Vertical Acceleration: From Algorithms to Logic Gates

considering Economics of Computation

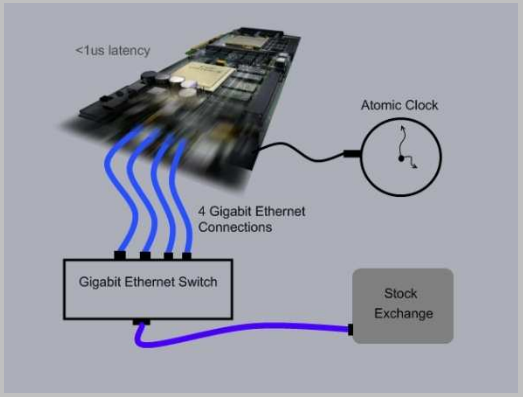
Oskar Mencer

Beograd, August 2010

MAXELER Technologies
MAXIMUM PERFORMANCE COMPUTING

Maxeler Technologies

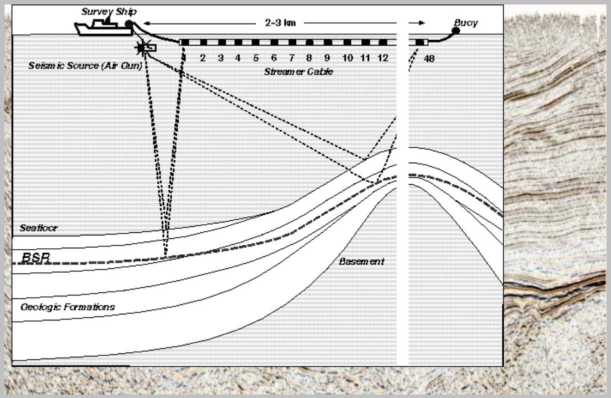
MaxCard
e.g. HFT Solution



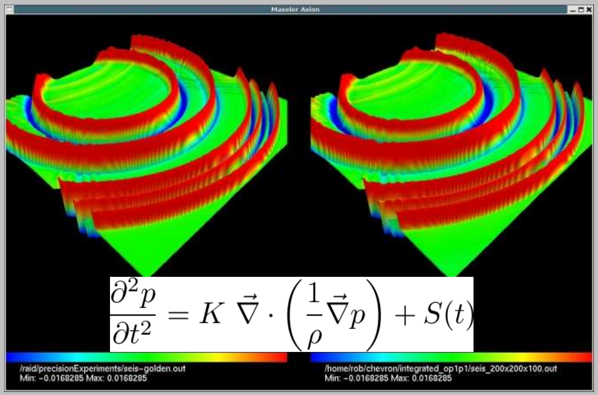
MaxBox
4 MaxCards in a 1U box

MaxRack
Storage, Network and Compute

Real-time trace processing



Finite Difference (with Chevron)

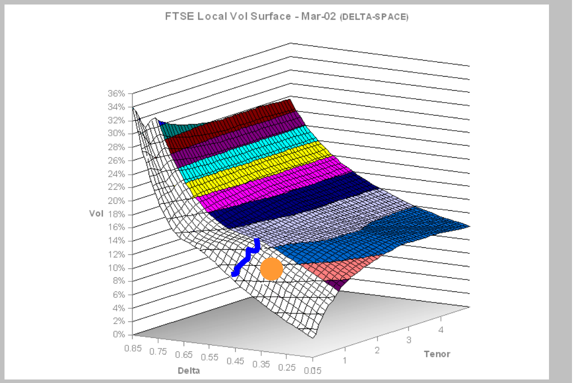


$$\frac{\partial^2 p}{\partial t^2} = K\ \vec{\nabla} \cdot \left(\frac{1}{\rho} \vec{\nabla} p \right) + S(t)$$

/raid/precisionExperiments/seis-golden.out
Min: -0.0168285 Max: 0.0168285

/home/rob/chevron/integrated_op1p1/seis_200x200x100.out
Min: -0.0168285 Max: 0.0168285

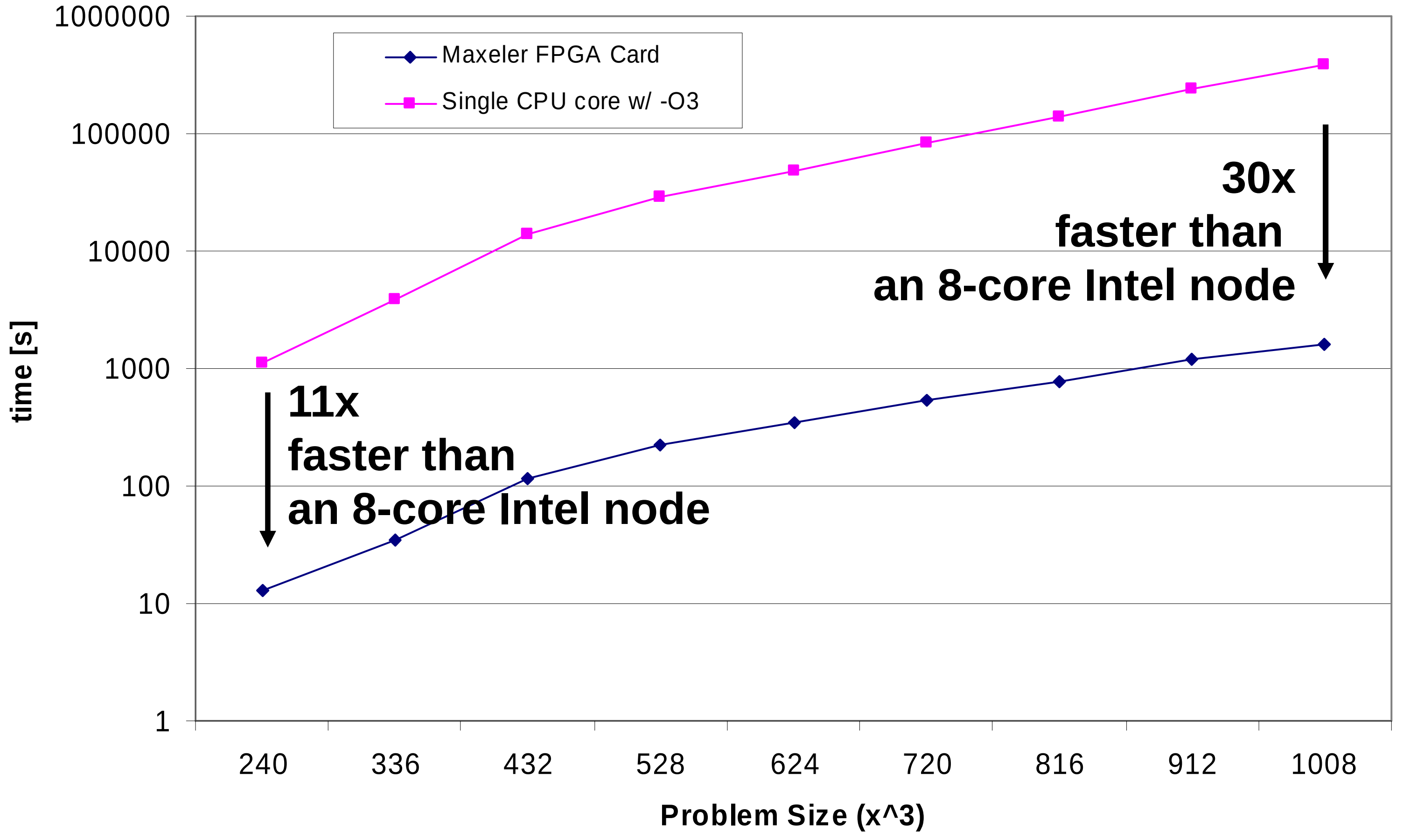
Local Vol Approximation



MAXELER Technologies
MAXIMUM PERFORMANCE COMPUTING

Speedup of 3D Finite Difference

Chevron Case Study $$\frac{\partial^2 p}{\partial t^2} = K\,\vec{\nabla}\cdot\left(\frac{1}{\rho}\vec{\nabla}p\right) + S(t)$$



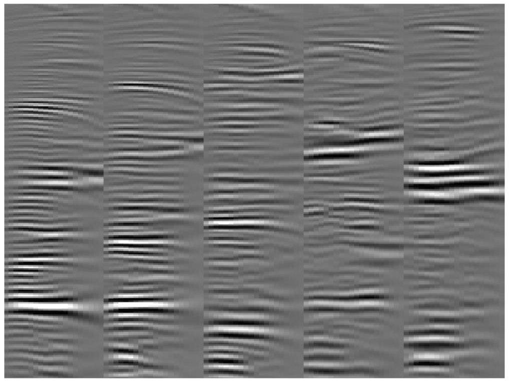
30x faster than an 8-core Intel node

11x faster than an 8-core Intel node

* published by the Society of Exploration Geophysicists in 2008

48x Speedup of Angle Gathers

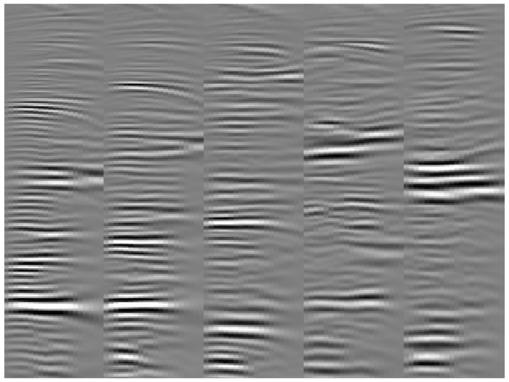
with Stanford Center for Earth and Environmental Sciences *)



Angle gathers from CPU computed subsurface offsets

- Can be dominant cost in shot profile migration
- Cross-correlating two fields by various shifts:

$$I(h,x,z) = \sum_s \sum_w S(x-h,z,w,s) \cdot G^*(x+h,z,w,s)$$



Angle gathers from FPGA computed subsurface offsets

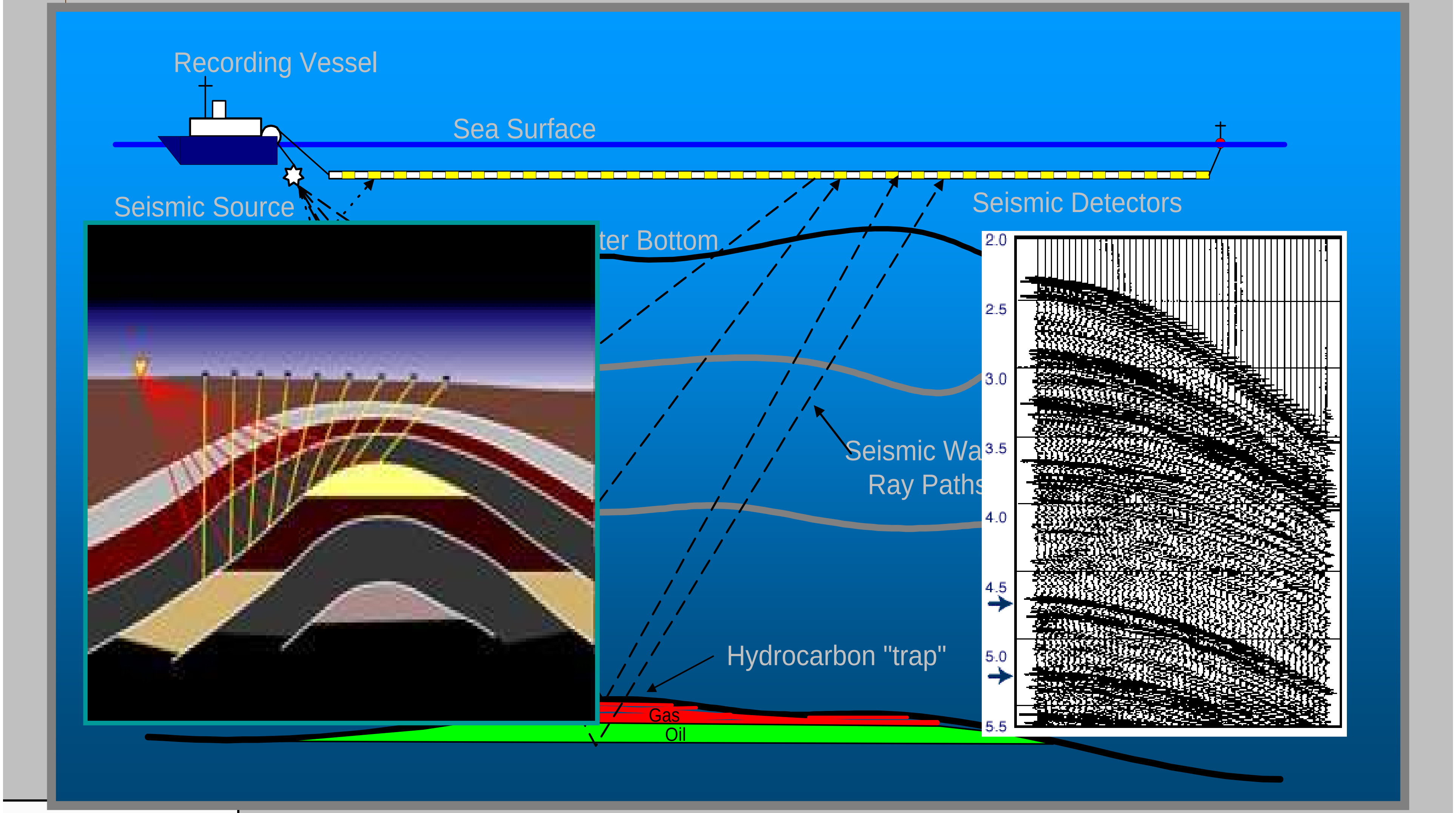
SPEEDUP RESULTS FROM CUSTOM TRACE MEMORY SYSTEM:

- **Trace = Unit of Transfer**
- **Buffers Prefetch Right Traces in Advance**

*) Case Study: Subsurface Offset Gathers, presented at SEG 2008

MAXELER Technologies
MAXIMUM PERFORMANCE COMPUTING

Seismic Data Acquisition

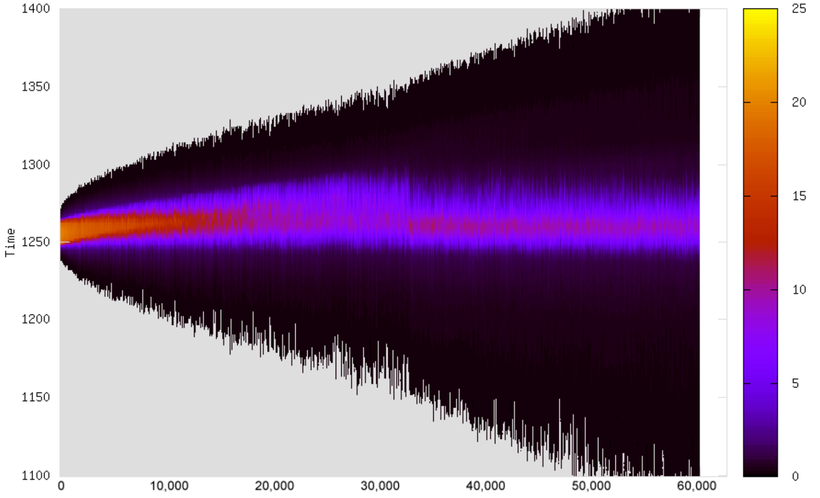


Courtesy of Schlumberger

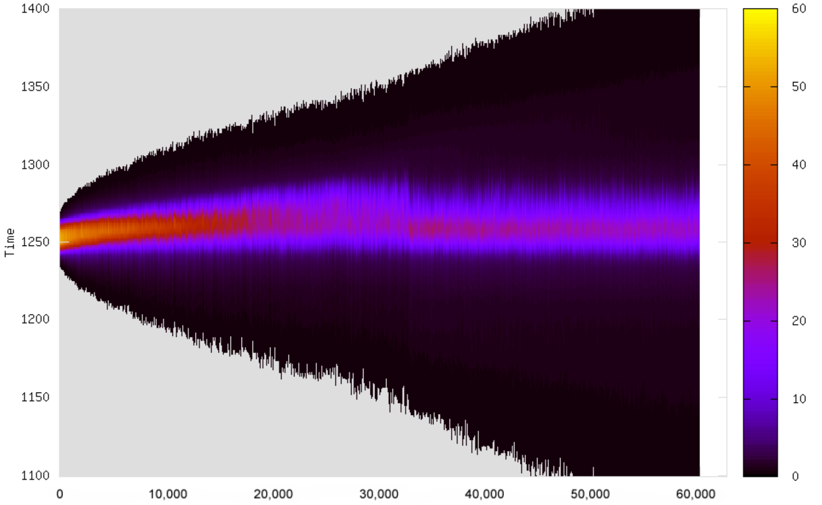
ENI-AGIP Seismic Trace App: Conjugate Gradient Optimization

100 MAX2 cards delivering performance of 21,800 CPU cores[EAGE2010]

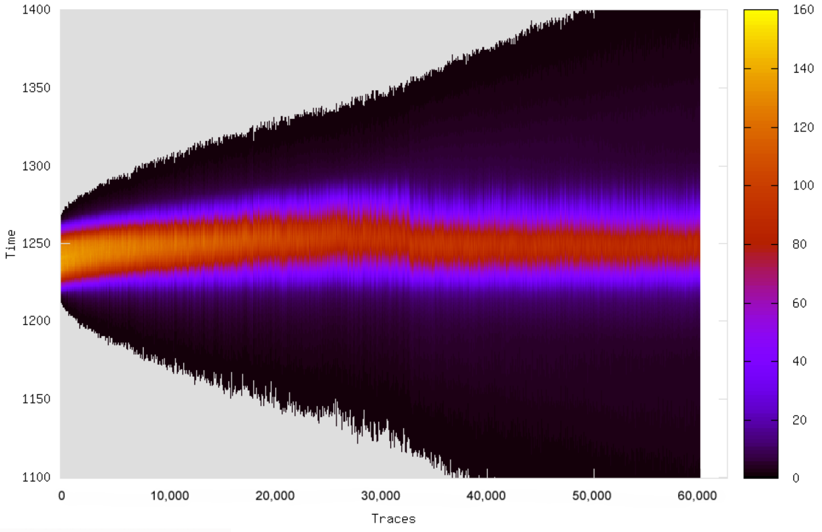
Data Use with 1 t_0



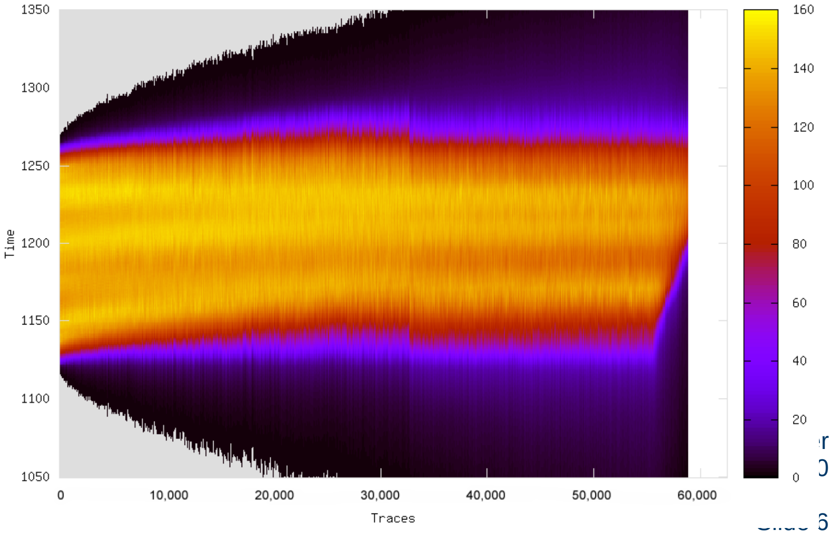
Data Use with 4 t_0



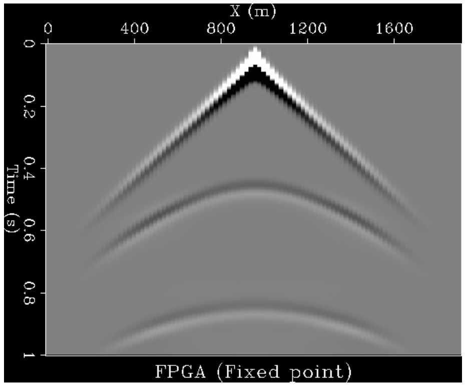
Data Use with 16 t_0



Data Use with 64 t_0

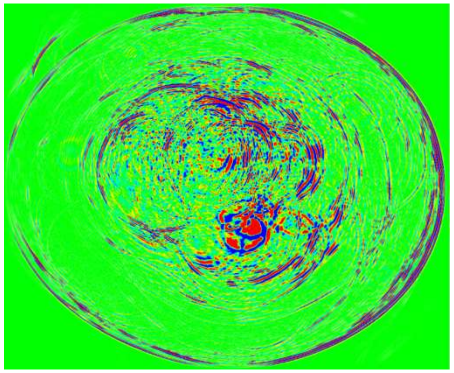


Acceleration Projects end up at 20-30x



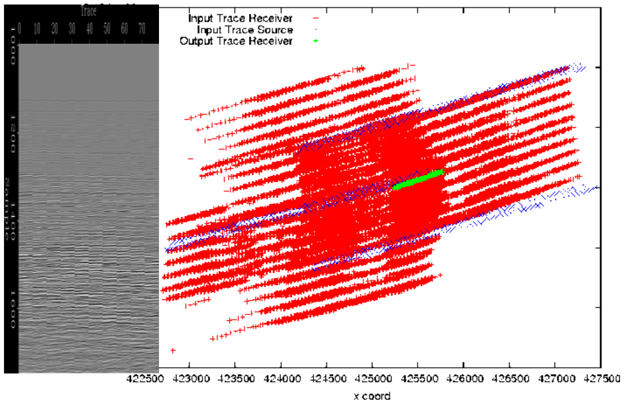
Customer A

App1: 19x, App2: 25x



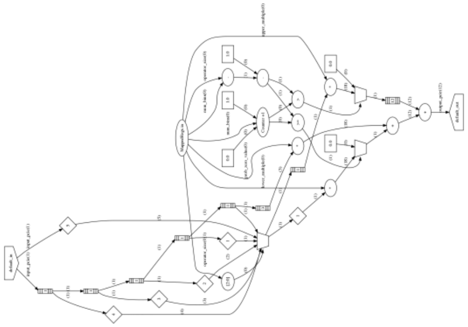
Customer B

1.2GB/s per card



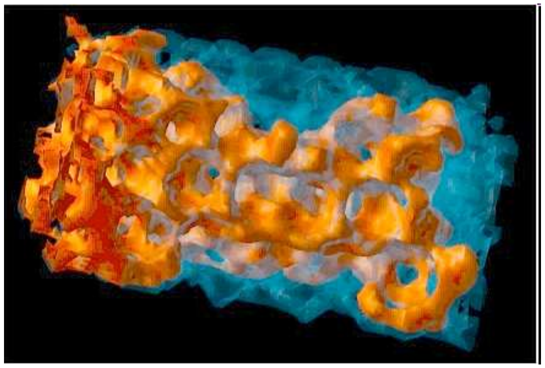
Customer C

App1: 22x, App2: 22x



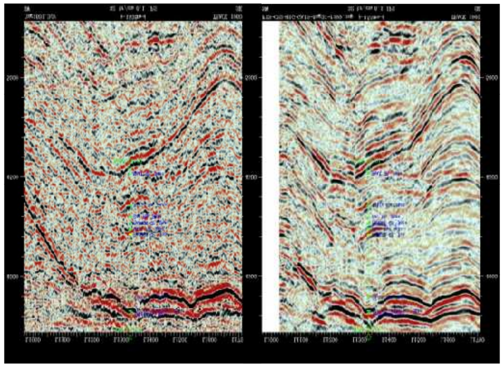
Customer D

App1: 32x, App2: 29x



Customer E

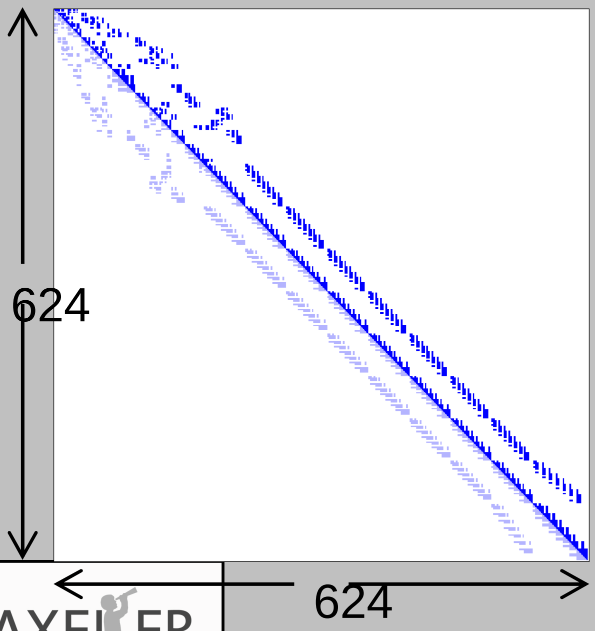
App1: 30x

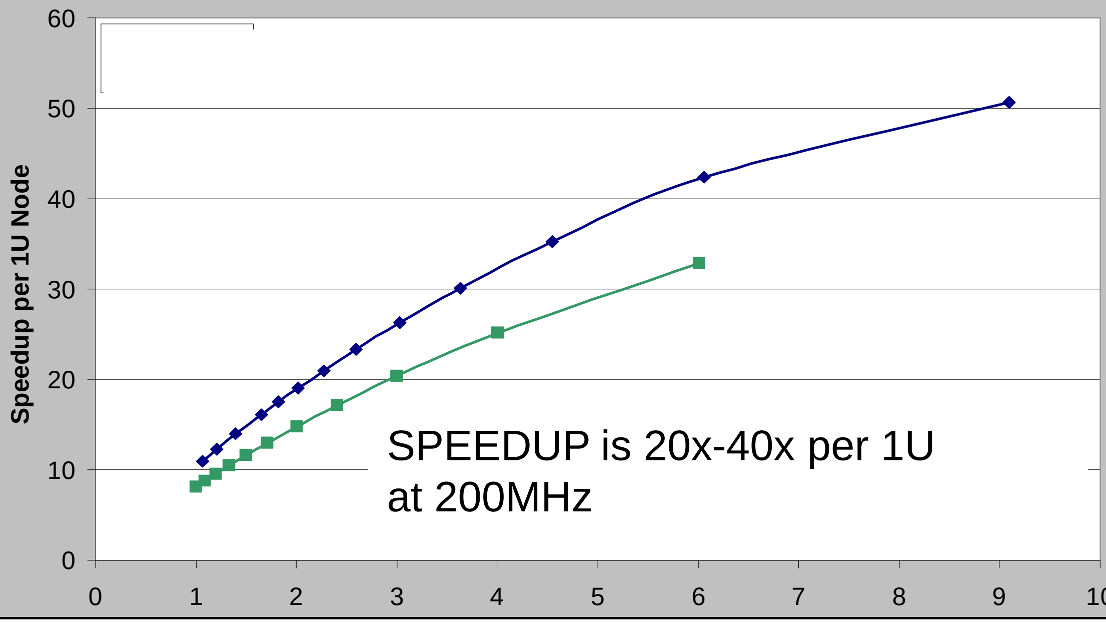


Customer F

App1: 26x, App2: 30x

A Maxeler Sparse Matrix Solution



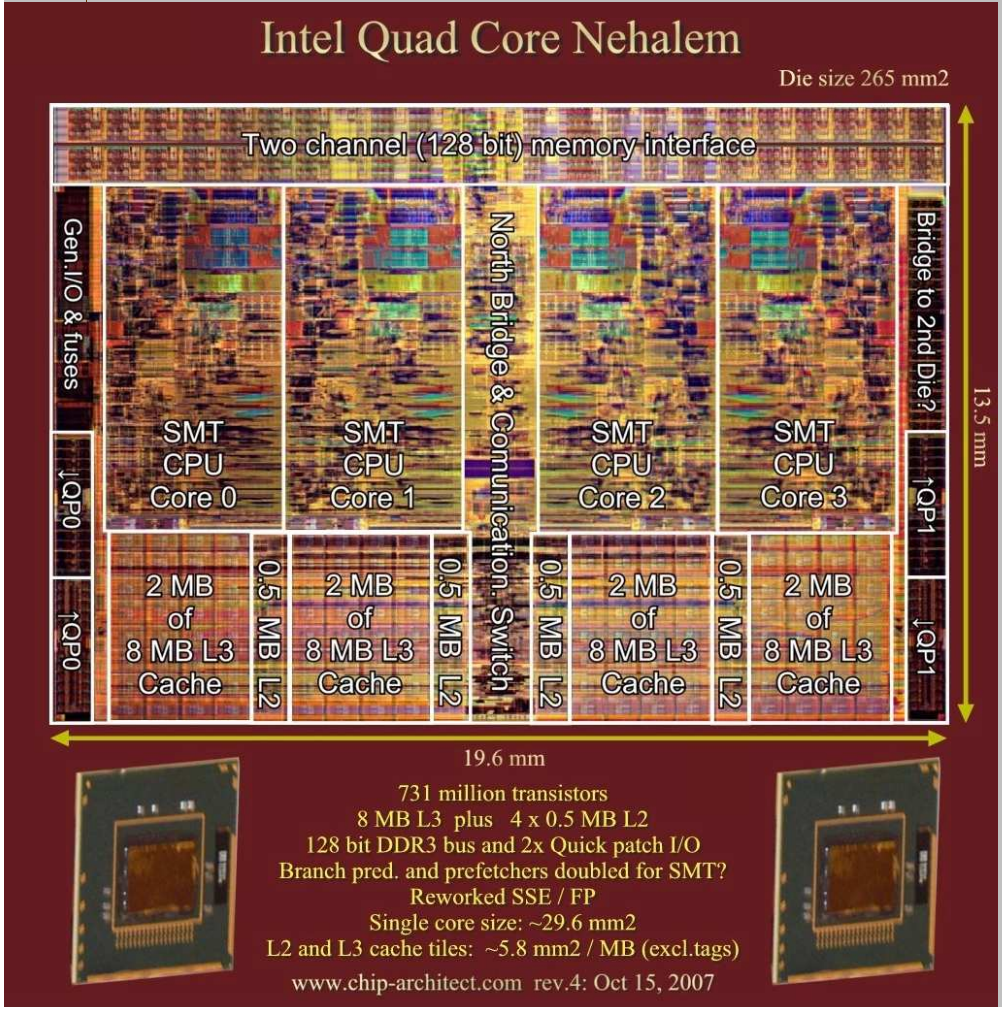


SPEEDUP is 20x-40x per 1U at 200MHz

Maxeler Domain Specific Address and Data Encoding

Computing Technology 2010

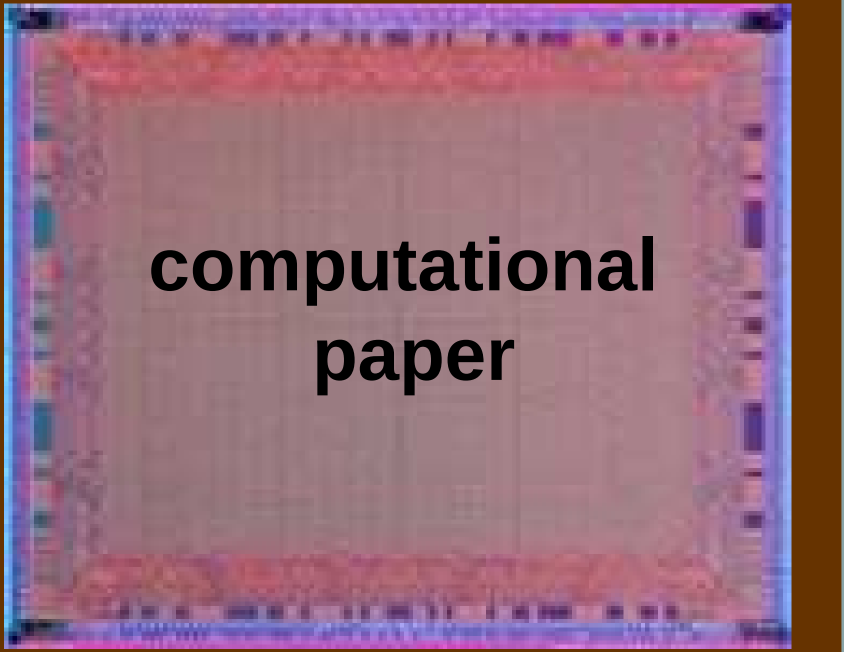
Microprocessor



731 million transistors
8 MB L3 plus 4 x 0.5 MB L2
128 bit DDR3 bus and 2x Quick patch I/O
Branch pred. and prefetchers doubled for SMT?
Reworked SSE / FP
Single core size: ~29.6 mm2
L2 and L3 cache tiles: ~5.8 mm2 / MB (excl.tags)
www.chip-architect.com rev.4: Oct 15, 2007

FPGA

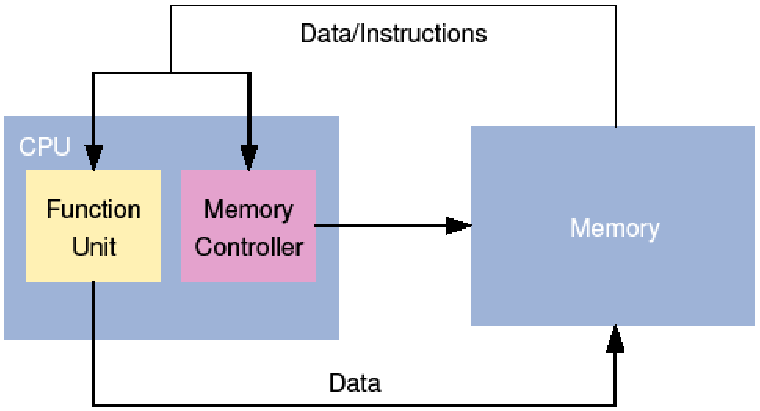
Xilinx Virtex-6 FPGA

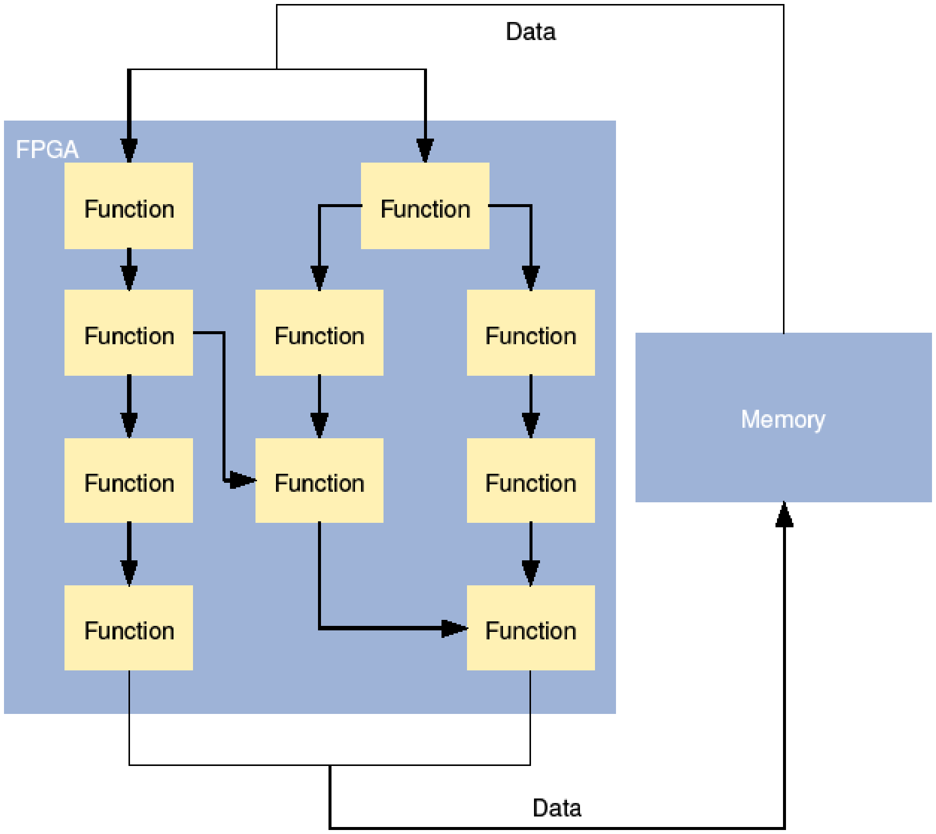


computational paper

- 5MB at >1TB/s
- 2000 multipliers
- ~1M logic elements

Computing with CPUs versus FPGAs

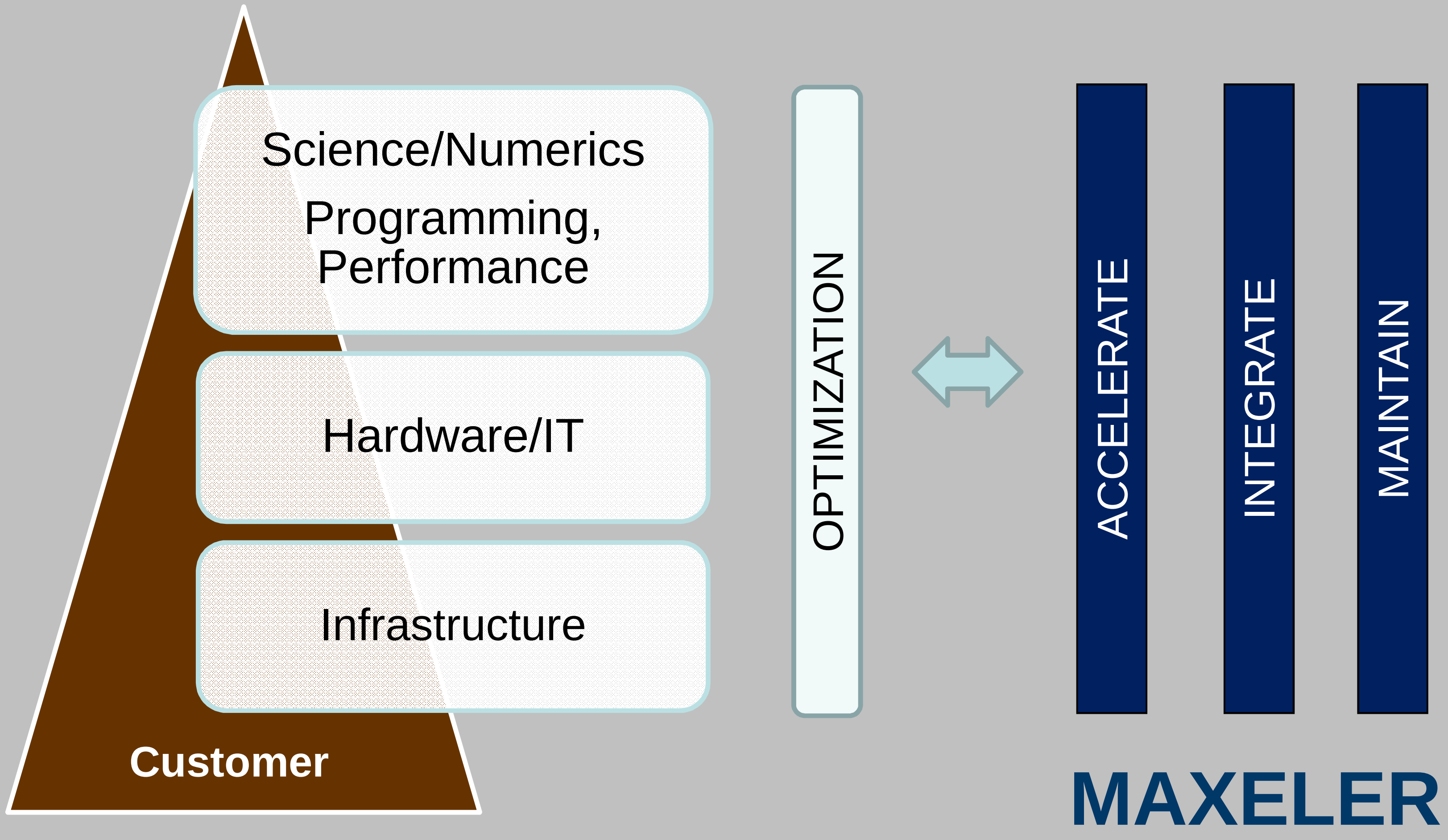




Streaming Data through a data flow machine

Acceleration is Hard

Vertical Optimizations in a Horizontal world



Science/Numerics

Programming, Performance

Hardware/IT

Infrastructure

Customer

OPTIMIZATION

ACCELERATE

INTEGRATE

MAINTAIN

MAXELER

Managing a Complex C++ Acceleration Project

“With C you can shoot yourself in the foot.
with C++ you can blow your whole leg away.”

Abstraction

- The enemy of acceleration?
- We balance abstraction with modelling of underlying dataflow
- In the process we make the code easier to read and maintain

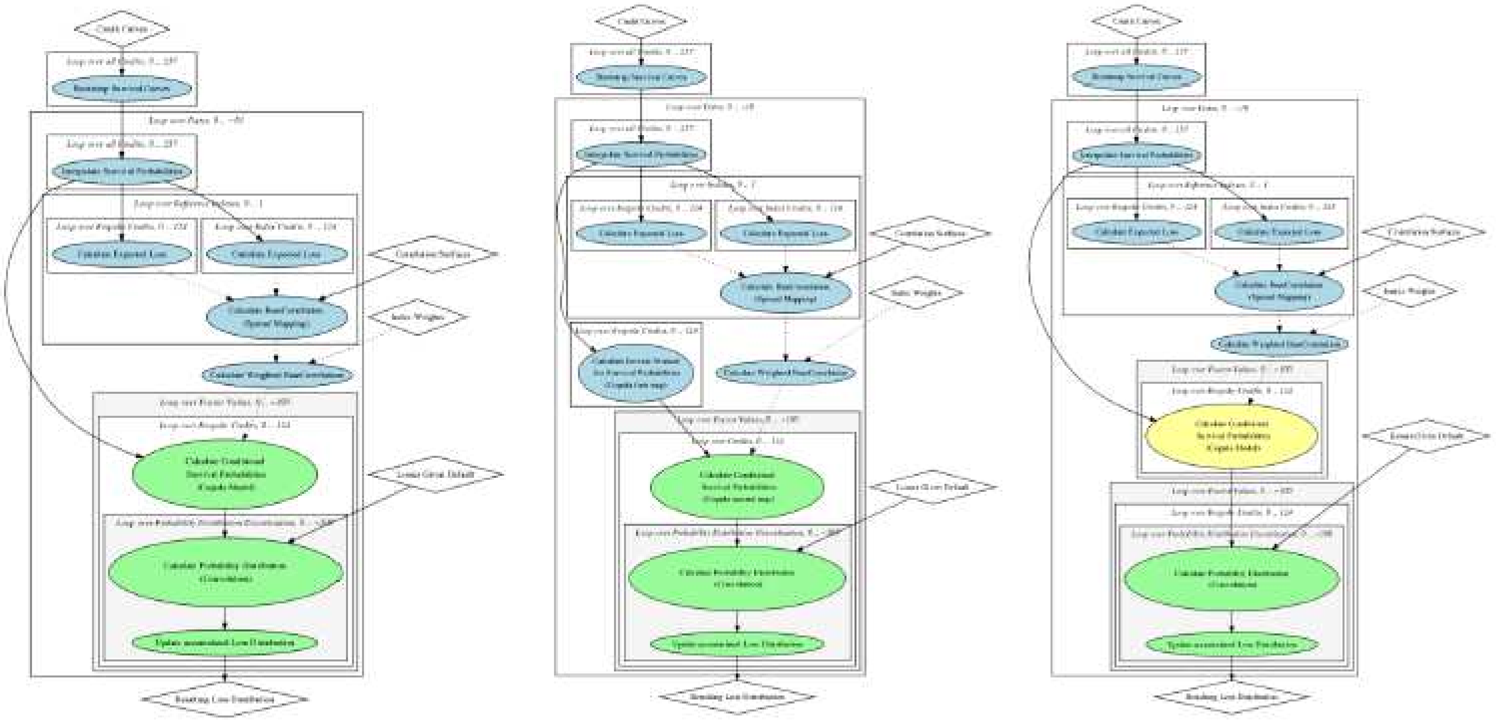
Software Transformation

- Requires the support of the people who wrote the code
- Focus on large “chunk size”, the unit of data and computation
- Requires adaption of the job distribution system

Fast & Reliable

- Acceleration has to track a moving target; ongoing development
- Accelerating the acceleration process: Agile Programming
- Coding standards to achieve fast deliveries

Maxeler Loop Flow Graphs for JP Morgan Credit Derivatives Transformation Options



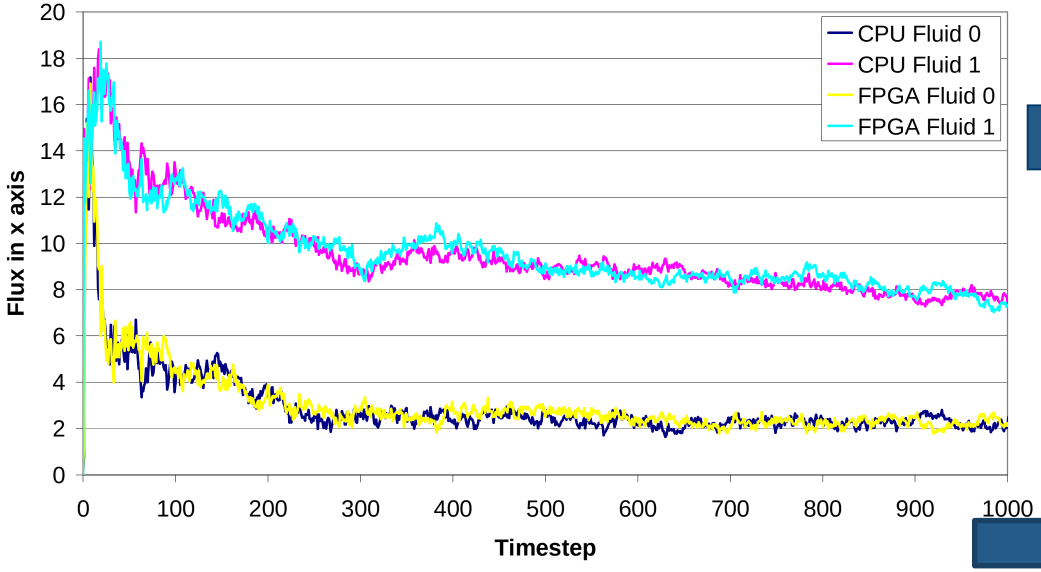
Option 1

Option 2

Option 3

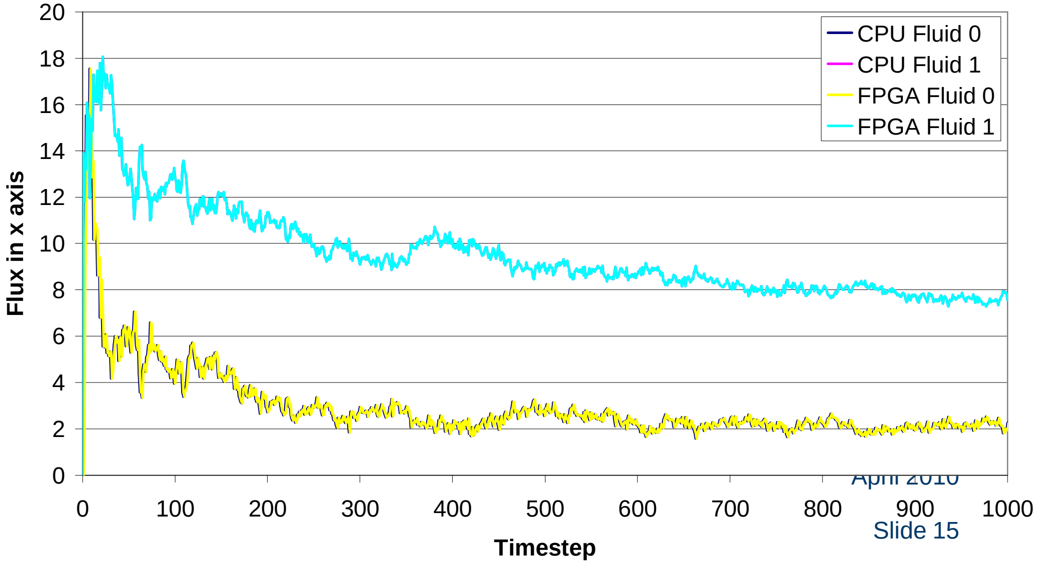
MAXELER Technologies MAXIMUM PERFORMANCE COMPUTING

30x Accelerated 2-Fluid Lattice Boltzmann: Changing the sorting step inside the iteration



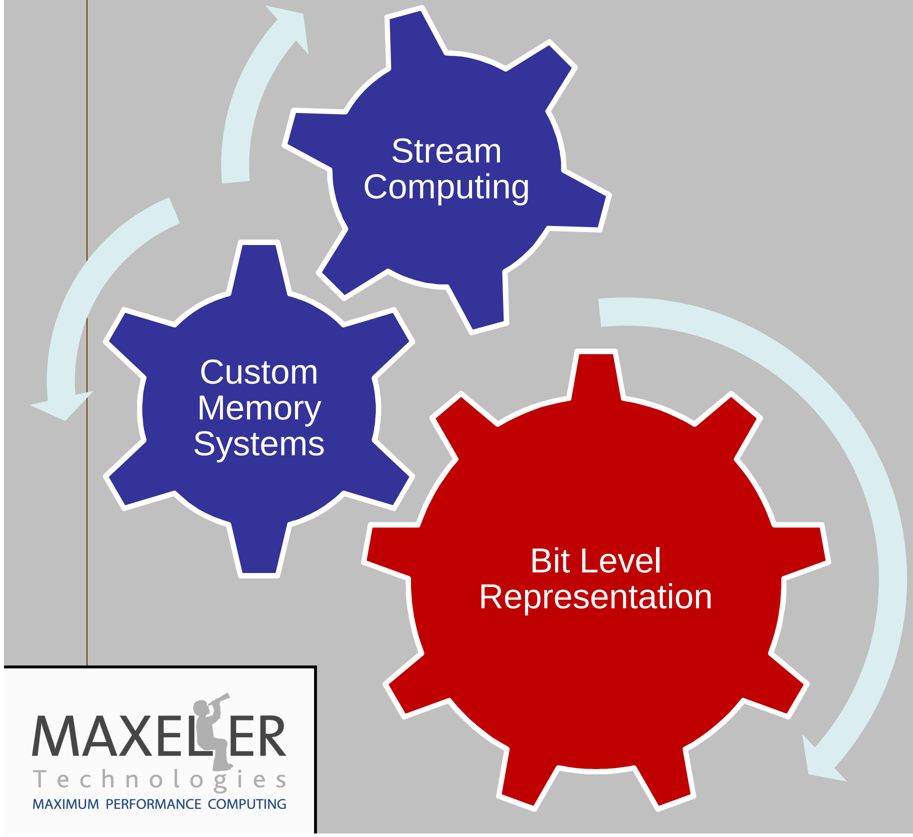
Effect of modifying sort in both software and FPGA

	Avg. Rel. Diff. CPU vs. FPGA
Original sort	7.9E-2
Modified sort	6.9E-7



Combining Algorithms & Acceleration

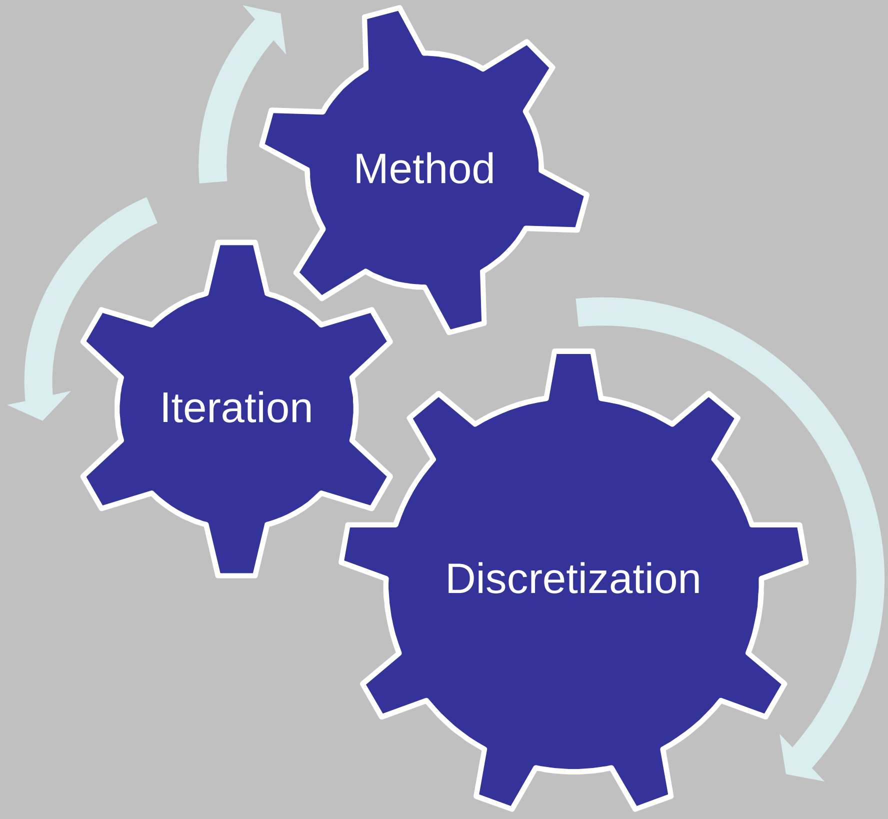
ACCELERATION



Stream Computing

Custom Memory Systems

Bit Level Representation



Method

Iteration

Discretization

NUMERICS

MAXELER Technologies
MAXIMUM PERFORMANCE COMPUTING

Slow and Fast Memories

John von Neumann, 1946:

“We are forced to recognize the possibility of constructing a hierarchy of memories, each of which has greater capacity than the preceding, but which is less quickly accessible.”

SABR model:

$$dF_t = \sigma_t F_t^{\beta} dW_t$$

$$d\sigma_t = \alpha \sigma_t dZ_t$$

$$< dW, dZ >= \rho dt$$

we integrate in time (Euler in log-forward, Milstein in vol.)

$$\ln F_{t+1} = \ln F_t - \tfrac{1}{2}(\sigma_t \exp((\beta - 1)\ln F_t))^2 .dt + \sigma_t \exp((\beta - 1)\ln F_t)\Delta W_t$$

$$\sigma_{t+1} = \sigma_t + \alpha\sigma_t \Delta Z_t + \tfrac{1}{2}(\alpha\sigma_t)(\alpha)\left(\Delta Z_t^2 - dt\right)$$

For each path, initialize: $\ln F_0, \quad \sigma_0$

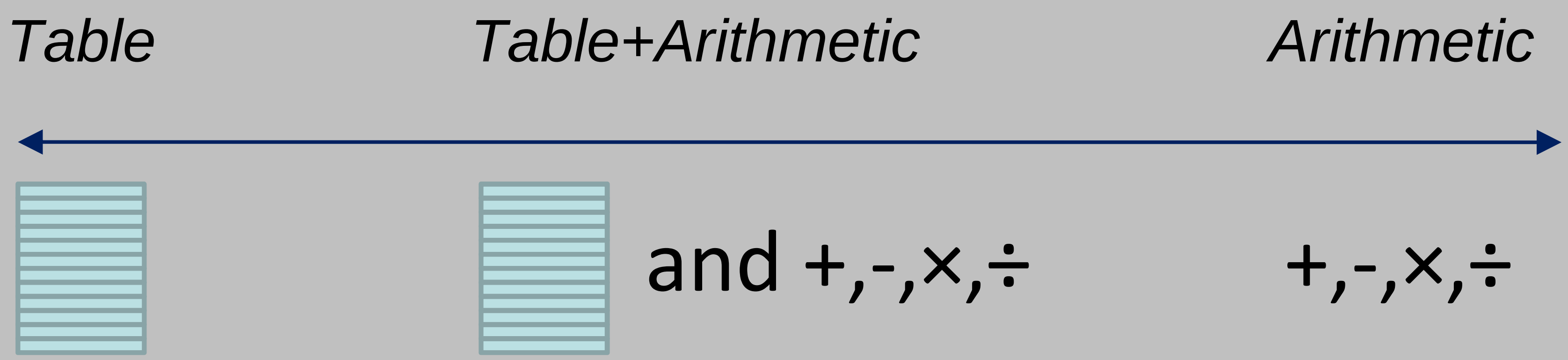
For each t: $\ln F_{t+1} = G(\sigma_t, \ln F_t, \Delta W_t)$

$$\sigma_{t+1} = H(\sigma_t, \Delta Z_t)$$

… How do we compute G and H optimally?

Design Space for Function Evaluation

Computing f(x) in the range [a,b] with $|E| \le 2^{-n}$



Table

Table+Arithmetic

Arithmetic

and +,-,×,÷

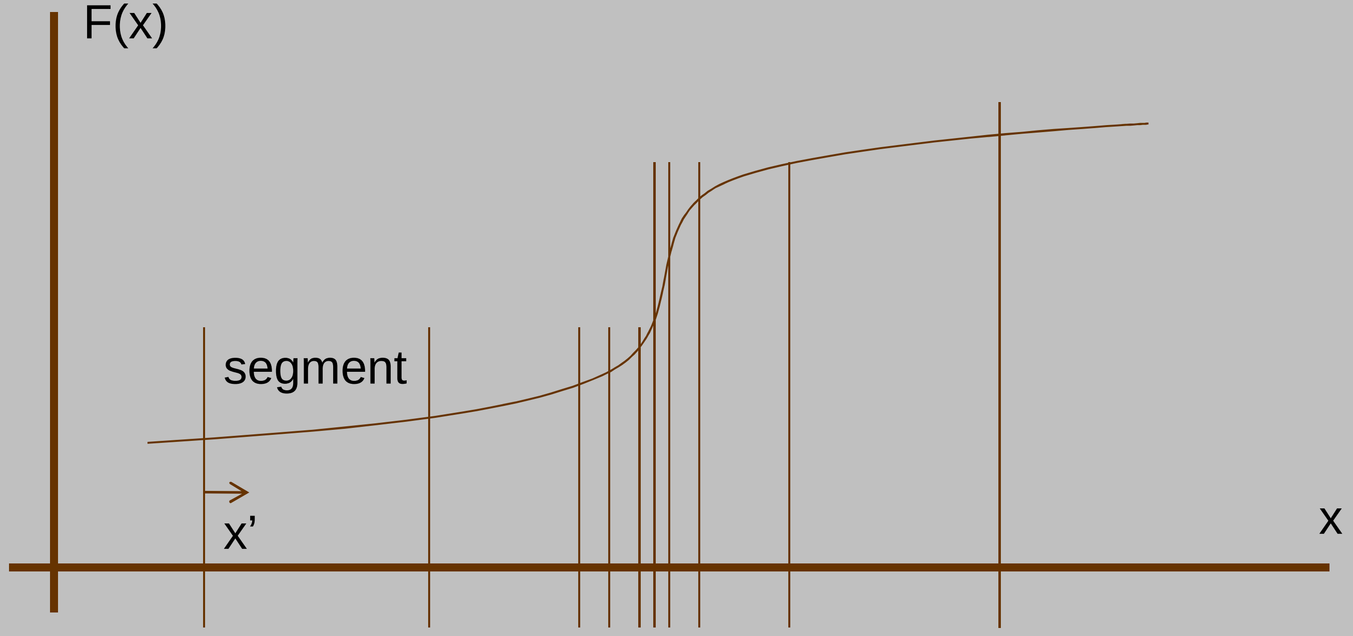
+,-,×,÷

- uniform vs non-uniform
- number of table entries
- how many coefficients

- polynomial or rational approx
- continued fractions
- multi-partite tables

Underlying hardware/technology changes the optimum

Variable Segments and Approximations



Approximate in each segment separately:

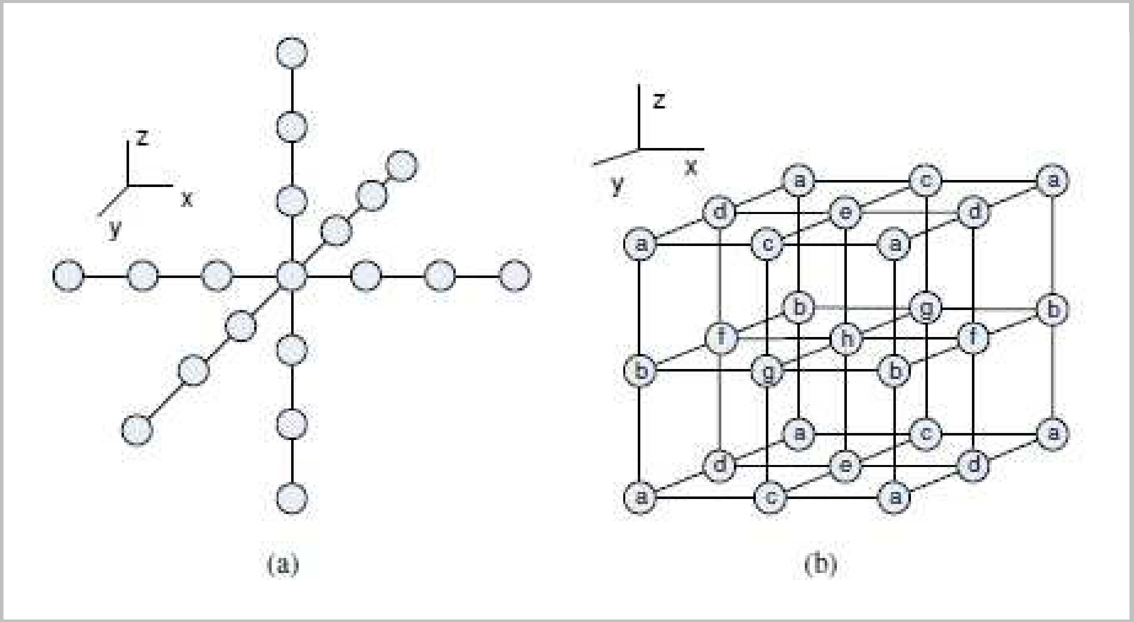
$$p(x') = c_0 + c_1 x' + c_2 x'^2 + c_3 x'^3 \cdots \qquad r(x) = \frac{a_0 + a_1 x + a_2 x^2 + a_3 x^3 \cdots}{b_0 + b_1 x + b_2 x^2 + b_3 x^3 \cdots}$$

Compute Coefficients: Taylor Series, MiniMax (Remez), Splines,...
Precision: How many coefficients AND how many bits per variable?

MAXELER Technologies
MAXIMUM PERFORMANCE COMPUTING

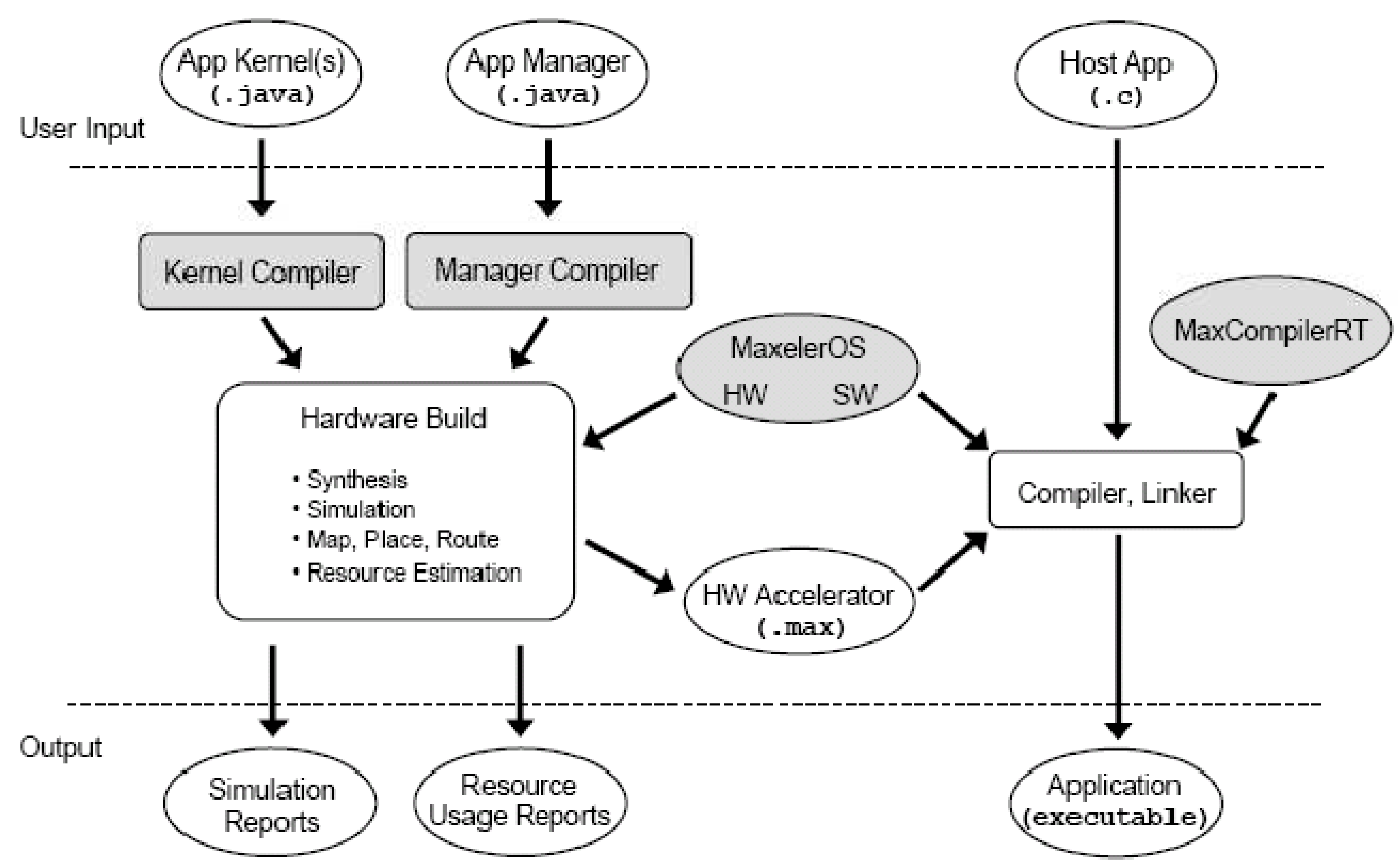
Finite Difference Stencils and Coefficients

- Monte Carlo vs Finite Difference vs Finite Elements
- Explicit versus implicit
- Discretization, delta_t versus accuracy
- Example: 3D stencil shape for Finite Difference



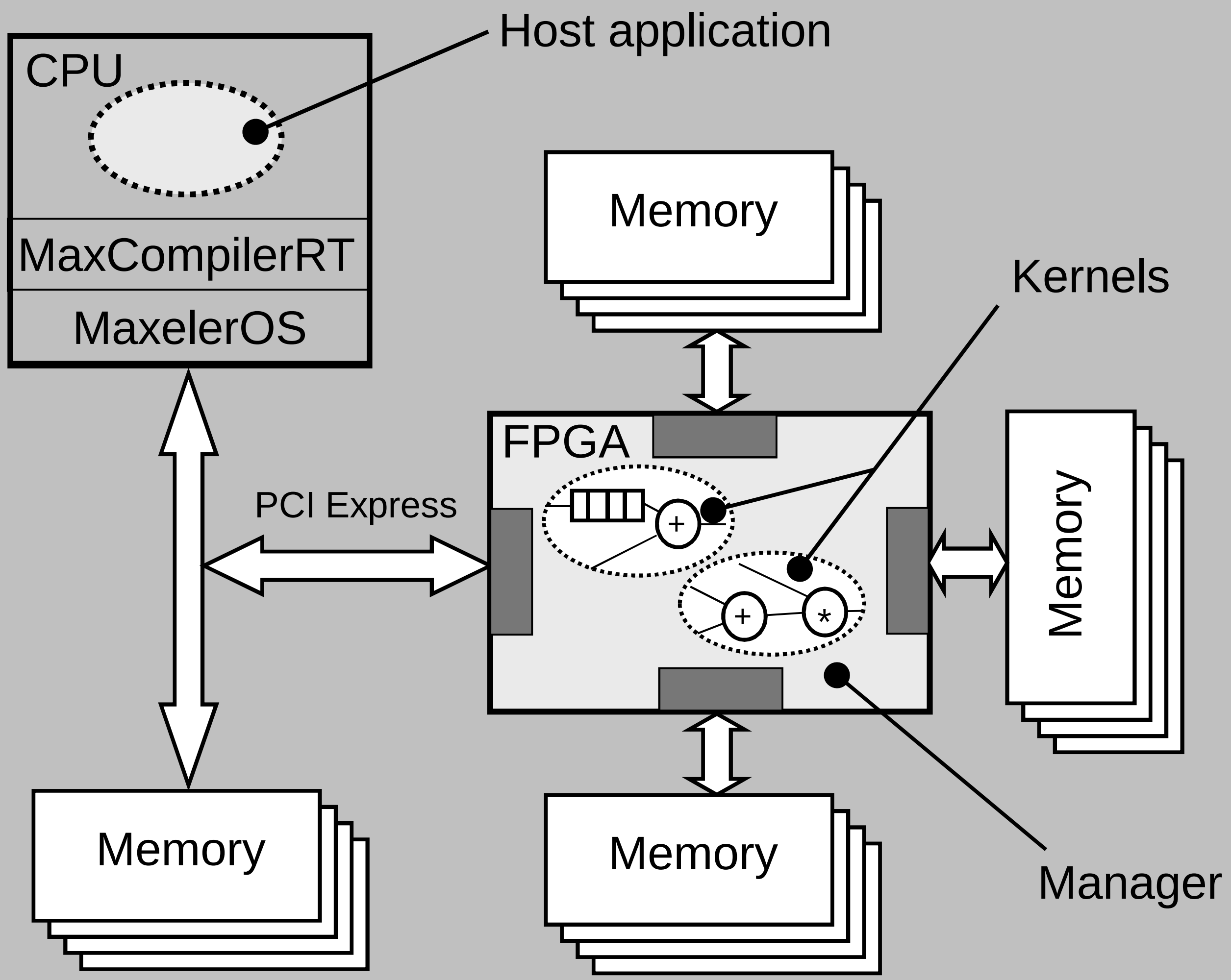
joint Stanford/Maxeler presentation at the SEG meeting 2009.

MaxCompiler

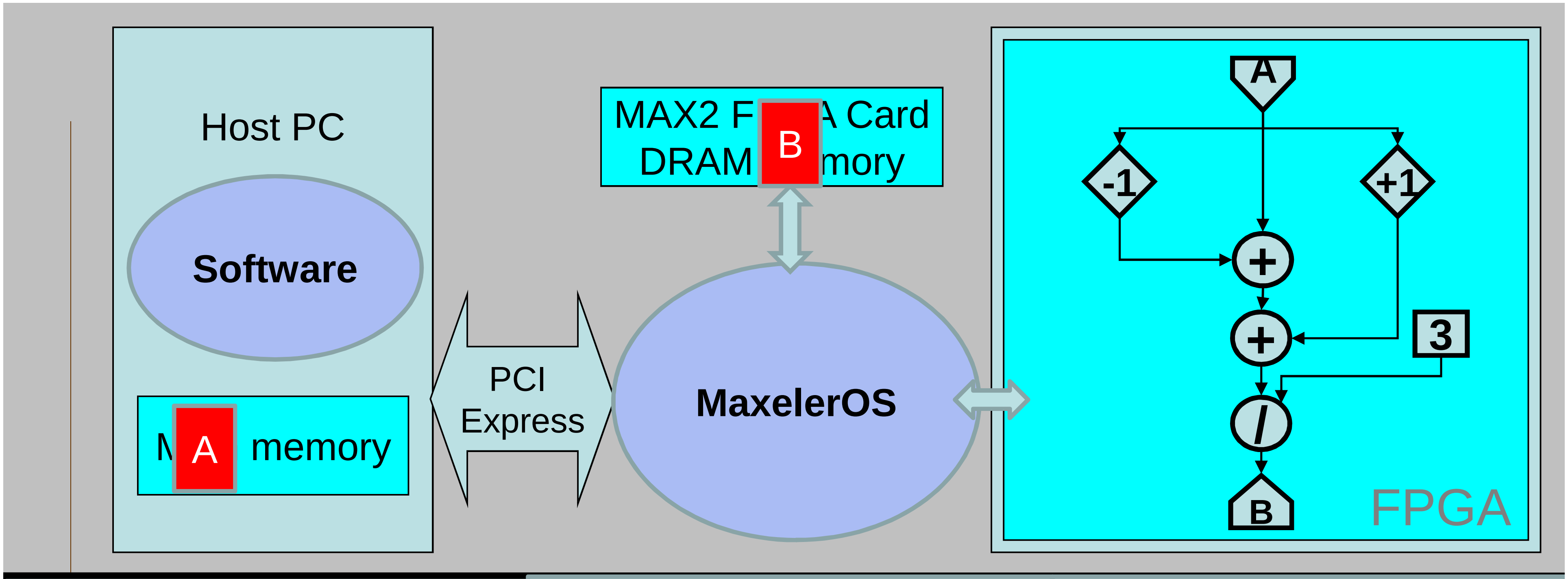


MAXELER Technologies
MAXIMUM PERFORMANCE COMPUTING

Generic Acceleration Architecture



MAXELER Technologies
MAXIMUM PERFORMANCE COMPUTING



Software

C

```
device = max_open_device(
            maxfile, "/dev/max0");
float A[SIZE];
…
stream_data(device, A);

~~for (int i=0; i<SIZE; ++i) {~~
~~  B[i] = ( A[i-1] + A[i] + A[i+1] )/3;~~
~~}~~

…
```

Manager

Java

```
Manager m = new
     Manager("Loop", MAX2);

m.kernel(mav_kernel,
       link("A", PCIE),
       link("B", DRAM(LINEAR));

m.build();
```

Kernel

MaxJava

```
class mav_kernel
                  extends kernel{

  input ("A",hwFloat (12 ,52) ) ;
  output ("B",hwFloat (12 ,52) ) ;

  A_prev=streamOffset(-1,A);
  A_next=streamOffset(1,A);

  B = (A_prev+A+A_next) / 3 ;
}
```

Providing Complete Solutions

Maxeler offers complete hardware, software and application acceleration solutions for high performance computing

Hardware

- Card: PCI Express x16, compute, memory and local interconnect
- Box: 1U solutions with 1 or 4 Cards
- Rack: 10U, 20U or 40U, balancing compute, storage & network

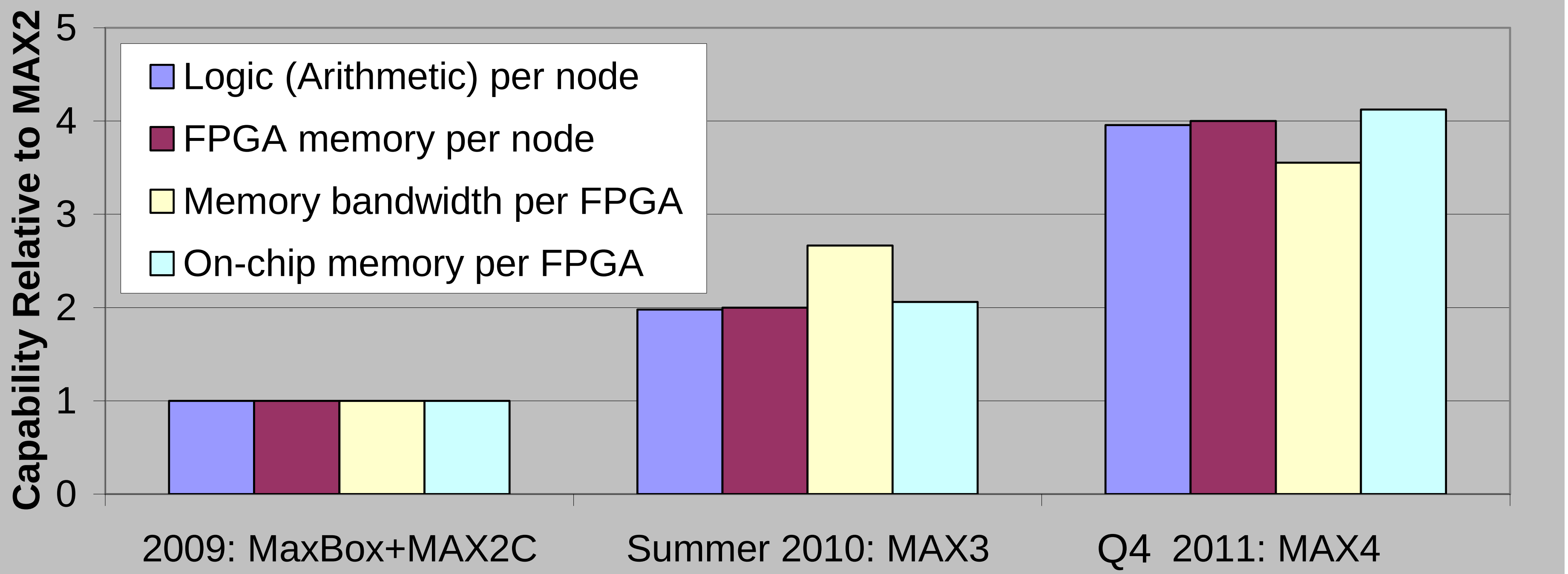
Software

- MaxelerOS: Resource management of Stream Computing
- Runtime support: memory management and data choreography
- MaxCompiler: providing programmability

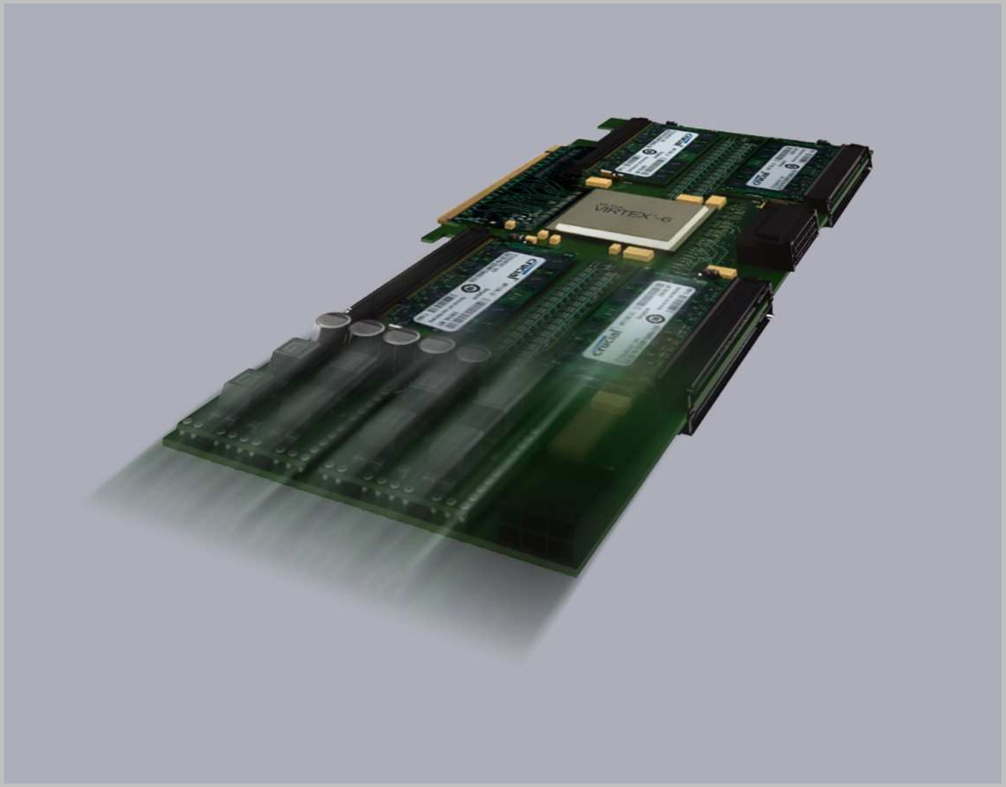
Consulting

- HPC System Performance Architecture
- Algorithms and Numerical Optimization
- Integration into business and technical processes

MAXELER Technology Roadmap



	2009: MaxBox+MAX2C	Summer 2010: MAX3	Q4 2011: MAX4
Silicon process	65nm	40nm	32nm
MaxCard compute capability per node	1,920 LC	3,800 LC	7,600 LC
MaxCard mem / node	96GB	192GB	384GB
DRAM bandwidth	15GB/s	39GB/s	51GB/s
Local On-chip memory	2.25MB	4.67MB	9.4MB
Power Consumption	417W	550W	680W



MAXELER
Technologies
MAXIMUM PERFORMANCE COMPUTING