

# Quantifying Productivity – Towards Development Effort Estimation in HPC



## Introduction

In pursuit of exaflop computing, the expenses of HPC centers increase in terms of acquisition, energy, and programming. Thus, a quantifiable productivity metric is needed to make an informed decision on how to invest available budgets<sup>1</sup>.

$$productivity = \frac{value}{cost} = \frac{\#app. runs}{total cost of ownership [\$]}$$

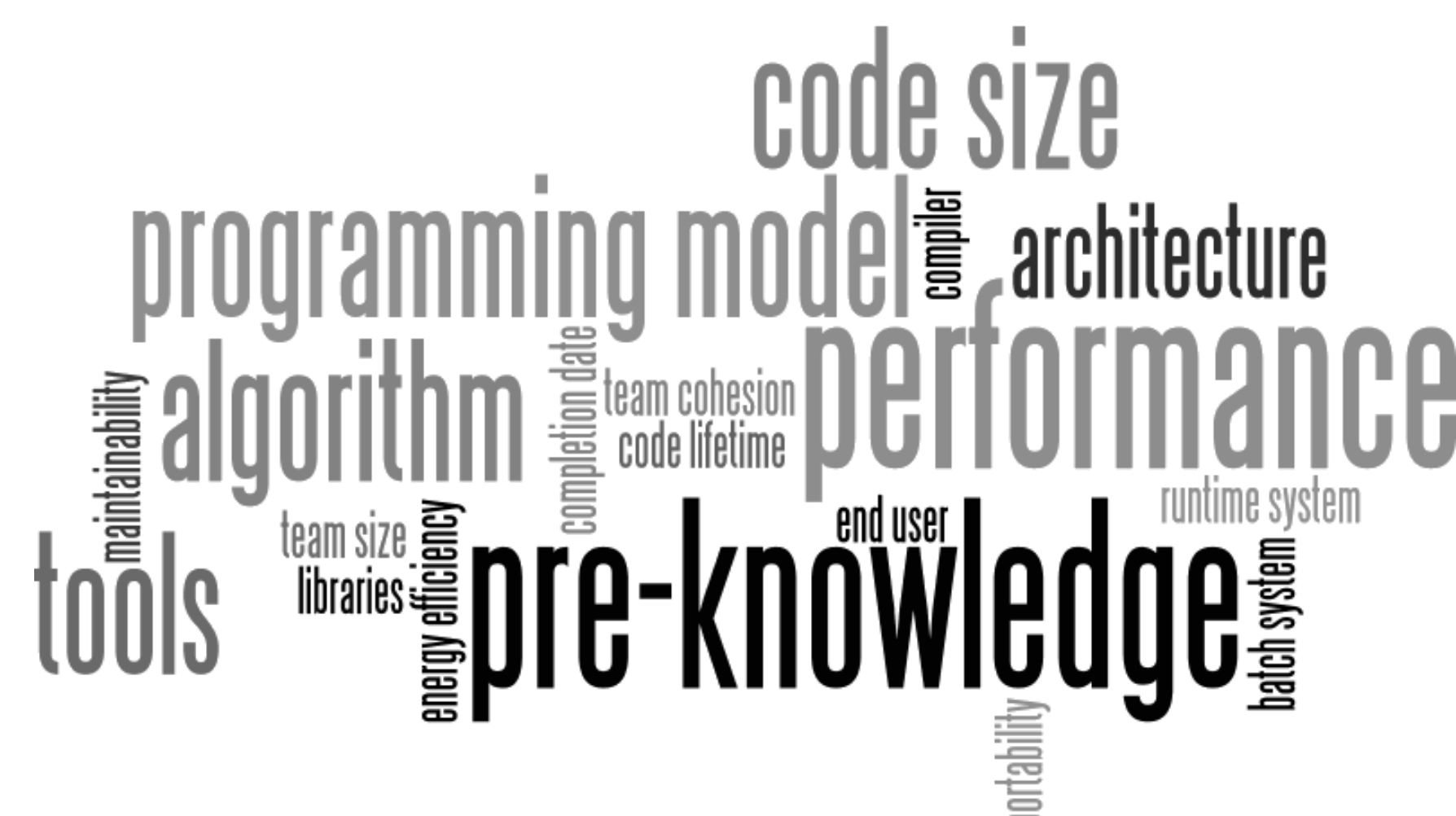
$$\$ = HW + energy + \underbrace{development cost}_{development\ effort [day] * salary [\$/day]} + \dots$$

How to estimate development effort in HPC?

## Development Effort Estimation

- Models exist in SW engineering
- Not directly applicable to (non-commercial) HPC setups due to particular target “performance”

**Aim:** create model for use cases in HPC  
**Challenge:** numerous impact factors on effort



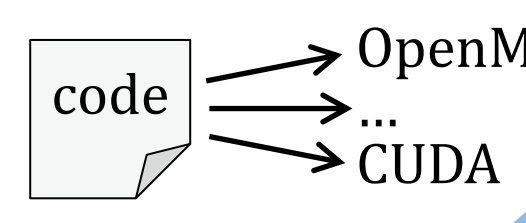
## Methodology

In our preliminary work, we try to qualify & quantify development effort & productivity parameters to estimate them by regression analysis.

### Qualifying...

Impact factors & their level of impact  
 ▶ Questionnaires: gathering & weighting parameters; correlation analysis

1. focus: Impact of programming model  
 ▶ Statistical significance in performance & effort



2. focus: Impact of algorithm  
 ▶ Analysis of effort-intensive code patterns in proxy apps<sup>2</sup>

### Quantifying by...

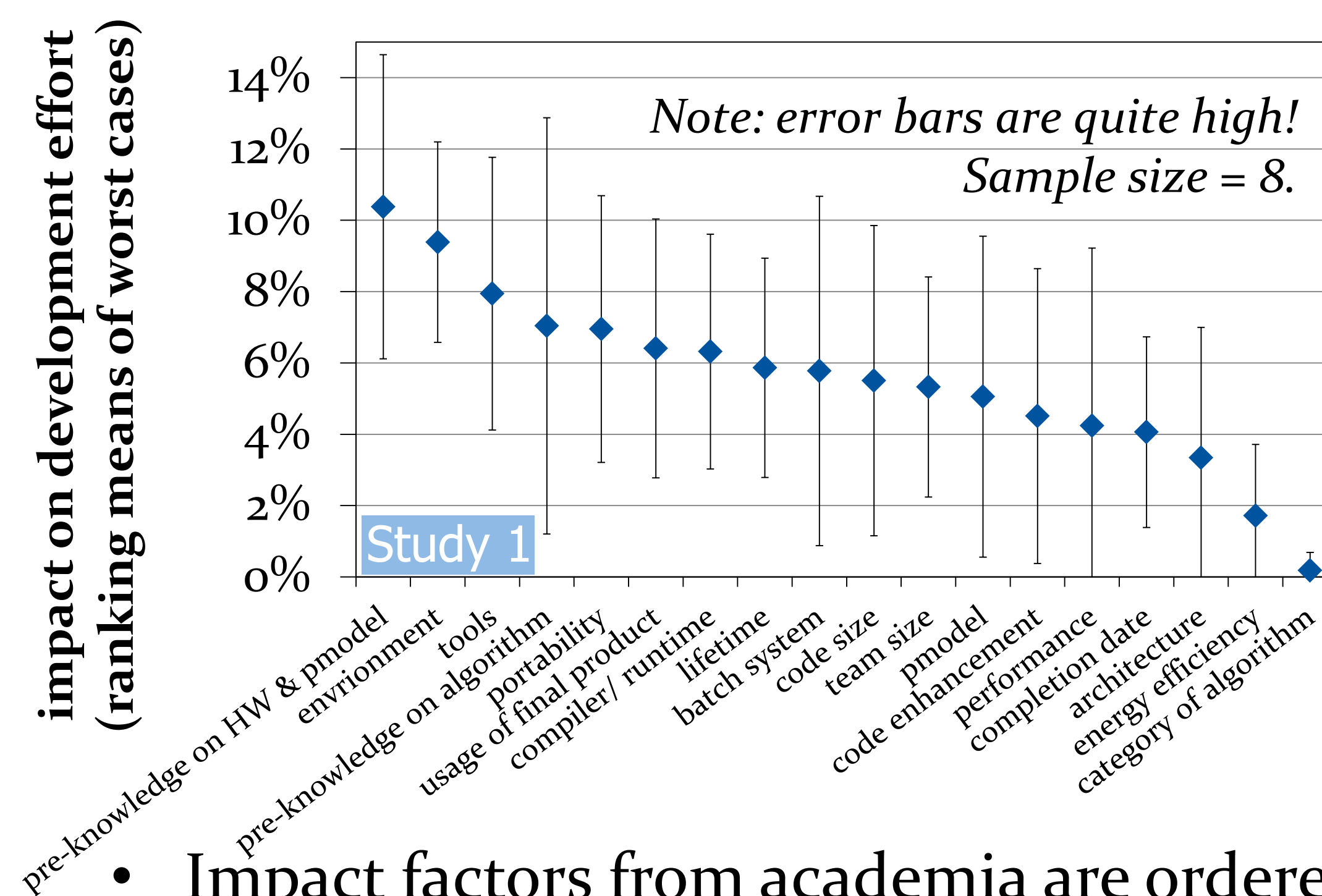
Development diaries, development effort tracking tool, error propagation

Regression analysis, guide on how to add new programming models

Instrument code patterns for development effort tracking

## First Results

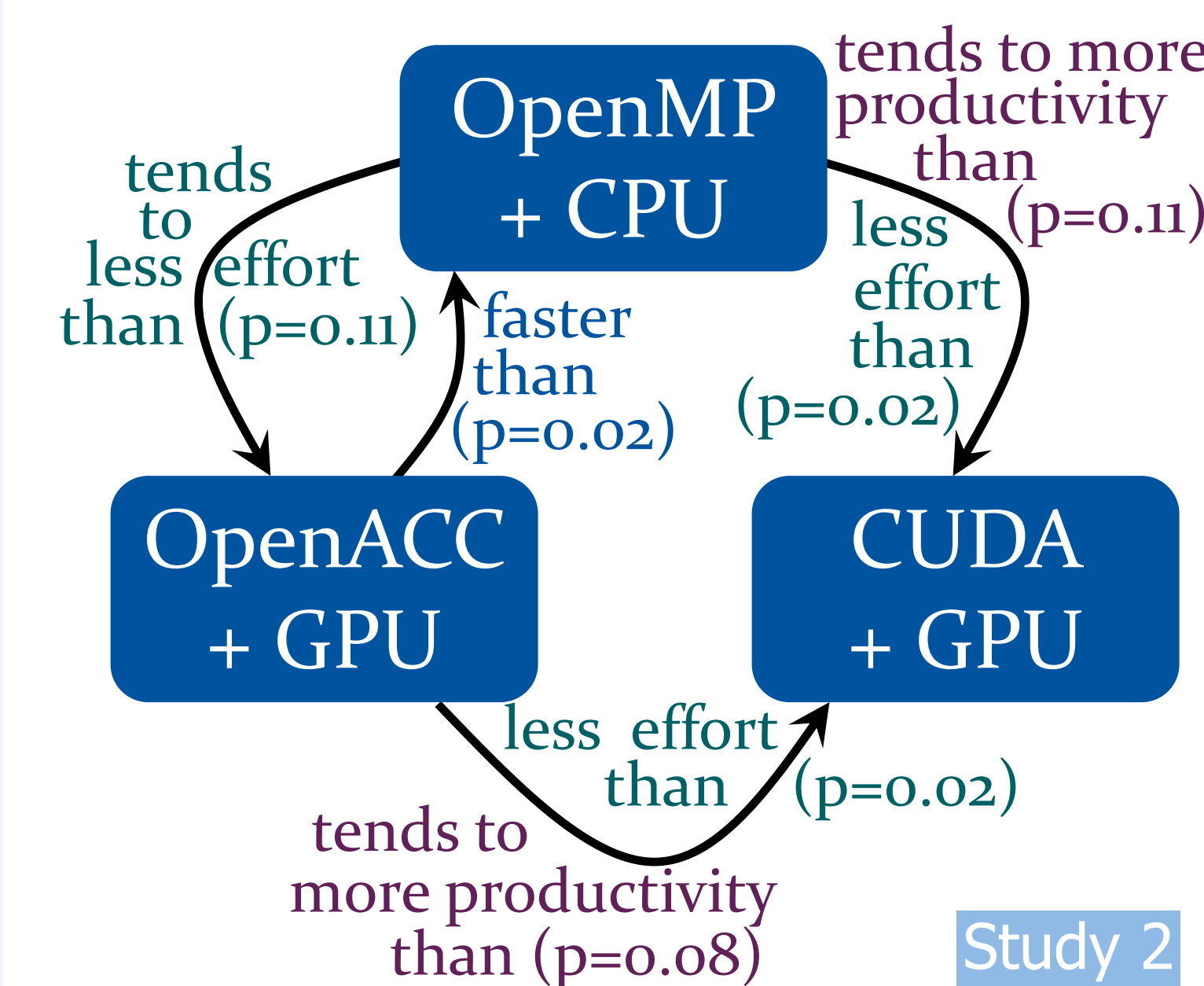
### Questionnaire on impact level of parameters<sup>3</sup>



- Impact factors from academia are ordered differently to company ones
- Some questions must be re-formulated (e.g., category of algorithm)
- First 3 parameters show tendency of importance: *pre-knowledge* as expected, *environment* & *tools* surprisingly high

### Statistical comparison of programming models on HW<sup>4</sup>

- One-sided signed Wilcoxon rank sum test
- 5% and 15% significance



- Programming models play a role in productivity analysis

## Conclusions

Our work-in-progress presents a methodology to quantify development effort & productivity in HPC. Various empirical studies must be conducted to gain a database. Statistical methods & regression analysis help to analyze this data and to estimate development effort.

First results show some applicability of our methodology. However, future analysis must rely on bigger samples, e.g. gathered through software labs at RWTH & LLNL, Hackathons and questionnaires.

## Get Involved

- Your definition of productivity & development effort?
- Your ranking of impact factors?
- Your parallelization effort?
- Your usage of SW cost estimation models?

Let us know!

Take part at our questionnaire<sup>5</sup>!



## References

<sup>1</sup>S. Wienke, H. Iliev, D. an Mey, M.S. Müller: Modeling the Productivity of HPC Systems on a Computing Center Scale. In: Kunkel, J.M., Ludwig, T. (eds.) High Performance Computing, LNCS 9137, pp. 358–375. Springer International Publishing (2015).  
<sup>2</sup>Lawrence Livermore National Laboratory: LLNL ASC Proxy Apps, <https://codesign.llnl.gov/proxy-apps.php> (viewed July 2015).  
<sup>3</sup>Lawrence Livermore National Laboratory: Hackathon in Computation (July 2015).  
<sup>4</sup>RWTH Aachen University: Software Lab of Parallel Programming Models for Applications of High-Performance Computing (2014).  
<sup>5</sup>RWTH Aachen University: Productivity Questionnaire in HPC, <http://www.hpc.rwth-aachen.de/research/tco/productivity/survey> (2015).