# The benefit of GPU HPC computation

Gilles Fourestey - EPFL/SCITAS

SKA Days 2019

INFORMATIQUE SCIENTIFIQUE & SUPPORT APPLICATIF
SCITAS
SCIENTIFIC IT AND APPLICATION SUPPORT

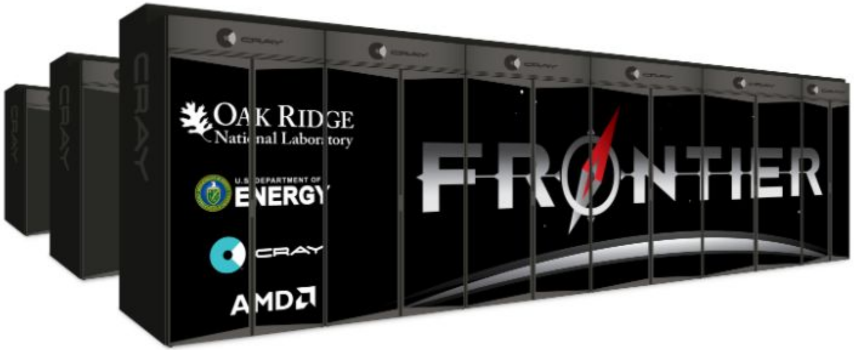
# GPUs will drive exascale

## U.S. Department of Energy and Cray to Deliver Record-Setting Frontier Supercomputer at ORNL

### Exascale system expected to be world's most powerful computer for science and innovation

**Topic:** Supercomputing

May 7, 2019




| Rank | System | Cores | (TFlop/s) | (TFlop/s) | (kW) |
|---|---|---|---|---|---|
| 1 | **Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , IBM DOE/SC/Oak Ridge National Laboratory United States | 2,414,592 | 148,600.0 | 200,794.9 | 10,096 |
| 2 | **Sierra** - IBM Power System S922LC, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States | 1,572,480 | 94,640.0 | 125,712.0 | 7,438 |
| 3 | **Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway , NRCPC National Supercomputing Center in Wuxi China | 10,649,600 | 93,014.6 | 125,435.9 | 15,371 |
| 4 | **Tianhe-2A** - TH-IVB-FEP Cluster, Intel Xeon E5-2692v2 12C 2.2GHz, TH Express-2, Matrix-2000 , NUDT National Super Computer Center in Guangzhou China | 4,981,760 | 61,444.5 | 100,678.7 | 18,482 |
| 5 | **Frontera** - Dell C6420, Xeon Platinum 8280 28C 2.7GHz, Mellanox InfiniBand HDR , Dell EMC Texas Advanced Computing Center/Univ. of Texas United States | 448,448 | 23,516.4 | 38,745.9 | |
| 6 | **Piz Daint** - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100 , Cray Inc. Swiss National Supercomputing Centre (CSCS) Switzerland | 387,872 | 21,230.0 | 27,154.3 | 2,384 |
| 7 | **Trinity** - Cray XC40, Xeon E5-2698v3 16C 2.3GHz, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect , Cray Inc. DOE/NNSA/LANL/SNL United States | 979,072 | 20,158.7 | 41,461.2 | 7,578 |
| 8 | **AI Bridging Cloud Infrastructure (ABCI)** - PRIMERGY CX2570 M4, Xeon Gold 6148 20C 2.4GHz, NVIDIA Tesla V100 SXM2, Infiniband EDR , Fujitsu National Institute of Advanced Industrial Science and Technology (AIST) Japan | 391,680 | 19,880.0 | 32,576.6 | 1,649 |
| 9 | **SuperMUC-NG** - ThinkSystem SD650, Xeon Platinum 8174 24C 3.1GHz, Intel Omni-Path , Lenovo Leibniz Rechenzentrum Germany | 305,856 | 19,476.6 | 26,873.9 | |
| 10 | **Lassen** - IBM Power System S922LC, IBM POWER9 22C 3.1GHz, Dual-rail Mellanox EDR Infiniband, NVIDIA Tesla V100 , IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States | 288,288 | 18,200.0 | 23,047.2 | |

# GPUs will drive exascale

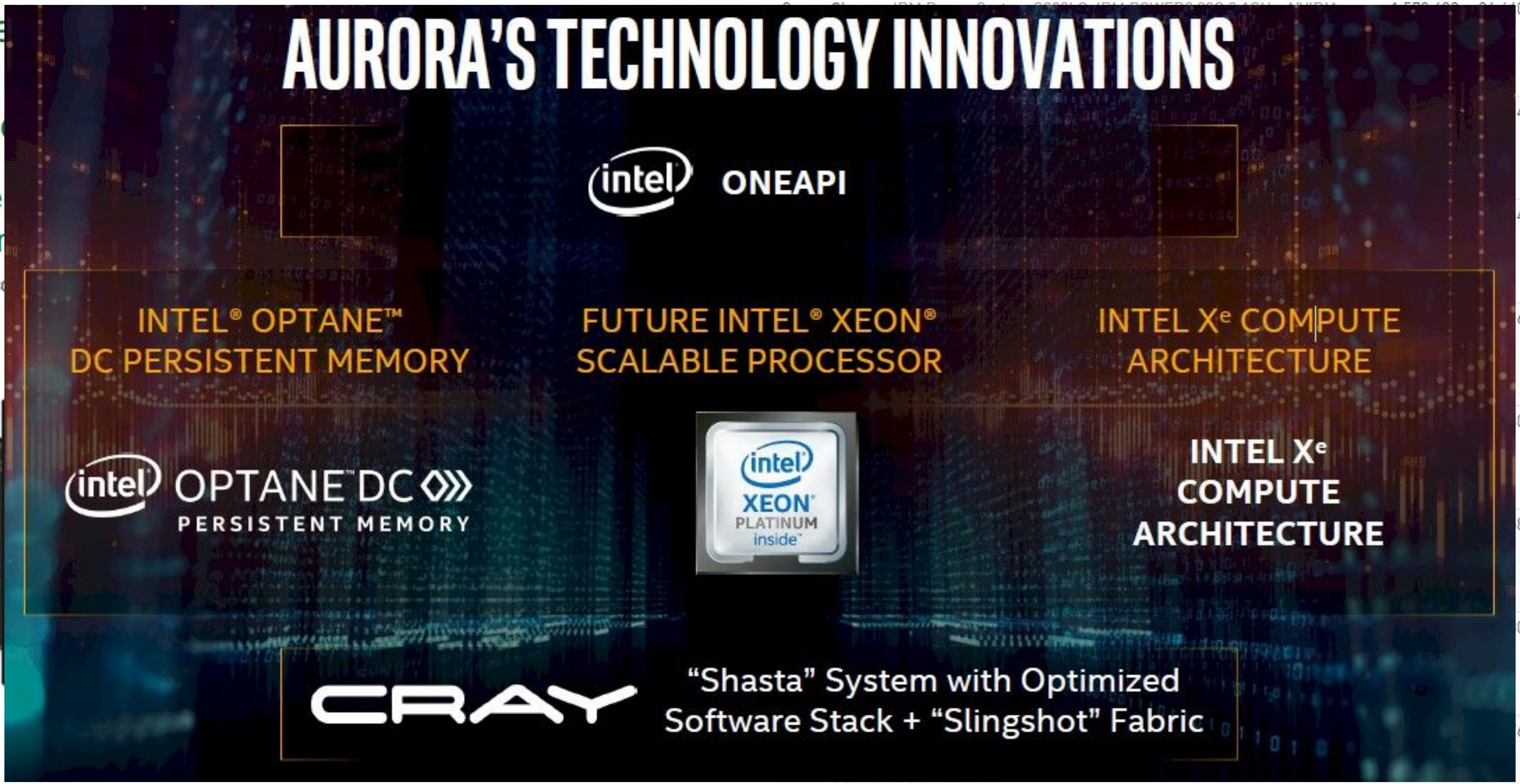
U.S. De
Deliver
Superc

Exascale
computer

**Topic:** Superc

May 7, 2019

| Rank | System | Cores | (TFlop/s) | (TFlop/s) | (kW) |
|---|---|---|---|---|---|
| 1 | **Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , IBM DOE/SC/Oak Ridge National Laboratory United States | 2,414,592 | 148,600.0 | 200,794.9 | 10,096 |
| | | | 0.0 | 125,712.0 | 7,438 |
| | | | 4.6 | 125,435.9 | 15,371 |
| | | | 4.5 | 100,678.7 | 18,482 |
| | | | 6.4 | 38,745.9 | |
| | | | 0.0 | 27,154.3 | 2,384 |
| | | | 8.7 | 41,461.2 | 7,578 |
| | | | 0.0 | 32,576.6 | 1,649 |
| | | | 6.6 | 26,873.9 | |
| | Germany | | | | |
| 10 | **Lassen** - IBM Power System S922LC, IBM POWER9 22C 3.1GHz, Dual-rail Mellanox EDR Infiniband, NVIDIA Tesla V100 , IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States | 288,288 | 18,200.0 | 23,047.2 | |



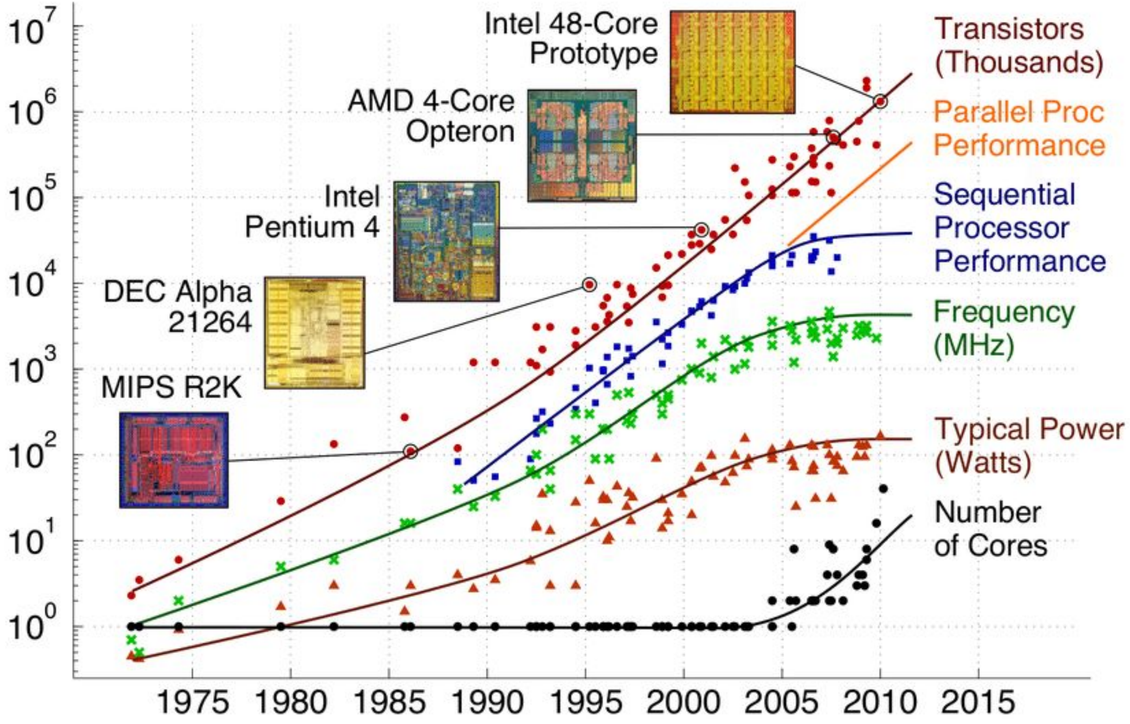
# Benefits of using GPUs - Another level of parallelism

**Little's Law: Arithmetic throughput = concurrency/latency**

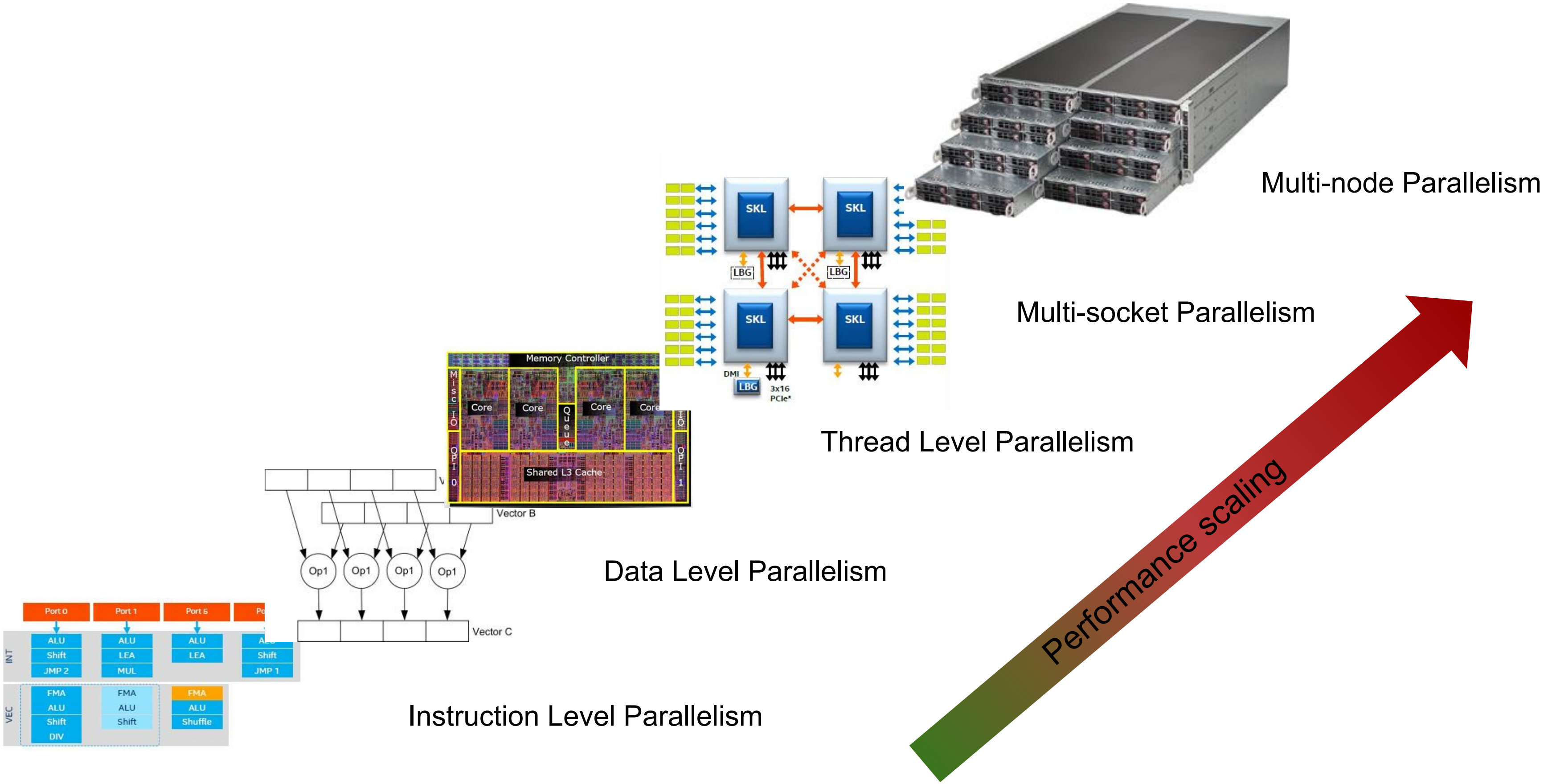
- Frequency scaling era **before 2005**

  (AKA free lunch era)

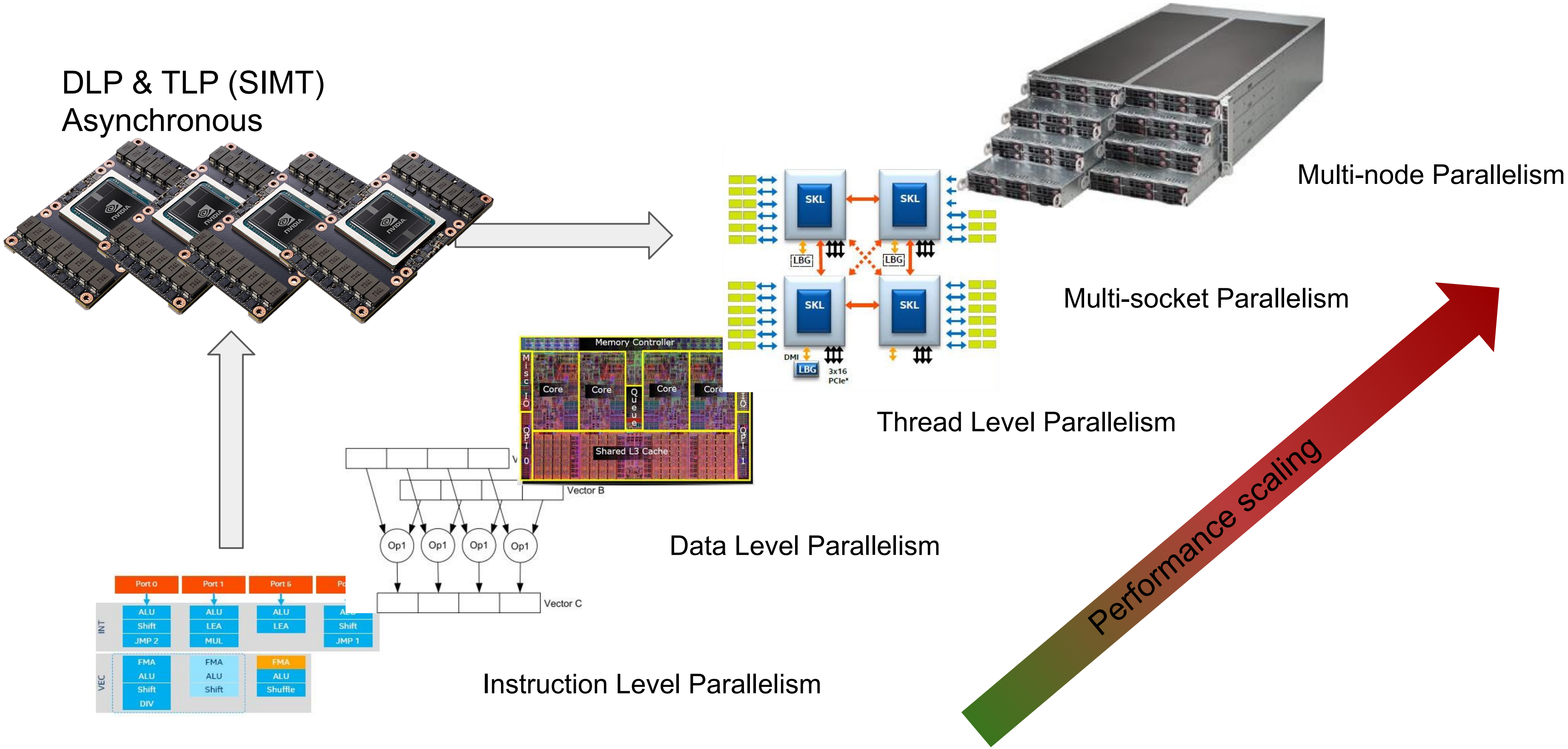
- **Since 2005**, performance increase is based on increasing parallelism



# Benefits of using GPUs - Another level of parallelism



# Benefits of using GPUs - Another level of parallelism



# Benefits of using GPUs - Another level of parallelism

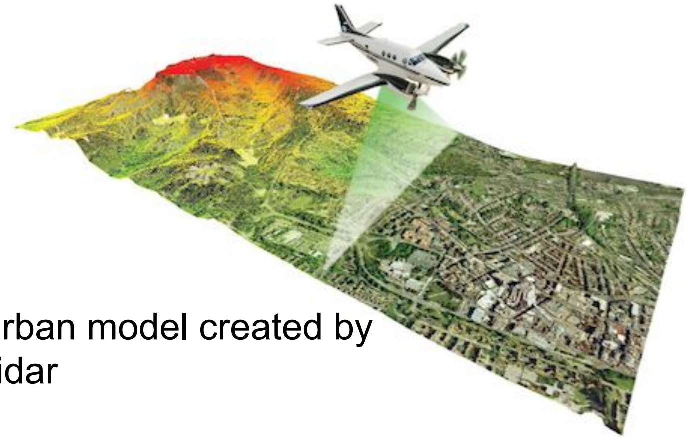
| | CPU (SKL) | GPU (V100) |
|---|---|---|
| FP64 perf | ~ 3 Tflops/s | ~ 7 Tflops |
| DRAM bandwidth | ~180 GB/s | ~750 GB/s |
| Memory | 100s of GB | 32 GB |
| Frequency | 2.0 ~ 4.0 Ghz | 1.3 ~ 1.45 Ghz |
| Characteristics | Low compute density<br>Low latency<br>Moderate throughput<br>Moderate concurrency<br>**Serial operations**<br>**Irregular patterns** | High compute density<br>High latency<br>High throughput<br>High concurrency<br>**Parallel operations**<br>**Regular patterns** |

# IceBound Project: computing rooftops global sunlight illumination

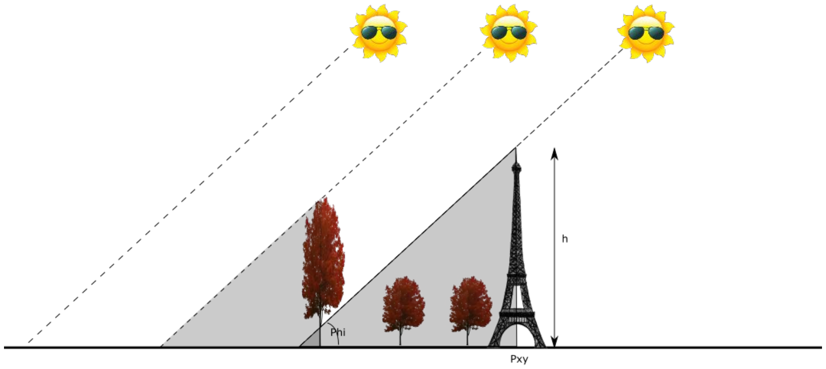
A time-lapse view of Madison Square Park. Chang W. Lee

EnhanceR | Enhancing Research through IT Expertise

# Shadow Model - Bresenham’s line algorithm

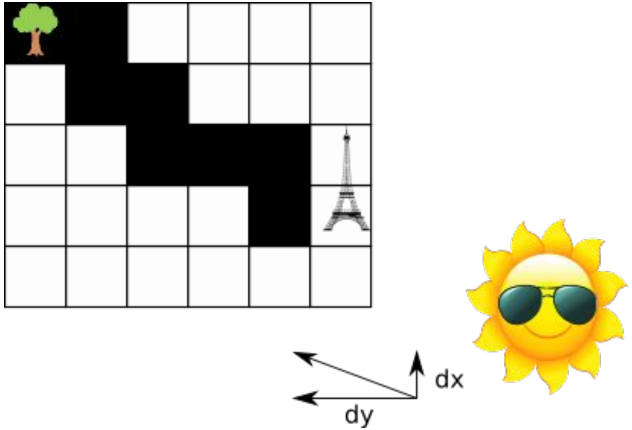


Urban model created by Lidar




- The direction of the light source beam is expressed as
- the light source beam vector [dx, dy, dz]
- The neighbour P' of a given point P at given step i is given by
    - P’x = Px - round[i*dx]
    - P’y = Py - round[i*dy]
- The height of P’ (h) is given by the urban model

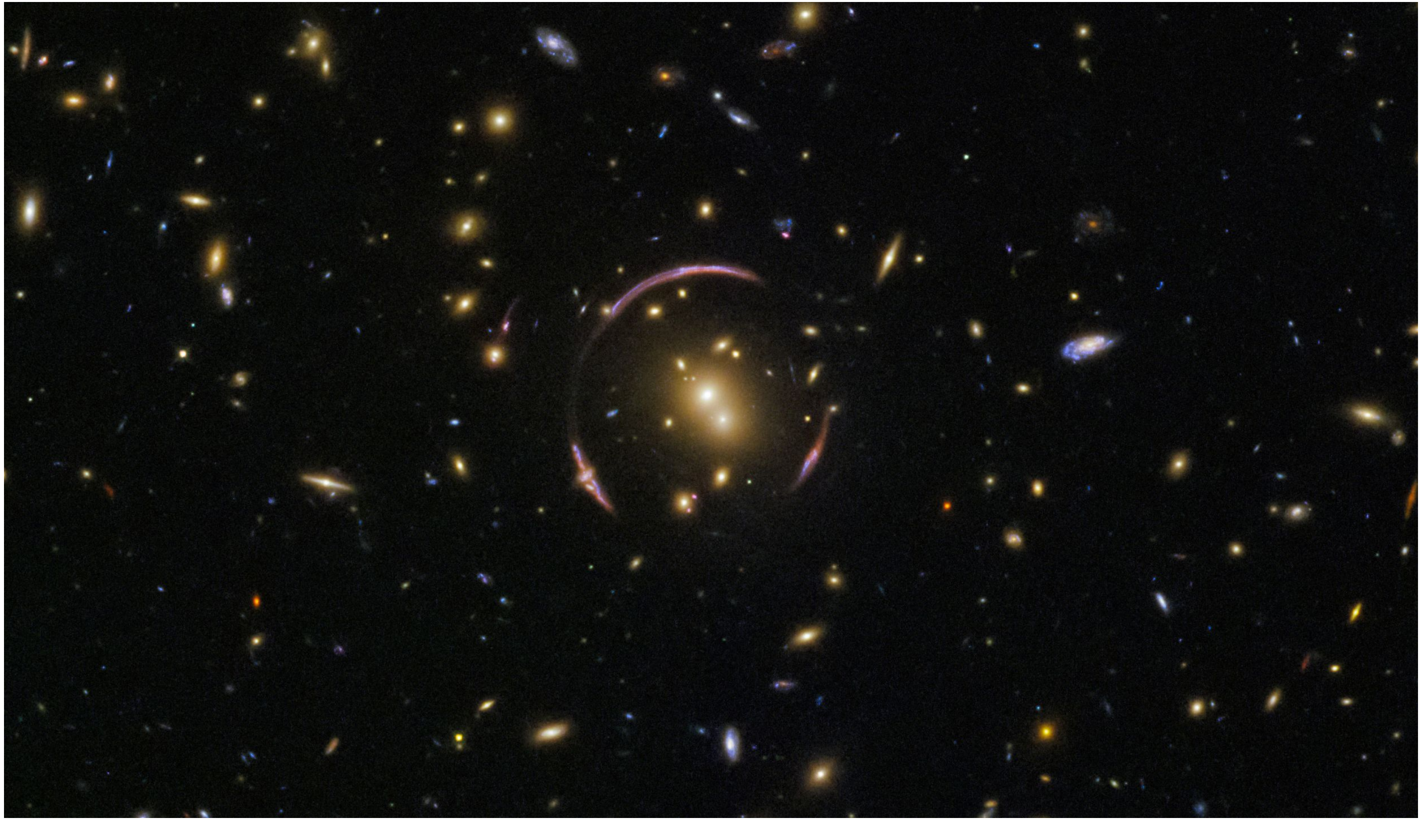
**P’ is shaded if: h < h0 - i*dz**



# Results - 5000x5000 map

| Platform | TTS (s) | Speedup |
|---|---|---|
| Broadwell (20 cores) | 10.8 | 1x |
| AWS g2.2xlarge (K520) | 2.4 (kernel) | 4.5x |
| Cuda K80 (deneb@SCITAS) | 0.64 (w/ data xfer)<br>0.5 (kernel) | 16.9x<br>21.6x |
| Cuda P100 (greina@CSCS) | 0.12 (w/ data xfer)<br>0.09 (kernel) | 90x<br>120x |

# Gravitational lensing computation code (LenstoolHPC)

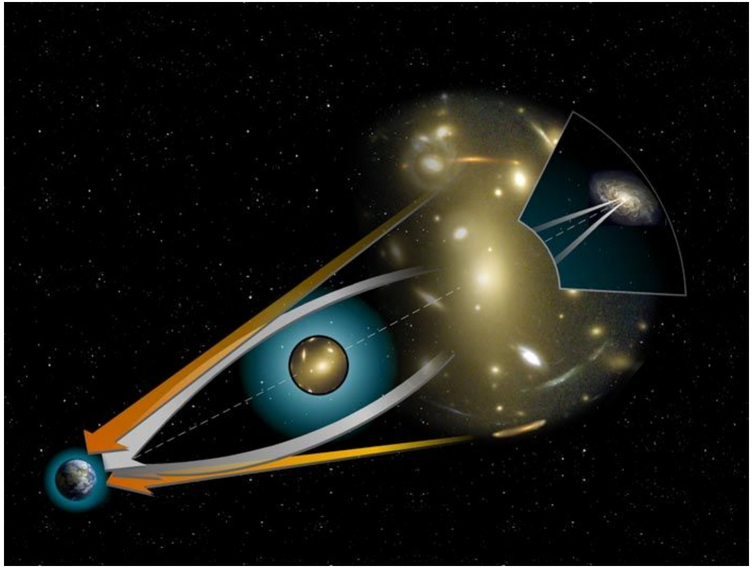


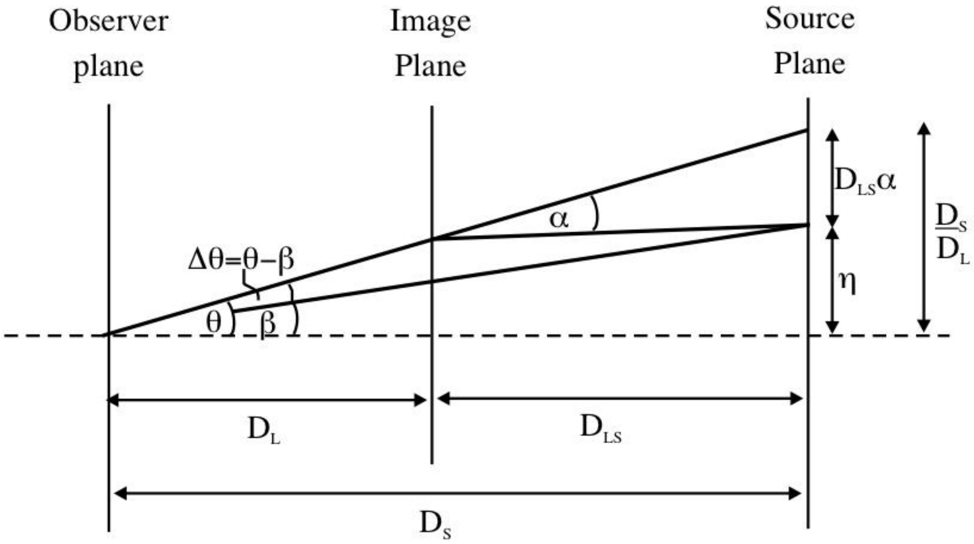
Einstein ring (credit: **Nasa/Hubble**)

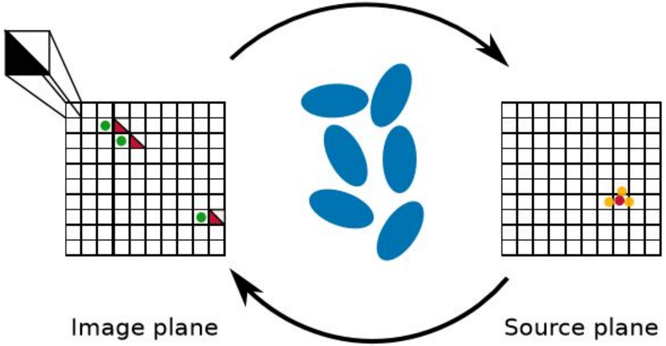
Collaboration with Christoph Schäfer

EPFL - LABORATOIRE D'ASTROPHYSIQUE

# Formalism






**2D lensing potential:**

$$\overrightarrow{\Delta\theta}(\vec{\theta}) = \frac{D_{LS}}{D_S}\vec{\alpha}(D_L\vec{\theta}) = \nabla\psi(\vec{\theta})$$

Example of gradient:

$$I^*_{\omega,1/2} = \frac{(1-\epsilon^2)\kappa_o\omega}{\imath\sqrt{\epsilon}}\ln\frac{\frac{1-\epsilon}{1+\epsilon}x - \imath\frac{1+\epsilon}{1-\epsilon}y + 2\imath\sqrt{\epsilon}\sqrt{\omega^2+\frac{x^2}{(1+\epsilon)^2}+\frac{y^2}{(1-\epsilon)^2}}}{x-\imath y+2\imath\omega\sqrt{\epsilon}}.$$

# Strong Lensing Algorithm

Given **a parametric model** for all the lens types

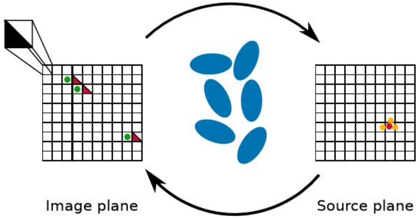
Step 0: Compute all the gradients
- **High compute intensity**
- **Regular**

Gradient Computation

$$I^*_{\omega,1/2} = \frac{(1-\epsilon^2)\kappa_o\omega}{\imath\sqrt{\epsilon}} \ln \frac{\frac{1-\epsilon}{1+\epsilon}x - \imath\frac{1+\epsilon}{1-\epsilon}y + 2\imath\sqrt{\epsilon}\sqrt{\omega^2 + \frac{x^2}{(1+\epsilon)^2} + \frac{y^2}{(1-\epsilon)^2}}}{x - \imath y + 2\imath\omega\sqrt{\epsilon}}.$$

Step 1a: Computing sources
- lensing the images to the source plane
- **Low compute intensity, regular**
- **independent of Step 0**

Computing sources



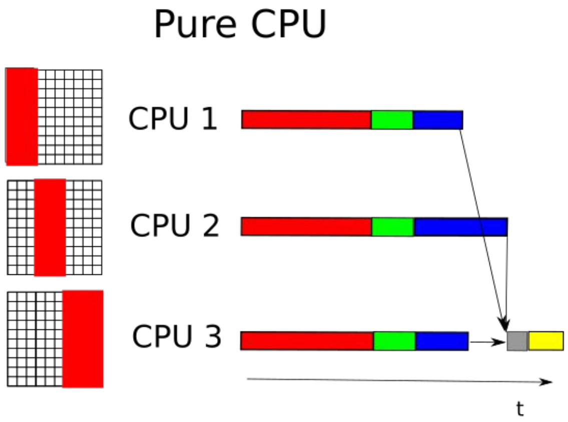

Step 1b: Searching images, needs Step 0 and 1a
- Lense image plane grid into source plate, **regular**
- Checking source plane grid for sources, **irregular**
- **Moderate compute intensity**

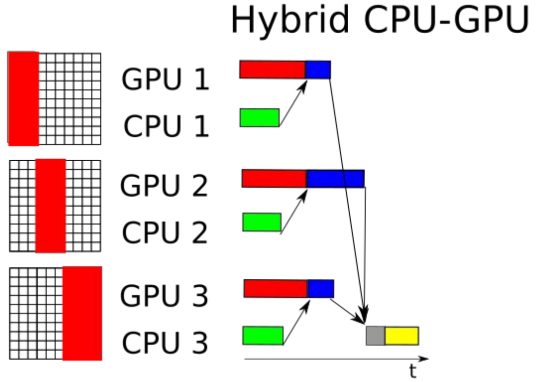
Searching for images

Step 2: reduce and compute Chi^2
- **Low compute intensity**
- **Irregular (MPI)**

Communication Chi2 computation

$$\chi^2 = \sum_i \sum_j \frac{(x_{\mathrm{obs},ij} - x_{ij})^2}{\sigma_{ij}^2}$$



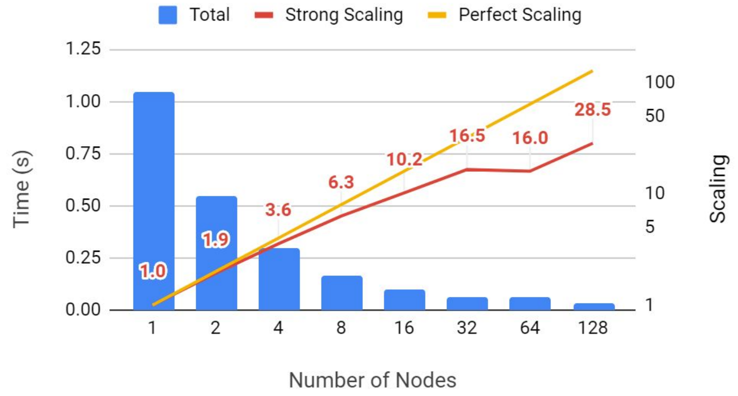
# Performance and Strong Scaling, Piz Daint@CSCS

m1931_modif_217Pot benchmark, 5kx5k image

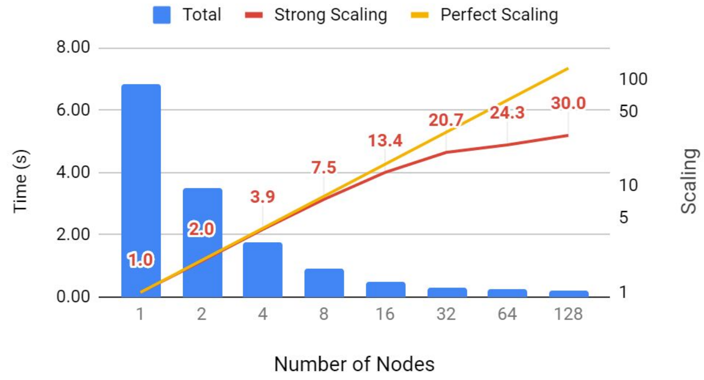
| Number of Nodes | Total | Strong Scaling |
|---|---|---|
| 1 | 1.05 | 1.0 |
| 2 | 0.54 | 1.9 |
| 4 | 0.29 | 3.6 |
| 8 | 0.17 | 6.3 |
| 16 | 0.10 | 10.2 |
| 32 | 0.06 | 16.5 |
| 64 | 0.07 | 16.0 |
| 128 | 0.04 | 28.5 |

| Number of Nodes | Total | Strong Scaling |
|---|---|---|
| 1 | 6.82 | 1.0 |
| 2 | 3.48 | 2.0 |
| 4 | 1.75 | 3.9 |
| 8 | 0.91 | 7.5 |
| 16 | 0.51 | 13.4 |
| 32 | 0.33 | 20.7 |
| 64 | 0.28 | 24.3 |
| 128 | 0.23 | 30.0 |

Piz Daint GPU - Total Time to Solution and Strong Scaling




Piz Daint CPU - Total Time to Solution



# Take home message

- CPUs and GPUs can work very well together
  - CPUs are designed for single thread performance and irregular patterns
  - GPUs are designed for parallelism and regular patterns
- Harnessing both is the key to exascale...