JÜLICH
APPLICATION LAB
NVIDIA

Mitglied der Helmholtz-Gemeinschaft

JÜLICH
FORSCHUNGSZENTRUM

NVIDIA Application Lab at Jülich

Dirk Pleiter | Jülich Supercomputing Centre (JSC)

Forschungszentrum Jülich at a Glance (status 2010)

- **Budget:** 450 mio Euro
- **Staff:** 4,800 (thereof 1,630 scientists)
- **Visiting scientists:** 900 per year
- **Trainees:** 90
- **Publications:** 1,800
- **Protective rights and licences:** 14,800
- **Research fields:** health, energy and environment, and information technology; key technologies for tomorrow

Jülich Supercomputing Centre

Supercomputer operation for:

- Centre – FZJ,
- Regional – JARA
- Helmholtz & National – NIC, GCS
- Europe – PRACE, EU projects

Application support

- User support; coordination with SimLabs
- Scientific Visualization
- Peer review support and coordination

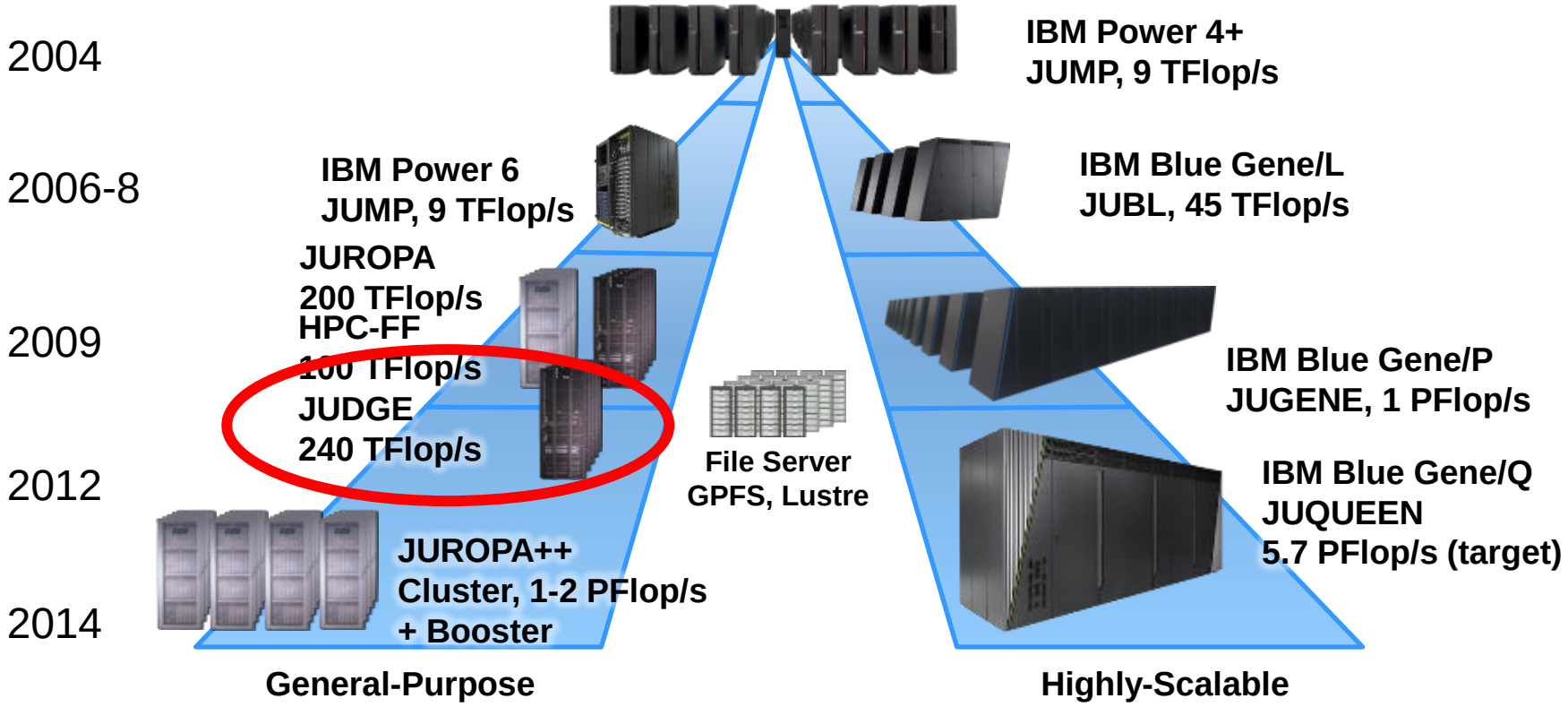
R&D work

- Algorithms, performance analysis and tools
- Community data management service
- Computer architectures, Exascale Laboratories: EIC, ECL, NVIDIA

Education and Training

JÜLICH APPLICATION LAB NVIDIA

Supercomputer Systems: Dual Track Approach



JUDGE Cluster

System

- 206 IBM iDataPlex nodes
  - *2 Tesla M2050 or M2070 per node*
- Infiniband QDR network
- Peak performance: 239 Tflops

Users

- Institute for Advanced Simulations
  - *Molecular dynamics and mechanics, micro-magnetism simulations, medical image reconstruction*
- JuBrain partition
- Milkey Way partition

NVIDIA Application Lab at Jülich

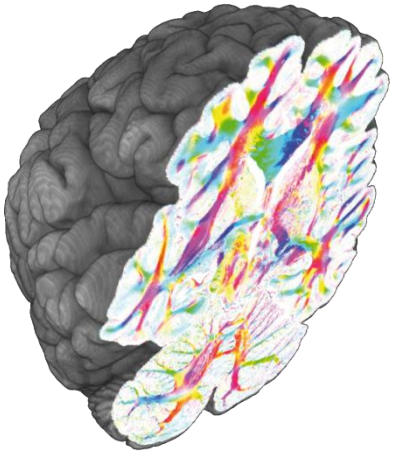
Collaboration between JSC and NVIDIA since July 2012

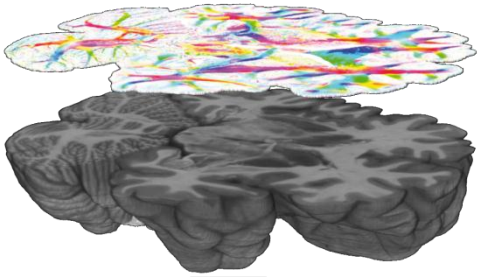
- Enable scientific applications for GPU-based architectures
- Provide support for their optimization
- Investigate performance and scaling

Work focus

- Application requirements analysis
- Kepler and CUDA feature analysis
- Parallelization on many GPUs
- Collaboration with performance tools developers
- Training

Pilot Application: JuBrain





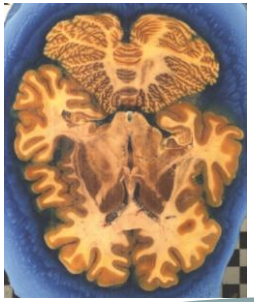
Application developed at the Institute of Neuroscience and Medicine (INM-1) at Forschungszentrum Jülich:

Katrin Amunts, Markus Axer, Marcel Huysegoms

Research goal

Accurate, highly detailed computer model of the human brain

Brain Section Images



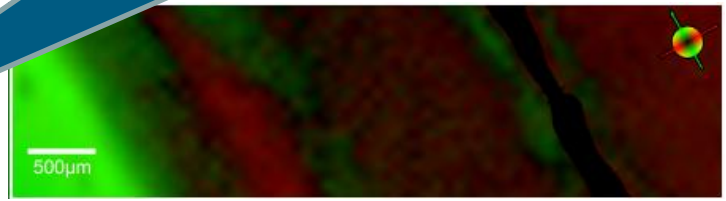
Blockface pictures

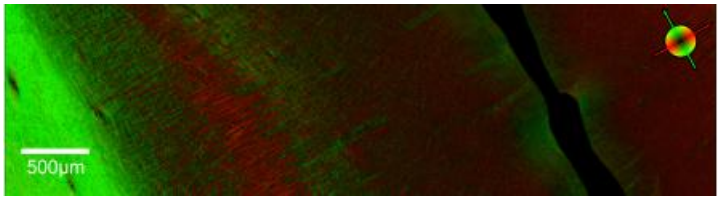
- Created while cutting brain in sections

Histological images

- Polarized light images
- Low resolution vs. high resolution
  - *100 µm → 3 µm pixel size*
  - *30 MBytes → 40 Gbytes data*

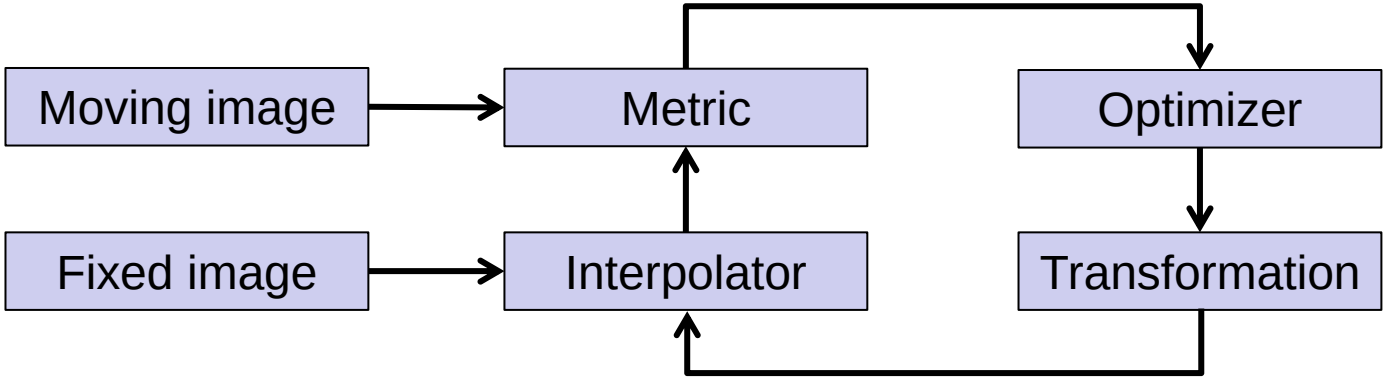
Exceeds GPU memory capacity





Challenge: 3d reconstruction

3D Reconstruction



Registration algorithms

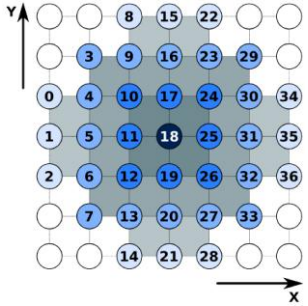
- Rigid registration → 3 parameters
- Afine registration → 6 parameters
- Elastic registration → O(100) parameters

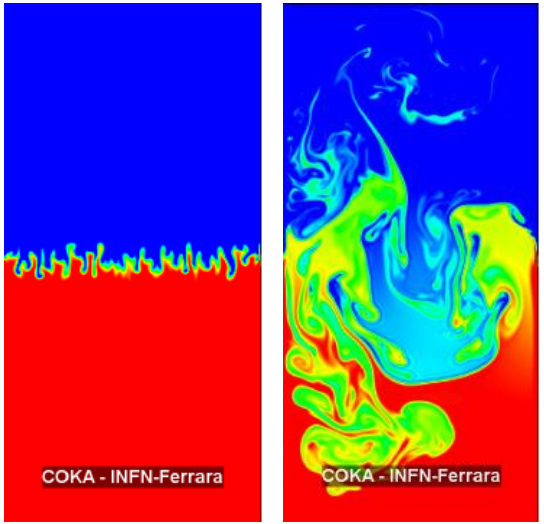
O(30) speedup on GPU

Fluid dynamics on Fermi and Kepler

Lattice Boltzmann method

- D2Q37 model
- Application developed at
  U Rome Tore Vergata/INFN,
  U Ferrara/INFN, TU Eindhoven
- Reproduce dynamics of fluid by simulating virtual particles which collide and propagate
- Simulation of large systems requires double precision computation on many GPUs

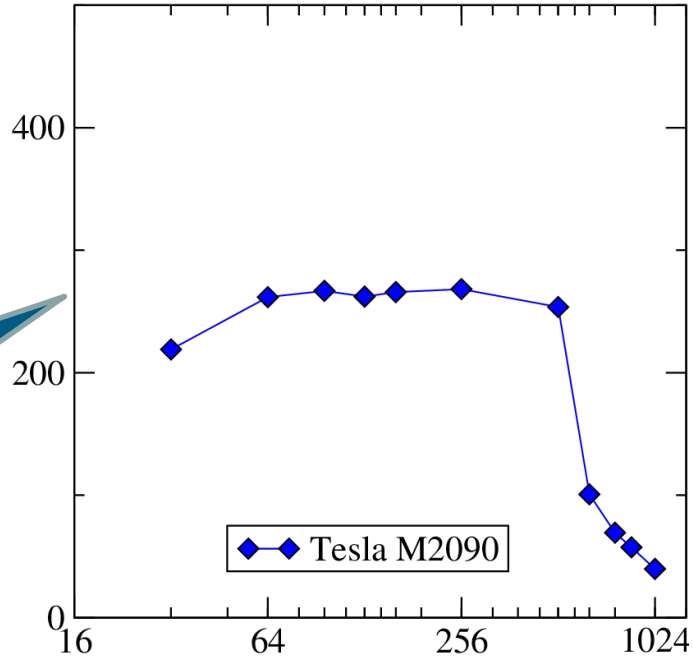




Collide kernel on Fermi

- Kernel dominated by arithmetic operations
- Floating-point performance as a function of the number of threads/block [GFlop/s]

Excellent performance on Fermi



Implementation:
F. Schifano (U Ferrara/INFN)

Kepler Performance Tuning

Performance analysis observations

- Significant increase of L1 cache misses
- 17% (Tesla M2090) → 67% (Tesla K20)

SM performance increased, but L1 cache capacity remained unchanged

```
for (i = 0; i < NPOP-1; i++) {
  lPop = p_prv[i*NX*NY + idx];
  u = u + param_cx[i] * lPop;
  v = v + param_cy[i] * lPop;
}
```

```
#pragma unroll
for (i = 0; i < NPOP-1; i++) {
  lPop = p_prv[i*NX*NY + idx];
  u = u + param_cx[i] * lPop;
  v = v + param_cy[i] * lPop;
}
```

Problem mitigation by simple code change

Enforce loop unrolling to eliminate indirect memory accesses

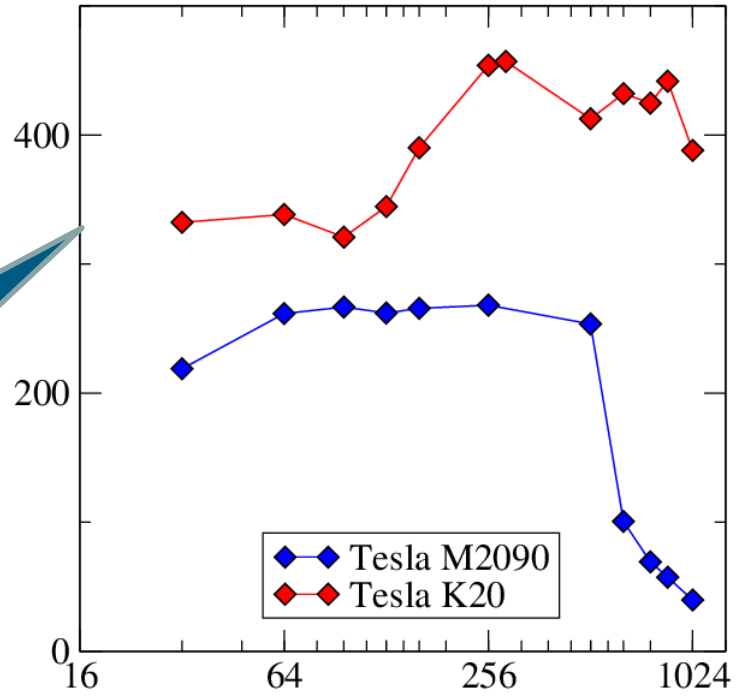
J. Kraus (NVIDIA Lab)

Collide kernel on Kepler GK110

Comparison Fermi vs. Kepler

- Grid size considered here: 252 x 16384
- Floating-point performance as a function of the number of threads/block

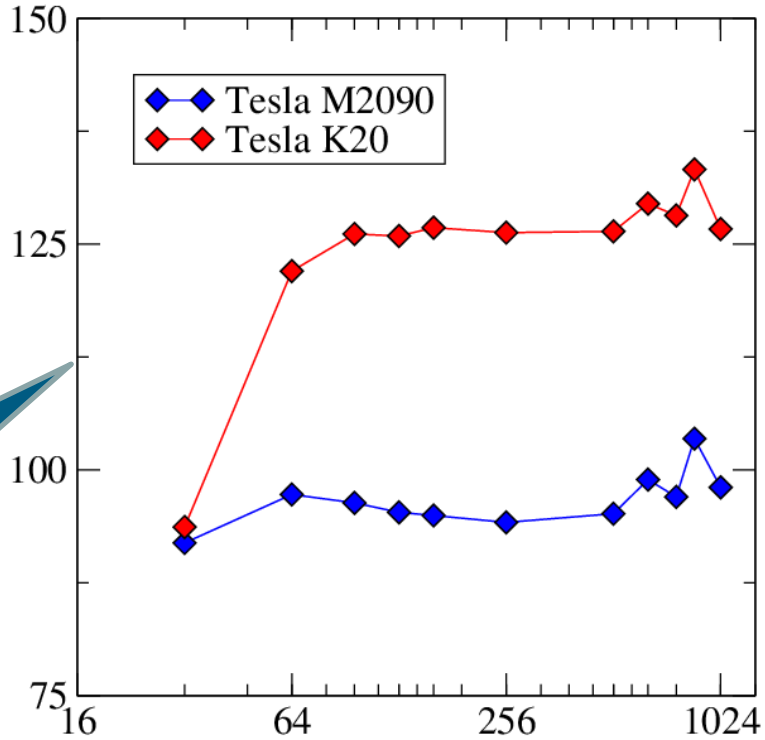
Performance improvement 1.7x



Propagate kernel

Kernel dominated by memory access

- Grid size considered here: 252 x 16384
- Memory bandwidth [GByte/s] as a function of the number of threads/block



Performance improvement 1.4x

Summary

NVIDIA Application Lab at Jülich

- New and fruitful model for collaboration
- We are just at the beginning ...

Application requirements analysis

- JuBrain: Project aiming for realistic model of the human brain

Kepler feature analysis

- Initial performance results for Lattice Boltzmann application on GK110
- Very high performance level reached on Fermi can be sustained