Report on the Birds-of-a-Feather Session at SC´13 in Denver

Cost-Benefit Quantification for HPC: An Inevitable Challenge

(Wednesday, November 20, 2013, bof167)

Contributors

- Thomas Ludwig, German Climate Computing Centre, Hamburg, Germany
- Sandra Wienke, RWTH Aachen University, Aachen, Germany
- Albert Reuther, MIT Lincoln Laboratory, Lexington, MA, USA
- Amy Apon, Clemson University, Clemson, SC, USA
- Earl Joseph, IDC Corporate, Framingham, MA, USA

Audience: the BoF attracted 50 participants. Most of them stayed until the end of the discussion.

High Performance Computing (HPC) is an indispensable though expensive technology. While the discussion of cost has been part of the conversation for many years, only recently have we also seen the analyses of its benefits. Quantification is difficult, and there is a need for robust economic models to evaluate the cost-benefit ratio and effects to efficiency in the industrial process and in scientific development. The BoF presented approaches to this issue and fostered a discussion between users and providers. A better understanding of the methods to quantify and qualify the benefit and cost aspects will result in cost efficient computational science and engineering.

The BoF started with a short presentation by Thomas Ludwig about the costs and benefits of HPC in computational science and engineering. Sandra Wienke talked about the cost-benefit aspects of using special hardware accelerators for parallel programs. Albert Reuther provided an overview of various ways to capture and quantify benefits of HPC for organizations. Amy Apon presented non-parametric efficiency estimators and their application in analyzing their use in assessing the effect of locally-available HPC resources on university research productivity. Earl Joseph presented IDC´s new project on this issue.

Presentations lasted about 65 minutes. After each talk we allowed a short three minutes discussion round. After the final presentation we had a general discussion of about 25 minutes.

The discussion covered many specific aspects of the individual talks. Several more general aspects were discussed and agreed upon by a large part of the audience:

- Academia knows very well how to buy huge systems. The knowledge in operating it cost-efficiently and also in using it cost-efficiently should be improved. Awareness on the side of the end users is still low.
- Experts in using HPC environments must pass this knowledge on to the graduate students.
- The BoF helps to raise the awareness for the issue of benefit and efficiency quantification.

New resources:
People in the audience asked for possibilities for further discussions and for further information material. A mailing list for discussions and information dissemination will be created in Jan 2014, and a web site will be launched by end of Jan 2014.

- Mailing list at http://wr.informatik.uni-hamburg.de/cgi-bin/mailman/listinfo/cbq4hpc
- Small web site with additional material at http://www.cbq4hpc.org/

http://bit.ly/sc13-eval

Birds-of-a-Feather Session

Cost-Benefit Quantification for HPC

An Inevitable Challenge

Thomas Ludwig – German Climate Computing Center, Germany
Sandra Wienke – RWTH Aachen, Germany
Albert Reuther – MIT Lincoln Laboratory, USA
Amy Apon – Clemson University, USA
Earl Joseph – IDC Corporate, USA

Abstract

HPC is an indispensable though expensive technology. While the discussion of cost has been part of the conversation for many years, only recently have we also seen the analyses of its benefits. Quantification is difficult and we still lack an economical model to evaluate the cost-benefit ratio, and effects to efficiency, in the industrial process and in scientific development.

This BoF will present approaches to this issue and foster a discussion between investors and users.

Better understanding of the methods to quantify and qualify the benefit and cost aspects will result in cost efficient computational science and engineering.

Resources

- BoF Evaluation Form (please, please, please fill in!)
  http://bit.ly/sc13-eval
- SC´13 program with BoF announcement
  http://sc13.supercomputing.org/schedule/event_detail.php?evid=bof167
- Mailing list Cost-Benefit Quantification for HPC
  Subscribe here:
  http://wr.informatik.uni-hamburg.de/cgi-bin/mailman/listinfo/cbq4hpc
  in order to be in the CBQ4HPC mailing list
- We will set up a small web site with a collection of information and links
  http://www.cbq4hpc.org/

How it Began

SC´12 in Salt Lake City

Invited talk of Thomas Ludwig with the title

“The Costs of HPC-Based Science in the Exascale Era”

Cited some interesting research papers and agreed with the authors upon the submission of a BoF proposal for SC´13

Why this Topic?

- HPC is very costly and will become extremely expensive in the Exascale era – top systems will cost several $100 millions
- HPC is indispensable for progress in science
- HPC is indispensable for production in industry

- As it is so expensive we should try to maximize its benefit
- However, how can benefit be quantified?
- And can we properly quantify the costs?
- We also need a metric for purchase decisions of systems
- Define a business modell for HPC-based science and engineering

Goal of the BoF

- Give an overview over existing ideas and approaches
- Stimulate discussions between investors, service providers, and users
- Get economy specialists involved
- Learn from business considerations in industry
- Make the business case for HPC-based science and engineering
- Optimize the benefit-cost ratio for HPC-based science and engineering

Protagonists

- Thomas Ludwig is CEO of DKRZ and professor for computer science at the University of Hamburg. His research focus is on parallel I/O and energy efficiency.
- Sandra Wienke is a research assistant at the Center for Computing and Communication at RWTH Aachen University, Germany. Her research interests include HPC on parallel heterogeneous architectures with emphasis on productivity and performance
- Dr. Albert Reuther is the Assistant Group Leader of the Computing and Analytics group at MIT Lincoln Laboratory. In this role, he leads projects in HPC for rapid-prototyping, Big Data analytics, and novel computer architectures.
- Dr. Amy Apon is Professor and Chair of the Computer Science Division in the School of Computing at Clemson University. She also leads the Big Data Systems Research lab and is active in the academic scientific computing community.
- Dr. Joseph heads up IDC's HPC programs. IDC tracks all computers sold around the world each quarter. Earl also leads IDC' HPC User Forum, which gas held 51 meetings. He also does consulting projects fir major HPC user sites and with vendors.

Who Will Present What?

- The BoF will start with a short presentation by Thomas **Ludwig**, German Climate Computing Centre (DKRZ), about the costs and benefits of HPC in computational science and engineering.
- Sandra **Wienke**, RWTH Aachen University, Germany, will talk about the cost-benefit aspects of using accelerators.
- Albert **Reuther**, MIT Lincoln Laboratory, will provide an overview of various ways to capture and quantify benefits of HPC for organizations.
- Amy **Apon**, Clemson University, will present nonparametric efficiency estimators and their application in analyzing their use in assessing the effect of locally-available HPC resources on university research productivity.
- Finally, Earl **Joseph** from IDC will present a new project on this issue.
- **We** will discuss with participants their suggestions and ideas to conduct further research in this important field. Presentation slides and results of the discussion will be publicly available.

Thomas Ludwig
German Climate Computing Center (DKRZ)
Hamburg, Germany

- DKRZ costs the tax payer €16 million per year
- It consumes 17 GWh of electricity per year which cost €2.5 million
- Climate researchers produce many papers to participate in the IPCC Assessment Reports
- What is the papers per kWh ratio?

Sandra Wienke
RWTH Aachen
Aachen, Germany

- Accelerators may boost performance in technical computing
- But, are they really worth the pain?
- Benefit-cost ratio approach by RWTH case study
    - o Benefit: scientific results (= application runs)
    - o Cost: total cost of ownership (TCO)

Albert Reuther
MIT Lincoln Laboratory
Lexington, MA, USA

- Each organization must weigh the impact priority of HPC activities
- Find a way to concretely quantify the benefit in order to compare costs
- The benefit of HPC to an organization depends on a number of factors
  - E.g. time saved per researcher or research team
  - E.g. impact on product sales.

Amy Apon
Clemson University
Clemson, SC, USA

- How to use nonparametric efficiency estimators to quantify the effect of locally-available HPC resources on research output
- Shows evidence that HPC instrumentation has a positive effect on universities' efficiency
- Locally-available HPC resources enhance the technical efficiency of research output in only some sciences
- The methods presented provide a critical step in a quantitative economic model for investments in HPC.

Earl Joseph
IDC Corporate
Framingham, MA, USA

- DOE funded research program aims at creating two macroeconomic models to show the linkage between investments in HPC and the resulting financial ROI and innovation
- The models also include a new innovation index

Questions

- Would you like to participate in the next phase of the ROI study?
- Would you like to become a spokesperson on the ROI topic in your sector or for your country?

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

DKRZ
DEUTSCHES KLIMARECHENZENTRUM

The Cost / Benefit Situation at DKRZ

Prof. Dr. Thomas Ludwig

German Climate Computing Centre & University of Hamburg
Hamburg, Germany
ludwig@dkrz.de

DKRZ in Hamburg

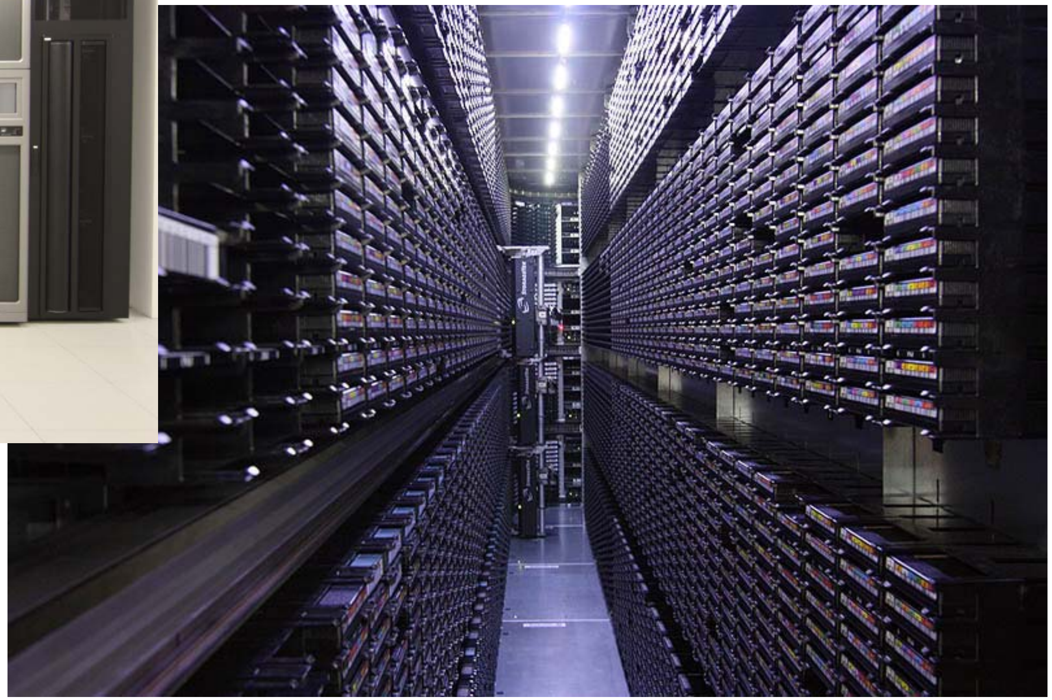
Deutsches Klimarechenzentrum (DKRZ)

German Climate Computing Centre

IBM Power6 Computer System

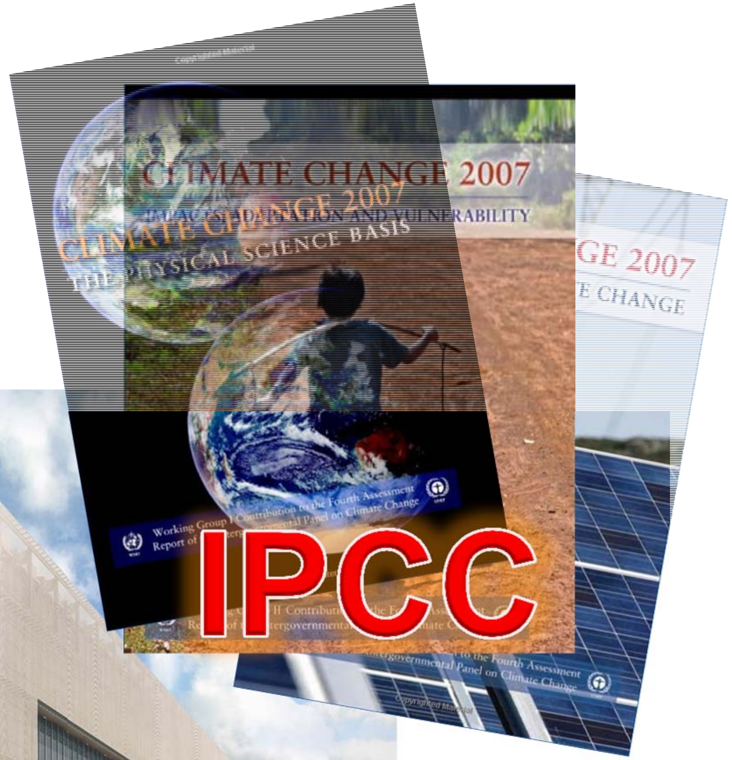
- Rank 368 in TOP500/Jun13
- 8,064 cores, 115 TFLOPS Linpack
- 6PB disks

Sun StorageTek Tape Library



- 100 PB storage capacity
- 90 tape drives
- HPSS HSM system

Cost-Benefit Modell at DKRZ



Energy Costs for Science

5th IPCC status report:

- German part uses ca. 30M corehours at DKRZ
- DKRZ offers ca. 60M corehours/y
- Energy costs for the German IPCC contribution: **ca. €1 m**
  - **9,000,000 kWh to solution** with DKRZ´s Blizzard system
  - 4,500 metric tons of CO_2 with regular German electricity

Climate researchers should predict the climate change…

… and not produce it!

Total Costs for Science Support at DKRZ

TCO of DKRZ per year: approximately **€16M**

€8M hardware, €2M electricity, €3M brainware

Publications per year: let´s assume 100

Mean price per publication: €160,000

+ costs for scientists ☺

It is tax money – society expects a benefit

Example IPCC AR5 production runs

- Remember

  Energy costs for the German IPCC contribution: ca. 1 M€

  – **9,000,000 kWh to solution** with DKRZ´s system

  – 4,500,000 kg of CO_2 with regular German electricity

- Approach: Tune program and save 10% runtime
  - Saves 900,000 kWh
  - Saves €100,000 (is one person year)
  - Saves 450 metric tons CO_2

Invest in people !

We need more HPC specialists

- Co-design and code development
- Tuning of applications
- many other things...

Gigadollars for iron and electricity

will not be the solution !

Accelerators in Technical Computing: Is it Worth the Pain?

A TCO Perspective

Sandra Wienke, wienke@rz.rwth-aachen.de

Dieter an Mey, Matthias S. Müller

Cost-Benefit Quantification for HPC - An Inevitable Challenge

BOF, SC 2013, Denver

Publication: ISC’13, LNCS 7905, pp. 330-342

Center for Computing and Communication, Rechen- und Kommunikationszentrum (RZ)

- **How it began**

Accelerators – Awesome or Harmful?

- **Decision making by "worth"**
  - → Purchase one system or another
  - → Use one programming model or another

$$\mathbf{CBR} = \frac{cost}{benefit} = \frac{TCO}{\#app.runs}$$

- **Cost-benefit quantification (case study)**
  - → **Cost**: TCO (total cost of ownership)
    - →Include costs for manpower (development, administration)
  - → **Benefit**: "amount of science"
    - →Expressed in number of results within time period
    - →Here: number of runs of a single application (w/o weights)

Methodology

- **Aim: compare architectures (& programming models)**
  - → Here: Intel Xeon server, NVIDIA Fermi GPU, Intel Xeon Phi
- **Modeling depends on architecture & application**
  - → Here: (small) engineering application with 90% kernel portion
- **Basis: single compute node ➔ extrapolate to cluster amount (no MPI)**
- **Investment** $\boldsymbol{I = \mathrm{TCO}(n,\tau) = C_{ot}(n) + C_{pa}(n) \cdot \tau}$
  - → One-time costs C_{ot}, e.g. HW acquisition, prog. effort
  - → Annual costs C_{pa}, e.g. compiler/software, power consumption, app. maintenance

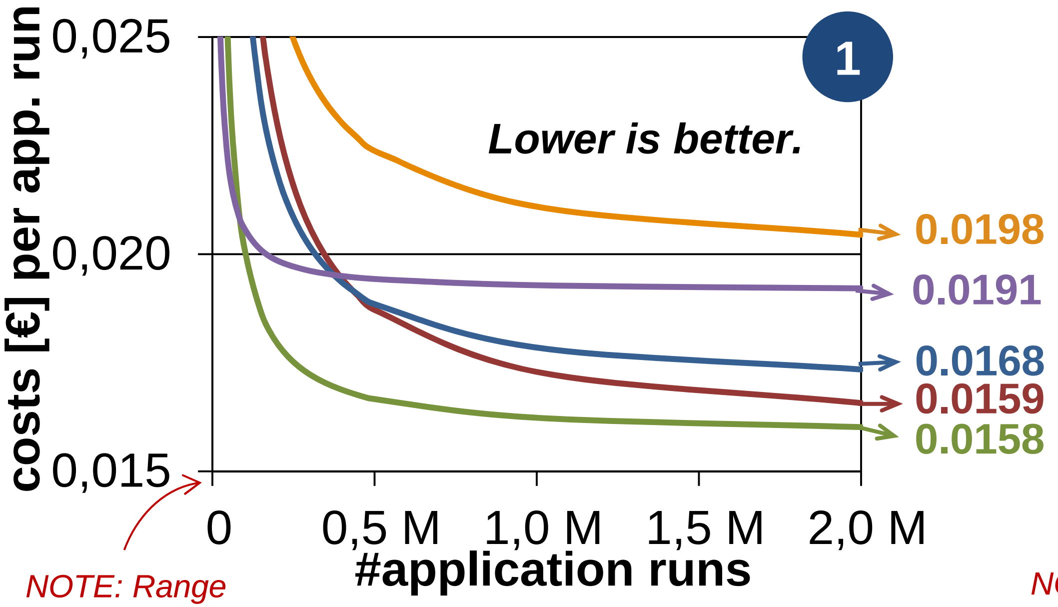
n :	number of nodes
τ :	system lifetime

- **Cost-benefit ratio (CBR): costs per application run C_{ar}**
  - → Includes investment/ TCO & application performance

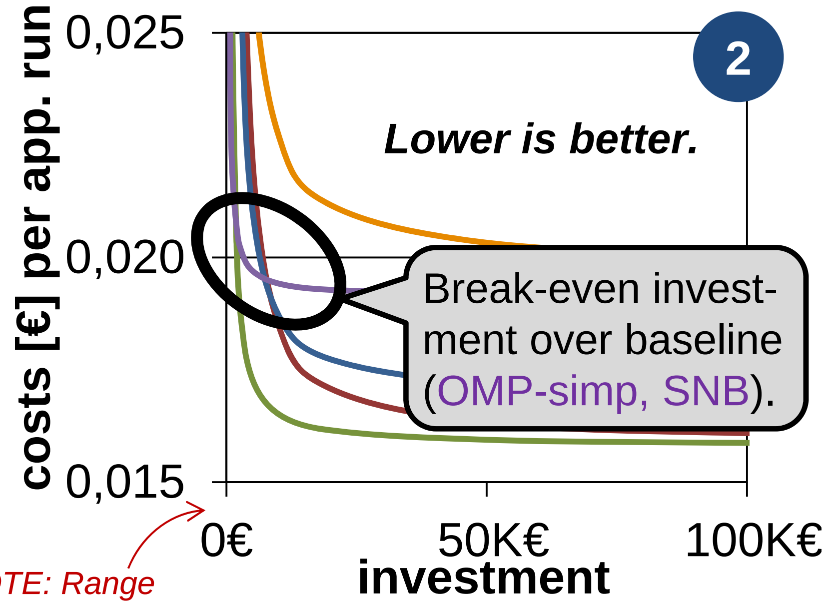
$$\mathbf{CBR} = \frac{cost}{benefit} = C_{ar}(n,\tau) = \frac{\mathrm{TCO}(n,\tau)}{\mathrm{ar}(n,\tau)} \quad \text{with} \quad \mathrm{ar}(n,\tau) = n \cdot \frac{\tau \cdot usagerate}{t_{par}}$$

Results of Case Study

- **Target: decrease** $\mathbf{CBR} = \frac{cost}{benefit} = \frac{TCO}{\#app.runs}$
  1. Keep benefit, decrease costs
  2. Keep cost, increase benefit
  3. Both: decrease costs + increase benefit
- **Case study**



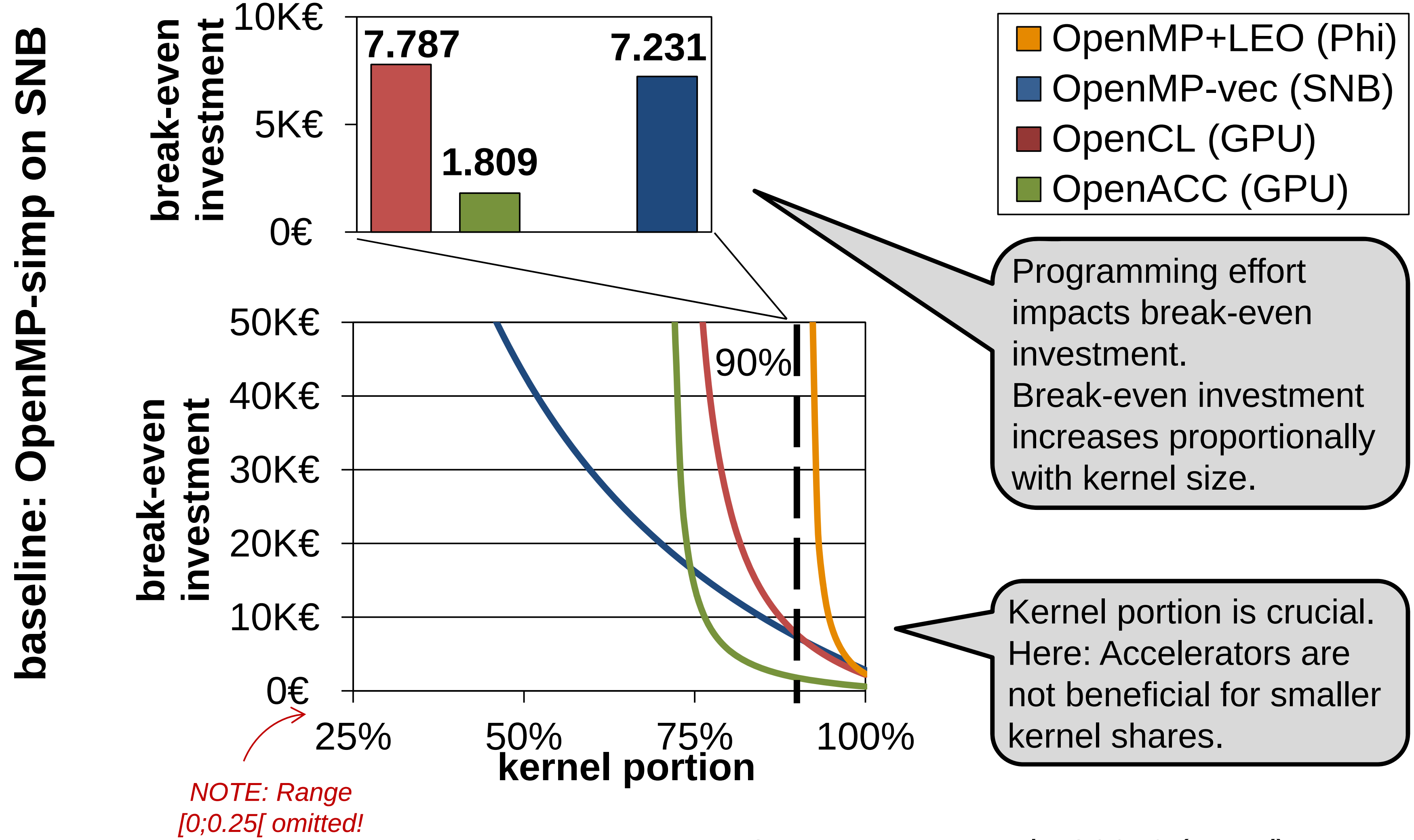
NOTE: Range [0;0.015[omitted!



NOTE: Range [0;0.015[omitted!

GPU: NVIDIA Tesla C2050 (Fermi)
Phi: Intel Xeon Phi 5110P, 60 cores
SNB: 2x Intel Sandy Bridge,16 cores, 2 GHz

Results of Case Study



Programming effort impacts break-even investment.
Break-even investment increases proportionally with kernel size.

Kernel portion is crucial.
Here: Accelerators are not beneficial for smaller kernel shares.

GPU: NVIDIA Tesla C2050 (Fermi)
Phi: Intel Xeon Phi 5110P, 60 cores
SNB: 2x Intel Sandy Bridge,16 cores, 2 GHz

Summary & Conclusion

Thank you for the attention!

- **Costs: TCO of HPC cluster environment**
  - → Including costs for manpower effort
  - → Difficult to quantify all values ⇔ but, (fix) metrics
- **Benefit: scientific results ~ #app. runs**
  - → Some weaknesses ⇔ but, easy number & useful for first approach
- **Cost-benefit ratio: cost per app. run**
  - → Depending on own environment & application
  - → Useful to compare systems (& prog. models)
- **Our CBR case study results (no general conclusion)**
  - → Fermi GPU good, Xeon Phi not
    - → Account sensitivity about input parameters!
      (e.g. Xeon Phi's high acquisition costs had main impact)
  - → Accelerators only worth w/ high amount of kernel portion

Spreadsheet for own modeling

Wienke, S., an Mey, D., Müller, M.S.: Accelerators for Technical Computing: Is it Worth the Pain? TCO Spreadsheet. https://sharepoint.campus.rwth-aachen.de/units/rz/HPC/public/Shared%20Documents/WienkeEtAl_Accelerators-TCO-Perspective.xlsx, 2013

http://bit.ly/sc13-eval

Quantifying Benefit for HPC

Albert Reuther
MIT Lincoln Laboratory

IEEE/ACM Supercomputing Birds of a Feather

November 20, 2013

LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

This work is sponsored by the Department of the Air Force under Air Force contract FA8721-05-C-0002. Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the United States Government.

Cost and Benefit

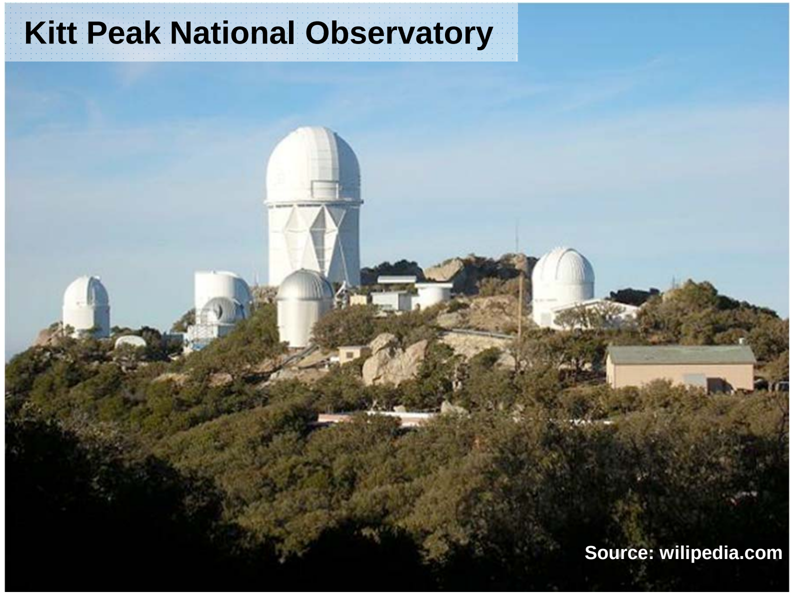
? ? ?

$$\text{ROI} = \frac{\text{Benefit}}{\text{Cost}}$$

Cost = (time to parallelize) + (time to train) + (time to launch) + (time to admin.) + (system cost)

- time to parallelize: Researcher parallel coding time
- time to train: Researcher training time
- time to launch: Queuing and launch overhead
- time to admin.: Sys-admin cost, Expert consultants
- system cost: Computer cost, Facilities cost, Utilities

Capability vs. Capacity Computing



Capability Computing

- National shared asset
- Enables previously unreachable results
- Focus: Wisely use precious resource

Capacity Computing

- Many more powerful binoculars
- Computational time/effort/force multiplier
- Focus: Wisely make users more productive

Various Methods of Benefits

Financial and Competitive Return

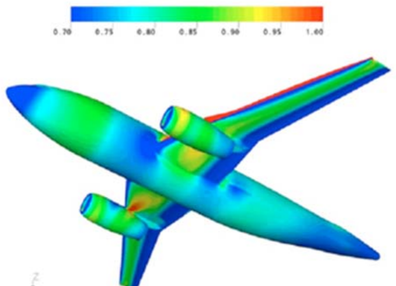
- **ROI**
  - Shorter time to market
  - Shortened product development cycle
  - Effective supply chain mgmt.
- **Reduced design costs**
- **Fewer physical tests**
  - Quality and reliability
  - Risk reduction
  - "Explore digitally, confirm physically" (mantra at P&G)
- **Critical to business/product development**
- **Breakthrough products**

Research Return

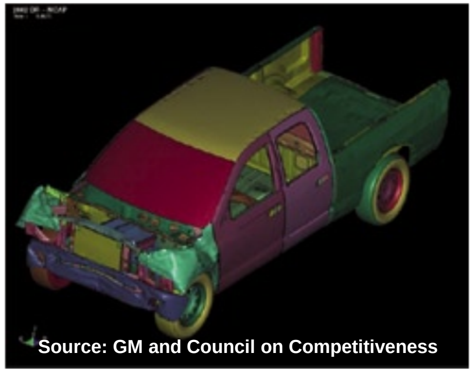
- **Research productivity**
- **Broader insight**
- **Research funding/competitiveness**
- **Entries in Top500 list**
- **Cover articles in *Science***
- **Number of NSF grants**

Council on Competitiveness: Compete.org

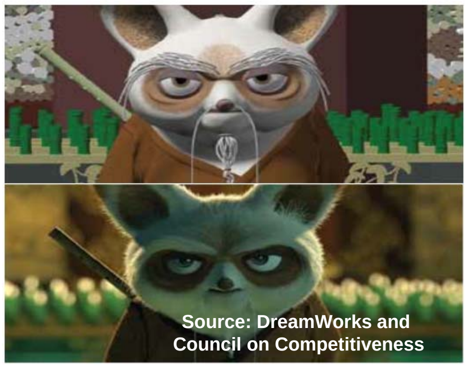
Industry Benefits – Examples



Source: Boeing and HPC for Energy

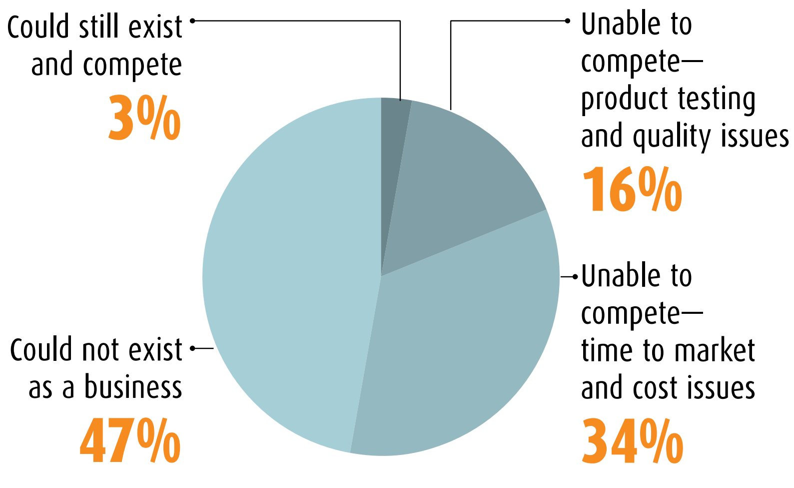


Source: GM and Council on Competitiveness

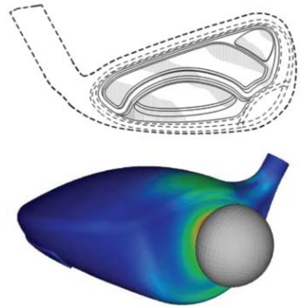


Source: DreamWorks and Council on Competitiveness

Why Does HPC Matter? (Industry Survey)



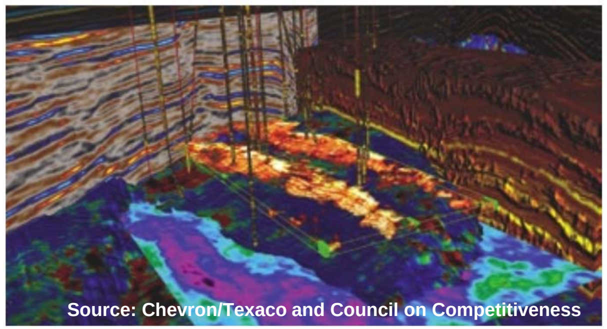
Source: The Council on Competitiveness and IDC



Source: PING and Council on Competitiveness

Source: Procter & Gamble Co. and Council on Competitiveness

- **Aircraft Design**
- **Vehicle Crash Tests**
- **Movie Rendering**
- **Sports Equipment Validation**
- **Container Optimization**
- **Oil Field Exploration**

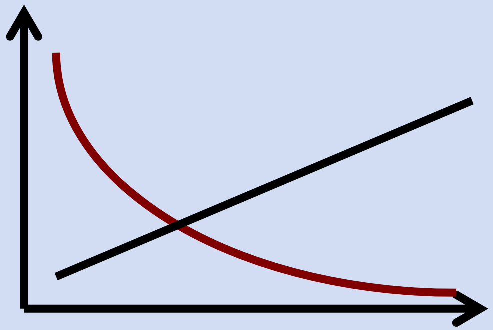


Source: Chevron/Texaco and Council on Competitiveness

Council on Competitiveness: http://www.compete.org/
HPC for Energy: http://hpc4energy.org/

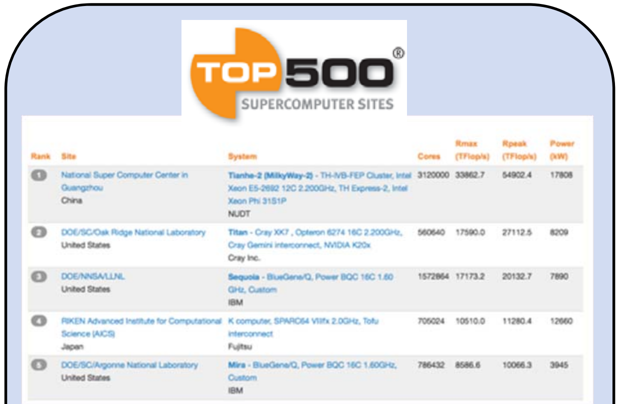
Research Benefits

Research Productivity



Broader insight/
risk reduction

Research
competitiveness



Rank	Site	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	National Super Computer Center in Guangzhou China	Tianhe-2 (MilkyWay-2) - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P NUDT	3120000	33862.7	54902.4	17808
2	DOE/SC/Oak Ridge National Laboratory United States	Titan - Cray XK7 , Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x Cray Inc.	560640	17590.0	27112.5	8209
3	DOE/NNSA/LLNL United States	Sequoia - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom IBM	1572864	17173.2	20132.7	7890
4	RIKEN Advanced Institute for Computational Science (AICS) Japan	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect Fujitsu	705024	10510.0	11280.4	12660
5	DOE/SC/Argonne National Laboratory United States	Mira - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM	786432	8586.6	10066.3	3945

Entries in Top500 list

Cover articles in
Science/Nature/Cell/etc.

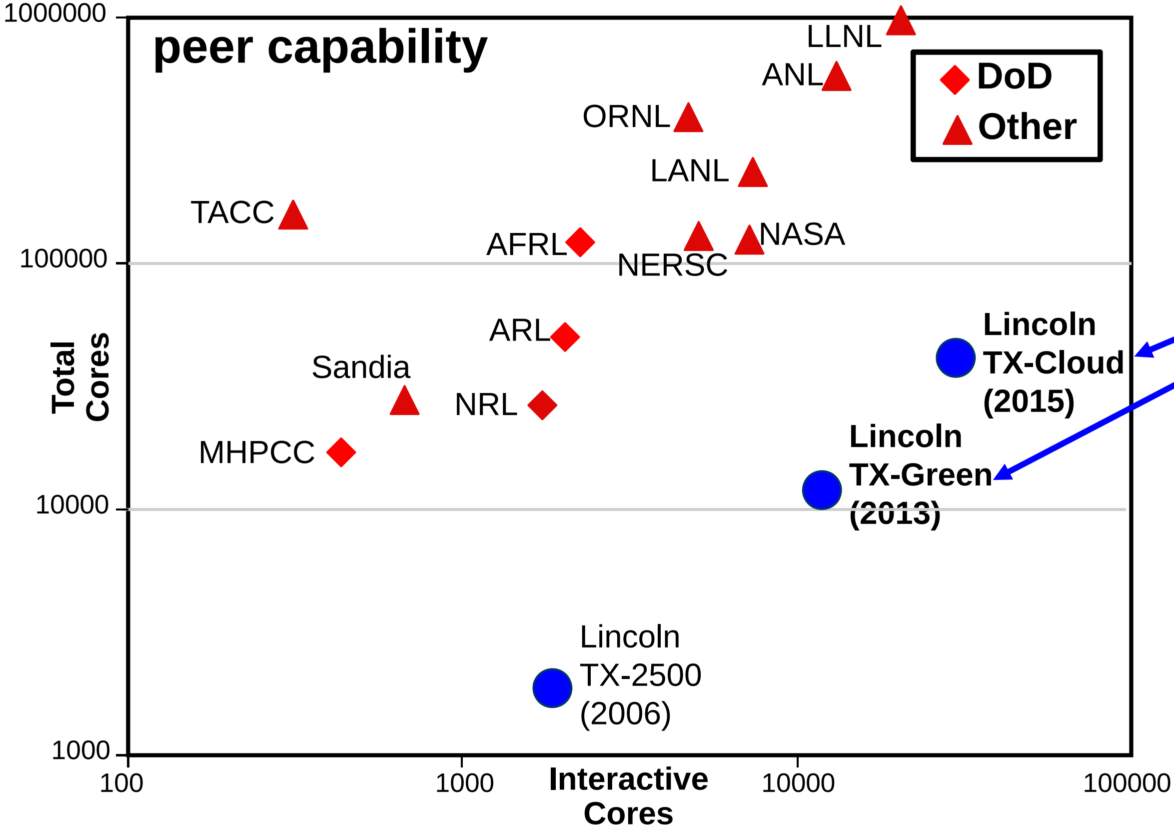
Number of NSF and
other grants

Council on Competitiveness: http://www.compete.org/

A. Apon, S. Ahalt, V. Dantuluri, C. Gurdiev, M. Limayem, L. Ngo, and M. Stealey. High Performance Computing and Research Productivity in U.S. Universities, *Journal of Information Technology Impact*, Vol. 10, Issue 2, 2010. pp. 87-98.

S. Tichenor and A. Reuther, "Making the Business Case for High Performance Computing: A Benefit-Cost Analysis Methodology," CyberTech. Watch Quarterly, Vol. 2,:4A, November 2006 A.

Lincoln's Leadership in Interactive, On-Demand Supercomputing



HP EcoPOD: high efficiency, deployable data center

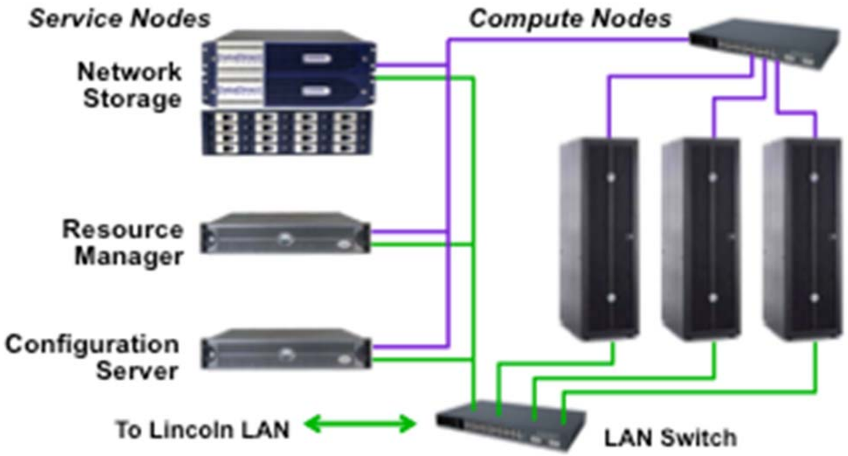
EcoPOD Capacity:
- **1500 nodes**
- **0.5 Petaflops**
- **0.4 Petabyte RAM**
- **15 Petabyte Disk**
- **44 racks**
- **1 Megawatt**

- **Lincoln's research mission requires interactive supercomputing**
  - **Primarily for the development of sensor processing systems**
- **Lincoln has the largest interactive capability in DoD**
- **EcoPOD facility allows competitive position to be maintained**

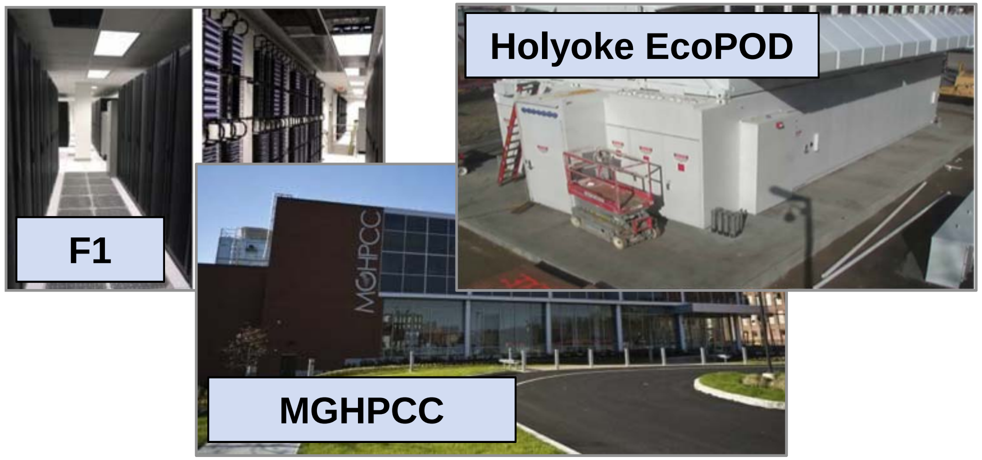
LLGrid Core Capabilities

LLGrid is a System-of-Systems

Common Interactive Architecture



Diverse Locations



Cluster(s)	TX-2500	TX-Green	~~TX-3D~~	TX-C	TX-2400	TX-BW	TX-CySA	TX-D2D	TX-SRC
Classification	Unclassified	Unclass	~~Class~~	Class	Class	External	CySA	External	Unclass
	Labwide	Labwide	~~Div3~~	PL-2 Labwide	PL2 SIPRNet	Open	Limited	U//FOUO	Labwide
	F1	EcoPOD	~~F1~~	F1 -> EcoPOD	F1->EcoPOD	MGHPCC	F1	F1	EcoPOD
Compute Nodes	~~578~~ 40	274	~~270~~	100	40	200	24	22	20
Compute Cores	~~1256~~ 1280	8768	~~540~~	3200	1280	6000	576	704	320
Peak Performance (TFLOPs)	~~16~~ 10	77.1	~~6.9~~	28.2	11	50	4.8	48.3	2
RAM (TB)	4	17.5	~~1.9~~	9.6	3.8	10	1.5	2.1	2.1
Central Storage (TB)	1,509	1,200	~~54~~	100	96	1,000	0	200	200
Distributed Storage (TB)	~~682~~ 360	2,466	~~108~~	900	240	1,000	288	132	132

Now **Now** **Winter** **Fall**

Quantifying the HPC Research Productivity Benefit

Reciprocal Jobsize Scaling

$$B = S \sum_{alljobs} (JobDur * (1 - \frac{1}{\#subjobs}))$$

Linear Jobsize Scaling

$$B = S \sum_{alljobs} (JobDur * \#subjobs)$$

Logarithmic Jobsize Scaling

$$B = S \sum_{alljobs} (JobDur * \log_2(\#subjobs))$$

$$ROI = \frac{Benefit}{Cost}$$

Cost = (time to parallelize) + (time to train) + (time to launch) + (time to admin.) + (system cost)

- time to parallelize: Researcher parallel coding time
- time to train: Researcher training time
- time to launch: Queuing and launch overhead
- time to admin.: Sys-admin cost, Expert consultants
- system cost: Computer cost, Facilities cost, Utilities

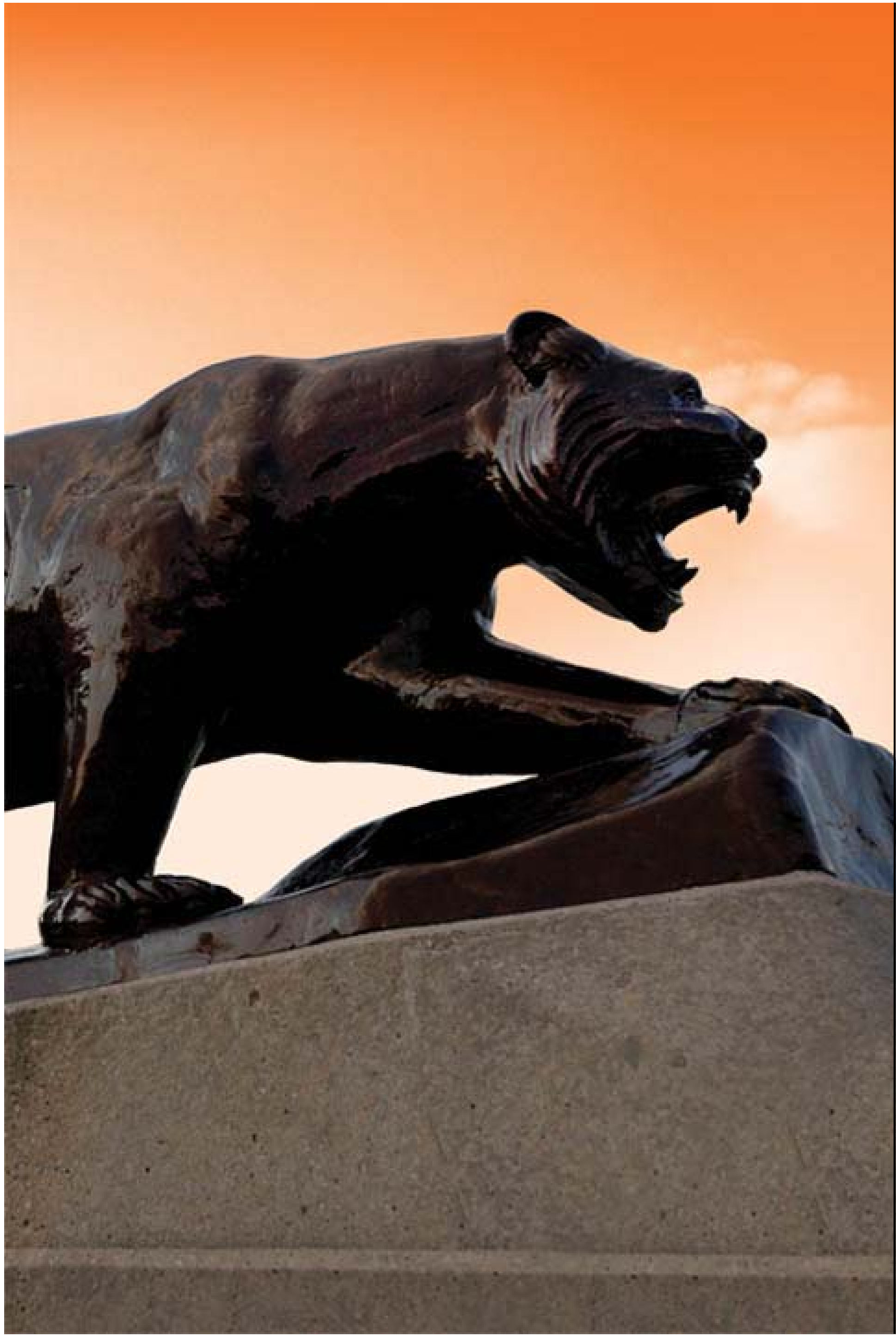
B = benefit
S = fully-burdened researcher salary

LLGrid Productivity Analysis

$$\text{productivity (ROI)} = \frac{\left(\text{time saved by users on system}\right)}{\left(\text{time to parallelize}\right) + \left(\text{time to train}\right) + \left(\text{time to launch}\right) + \left(\text{time to admin.}\right) + \left(\text{system cost}\right)}$$

Parameters	*FY2007 2 new systems*	*FY2010 no new systems*	*FY2013 3 new systems*
Lab-wide Users	364	711	1017
New Users in Latest Year	63	60	83
Active Users in Latest Year	105	153	218
Avg. Simultaneous Jobs	48	65	55
Avg. CPUs per Job	29	94	65
Total Job Launches	47,043	226,839	89,270
Number of System Administrators	7	6	7
Nodes in System	748	937	926
New Nodes in System (Latest Year)	748	0	414
Benefit/Cost (Linear: CPUs)	41.6	263.0	115.7
Benefit/Cost (Logarithmic: $\log_2$(CPUs))	4.6	27.6	9.1
Benefit/Cost (Reciprocal: 1-1/CPUs)	0.99	5.4	2.2

LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY



Efficiency as a Measure of Knowledge Production of Research Universities – Application to HPC ROI

Amy W. Apon* Linh B. Ngo*
Michael E. Payne* Paul W. Wilson+

School of Computing* and Department of Economics +

Clemson University

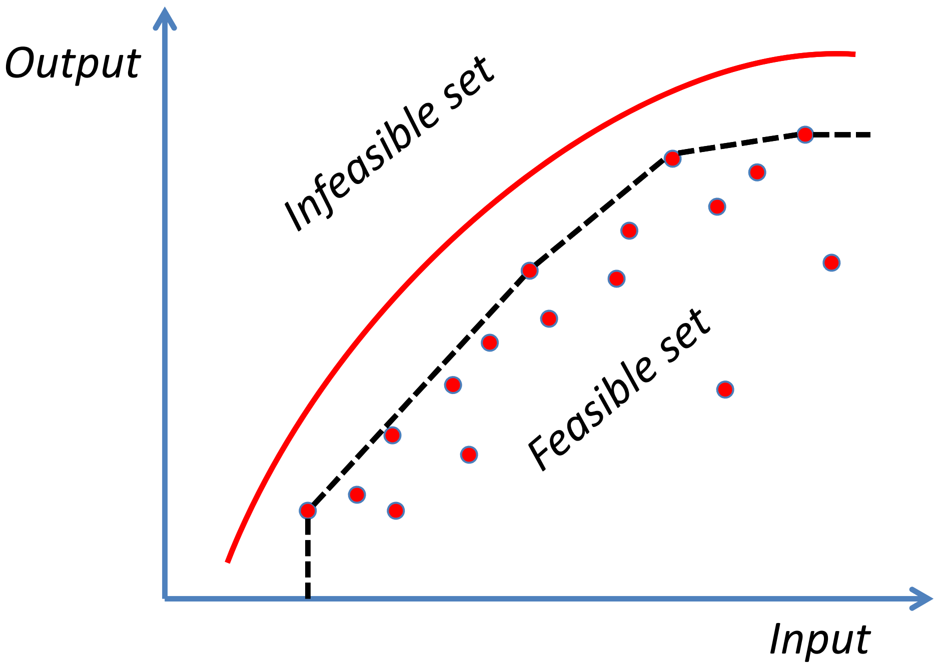
Using efficiency as a measure

- Extends traditional metrics
- Utilizes non-parametric estimation of relative efficiency of production units
- Avoids regression framework
- Takes scale of operation of institution into consideration
- Is supported by rigorous hypothesis testing

Efficiency is calculated as an estimate of a frontier of production

- We define P as the set of feasible combinations of *p* inputs and *q* outputs, also called the production set.

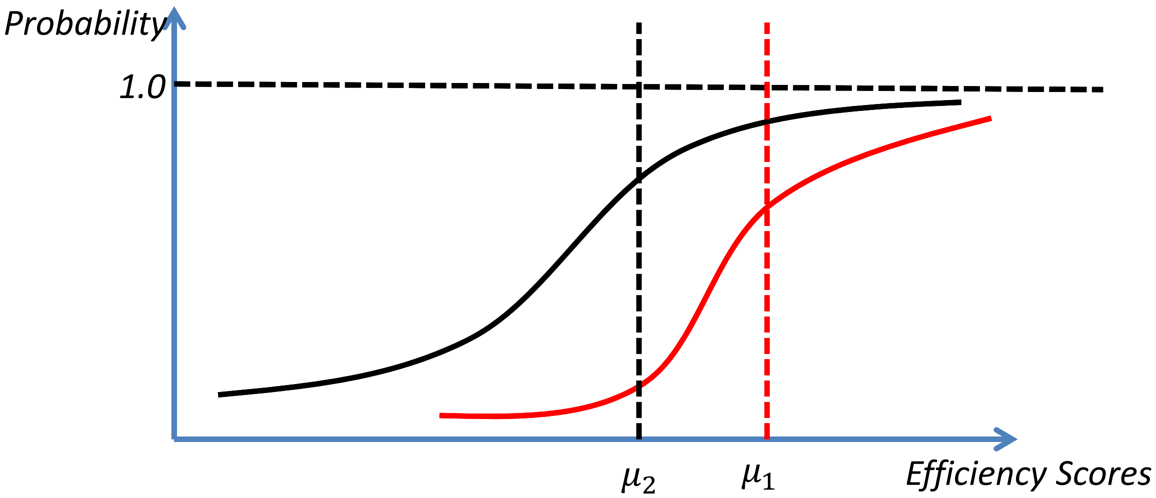
$$\mathcal{P} = \{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{p+q} \mid \boldsymbol{x} \text{ can produce } \boldsymbol{y}\}$$



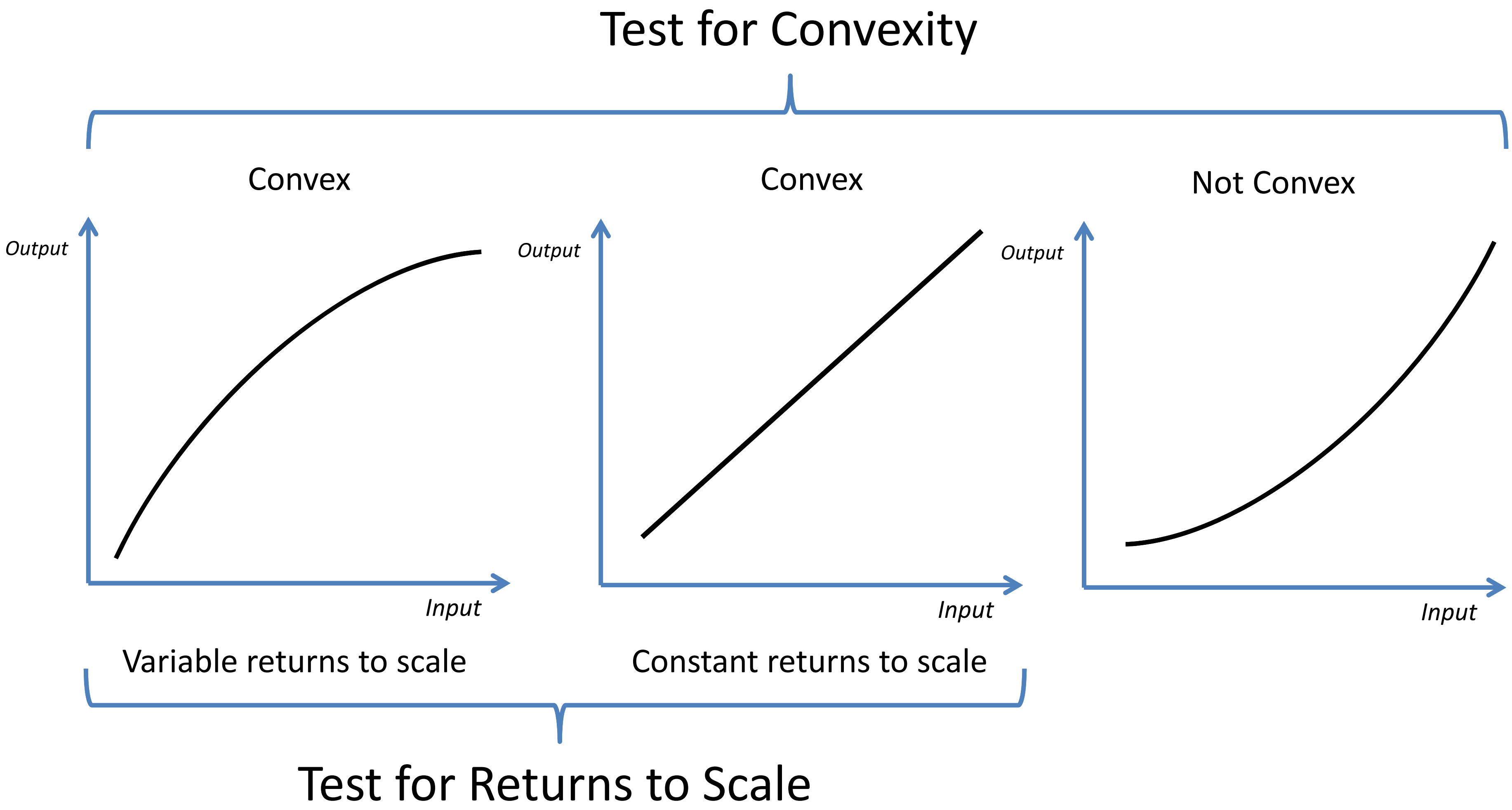
- The production set is closed.
- The efficiency score measures the distance from the frontier.
- Efficiency must be estimated since it is unobserved.

Tests for equivalent means and for stochastic dominance

- Test for equivalent means
  - $H_0: \mu_1 = \mu_2$ versus $H_1: \mu_1 > \mu_2$
- Test for first order stochastic dominance between the two efficiency distributions:
  - H_0: Distribution 1 does not dominate Distribution 2



Shape of frontier



HPC institutional data

- Leveraging the historical record of the Top 500 list
  - An historical record without comparison
- Steps in data preparation include corresponding the HPC site with its associated academic institutions
- Classification of HPC capability for our study
  - Have versus Have nots: Whether an institution has a Top500 system between 2000-2006

Department data description

- National Research Council: Data-Based Assessment of Research-Doctorate Programs in the U.S. for 2005-2006
- Input: Faculty Count, Average GRE Scores
- Output: PhD Graduates, Publication Counts
- 8 academic fields have sufficient data:
  - Chemistry
  - Civil and Environmental Engineering
  - Computer Science
  - Ecology and Evolutionary Biology
  - English
  - Economics
  - History
  - Physics

Tests for convexity, CRS

Department	Test for Convexity	Test for Constant Returns to Scale
Chemistry	0.47	0.06
Civil/Environmental	0.90	0.99
Computer Science	0.37	0.99
Ecology/Evolutionary Biology	0.73	0.99
Economics	0.44	0.28
English	< 0.01	--
History	< 0.01	--
Physics	0.21	0.99

Tests for equivalent means
Stochastic dominance

Have > Have not		
	Equiv. means	St. Dom.
Chemistry	< 0.01	< 0.01
Civil/Environmental	< 0.01	< 0.01
Computer Science	< 0.01	< 0.01
Ecology/Evolutionary Biology	0.99	--
Economics	1.00	--
English	< 0.01	0.95
History	< 0.01	0.65
Physics	< 0.01	< 0.01

Issues and Concerns

- Scarcity of useful data
  - Comparing different abstraction level needs datasets that can produce similar inputs/outputs
  - Longitudinal data
- Integration and curation of existing data
- A lower efficiency estimate does not mean a program is not doing well
  - Factors to consider include the characteristics of the admitted population, institutional goals and mission, etc.

Future Work

Our results raised interesting observations – why?

- What are the factors that make one set of institutions more efficient than the other?
- How is the nature of research different in more efficient institutions?

Efficiency as a Measure of Knowledge Production of Research Universities – Applications to HPC

Amy W. Apon* aapon@clemson.edu
Linh B. Ngo* lngo@clemson.edu
Michael E. Payne* mpayne3@clemson.edu
Paul W. Wilson+ pww@clemson.edu
School of Computing* and Department of Economics +

This research has been supported in part by NSF INSPIRE #123436 and the CI SEEDS project, NSF #1212680.

ROI And HPC Research Study:

Creating Economic Models For Financial ROI And Innovation From HPC Investments

November, 2013

Earl Joseph, ejoseph@idc.com
Steve Conway, sconway@idc.com
Chirag Dekate, cdekate@idc.com

Background: Project Overview

A pilot study that describes how HPC investments are related to improved economic success and increased scientific innovation

The study included creating two unique models:

1. A macroeconomic model which depicts how HPC investments result in economic advancements in the form of ROI, growth and jobs

2. An "Innovation Index" that measures and compares innovation levels, based on the level of applying HPC computing resources towards scientific and technical advancement

The Financial ROI Models That Were Developed

IDC
Analyze the Future

The Financial ROI models include:

1. **ROI based on revenues/GDP generated, divided by HPC investment**
2. **ROI based on profits generated, divided by HPC investment**
3. **ROI based on jobs created (and the HPC investment required per job created)**

The ROI models were tested for variances by:

- Industry sector
- Country
- Organization size

The Innovation Models That Were Developed

The Innovation models are of two main types:

1. **Basic Research / Major Innovations**
2. **Applied Research / Incremental Innovations**

These are captured as:

- Innovations in government & academia
- Innovations in industry

The Innovation models can be sorted for variances by:

- Industry sector
- Country
- Organization size
- Government, Industry and Academia

The Innovation Index Scale

10 = One of the top 2 to 3 innovations in the last decade

9 = One of the top 5 innovations in the last decade

8 = One of the top 10 innovations in the last decade

7 = One of the top 25 innovations in the last decade

6 = One of the top 50 innovations in the last decade

5 = It had a major impact and is useful to many organizations

4 = A minor innovation that is useful to many organizations

3 = A minor innovation or only useful to 2 -3 organizations

2 = A minor innovation or only useful to 1 organization

1 = An innovation that is recognized ONLY by experts in the field

Key Findings

- **IDC is able to collect the required data across a broad set of organizations, with enough detail to create the economic models and the innovation index**
- **Early results indicate very substantial returns related to investments in HPC, on average:**
  - $356.5 in revenue per dollar of HPC invested
  - $38.7 of profits/cost savings per dollar of HPC invested
  - The average HPC investment per innovation was $3.1M
- **Note that an additional outcome of this research is an expansive list of HPC success stories**

Sample Characteristics

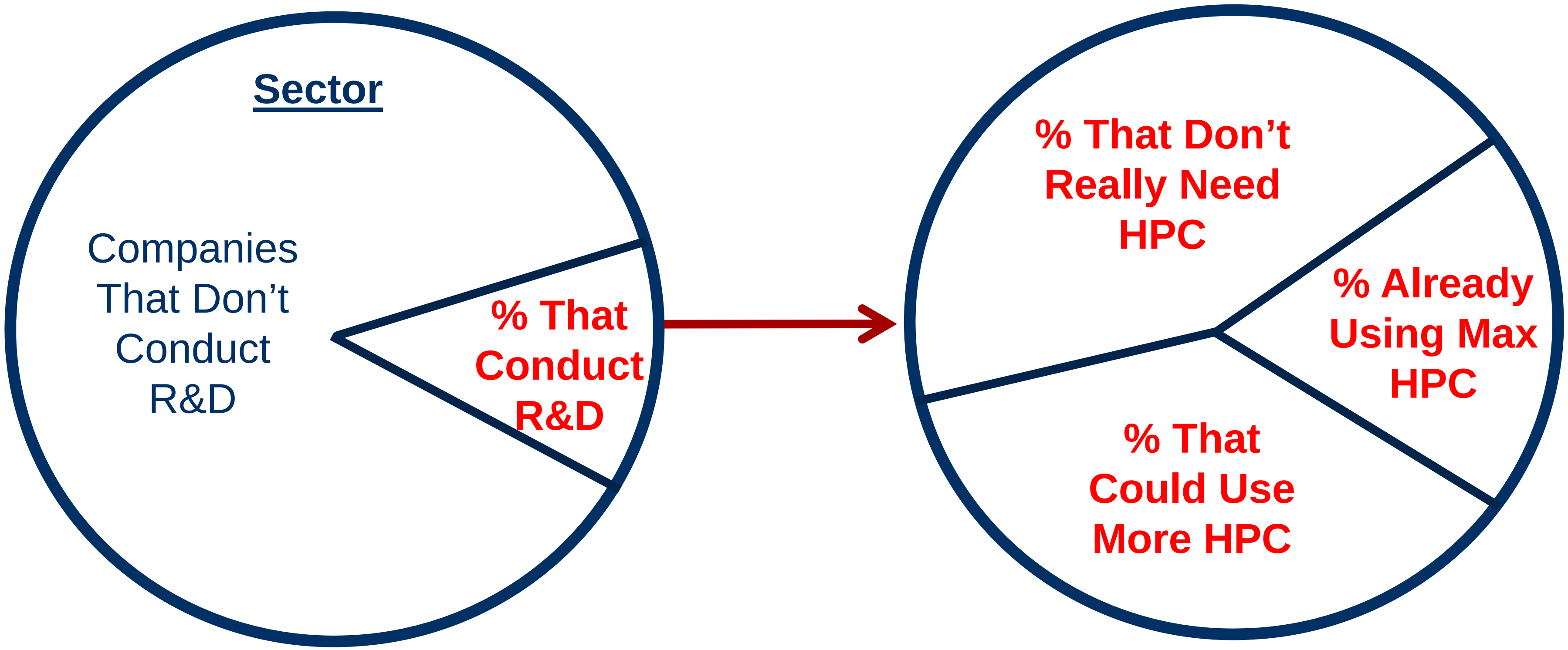
Sample demographics:

- **A total of <u>208 case study examples</u> of ROI and innovations were collected as part of the study:**
  - 67 financial ROI examples
  - 141 innovation examples

- In addition, a large number of micro-surveys were conduct to learn key ratios in order to eventually apply the results to large economic data sets.
  - Over 30,000 scientists and engineers were contacted, with over 1,500 completing the micro-survey.

Research Overview – Parameters Being Collected to Tie to Broader Economic Reports

For each sector we need 4 basic ratios



Note: IDC has conducted over 30,000 light phone calls for this data.
We will likely require 5x to 10x more surveys

Pilot Study Results: Financial ROI

Key Findings:
Primary Financial ROI Results

1. **IDC is able to collect the required data across a broad set of organizations with enough detail to create the two economic models and the innovation index**

2. **Early results indicate very substantial returns for investments in HPC:**
   - **$356 dollars on average in revenue** per dollar of HPC invested.
   - **$38 dollars on average of profits** (or cost savings) per dollar of HPC invested.

Key Findings:
The Financial ROI Model – By Sector

3	Organization Size: People	(All)
4	Organization Size in $M	(All)
5	Organization Size (S,M,L)	(All)
6	Industry	(All)
7	Innovation Level	(All)
8	Country	(All)
9	Years Before 1st Return	(All)
10	Applied	(All)
11	Basic	(All)
12	Accomplishment Type	Financial R
13	Total R&D	(All)
14	Employee Growth	(All)
15	Organization Type	(All)
16	Total HPC Investment	(All)

19	Sector	Count	Sum of Employee Growth	Years Before 1st Return	Average of Revenue $ per HPC $	Average of Profit $ per HPC $
20	Academic	12	2	1.8	37.4	70.8
21	Government	4	10	1.4	9.2	3.9
22	Industry	51	1,157	1.9	462.4	36.4
23	**Grand Total**	**67**	**1,169**	**1.9**	**356.5**	**38.7**

Key Findings:
The Financial ROI Model – By Country

IDC
Analyze the Future

3	Organization Size: People	(All)
4	Organization Size in $M	(All)
5	Organization Size (S,M,L)	(All)
6	Industry	(All)
7	Innovation Level	(All)
8	Years Before 1st Return	(All)
9	Applied	(All)
10	Basic	(All)
11	Sector	(All)
12	Accomplishment Type	Financial ROI
13	Total R&D	(All)
14	Employee Growth	(All)
15	Organization Type	(All)
16	Total HPC Investment	(All)

19	Country	Count	Sum of Employee Growth	Average Years Before 1st Return	Average of Revenue $ per HPC $	Average of Profit $ per HPC $
20	China	3	30	1.3	8.7	5.4
21	France	4		5.1	621.7	125.0
22	UK	31	896	1.6	366.5	26.7
23	US	27	243	1.8	373.3	49.8
24	Italy	2		1.0	10.0	7.5
25	**Grand Total**	**67**	**1,169**	**1.9**	**356.5**	**38.7**

Pilot Study Results: Innovation

Key Findings: Investments Per Innovation

The average HPC investment per innovation was $3.1 million.

- Overall $497 million in HPC investments were made to generate the 141 innovations in the pilot study.
- With many at under $1 million per innovation.

Key Findings: The Innovation Areas For The 141 Innovation Data Examples

Industry	(All)
Years Before 1st Return	(All)
Applied	(All)
Basic	(All)
Accomplishment Type	Innovation
Country	(All)
Total R&D	(All)
Organization Type	(All)

Primary Innovation / ROI Area	Count	Sum of Total HPC Investment	Average Years Before 1st Return	Average of HPC $M per Innovation
Better Products	54	$114 M	1.9	$4.2 M
Created New Approach	40	$15 M	1.2	$0.4 M
Discovered Something New	20	$46 M	1.8	$2.7 M
Helped Society	11	$66 M	1.0	$6.0 M
Cost Saving	6	$180 M	1.3	$2.1 M
Major Breakthrough	5	$3 M	3.2	$1.1 M
Helped Research Program	5	$71 M	1.5	$14.3 M
Grand Total	**141**	**$497 M**	**1.6**	**$3.1 M**

Key Findings:
The New Innovation Index Scores

The average innovation rating = 5.0

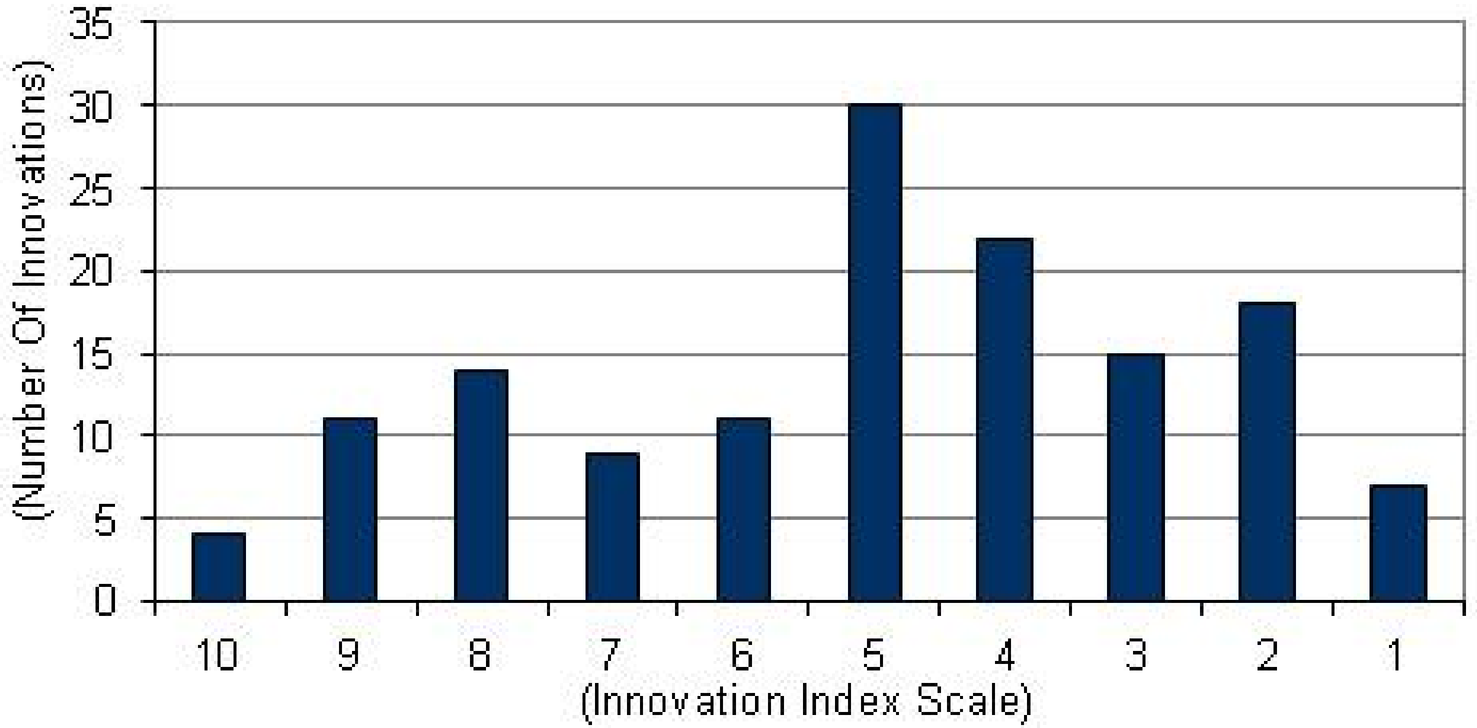
- 4.4 for the 67 basic research/major innovations
- 5.5 for the 74 applied research/incremental innovations

10 = One of the top 2 to 3 innovations in the last decade
9 = One of the top 5 innovations in the last decade
8 = One of the top 10 innovations in the last decade
7 = One of the top 25 innovations in the last decade
6 = One of the top 50 innovations in the last decade
5 = It had a major impact and is useful to many organizations
4 = A minor innovation that is useful to many organizations
3 = A minor innovation or only useful to 2 -3 organizations
2 = A minor innovation or only useful to 1 organization
1 = An innovation that is recognized ONLY by experts in the field

Key Findings: The New Innovation Index Scores – For All 141 Innovations

FIGURE 12

HPC Innovation Index Scale Results: All Respondents



N = 141

Source: IDC 2013

Key Findings:
Locations of the Innovations

IDC Analyze the Future

Academic and industrial sites represented the bulk of the innovations in the sample.

- Most of the basic research innovations were in academia, while most of the applied research innovations were in industry.

Government innovations were higher on the innovation index scale (averaging 7.0).

- Innovations in industry ranked lower at 5.7.
- Academic innovations averaged 3.9.

Key Findings: Success Stories

Note that an additional outcome of this research is an expansive list of HPC success stories.

- These can be used to help explain the importance of HPC to funding bodies, key decision makers and the broader public.
- IDC is writing up a number of them for broader dissemination.

Dissemination Program

- **The report and excel models will be posted at: www.hpcuserforum.com/ROI**
- **Once approved by DOE, IDC will send the report and models to ~4,500 people in the broader HPC community**
- **IDC will brief the community at SC13**
  - On Tuesday during the IDC breakfast briefing
  - On Wednesday, at 5:30 – 7:00pm, during the Cost-Benefit Quantification for HPC BOF
- **Press release**
- **Many briefings have already been given – does DOE have additional briefings that IDC should conduct?**

Future ROI Research Plans (Proposed)

Phase I (Year 1) – The goal is to create the actual ROI models with a full data set at least 2x in Phase I, growing to at least 8x in size by Phase III. This is needed to create predictive models a fuller understanding of the relationships, to provide enough data/analysis to start making predictive results, and to refine the models as needed.

- The goal is to have enough data to start making statistically sound correlations between industries, between countries and between different sizes of organizations.

Phase II (Year 2) – Expand the data set by at least 2x more, and including more countries and industries. Motivate a larger set of nations to contribute deeper data samples.

- The goal is to have enough data to make strong statistically sound correlations between industries, between countries and between different sizes of organizations – and cross-correlations like industries by country, and organization size by country.

Phase III (Year 3) – Expand the data set again by at least 2x more, and focus on hardening the predictive nature of the models by conducting additional research to "test" correlations. At this phase it the tie to large general economic data sets should be robust enough to be able to hand-off to other government organizations.

- The goal is to conduct enough research to show both strong statistical correlation – and causation between investments in HPC and the resulting ROI and innovation.
- In this phase the scenario testing should improve to be very robust and directly useful for making national policy decisions.

Questions?

Please email:
hpc@idc.com

Or check out:
www.hpcuserforum.com

Grant References

The authors thank DOE for its insights and guidance on and funding of this grant-based research project.

This study is based upon work funded by the U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research, and the National Nuclear Security Administration, under award number DE-SC0008540.

DOE Program Managers:

Christine Chalk, 301-903-5152, christine.chalk@science.doe.gov, and Barbara J. Helland, 301-903-3127, barbara.helland@science.doe.gov, U.S. Department of Energy Office of Science, Germantown Building, 1000 Independence Avenue, S.W. Washington, D.C., 20585-1290.
and Bob Meisner in National Nuclear security Administration (NNSA).

Administrator/Contracting Officer:

Warren Riley, 630-252-2485, warren.riley@ch.doe.gov, U.S. Department of Energy Office of Acquisition and Assistance, 9800 South Cass Avenue, Argonne, Illinois, 60439

IDC Reporting:

- Principal investigator: Earl C. Joseph, Ph.D., 612-812-5798, ejoseph@idc.com
- Senior technical project manager: John Daly, 508-935-4643, jdaly@idc.com

http://bit.ly/sc13-eval

Questions to the Audience

- What are your ideas to quantify the benefit of HPC?
  - Publications count?
  - Grant money?
- What is the trade-off between hardware and human resources?
  - Can we argue for more human resources?
- Would you like to participate in the IDC initiative?
- …

http://bit.ly/sc13-eval

Potential Future Activities

Alternatives

- Don´t worry – be happy
  - In academia it is "just" tax money
  - In industry the customer will pay for it

- Participate by joining this SIGCBQ
  - Have a follow-up BoF at SC´14 (if accepted)
  - Organize a workshop
  - Install a mailing list

As for Quantification…

Please fill in the evaluation form

http://bit.ly/sc13-eval