



# Methodology for Development Effort & Productivity Estimation in HPC – Getting HPC Developers Involved

Julian Miller, Sandra Wienke

IT Center, RWTH Aachen University, Germany

GPU Hackathon, Jülich, 3/6/2017

# Whoami?

---

- M.Sc. in Computational Engineering Science
- HPC Group, IT Center, RWTH Aachen University
  - Teaching assistant (HPC software labs & lectures)
  - Tutorials, workshops
  - GPU support for the RWTH Compute Cluster
- Research focus
  - Analysis of Total Cost of Ownership (TCO) & Productivity of HPC centers
  
- Why I came as mentor to the Hackathon?
  - Help researcher to accelerate their code
  - Own research on real development effort
    - Knowledge surveys
    - Productivity (electronic developer diaries)

# Productivity of HPC Centers

- In pursuit of exaflop computing: expenses of HPC centers increase
  - Acquisition, energy, staff/ labor costs, programming
- Informed decision on how to invest available budgets
  - Quantifiable metric for benefit cost ratio needed<sup>3</sup>

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

$$\text{\$} = HW + energy + \underline{development\ cost} + \dots$$



$$development\ effort\ [day] * salary\ [\$/day]$$

- **How to estimate development efforts in HPC?**

<sup>3</sup>S. Wienke, H. Iliev, D. an Mey, M.S. Müller: Modeling the Productivity of HPC Systems on a Computing Center Scale. LNCS, vol. 9137, pp. 358–375. Springer International Publishing (2015)

# Development Effort Estimation in HPC

---

- **Aim:** model for development effort in HPC
  - Models exist in SE, e.g., COCOMO:  $effort = A \cdot (KLOC)^B$
  - Results of the projects evaluated so far: not directly applicable to (non-industry) HPC setups due to particular target “performance”
- **Challenges:**
  - Which metric best captures the effort? Code size, performance, etc.
  - Numerous impact factors on effort



- **Needed:** data base to rely statistical methods on!

# Getting HPC Developers Involved

---

- Tools
  - **Knowledge Survey**<sup>5</sup> (<https://www.soscisurvey.de/ksHack17/?q=CUDA>,  
<https://www.soscisurvey.de/ksHack17/?q=OpenACC>)
    - Do not really answer questions, instead rate level of confidence to answer question
    - Impact of PRE-knowledge on development productivity (until Monday, 3/6/2017)
    - Questions on application, parallel programming, GPUs (~30 min)
  - **Effort Log Tool** (<https://github.com/julianmi/effort-log>)
    - Electronic developer diary (Linux, Windows, Mac)
    - Interval-based: What did you do? Tuning knobs?
    - Milestones for performance-effort relationship!
- **Mapping** of surveys, effort logs, performance makes our research more valuable
  - Please use the same (arbitrary) nick name (data will be anonymously treated)

***Thank you for your participation!***

## Selected Publications

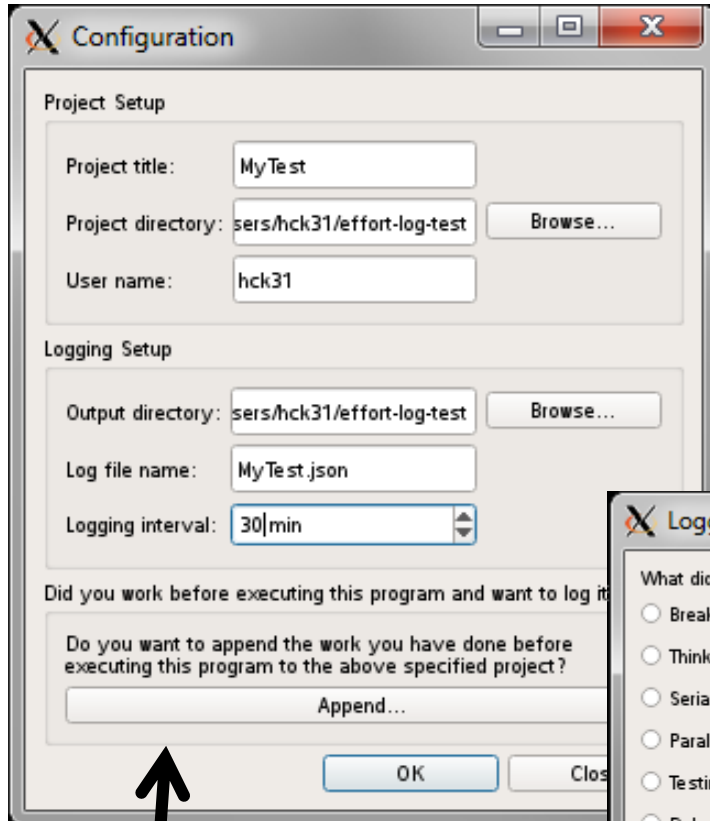
---

- 1) S. Wienke, J. Miller, M. Schulz, M.S. Müller: **Development Effort Estimation in HPC**. Paper accepted at International ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis (SC16), November 2016, Salt Lake City, UT, USA.
- 2) S. Wienke, T. Cramer, M.S. Müller, M. Schulz: **Quantifying Productivity - Towards Development Effort Estimation in HPC**. Scientific poster at the International ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis (SC15), November 2015, Austin, TX, USA.
- 3) 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, Lecture Notes in Computer Science, vol. 9137, pp. 358–375, Springer International Publishing, 2015.
- 4) S. Wienke, D. an Mey, D., M.S. Müller: **Accelerators for Technical Computing: Is It Worth the Pain? A TCO Perspective**. In: Kunkel, J., Ludwig, T., Meuer, H. (eds.) Supercomputing, Lecture Notes in Computer Science, vol. 7905, pp. 330–342, Springer Berlin Heidelberg, 2013.

# Further Information on *EffortLog*

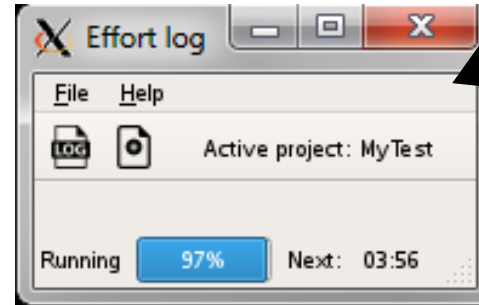
---

# EffortLog - Interface



## Setup frame

- Specify interval (e.g. 30min)
- Specify local file name

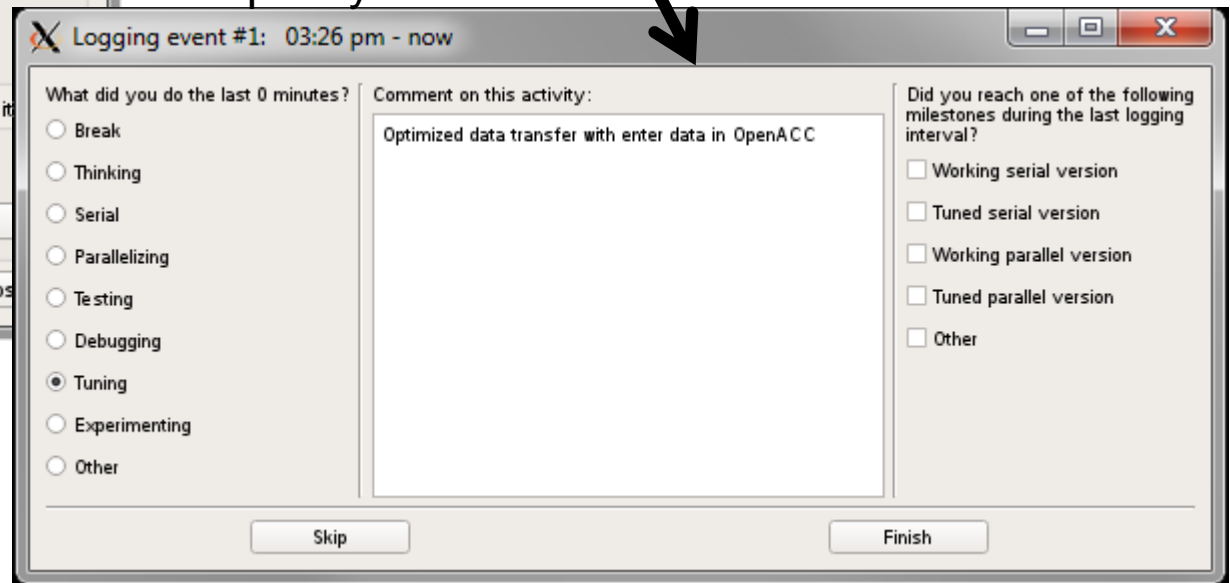


## Log Statistic

- Timer
- Manually log entries
- Tree view of previous entries
- JSON files

## Log Events

- Specify kind of activity & comment on it
- Specify milestones





## EffortLog - Milestones

- *Scratch*: No code base given - just the problem statement
- *Working serial version*: Serially-running version of your code that was tested for correctness, but not tuned for performance, e.g., 1st correct code implementation
- *Tuned serial version*: Serially-running version of your code that was tested for correctness and tuned for performance
- *Working parallel version*: In-parallel running version of your code that was tested for correctness, but not highly-tuned for performance, e.g. 1st correct parallel version of your code
- *Tuned parallel version*: In-parallel running version of your code that was tested for correctness and tuned for performance
- *Other*: Specify this code version in a later step.

Milestone: Tuned serial version

Please specify the following information regarding this milestone:

1. Performance  
Execution time in seconds  
20.00 sec
2. Architecture  
Intel Xeon Sandy Bridge + NVIDIA K20X
3. No threads, nodes, etc.  
Number of threads  
1 threads
4. Compiler  
Intel Compiler 16.0
5. Programming model  
OpenACC
6. Data size  
Same data (see last milestone)
7. Other comments  
General comment on this milestone

Skip Finish