RWTH AACHEN UNIVERSITY

Brainware for Green HPC

ENA-HPC

International Conference on Energy-Aware High Performance Computing
September 07–09, 2011

Hamburg

Christian Bischof

christian.bischof@tu-darmstadt.de

Dieter an Mey, Christian Iwainsky

{anmey,iwainsky}@rz.rwth-aachen.de

Overview

- TCO of HPC and Impact of Brainware
- Brainware Complexity
- HECToR dCSE Success Stories
- A Throughput Case Study
- Summary

Overview

- TCO of HPC and Impact of Brainware
- Brainware Complexity
- HECToR dCSE Success Stories
- A Throughput Case Study
- Summary

Total Cost of Ownership for HPC as a Service

RWTH AACHEN UNIVERSITY

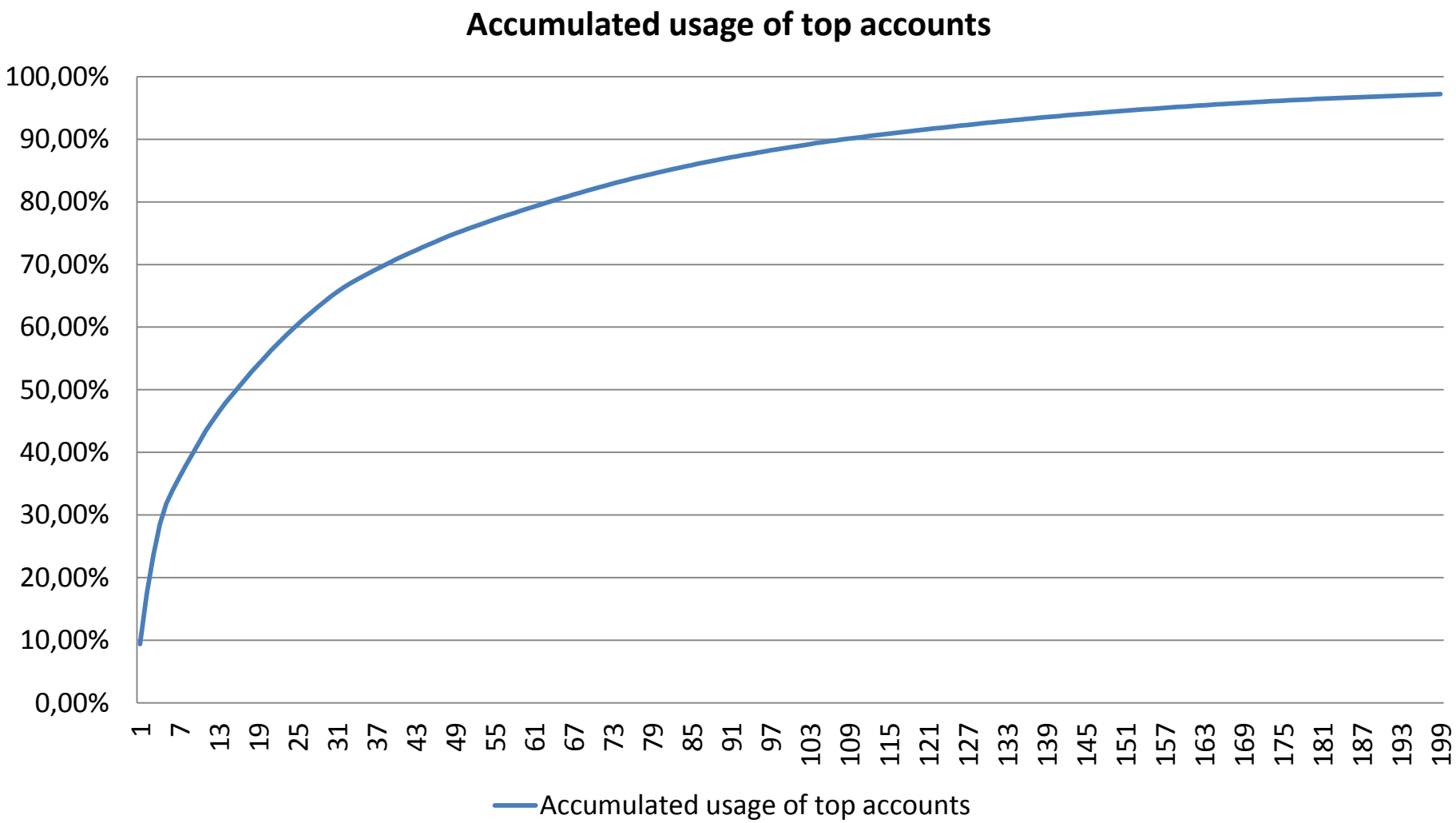
Assumptions

- 2 Mio € HW investment per year
- 5 years lifetime with 4 years maintenance through vendor
- 850 KW, PUE=1.5, 0.14€ per kWh => 1.5 Mio € per year
- ISV software provided by users
- Commercial batch system
- Free Linux distribution
- 4 FTE are for “brainware”

	costs per year	percentage
Building (7.5Mio / 25y)	300.000 €	5%
Investment compute servers	2.000.000 €	36%
hardware maintenance	800.000 €	14%
Power	1.564.000 €	28%
Linux	0 €	0%
Batch system	100.000	2%
ISV software	0 €	0 %
HPC software	50.000 €	1 %
Staff 12 FTE	720.000 €	13%
Sum	5.354.000 €	100%

Code Performance does not matter for TCO calculation

Usage Distribution



Does it pay to hire HPC Experts? – 1 of 2

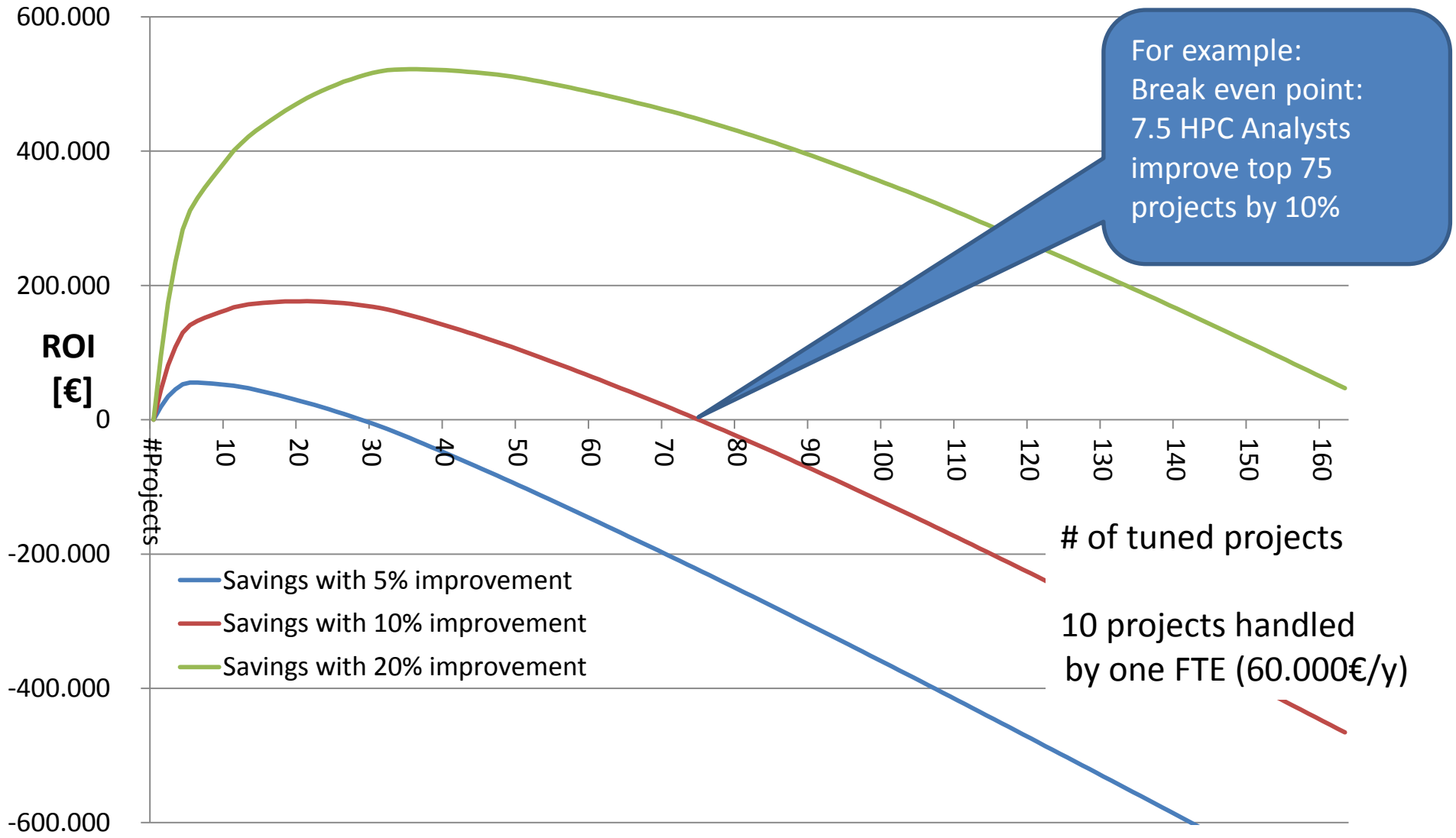
RWTH AACHEN UNIVERSITY

- **Start tuning top user projects first**
  - 15 projects account for 50% of the load
  - 64 projects account for 80% of the load

- **Assumptions**
  - It takes 2 months to tune one project
  - One analyst can handle 5 projects per year
  - A projects profits for 2 years
  - As a consequence one HPC expert

    can on average take care of 10 projects at a time in a year
  - One FTE costs 60,000€

Does it pay to hire HPC Experts? – 2 of 2



The Impact of Brainware

- **Brainware: Tuning Experts enhancing software performance and software life cycle in light of changing operating environments.**
- Even very moderate improvements in computational efficiency result in considerable savings.
- For example, a rather minuscule improvement of 5% on the top 30 projects "pays" for three HPC specialists.
  - If the performance is improved by 20 %, 0.5 Mio € are saved.
- **Energy savings account for a substantial part of the gain thus realized, i.e. brainware is an essential ingredient of green computing.**

Overview

- TCO of HPC and Impact of Brainware
- Tuning Complexity
- HECToR dCSE Success Stories
- A Throughput Case Study
- Summary

Opportunities for Tuning without Code Access

- **Sanity Check**
  - Use HW Counters
  - Employ Performance Analysis Tools
  - IO behavior
  - System call statistics
- **Hardware**
  - Choose the optimal hardware platform
  - File system, IO parameters
- **Parameterization**
  - Choose optimal number of threads / MPI processes
  - Thread / Process Placement (NUMA)
  - Mapping MPI topology to hardware topology
  - MPI parameterization (buffers, protocols)
  - Optimal libraries (MKL …)

Opportunities for Tuning with Code Access

- **Without Code Changes**
  - Choose the optimal compiler and optimal compiler options
  - Autoparallelization, compiler profile / feedback
  - Adapt dataset – partitioning / blocking – load balancing
- **Cache Tuning**
  - padding, blocking, loop based optimization techniques, inlining/outlining
- **MPI optimization**
  - Avoid global synchronization, coalesce communications
  - Hide / reduce communication overhead, Unblocking communications
- **OpenMP optimization**
  - Extend parallel regions, avoid false sharing
  - NUMA optimization: first touch, migration
- **In vogue: Add OpenMP to an MPI code to improve scalability**
- **Of Course: Crucial to choose the optimal algorithm**
  - To be handled by or with the domain expert

Building Brainware

- The skills just shown are typically not taught to code developers.
- It takes experience and skill to pick the most efficient tuning path & tools on a particular hardware platform.
- As academic computing is typically “free”, appreciation for those skills is often lacking.
- As a result, “tuning expert” is a rare career path at academic institutions.

- **Unless brainware becomes a standard ingredient in HPC operations (i.e. software is viewed as part of HPC infrastructure), money is being wasted.**

Overview

- TCO of HPC and Impact of Brainware
- Tuning Complexity
- HECToR dCSE Success Stories
- A Throughput Case Study
- Summary

HECToR Computational Science & Engineering Service

RWTH AACHEN UNIVERSITY

- **HECTor is the UK supercomputer service (Cray XE6 System).** ***http://www.hector.ac.uk/cse***

- **Part of the procurement was a service to make sure that users were supported/trained in making good use of the hardware**

- **This bid was won by the Numerical Algorithms Group (NAG).**
- **Central component (staff at NAG):**
  - Advice on using the system, non-invasive tuning, profiling
- **Distributed component (staff on-site)**
  - Panel reviews proposals for code improvement (software engineering, not implementation of new science).
  - Early grants up to 2 yrs, currently up to 1 yr.
  - Contracts (not grants) awarded

HECToR Distributed CSE Service Success Stories

RWTH AACHEN UNIVERSITY

Code	Domain	Effect	Effort	Saving
CASTEP	Key Materials Science	4x Speed and 4x Scalability	8 PMs	320k - 480k £ (p.a.)
NEMO	Oceanography	Speed and I/O-Perform.	6 PMs	95 k £ (p.a.)
CASINO	Quantum Monte-Carlo	4x Performance and 4x Scalability	12 PMs	760 k £ (p.a.)
CP2K	Materials Science	12 % Speed and Scalability	12 PMs	1500 k £ (in total)
GLOMAP/ TOMCAT	Atmospheric Chemistry	15 % Performance	?	
CITCOM	Geodynamic Thermal Convection	30% Performance	?	significant
Incompact 3D	Fluid Turbulence	6.75x Speed and 16x Scala-bility	12 PMs	
ChemShell	Catalytic Chemistry	8x Performance	9 PMs	
Fluidity-ICOM	Ocean Modelling	Scalability	?	
DL_POLY_3	Molecular Dynamics	20x Performance	6 PMs	
CARP	Heart Modelling	20x Performance		

http://www.hector.ac.uk/cse/reports/

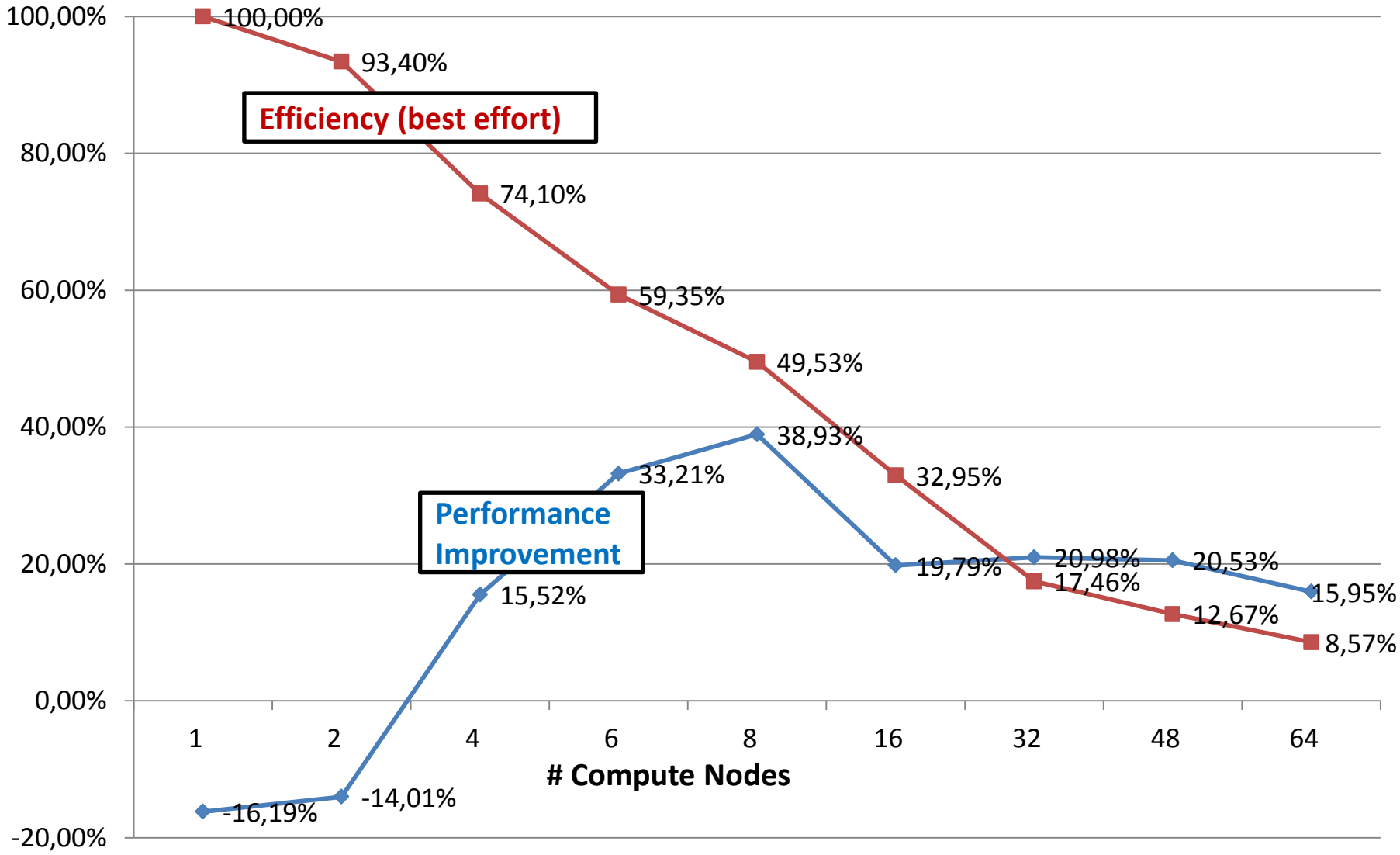
Overview

- TCO of HPC and Impact of Brainware
- Tuning Complexity
- HECToR dCSE Success Stories
- A Throughput Case Study
- Summary

Impact of Brainware on Throughput

- XNS code, developed at the Institute for Computer Analysis of Technical Systems at RWTH Aachen University (Prof. M. Behr, www.cats.rwth-aachen.de)
- Parallel finite element (FE) solver
- Satisfactory scalability on up to 4096 processors on a Blue Gene/L system using MPI parallelization.
- **Also extensively used in parameter studies involving smaller problems on a few cluster nodes.**
- In an effort off roughly six weeks, nine parallel regions were introduced into the most compute intense program parts.
- Experimental Results on QDR Infiniband-Cluster, nodes with two Nehalem EP processors each (3 GHz, 4 cores per processor chip). Serial time ~ 20 Minutes.

XNS: Impact of Hybrid Parallelization



Impact of Brainware on Throughput Computing

RWTH AACHEN UNIVERSITY

- Interested in the impact of code tuning on configurations where the parallel efficiency is relatively high (i.e. adding hardware is an economically sensible way to improve code performance)

- If we accept a decline of efficiency down to 50 percent, then the tuning effort delivers an improvement of up to 39 percent on 8 nodes.

- **So brainware is as important for capacity computing as it is for capability computing.**

Overview

- TCO of HPC and Impact of Brainware
- Tuning Complexity
- HECToR dCSE Success Stories
- A Throughput Case Study
- Summary

Summary

- We need to take a holistic view of cost effectiveness and computing efficiency: It makes more sense to invest in brainware rather than buy more inefficiently used “green” hardware.
- Higher investment in brainware pays off.
- HPC experts are a rare species requiring extensive training.
- Current and upcoming architectures require even more expertise (e.g. vector/multicore/distributed/cloud programming paradigms) so the brainware component becomes ever more important.
- HPC funding policies, educational curricula, and career development paths must recognize need for brainware.

Acknowledgements

- **Thanks to**

  N. Berr, J. Dietter, A. ElShekh, I. Ierotheou, C. Iwainsky, L. Jerabkova, S. Johnson, A. Gerndt, J.H. Goebbert, T. Haarmann, I. Hörschler, P. Leggett, D. Schmidl, Z. Peng, H. Pflug, T. Preuß, S. Sarholz, S. Siehoff, A. Spiegel, A. Wolf