

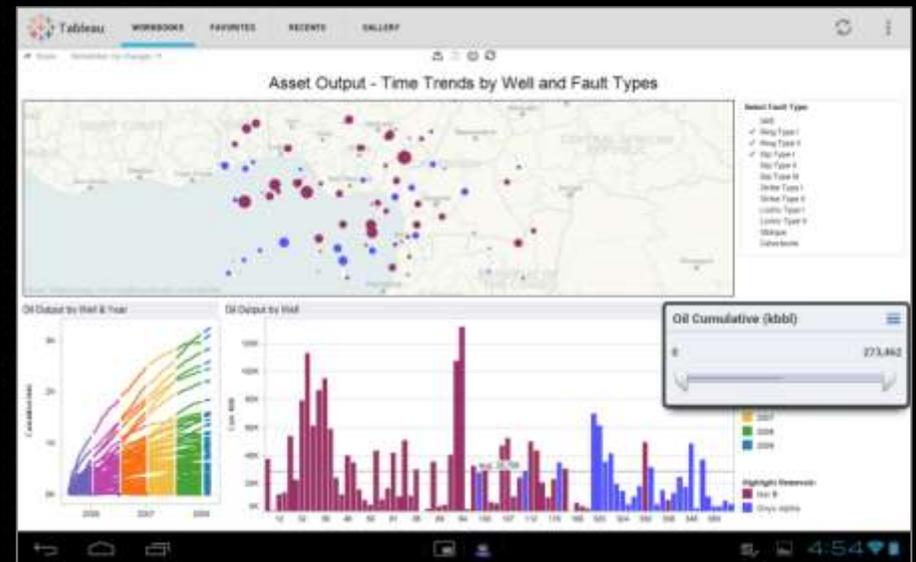
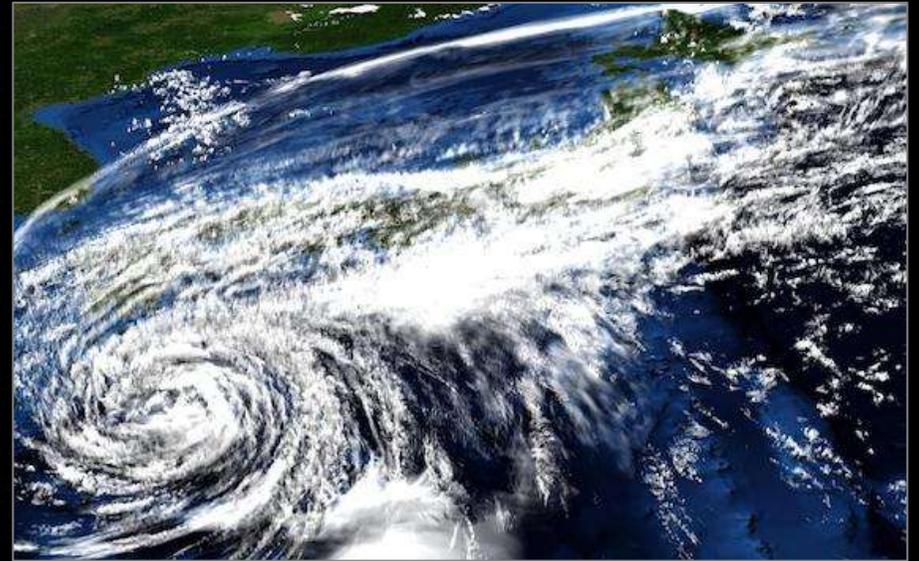


NVIDIA®

GPU - БУДУЩЕЕ ГИБРИДНЫХ ВС

Антон Джораев

Современный мир
требует все больших
вычислительных
ресурсов

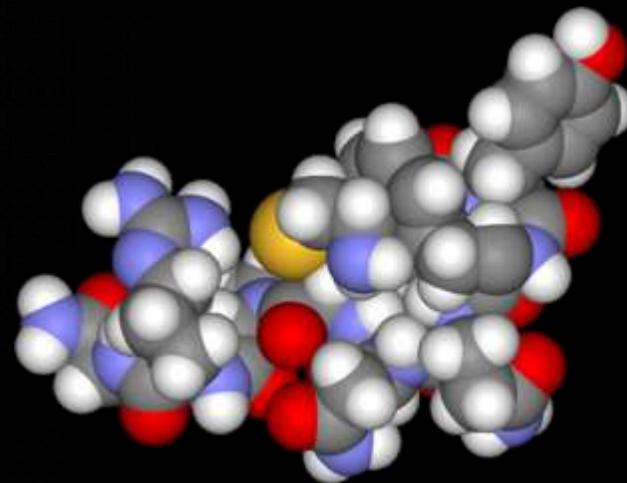
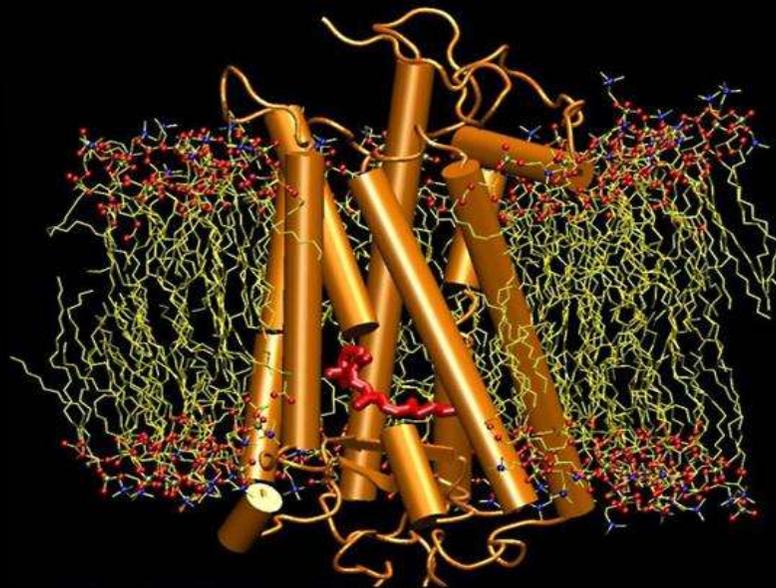


Современный мир
требует все больших
вычислительных
ресурсов

Google

Яндекс

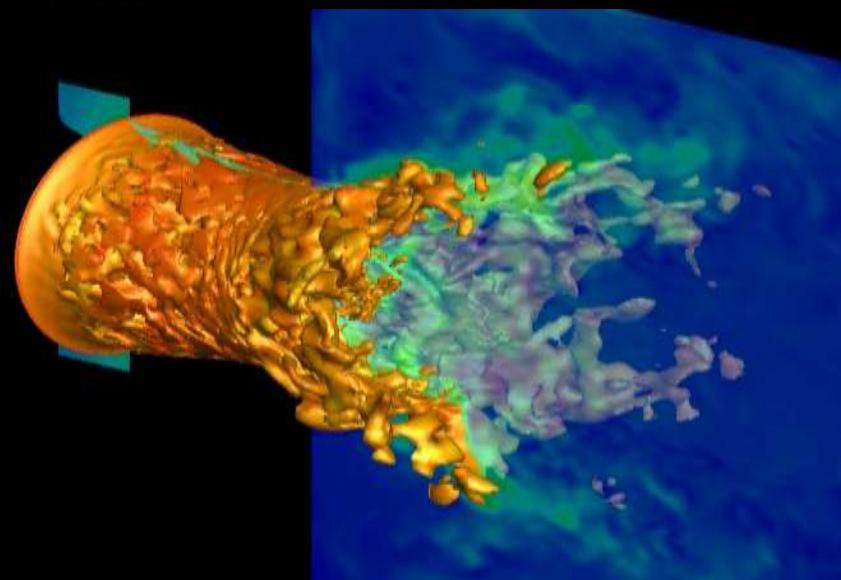
Современный мир
требует все больших
вычислительных
ресурсов



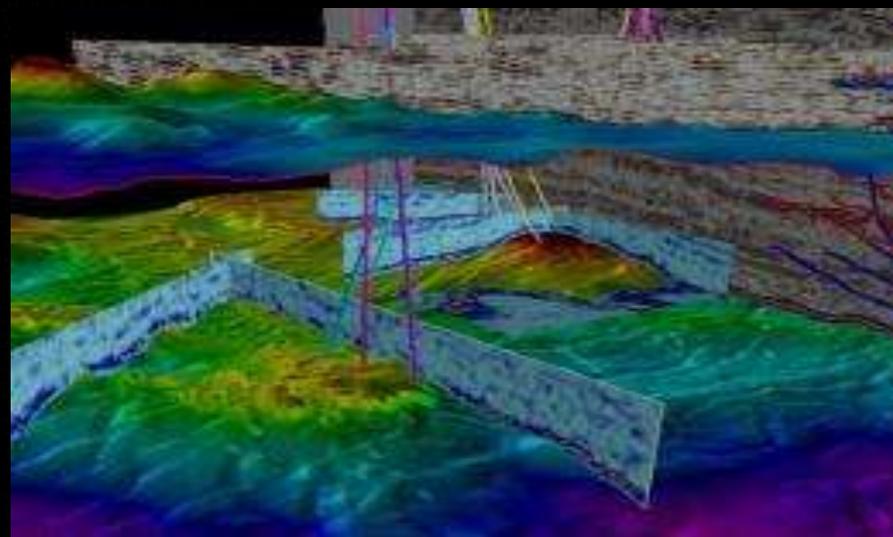
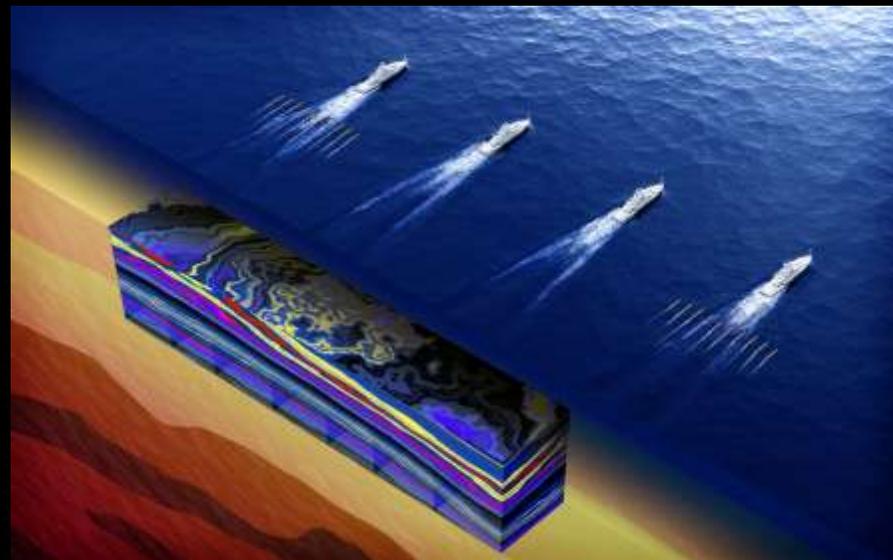
Современный мир
требует все больших
вычислительных
ресурсов



Современный мир
требует все больших
вычислительных
ресурсов



Современный мир
требует все больших
вычислительных
ресурсов



Современный мир
требует все больших
вычислительных
ресурсов



Проблемы архитектуры x86

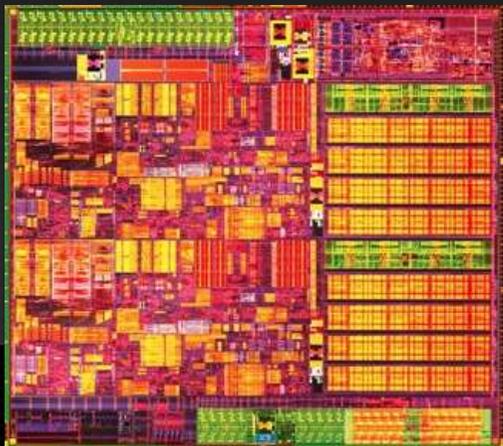
- Оптимизация под последовательный код, низкая произв. SIMD
- Низкая эффективность - стоимость, энергетика, охлаждение
- Огромное количество ПО написанного давно и неэффективно

CPU

1690 pJ/flop

Оптимизирован для исполнения
последовательного кода

Неконтролируемый кэш

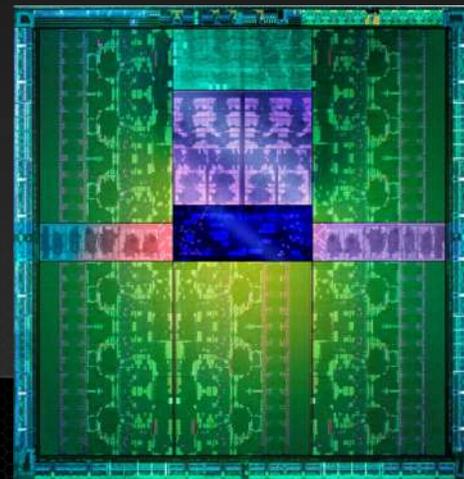


GPU

140 pJ/flop

Оптимизирован для работы с
большим количеством данных

Явное управление памятью GPU

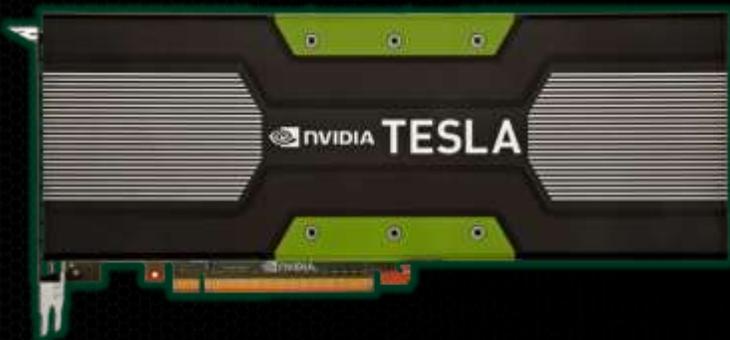


GPU - будущее HPC

- GPU - самая производительная архитектура
- GPU - наиболее энергоэффективная архитектура
- GPU доступны повсеместно

GPU - самый производительный процессор

Tesla K20X



Кол-во CUDA ядер	2688
Пиковая произв-ть DP DGEMM	1.32 TF 1.22 TF
Пиковая произв-ть SP SGEMM	3.95 TF 2.90 TF
Пропускная способность памяти	250 GB/s
Объем памяти	6 GB
Потребление	235W

Самый энергоэффективный компьютер мира

3150 MFLOPS/Watt

128 Tesla K20

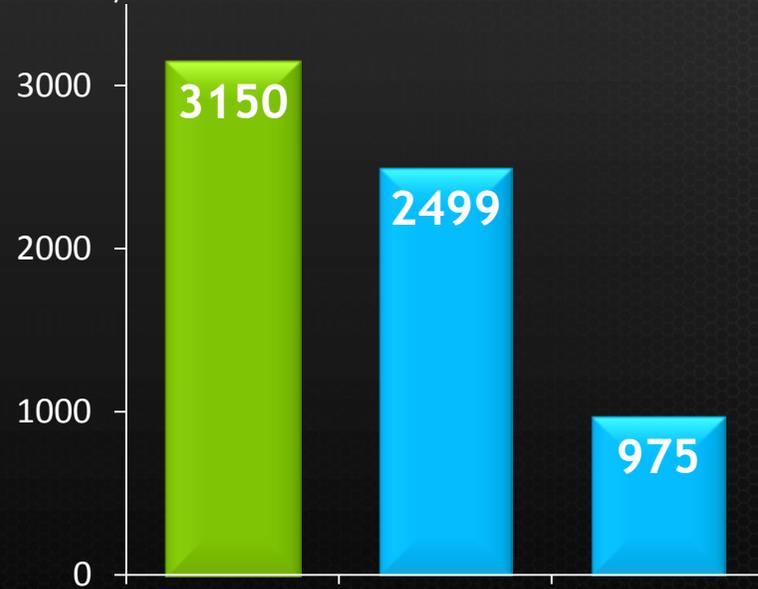


CINECA Eurora

"Liquid-Cooled" Eurotech Aurora Tigon

Эффективнее Xeon Phi и Xeon

MFLOPS/Watt

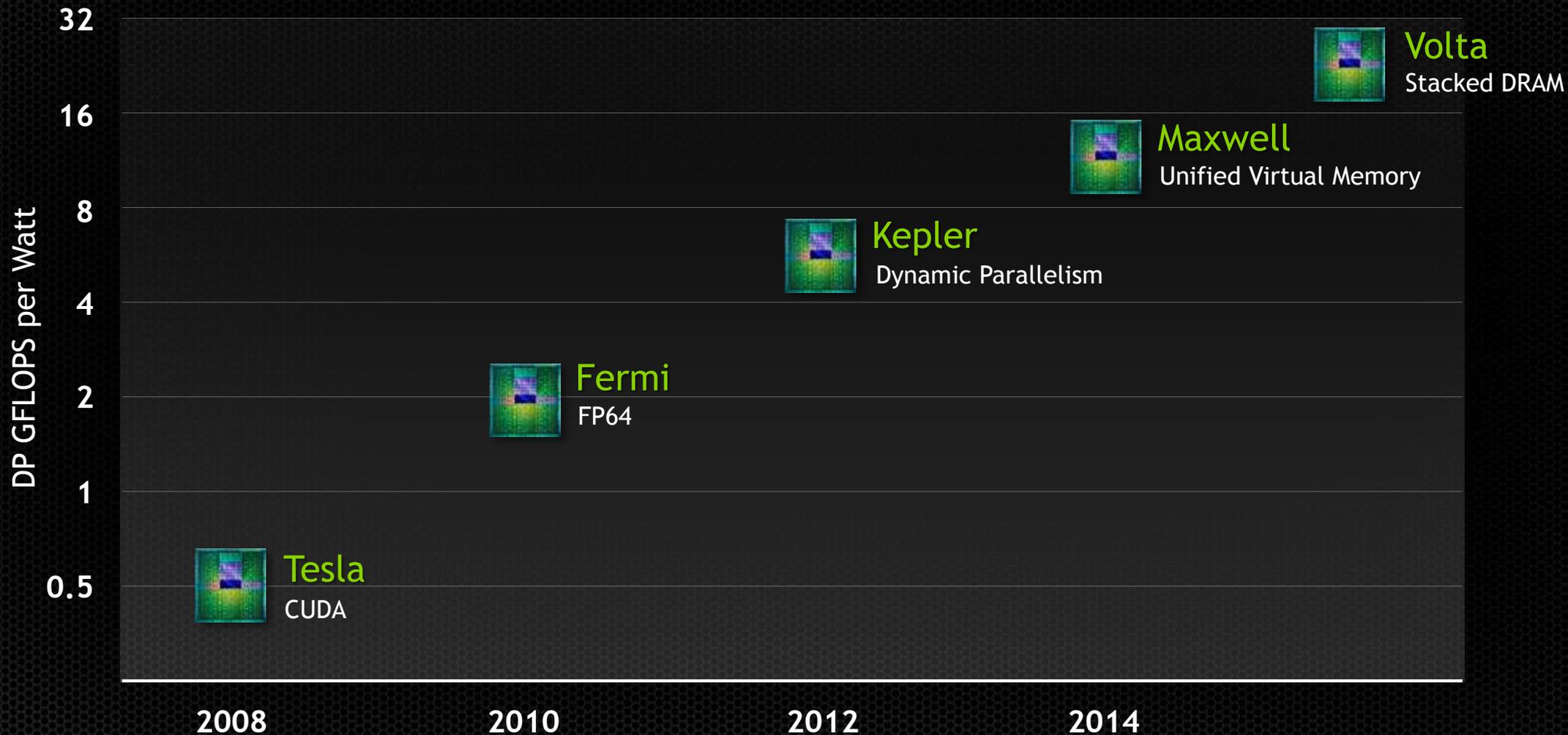


CINECA Eurora-Tesla K20

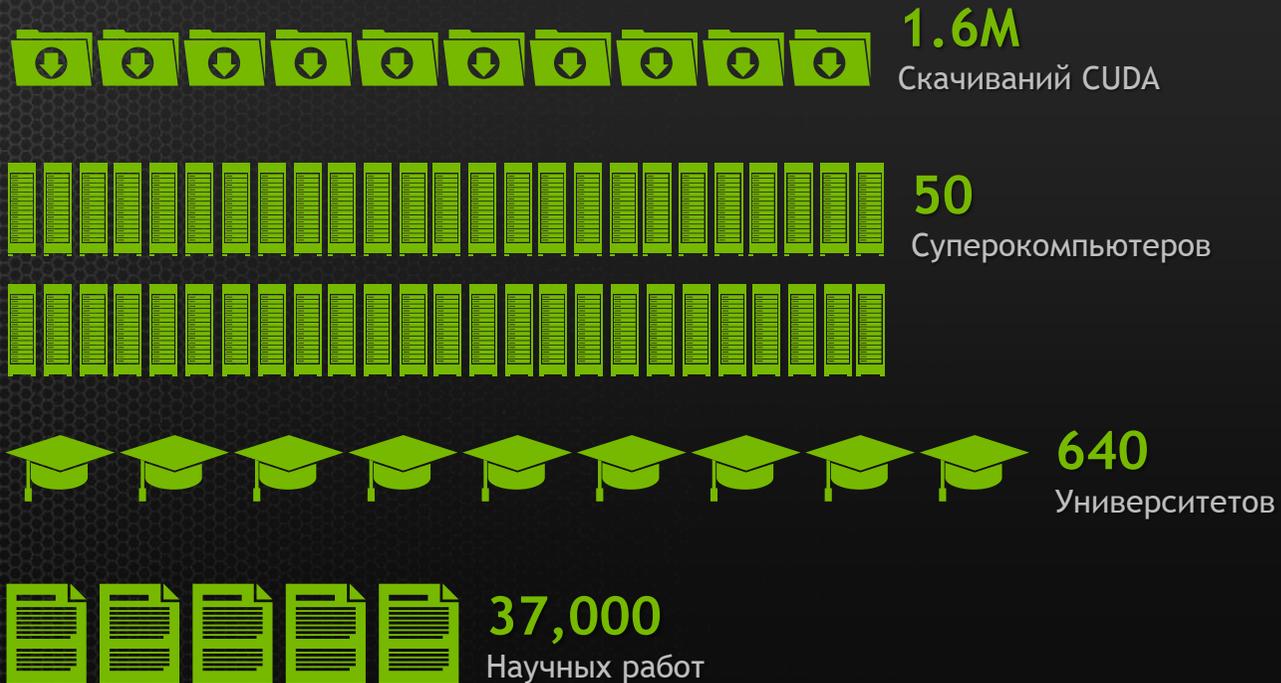
NICS Beacon-Greenest Xeon Phi System

C-DAC-Greenest CPU System

GPU Roadmap



Распространенность вычислений на GPU



2013

Вычисления на GPU в России

 **20M+**
GPU с CUDA

 **16**
Суперкомпьютеров



 **20+**
Университетов

 **500+**
Научных работ

2013

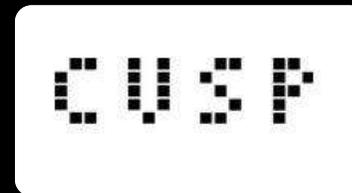
GPU доступны каждому исследователю

- 30% (120 из 400) институтов РАН решают вычислительные задачи
- 65% (78) из них применяют вычисления на GPU

GPU Accelerated Libraries

“Drop-in” Acceleration for your Applications

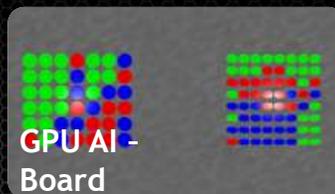
Linear Algebra
FFT, BLAS,
SPARSE, Matrix



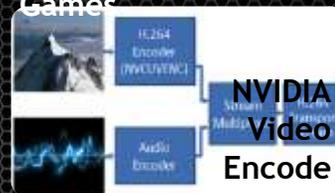
Numerical & Math
RAND, Statistics



Data Struct. & AI
Sort, Scan, Zero Sum



Visual Processing
Image & Video





POPULAR GPU-ACCELERATED APPLICATIONS

CONTENTS

- 02 Research: Higher Education and Supercomputing
 - COMPUTATIONAL CHEMISTRY AND BIOLOGY
 - BIOMEDICAL ANALYTICS
 - PHYSICS
 - WEATHER AND CLIMATE FORECASTING
- 06 Defense and Intelligence
- 07 Computational Finance
- 08 Manufacturing: CAD and CAE
- COMPUTER-AIDED DESIGN
- COMPUTATIONAL FLUID DYNAMICS
- COMPUTATIONAL STRUCTURAL MECHANICS
- ELECTRONIC DESIGN ASSISTANCE
- 10 Media and Entertainment
 - ANIMATION, MODELING AND RENDERING
 - COLOR CORRECTION AND GRAIN MANAGEMENT
 - COMPOSITING, FINISHING AND EFFECTS
 - EDITING
 - ENCODING AND DIGITAL DISTRIBUTION
 - ON-SET GRAPHICS
 - ON-SET, REVIEW AND STEREO TOOLS
 - SIMULATION
 - WEATHER GRAPHICS
- 14 Oil and Gas

Research: Higher Education and Supercomputing

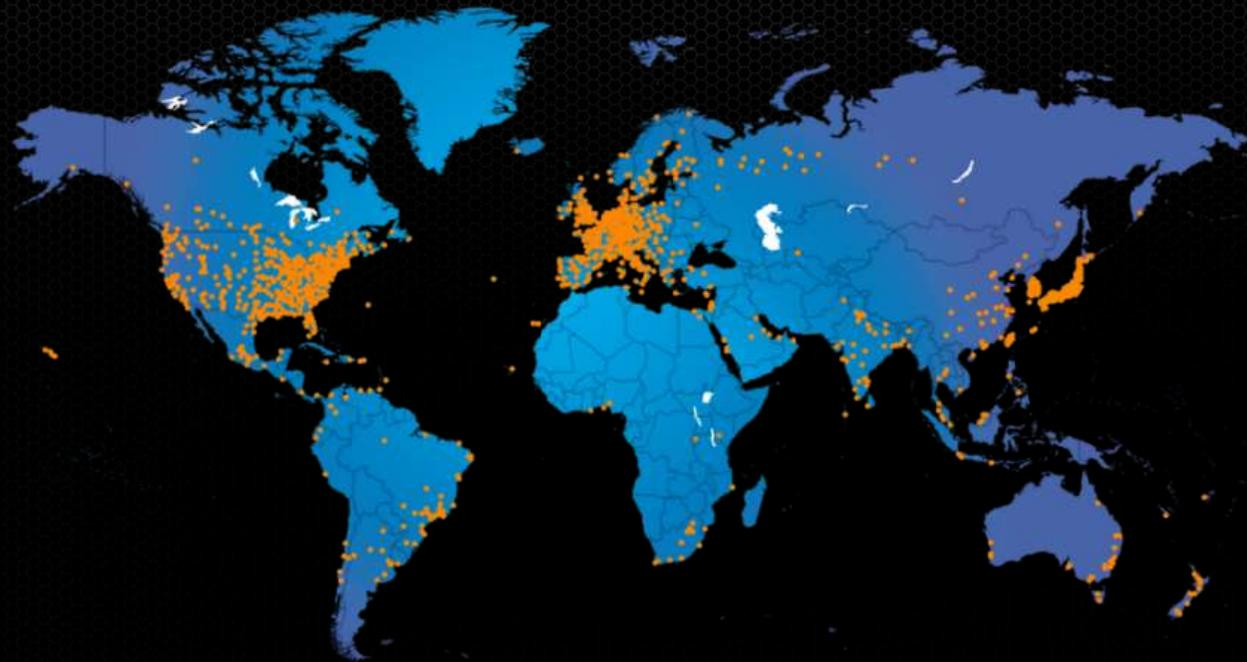
COMPUTATIONAL CHEMISTRY AND BIOLOGY

Application	Description	Architecture	GPU Support	Availability	
BerryCUDA	Sequence mapping software	Alignment of short sequencing reads	6-10x	T 2015, 2016, K10, K20, K20X	Available now
CUDA-M++	Open source software for Smith-Waterman protein-database searches on GPUs	Parallel search of Smith-Waterman database	10-100x	T 2015, 2016, K10, K20, K20X	Available now
CUSIMM	Parallelized short read aligner	Parallel, accurate long read aligner - gapped alignments to large genomes	15x	T 2015, 2016, K10, K20, K20X	Available now
GPU-BLAST	Local search with fast k-tuple heuristic	Protein alignment according to heuristic, much spm threads	9-14x	T 2015, 2016, K10, K20, K20X	Single only
GPU-INDEXER	Parallelized local and global search with profile hidden Markov models	Parallel local and global search of hidden Markov Models	10-100x	T 2015, 2016, K10, K20, K20X	Available now
ncCUDA-MEME	GPU accelerated motif discovery algorithm based on MEME	Scalable motif discovery algorithm based on MEME	6-10x	T 2015, 2016, K10, K20, K20X	Available now
SeqFind	A GPU Accelerated Sequence Analysis Toolkit	Reference assembly, blast, Smith-Waterman, k-mer, de novo assembly	100x	T 2015, 2016, K10, K20, K20X	Available now
MEME	Systematic Smith-Ablorsson for CUDA/CUDA, suffix array based regular index and support	Fast short read alignment	4-5x	T 2015, 2016, K10, K20, K20X	Available now
WMLM	File formats linear models to a fast design and response	Parallel linear regression on multiple similarly-shaped models	150x	T 2015, 2016, K10, K20, K20X	Available now

Application	Description	Architecture	GPU Support	Availability	
Alanine	Modes molecular dynamics of ligand-protein for simulations of proteins, DNA and ligands	Computations for 100k GPU	4-21x	T 2015, 2016, K10, K20, K20X	Available now
ACEMD	GPU simulation of molecular mechanics force fields, implicit and explicit solvent	Written for use on GPUs	140x (vs CPU)	T 2015, 2016, K10, K20, K20X	Available now
AMBER	Suite of programs to simulate molecular dynamics on GPUs	PME: implicit and explicit solvent	95-100x (vs CPU)	T 2015, 2016, K10, K20, K20X	Available now
DL-POLY	Simulate macromolecules, polymers, ionic systems, etc on a distributed memory parallel computer	Two-body forces, Link-rod pairs, Bonded SPME forces, Shell-CP	4x	T 2015, 2016, K10, K20, K20X	Available now
CHARMM	MD package to simulate molecular dynamics on GPUs	Implicit Sol, Explicit Sol, Coarse-grained	70x	T 2015, 2016, K10, K20, K20X	Available now
DRYMACS	Simulation of biomolecules, macromolecules, and complexed biomolecular interactions	Implicit Sol, Explicit Sol, Coarse-grained	140x (vs CPU)	T 2015, 2016, K10, K20, K20X	Single only
HOLO-Blue	Particle dynamics package written primarily for GPUs	Written for GPUs	2x	T 2015, 2016, K10, K20, K20X	Available now
LAMMPS	Classical molecular dynamics package	Lennard-Jones, Morse, Buckingham, CHARMM, Substanced, Coarse-grain, SHC, Anisotropic Dipole, etc. (various)	3-10x	T 2015, 2016, K10, K20, K20X	Available now
HOLO	Designed for high-performance simulation of large molecular systems	HOLO atom capable	6-24x (vs CPU)	T 2015, 2016, K10, K20, K20X	Available now
HOLO	Library and application for molecular dynamics for HPC with GPUs	Implicit and explicit solvent, custom forces	100x (vs CPU)	T 2015, 2016, K10, K20, K20X	Available now

242 GPU-Accelerated Applications
www.nvidia.com/appscatalog

NVIDIA приобрела The Portland Group (PGI)



PGI has over 25,000 users at over 5,000 sites worldwide

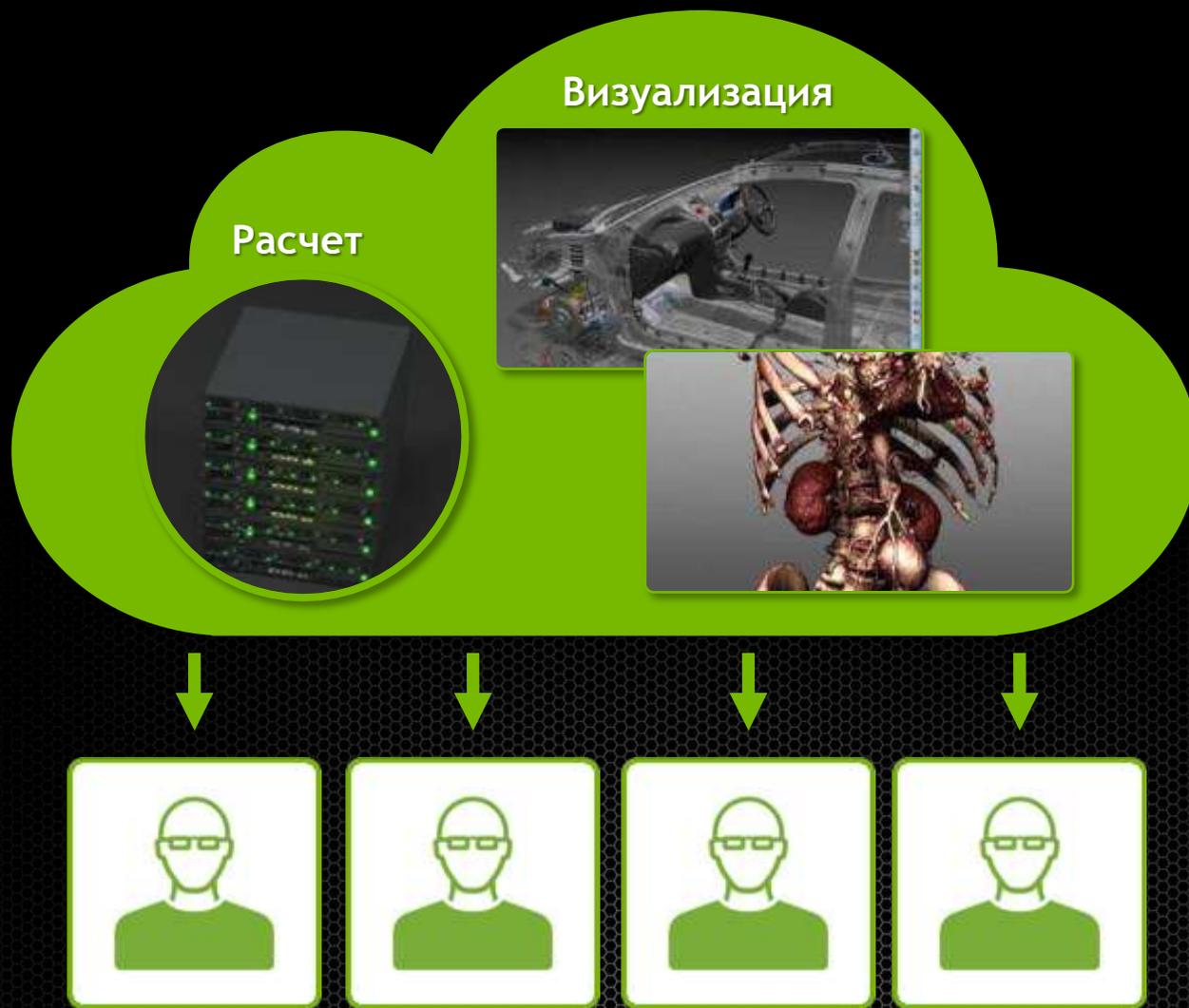
Для того чтобы:

- **привнести инвестиции в развитие OpenACC и CUDA Fortran**
- **усилить экспертизу в области разработки компиляторов**
- **сделать средства разработки для CPU/GPU еще более удобными и эффективными**

GPU - ответ на вызовы будущего

- Максимальная производительность
- Самая высокая энергоэффективность
- Большое количество ПО и библиотек

Виртуальные рабочие места - NVIDIA GRID



Полноценный ПК

Доступен на любом персональном устройстве, которое пользователь может принести на работу.



nVIDIA®

АНТОН ДЖОРАЕВ
adzhoraev@nvidia.com

Семинары AbraCUDAbra

- Среда, 25 сентября, 21:30 - 23:00
 - Введения в вычисления на GPU
- Пятница, 27 сентября, 16:30 - 18:00
 - Математические библиотеки на GPU: производительность
 - Директивы OpenACC - стандарт программирования параллельных вычислений

Всероссийский конкурс GPU: серьезные ускорители для больших задач

- Информация на сайтах:
 - ccoe.msu.ru
 - www.nvidia.ru
- Различные номинации, интересные призы...

IBM Open Power Consortium

Благодаря сотрудничеству NVIDIA и IBM, системам на базе процессоров Power будет доступно ускорение на GPU.

www.open-power.org