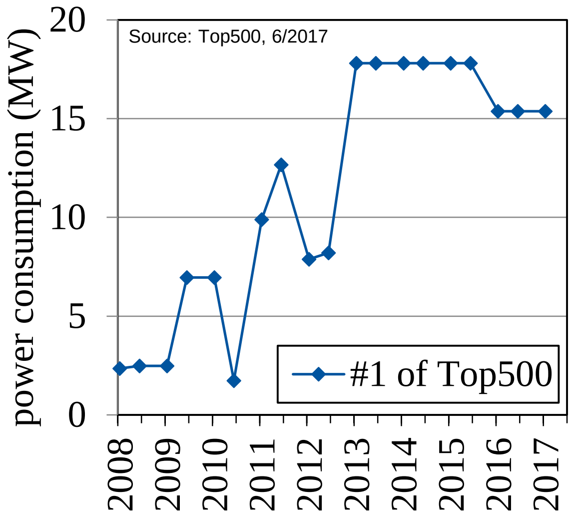
Productivity and Software Development Effort Estimation in High-Performance Computing

Sandra Wienke
IT Center & Chair for High Performance Computing, RWTH Aachen University

SC Doctoral Showcase; Denver, CO, USA; November 2017

Motivation

- Ever increasing demands for computational power
  - Increasing expenses for, e.g., HW acquisition, electrical power, software development



- HPC procurements
  - Quantifiable metric for informed decisions on how to invest available budgets needed

- Comprehensive metric
  - Focus here: German university HPC centers (in production)
  - Advisory council on scientific matters recommends (amongst others) the integration of energy and personnel costs into funding lines for the German National High-Performance Computing (NHR)
  - My approach: productivity model with predictive power

Source: Gauß Allianz, https://gauss-allianz.de

Agenda

- Productivity
  - Single-application Perspective
  - Multi-application Perspective (Job Mix)
  - Sensitivity Analysis
- Total Cost of Ownership
- Software Development Effort
  - Performance Life-Cycle
  - Identification of Impact Factors
  - Quantification of Factor "Pre-knowledge"
- Case Study: Aeroacoustics Simulations
- Conclusion

$$\text{productivity} = \frac{\text{value}}{\text{cost } [€]}$$

$$€ = \text{HW} + \text{energy} + \text{development costs} + \ldots$$

$$\text{development effort [days]} * \text{salary } \left[\frac{€}{\text{days}}\right]$$

Agenda

- Productivity
  - Single-application Perspective
  - Multi-application Perspective (Job Mix)
  - Sensitivity Analysis
- Total Cost of Ownership
- Software Development Effort
  - Performance Life-Cycle
  - Identification of Impact Factors
  - Quantification of Factor “Pre-knowledge
- Case Study: Aeroacoustics Simulations
- Conclusion

$$\text{productivity} = \frac{\text{value}}{\text{cost [€]}}$$

€ = HW + energy + development costs + ...

development effort [days] * salary [$\frac{€}{\text{days}}$]

IT Center

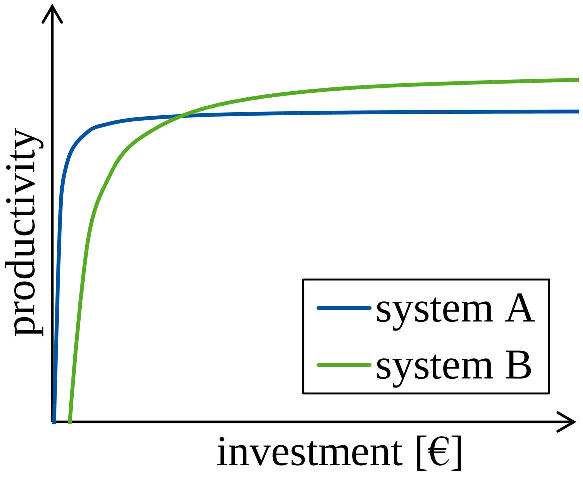
RWTH AACHEN UNIVERSITY

Productivity

[1] Wienke, S., et al. (2013). Accelerators for Technical Computing: Is It Worth the Pain? A TCO Perspective.
[2] Wienke, S., et al. (2015). Modeling the Productivity of HPC Systems on a Computing Center Scale.
[3] Schneider, F.P., et al. (2017). Operational Concepts of GPU Systems in HPC Centers: TCO and Productivity. Accepted.

- Economics

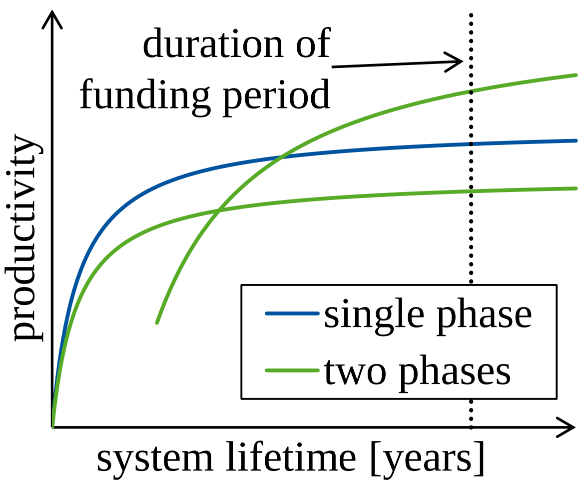
$$\text{productivity } \Psi = \frac{\text{output}}{\text{input}} = \frac{\text{value}}{\text{cost [€]}}$$



- My adaptions for HPC purposes

$$\Psi(n,\tau) = \frac{\text{scientific outcome}}{\text{total costs [€]}} = \frac{\sum_i r_{app,i}(n,\tau)}{\text{TCO}(n,\tau)}$$

n	no. of compute nodes	$r_{app,i}$	no. of application runs
τ	system lifetime	TCO	total cost of ownership



[1] Wienke, S., et al. (2013). Accelerators for Technical Computing: Is It Worth the Pain? A TCO Perspective.
[2] Wienke, S., et al. (2015). Modeling the Productivity of HPC Systems on a Computing Center Scale.
[3] Schneider, F.P., et al. (2017). Operational Concepts of GPU Systems in HPC Centers: TCO and Productivity. Accepted.

Productivity: Single-application Perspective

- **Value**: Number of simulation code runs of applications i

$$r_{app,i}(n,\tau) \sim \frac{\alpha \cdot \tau}{t_{app,i}(n)}$$

  – In addition: quality weighting factor for high-scaling app.

n	no. of compute nodes
τ	system lifetime
α	system availability
$t_{app,i}$	application runtime
C^{ot}	one-time costs
C^{pa}	annual costs

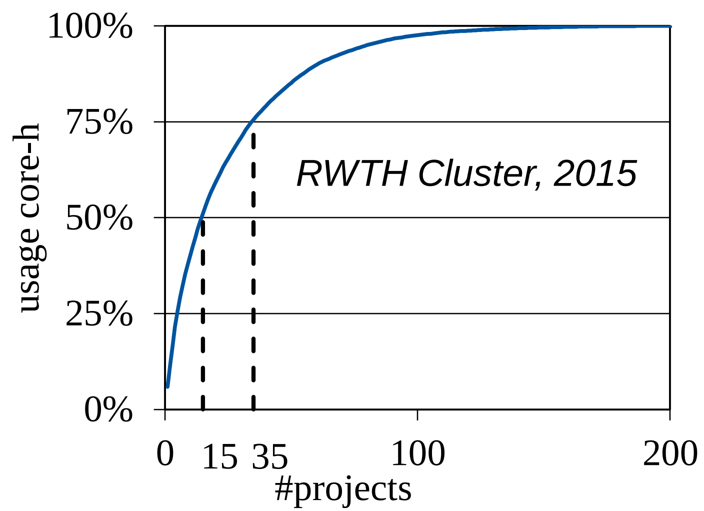
- **Cost**: Total Cost of Ownership (TCO)
  – Per node and per node type

$$\mathrm{TCO}(n,\tau) = \underbrace{C^{ot}(n)} + \underbrace{C^{pa}(n)} \cdot \tau$$

… + HW acquisition + (initial) dev. effort + HW maintenance + energy + SW licenses + …

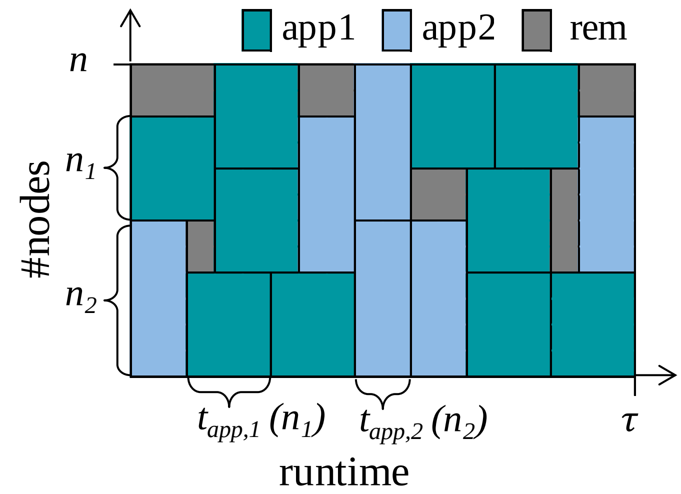
Productivity: Multi-application Perspective (Job Mix)

- Number of applications
  - Reduction to m „relevant“ applications
  - E.g. tendering process, cluster statistics



- **Value**: Sum of all application runs
  - Comprehensive metric (all kinds of applications)
  - Capacity-based weighting factor p_i

$$\sum_{i=1}^{m} r_{app,i}(n,\tau) \sim \sum_{i=1}^{m} \left(\frac{\alpha \cdot \tau}{t_{app,i}(n_i)} \cdot \frac{n}{n_i} \cdot p_i \right)$$



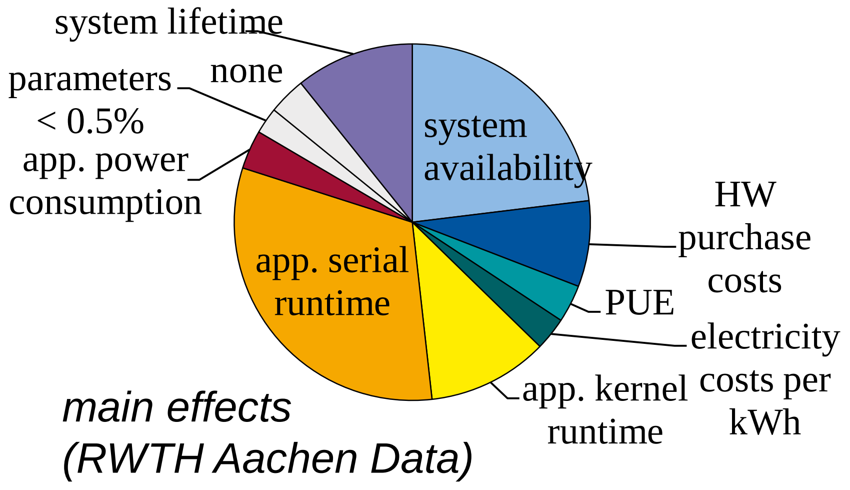
- **Cost**: Application-dependent factors
  - Development effort: sum over all apps i
  - Power consumption: consider capacity factor p_i

Sensitivity of Productivity Model

- Assumptions in model parameters → variances in productivity model

- Sensitivity analysis
  - How uncertainty in productivity model can be apportioned to different model parameters
  - Here: simulation- und variance-based global sensitivity analysis (Saltelli et al.)

- Productivity model
  - Only few (well-understood) parameters must be accurately predicted
  - → Model is robust (within the given conditions)



Agenda

- Productivity
  - Single-application Perspective
  - Multi-application Perspective (Job Mix)
  - Sensitivity Analysis
- Total Cost of Ownership
- **Software Development Effort**
  - Performance Life-Cycle
  - Identification of Impact Factors
  - Quantification of Factor "Pre-knowledge
- Case Study: Aeroacoustics Simulations
- Conclusion

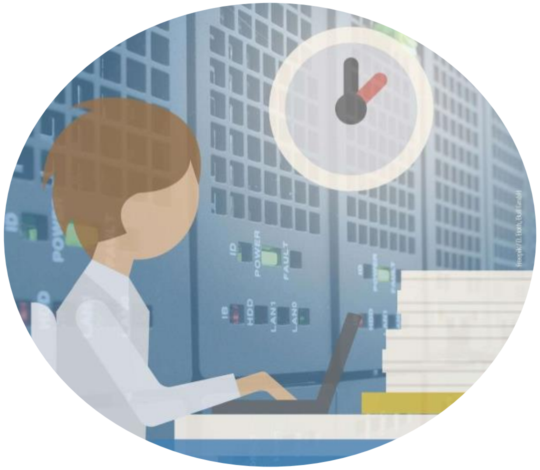
$$\text{productivity} = \frac{\text{value}}{\text{cost [€]}}$$

$$\text{€} = \text{HW} + \text{energy} + \underline{\text{development costs}} + \ldots$$

$$\text{development effort [days]} * \text{salary } [\frac{\text{€}}{\text{days}}]$$

Estimation of Software Development Effort

[4] Nicolini, M., et al. (2016). Software Cost Analysis of GPU-Accelerated Aeroacoustics Simulations in C++ with OpenACC.
[5] Miller, J., et al. (2017). Applicability of the Software Cost Model COCOMO II to HPC Projects. Accepted.



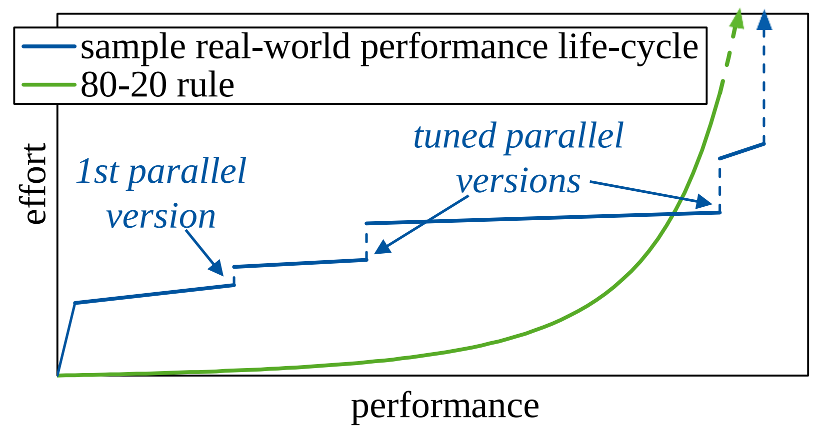
- Increasing demands for computational power →
  - Increasing HW and SW complexity in HPC
  - Increasing application development effort
- Integration of effort into productivity model becomes more important
  - Quantifiable metric for effort estimation needed
- Used definition of development effort
  - Needed effort for HPC-typical activities such as parallelization, optimization, port of simulation codes, performance analysis, debugging
- Software engineering (SE): software cost models
  - Example: COCOMO II (focus on code size)
  - But: direct applicability of COCOMO II to (investigated) HPC projects not feasible (focus on performance)

Performance Life-Cycle

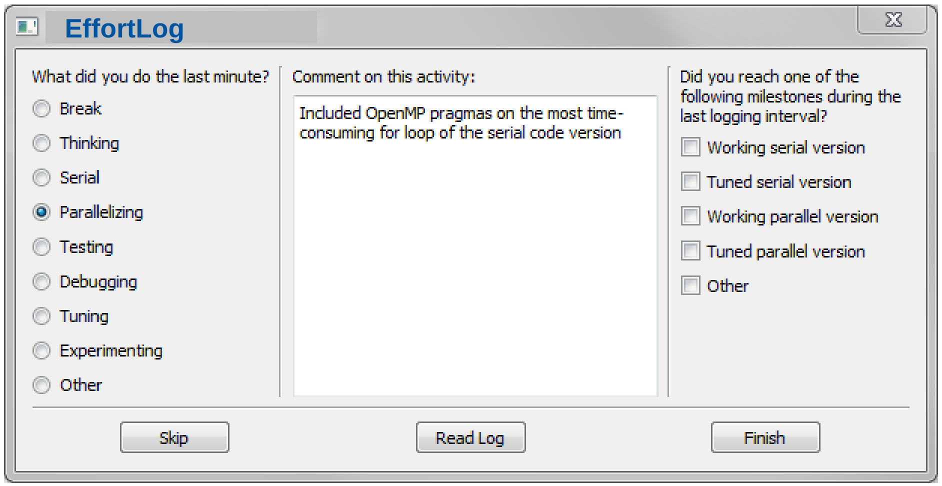
[6] Wienke, S., et al. (2016). Development Effort Estimation in HPC.

- Model of relationship of performance and corresponding development effort

$$\text{effort} = S \cdot f(\text{performance})^{R} + T$$



- Method: regression analysis from collected data
- Data collection: human-subject studies
  - Electronic development diary



- Numerous impact factors on software development effort (captured in S, R, T)

Identification of Impact Factors on Effort

[6] Wienke, S., et al. (2016). Development Effort Estimation in HPC.
[7] Wienke, S., et al. (2015). Quantifying Productivity – Towards Development Effort Estimation in HPC. Poster.

more impact

- Pre-knowledge on HW & parallel prog. model
- Code work
- Pre-knowledge on numerical algorithm used
- Parallel prog. model & compiler/ runtime system
- Performance
- Architecture/ hardware
- Tools
- Kind of algorithm
- Code size
- Portability & maintainability over code's lifetime
- Energy efficiency

less impact

Impact factors on effort
(derived by rank sums of 44 surveys)

- Key drivers
  - Aim: focus on most influencing factors first
- Method
  - Ranking of factors based on surveys
  - Combine/ eliminate factors with statistical analysis
- Realization
  - Start with set of 11 factors & ask for missing factors
  - Here: professional and student developers in HPC
  - Statistical significance testing:
    one-sided Wilcoxon rank test with Holm correction

Quantification of Impact Factor “Pre-knowledge”

[6] Wienke, S., et al. (2016). Development Effort Estimation in HPC.

- Method: knowledge surveys (KS)
  - Knowledge questions are not really answered, but participants rate their confidence in answering the questions

A	I am **confident** that I can adequately answer the question for graded test purposes at this time.
B	I can now answer **at least 50%** of the question or know precisely where I can quickly get the information needed and return here in 20 min or less to provide a complete answer for graded test purposes.
C	I am **not confident** to answer the question sufficiently for graded test purposes at this time.

- Realization
  - 40 questions: parallel computer architecture/ programming models, algorithm
  - Distribution of questions by Bloom’s taxonomy
  - Usage of Ø (pre-)KS results in performance life-cycle

Agenda

- Productivity
  - Single-application Perspective
  - Multi-application Perspective (Job Mix)
  - Sensitivity Analysis
- Total Cost of Ownership
- Software Development Effort
  - Performance Life-Cycle
  - Identification of Impact Factors
  - Quantification of Factor “Pre-knowledge
- Case Study: Aeroacoustics Simulations
- **Conclusion**

$$\text{productivity} = \frac{\text{value}}{\text{cost [€]}}$$

$$\text{€} = \text{HW} + \text{energy} + \underline{\text{development costs} + \ldots}$$

$$\text{development effort [days]} * \text{salary} \left[\frac{\text{€}}{\text{days}}\right]$$

Conclusion

- Methodologies to support informed HPC procurements

- Productivity metric for HPC environments
  - Applicability in real-world multi-job setups
  - Robust to errors in assumptions for parameters
  - Used in procurement process of RWTH Cluster CLAIX (2016)
- Software development effort estimation in HPC
  - Focus on performance: performance life-cycle
  - Identification and quantification of key drivers
  - Support of data collection through tools/ material

Outlook

- Continuing data collection with human-subject research
  - Aim: HPC community approach
- Conditional refinement of productivity model

References

1) S. Wienke, D. an Mey, and M. S. Müller, "Accelerators for Technical Computing: Is It Worth the Pain? A TCO Perspective," in *Supercomputing* vol. 7905, ed: Springer Berlin Heidelberg, 2013, pp. 330-342.
2) S. Wienke, H. Iliev, D. an Mey, and M. S. Müller, "Modeling the Productivity of HPC Systems on a Computing Center Scale," in *High Performance Computing* vol. 9137, ed: Springer International Publishing, 2015, pp. 358-375.
3) F. P. Schneider, S. Wienke, and M. S. Müller, "Operational Concepts of GPU Systems in HPC Centers: TCO and Productivity," in *15th International Workshop on Algorithms, Models and Tools for Parallel Computing on Heterogeneous Platform (HeteroPar 2017)*, 2017. Accepted.
4) M. Nicolini, J. Miller, S. Wienke, M. Schlottke-Lakemper, M. Meinke, and M. S. Müller, "Software Cost Analysis of GPU-Accelerated Aeroacoustics Simulations in C++ with OpenACC," in *High Performance Computing: ISC High Performance 2016 International Workshops, ExaComm, E-MuCoCoS, HPC-IODC, IXPUG, IWOPH, P^3MA, VHPC, WOPSSS, Frankfurt, Germany, June 19–23, 2016, Revised Selected Papers*, M. Taufer, B. Mohr, and M. J. Kunkel, Eds., ed Cham: Springer International Publishing, 2016, pp. 524-543.
5) J. Miller, S. Wienke, M. Schlottke-Lakemper, M. Meinke, and M. S. Müller, "Applicability of the Software Cost Model COCOMO II to HPC Projects," in *International Journal of Computational Science and Engineering*, 2017. Accepted.
6) S. Wienke, J. Miller, M. Schulz, and M. S. Müller, "Development Effort Estimation in HPC," in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC16)*, IEEE Press, 2016, pp. 107-118.
7) S. Wienke, T. Cramer, M. S. Müller, and M. Schulz, "Quantifying Productivity-Towards Development Effort Estimation in HPC," Poster at *the International Conference for High Performance Computing, Networking, Storage and Analysis (SC15)*, 2015.