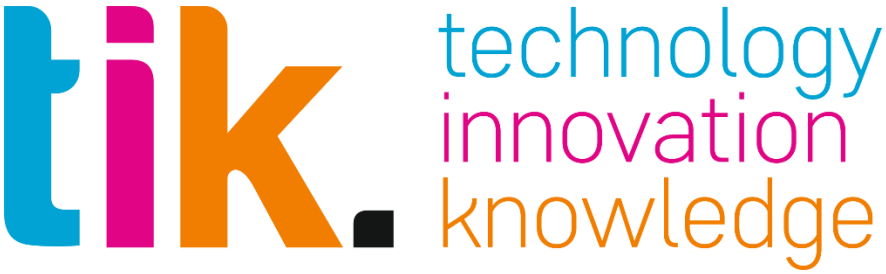
An Introduction to Parallelism, Concurrency and Acceleration (2)

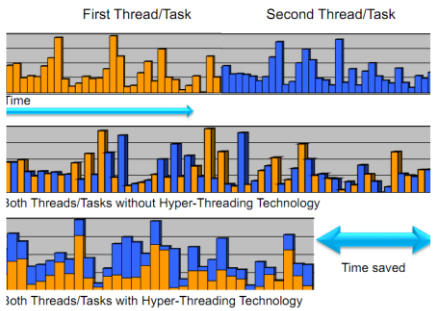
CERN Academic Training – Jan 2016

Andrzej Nowak

http://tik.services


tik.
technology
innovation
knowledge

Outline



Day 1:

Concurrency and Parallelism

Day 2:

Acceleration

Acceleration and co-processing

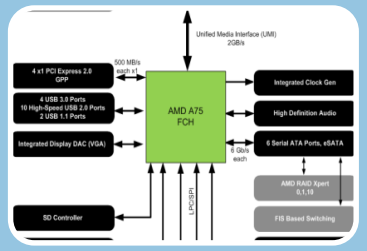
Heterogeneity scenarios

Cluster level

- Non-homogeneous nodes
- Large scale, expensive interconnect

Node level

- Non-homogeneous components of a node
- Standard platform interconnect



Chip level

- Non-homogeneous components in a package/chip
- On-chip interconnect or standard bus

i387

Image: cpu-world.com

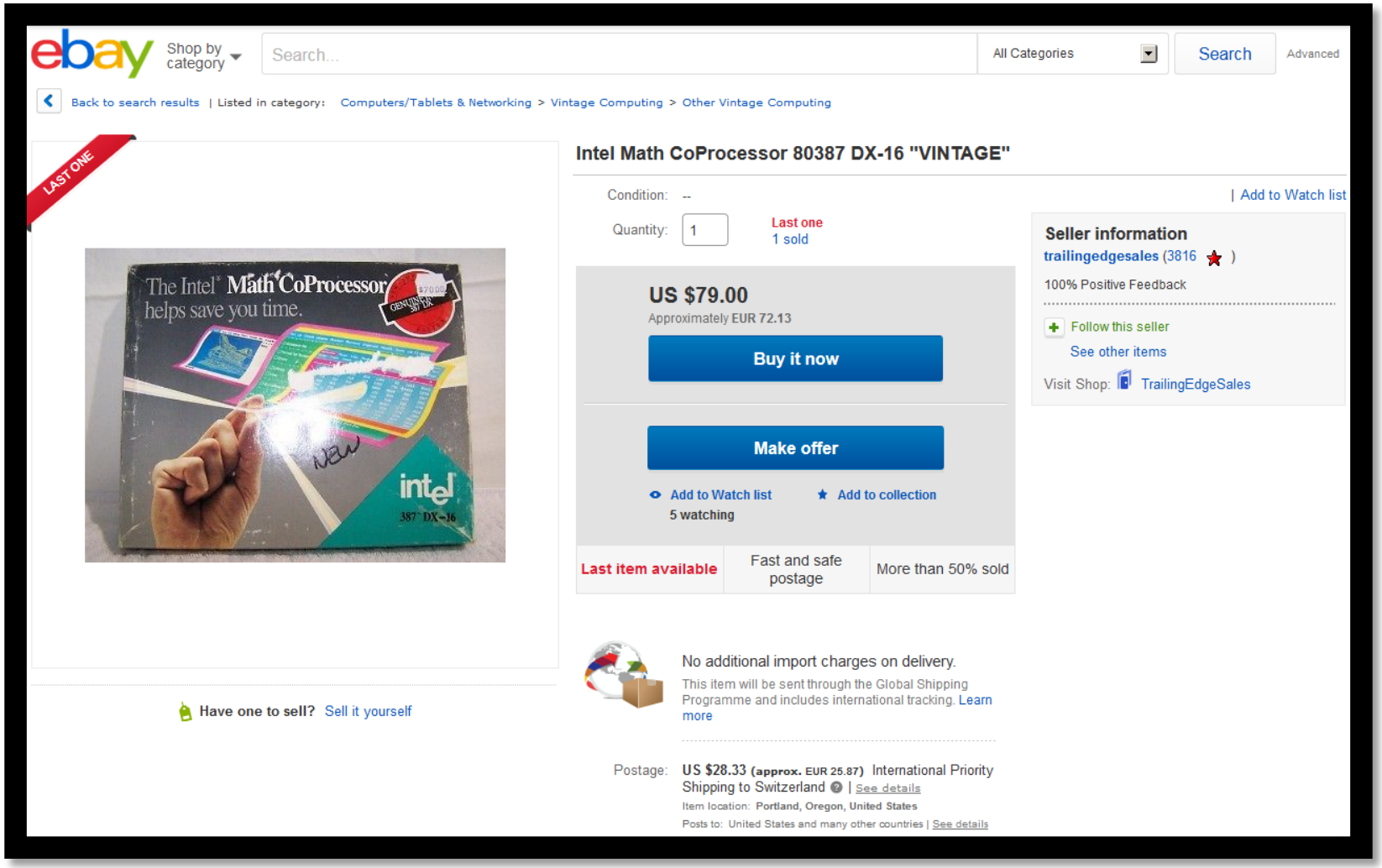
i387

intel®

MILITARY i387™ MATH COPROCESSOR

- High Performance 80-Bit Internal Architecture
- Implements ANSI/IEEE Standard 754-1985 for Binary Floating-Point Arithmetic
- Five to Seven Times M8087/M80287 Performance
- Upward Object-Code Compatible from M8087 and M80287
- Expands i386™ Microprocessor Data Types to Include 32-, 64-, 80-Bit Floating Point, 32-, 64-Bit Integers and 18-Digit BCD Operands
- Directly Extends i386 Processor Instruction Set to Include Trigonometric, Logarithmic, Exponential and Arithmetic Instructions for All Data Types
- Full-Range Transcendental Operations for SINE, COSINE, TANGENT, ARCTANGENT and LOGARITHM
- Built-In Exception Handling
- Operates Independently of Real, Protected and Virtual-8086 Modes of the i386 Microprocessor
- Eight 80-Bit Numeric Registers, Usable as Individually Addressable General Registers or as a Register Stack
- Available in 68-Pin PGA Package and 68-Lead Ceramic Quad Flat Pack
  (See Packaging Spec: Order #231369)
- Available in Three Product Grades:
  — MIL-STD-883, −55°C to +125°C (T_C)
  — Military Temperature Only, −55°C to +125°C (T_C)
  — Extended Temperature, −40°C to +110°C (T_C)

...that you can still buy



The Rapid Mind platform

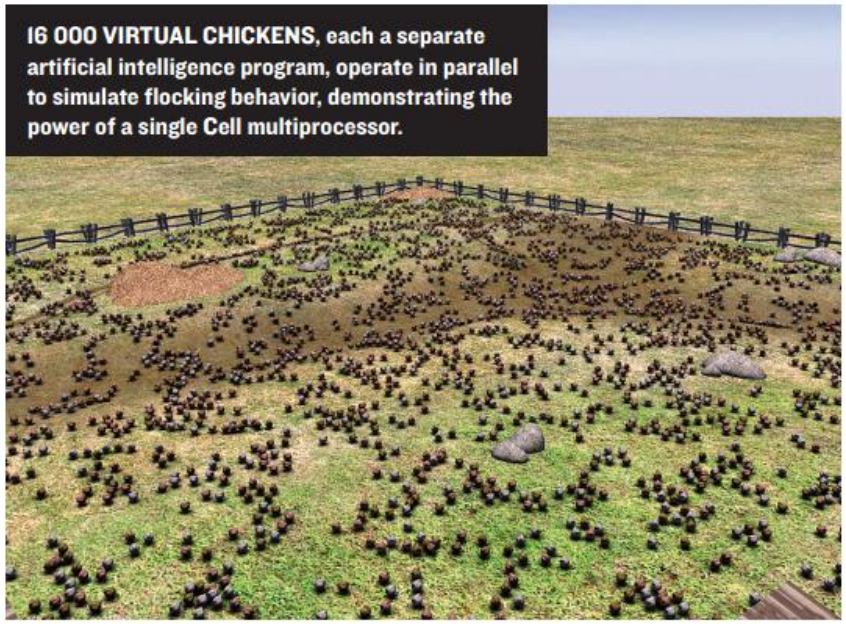


Image: IEEE/RapidMind

FPGAs

Image: Nallatech

Clearspeed

Software Development Kit with integrated visual profiling

Optimizing C compiler

Rapid application development

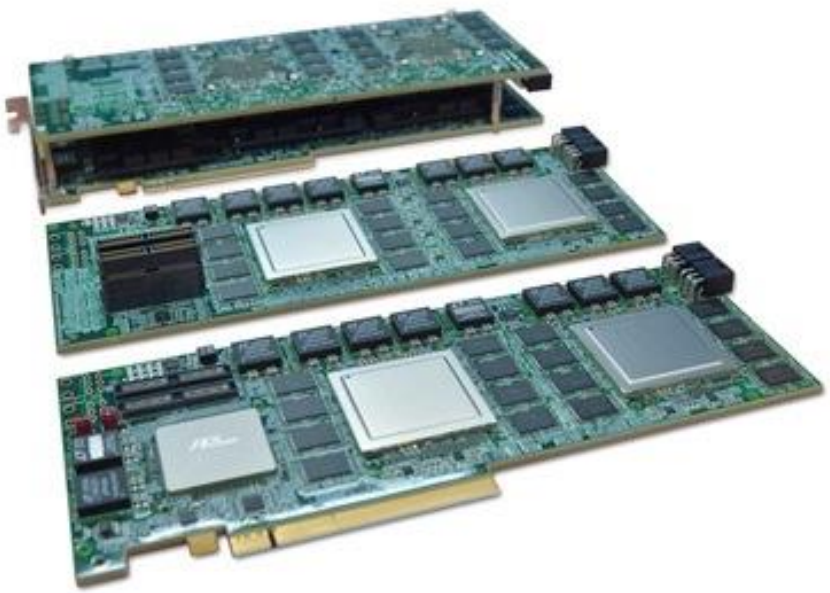
96 GFLOPS, 32 & 64-bit IEE 754 floating point

48 GMAC/s integer performance

Low power: 9W typical

Image: ClearSpeed

PEZY-SC



PEZY-SC Quad Board

Mother Board

Images: PEZY

NVIDIA

SMX: 192 single-precision CUDA cores, 64 double-precision units, 32 special function units (SFU), and 32 load/store units (LD/ST).

Image; NVIDIA

AMD
ex-ATI

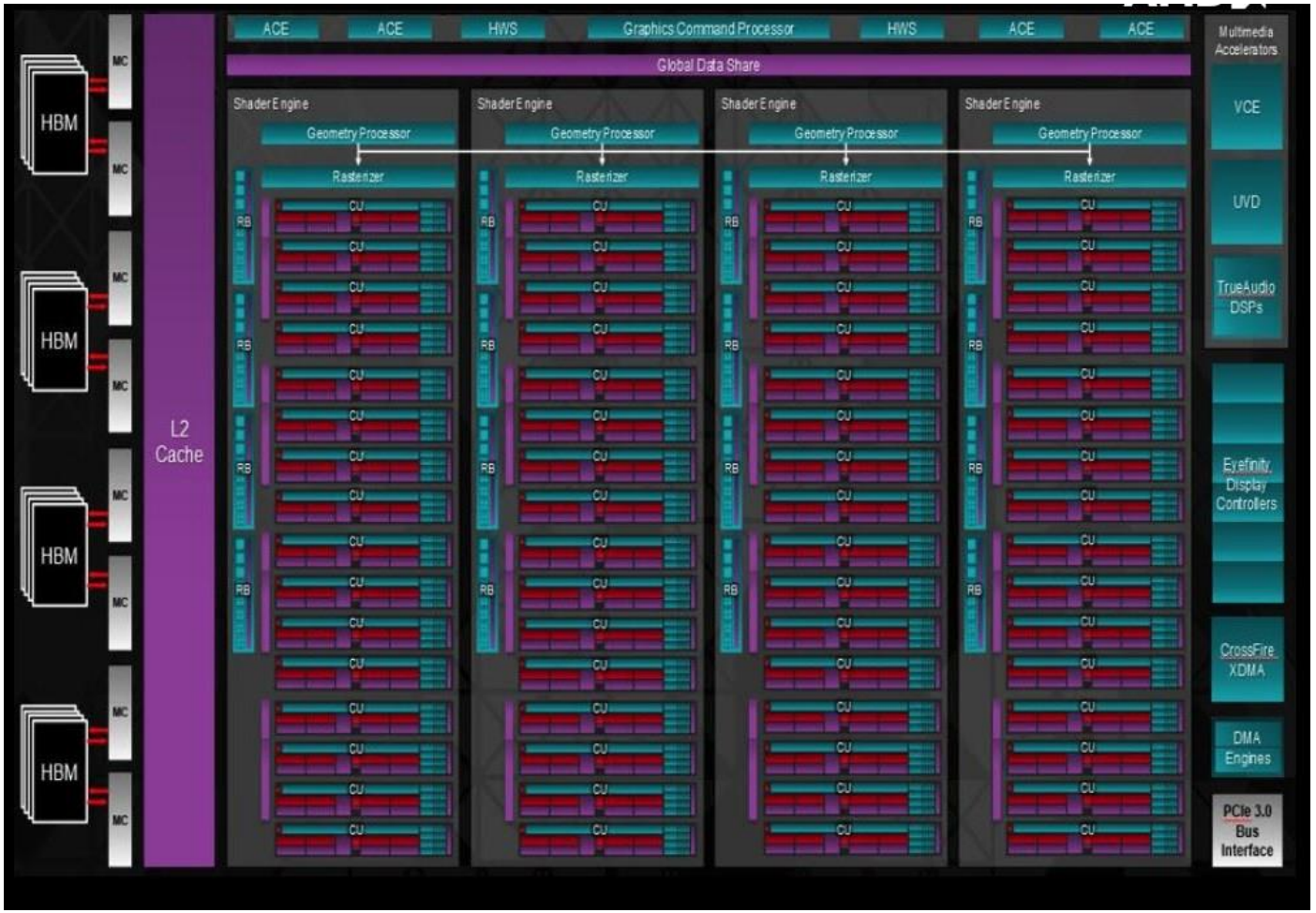
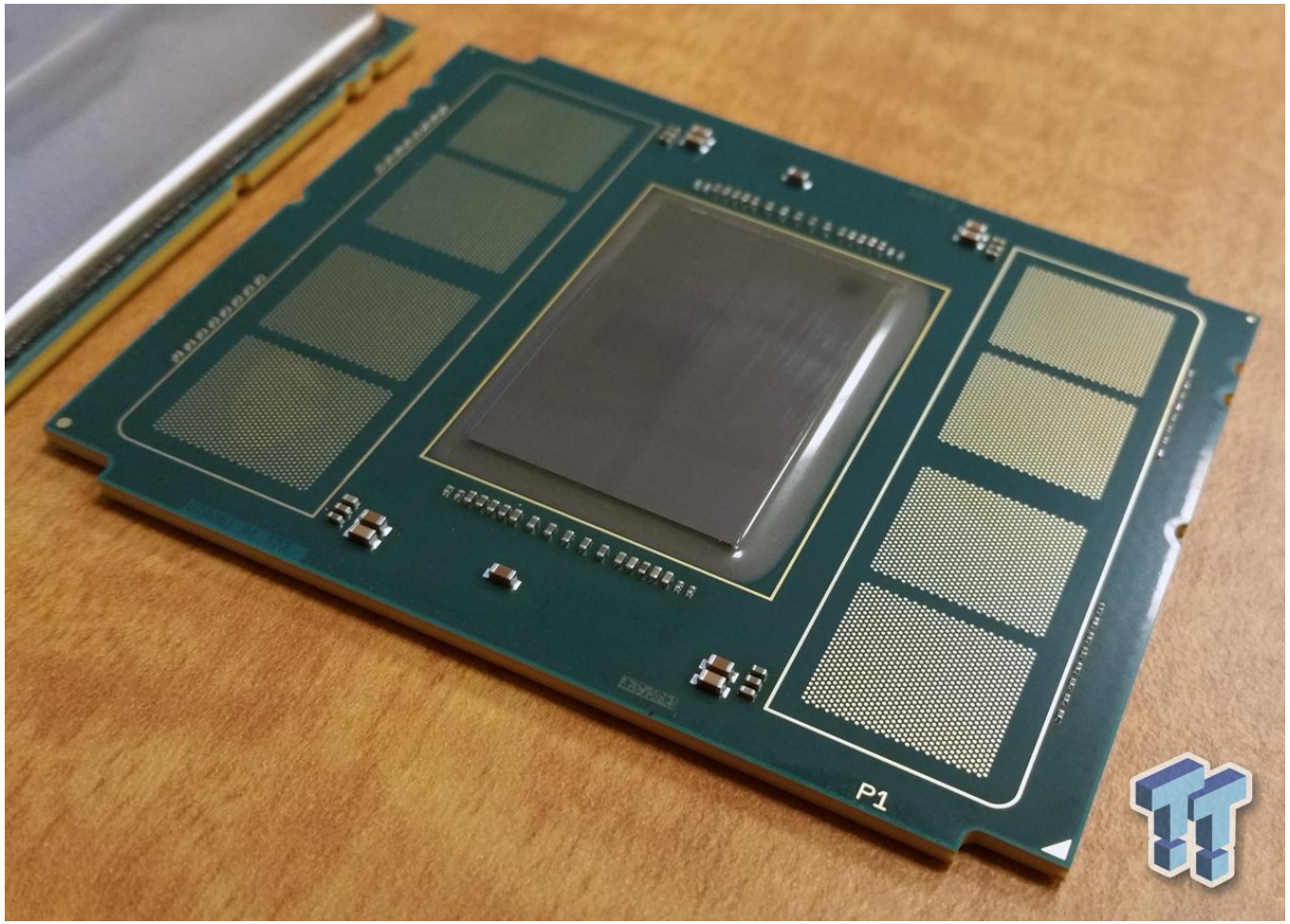


Image: AMD

Intel Xeon Phi

Aka Larrabee aka Many Integrated Core aka Knights



Image; Tweak Town

Intel Knights Landing

3 Knights Landing Products

A Paradigm Shift for Highly-Parallel

KNL Coprocessor

Host Processor

Host Processor with Integrated Fabric

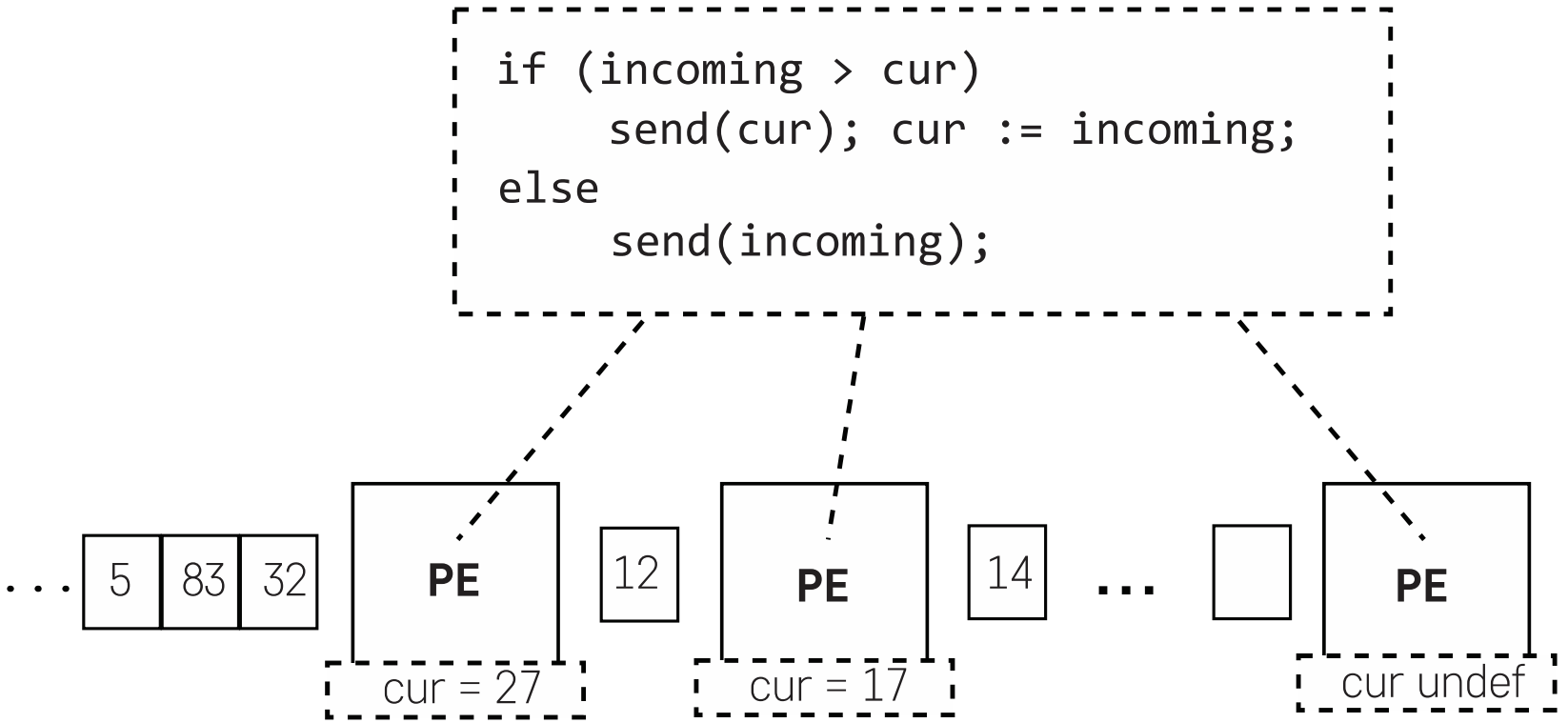
Programming Model	Intel® 64 / AVX-512	Intel® 64 / AVX-512	Intel® 64 / AVX-512
I/O	PCIe	Fabric	Integrated Fabric
Power Efficiency	Baseline	>25% Better[1]	>25% Better[1]
Resiliency	Baseline	Intel server-class	Intel server-class
Performance	>3 TF[2]	>3 TF[2]	>3 TF[2]
Memory Capacity	up to 16GB	up to 400GB[3]	up to 400GB[3]
Memory Bandwidth	>5x STREAM vs. DDR4[4]	>5x STREAM vs. DDR4[4]	>5x STREAM vs. DDR4[4]

Image: Intel

Intel Knights Landing (2015+)

KNC 2012	KNL 2015+
Core	
61 cores	72-76 cores
P54C (Pentium Pro)	Silvermont (Atom) – OoO!
1-1.2 GHz	? (1-1.3 GHz?)
Memory	
16 GB GDDR	16 GB MCDRAM (3D)
0 GB DDR3	Up to 384 GB DDR4
Package	
22nm	14nm
1 TF DP	3 TF DP
PCIe gen2	PCIe gen3, Socket, 100Gb/s

Spatial architectures
Triggered instructions



Adapted from IEEE Micro/MIT

Considerations

Hardware aspects

- Differing architectures talking to each other
  - Incompatible instructions, registers, binaries
- Limited memory
  - Host sharing might be possible
- Limited communication opportunities
  - Larger node distance
  - Memory transfer costly but flops are cheap
  - Need standard interconnect/communication
- Previously synchronous behavior might become asynchronous

Programming options (1)

Summing vector elements in C using OpenMP – openmp.org

```
#pragma omp parallel for reduction(+: s)
for (int i = 0; i < n; i++) {
  s += x[i];
}
```

Per element multiply in C++ using Intel® Array Building Blocks – intel.com/go/arbb

```
void products( const arbb::dense<arbb::f32>& a,
               const arbb::dense<arbb::f32>& b,
               arbb::dense<arbb::f32>& c) {
  c = a * b;
}
```

Dot product in Fortran using OpenMP – openmp.org

```
!$omp parallel do reduction ( + : adotb )
  do j = 1, n
    adotb = adotb + a(j) * b(j)
  end do
!$omp end parallel do
```

Sum in Fortran, using co-array feature – intel.com/software/products

```
REAL SUM[*]
CALL SYNC_ALL( WAIT=1 )
DO IMG= 2,NUM_IMAGES()
  IF (IMG==THIS_IMAGE()) THEN
    SUM = SUM + SUM[IMG-1]
  ENDIF
  CALL SYNC_ALL( WAIT=IMG )
ENDDO
```

Parallel function invocation in C using Intel® Cilk™ Plus – cilk.org

```
cilk_for (int i=0; i<n; ++i) {
  Foo(a[i]);
}
```

Parallel function invocation in C++ using Intel® Threading Building Blocks – threadingbuildingblocks.org

```
parallel_for (0, n,
  [=](int i) { Foo(a[i]); }
);
```

Programming options (2)

MPI code in C for clusters – intel.com/go/mpi

```
for (d=1; d<ntasks; d++) {
  rows = (d <= extra) ? avrow+1 : avrow;
  printf(" sending %d rows to task %d\n", rows, dest);
  MPI_Send(&offset, 1, MPI_INT, d, mtype, MPI_COMM_WORLD);
  MPI_Send(&rows, 1, MPI_INT, d, mtype, MPI_COMM_WORLD);
  MPI_Send(&a[offset][0], rows*NCA, MPI_DOUBLE, d, mtype, MPI_COMM_WORLD);
  MPI_Send(&b, NCA*NCB, MPI_DOUBLE, d, mtype, MPI_COMM_WORLD);
  offset = offset + rows;
}
```

Per element multiply in C using OpenCL – intel.com/go/opencl

```
kernel void
  dotprod( global const float *a,
           global const float *b,
           global float *c) {
    int myid = get_global_id(0);
    c[myid] = a[myid] * b[myid];
}
```

Matrix Multiply in Fortran using Intel® Math Kernel Library – intel.com/software/products

```
      call DGEMM(transa,transb,m,n,k,alpha,a,lda,b,ldb,beta,c,ldc)
```

Floating Point

$$x = (-)^s \beta^e \sum_{i=0}^{p-1} x_i \beta^{-i}$$

- Floating point results will differ
  - Does the algorithm support that?
  - Is it possible to live with different results?
  - Reproducibility
- Math functions will vary in precision across platforms
  - "Fast" functions might be used as defaults!
- Results can depend on data and operation ordering
  - E.g. "if two items are equal, choose the first one" (risky)
  - Solution: sort or identify data uniquely
- Associativity may vary across platforms
  - Outputs differ as a result

Energy efficiency

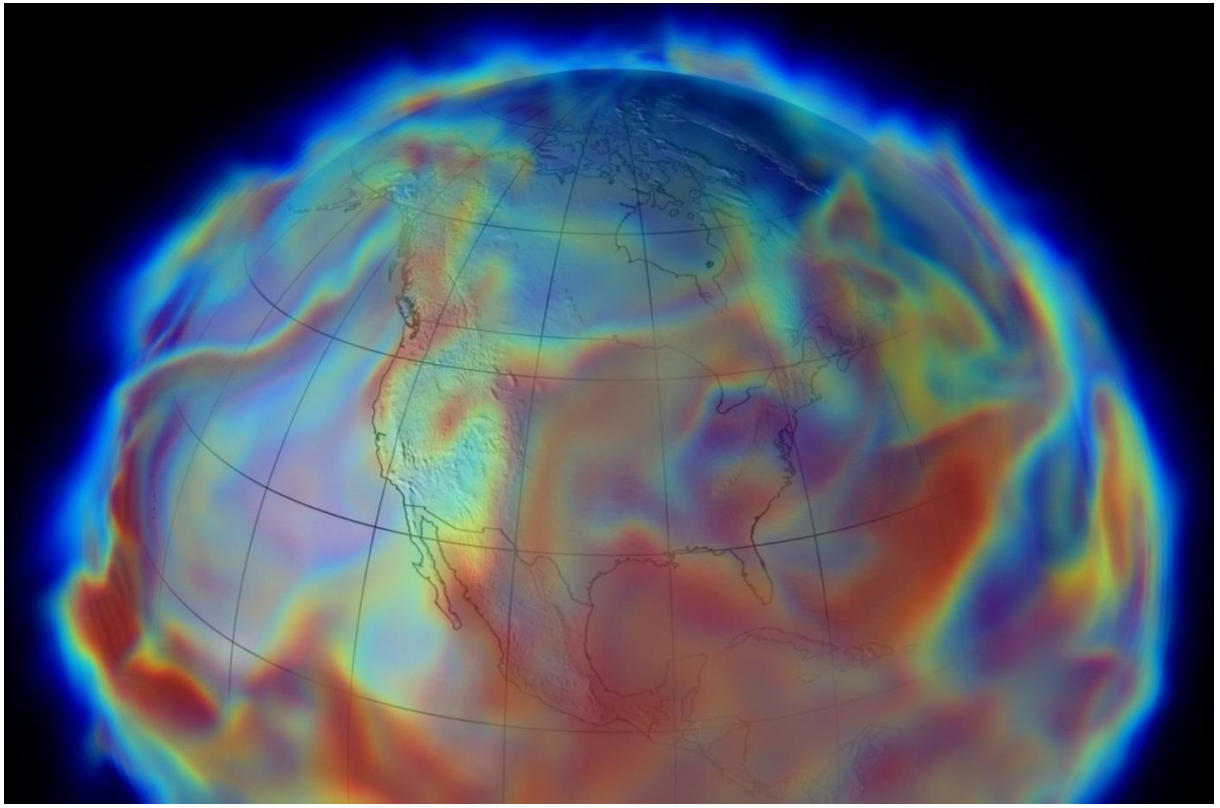
Operation	Energy cost
L1 access	2 pJ
L2 access	150 pJ
RAM	2 000 pJ
FLOP	?

Green500

Green500 Rank	MFLOPS/W	Site*	Computer*	Total Power (kW)
1	7,031.58	Institute of Physical and Chemical Research (RIKEN)	Shoubu - ExaScaler-1.4 80Brick, Xeon E5-2618Lv3 8C 2.3GHz, Infiniband FDR, PEZY-SC	50.32
2	5,331.79	GSIC Center, Tokyo Institute of Technology	TSUBAME-KFC/DL - LX 1U-4GPU/104Re-1G Cluster, Intel Xeon E5-2620v2 6C 2.1GHz, Infiniband FDR, NVIDIA Tesla K80	51.13
3	5,271.81	GSI Helmholtz Center	ASUS ESC4000 FDR/G2S, Intel Xeon E5-2690v2 10C 3GHz, Infiniband FDR, AMD FirePro S9150	57.15
4	4,778.46	Institute of Modern Physics (IMP), Chinese Academy of Sciences	Sugon Cluster W780I, Xeon E5-2640v3 8C 2.6GHz, Infiniband QDR, NVIDIA Tesla K80	65.00
5	4,112.11	Stanford Research Computing Center	XStream - Cray CS-Storm, Intel Xeon E5-2680v2 10C 2.8GHz, Infiniband FDR, Nvidia K80	190.00
6	3,856.90	IT Company	Inspur TS10000 HPC Server, Xeon E5-2620v3 6C 2.4GHz, 10G Ethernet, NVIDIA Tesla K40	58.00
7	3,775.45	Internet Service	Inspur TS10000 HPC Server, Intel Xeon E5-2620v2 6C 2.1GHz, 10G Ethernet, NVIDIA Tesla K40	110.00
8	3,775.45	Internet Service	Inspur TS10000 HPC Server, Intel Xeon E5-2620v2 6C 2.1GHz, 10G Ethernet, NVIDIA Tesla K40	110.00
9	3,775.45	Internet Service	Inspur TS10000 HPC Server, Intel Xeon E5-2620v2 6C 2.1GHz, 10G Ethernet, NVIDIA Tesla K40	110.00
10	3,775.45	Internet Service	Inspur TS10000 HPC Server, Intel Xeon E5-2620v2 6C 2.1GHz, 10G Ethernet, NVIDIA Tesla K40	110.00

The Outcome

Natural threat exploration



Image; NASA

“Geological” exploration

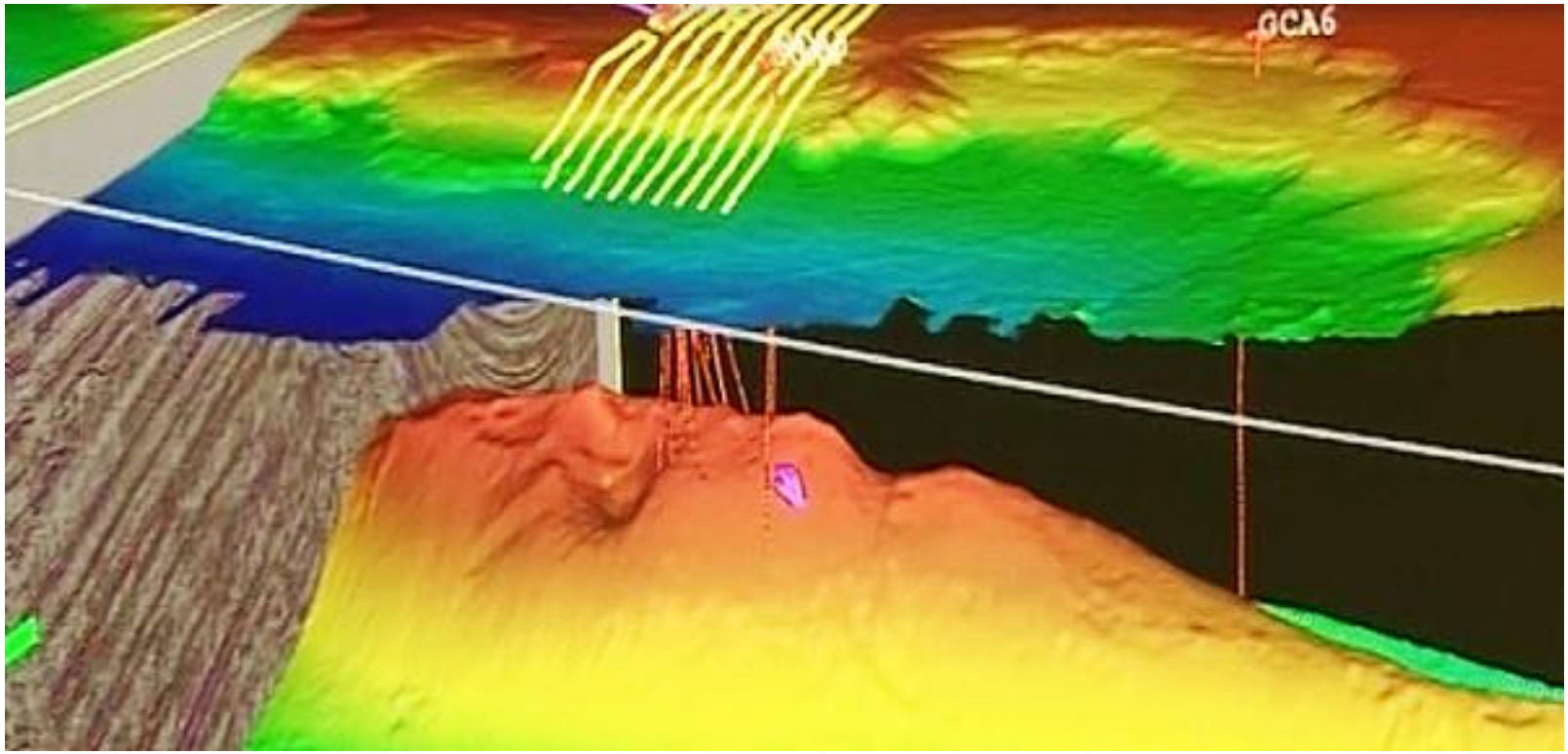


Image: BP

Scientific exploration

Image: WLCG / B. Dice, V. Mancinelli

Top500

TOP 500
The List.

HOME LISTS STATISTICS FEATURES RESOURCES ABOUT Search Go

TOP500 SUPERCOMPUTING SITES

CURRENT LIST NOVEMBER 2015

TOP #1 SYSTEMS

NEWS

K COMPUTER RUNS WORLD'S LARGEST-SCALE MAGNETIC-REVERSAL SIMULATOR

Posted 5 days, 11 hours ago

Today Fujitsu Limited announced it has developed the world's largest magnetic-reversal simulator. Developed in joint research with the National Institute for Materials Science (NIMS), the simulator runs on the famous K computer using a mesh covering more than 300 million micro-regions.

read more

5 PREDICTIONS: WHERE SUPERCOMPUTING IS HEADING IN 2016

Posted 1 week ago

LIST HIGHLIGHTS

1

Tianhe-2 (MilkyWay-2) - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P
NUDT

2

Titan - Cray XK7 , Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x
Cray Inc.

3

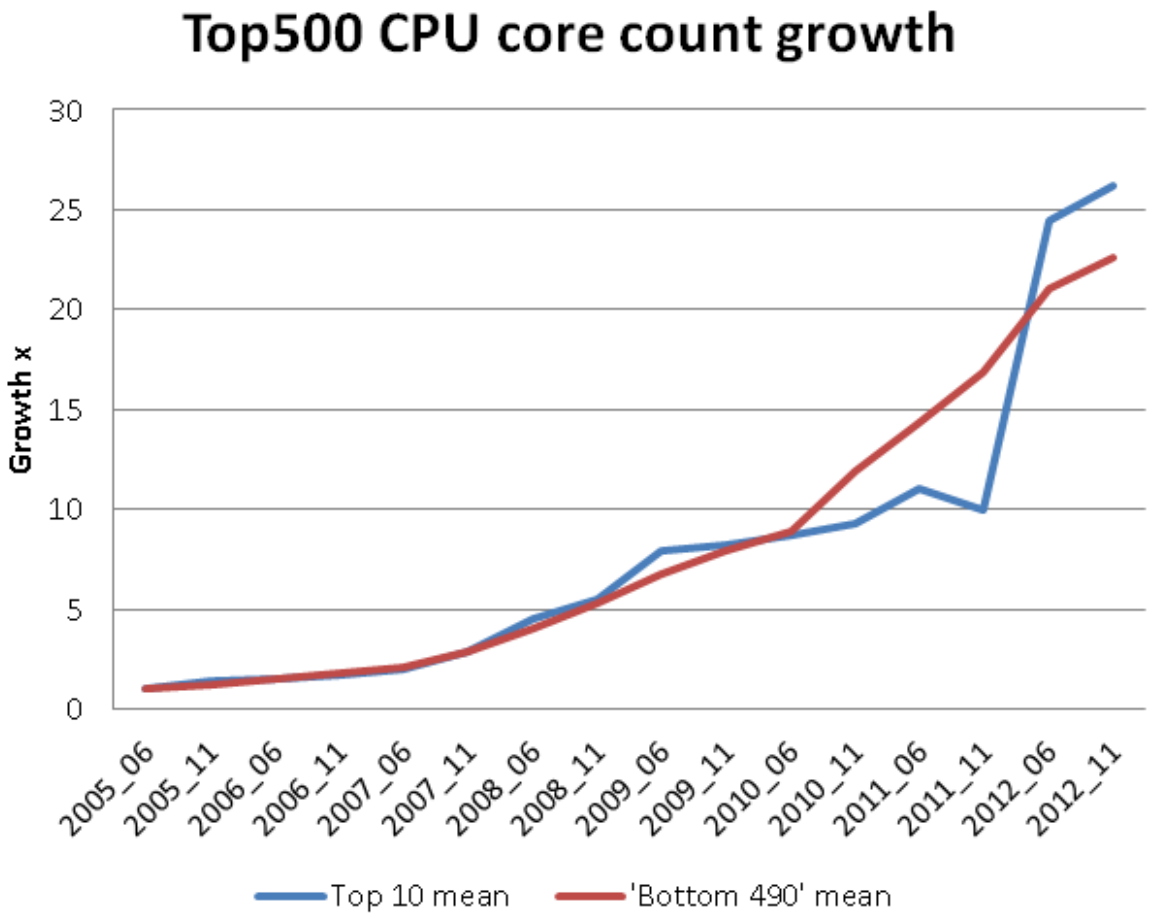
Sequoia - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom
IBM

4

K computer, SPARC64 VIIIfx 2.0GHz, Tofu

Top500 - tidbits

Top500 - tidbits



Modeled after Lehto, Manninen, von Alfthan

Top500 - tidbits

PROCESSOR GENERATION	COUNT	SYSTEM SHARE (%)	RMAX (GFLOPS)	RPEAK (GFLOPS)	CORES
Intel Xeon E5 (IvyBridge)	161	32.2	123,824,132	206,197,286	9,711,630
Intel Xeon E5 (Haswell)	137	27.4	117,484,829	187,232,555	6,026,550
Intel Xeon E5 (SandyBridge)	118	23.6	62,585,931	88,627,249	4,305,047
Power BQC	19	3.8	47,047,481	55,155,100	4,308,992
Xeon 5600-series (Westmere-EP)	16	3.2	11,323,706	22,080,558	815,644
Opteron 6200 Series "Interlagos"	8	1.6	20,967,117	31,699,293	934,424
Opteron 6100-series "Magny-Cours"	7	1.4	3,443,900	4,562,726	492,000
POWER7	7	1.4	3,455,442	4,146,720	139,072
Opteron 6300 Series "Abu Dhabi"	6	1.2	2,140,661	7,159,578	699,328
SPARC64 XIfx	5	1	8,454,200	9,359,769	285,696
Intel Xeon E7 (IvyBridge)	4	0.8	1,583,968	3,954,432	229,920
Xeon 5500-series (Nehalem-EP)	4	0.8	1,466,431	2,130,314	129,024
Intel Xeon Phi	1	0.2	209,088	463,983	23,424
Xeon 5400-series "Harpertown"	1	0.2	237,800	267,878	23,040
Intel Xeon E3 (Haswell)	1	0.2	238,226	466,560	16,200
Xeon E7 (Westmere-EX)	1	0.2	354,300	655,258	38,400
ShenWei	1	0.2	795,900	1,070,160	137,200
SPARC64 IXfx	1	0.2	1,043,000	1,135,411	76,800
Xeon 5500-series (Nehalem-EX)	1	0.2	1,050,000	1,254,550	138,368
SPARC64 VIIIfx	1	0.2	10,510,000	11,280,384	705,024

Image: Nextplatform/Top500

Top500 - tidbits

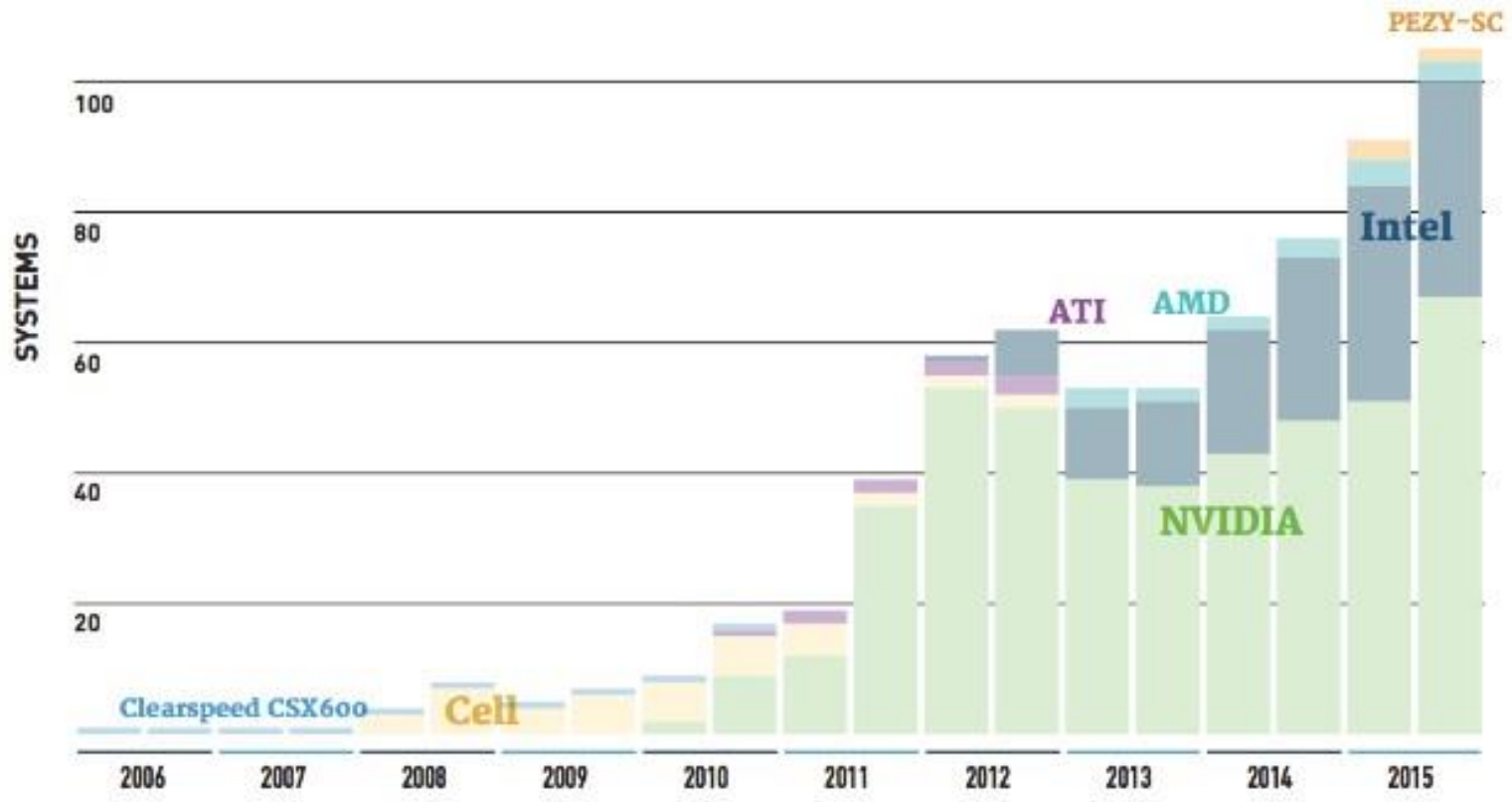


Image: Top500

Top500 - tidbits

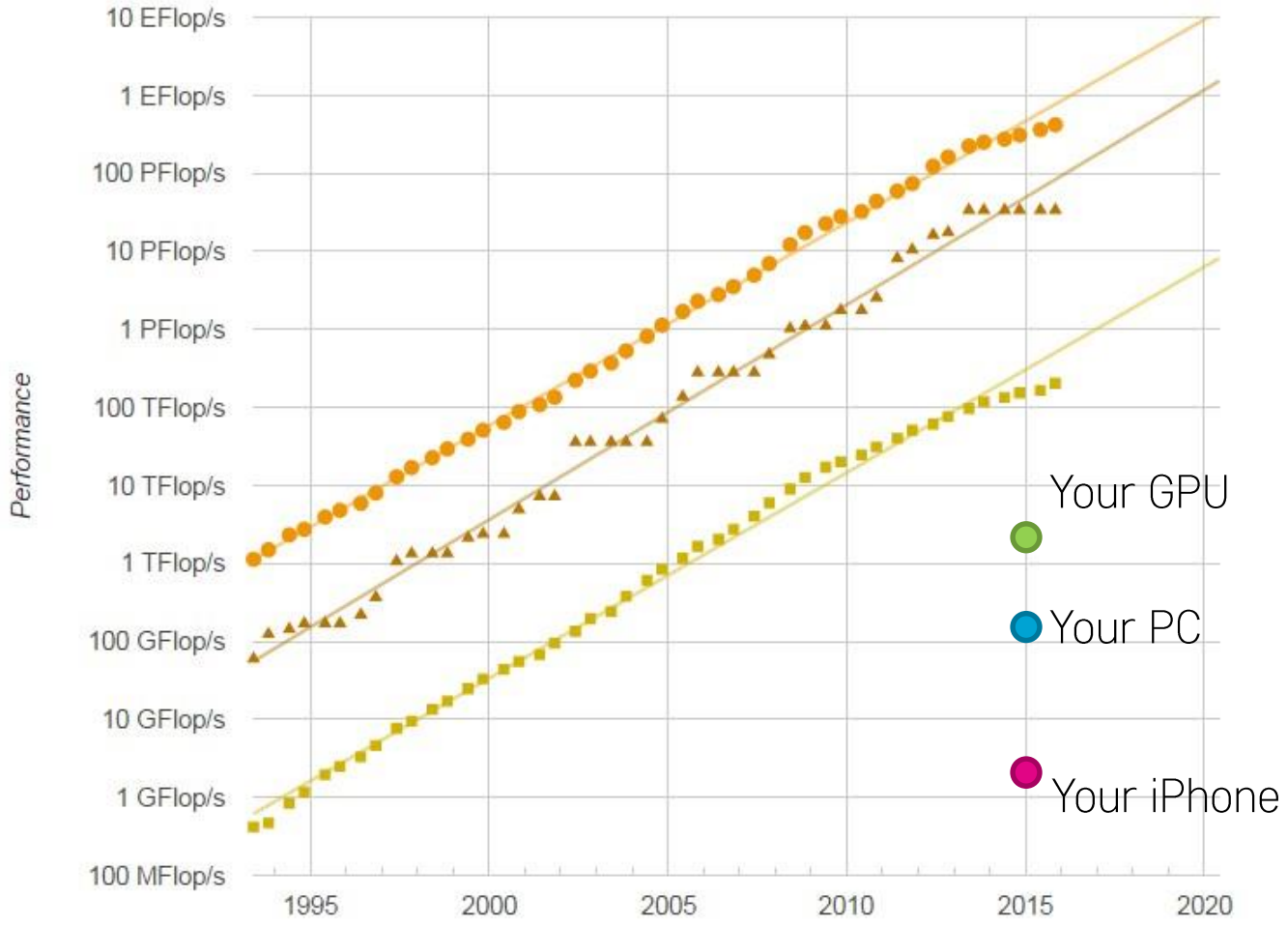


Image: Nextplatform/Top500

Summary
Key considerations for parallelism

Thank you

e-mail: an@tik.services

http://tik.services