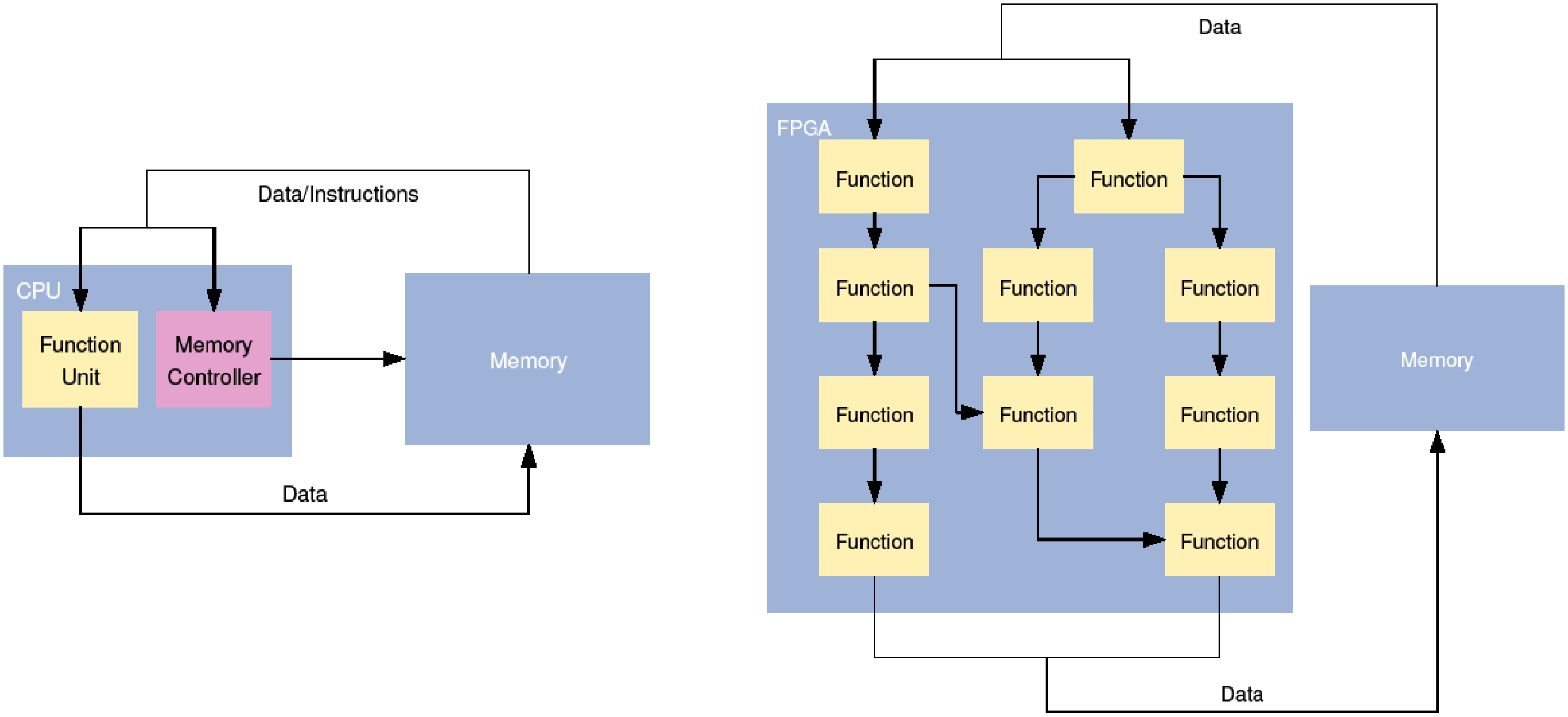
Surviving the End of Scaling of Traditional Micro Processors in HPC

Olav Lindtjørn (Schlumberger, Stanford), Robert G. Clapp (Stanford), Oliver Pell, Oskar Mencer, Michael J Flynn (Maxeler)

The Memory Wall and the Power Wall

- Moore's Law continues to deliver double the transistors on a chip every 18-24 months
  - But we are having trouble making those extra transistors deliver performance.
- Memory Wall
  - Parallel processing elements on-chip must share the same off-chip bandwidth
- Power Wall
  - Chips still need to be cooled in the same physical space

CPUs vs. FPGA Processing



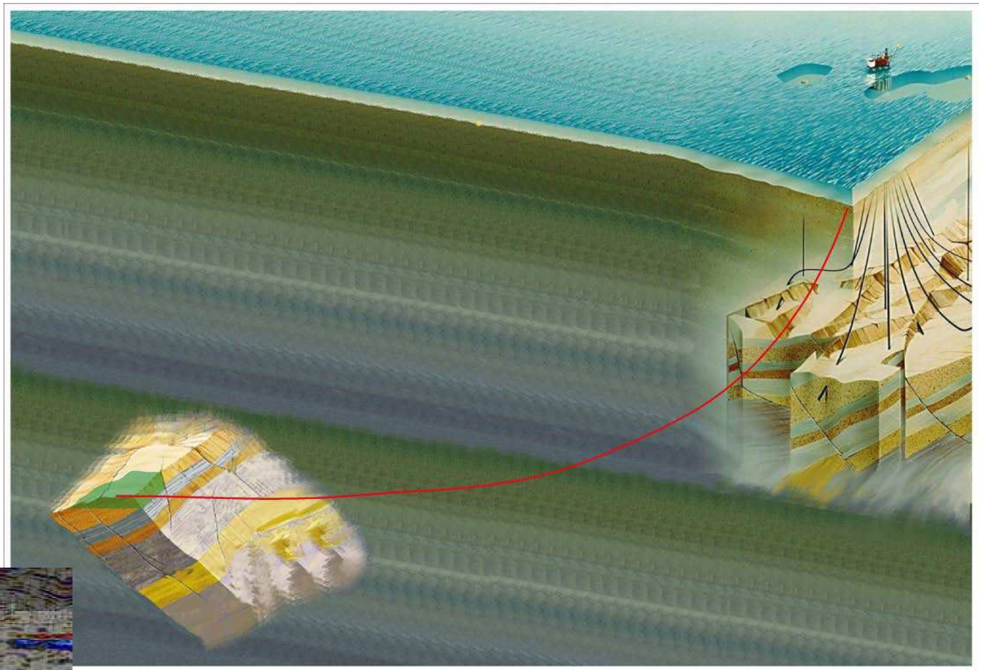
Streaming Data through a data flow machine

Outline

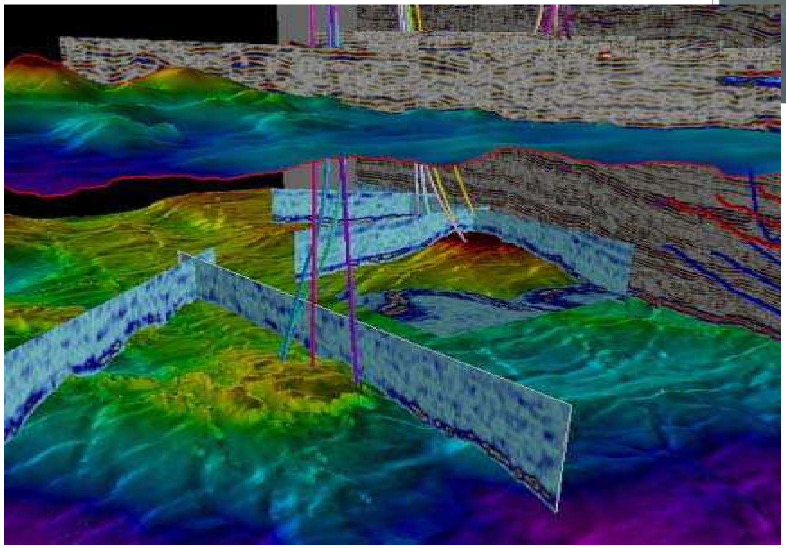
- Oil and Gas HPC applications
- Maxeler FPGA Compiler and Accelerators
- Key Computational Kernels in Oil&Gas
  – Sparse Matrix
  – Convolution based methods
- Applications scalability – Technology trends
- Conclusions

HPC – Its role in Oil & Gas exploration

- Identify resources
- Access resources
- Maximize recovery

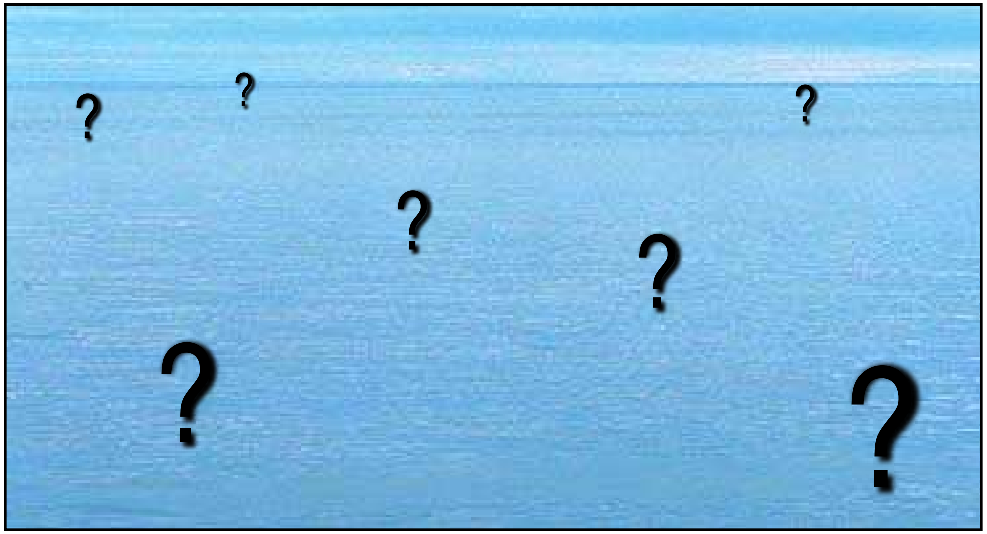


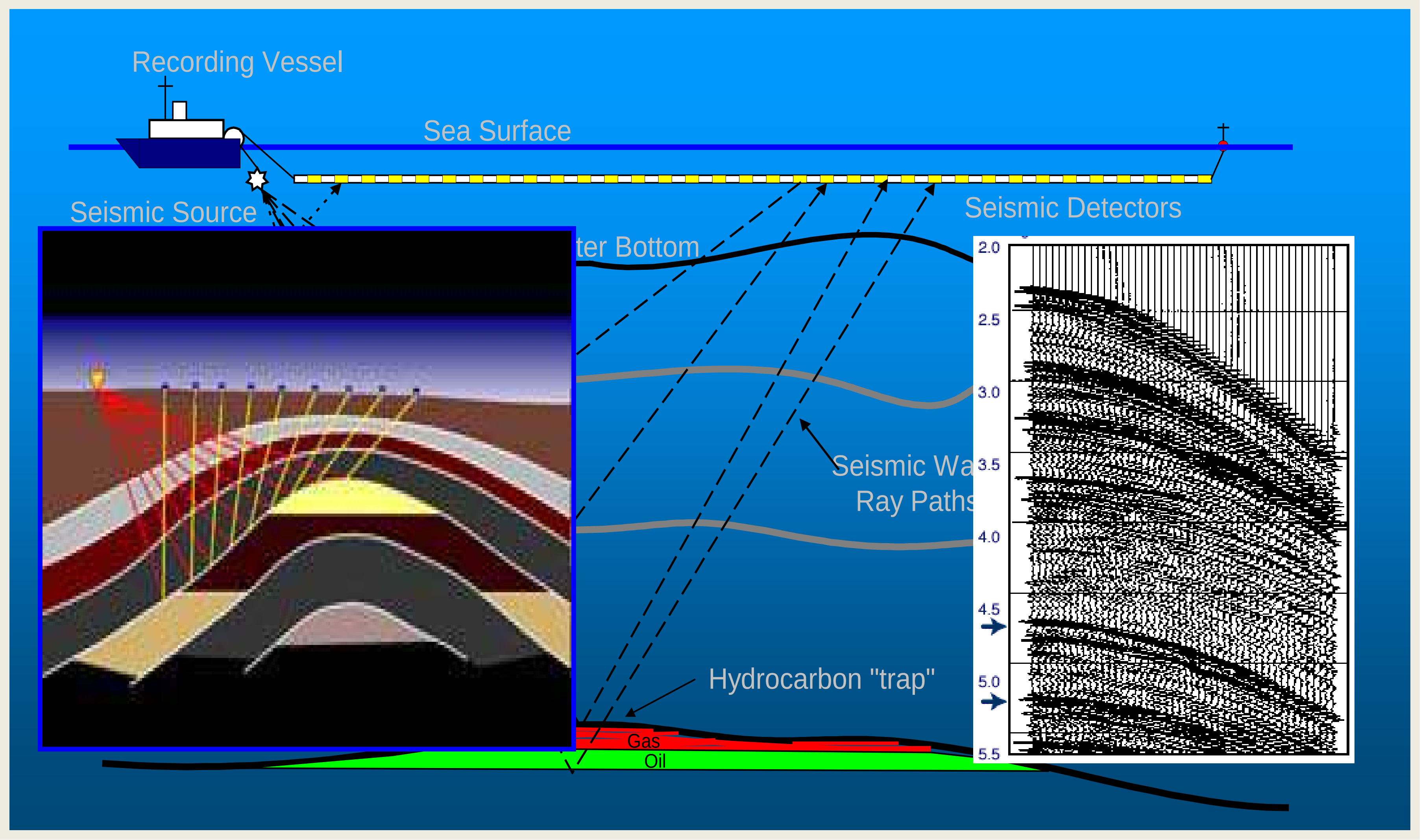
Courtesy of Statoil



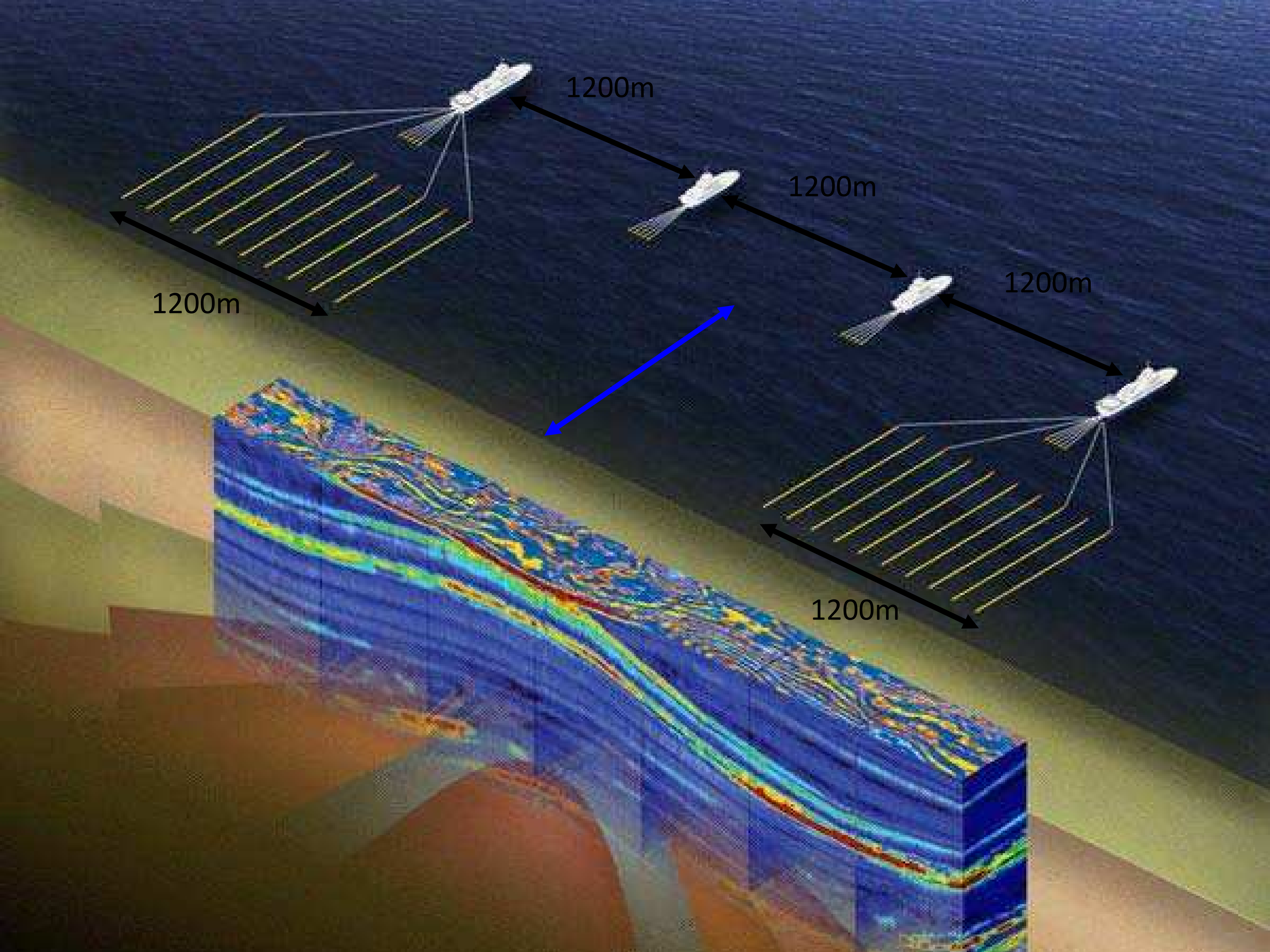
Where to Drill

- Seismic –Acoustic measurement
- Electromagnetic
- Gravity

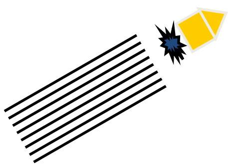


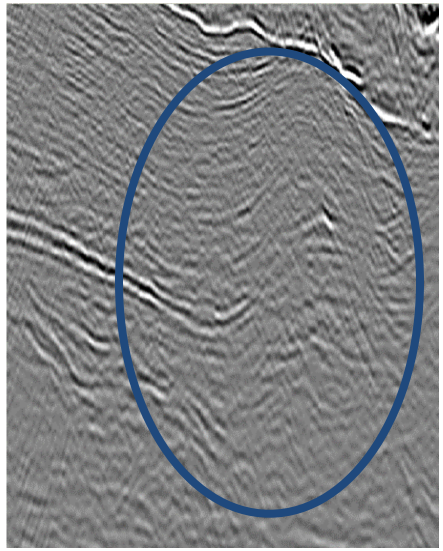


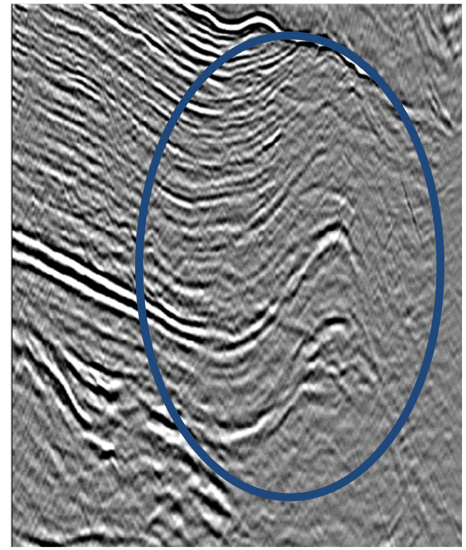
Recording Vessel
Sea Surface
Seismic Source
ter Bottom
Seismic Detectors
Seismic Wa
Ray Paths
Hydrocarbon "trap"
Gas
Oil
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5

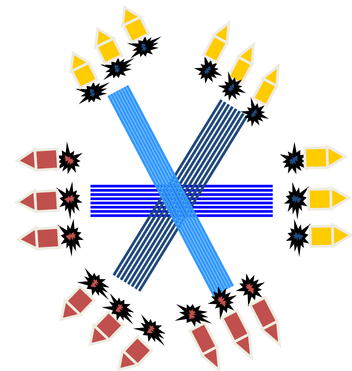

1200m
1200m
1200m
1200m
1200m

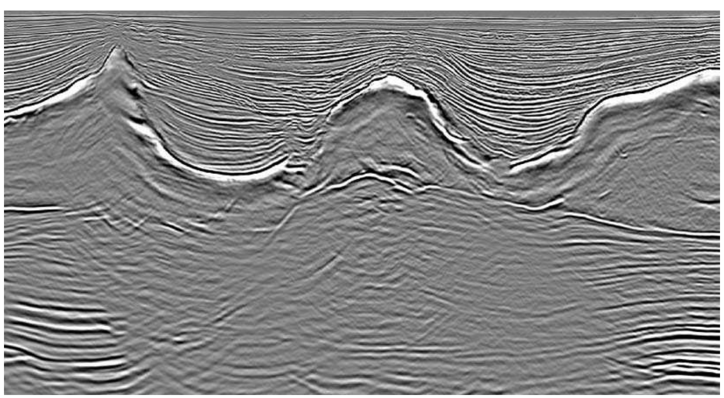
Data Intensity and Complex Physics

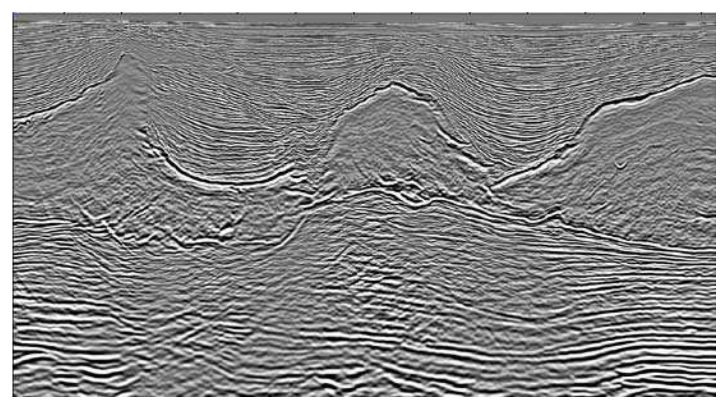








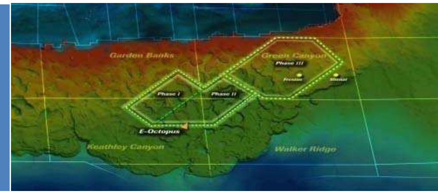


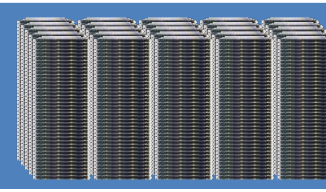


Isotropic

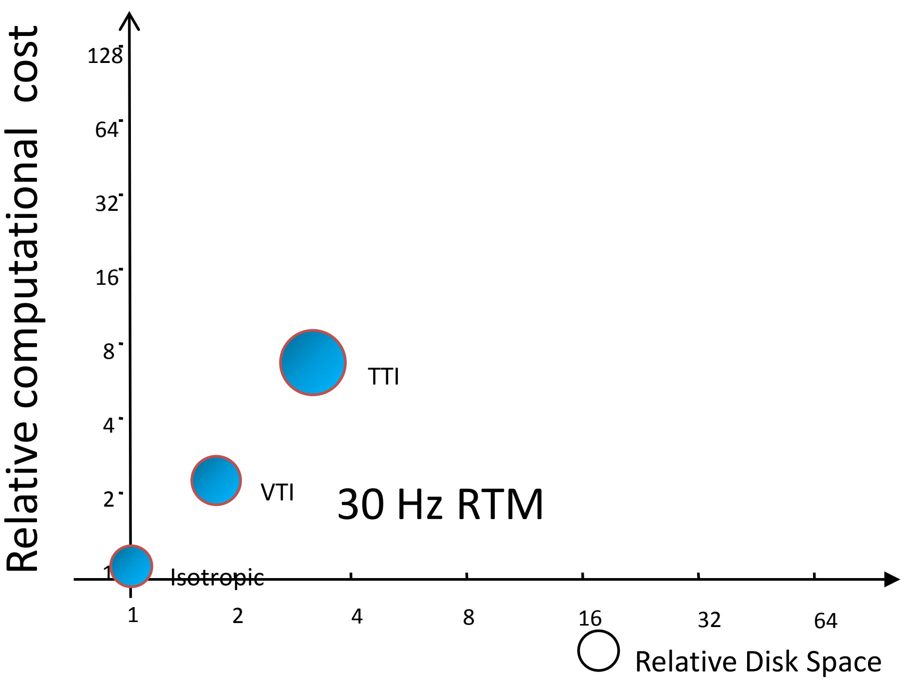
Anisotropic

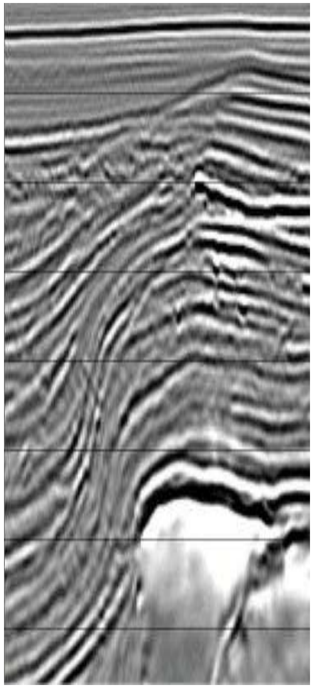
Data Rates and Computational needs



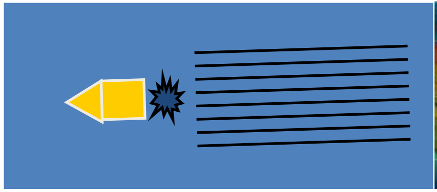
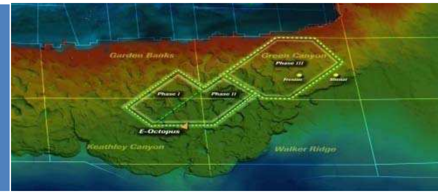
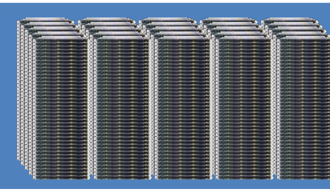


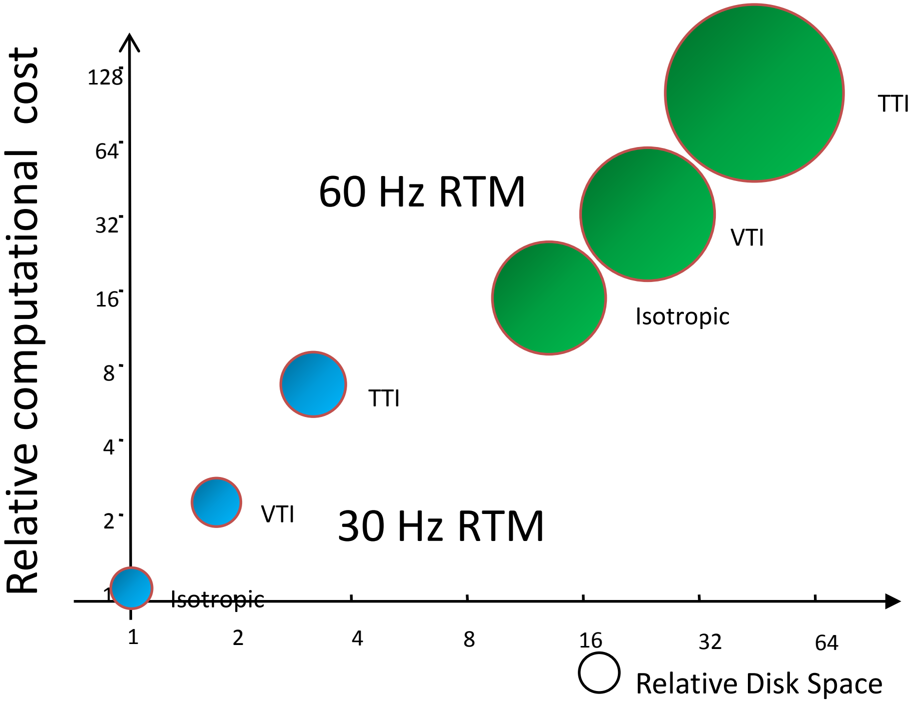
20 – 25,000 sensors 500 MB – 2 GB	50 – 200,000 shots 50 – 200 TB Data	1000s node 5 – 7 days	

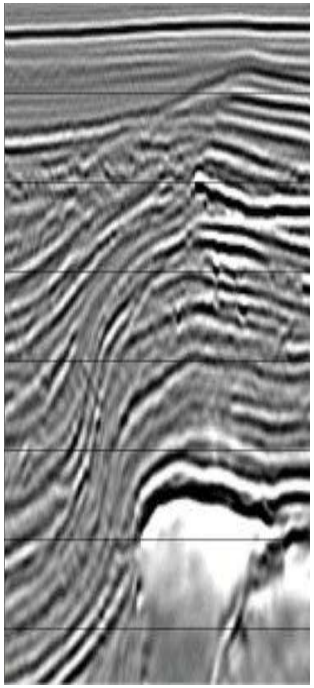


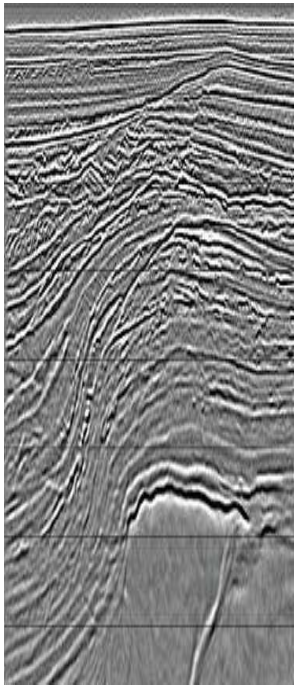


Data Rates and Computational needs

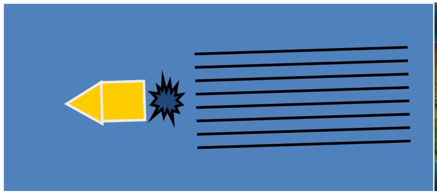
			
20 – 25,000 sensors 500 MB – 2 GB	50 – 200,000 shots 50 – 200 TB Data	1000s node 5 – 7 days	

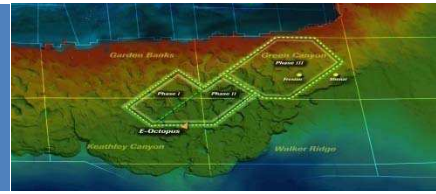


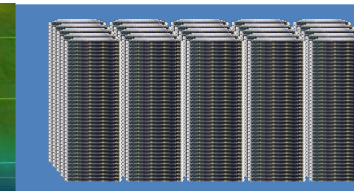


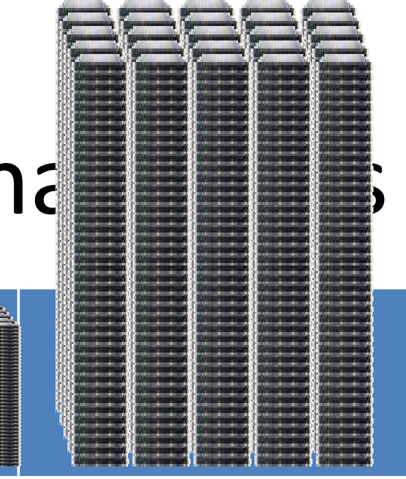


Data Rates and Computationa s

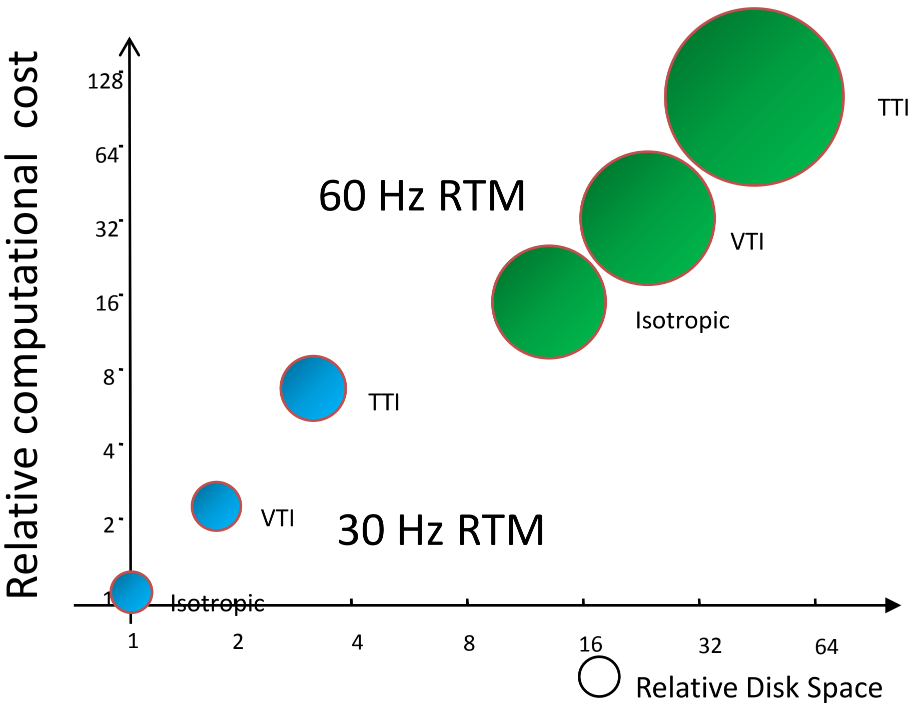


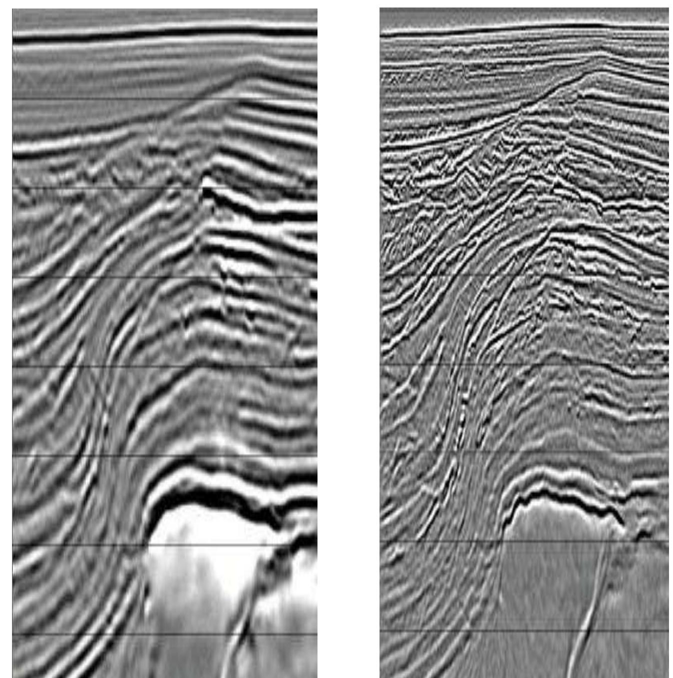




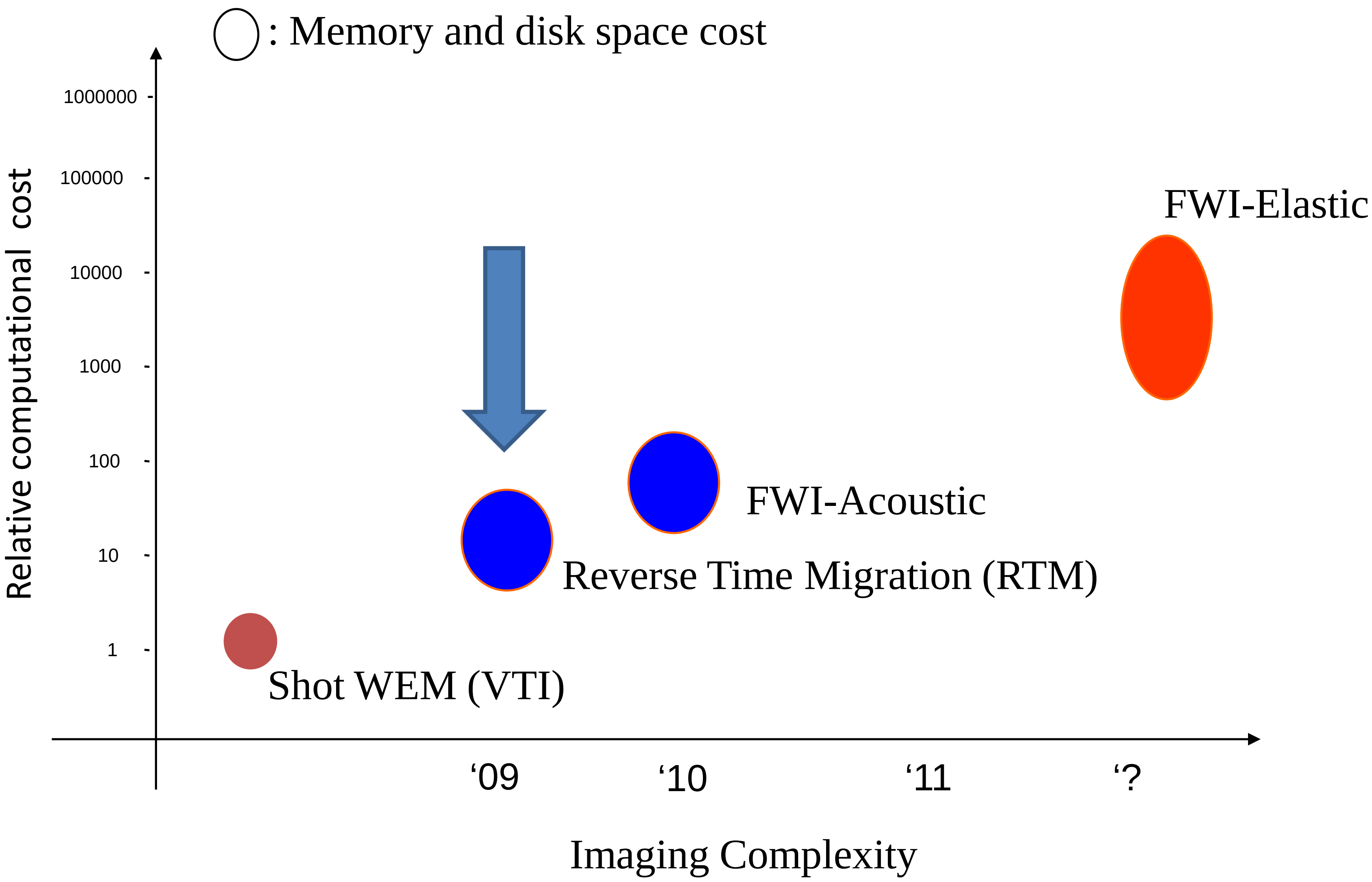


20 – 25,000 sensors 500 MB – 2 GB	50 – 200,000 shots 50 – 200 TB Data	1000s node 5 – 7 days	15 -20,000 nodes Days - weeks



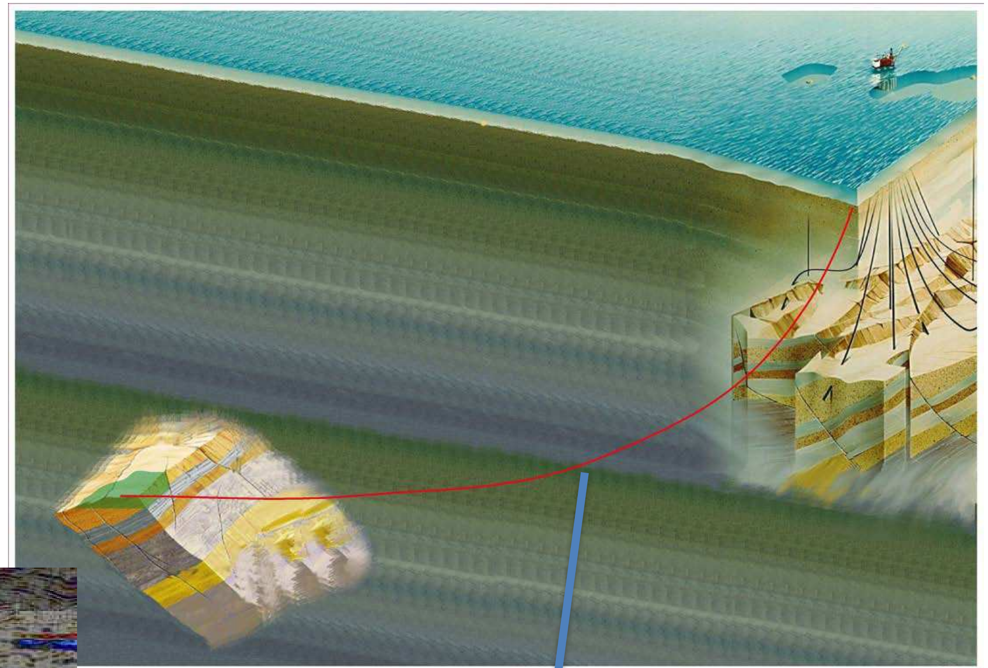


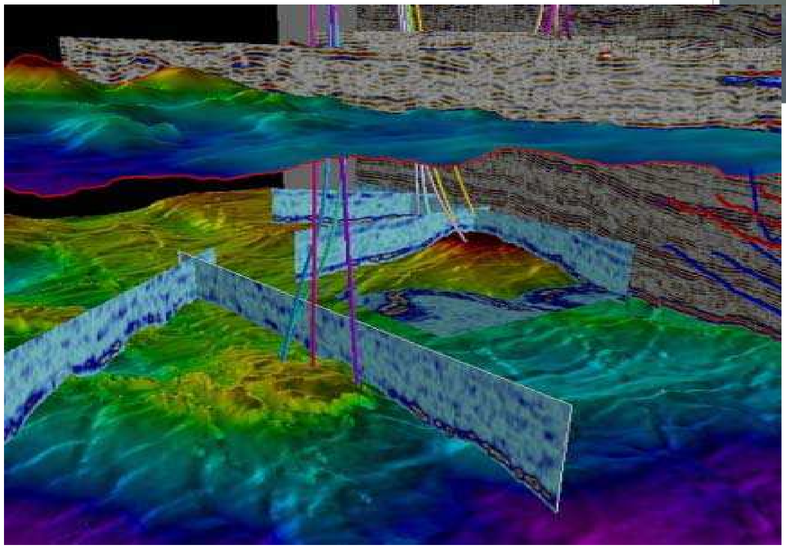
Cost of Imaging Algorithms

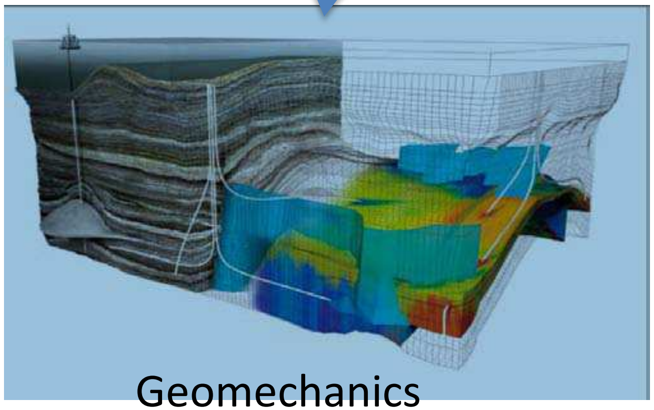


HPC – Its role in Hydrocarbon exploration

- Identify resources
- Access resources



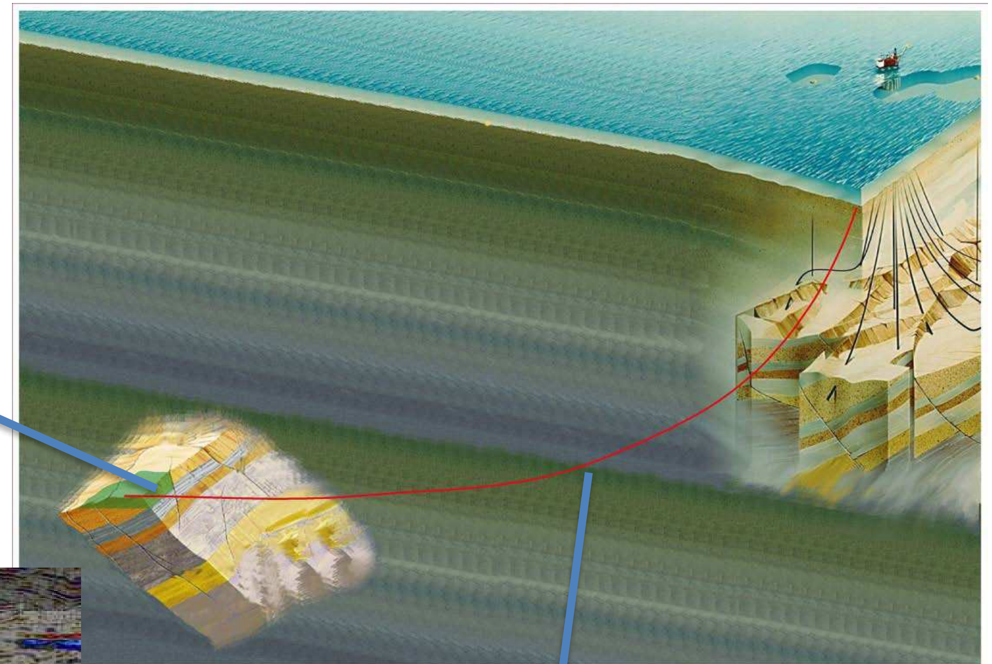


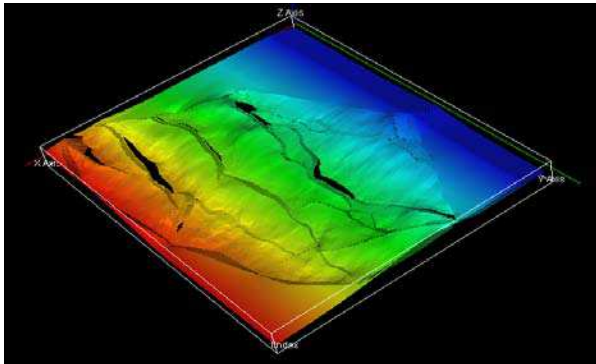


Geomechanics

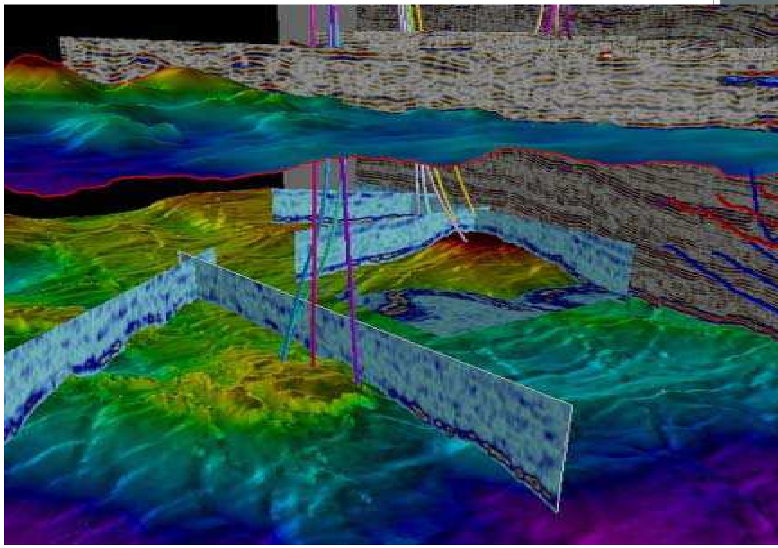
HPC – Its role in Hydrocarbon exploration

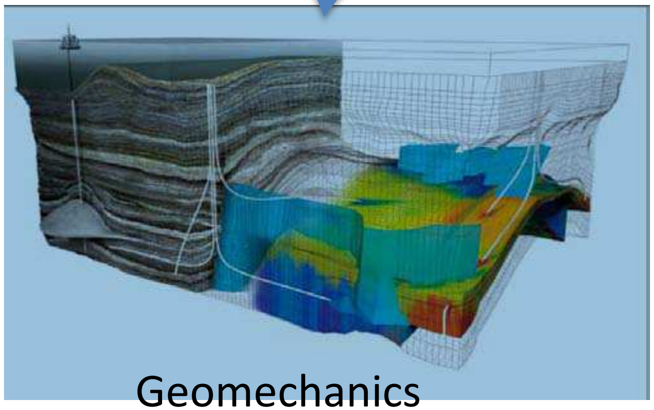
- Identify
-
- Maximize recovery





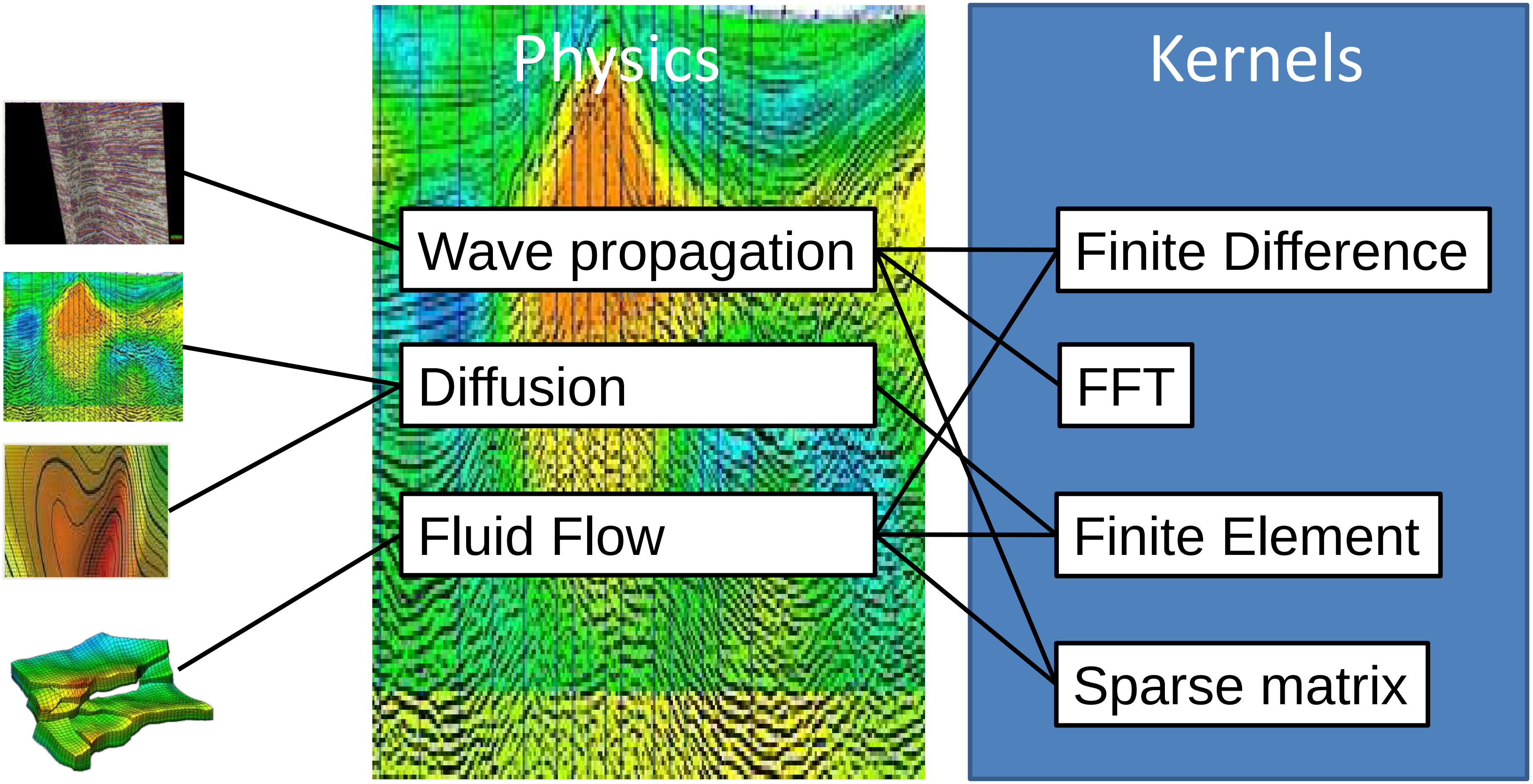
Reservoir Flow Simulation



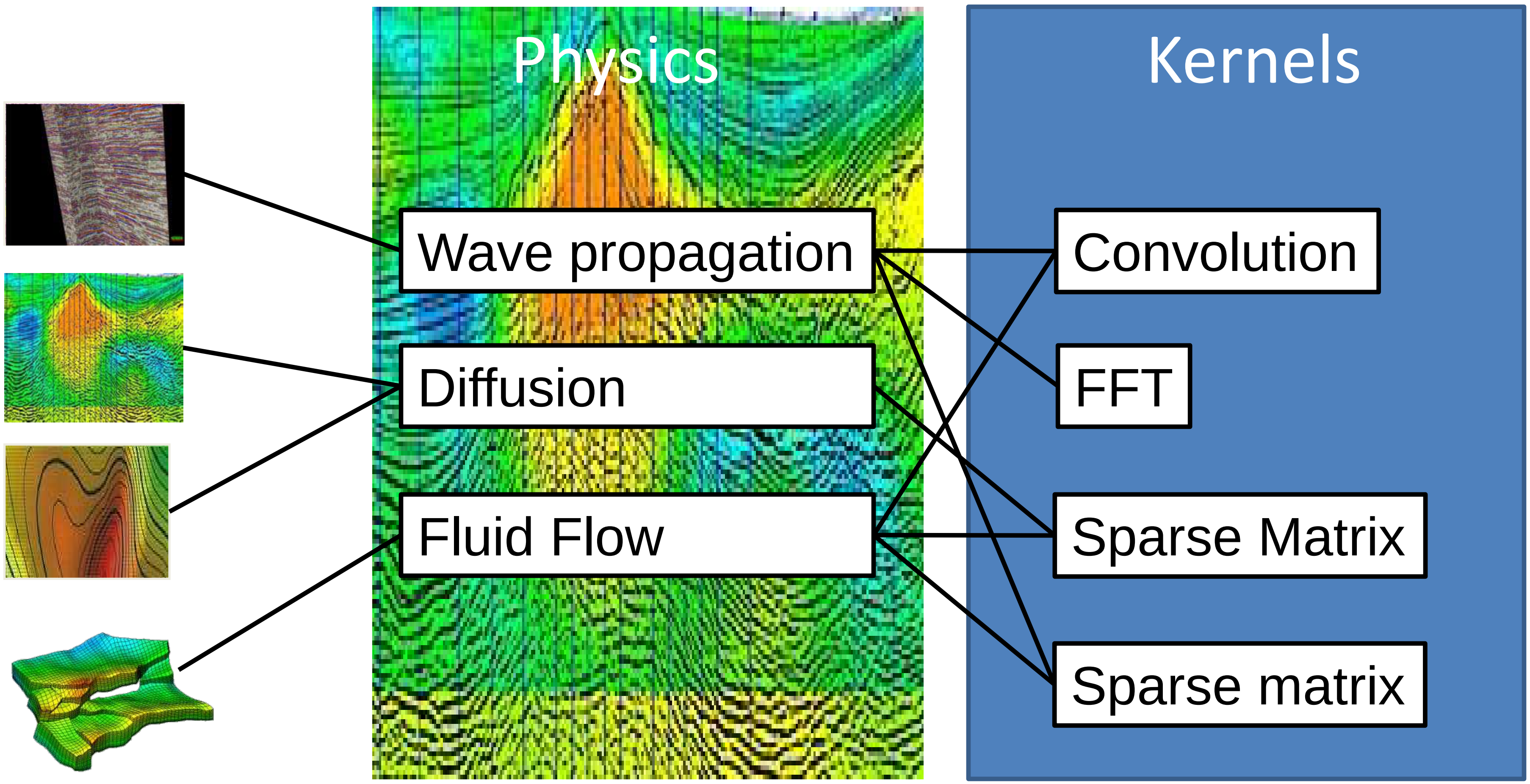


Geomechanics

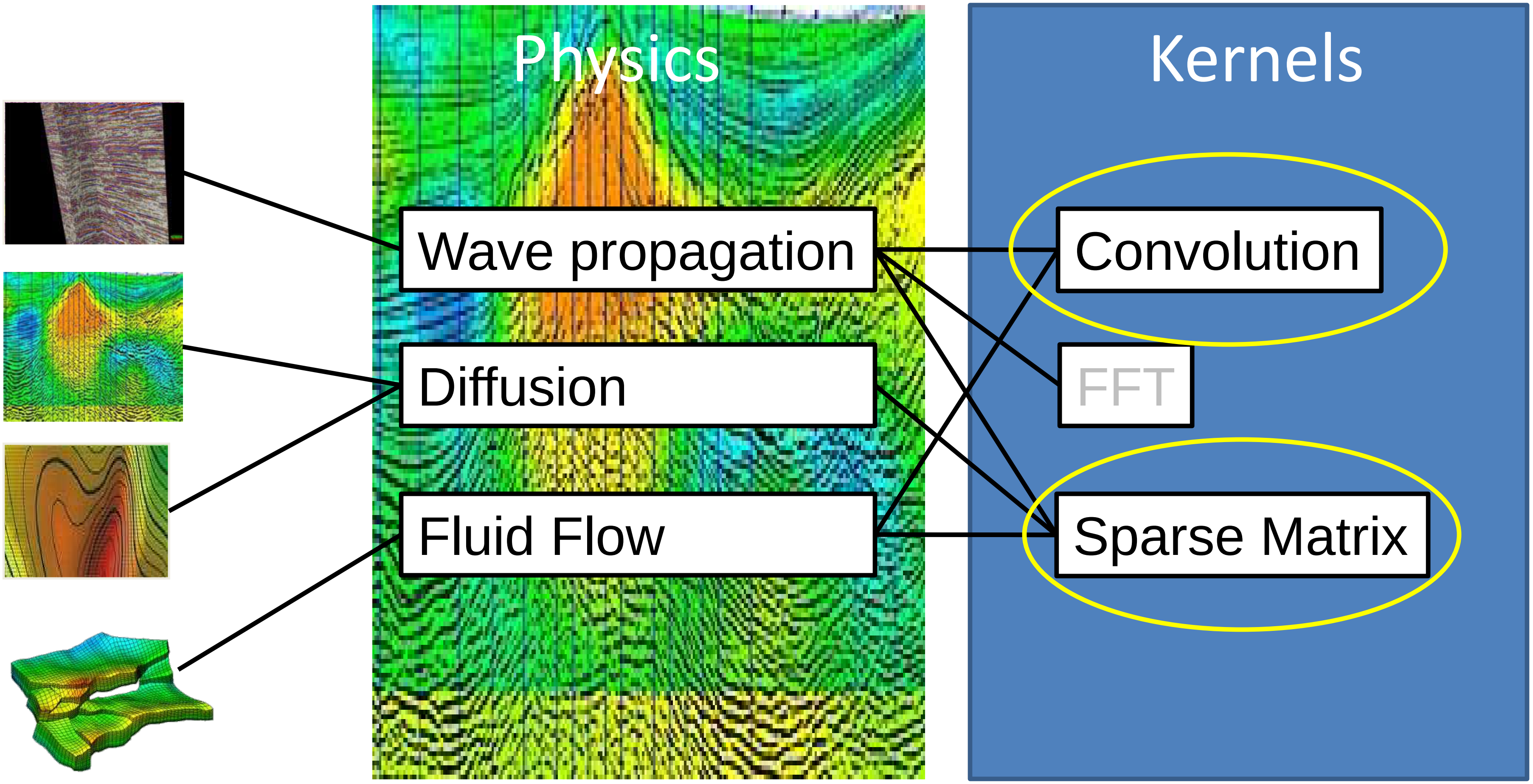
Oil and Gas Computational Kernels



Oil and Gas Computational Kernels



Oil and Gas Computational Kernels



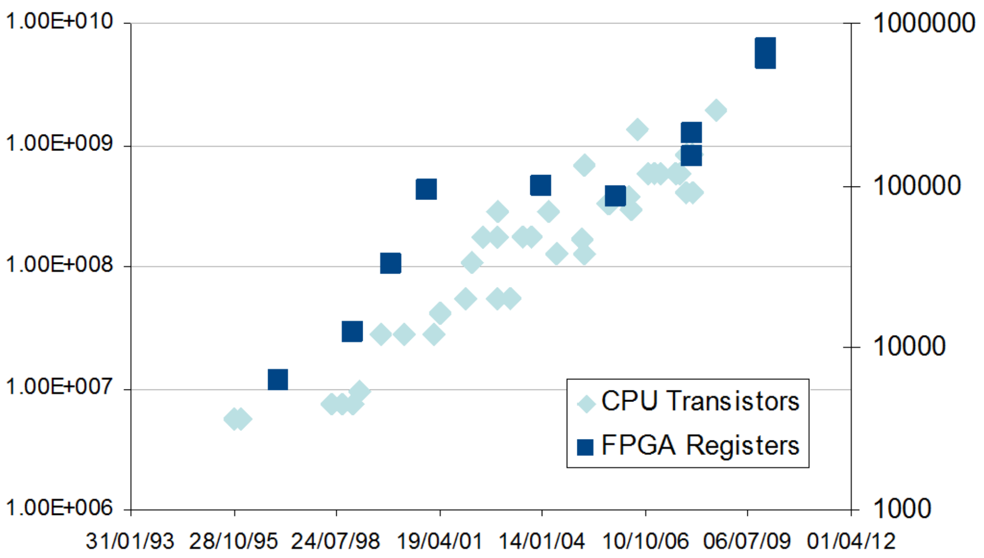
Outline

- Oil and Gas HPC applications
- Maxeler FPGA Compiler and Accelerators
- Key Computational Kernels in Geophysics
  - Sparse Matrix
  - Convolution based methods
- Applications scalability – Technology trends
- Conclusions

Accelerating Convolution and Sparse Matrix in the Maxeler Environment

Maxeler Accelerators

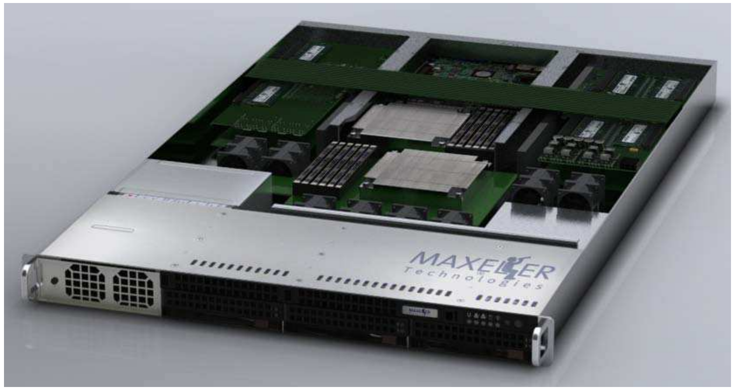
- Commodity silicon chips configurable to implement any digital circuit
  - ~10^6 small processing elements that operate in parallel
  - Several megabytes of on-chip memory
  - Run at several hundred megahertz
  - Support large on-board memory (24GB+)

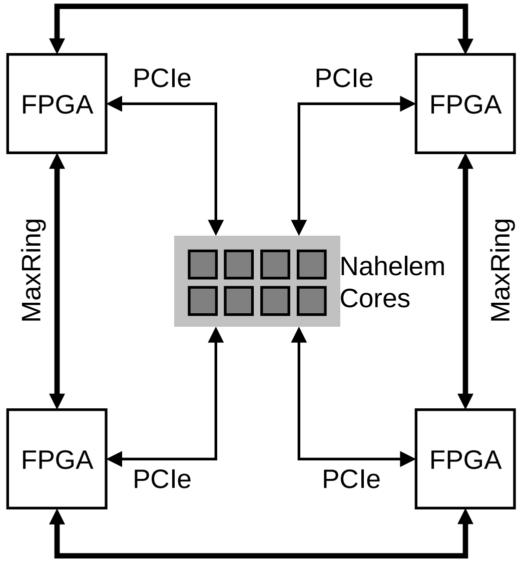


MaxNode with MAX3

Specifications:

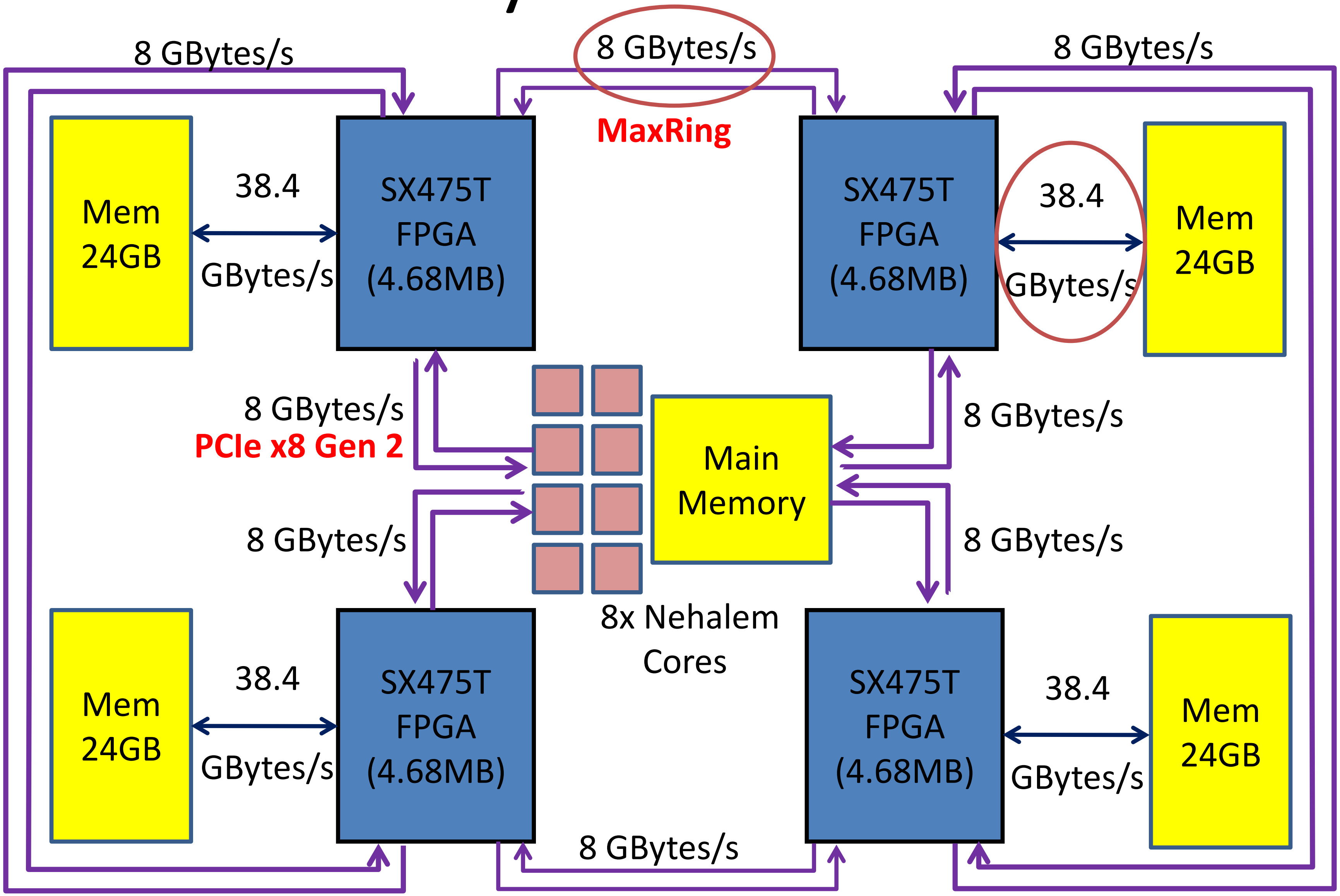
Compute	8x 2.8GHz Nahelem Cores 4x Virtex 6-SX475T FPGAs
Interconnect	PCI-Express Gen. 2 MaxRing Gigabit Ethernet
Storage	3x 2TB Hard disks
Memory	96GB DRAM
Form Factor	1U





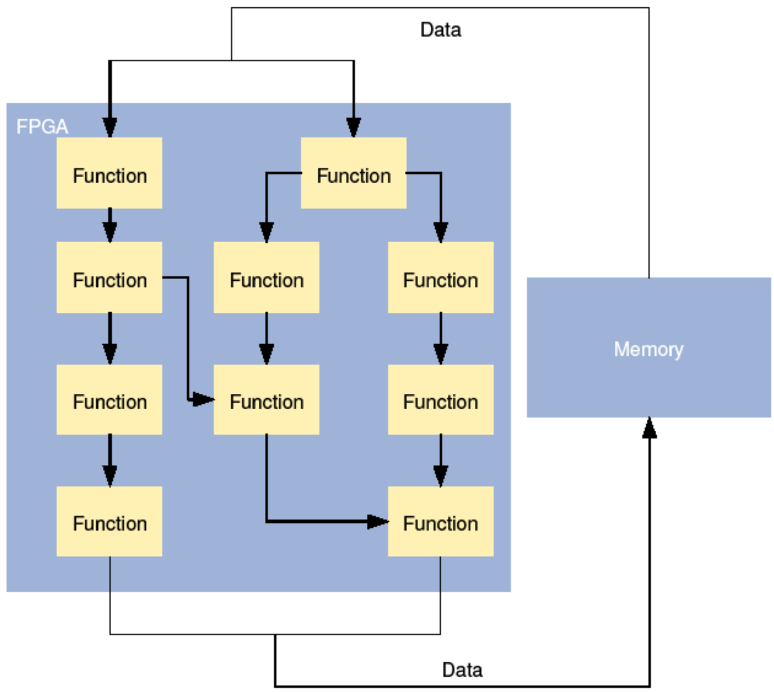
MAX3 Node Architecture

MAX3 System Bandwidths

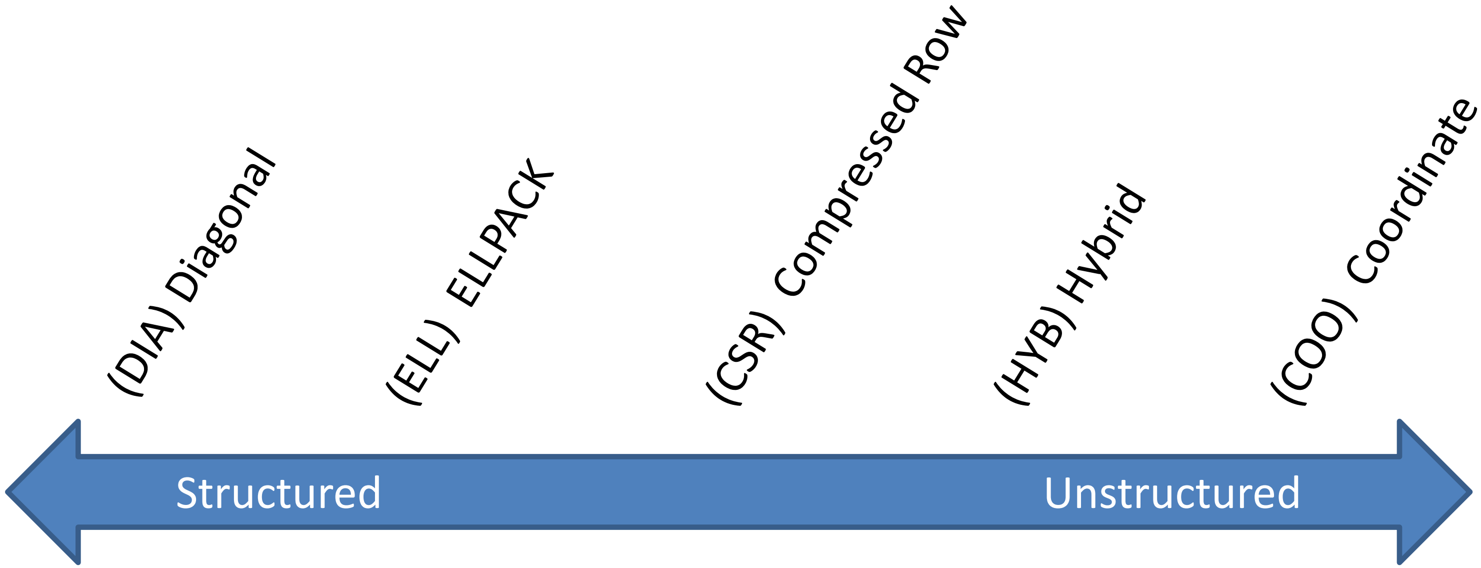


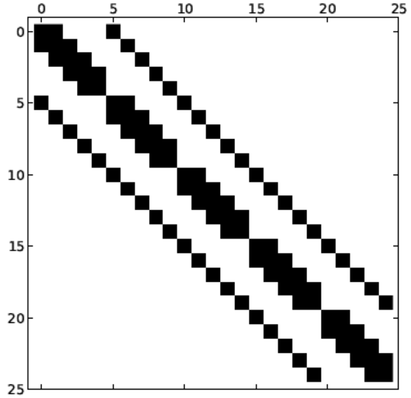
Maxeler Programming Paradigm

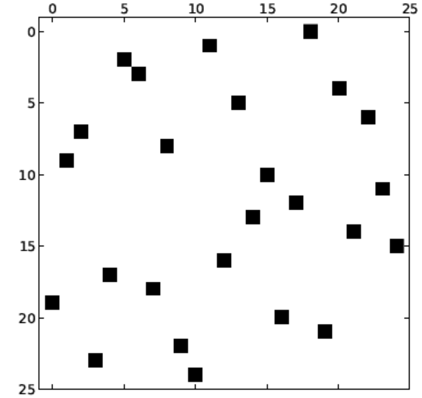
```java
public class MovingAverageKernel extends Kernel {

    public MovingAverageKernel(KernelParameters parameters, int N) {
        super(parameters);

        // Input
        HWVar x = io.input("x", hwFloat(8, 24));

        // Data
        HWVar prev = stream.offset(x, −1);

        HWVar next = stream.offset(x, 1);

        HWVar sum = prev+x+next;

        HWVar result = sum/3;

        // Output
        io.output("y", result, hwFloat(8, 24));
    }
}
```



Sparse Matrix Format



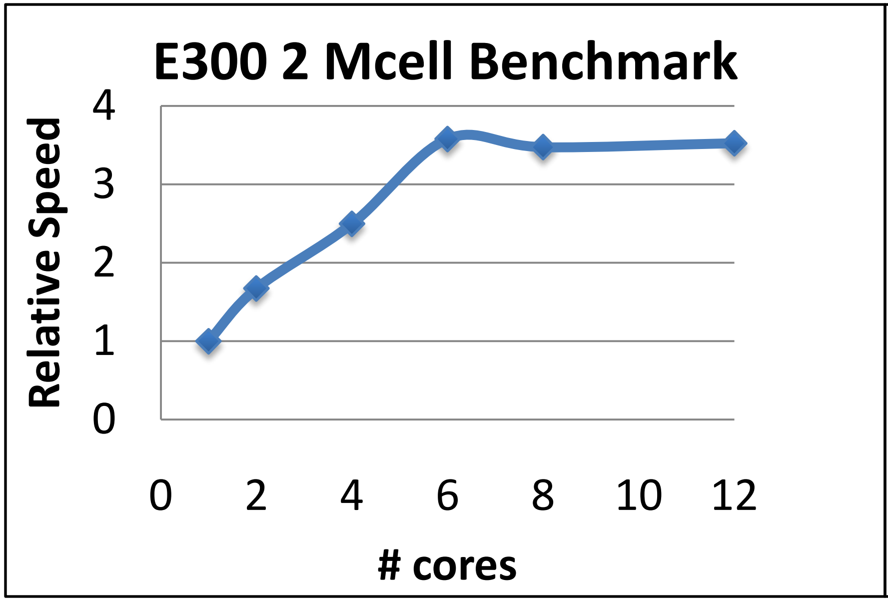




Typical scalability of SLB Sparse Matrix Applications

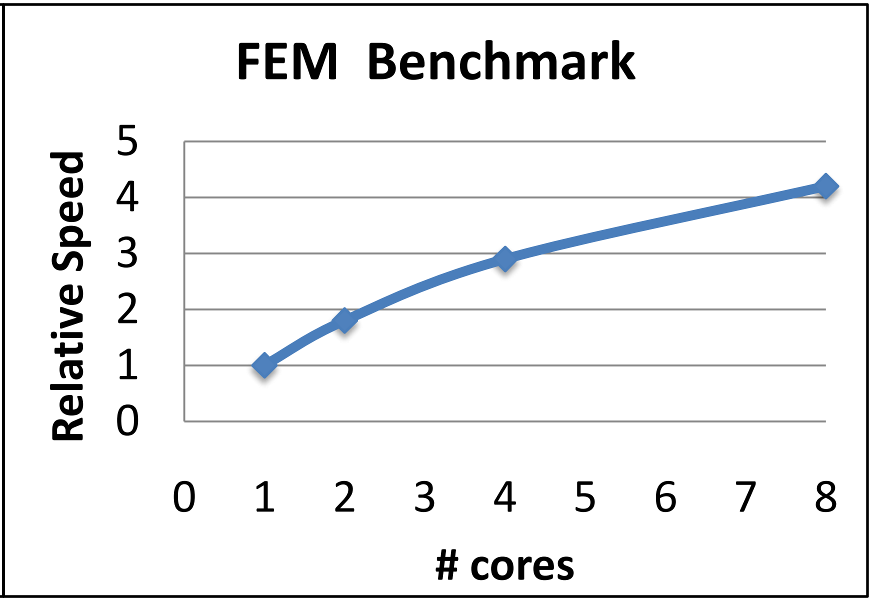
Eclipse Benchmark

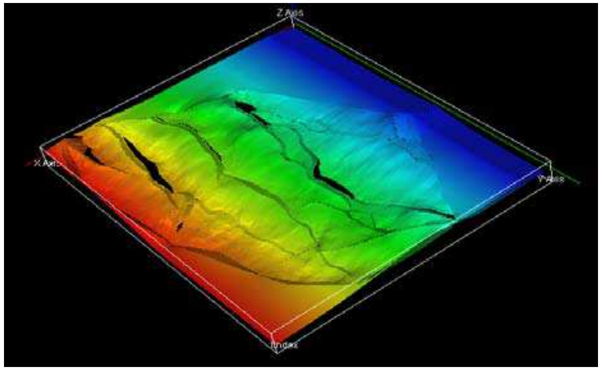
(2 node Westmere 3.06 GHz)

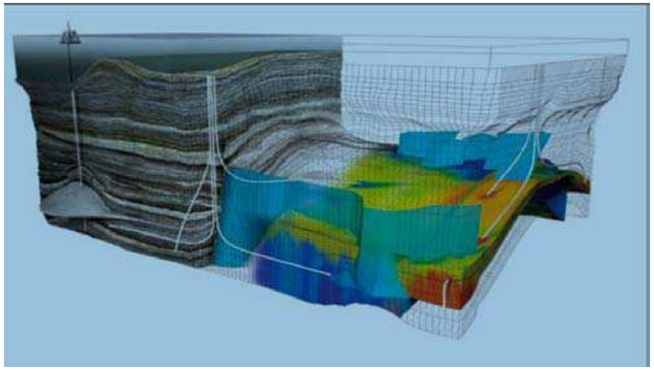


Visage – Geomechanics

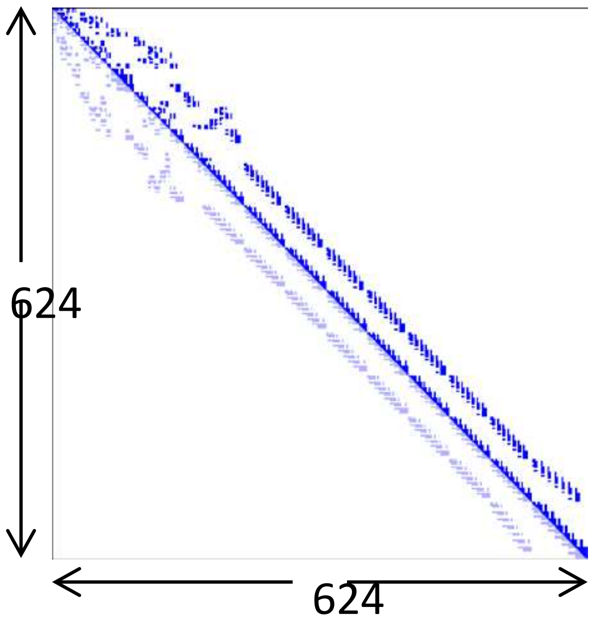
(2 node Nehalem 2.93 GHz)



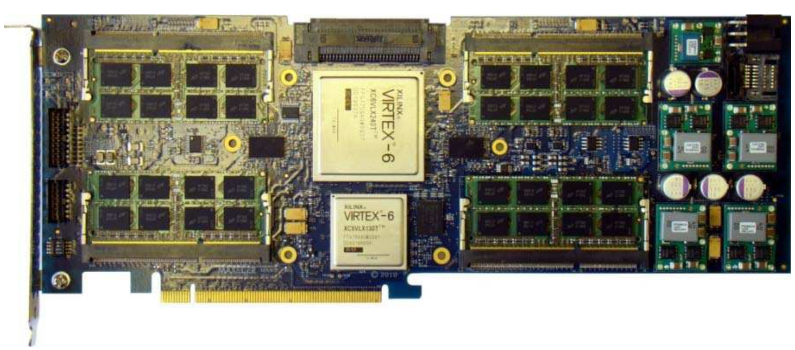


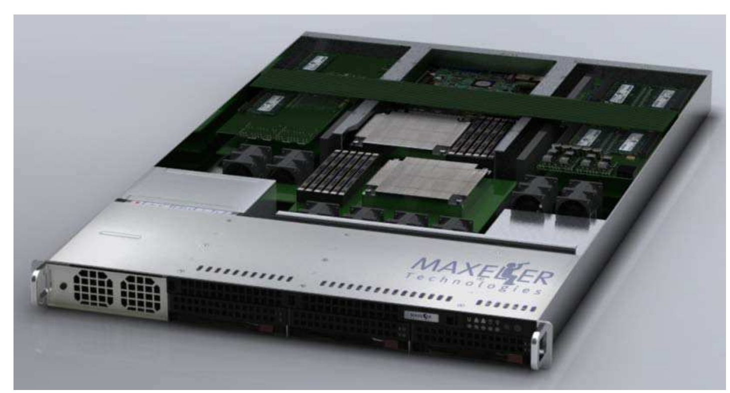


Sparse Matrix on FPGAs

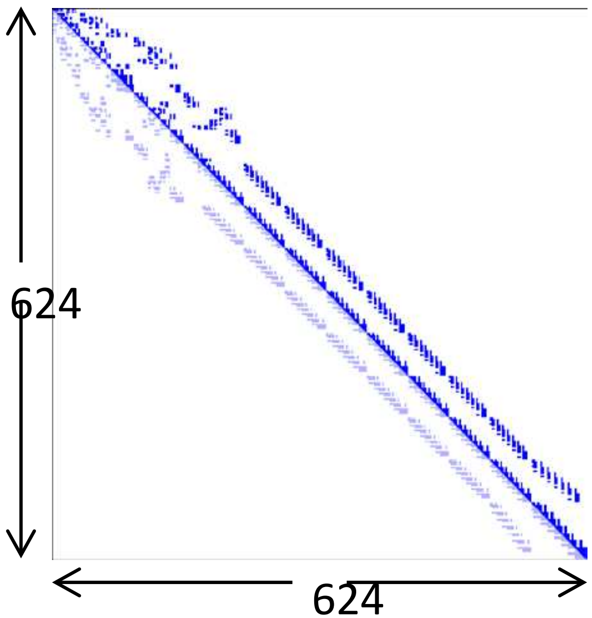


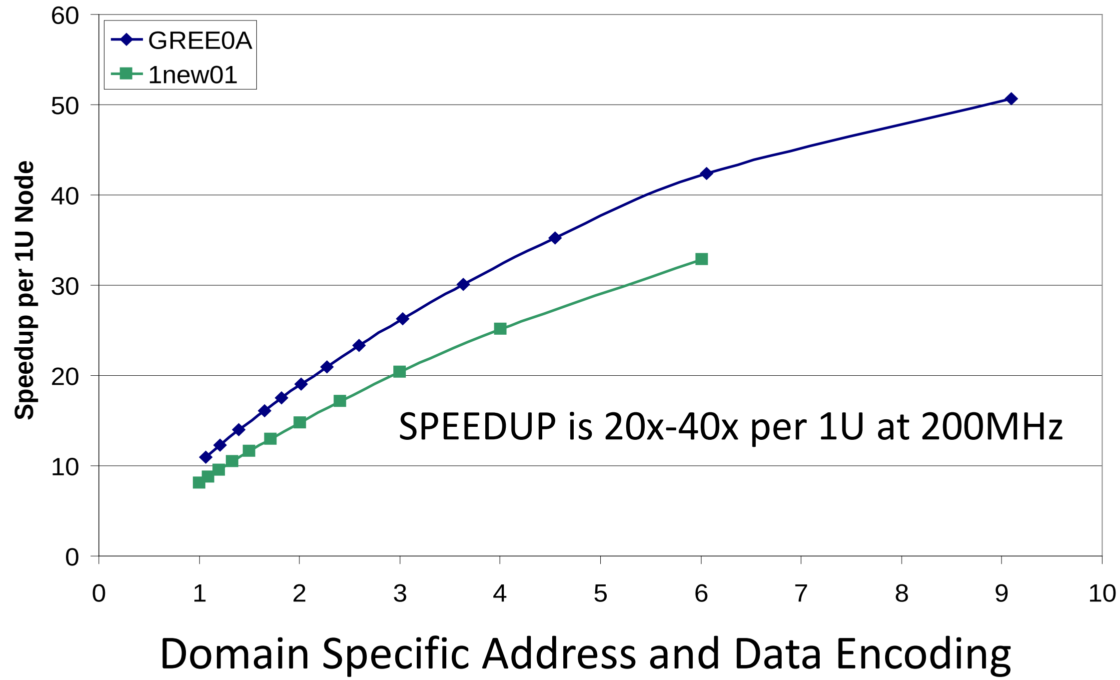
- 4 MB BLK RAM
- Pipelining
- Addressing scheme optimized for Matrix structure
- Domain Specific Data Encoding

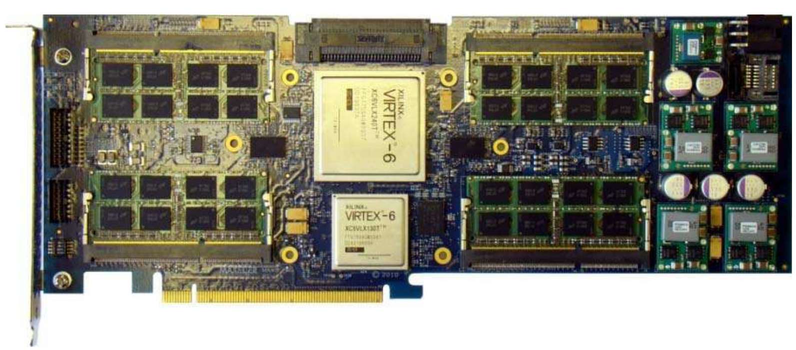


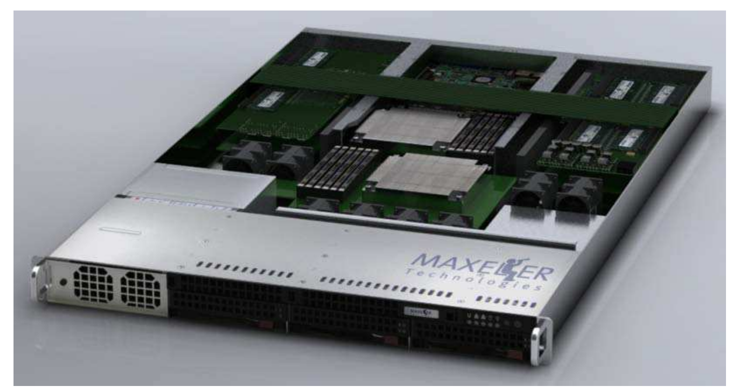


Sparse Matrix on FPGAs

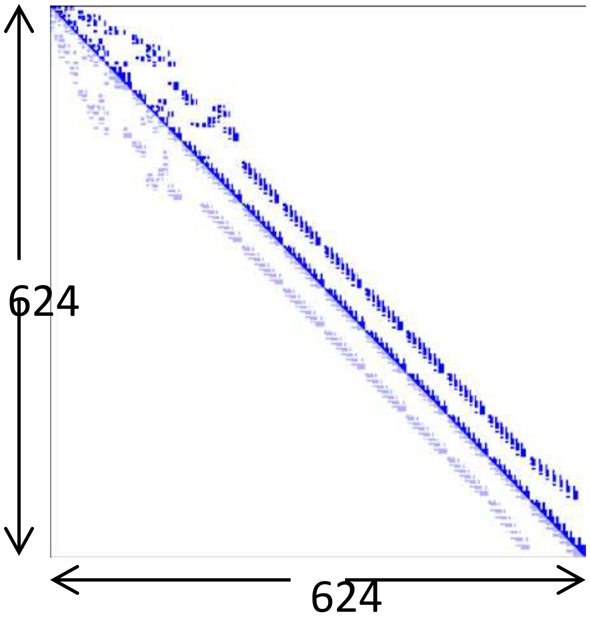


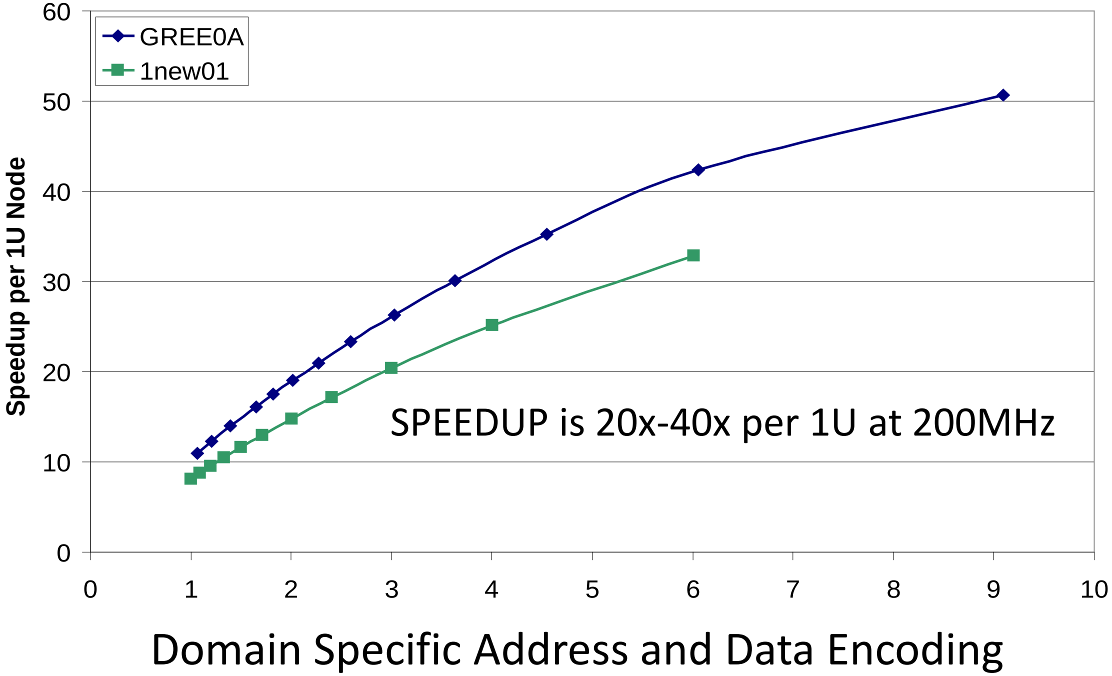


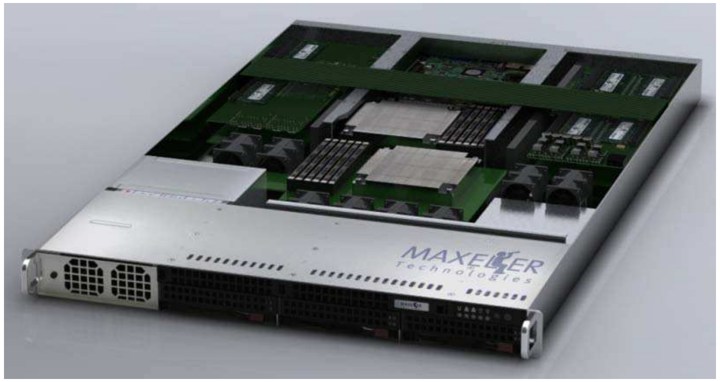




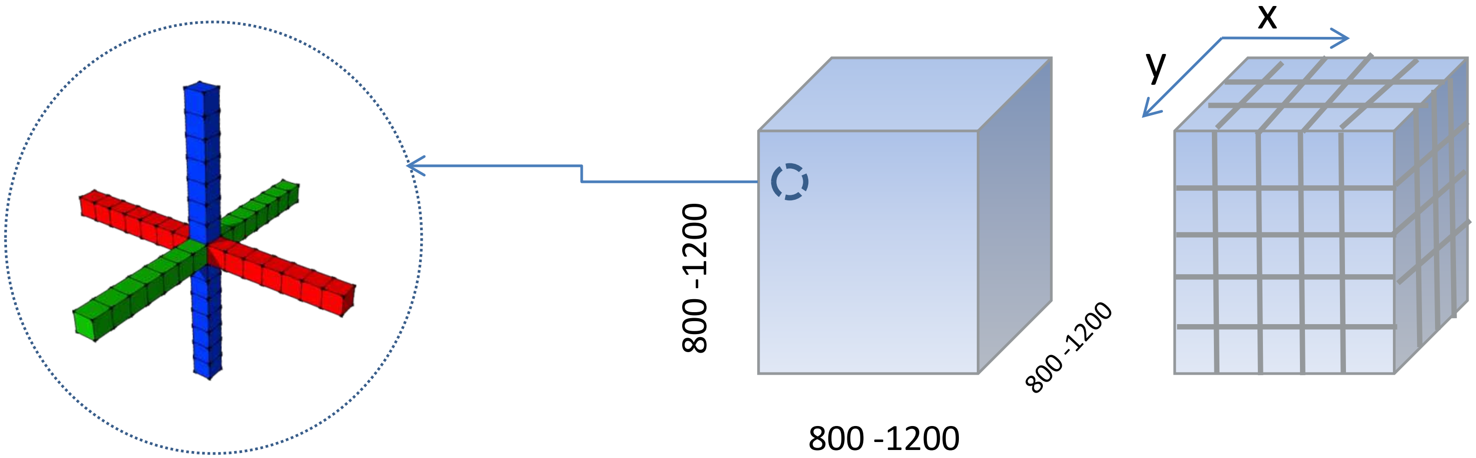
Sparse Matrix on FPGAs







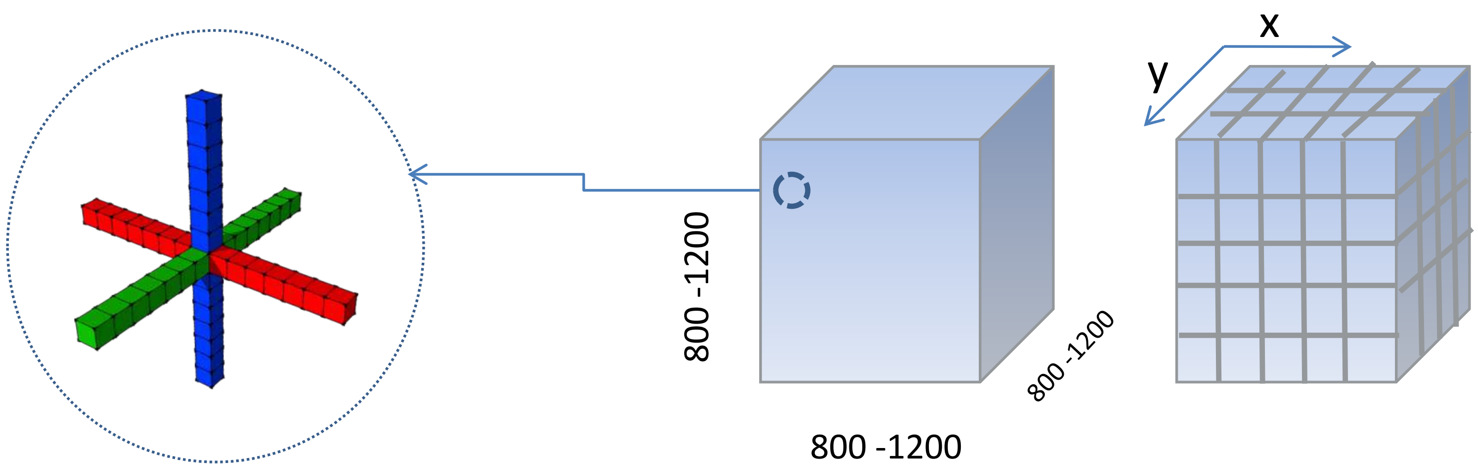
3D Convolution



- Low Flop/Byte ratio
- Sparse structure requires large streaming memory buffers ($14 \times nx \times ny$ for 14th order in space).
- Data Structure >> Data Caches

- CPUs:
- Constrained by:
  - Small L1/L2 cache
  - Limited utilization of pipeline
  - Limited by Streaming BW
  - Limited data element reuse
  - → Fraction of peak performance

FPGA Opportunities



- FPGA opportunities
- 4 MB on-chip Memory
- Hundreds of pipeline stages
- Optimal trade off between streams for BW utilization and Pipe line depth

- CPU limits:
- Constrained by:
  - Small L1/L2 cache
  - Limited depth of pipeline
  - Limited by Streaming BW
  - Limited data element reuse
  - → Fraction of peak performance

Performance

Algorithm	**Hardware**	**Design**	**Speedup** 8-core Nehalem 2.93 GHz 1U server *vs* 1U MaxNode
Star stencil	VIRTEX 5	3 pipe	20x
Star stencil	VIRTEX 6	9 pipe	73x

Outline

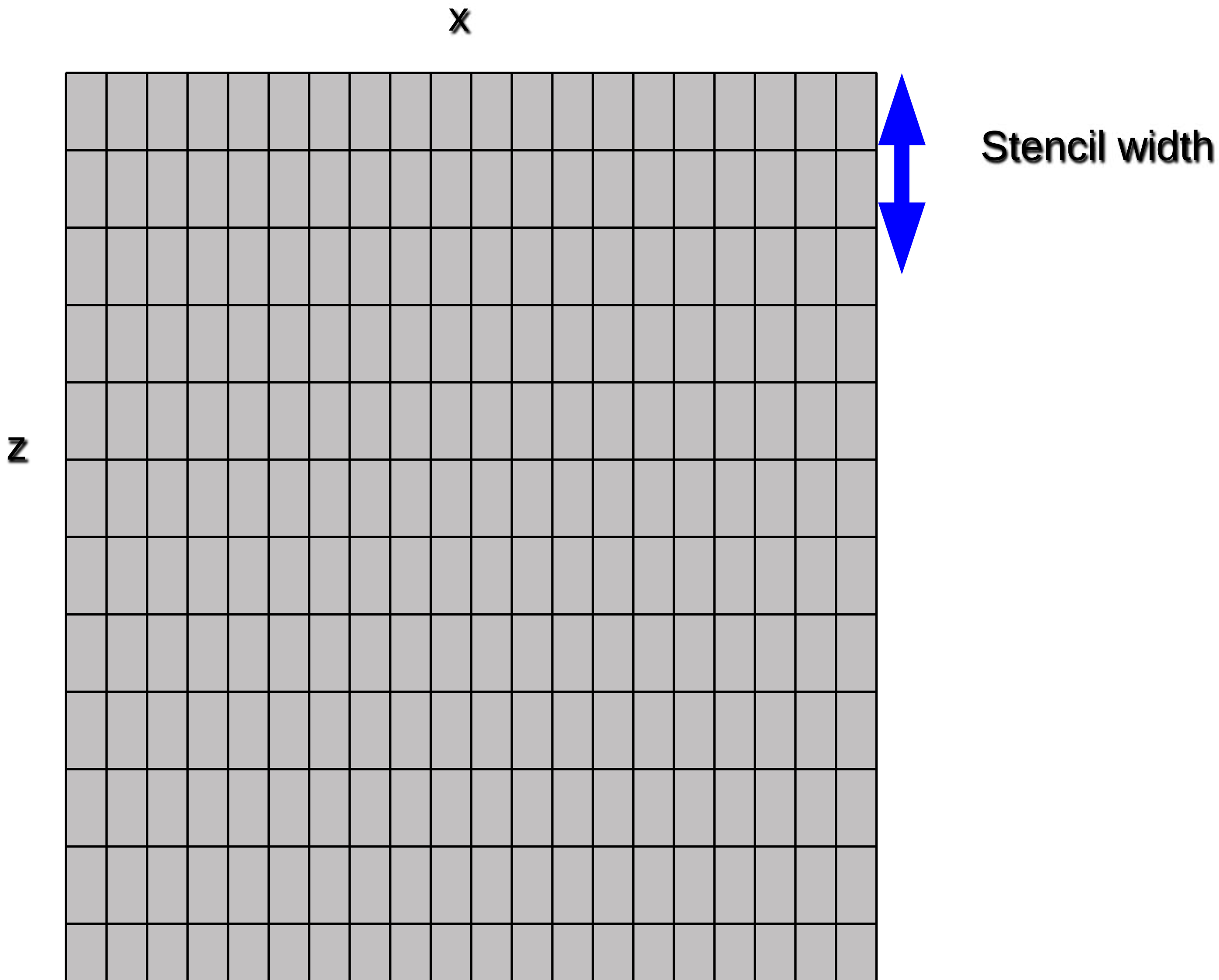
- Oil and Gas HPC applications
- Maxeler FPGA Compiler and Accelerators
- Key Computational Kernels in Geophysics
  – Sparse Matrix
  – Convolution based methods
- Applications scalability – Technology trends
- Conclusions

Application scalability and Technology trends

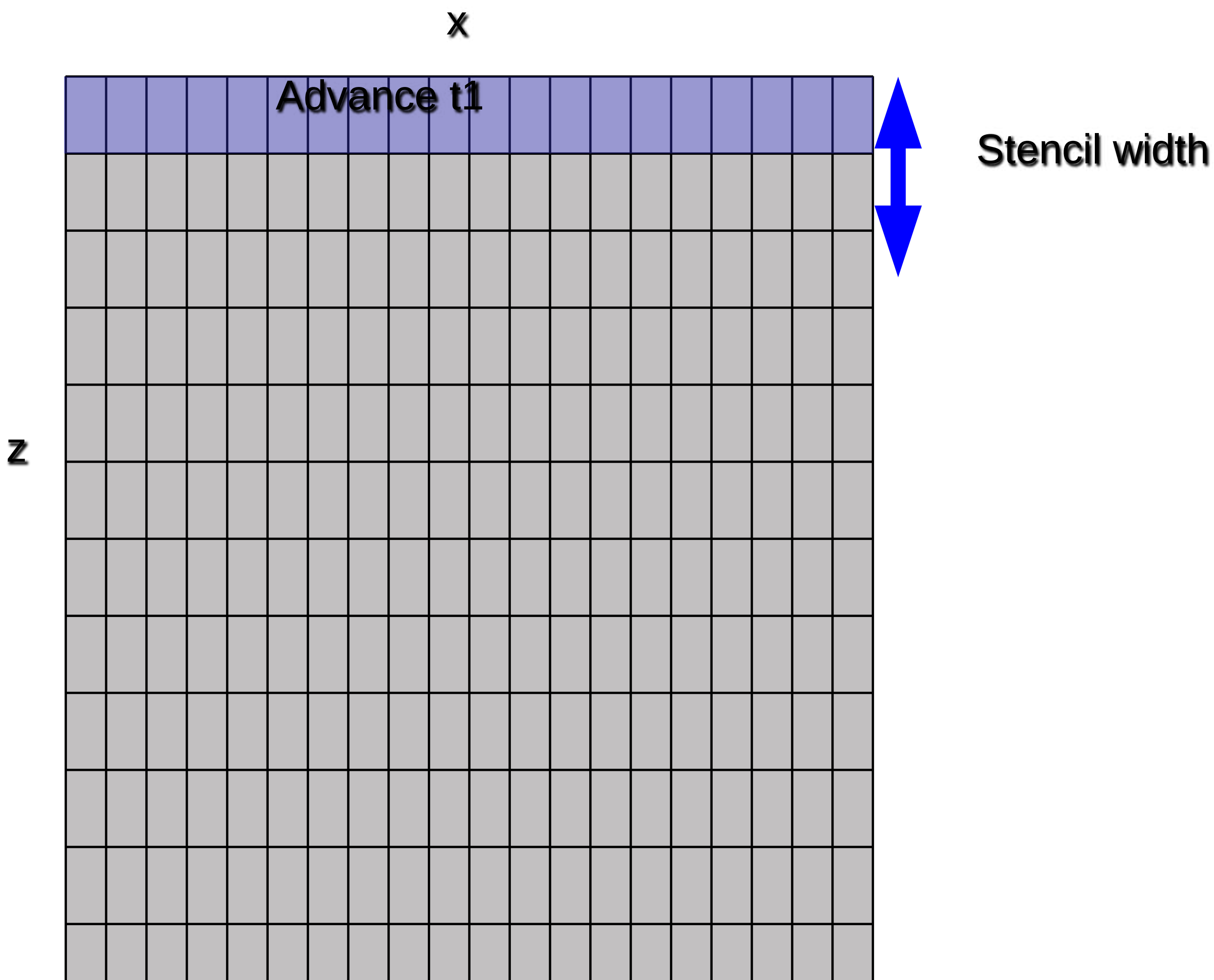
- Transistor count keeps increasing
- Memory BW continues to trail
- How will our algorithms scale?

- Convolution:
  - Deeper pipelines:
    - An example: Cascading multiple time steps
  - Specialized macros on FPGAs

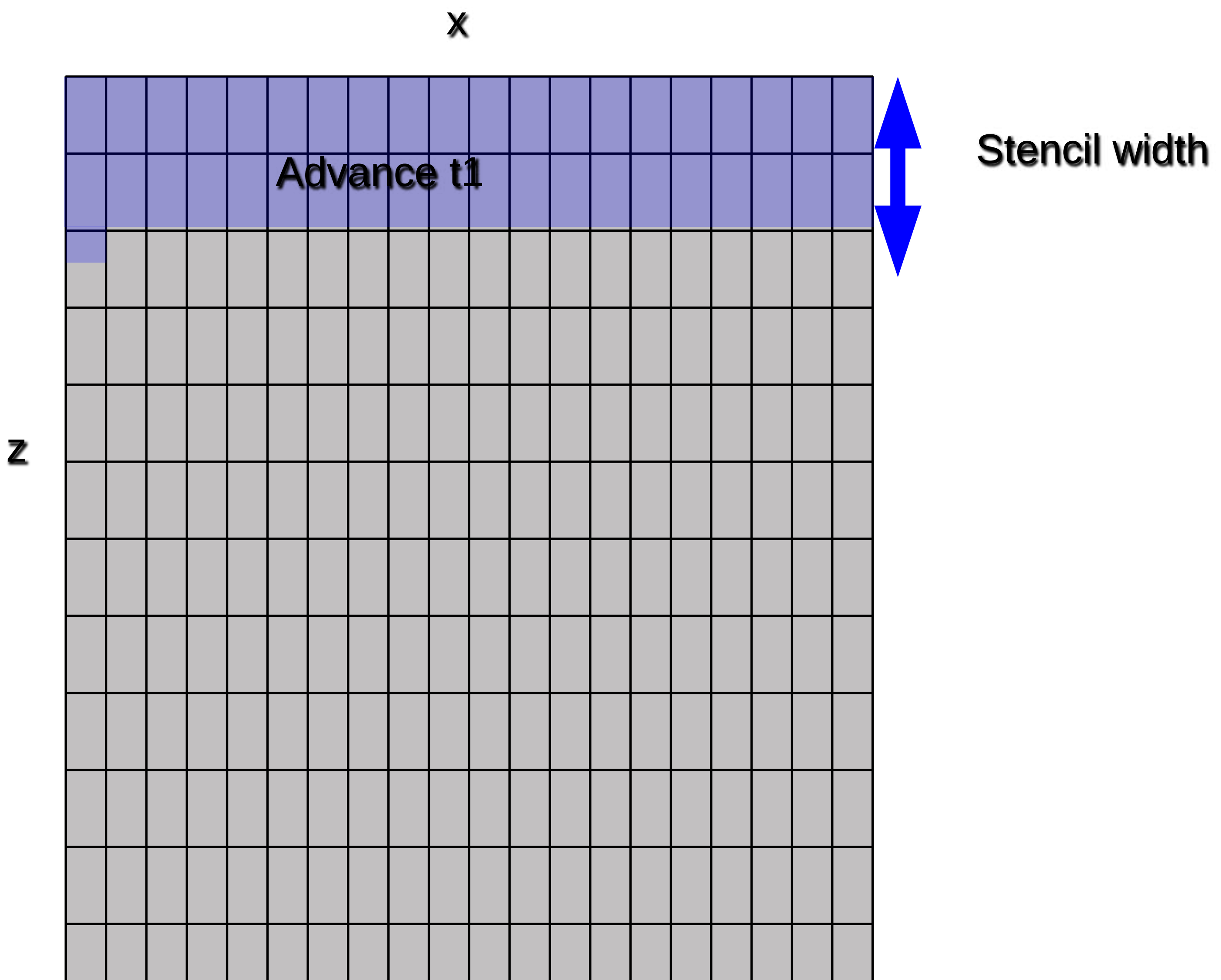
FPGA: Time step Cascading



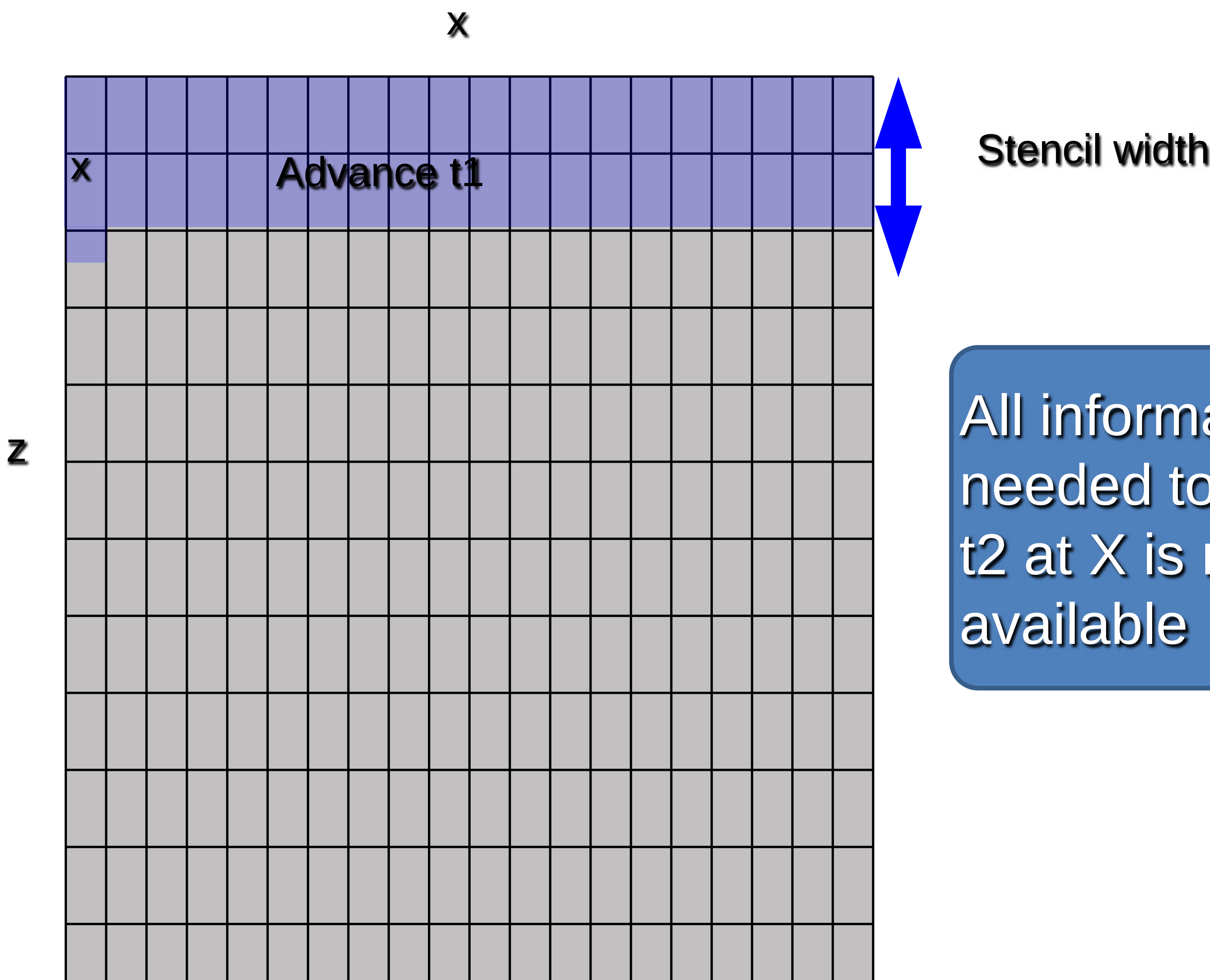
FPGA: Time step Cascading



FPGA: Time step Cascading

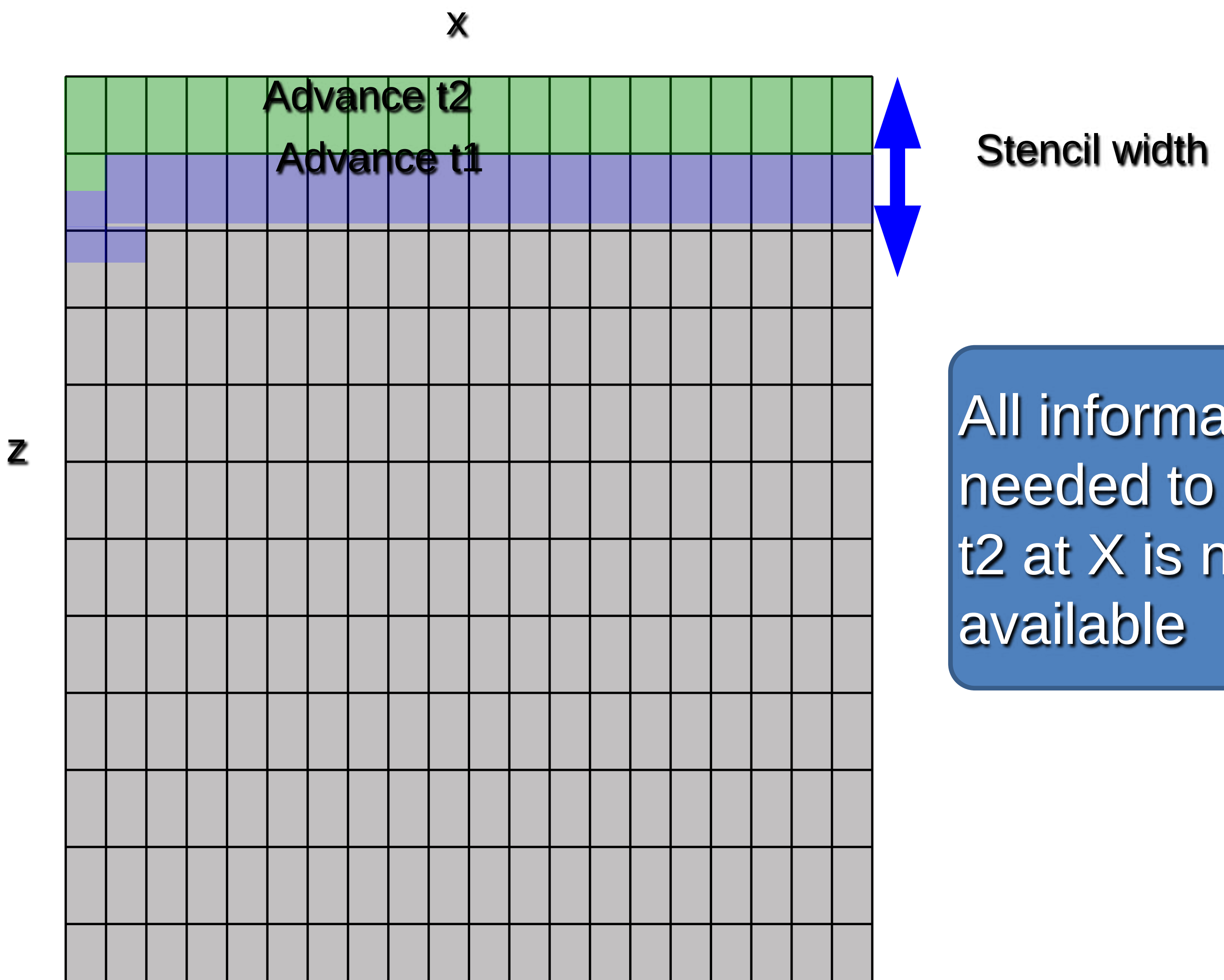


FPGA: Time step Cascading



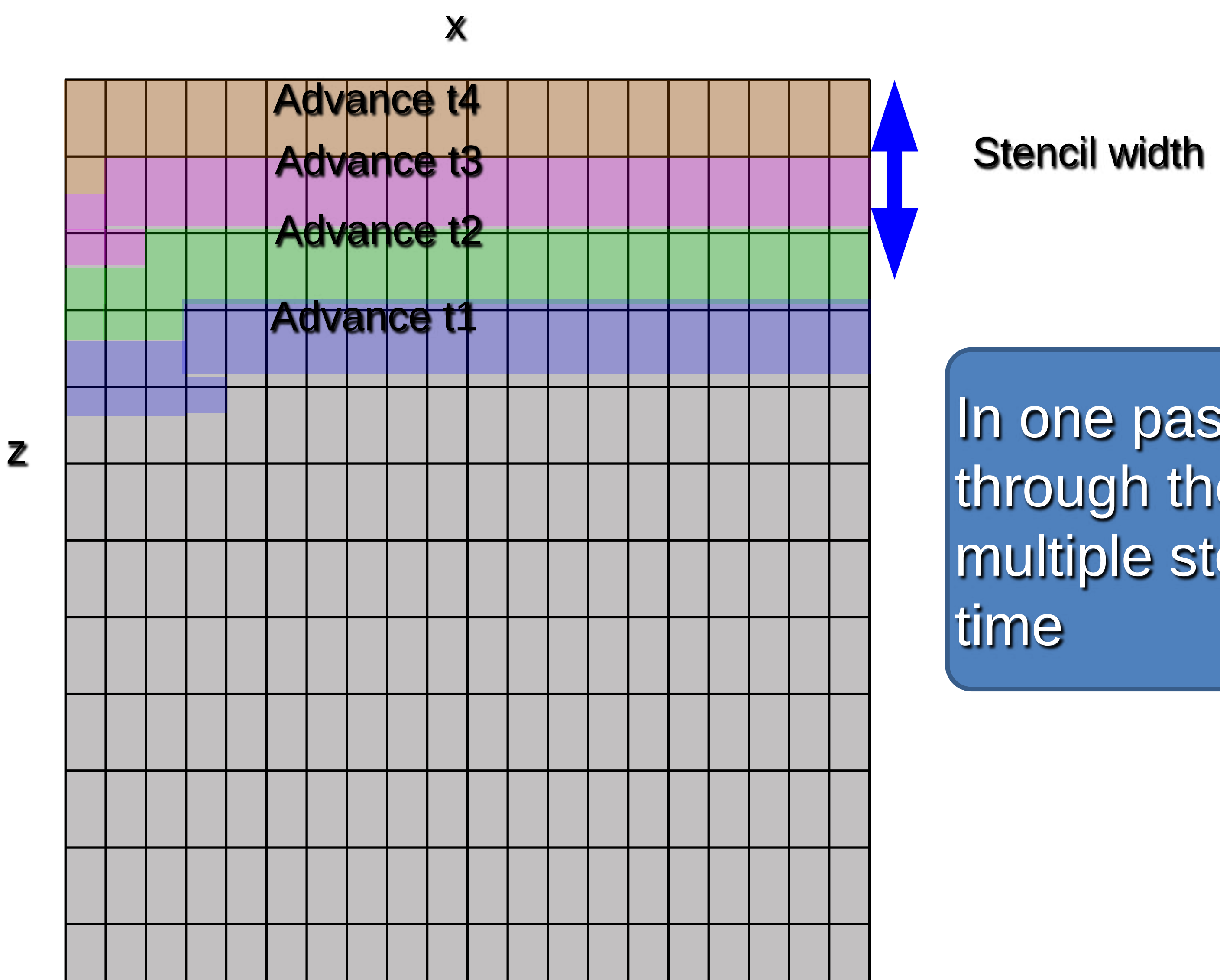
All information needed to update t2 at X is now available

FPGA: Time step Cascading



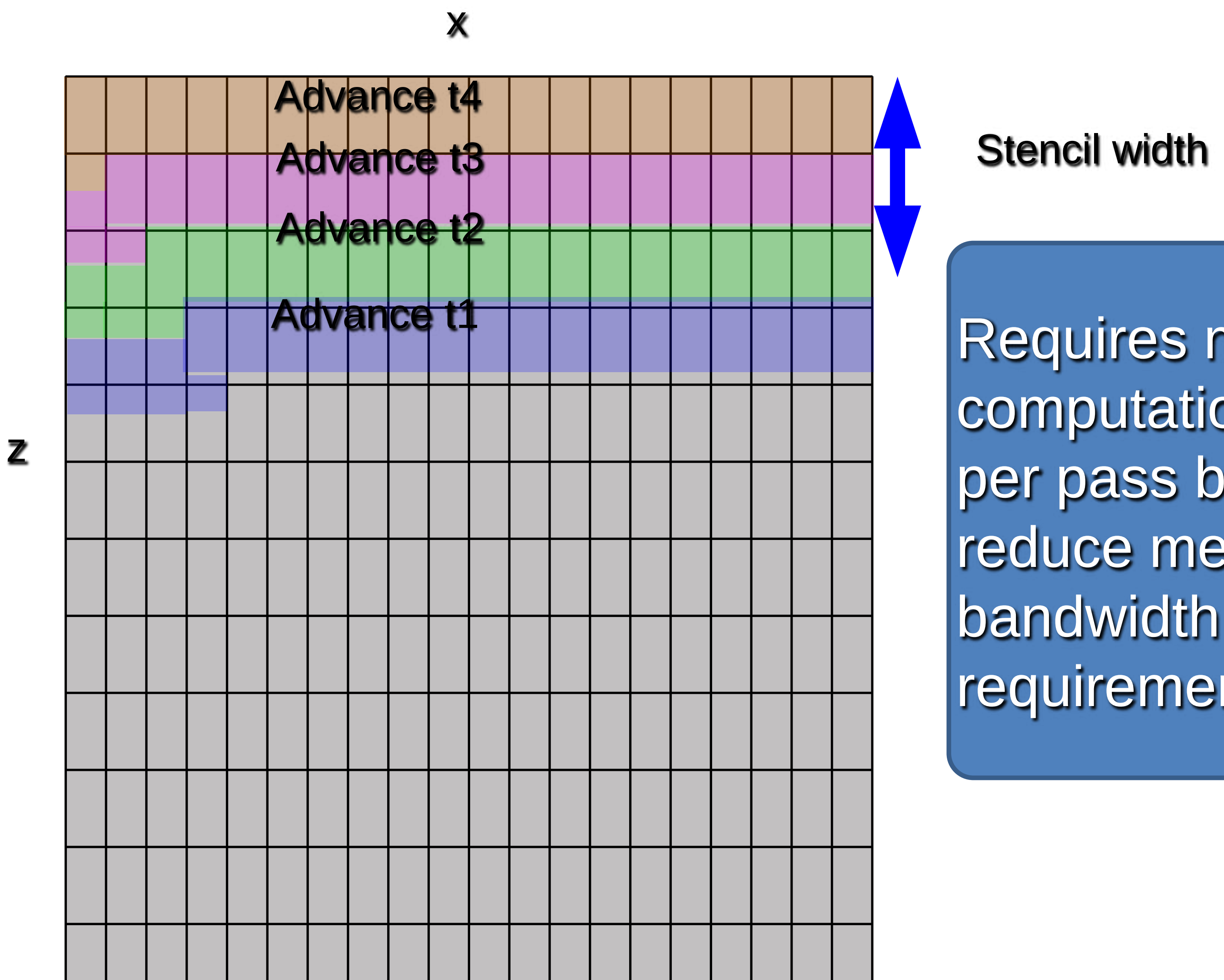
All information needed to update t2 at X is now available

FPGA: Time step Cascading



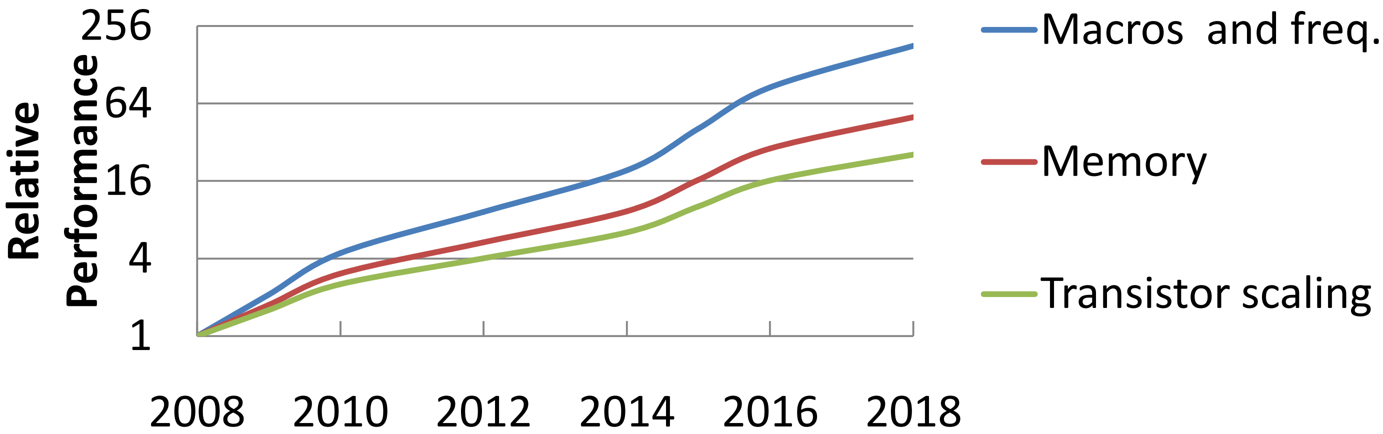
In one pass through the data to multiple steps in time

FPGA: Time step Cascading



Requires more computational units per pass but reduce memory bandwidth requirements

Technology opportunities



- Added Resources (Transistor scaling) translates directly into performance using Multiple time step techniques
- Independent of Memory BW increase

Resource costs for a symmetric 15-point stencil:

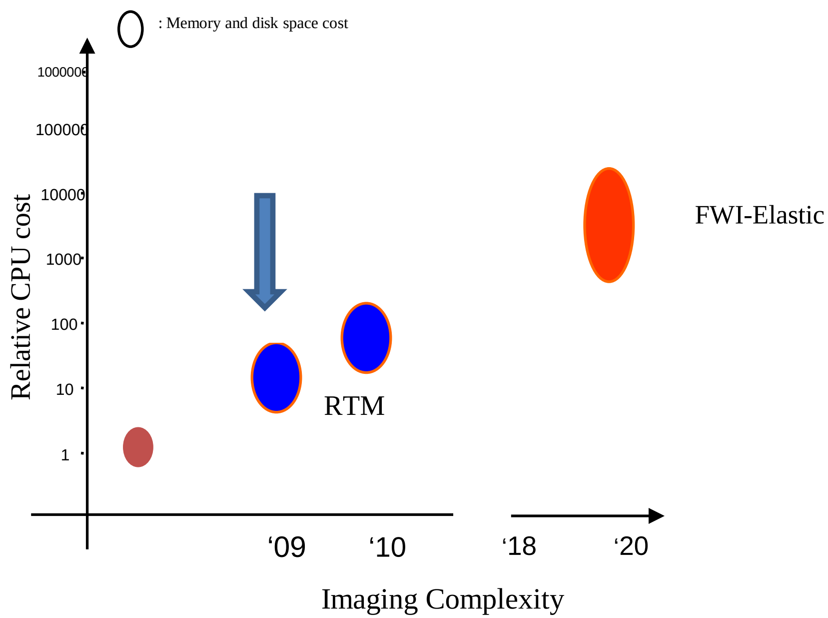
	LUT/FFs	DSPs
MaxGenFD on Virtex-5	207	8
MaxGenFD on Virtex-6	33	8
Resulting perf. improvement	50 %	

Virtex-6 DSP enhanced with Pre-Adder

Outline

- Oil and Gas HPC applications
- Maxeler FPGA Compiler and Accelerators
- Key Computational Kernels in Geophysics
  – Sparse Matrix
  – Convolution based methods
- Applications scalability – Technology trends
- Conclusions

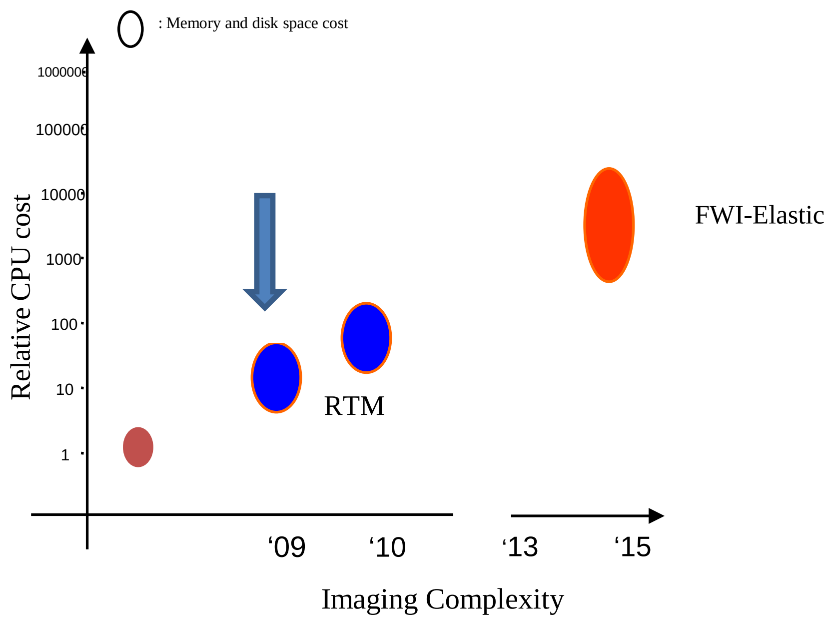
Surviving the End of Scaling of Traditional Micro Processors in HPC



Conventional Road Map

- Conclusions:
  - FPGA Streaming has come of age
  - Development Environment is here today
  - Application will scale with predicted technology evolution
  - Considerable upside for "smart macros"

Surviving the End of Scaling of Traditional Micro Processors in HPC



FPGA road maps

- Conclusions:
  - FPGA Streaming has come of age
  - Development Environment is here today
  - Application will scale with predicted technology evolution
  - Considerable upside for "smart macros"

Thank You

GPU Comments