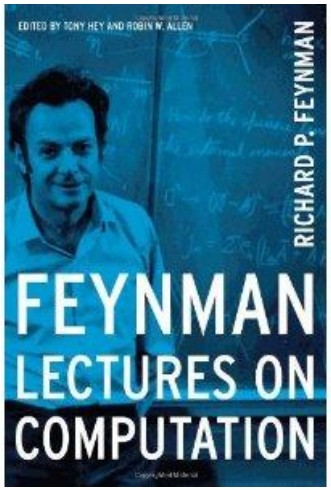
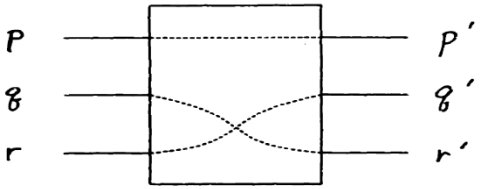
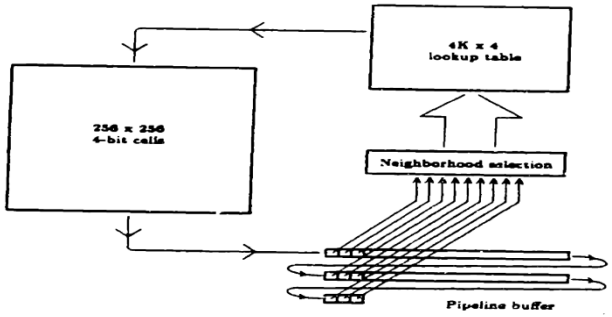
Dataflow

Some Dataflow History





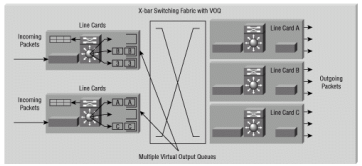
Toffoli=>Fredkin Gate



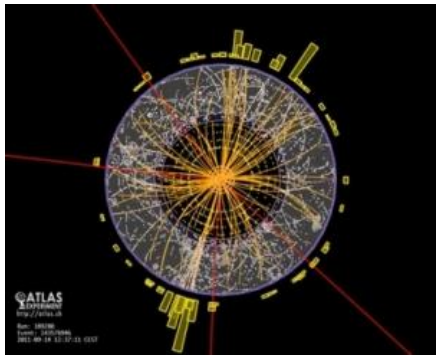
CAM architecture

PhD Thesis of Norman Margolus, MIT 1987

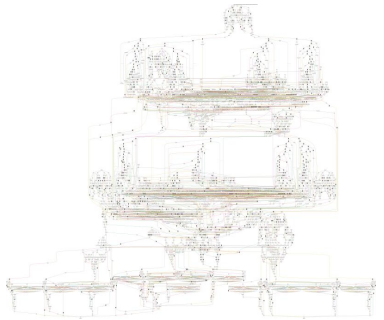
Today:



CISCO Internet Switch



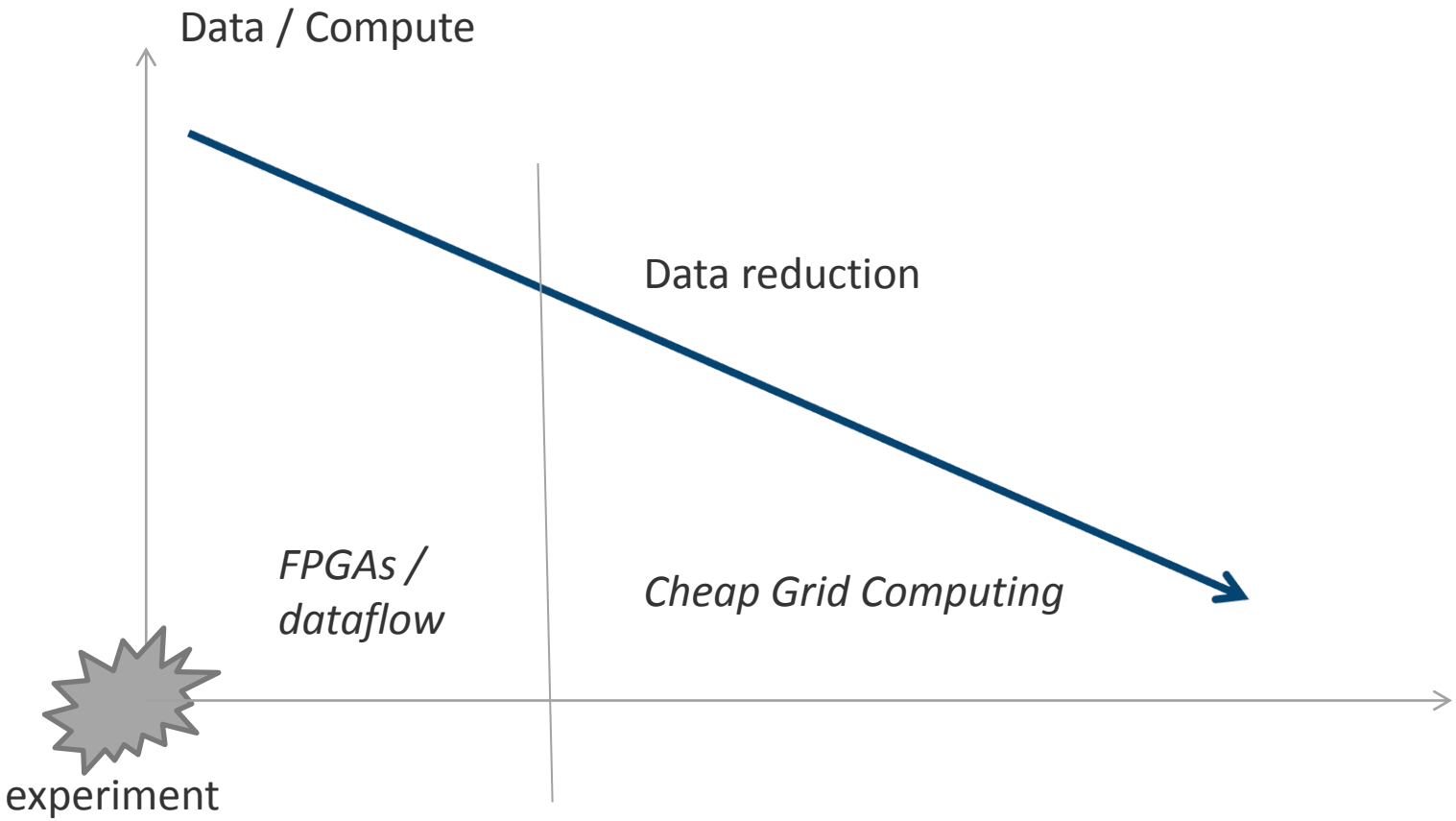
Data reduction in Physics



Maxeler Dataflow Engines

MAXELER Technologies

CERN Large Hadron Collider



“Is the LHC throwing away too much data?” New Scientist 2012.

Future of Medicine...

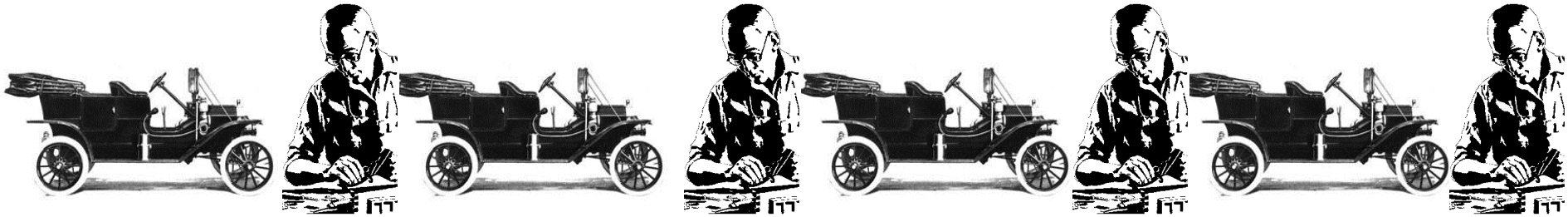
- Cancer: multiple genomes per tissue
- Spain: 28PB for 1 genome per cancer

- Max Planck Institute: 10^18 core-hours for Europe
- Cancer, Diabetes, etc...

- Healthcare requires complete medical records
- Images, genomes, transcriptome, proteom, ...

Explaining Control Flow versus Data Flow

The Ford Production Line



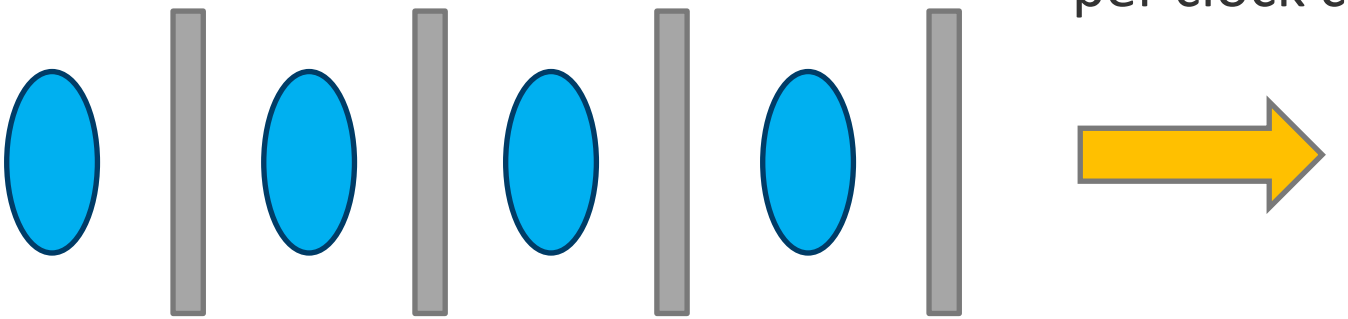
- Experts are expensive and slow (control flow)
- Many specialized workers are more efficient (data flow)

Technology

resulting from our research at Stanford, Bell Labs, and Imperial College

MAXELER DATAFLOW COMPUTING

One result
per clock cycle



Dynamic (switching) Power Consumption:

$$P_{avg} = C_{load} \cdot V_{DD}^{\ 2} \cdot f$$

Minimal frequency f achieves maximal performance, thus for a given power budget, we get Maximum Performance Computing (MPC)!

Maxeler Architecture Series 2012

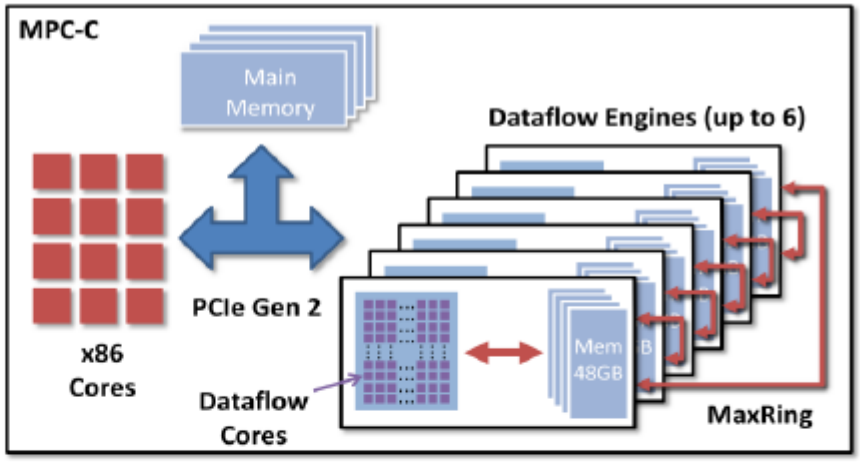


Figure 3. MPC-C Series Architecture

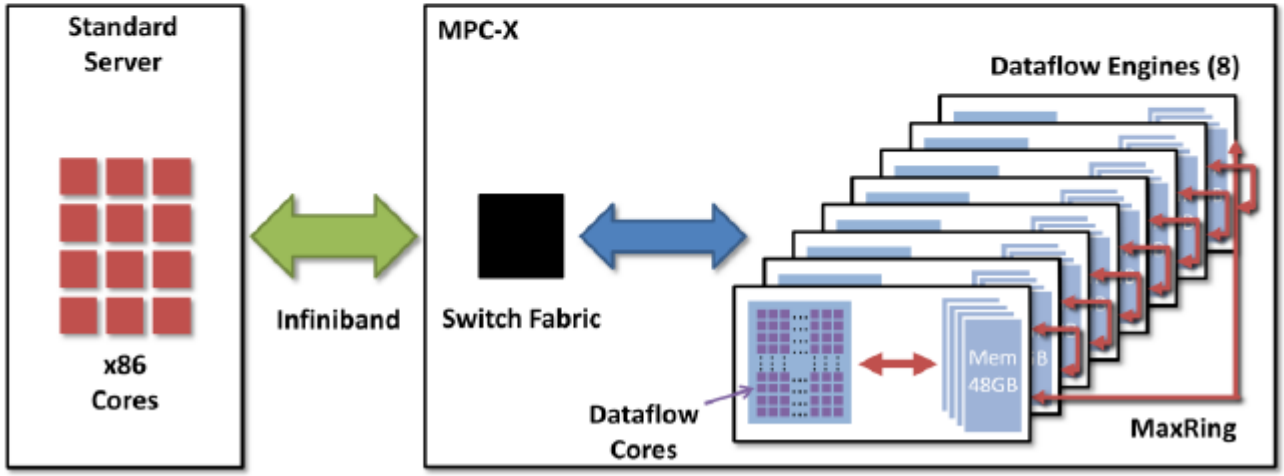
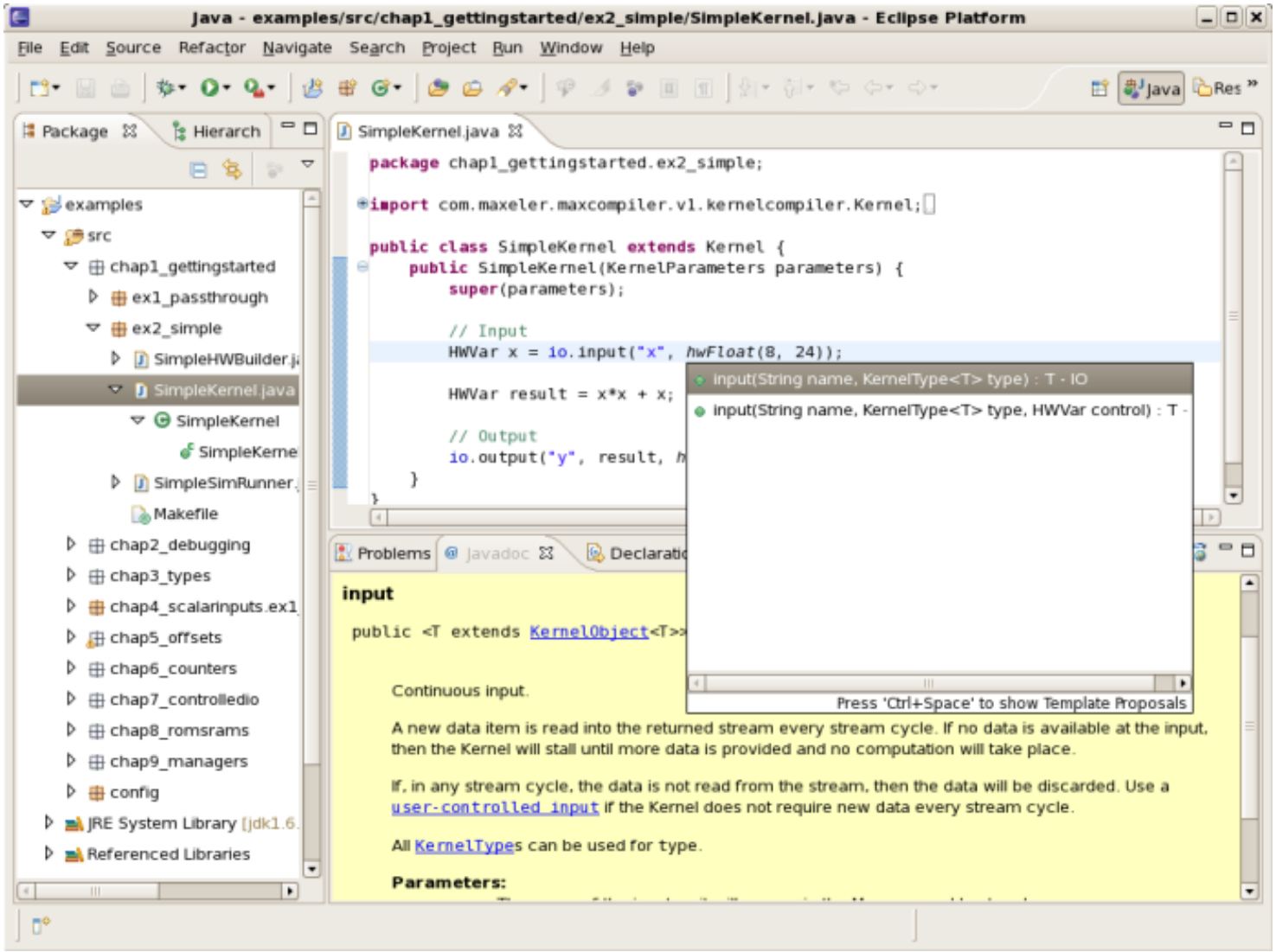


Figure 4. MPC-X Series Architecture

MAXELER Technologies

Maxeler Programming Environment in Java



MAXELER Technologies

Cox-Ingersol (CIR) Interest Rate Model

$$dr(t) = \alpha(b - r(t))dt + \sigma\sqrt{r(t)}dW(t)$$

$$r(t_{i+1}) = r(t_i) + \alpha(b(t_i) - r(t_i))[t_{i+1} - t_i] + \sigma\sqrt{r(t_i)^+}\sqrt{t_{i+1} - t_i}Z_{i+1}$$

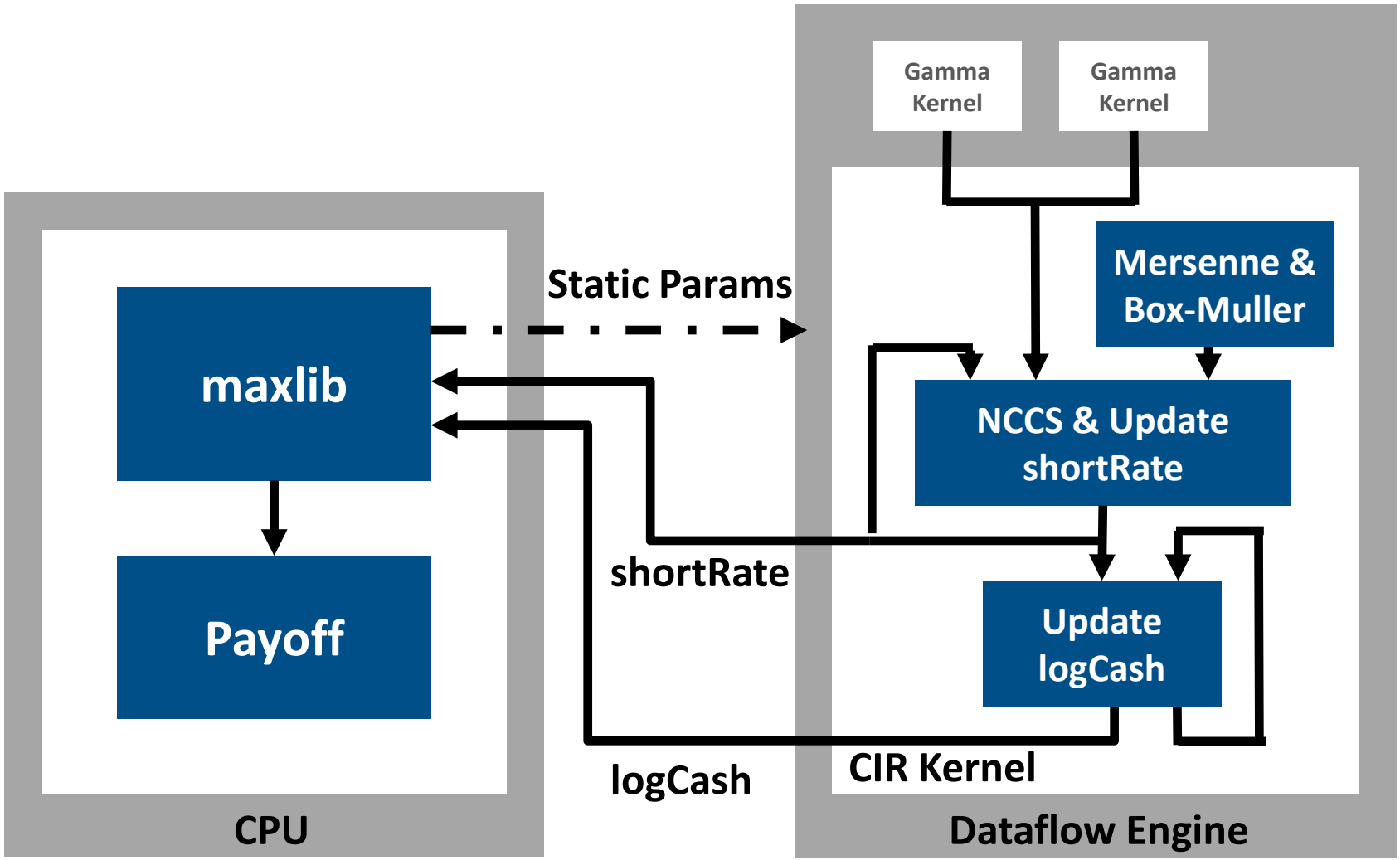
$$r(t_{i+1}) = \frac{\sigma^2(1-e^{-\alpha(t_{i+1}-t_i)})}{4\alpha}\chi'^2_d\left(\frac{4\alpha e^{-\alpha(t_{i+1}-t_i)}}{\sigma^2(1-e^{-\alpha(t_{i+1}-t_i)})}r(t_i)\right),\ d = 4\bar{b}\alpha/\sigma^2$$

$$B(t,T) = e^{-A(t,T)r(t)+C(t,T)}$$

$$A(t,T) = \frac{2(e^{\gamma(T-t)}-1)}{(\gamma+\alpha)(e^{\gamma(T-t)}-1)+2\gamma}$$

$$C(t,T) = \frac{2\alpha b}{\sigma^2}\log\left(\frac{2\gamma e^{(\alpha+\gamma)(T-t)/2}}{(\gamma+\alpha)(e^{\gamma(T-t)}-1)+2\gamma}\right)$$

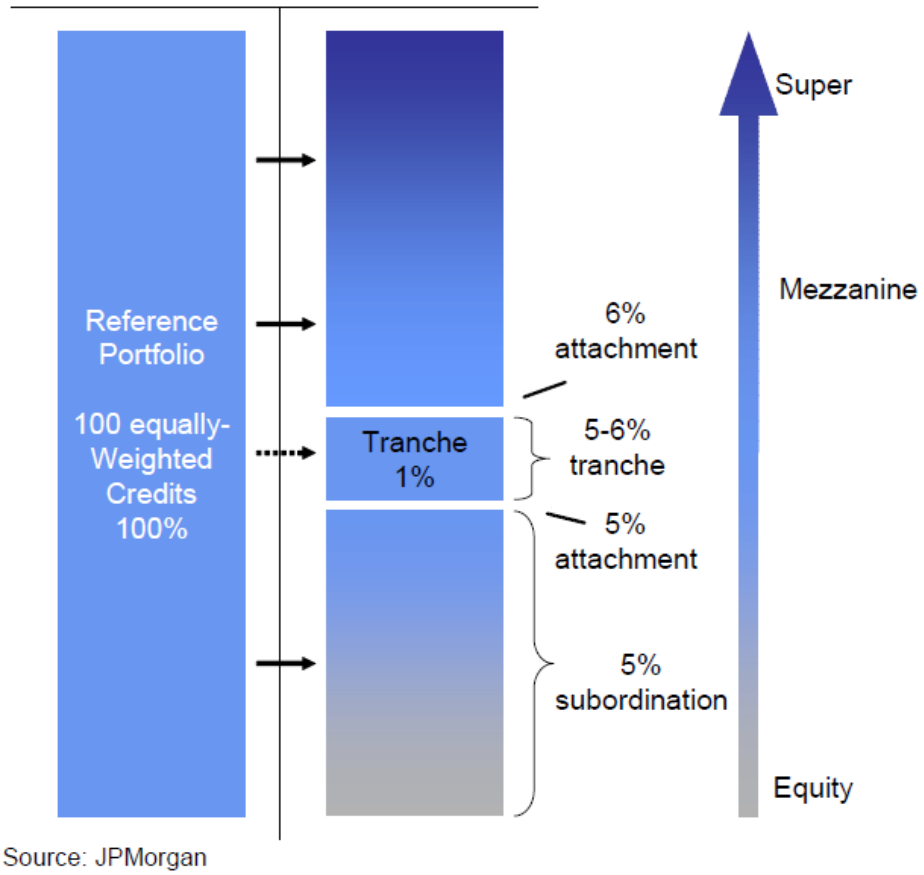
CIR Monte Carlo Dataflow Architecture



JP Morgan Credit Derivatives Pricing

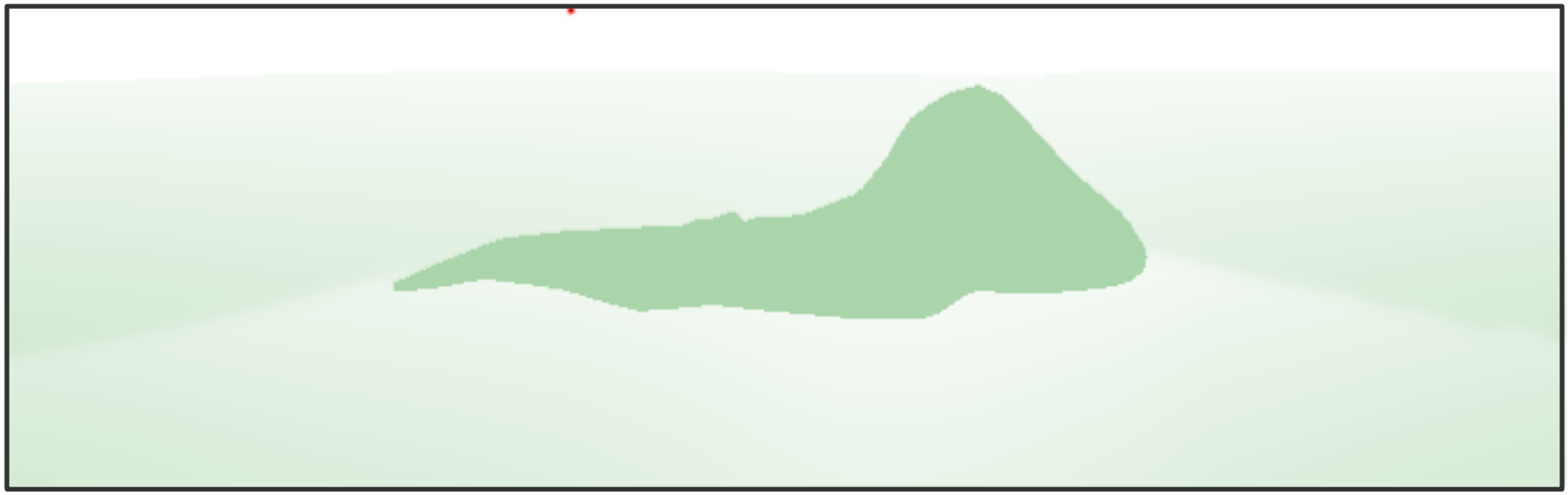
O. Mencer and S. Weston, 2010

- Compute value of complex financial derivatives (CDOs)
- Typically run overnight, but beneficial to compute in real-time
- Many independent jobs
- **Speedup: 220-270x**
- Power consumption per node drops from 250W to 235W/node



Source: JPMorgan

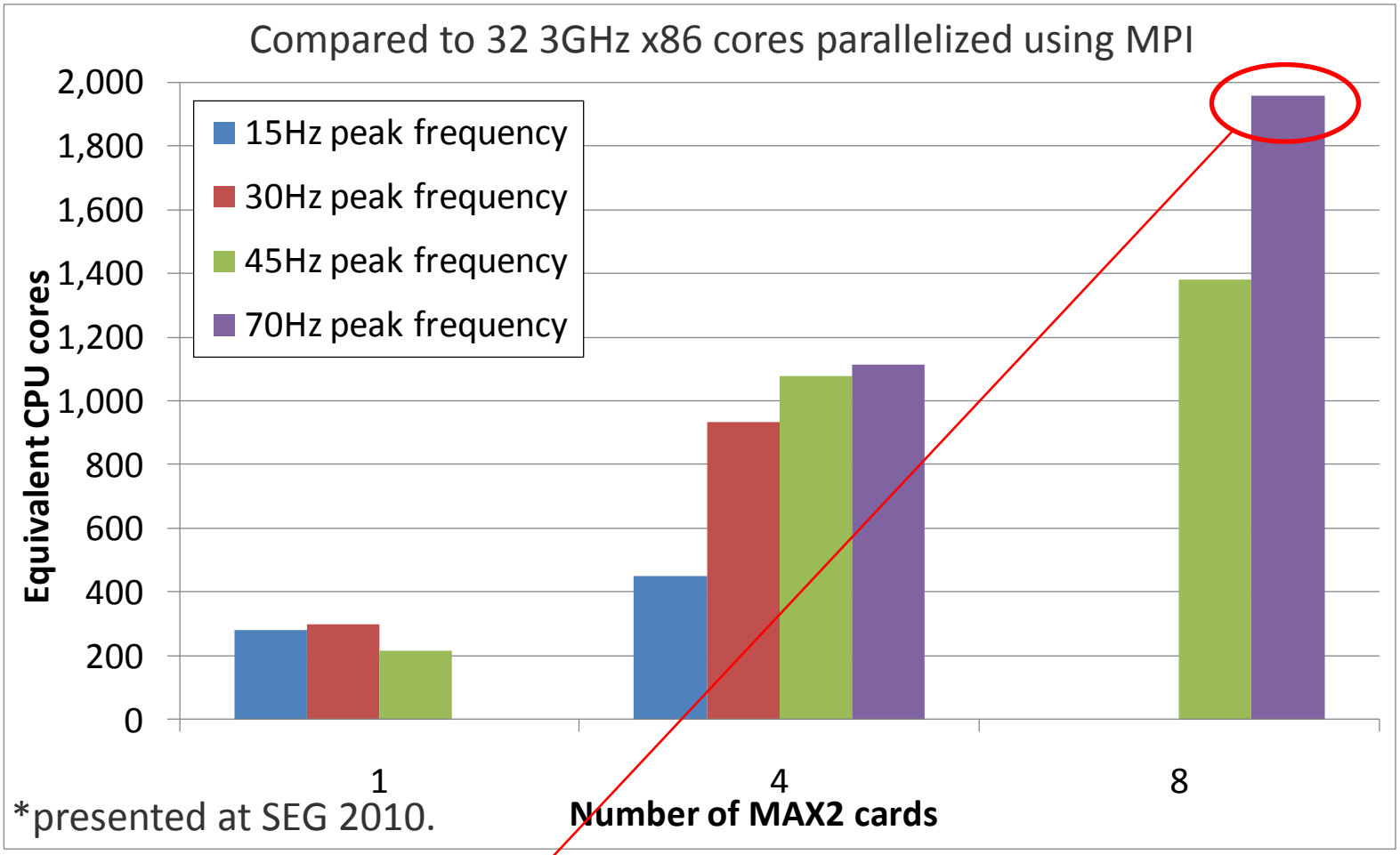
Running with SEG Salt Model



- Running on MaxNode servers
  - 8 parallel compute pipelines per chip
  - 150MHz => low power consumption!
  - 30x faster than microprocessors

Beyond Traditional Microprocessors for Geoscience High-Performance Computing Applications O. Lindtjorn[†], R. G. Clapp[†], O. Pell[‡], O. Mencer[‡], M. J. Flynn[‡] and H. Fu[§]. [†]Stanford, [‡]Maxeler, [§]Tsinghua, IEEE Micro, March 2011.

3000^3 Modeling

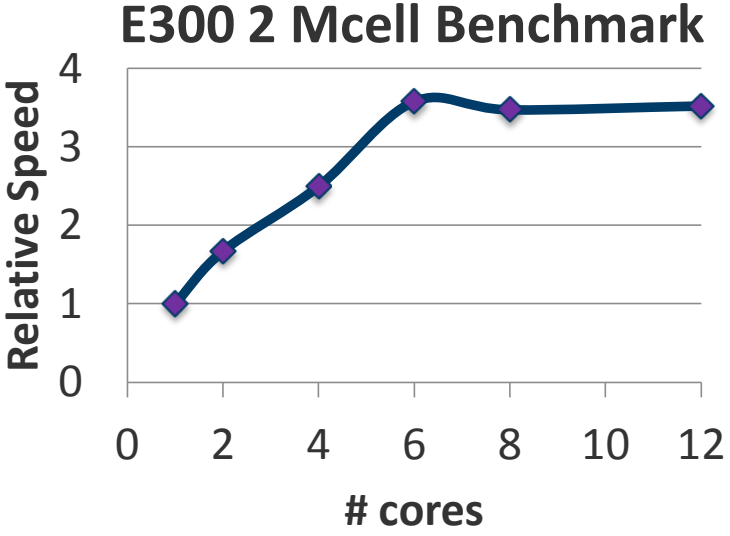


8 Full Intel Racks ~100kWatts => 2 MaxNodes (2U) Maxeler System <1kWatt

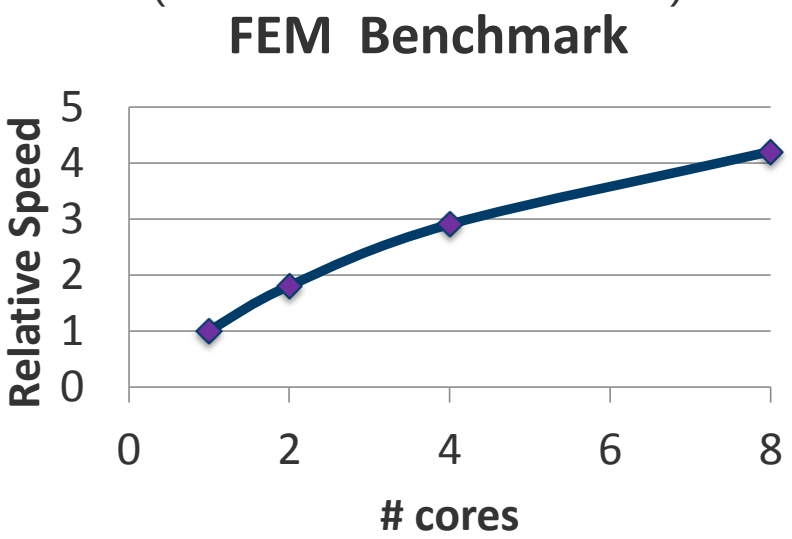
MAXELER Technologies

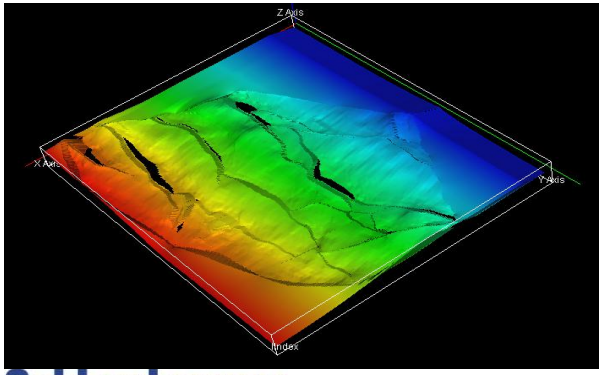
Typical Scalability of Sparse Matrix

Eclipse Benchmark
(2 node Westmere 3.06 GHz)

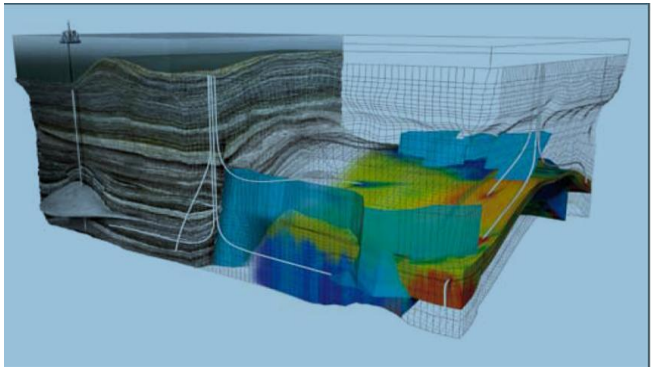


Visage – Geomechanics
(2 node Nehalem 2.93 GHz)





Schlumberger



MAXELER Technologies

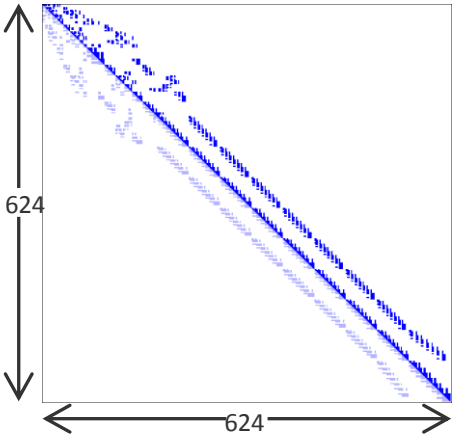
Sparse Matrix Solving

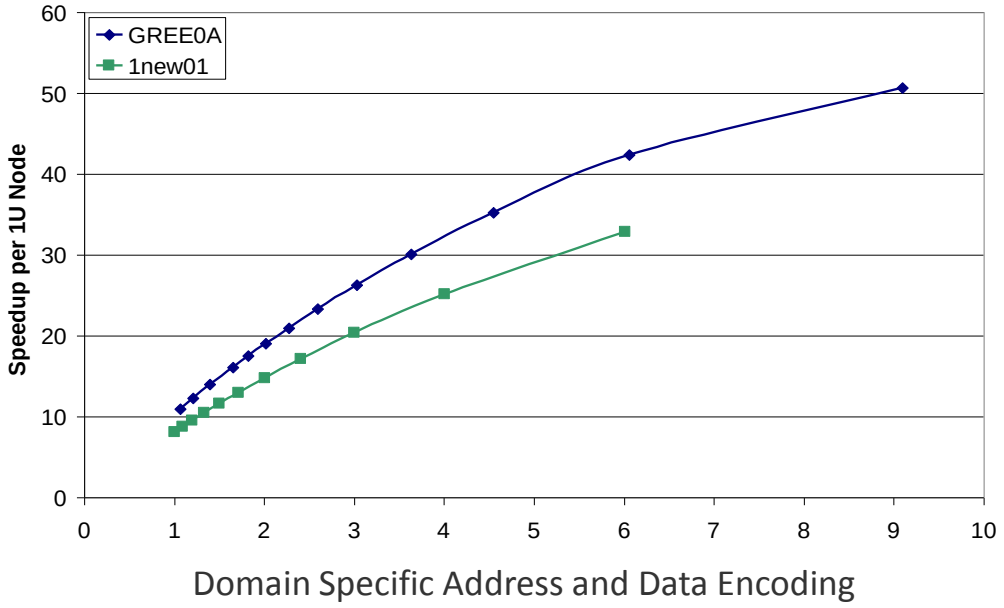
O. Lindtjorn et al, 2010

- Given matrix **A**, vector *b*, find vector *x* in:

  $\mathbf{A}x = b$

- Typically memory bound, not parallelisable.
- 1 MaxNode achieved 20-40x the performance of an Intel node.





MAXELER Technologies

Economics of Computation (TCO)

Example Solution 1

- **50x** Speed-up per 1U server node
- **32** Maxeler Node Solution
- Equivalent to **1600** CPU-only Nodes
- **$3.2m** Operational cost savings over 3 years

Example Solution 2

- **30x** Speed-up per 1U server node
- **40** Maxeler Node Solution
- Equivalent to **1200** CPU-only Nodes
- **$1.8m** Operational cost savings over 3 years

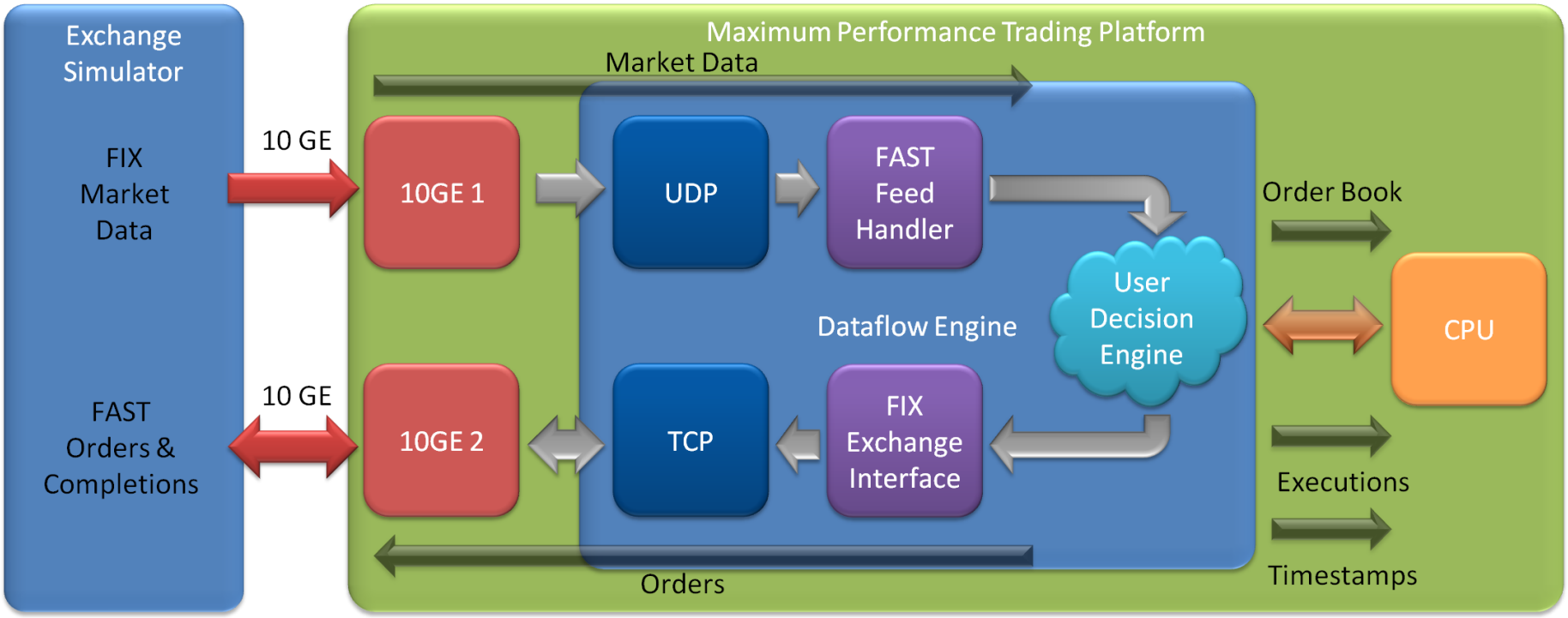
Example Solution 3

- **20x** Speed-up per 1U server node
- **50** Maxeler Node Solution
- Equivalent to **1000** CPU-only Nodes
- **$1.7m** Operational cost savings over 3 years

Example Solution 4

- **40x** Speed-up per 1U server node
- **32** Maxeler Node Solution
- Equivalent to **1280** CPU-only Nodes
- **$2.6m** Operational cost savings over 3 years

Trading: Direct Market Access and HFT



Maxeler Hardware Solutions 2012

MPC C Series

CPUs plus DFEs

Intel Xeon CPU cores and up to 6 DFEs with 288GB of RAM

MPC X Series

DFEs shared over Infiniband

Up to 8 DFEs with 384GB of RAM and dynamic allocation of DFEs to CPU servers

MPC N Series

Low latency connectivity

Intel Xeon CPUs and 1-2 DFEs with up to six 10Gbit Ethernet connections

MaxWorkstation

Desktop development system

MaxCloud

On-demand scalable accelerated compute resource, hosted in London

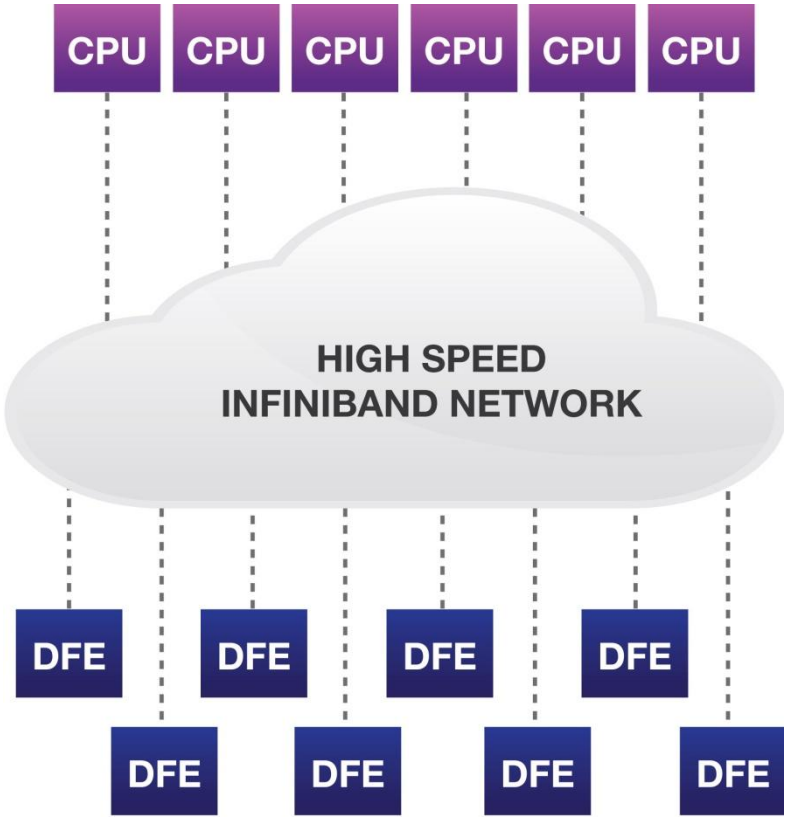
MAXELER Technologies

MPC X Series

1U dataflow cloud providing dynamically scalable compute capability over Infiniband

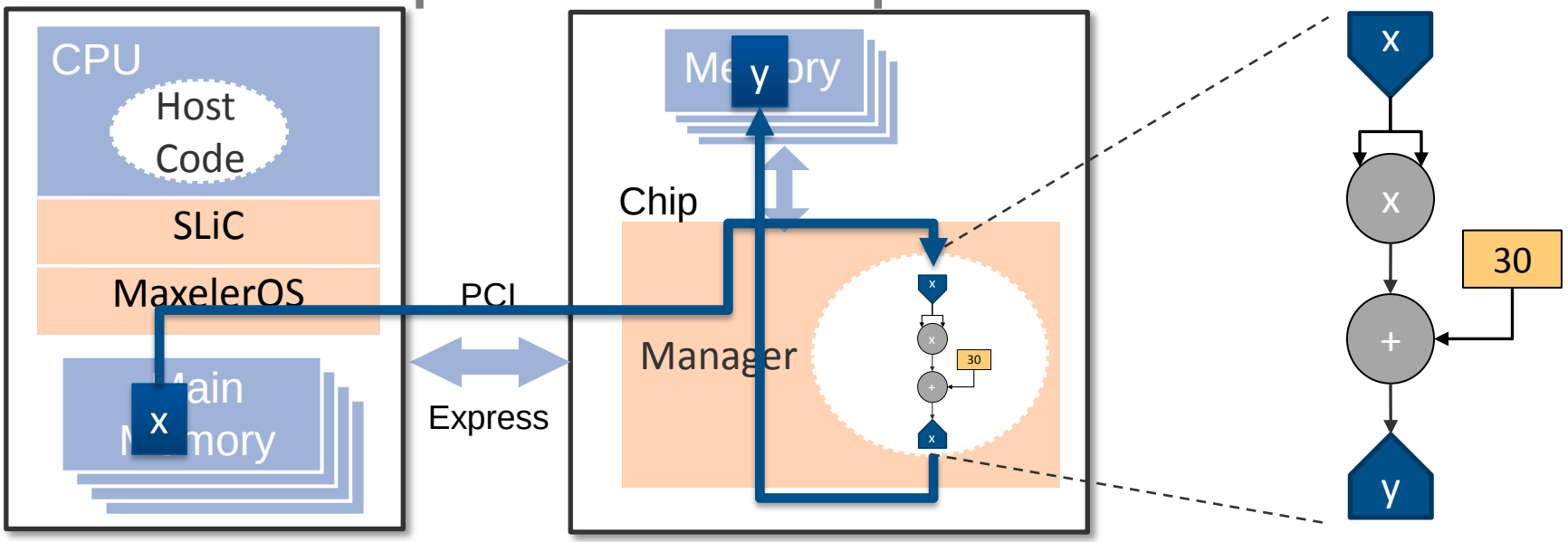
MPC-X1000

- 8 *vectis* dataflow engines (DFEs)
- 192GB of DFE RAM
- Dynamic allocation of DFEs to conventional CPU servers
  - Zero-copy RDMA between CPUs and DFEs over Infiniband
- Equivalent performance to 40-60 x86 servers



MAXELER Technologies

MaxCompiler Development Process



CPUCode (.c)

```
#include “MaxSLiCInterface.h”
#include “Calc.max”
int *x, *y;


Calc(x, DATA_SIZE)
```

Manager (.java)

```
Manager m = new Manager();
Kernel k =
    new MyKernel();

m.setKernel(k);
m.setIO(
   link(“x", PCIE),
   link(“y", DRAM_LINEAR1D));
m.addMode(modeDefault());
m.build();
```

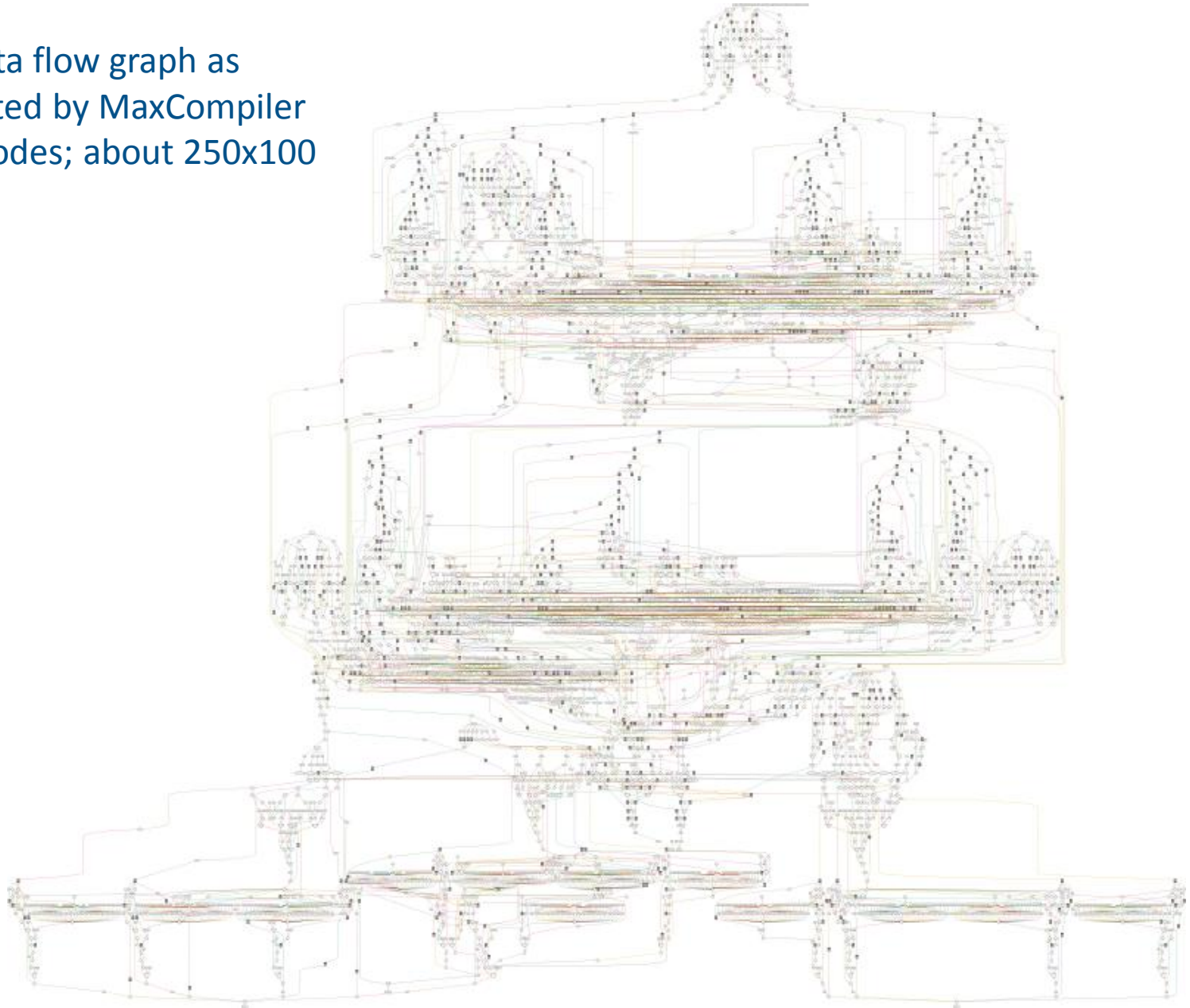
MyKernel (.java)

```
HWVar x = io.input("x", hwInt(32));

HWVar result = x * x + 30;

io.output("y", result, hwInt(32));
```

Data flow graph as generated by MaxCompiler
4866 nodes; about 250x100



Chip Resource Usage

The goal is to maximize utilization of resources on the chip, and bandwidth on the memory bus.

```
  LUTs      FFs    BRAMs     DSPs : MyKernel.java
   727      871      1.0        2 : resources used by this file
 0.24%    0.15%    0.09%    0.10% : % of available
71.41%   61.82%  100.00%  100.00% : % of total used
94.29%   97.21%  100.00%  100.00% : % of user resources
                                  :
                                  : public class MyKernel extends Kernel {
                                  :   public MyKernel (KernelParameters parameters) {
                                  :     super(parameters);
     1       31      0.0        0 :     HWVar p = io.input("p", hwFloat(8,24));
     2        9      0.0        0 :     HWVar q = io.input("q", hwUInt(8));
                                  :     HWVar offset = io.scalarInput("offset", hwUInt(8));
     8        8      0.0        0 :     HWVar addr = offset + q;
    18       40      1.0        0 :     HWVar v = mem.romMapped("table", addr,
                                  :                             hwFloat(8,24), 256);
   139      145      0.0        2 :     p = p * p;
   401      541      0.0        0 :     p = p + v;
                                  :     io.output("r", p, hwFloat(8,24));
                                  :   }
                                  : }
```

Measuring Utilization

- *Top* measures % of time CPU is running
- *Maxtop* monitors % of time the DFE is running

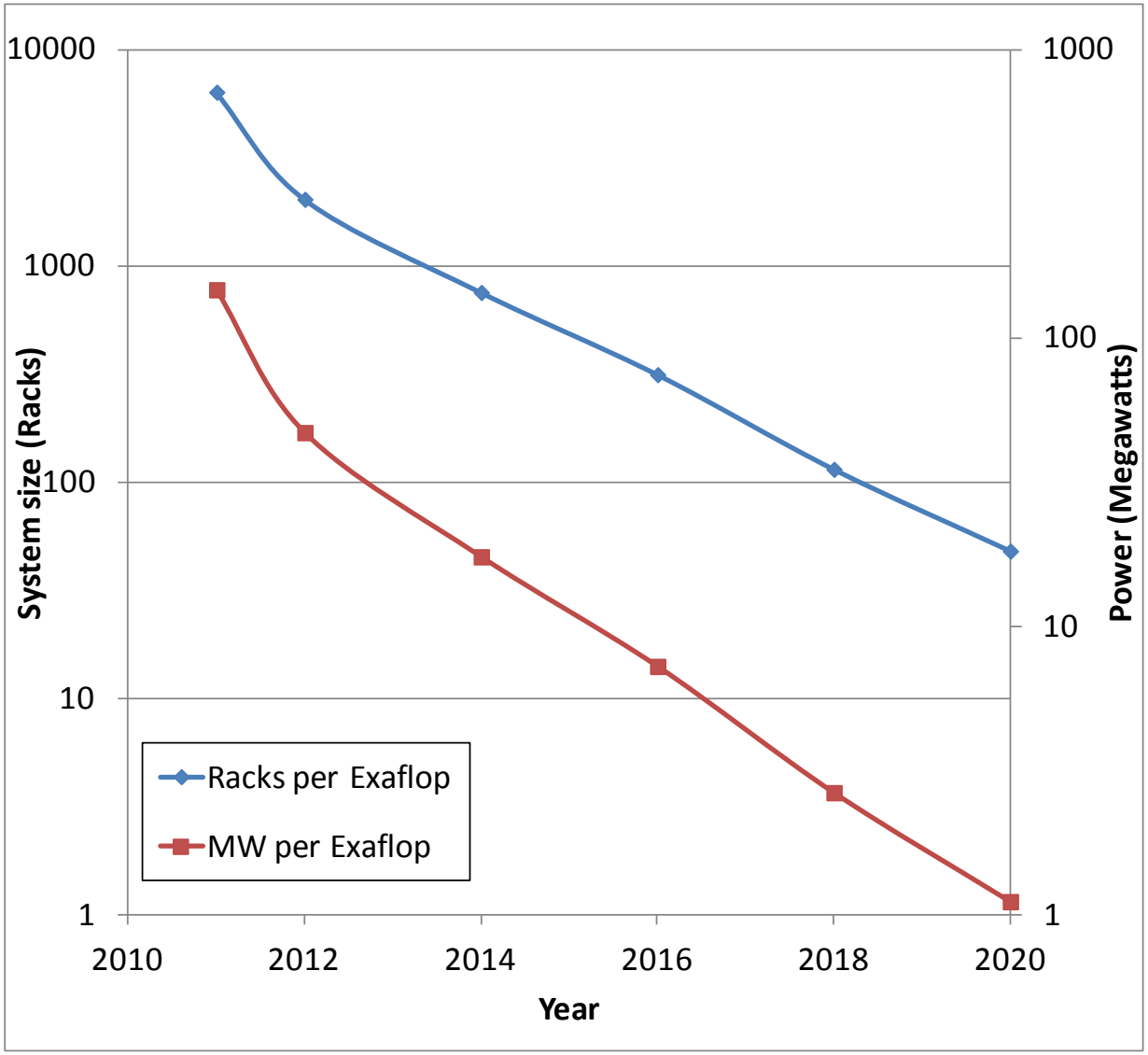
```
MaxTop Tool 2011.2
Found 2 Maxeler card(s) running MaxelerOS 2011.2
Card 0: MAX3A (P/N: 13424) S/N: 219270088 Mem: 24GB DFE(s): 1 /dev/maxeler0
Card 1: MAX3A (P/N: 13424) S/N: 000025559 Mem: 24GB DFE(s): 1 /dev/maxeler1

DEVICE     %DFE    TEMP   BITSTREAM      PID    USER        TIME       COMMAND
maxeler0   66.6%   57.1C  9d9de1...      12333  jspooner    00:00:39   model
maxeler1   0.0%    54.6C  9d9de1...      -      -           -          -
```

The Exascale Supercomputer (2018)

- 1 exaflop = 10^{18} FLOPS
- Using processor cores with 8FLOPS/clock at 2.5GHz
- **50M CPU cores**
- What about power?
  - Assume power envelope of 100W per chip
  - Moore's Law scaling: 6 cores today → ~100 cores/chip
  - 500k CPU chips
- **50MW** (*just for CPUs!)* → **100MW likely**
- 'Jaguar' power consumption: 6MW for 2 Petaflops

Exascale Dataflow Computing



- Exascale application performance by 2018
  - 100 racks
  - <3MW power
- A custom silicon solution could improve this by 5x
  - 20 racks
  - <1MW power

MAXELER Technologies

Recent Scientific Publications with Clients

- **Rapid Computation of Value and Risk for Derivatives Portfolios**

  S. Weston[†], J. Spooner[‡], S. Racanière[‡] and O. Mencer[‡§].
  [†]JPMorgan, [‡]Maxeler Technologies, [§]Imperial College London

  Concurrency and Computation: Practice and Experience, Special Issue Paper, July 2007, doi: 10.1002/cpe.1778.

- **Beyond Traditional Microprocessors for Geoscience High-Performance Computing Applications**

  O. Lindtjorn[†], R. G. Clapp[†], O. Pell[‡], O. Mencer[‡], M. J. Flynn[‡] and H. Fu[§].
  [†]Stanford University and Schlumberger, [‡]Maxeler Technologies, [§]Tsinghua University

  IEEE Micro, vol. 31, no. 2, March/April 2011.

- **Accelerating the Computation of Portfolios of Tranched Credit Derivatives**

  S. Weston[†], J-T. Marin[†], J. Spooner[‡], O. Pell[‡] and O. Mencer[‡].
  [†]JPMorgan, [‡]Maxeler Technologies

  IEEE Workshop on High Performance Computational Finance, New Orleans, USA, November 2010.

- **FD modeling beyond 70Hz with FPGA acceleration**

  D. Oriato[†], O. Pell[†], C. Andreoletti[‡] and N. Bienati[‡]. [†]Maxeler Technologies, [‡]Eni E&P Division

  SEG 2010 HPC Workshop, Denver, USA, October 2010.

- **Surviving the End of Scaling of Traditional Microprocessors in HPC**

  O. Lindtjorn[†‡], R. G. Clapp[‡], O. Pell[§], O. Mencer[§] and M. J. Flynn[§]. [†]Schlumberger, [‡]Stanford University, [§]Maxeler Technologies

  IEEE HOT CHIPS 22, Stanford, USA, August 2010.

- **Fast 3D ZO CRS Stack - An FPGA Implementation of an Optimization Based on the Simultaneous Estimate of Eight Parameters**

  P. Marchetti[†], D. Oriato[‡], O. Pell[‡], A.M. Cristini[§] and D. Theis[§].
  [†]Eni E&P Division, [‡]Maxeler Technologies, [§]CRS4

  72nd European Association of Geoscientists and Engineers (EAGE) Conference, Barcelona, June 2010.

- **Computational acceleration of credit and interest rate derivatives**

  O. Mencer[†] and S. Weston[‡]. [†]Maxeler Technologies, [‡]JPMorgan

  Global Derivatives Trading and Risk Management, Paris, May 2010.

MAX-UP: Maxeler University Program

- Founding Members: Imperial, Tsinghua, U Tokyo and Stanford
- MaxAcademy course material for teaching
- MaxTraining at Maxeler: learn dataflow programming
- Internships and collaboration
- Joint research grant applications
- Demos

Imperial College London

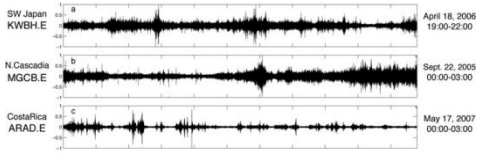
Tsinghua University

THE UNIVERSITY OF TOKYO

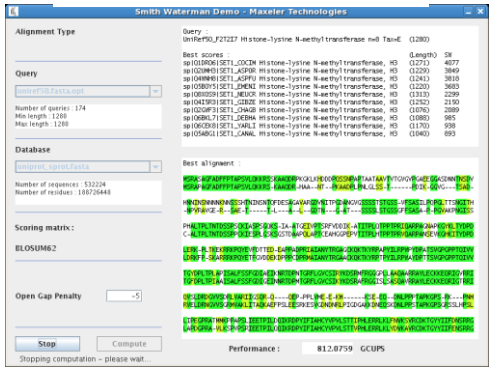
LELAND STANFORD JUNIOR UNIVERSITY

MaxWorkstation

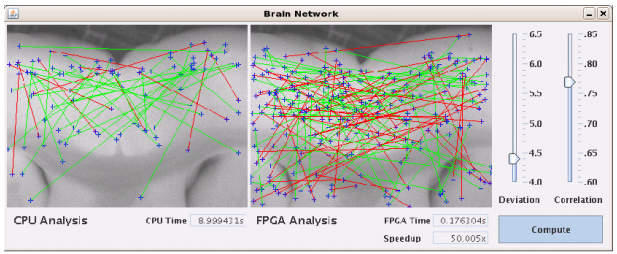
Earthquake Prediction



Smith-Waterman



Brain Network



MAXELER Technologies