NTNU – Trondheim
Norwegian University of
Science and Technology

Multi-core HW/SW interplay and energy efficiency — examples and ideas

Lasse Natvig
CARD group, Dept. of comp.sci. (IDI) - NTNU
& HPC-section – NTNU

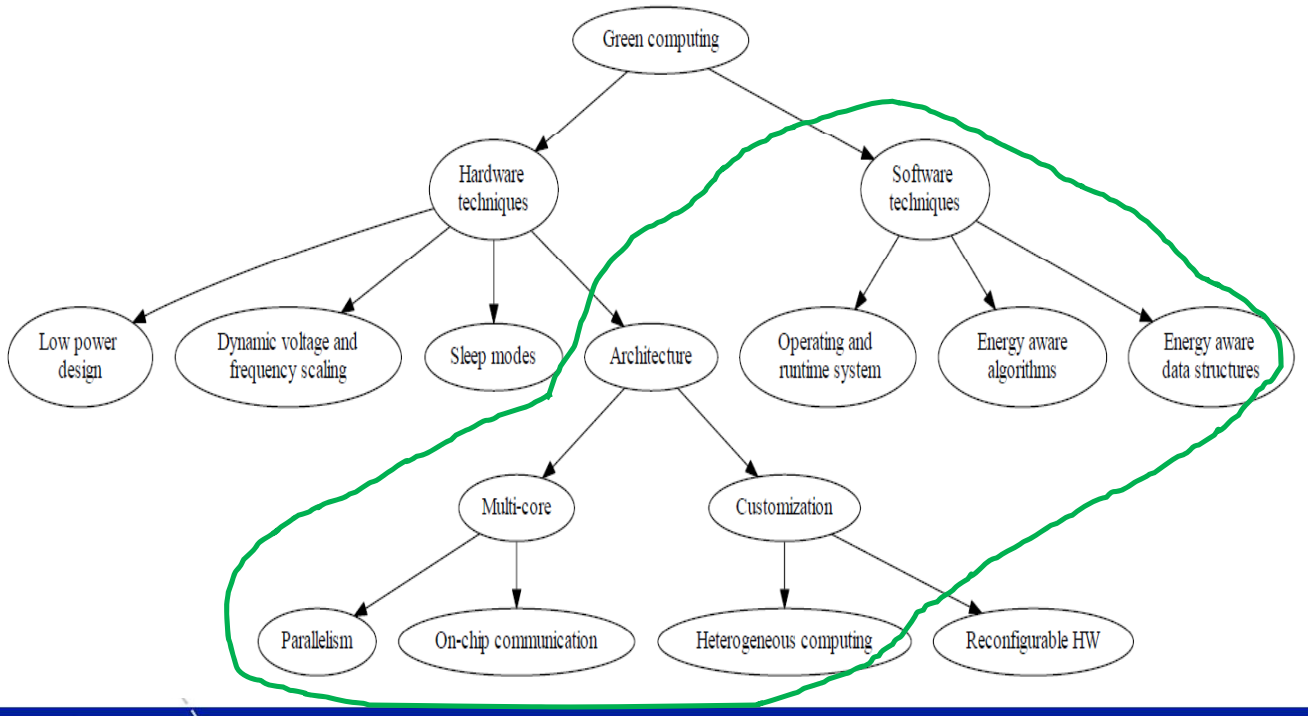
HiPEAC3 – Gøteborg 24/4-2012

Too large design space?

Very young and highly dynamic technology ⇒ far too many degrees of freedom !?

Instability gives opportunities

Energy efficiency design space

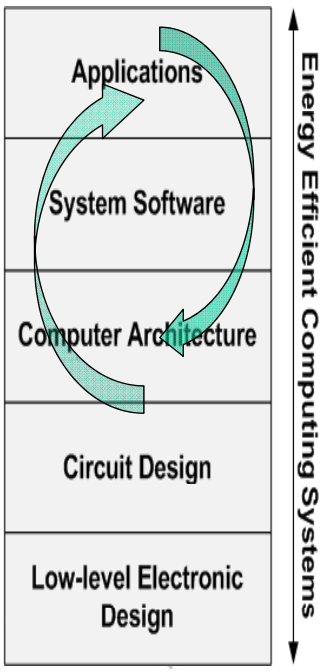


- **Claim:**
  Dynamic adaptivity in the HW/SW interface is a central opportunity for improved multi-core energy efficiency

Presentation Overview

- Introduction
- HW/SW interplay
  - Levels and parameters
  - Successful examples
- Some recent energy efficiency results
  - OmpSs on Intel Sandy Bridge
- Two proposals/ideas for cTuning.org version (i+1)

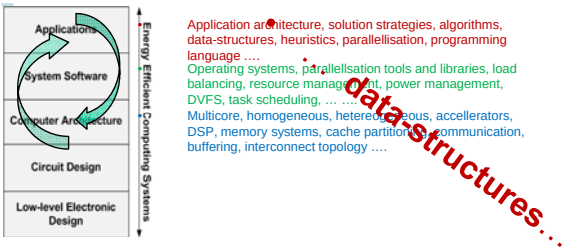
HW/SW interplay – levels



Application architecture, solution strategies, algorithms, data-structures, heuristics, parallellisation, programming language ….

Operating systems, parallellsation tools and libraries, compilers, compiler options, profilers, analysis tools, load balancing, resource management, power management, DVFS, task scheduling, … ….

Multicore, homogeneous, hetereogeneous, accellerators, DSP, memory systems, cache partitioning, communication, buffering, interconnect topology ….



… and parameters

- Lock-free data structures;
  - Reduced serialization
  - Concurrent updates
  - ⇒ looser "unordered" concurrent constructs based on distribution and randomization [Shav11]
- **Compressed** data structures?

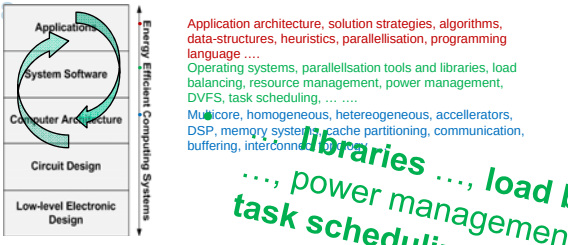
DOI:10.1145/1897852.1897873

The advent of multicore processors as the standard computing platform will force major changes in software design.

BY NIR SHAVIT

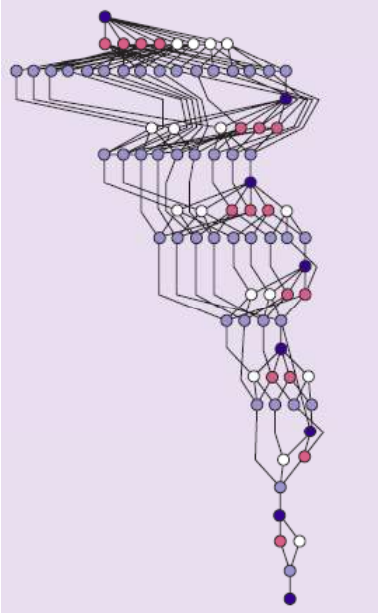
Data Structures in the Multicore Age

Trade-off:
complexity vs. energy

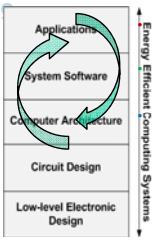


… and more

- **Libraries** with different implementations for different energy budgets (algorithms, level of parallelism)
- OmpSs **task scheduling** - alternatives



NTNU – Trondheim
Norwegian University of Science and Technology



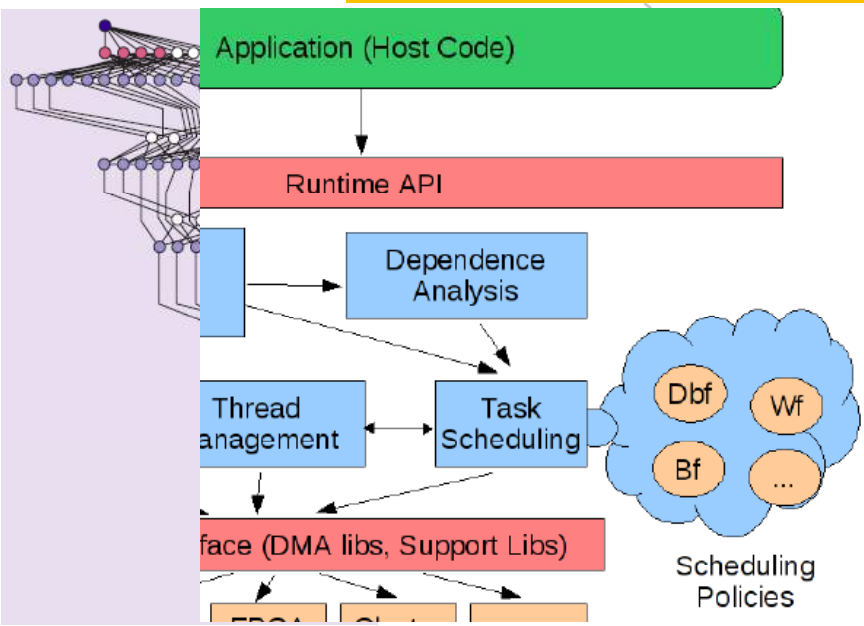
Application architecture, solution strategies, algorithms, data-structures, heuristics, parallellisation, programming language ….
Operating systems, parallellsation tools and libraries, load balancing, resource management, power management, DVFS, task scheduling, … ….
Multicore, homogeneous, hetereogeneous, accelerators, DSP, memory systems, cache partitioning, communication, buffering, interconnect topology ….

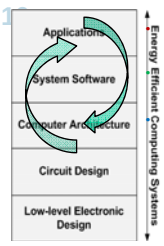
libraries …, load balancing, …, power management, DVFS, task scheduling, … ….

… and more

Trade-off:
complexity vs. energy

- **Libraries** with different implementations for different energy budgets (algorithms, level of parallelism)
- OmpSs **task scheduling** - alternatives

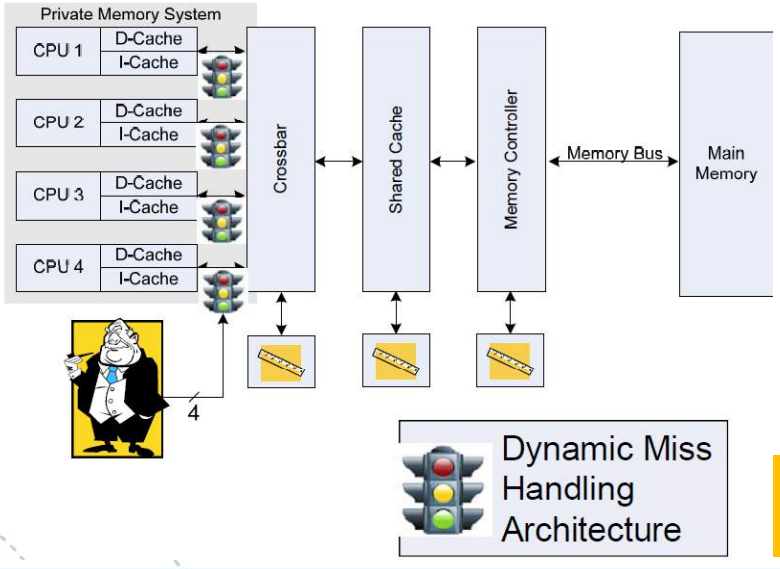




Application architecture, solution strategies, algorithms, data-structures, heuristics, parallellisation, programming language ….
Operating systems, parallellsation tools and libraries, load balancing, resource management, power management, DVFS, task scheduling, … ….
Multicore, homogeneous, hetereogeneous, accelerators, DSP, memory systems, cache partitioning, communication, buffering, interconnect topology ….

… and even more

- Multicore, … **dynamic cache partitioning**, **miss handling architectures**, communication, buffering, **prefetching**, interconnect topology ….

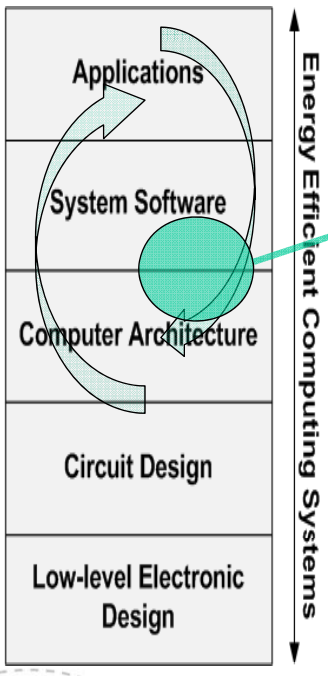


A High Performance Adaptive Miss Handling Architecture for Chip Multiprocessors,
M. Jahre and L. Natvig,
In Trans. on HiPEAC 2009

Add energy studies

Is this huge variety a necessity?

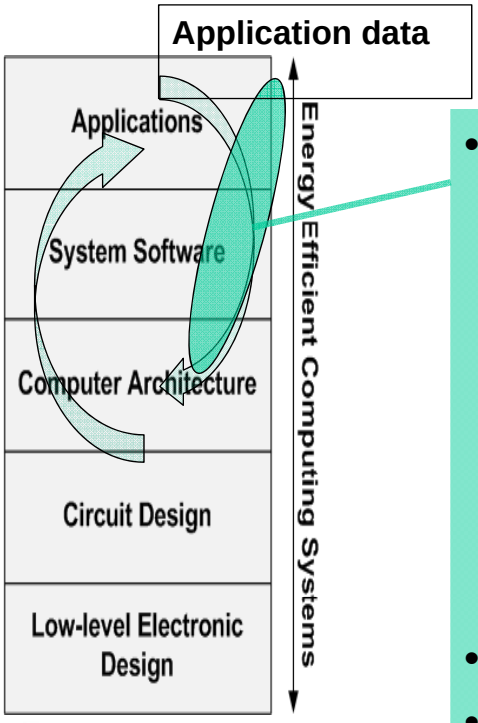
Example: ATLAS project



- Example of Automated Empirical Optimization of Software (AEOS)
  - Replaces hand-coded optimization for one specific platform with empirical, automated performance tuning for a large set of platforms
  - Uses code generators
- ⇒Portable performance
- Static (compile time) optimization
- [WPD01]

NTNU – Trondheim
Norwegian University of Science and Technology

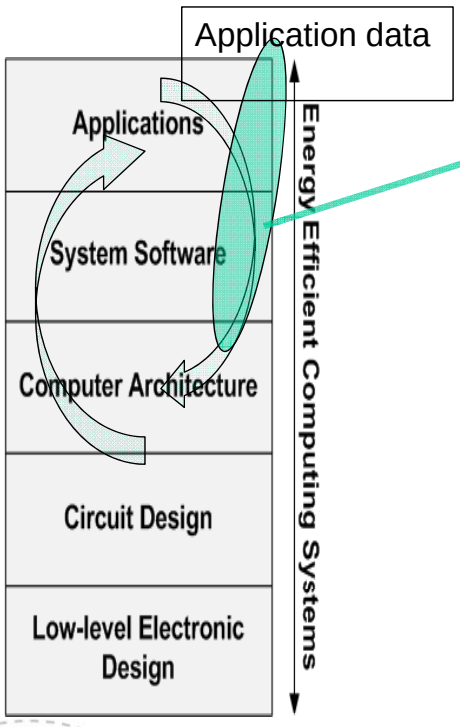
Example: FFTW



- Similar approach with ATLAS
  - Adaptive FFT program that tunes the computation automatically for the HW
  - Retains complete portability
  - Competetive with or faster than codes optimized for a single machine
  - Composable building blocks (codelets)
  - Planner determines execution plan at runtime
    - Plan can be stored on disk
- ”Semi-static optimization”
- [FJ05]

NTNU – Trondheim
Norwegian University of Science and Technology

Example: CSX

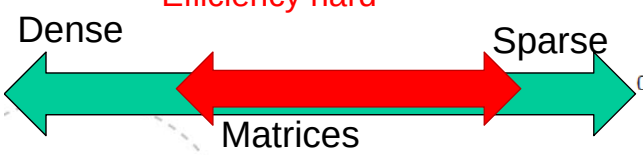


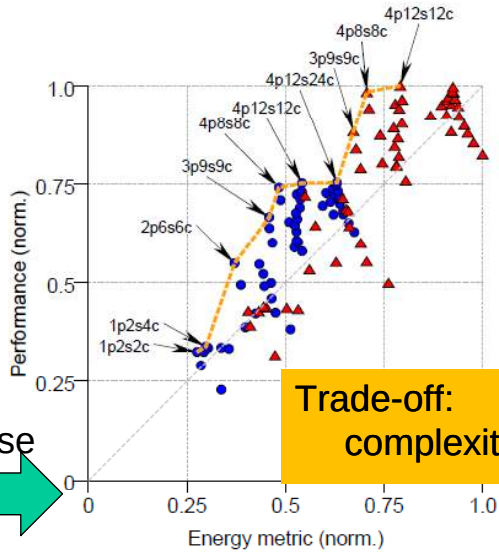
- Compressed Sparse eXtended (CSX)
  - For Sparse Matrix-Vector multipl. (SpMV)
  - Sparse m. formats includes metadata
  - CSX compress metadata by exploiting substructures in the matrix
  - Uses runtime code generation to construct specialized SpMV routines for each matrix
  - General approach
  - Tradeoff between performance and preprocessing cost
- Dynamic optimization
- [KKGK10]

NTNU – Trondheim
Norwegian University of Science and Technology

Future work; CSX and energy performance tradeoff studies

- Joint work NTNU-Trondheim and NTU-Athens
  - In scope of PRACE IP2 – WP12, task 12.1 autotuning
- Sparsity
  - Efficiency easy
  - Efiiciency hard





Trade-off:
complexity vs. energy

Presentation Overview

- Introduction
- HW/SW interplay
  - Levels and parameters
  - Successfull examples
- Some recent energy efficiency results
  - OmpSs on Sandy Bridge
- Proposals/ideas for cTuning.org version (i+1)

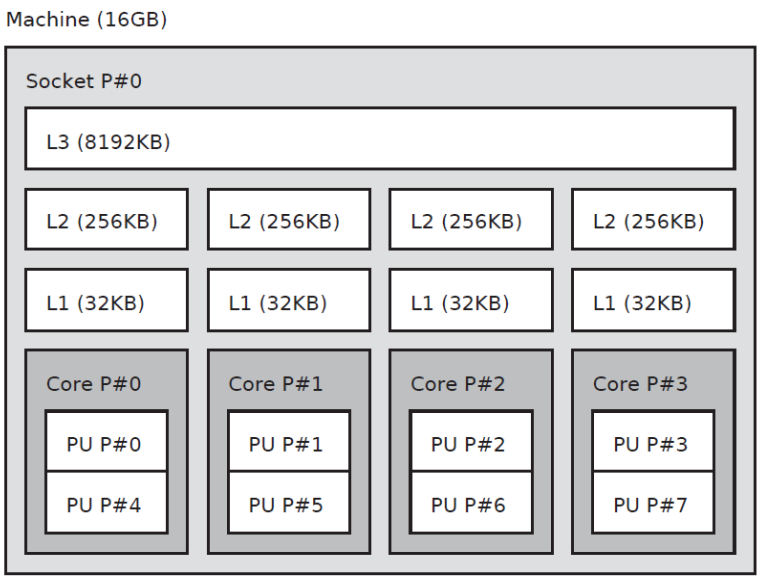
NTNU – Trondheim
Norwegian University of Science and Technology

Power, energy, performance – metrics

- (Energy = power × time)
- Examples
  - Energy Delay Product (EDP)
  - Performance per Watt
  - Speedup per Watt [MLH10]
- Balance performance against energy
  - Performance^N / Watt
    - N = 2 gives EDP
    - N = 0 gives Power alone
  - Different optimization criteria
    - Performance (Classical HPC view)
    - Energy (Eg. Offshore senSor-newtworks)
- **GFLOPS/Watt**
  - **Used in Green500, targetted by Mont Blanc project**

Experimental setup

- Energy consumption from Machine State Registers (MSRs)
- Available in Running Average Power Limit (RAPL) interface
- Observe L3-cache miss rate (keep low)
  ⇒"on-chip energy efficiency"
- Median of 10 runs, relative STDEV < 3 %

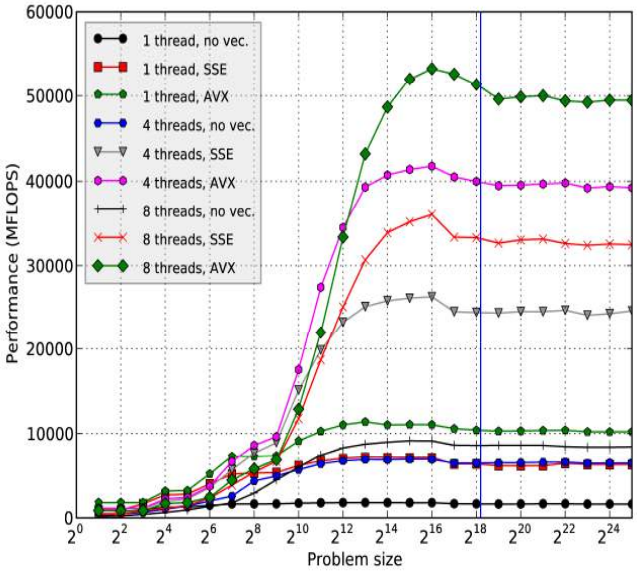


- CPU Core C-states
- CPU/ PG/ Package power
- Built-in power monitoring
- Power Budget Management
- Platform Control (EC / VR)

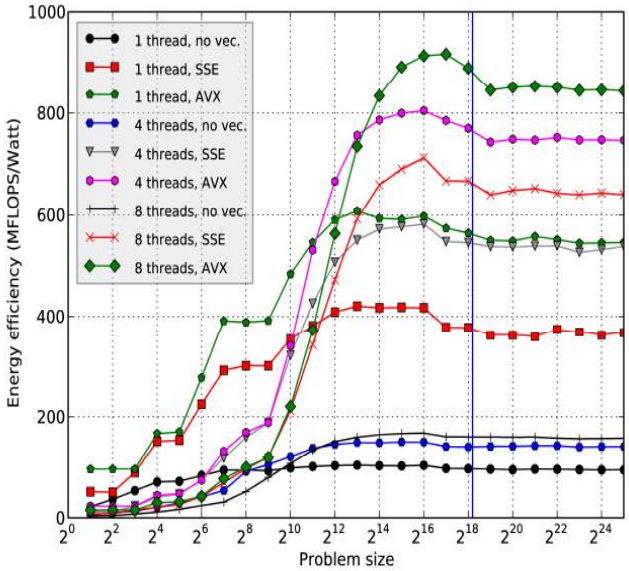
- HW controlled power sharing between CPU - PG
- Brief turbo above TDP → dynamic Turbo
- More platform control via PECI 3.0 and SVID

OmpSs – Black Scholes

- On-chip energy efficiency

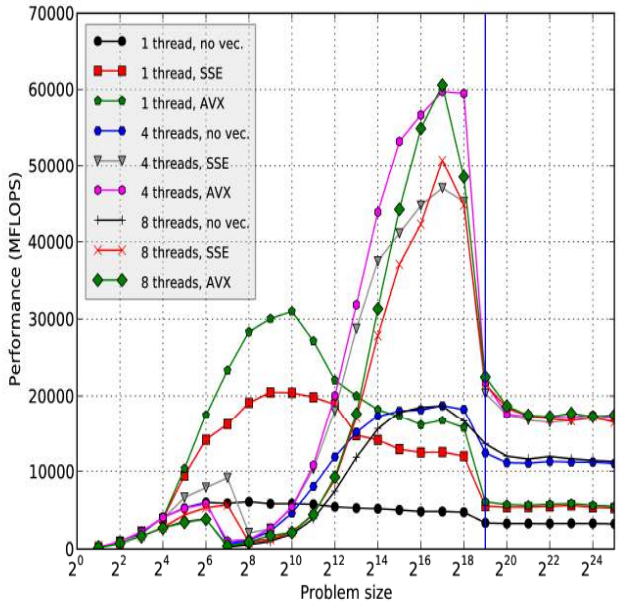


(a) Performance vs. problem size

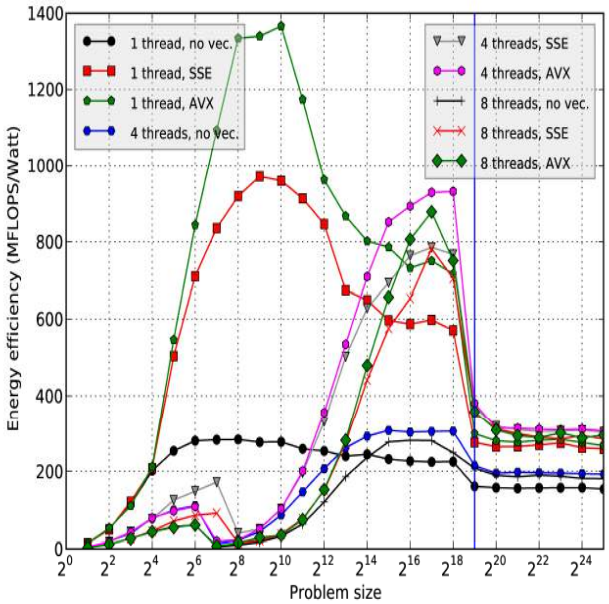


(b) Energy efficiency vs. problem size

OmpSs – FFTW

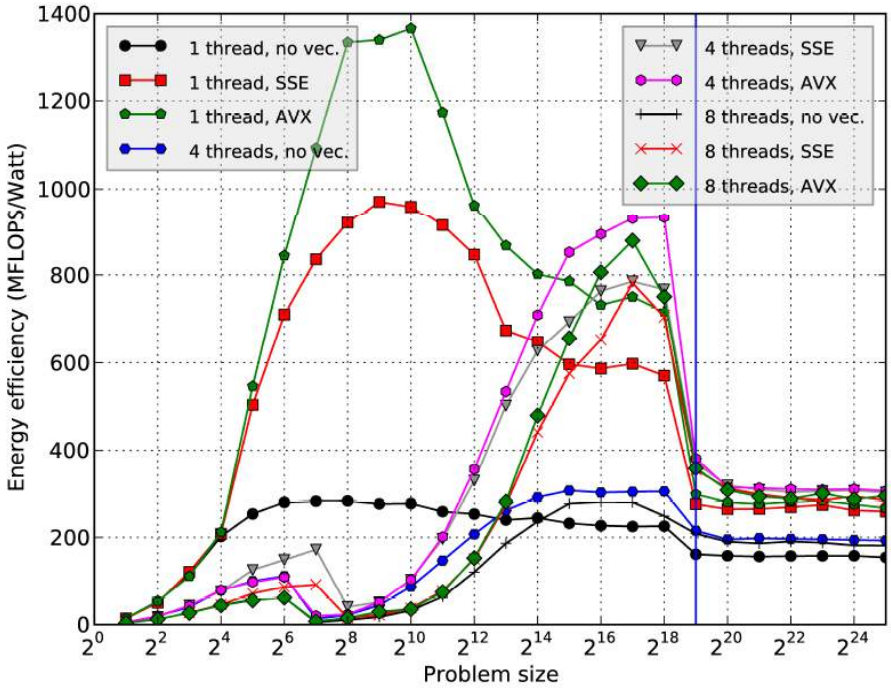


(a) Performance vs. problem size



(b) Energy efficiency vs. problem size

OmpSs – FFTW on-chip energy efficiency



(b) Energy efficiency in MFLOPS/watt

Main observations, results

- Peak on-chip GFLOPS/W rates:
  - 0.89 (Black Scholes)
  - 1.38 (FTTW)
  - 1.97 (Matrix Multiply)
- Green500 November 2011
  - 2.03 (LINPACK), but …
- SSE and AVX vectorization favorable
- FFTW
  - No benefit from hyperthreading
  - Bandwidth bound for problems larger than cache
- [Lien12]

To be continued …

- SGI supercomputer based on Sandy Bridge
  - Operational from June 2012, first in Europe?
- Node specification
  - 2 x 8 core nodes
  - 8-core E5-2600 socket R «2-3 GHz»
  - 20 MiB shared L3/chip.
  - AVX floating-point
- > 20 000 cores

Norwegian Meteorological Institute
met.no

NTNU – Trondheim
Norwegian University of Science and Technology

Presentation Overview

- Introduction
- HW/SW interplay
  - Levels and parameters
  - Successfull examples
- Some recent energy efficiency results
  - OmpSs on Sandy Bridge
- Proposals/ideas for cTuning.org version (i+1)

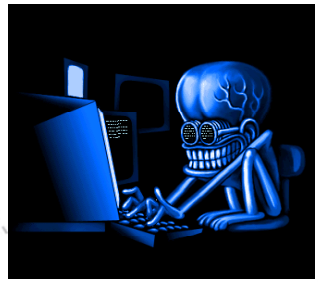
NTNU – Trondheim
Norwegian University of Science and Technology

Ideas for cTuning.org version (i +1)

- Energy
  - Various metrics
- Data structure exploration
  - Alternative structures
  - Degree of loss-less compression
  - Example: matrices
    - Dense …. Sparse
    - Different formats
    - Different compression methods

NTNU – Trondheim
Norwegian University of Science and Technology

> gcc –O young_student



- Could we exploit the masses of young students in improving our knowledge in energy efficient computing?

IDI Open 2012

Online Judge

Home ▸ Site Statistics ▸ Last 50 Submissions

Google™ Custom Search
Search

Main Menu

- Home
- My Account
- Contact Us
- TOOLS on the Old UVa OJ Site
- ACM-ICPC Live Archive
- Logout

Online Judge

Last 50 Submissions

#		Problem	User	Verdict	Language	Run Time	Submission Date
10025305	10245	The Closest Pai...	Ahmad Harf...	Running	C++	0.000	2012-04-23 10:17:55
10025304	1152	4 Values whose...	u_aizu	Accepted	C++	4.892	2012-04-23 10:17:54
10025303	572	Oil Deposits	黄汉升	Accepted	C++	0.008	2012-04-23 10:17:40
10025302	11854	Egypt	Faisal Ahm...	Accepted	ANSI C	0.012	2012-04-23 10:17:40
10025301	11639	Guard the Land	张翼德	Wrong answer	C++	0.004	2012-04-23 10:17:37
10025300	10071	Back to High Sc...	heartofsto...		ANSI C	0.000	2012-04-23 10:17:34
10025299	572	Oil Deposits	马孟起	Wrong answer	C++	0.008	2012-04-23 10:17:19
10025298	10075	Airlines	黄汉升	Wrong answer	C++	0.096	2012-04-23 10:17:19
10025295	991	Safe Salutation...	张翼德	Compilation error	ANSI C	0.000	2012-04-23 10:17:19

Questions / Discussion

Visit our website:
http://research.idi.ntnu.no/multicore/

Contact:
Lasse.Natvig
@idi.ntnu.no

References

- [Sha11] Shavit Nir, *Data structures in the multicore age*, Commun. ACM, March 2011.
- [WPD01] Automated empirical optimizations of software and the ATLAS project, R. Clint Whaley, Antoine Petitet, Jack J. Dongarra, Parallel Computing 2001
- [FJ05] Matteo Frigo and Steven G. Johnson, The Design and Implementation of FFTW3, Proc. of the IEEE 93 (2), 216–231 (2005).
- [KKGK10] Kornilios Kourtis, Vasileios Karakasis, Georgios I. Goumas, Nectarios Koziris: CSX: an extended compression format for spmv on shared memory systems. PPOPP 2011: 247-256
- [Amun11] Jørn Amundsen, VILJE – the new supercomputer at NTNU, NOTUR Meta magazine, http://www.notur.no/publications/magazine/pdf/meta_2011_4.pdf
- [Lien12] Hallgeir Lien, Master Thesis, IDI – NTNU, July 2012
- [MLH10] Metrics and Task Scheduling Policies for Energy Saving in Multicore Computers, J. Mair, K. Leung, Z. Huang

NTNU – Trondheim
Norwegian University of
Science and Technology