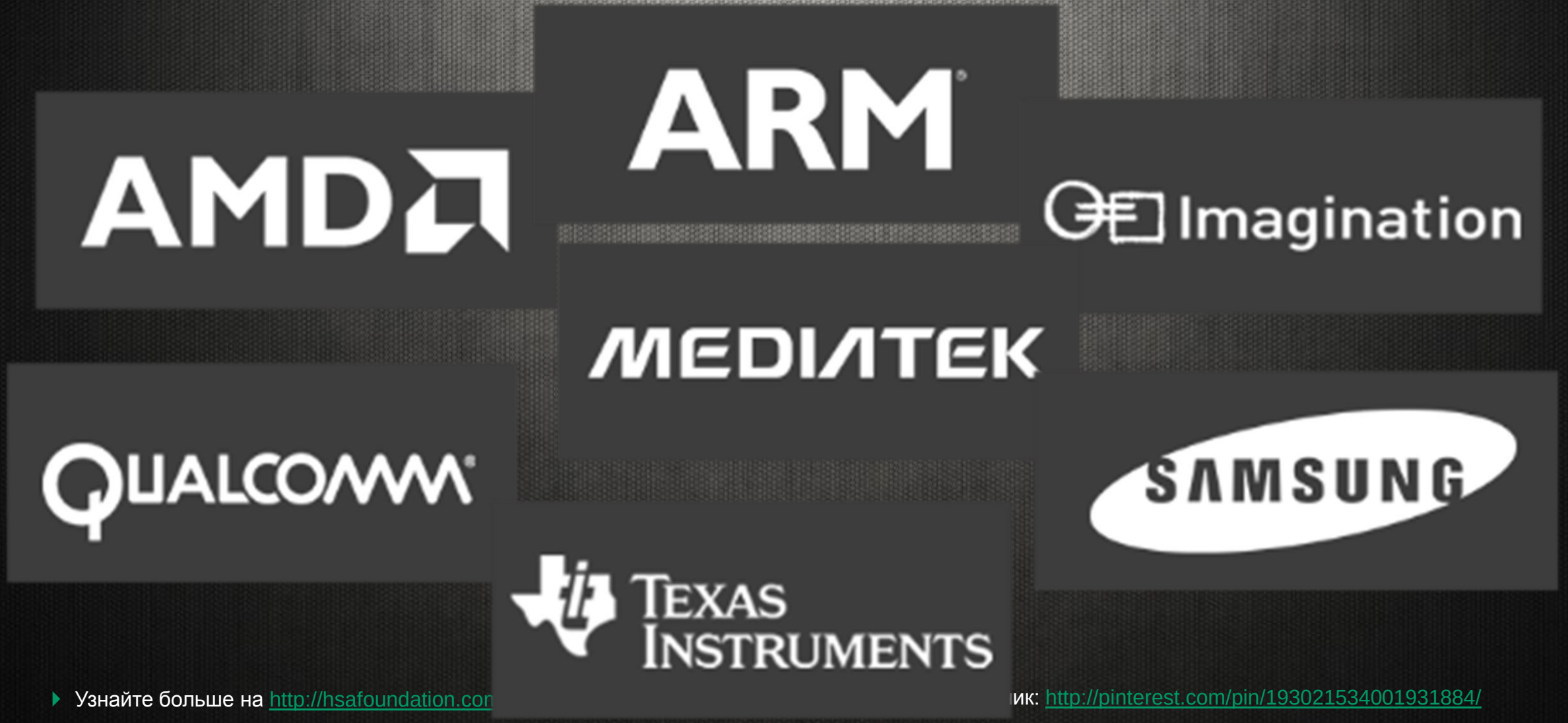
УЛУЧШЕННАЯ ПРОИЗВОДИТЕЛЬНОСТЬ

Получите больше производительности
в прежнем форм-факторе

УЛУЧШЕНИЕ ЭНЕРГОЭФФЕКТИВНОСТИ

Меньшее энергопотребление
при прежней производительности

ПОДДЕРЖКА ВЕДУЩИМИ ИГРОКАМИ РЫНКА



- Узнайте больше на http://hsafoundation.com
- ик: http://pinterest.com/pin/193021534001931884/

ПОТЕНЦИАЛЬНЫЙ РЫНОК ОГРОМЕН


AMD
СПАСИБО

GFLOPS

AMD

Год	CPU	CPU GFLOPS	GPU (RADEON)	GPU GFLOPS
2002	Pentium 4 (Northwood)	12.24	9700 Pro	31.2
2003	Pentium 4 (Northwood)	12.8	9800 XT	36.48
2004	Pentium 4 (Prescott	15.2	X850 XT	103.68
2005		15.2	X1800 XT	134.4
2006	Core 2 Duo	23.44	X1950	375
2007	Core 2 Quad	48	HD 2900 XT	473.6
2008	Q9650	96	HD 4870	1200
2009	Core i7 960	102.4	HD 5870	2720
2010	Core i7 970	153.6	HD 6970	2703
2011	Core i7 3960X	316.8	HD7970	3789
2012	Core i7 3970X	336	HD 7970 GHz Edition	4301

ПОТЕНЦИАЛЬНЫЙ РЫНОК ОГРОМЕН

OPENCL™ GAINS MOMENTUM

AMD

Adobe

Autodesk

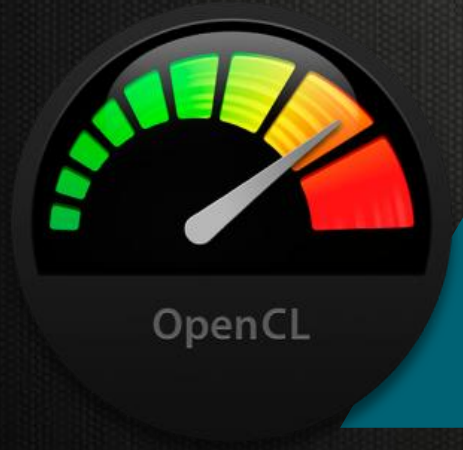
MAIN CONCEPT

DASSAULT SYSTEMES

DEMSolutions

OPENCASCADE

OPTIS



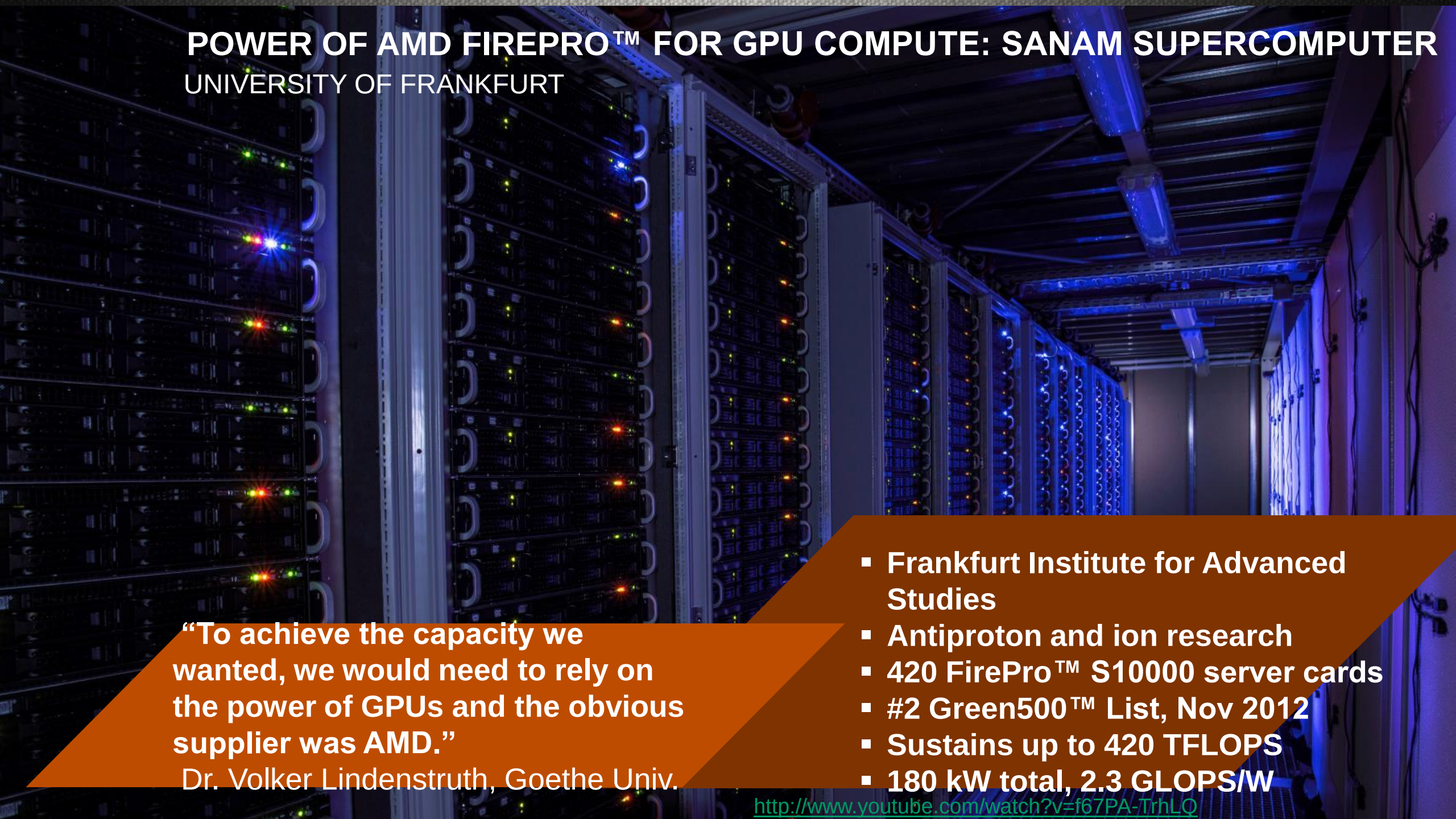
“OpenCL continues to gather momentum on both desktop and mobile devices. In addition to enabling application developers it is providing foundational, portable acceleration for middleware libraries, engines and higher-level programming languages that need to take advantage of heterogeneous compute resources including CPUs, GPUs, DSPs and FPGAs.”

Neil Trevett, chair of the OpenCL™ Working Group and President of the Khronos Group

https://www.khronos.org/news/press/khronos-releases-opencl-2.0

POWER OF AMD FIREPRO™ FOR GPU COMPUTE: SANAM SUPERCOMPUTER

UNIVERSITY OF FRANKFURT



“To achieve the capacity we wanted, we would need to rely on the power of GPUs and the obvious supplier was AMD.”

Dr. Volker Lindenstruth, Goethe Univ.

- **Frankfurt Institute for Advanced Studies**
- **Antiproton and ion research**
- **420 FirePro™ S10000 server cards**
- **#2 Green500™ List, Nov 2012**
- **Sustains up to 420 TFLOPS**
- **180 kW total, 2.3 GLOPS/W**

http://www.youtube.com/watch?v=f67PA-TrhLQ

POWER OF AMD FIREPRO™ FOR DISPLAY WALLS: STONY BROOK UNIV.

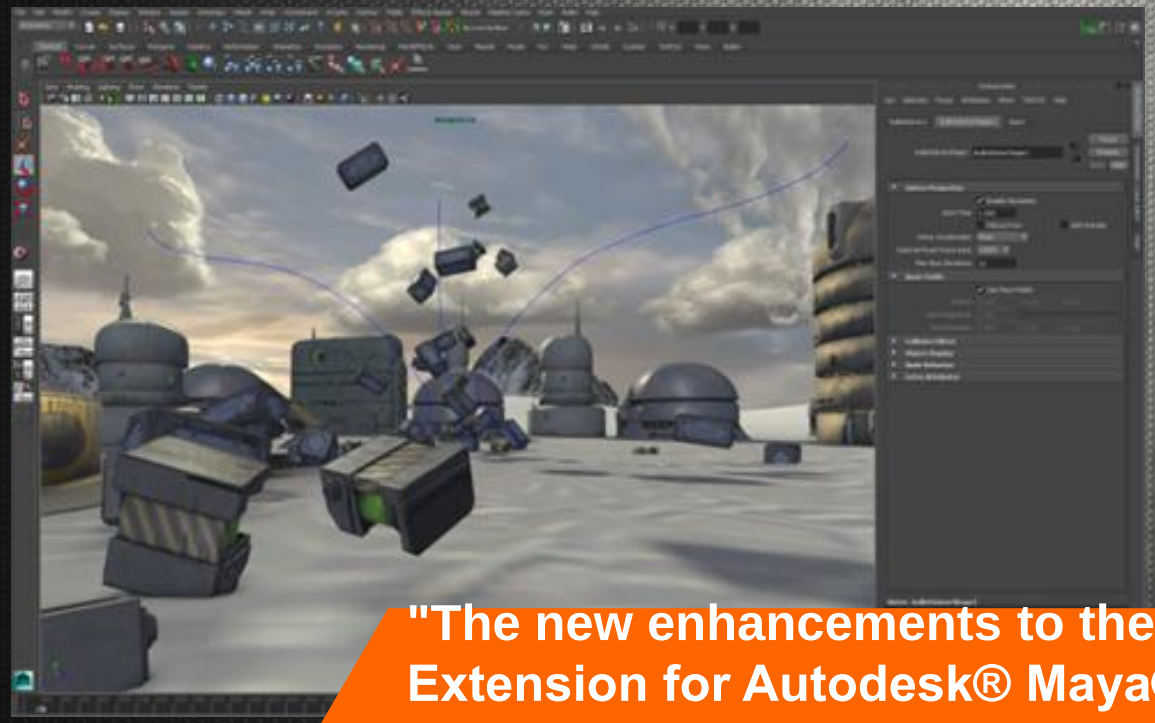
"AMD FirePro™ workstation graphics are central to this immersive experience."

– Dr. Arie Kaufman, Project Director

- The Reality Deck
- 1.5 billion pixels
- Visualize large amounts of data
- 72 AMD FirePro™ graphics cards
- 16 ATI FirePro™ S400 sync cards
- 416 HD displays (2560x1440)

www.youtube.com/watch?v=E9KNLnzaHO

AMD FIREPRO™ ACCELERATES AUTODESK® MAYA® 2014 WITH OPENCL™



AUTODESK MAYA

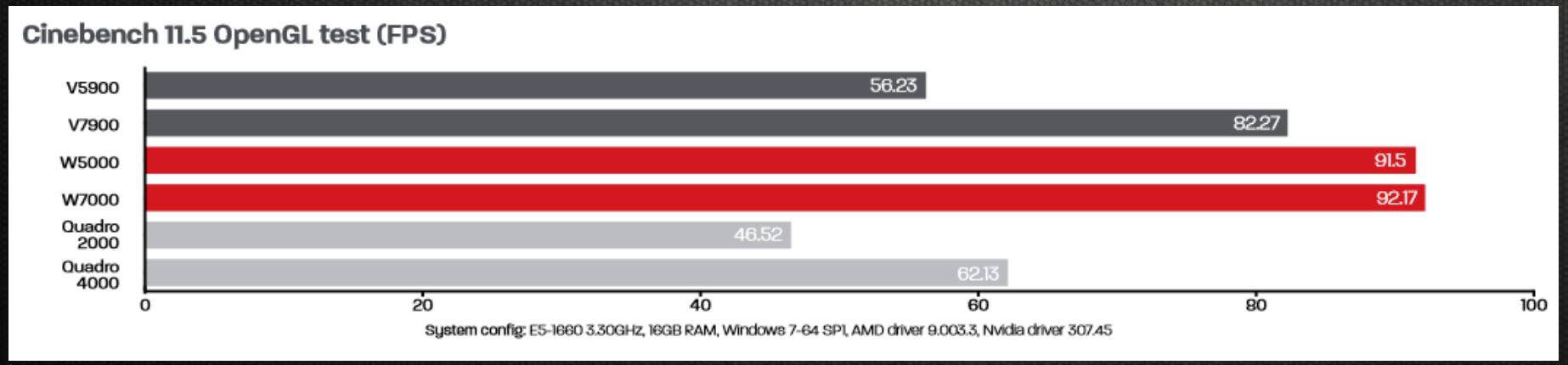
Bullet Physics is an open source library GPU accelerated with OpenCL™

"The new enhancements to the Bullet Physics engine in the Extension for Autodesk® Maya® 2014, empower artists to create large-scale, highly realistic dynamic and kinematic simulations. Plus the AMD FirePro™ accelerates the Bullet plug-in for increased capability for compound collision shapes from multiple meshes, additional collisions with concave shapes, along with rigid set support for increased scalability."

Rob Hoffmann, Senior Product Marketing Manager,
Autodesk Media & Entertainment

GPU ACCELERATED MAXON CINEMA 4D

- Cinema 4D uses OpenCL™ accelerated Bullet engine for physic
  calculations
- AMD FirePro™ drivers have special tuning to give excellent
  performance
  - AMD GeometryBoost technology and leading-edge graphics
    memory bandwidth help users take new sculpting, physics
    and global illumination capabilities to the max in Cinema 4D



CINEMA 4D by MAXON

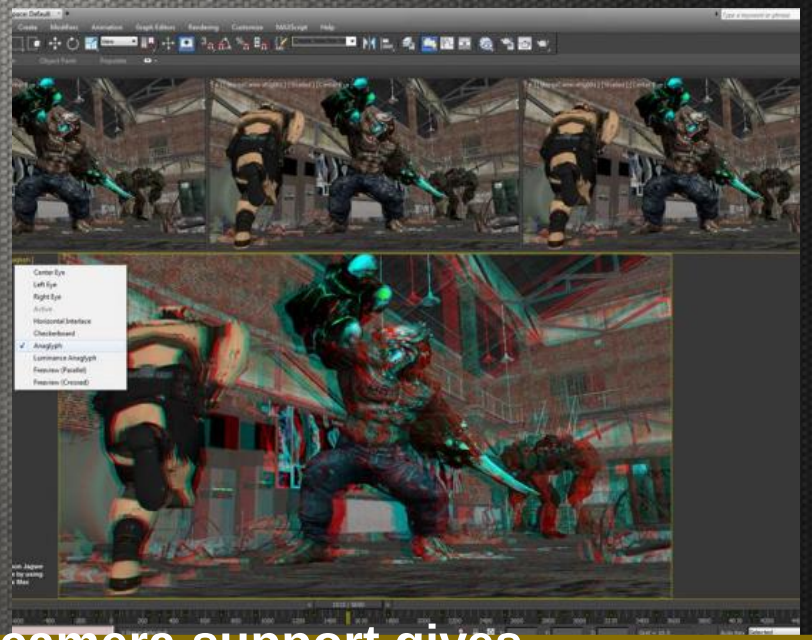
http://www.fireprographics.com/maxon/index.asp

AUTODESK® 3DS MAX® EXTENSION: ACTIVE STEREO

EXCLUSIVELY FOR AMD FIREPRO™ USERS

- First time a complete integrated stereo workflow available in 3ds Max
- Requires subscription and HD3D supported stereo display

AUTODESK 3DS MAX



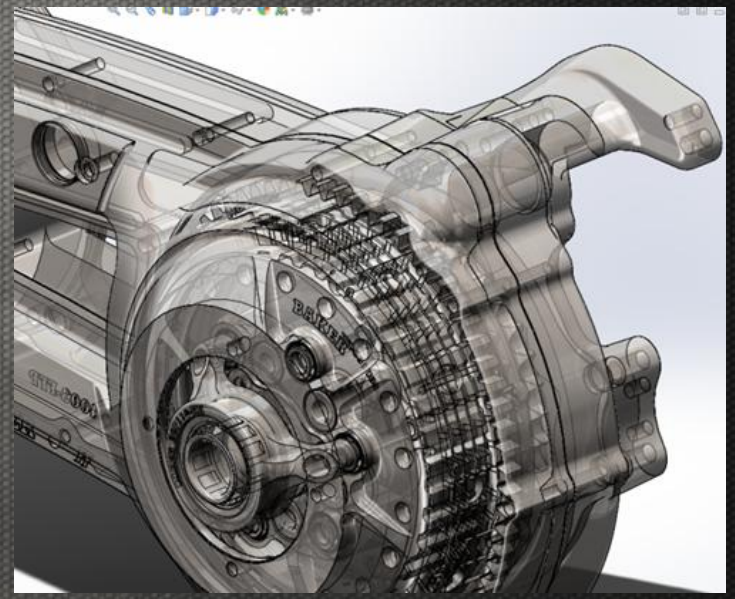
“The new 3ds Max extension with Stereo camera support gives added functionality for increased productivity to AMD FirePro™ users on Autodesk Subscription. Active Stereo is also available and only supported on AMD FirePro™ graphics cards. Artists can begin to take advantage of the new functionality for 3ds Max with AMD hardware to enable a broader set of tasks for creating exceptional content."

Rob Hoffmann, Senior Product Marketing Manager,
Autodesk Media & Entertainment

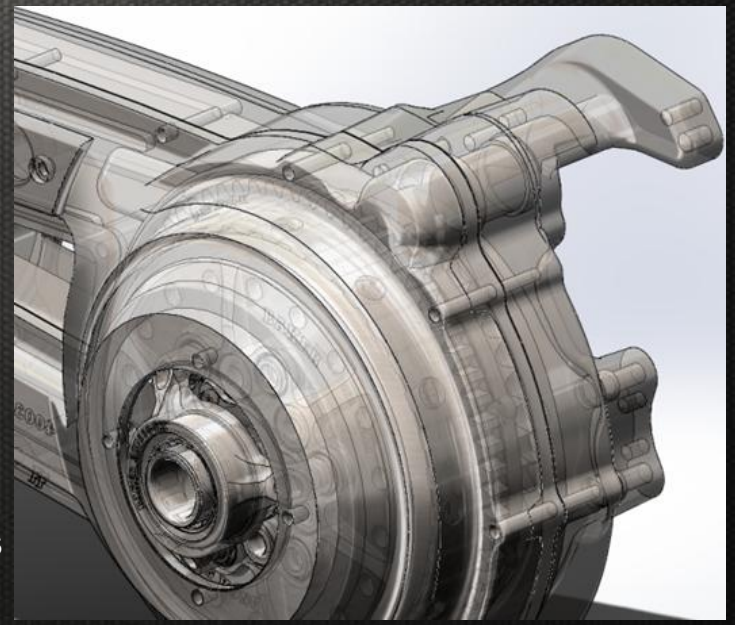
SOLIDWORKS® 2014 PREVIEW

- GPU accelerated OIT
- Order Independent Transparency (OIT) assembles a “pixel-accurate” representation of the model and its surrounding geometry in the GPU memory
  - Users can see objects closer to the screen more accurately
  - Helps improve the designer’s sense of “design intuition” and aids in better decision-making throughout the product development stages

With OIT



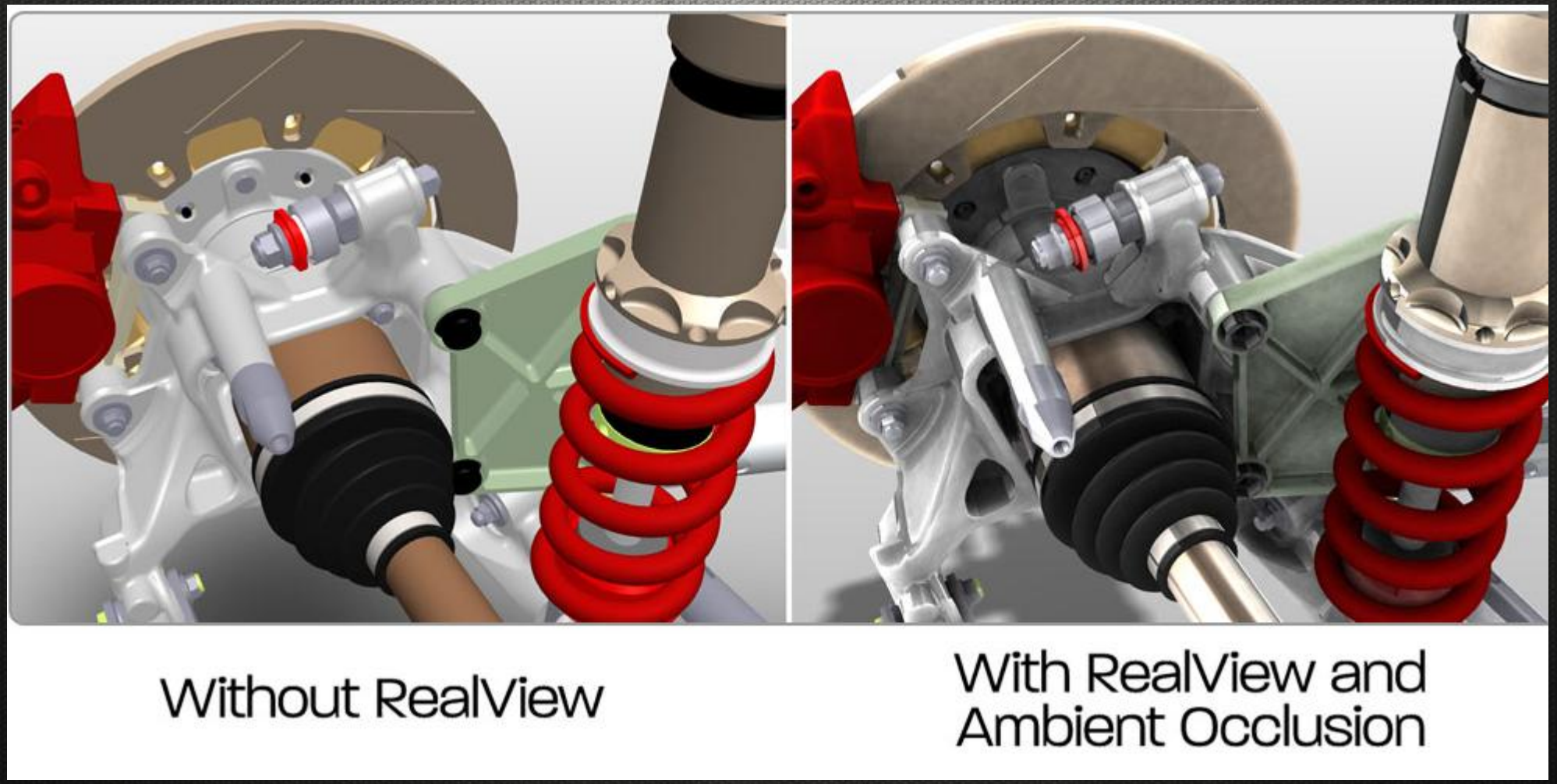
Without OIT



Images Courtesy of Dassault Systemès

SOLIDWORKS® 2013

POWERFUL REAL-TIME PREVIEWS WITH REALVIEW™



www.fireprographics.com/solidworks

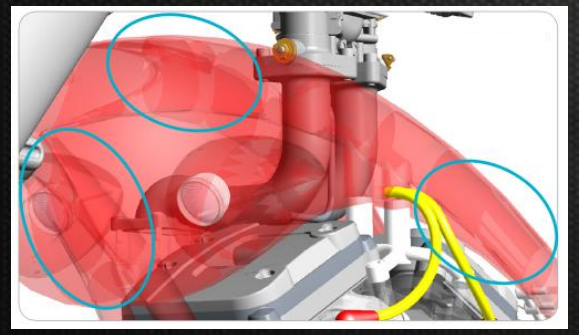
Workstation with dual Intel CPUs E5-1660 @ 3.30GHz, 16GB RAM, Win7 64-bit SP1, AMD driver 9.003.3.3

AMD FIREPRO™ PRO GRAPHICS ACCELERATE OIT IN PTC® CREO® 2.0

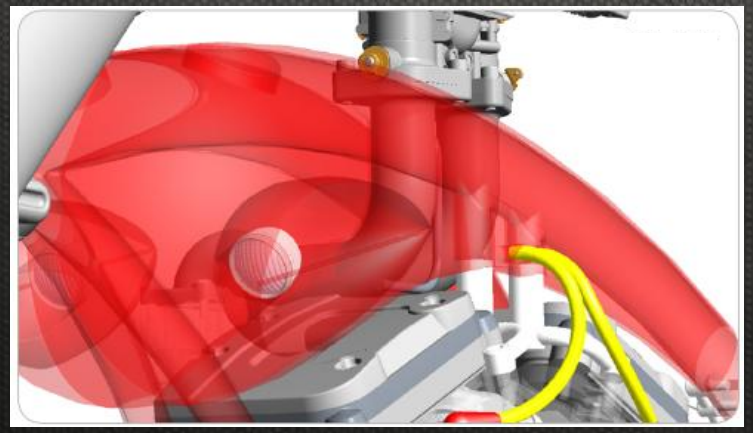
- AMD FirePro™ provides certified workflow performance gains for PTC Creo Parametric 2.0 using the new and exclusive GPU-accelerated Order Independent Transparency mode (OIT)
  - Provides greater “design intuition” for faster insight
  - Up to 17x as fast as blended mode and no visual artifacts² (with an AMD FirePro W7000)
  - AMD Catalyst™ Pro drivers optimized and certified for PTC Creo Parametric 2.0

Exclusively for AMD FirePro™ Users

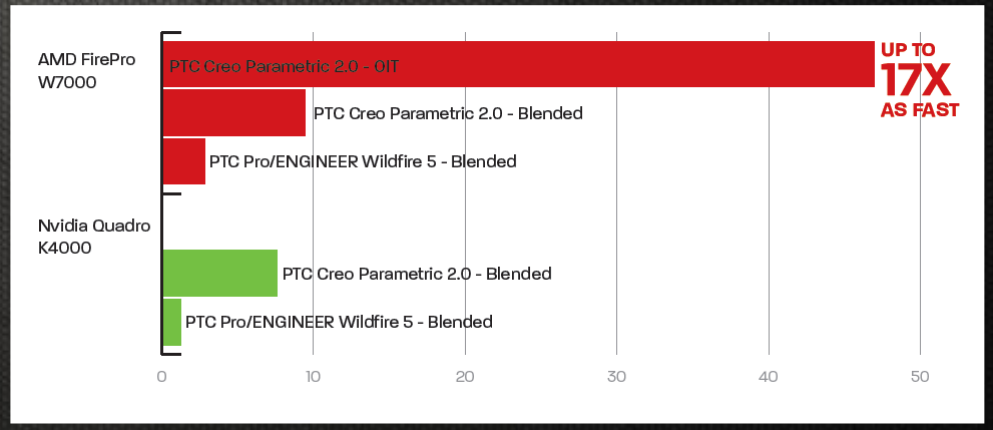
PTC Creo®



Without OIT



With OIT



Benchmarking graph showing up to 17x faster viewport performance with PTC Creo®Parametric® 2.0 "OIT" accelerated transparency mode²

AMD FIREPRO™ ACCELERATES SIEMENS PLM SOFTWARE NX WORKFLOWS

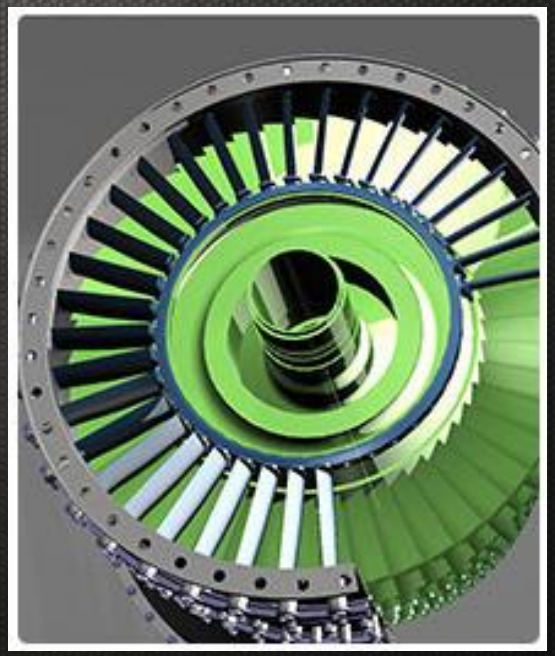
AMD

- Support for the new Advanced Studio mode for realistic representations of designs without losing interactivity
  - 28nm AMD GCN architecture delivers high in-app shader performance
  - AMD Catalyst™ Pro drivers certified and optimized for NX 8.5
  - AMD Eyefinity technology ideal for PLM workflows using multiple apps

SIEMENS NX

Studio Mode with True Shading

Advanced Studio Mode



http://www.fireprographics.com/siemens/index.asp

DISCLAIMER

The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions and typographical errors.

The information contained herein is subject to change and may be rendered inaccurate for many reasons, including but not limited to product and roadmap changes, component and motherboard version changes, new model and/or product releases, product differences between differing manufacturers, software changes, BIOS flashes, firmware upgrades, or the like. AMD assumes no obligation to update or otherwise correct or revise this information. However, AMD reserves the right to revise this information and to make changes from time to time to the content hereof without obligation of AMD to notify any person of such revisions or changes.

AMD MAKES NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE CONTENTS HEREOF AND ASSUMES NO RESPONSIBILITY FOR ANY INACCURACIES, ERRORS OR OMISSIONS THAT MAY APPEAR IN THIS INFORMATION.

AMD SPECIFICALLY DISCLAIMS ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR ANY PARTICULAR PURPOSE. IN NO EVENT WILL AMD BE LIABLE TO ANY PERSON FOR ANY DIRECT, INDIRECT, SPECIAL OR OTHER CONSEQUENTIAL DAMAGES ARISING FROM THE USE OF ANY INFORMATION CONTAINED HEREIN, EVEN IF AMD IS EXPRESSLY ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

ATTRIBUTION