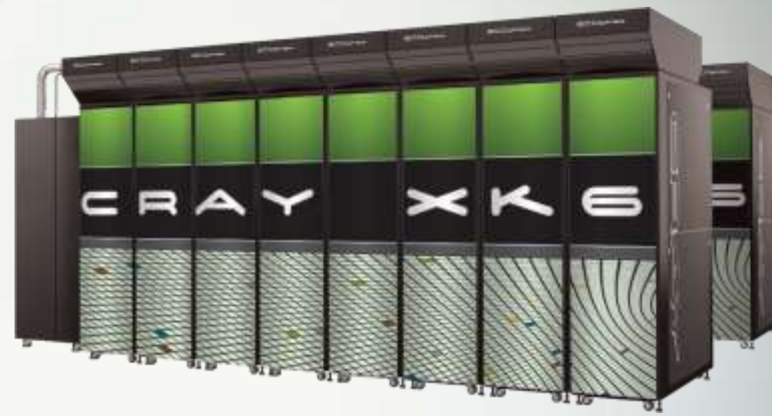


Energy Efficiency Metrics and Cray XE6 Application Performance

Wilfried Oed
Principal Engineer

Taking a Step back in History

- What made this machine so unique ?
- Some answers
  - Novel vector architecture
  - Packaging
  - Cooling
  - Fastest scalar machine !!!
  - High productivity for users
    - Autovectorizing compiler
    - Performance analysis tool
    - Simple OS

and no one cared about the power consumption

Progress in Energy Efficiency

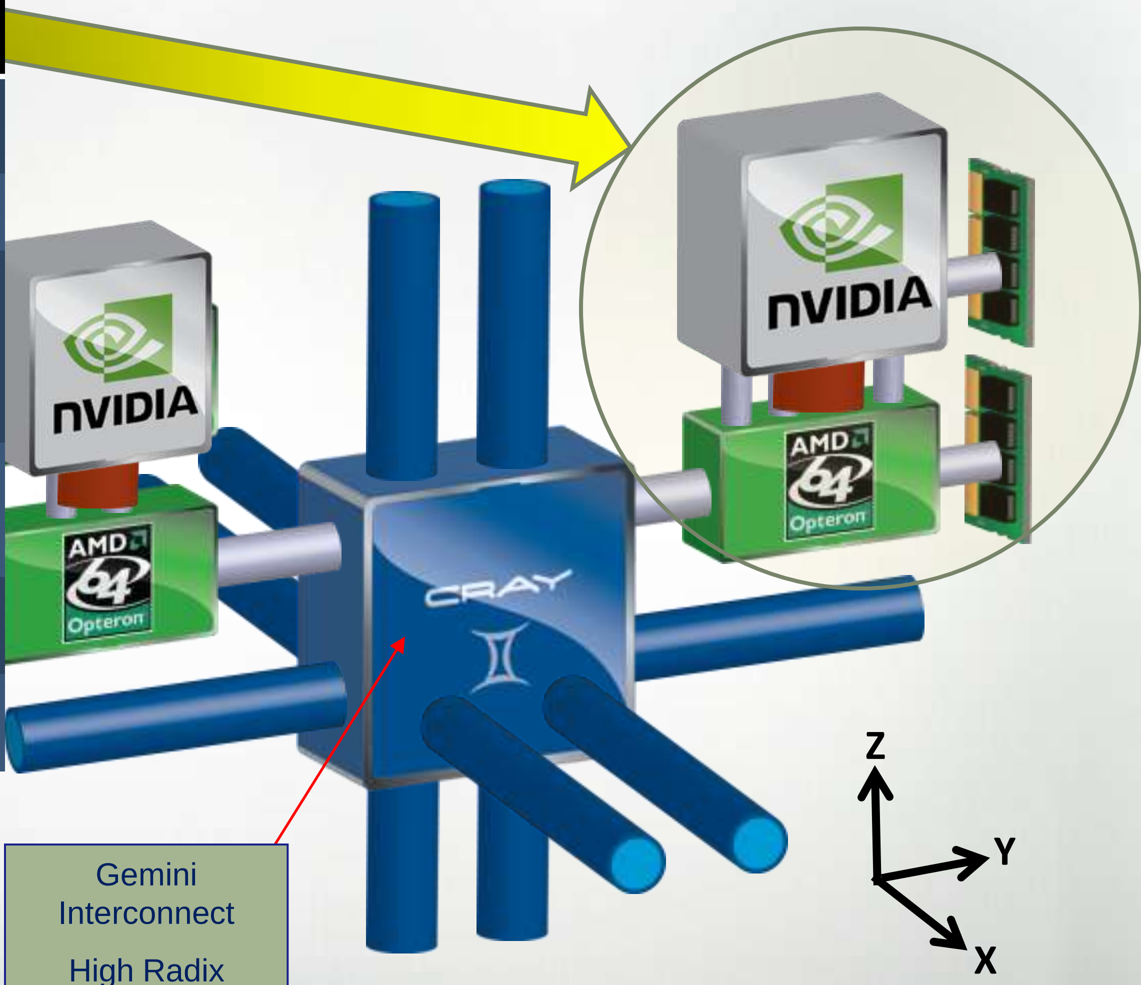
Power Consumption for Cray Systems	1978	1988	1998	2008
	Cray-1	Cray Y-MP 8	Cray T3E	Cray XT5
number processors / cores	1	8	1,024	150,152
power consumption (kW)	140	200	220	6,500
Rmax PF	1.50E-07	2.10E-06	8.92E-04	1.06E+00
Flop / Watt	~ 0.001 MF	~ 0.01 MF	~ 4 MF	~ 150 MF
Efficiency improvement	1	10	~ 4,000	~ 150,000

- An improvement of 150 thousand in 30 years – and still no end in sight !
  - Cray XE6 is ~ 600 MF / W
  - Cray XK6 is ~ 1200 MF / W
- So where's the problem ?
  - Price performance has improved even more dramatically
    - Computing has become ubiquitous
  - The combined systems of the current Green500 require 340 MW
    - That's up 50 MW from previous list
    - Largest system @ 10 MW
  - Supercomputing and HPC are vital tools for science
- An interesting article – especially the focus on software

Andrew Jones, Vice-President of HPC Services and Consulting, Numerical Algorithms Group
http://www.hpcwire.com/hpcwire/2011-08-29/exascale:_power_is_not_the_problem_.html

Cray XE6 Node   Cray XK6 Node

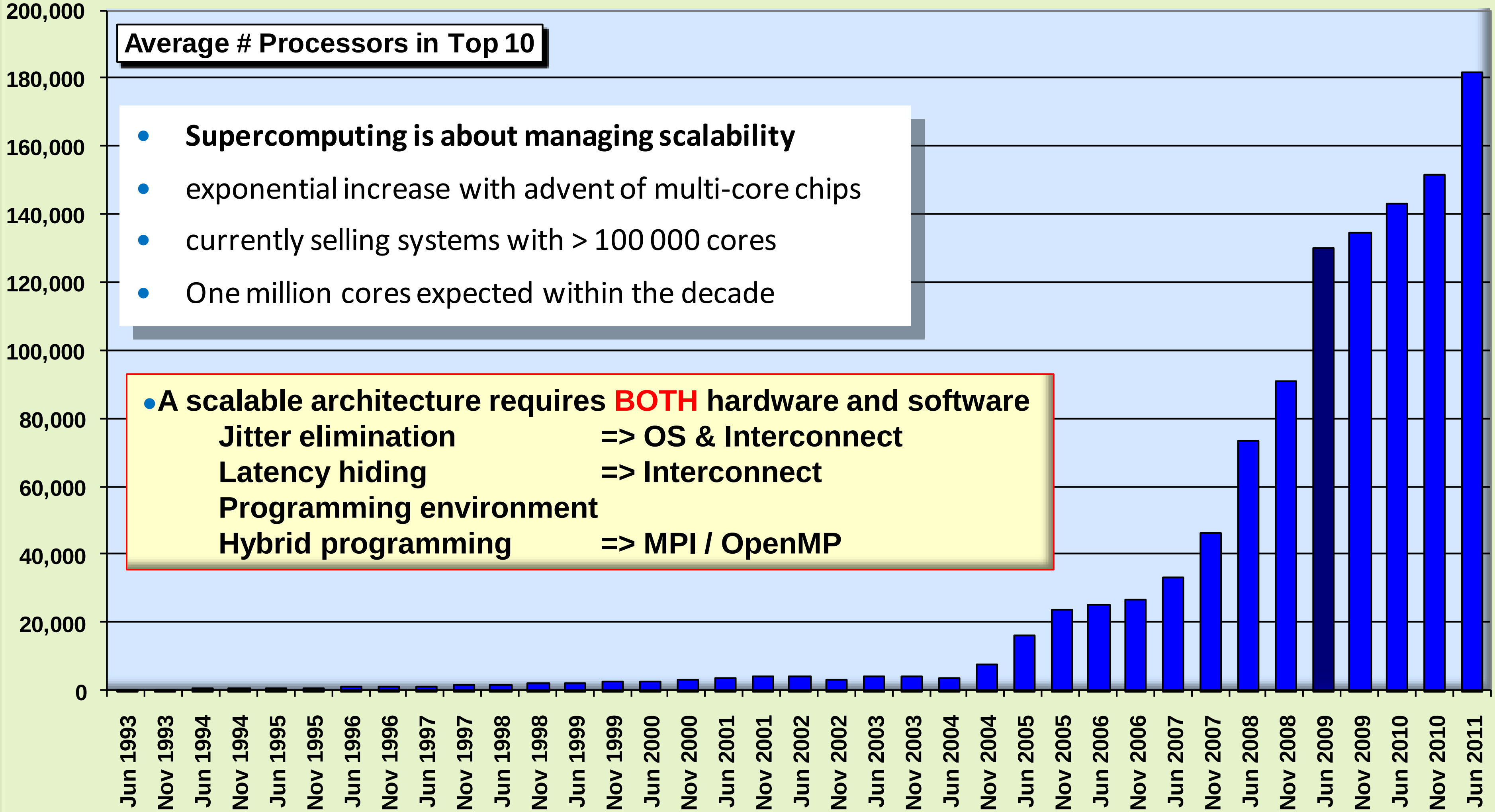
XK6 Compute Node Characteristics
AMD Series 6200 (Interlagos)
NVIDIA Tesla X2090
Host Memory 16 or 32GB 1600 MHz DDR3
NVIDIA Tesla X2090 Memory 6GB GDDR5 capacity
Gemini High Speed Interconnect
Upgradeable to future GPUs



Gemini Interconnect

High Radix YARC Router with adaptive Routing

Supercomputing Today



- **Supercomputing is about managing scalability**
- exponential increase with advent of multi-core chips
- currently selling systems with > 100 000 cores
- One million cores expected within the decade

- **A scalable architecture requires BOTH hardware and software**
  - **Jitter elimination => OS & Interconnect**
  - **Latency hiding => Interconnect**
  - **Programming environment**
  - **Hybrid programming => MPI / OpenMP**

Early Science Applications (Cray XT5)

Eight Application World Records Set in First Week (Nov. 2008)!

Science Area	Code	Contact	Cores	Total Perf	Notes	Scaling
Materials	DCA++	Schulthess	150,144	1.3 PF*	Gordon Bell Winner	Weak
Materials	LSMS/WL	ORNL	149,580	1.05 PF	64 bit	Weak
Seismology	SPECFEM3D	UCSD	149,784	165 TF	Gordon Bell Finalist	Weak
Weather	WRF	Michalakes	150,000	50 TF	Size of Data	Strong
Climate	POP	Jones	18,000	20 sim yrs/ CPU day	Size of Data	Strong
Combustion	S3D	Chen	144,000	83 TF		Weak
Fusion	GTC	UC Irvine	102,000	20 billion Particles / sec	Code Limit	Weak
Materials	LS3DF	Lin-Wang Wang	147,456	442 TF	Gordon Bell Winner	Weak

Energy Efficiency Metrics

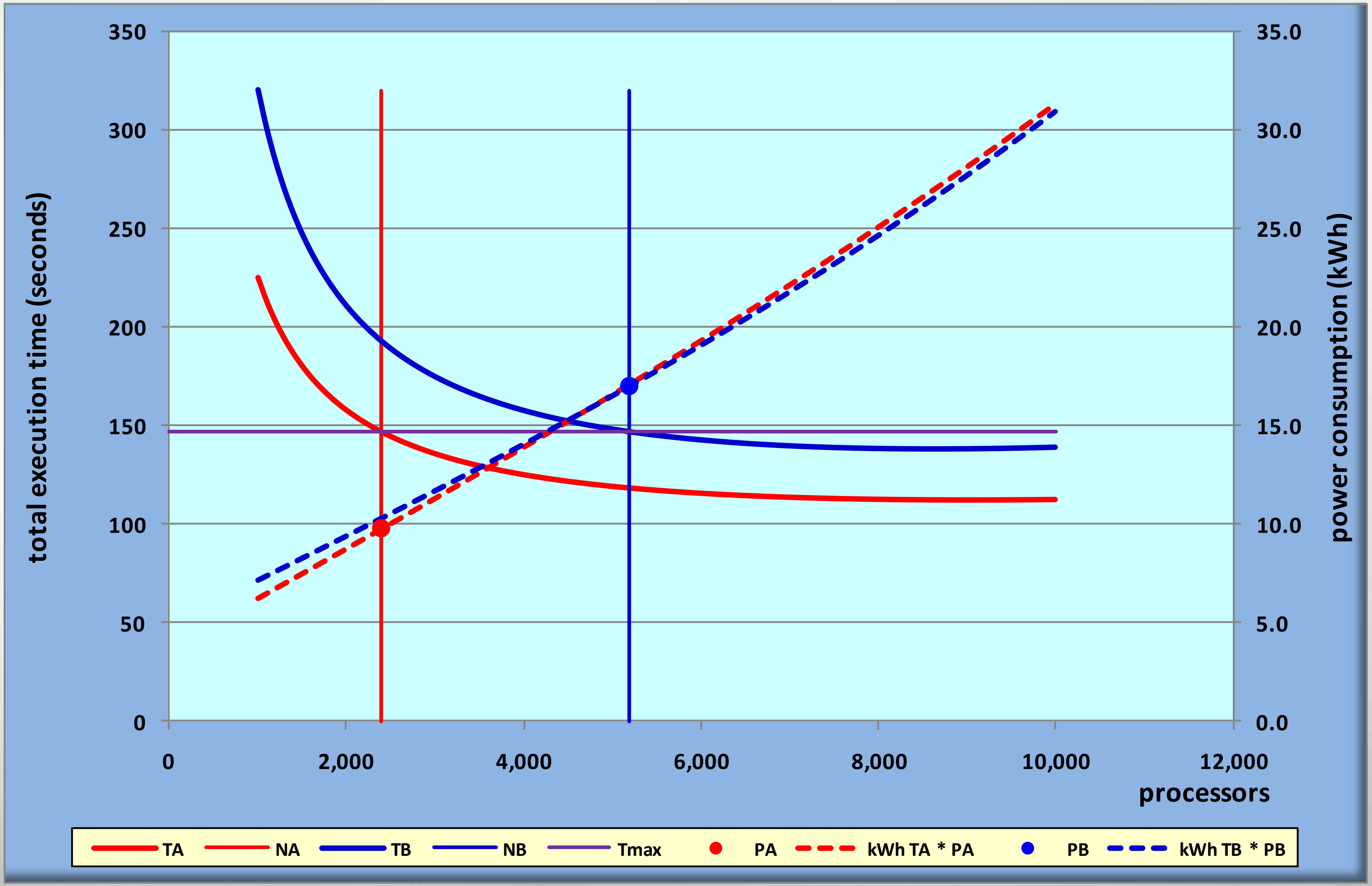
- Power Usage Effectiveness (PUE)
  - Reflects how well a system is being cooled
    - A poorly designed system can still have a wonderful PUE if cooling is efficient
  - Need to define the components that account for "power usage"
- MFLOPS per Watt
  - Reflected in the Green500
  - Emphasizes pure floating-point (HPL)
- Time to Solution (sustained performance) per Watt
  - Supercomputers are there to solve big problems (aka Grand Challenges)
    - An extremely high degree of parallelism is required
    - Besides floating-point, real applications have to deal with communication, organization, load balance
  - Power consumption [kWh] = $N_{proc} * P_{proc} * T_{max}$ [kWh]
    - T_{max} time allowed to finish the problem
    - N_{proc} number of processors (cores) utilized to finish within T_{max}
    - P_{proc} power utilized per processor (core)
  - This metric is problem oriented and can be applied across various architectures
    - Can also be based on power per *node* for comparing vastly different archictures (e.g. Cray XK6 using hybrid CPU / GPU nodes)

Comparison at any T_{max}

- The lower power processor has the same power on a per core basis
- Despite being a lower power processor and having similar scalability, the higher core count required makes it less efficient regardless of the desired solution time

Note: this is an arbitrary example for demonstrating certain effects

neither based on actual systems nor applications

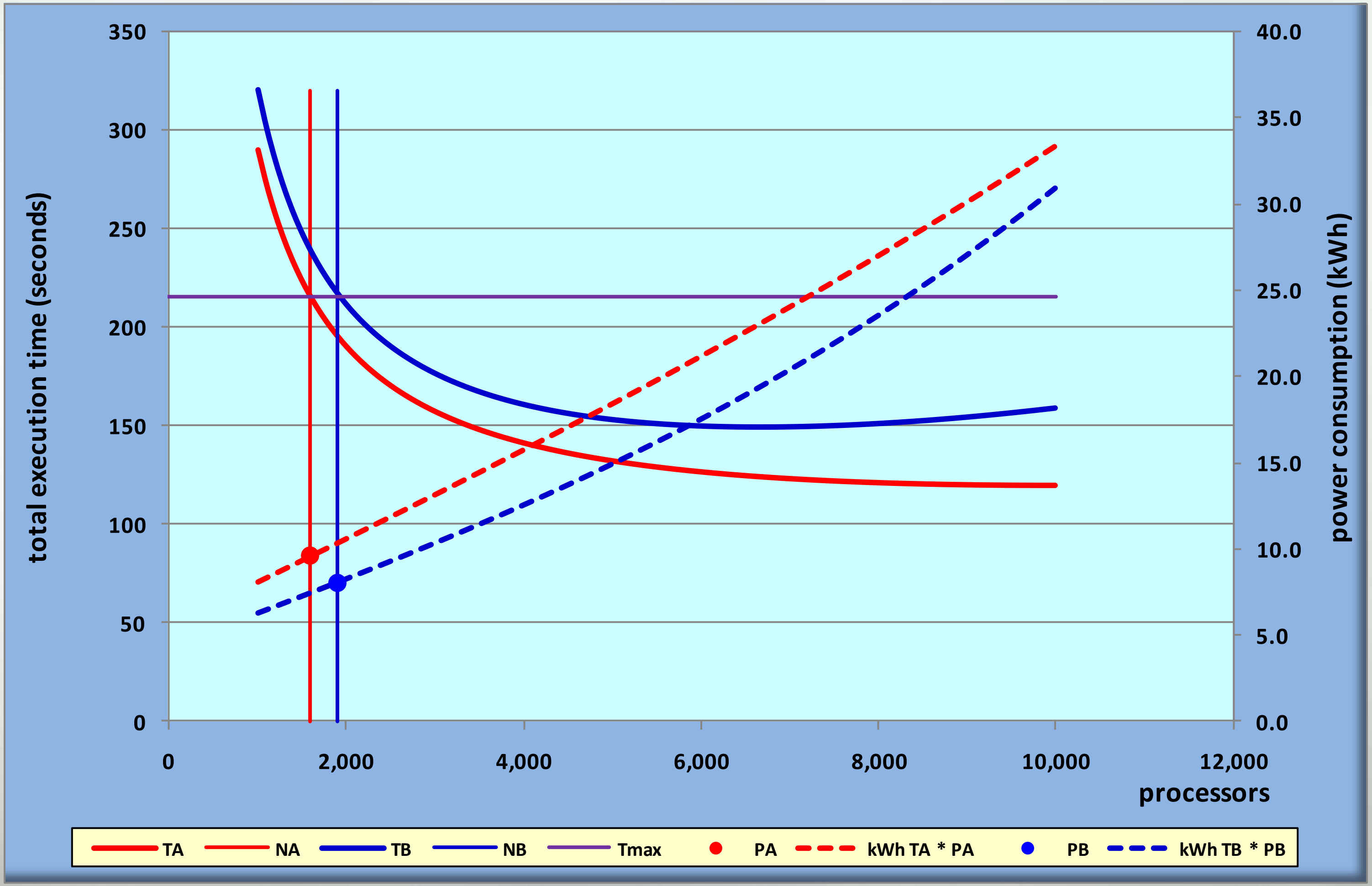


Comparison at higher T_{max}

- The lower power processor always requires less power on a per core basis
- At low core counts (higher time to solution) the lower powered processor is more energy efficient , as only a few additional cores are required

Note: this is an arbitrary example for demonstrating certain effects

neither based on actual systems nor applications

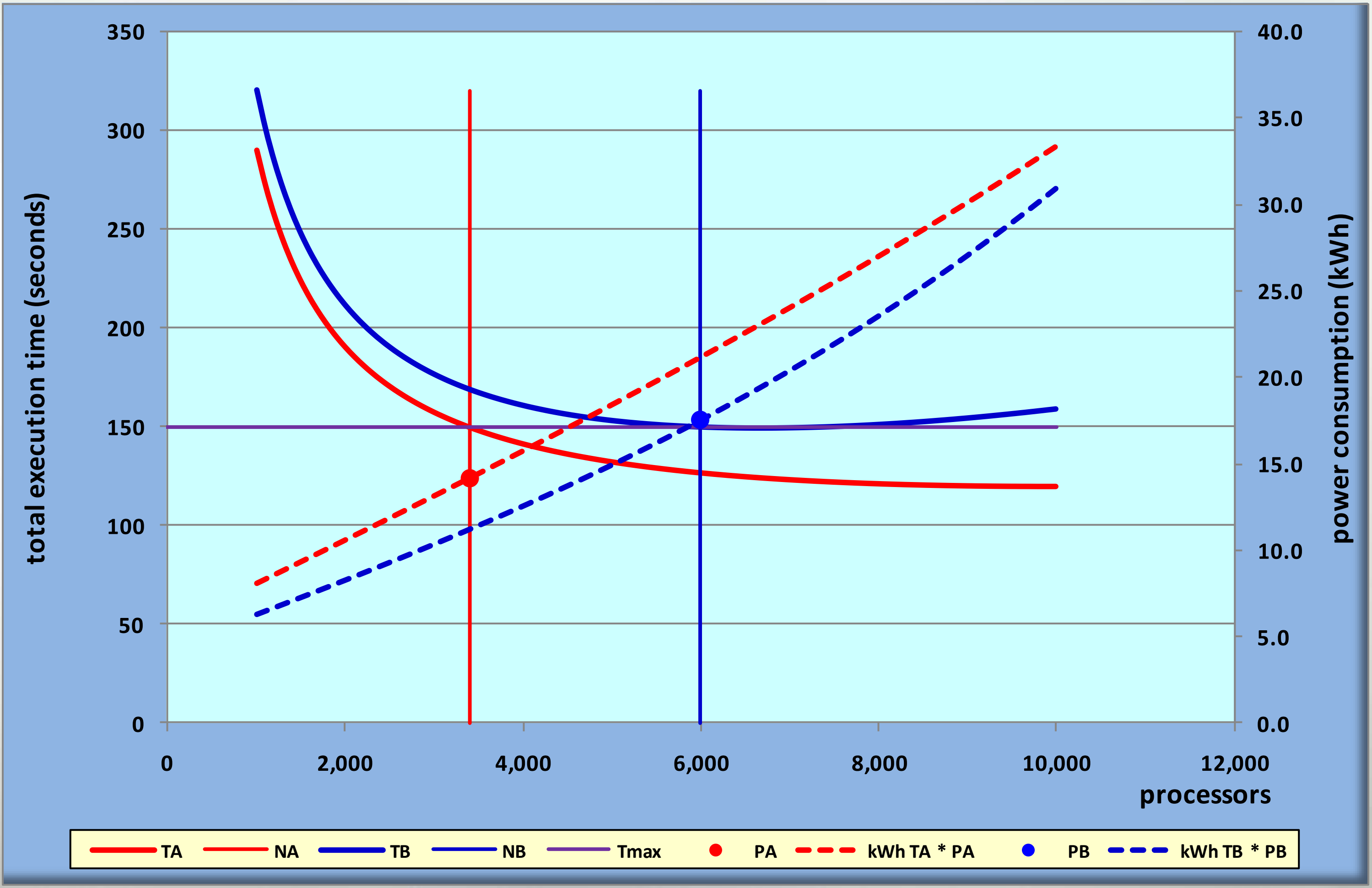


Comparison at lower T_{max}

- The lower power processor always requires less power on a per core basis
- At higher core counts (lower time to solution) the lower powered processor is less energy efficient, as far more cores are required

Note: this is an arbitrary example for demonstrating certain effects

neither based on actual systems nor applications



Applications at EPCC on the HECToR System (Cray XE6)

- A set of scientific applications running on a regular basis at high core counts at EPCC

Science Area	Code	Nodes	Cores
Combustion	Senga	844	20,256
Materials and MD	CASTEP	1,024	24,576
fluid flow/lattice-boltzmann method	Heme1b	1,024	24,576
Materials	CRYSTAL	1,024	24,576
Quantum Monte Carlo	CASINO	664	15,936
MD	DL_POLY_4	683	16,392
Chemistry	Sparkle	683	16,392

Summary & Outlook

- Despite huge progress let‘s not rest
  - The biggest innovations will have to come from technology
    - Remember: the goal for EXAflop is 20 MW or 50 GF / W
    - Which may questionable => keynote: Jens Wiebe
  - Reclaim energy => driving towards PUE < 1
    - Heating your office is not the answer
  - Throttling CPU performance if higher T_{max} can be tolerated
    - Current processors have the ability to operate at different clock speeds already
    - But beware, your overall power consumption may end up to be higher
- Applying the metrics
  - Required is the ability to measure performance on an application level
    - James H. Laros III, Kevin T. Pedretti, Suzanne M. Kelly, John P. Vandyke, Kurt B. Ferreira, Courtenay T. Vaughan, Mark Swan. Topics on Measuring Real Power Usage on High Performance Computing Platforms, IEEE International
  - Energy aware scheduling
- TUNE your application (a truck has good mileage only if fully loaded)
  - Scalability is a decisive factor on time to solution and consequently on power efficiency

Thank You!


JAGUAR
APPLIED MECHANICS
Railway Technical Research Institute
HLRS
CRAY XT5m
KRAKEN