Autotuning: Study of existing

Steven QUINITO MASNADA

March 4, 2016

Topic

Laboratory book

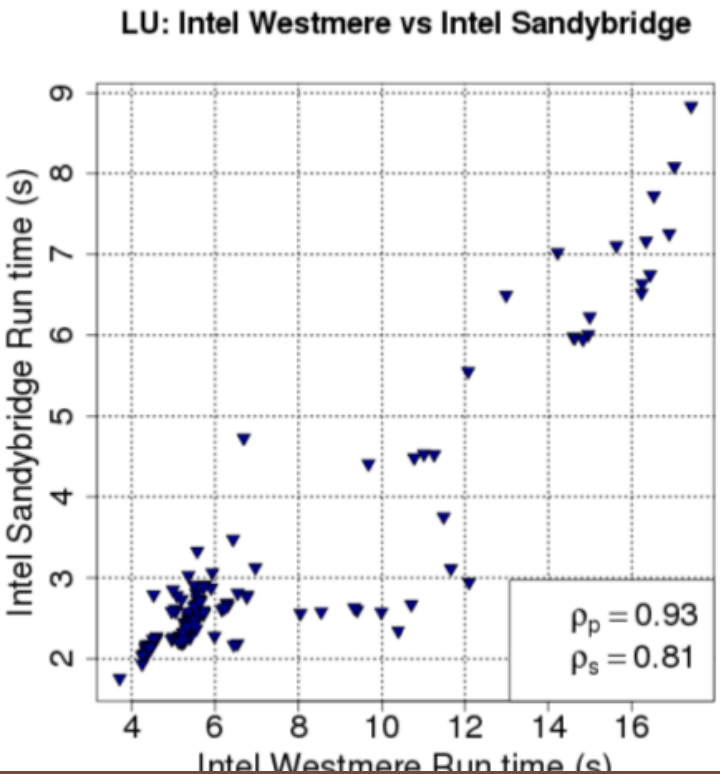
- Available at https://github.com/swhatelse/M2_internship
- Journal and notes

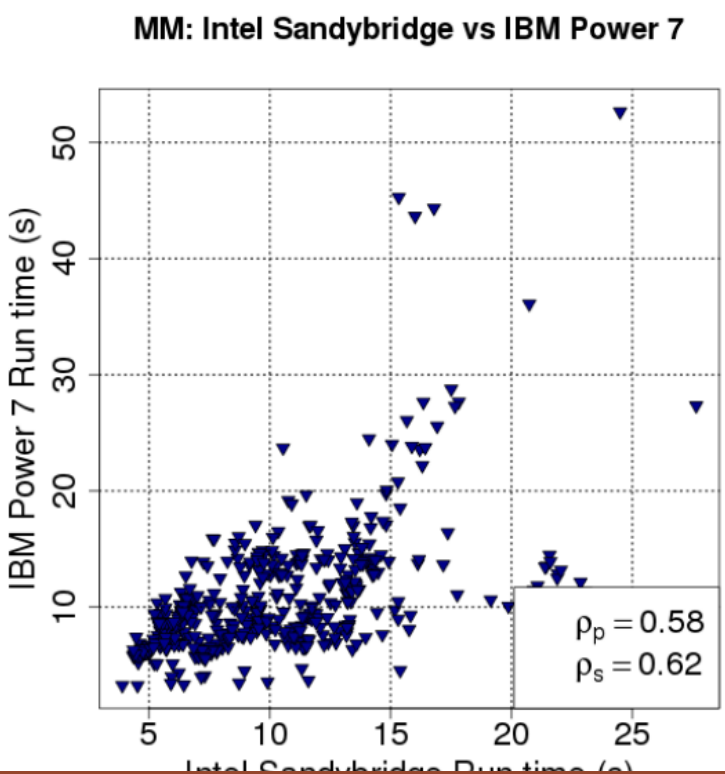
Topic

Jack Dongarra

- ATLAS $\rightarrow$ specialized autotuner for dense linear algebra
- Relaxed 1-D linear search $\rightarrow$ require to know good starting point and interaction between optimizations

- Discrete, derivative-free, and constraint optimization
- Correlation between peformance and tuning across platforms [Roy et al., 2015]
- Learning good combinations on one architecture and try to apply to another one
- Strong correlation with similar architectures





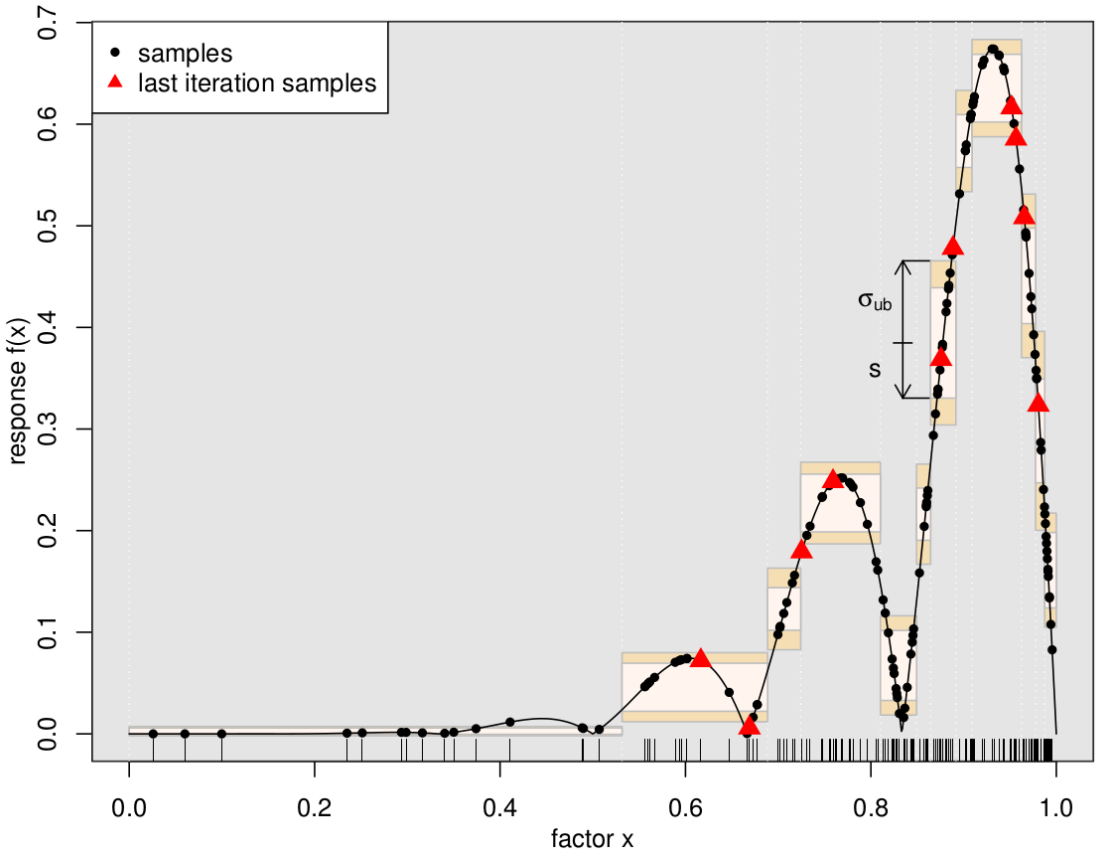
Grigory Fursin

- Discrete & Multi-objective optimization (e.g: size and performances)
- Possible to build model of compiler flags that works across applications → Milepost GCC [Fursin et al., 2011]
- Use of machine learning technics
- Models build using random search

- Training overhead reduce by Collective optimization → cTunning [Memon and Fursin, 2013]

ASK

- Exascale Computing research, UVSQ [De Oliveira Castro et al., 2012]
- Accuracy of a model relies on the sampling
- Partitioning search into regions with different level of variance
- Region with more variance are allocated more samples



OpenTuner

- MIT - CSAIL [Ansel et al., 2014]
- Discontinuous, non-smooth optimization
- Efficient of a search technic depends on the problem
- Adatping the search method to the particularities of the search space
- Testing multiple methods at the same time and keep those which performs better. Improvment are shared between methods.

BOAST

- specific generation/metaprogramming of kernels
- requires involvement of the developper and code restructuration
- allows optimizations that no compiler would dare
- At the moment, no strategy for tuning code

Topic

Autotuning optimization problem

- Nonsmooth and Empirical objective function and constraints
- Both discrete and continuous parameters
- Large optimization space with potential interactions between parameters

A priori on the objective function

- Exploit information about the problem (e.g., convexity, locality)
- Derivative based methods (local search generaly based on gradient descent)
  - Non convex (hence local minimum): randomized strategies (e.g., simulated annealing)
  - If derivation is not available (function too complex)
    - Direct search (e.g., Nelder-Mead, pattern search)
    - plus randomized strategies
  - If derivation is not possible (empirical function): estimate with regressions (e.g., surrogates-based search)
  - If evaluation is costly: meta-models (e.g., krigging) but derivation of interpolation is dubious...
- Additional difficulty depending on whether the parameters are constrained or not

"No a priori" (or other kind of a priori)

- Other kind of (discrete) structure (e.g. permutation, binary vector, tree, ...)
- Different notion of locality, hence need to cover a larger part of the search space
- Based on heuristics: Naive sampling, Genetic algorithm, tabu search, ant colonies, swarm

Many combinations of heuristics

Generalized Pattern Search [Abramson et al., 2004]:

- Global exploration $\rightarrow$ SEARCH
  - Possible to use any technics (e.g. randomized, surrogate-based, etc. . .)
  - Try find better solution than the local optimal elsewhere
- Local exploration $\rightarrow$ POLL
  - Try refine the solution by exploring a mesh
- Use derivative informations when available to speedup the exploration

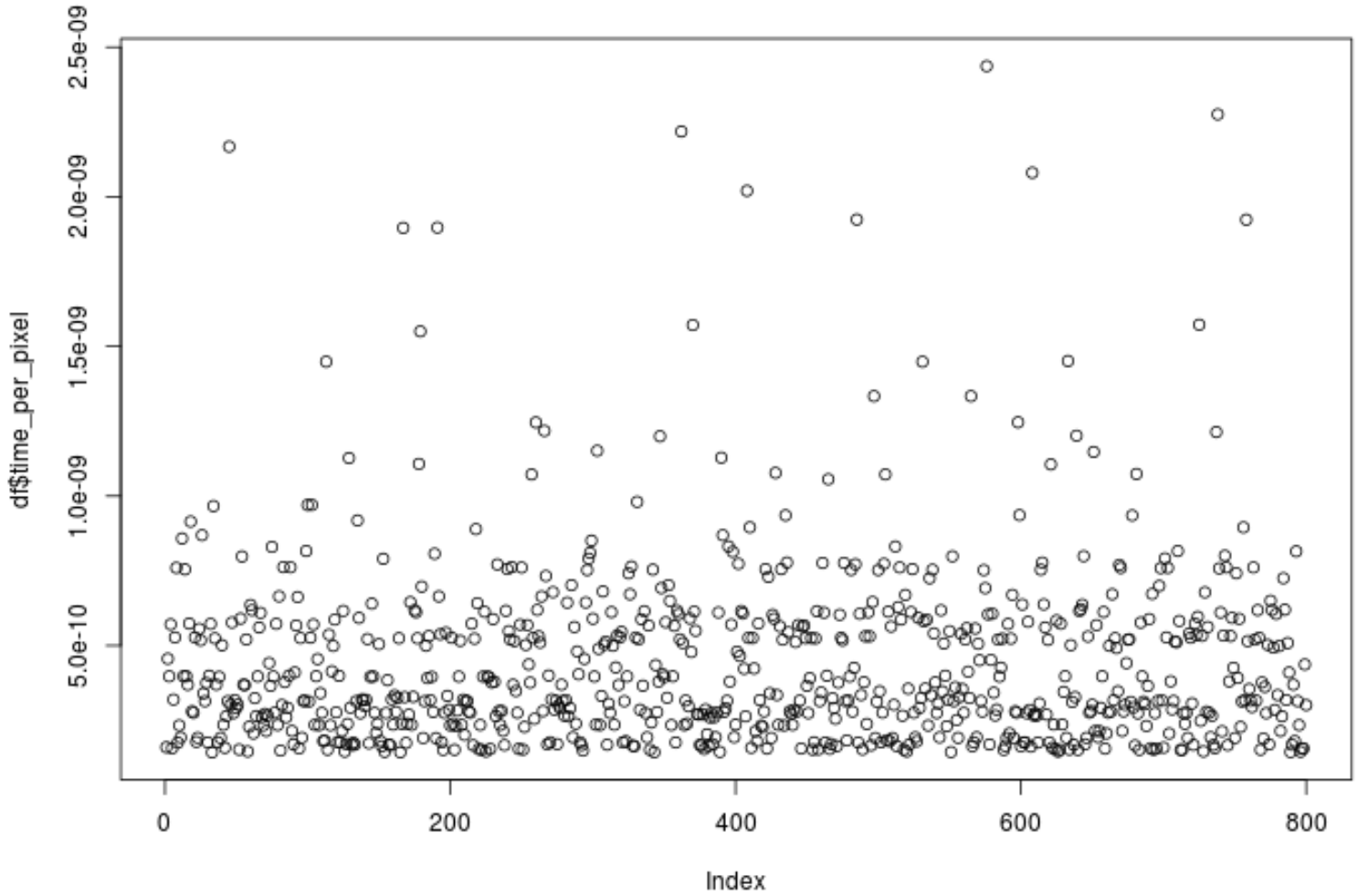
Topic

Laplacian

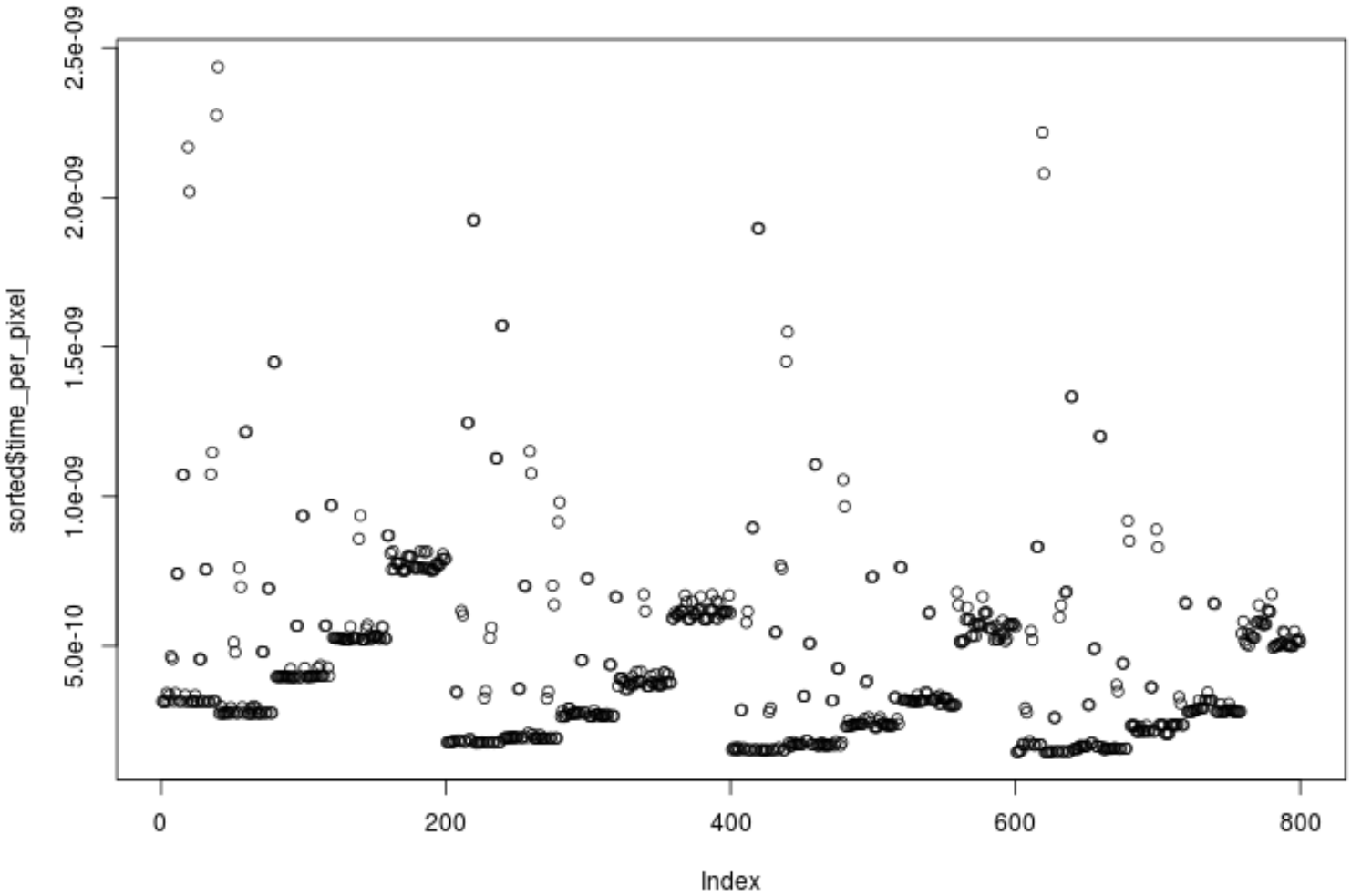
- Parameters:
  - x_component_number [1,2,4,8,16]
  - vector_length [1,2,4,8,16]
  - y_component_number [1,2,3,4]
  - temporary_size [2,4]
  - vector_recompute [true,false]
  - load_overlap [true,false]
- OpenCL Nvidia implementation

Brute force exploration on GPU NVIDIA

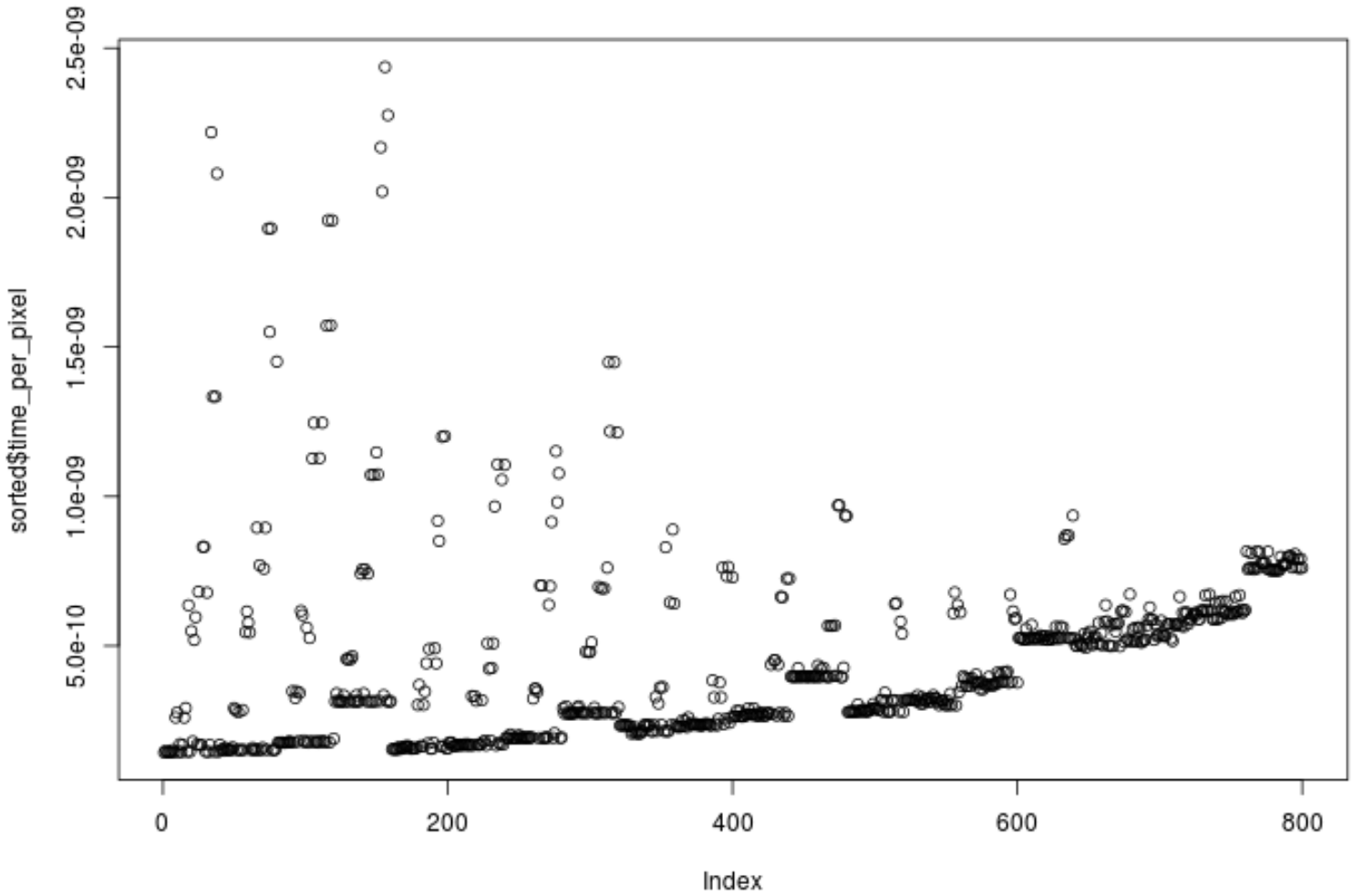
- Search space = 800
- Each version tested 4 time on 4 image sizes.



One possible order



And another one



Can we achieve an efficient "search space simplification" ?

- Facilitates the search
- Do we still need for complex exploration scheme ?
- Comparison between search technics which are correctly adapted to the autotuning search problem, with a correctly presented problem (especially with the random search with high number of parameters)

Topic

Objectives

- Find how to present the problem to have the nicest shape as possible in order to facilitate the search
- Characterization of the autotuning optimization search problems
- Which algorithms are the most suited for each kind of problems
- Devise an adaptive approach

Topic

End

Thank you for your attention

References I


Abramson, M. A., Audet, C., and Jr., J. E. D. (2004).
Generalized pattern searches with derivative information.
Math. Program., 100(1):3–25.


Ansel, J., Kamil, S., Veeramachaneni, K., Ragan-Kelley, J., Bosboom, J., O'Reilly, U.-M., and Amarasinghe, S. (2014).
Opentuner: An extensible framework for program autotuning.
In *Proceedings of the 23rd International Conference on Parallel Architectures and Compilation*, PACT '14, pages 303–316, New York, NY, USA. ACM.


De Oliveira Castro, P., Petit, E., Beyler, J. C., and Jalby, W. (2012).
ASK: Adaptive Sampling Kit for Performance Characterization.
In *Euro-Par 2012 Parallel Processing*, volume 7484, pages 89–101, Greece. springer.

References II

Fursin, G., Kashnikov, Y., Memon, A. W., Chamski, Z., Temam, O., Namolaru, M., Yom-Tov, E., Mendelson, B., Zaks, A., Courtois, E., Bodin, F., Barnard, P., Ashton, E., Bonilla, E., Thomson, J., Williams, C. K. I., and O'Boyle, M. (2011).
Milepost GCC: Machine Learning Enabled Self-tuning Compiler.
International Journal of Parallel Programming, 39:296–327.

Memon, A. W. and Fursin, G. (2013).
Crowdtuning: systematizing auto-tuning using predictive modeling and crowdsourcing.
In *PARCO mini-symposium on "Application Autotuning for HPC (Architectures)"*, Munich, Germany.

References III


Roy, A., Balaprakash, P., Hovland, P. D., and Wild, S. M. (2015). Exploiting performance portability in search algorithms for autotuning. Technical Report ANL/MCS-P5400-0915, Argonne National Laboratory.