cTuning.org

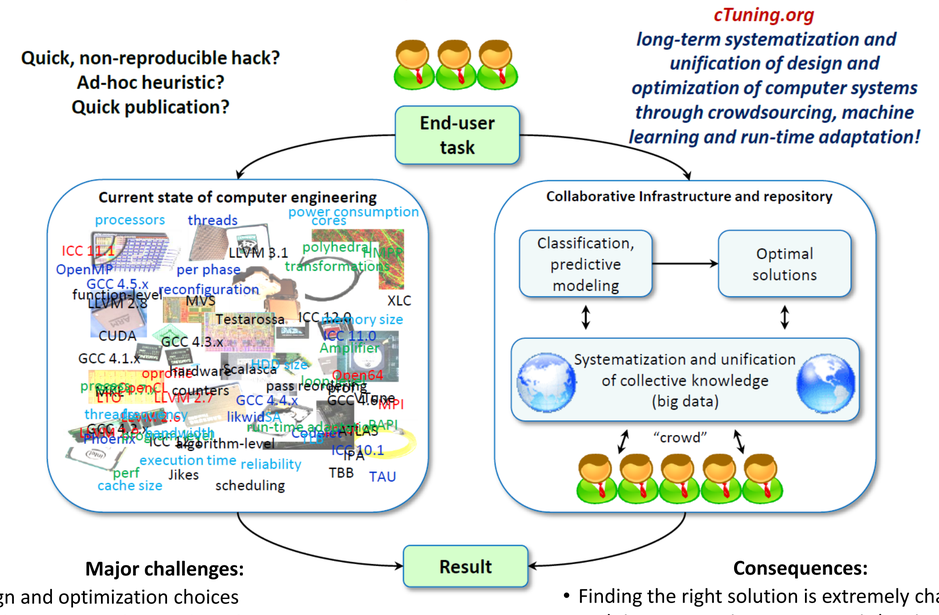
novel extensible methodology, framework and public repository to collaboratively address Exascale challenges

Grigori Fursin
INRIA, France

Acknowledgments

Path towards Exascale computing (2012-2018)

Which path to choose?



Major challenges:

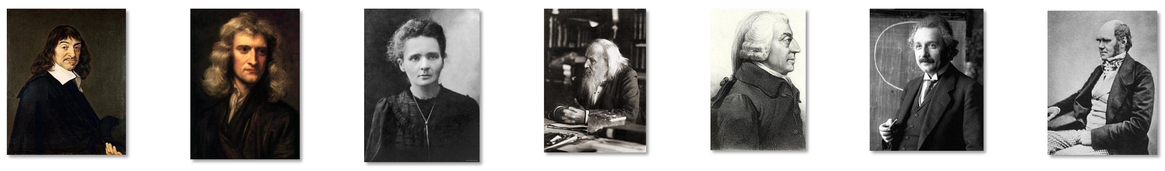
- Too many design and optimization choices
- Multiple user objectives vs. choices; benefit vs. costs
- Complex relationship and interactions between ALL software and hardware components.
- Too many tools with non-unified interfaces changing from version to version: technological chaos
- Too many papers with non-reproducible results: technological chaos

Consequences:

- Finding the right solution is extremely challenging and time consuming - everyone is lost in choices
- Slow time to market, low ROI
- Underperforming systems wasting expensive energy
- Attracting new students is difficult: bad image of ad-hoc solutions and hacking
- Slowing innovation in science and technology

cTuning long term interdisciplinary vision

Take the best of existing sciences that deal with complex systems:
computer science, physics, mathematics, chemistry, biology, etc

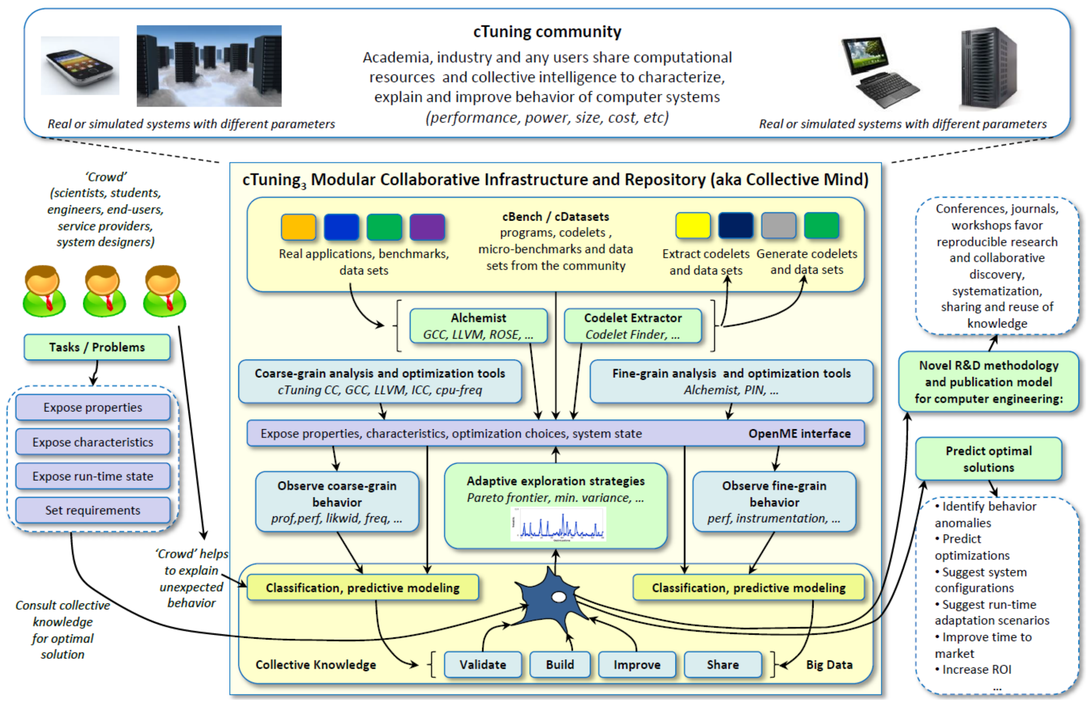


cTuning framework includes methodology, tools, and repository to systematize, quantify, unify and automate architecture and code design, optimization and run-time adaptation based on empirical, analytical and statistical techniques combined with learning, classification, predictive modeling and expert advice web services:

- Extensible and collaborative infrastructure and repository to record information flow within computer systems
- Continuous data collection and sharing from multiple users
- Collection of unified benchmarks, codelets, micro-benchmarks and datasets
- Continuous exploration of multiple design and optimization dimensions
- Plugins for online data mining and machine learning techniques to extrapolate existing knowledge to build faster, more power efficient and reliable devices
- Public web-services to suggest optimal program optimizations or architecture designs
- New publication model to reproduce experimental results by the community

Major publications available at http://cTuning.org/lab/education

Collective Mind infrastructure and repository for online auto-tuning and learning



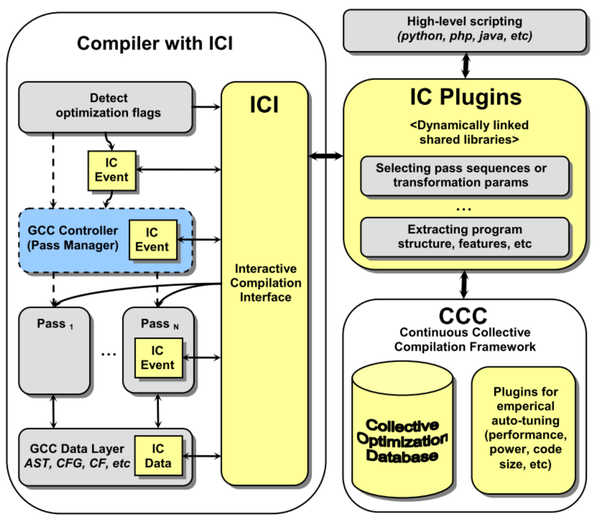
cM is the first framework to our knowledge that easily integrates various existing tools and techniques through plugins to crowdsource multi-objective auto-tuning and learning.

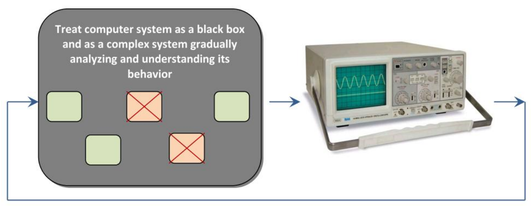
Empirical analysis and auto-tuning using interactive compilers

Novel concept to convert rigid tools into powerful interactive toolsets using light-weight event-based plugin framework.

We developed Interactive Compilation Interface (ICI) to "open up" existing compilers such as Open64 and GCC, and to enable transparent for end-users empirical multi-objective auto-tuning on the fly (exploration of large optimization spaces) and extraction of program features to enable predictive modelling.

In 2010, ICI (plugin framework) was added to the mainline GCC. We are developing a new version of the plugin framework ("OpenME") to be able to open up any existing compiler or tool.

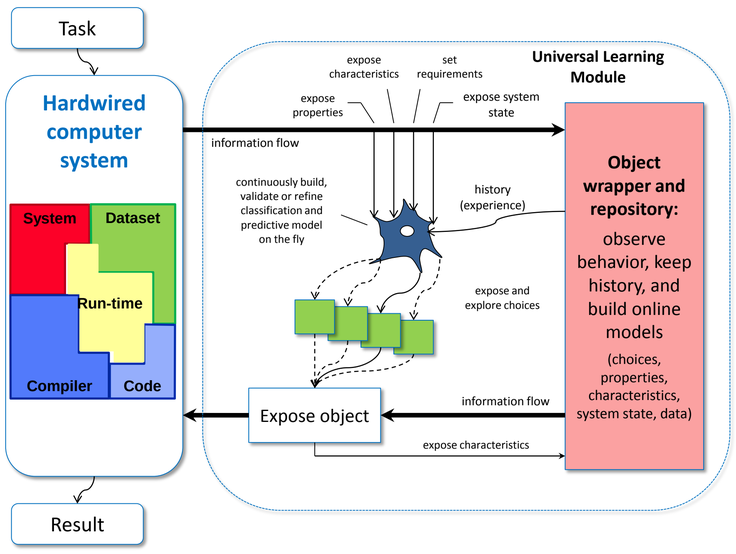




We implemented another novel concept to statistically characterize programs and architectures through reactions to optimizations or even semantically-non equivalent code modifications similar to physics.

(removing or adding individual instructions , code segments, threads, etc: for example to detect memory and cache bottlenecks or contentions)

Continuous online learning of a behavior of computer systems at all levels using statistical analysis and predictive modeling



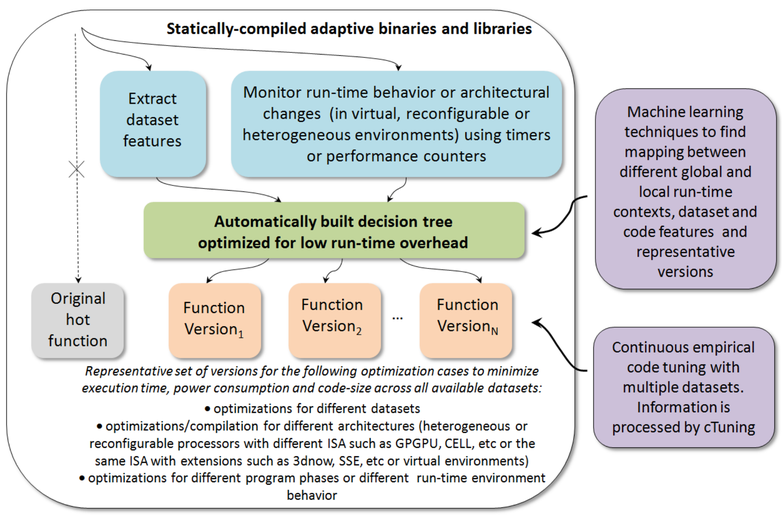
Collecting data from multiple users in a unified way allows to continuously apply and compare various machine learning techniques to correlate program and architecture behaviour, static and dynamic features, designs and optimizations.

We use continuously updated predictive models (accessible through online web-services) to quickly suggest better optimizations for a given user program, codelet, dataset and architecture to balance multiple objectives such as performance, power, compilation time, code size, scheduling on GPGPUs, etc...

cTuning$_1$ proof-of-concept framework has been released in 2009 at cTuning.org (see IBM's world wide press release "*World's First Intelligent, Open Source Compiler Provides Automated Advice on Software Code Optimization*") . Since then, it has been extended within multiple international collaborative projects and Google Summer of Code program.

New version of cTuning aka Collective Mind to crowdsource auto-tuning will be released in 2013.

Building self-tuning computer systems (combining static and dynamic approaches with online learning)



We developed a novel approach to statically enable dynamic optimizations (UNIDAPT framework) by combining a small set of pre-optimized versions of a code with online learning plugins to quickly select the most appropriate versions at run-time adapting to a given (heterogeneous) architecture or varying program phase due to multiple datasets, contentions, etc.

Collective Mind framework uses machine learning techniques to continuously determining minimal representative sets of codelets and optimizations that cover varying program behavior due to different datasets, run-time program and system behavior, etc.

We gradually release all tools, plugins, benchmarks and datasets at cTuning.org to enable collaborative and reproducible R&D, and new publication model.

More information about cTuning.org, Collective Mind, ICI, UNIDAPT framework can be found in our publications at:
http://cTuning.org/lab/education#publications

Contact and further information:

Grigori Fursin: grigori.fursin@inria.fr ; http://cTuning.org/lab/people/gfursin

http://cTuning.org (tools, benchmarks, datasets, web-services, collaborative wiki)
http://groups.google.com/group/ctuning-discussions (public discussions)

Join!
Invest!