INTELLECTUAL PROPERTY RIGHTS NOTICE:

- The User may only download, make and retain a **copy of the materials for his/her use for non-commercial** and research purposes.
- The User **may not commercially use the material**, unless has been granted prior written consent by the Licensor to do so; and **cannot remove, obscure or modify copyright notices**, text acknowledging or other means of identification or disclaimers as they appear.
- For further details, please contact **BSC-CNS** patc@bsc.es

PATC training, Barcelona, May 2012 ‹#›

PRACE TRAINING COURSE under PRACE Advance Training Centre at BSC

BSC-CNS http://www.bsc.es/

PRACE project http://www.prace-ri.eu/

PRACE Training Portal http://www.training.prace-ri.eu/

PATC @ BSC Training Program

http://www.bsc.es/marenostrum-support-services/hpc-trainings/prace-trainings

PATC training, Barcelona, May 2012 ‹#›

Performance analysis with BSC-Tools

Judit Gimenez (judit@bsc.es)
BSC

Outline

Advanced Analysis
- Scaling model
- Structure detection: Clustering
- Folding: Instrumentation + sampling

Scaling model

Presenting application performance

Factors modeling parallel efficiency

- **Load balance** (LB)
- **Micro load balance** (μLB) or **serialization**
- **Transfer**

$$\eta = LB * \mu LB * Transfer$$

Factors describing serial behavior

- Performance: **IPC**

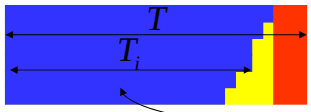
Scaling model

$$Sup = \frac{P}{P_0} * \frac{\eta}{\eta_0} * \frac{IPC}{IPC_0} * \frac{\#instr_0}{\#instr}$$

Scaling model

$$Sup = \frac{P}{P_0} * \frac{LB}{LB_0} * \frac{CommEff}{CommEff_0} * \frac{IPC}{IPC_0} * \frac{\#instr_0}{\#instr}$$

Directly from real execution metrics



$$eff_i = \frac{T_i}{T}$$

$$CommEff = \max(eff_i)$$

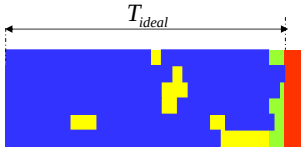
$$LB = \frac{\sum_{i=1}^{P} eff_i}{P * \max(eff_i)}$$

$$\frac{IPC}{\#instr}$$

$$microLB = \frac{\max(T_i)}{T_{ideal}} \qquad Transfer = \frac{T_{ideal}}{T}$$

$$Sup = \frac{P}{P_0} * \frac{macroLB}{macroLB_0} * \frac{microLB}{microLB_0} * \frac{Transfer}{Transfer_0} * \frac{IPC}{IPC_0} * \frac{\#instr_0}{\#instr}$$

Requires Dimemas simulation



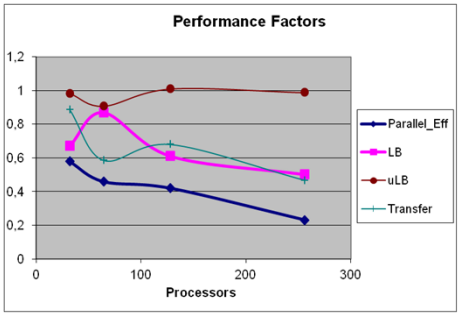
Migrating/local load imbalance
Serialization

BSC tools @ PATC, Barcelona May 21-22 2012 7

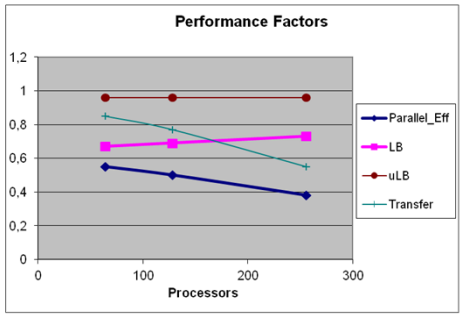
Parallel Performance Models

GROMACS

Original Version



v4.5



Old → new version: -43% instr, -52% time

BSC tools @ PATC, Barcelona May 21-22 2012 8

Structure detection: Clustering

Performance @ serial computation bursts

Burst = continuous computation region

between exit of an MPI call and entry to the next

Scatter plot representation of bursts

Collapse time dimension

N dimensional space of HWC derived relevant metrics

- Instructions: idea of computational complexity, computational load imbalance,…
- IPC: Idea of absolute performance and performance imbalance

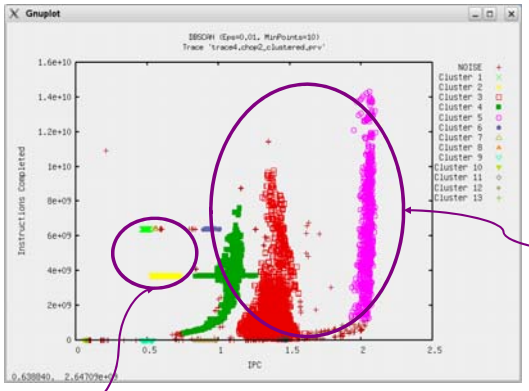
Structure

Clouds, clusters: Burst of “similar” characteristics

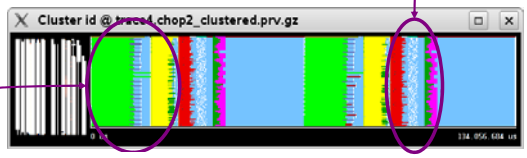
How those similarities spread in the timeline?

How does it relate to source code?

Clustering module – detecting structure



Huge variability on instructions



Most relevant computing regions obtain poor IPC

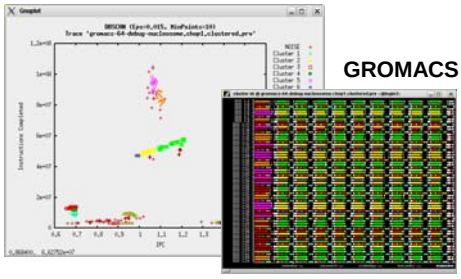
Unbalance due to different number of chunks

BSC tools @ PATC, Barcelona May 21-22 2012 11

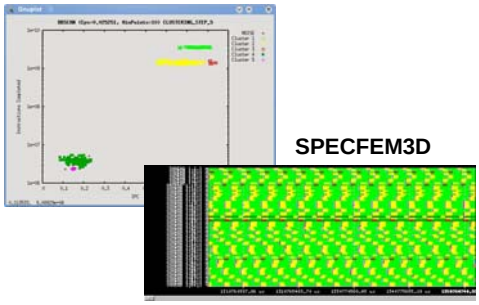
Clustering module – detecting structure

Highlight structure

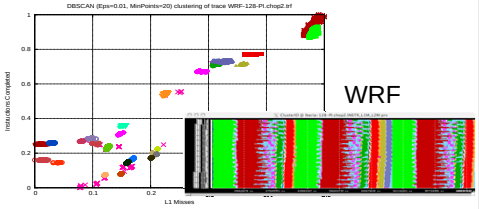
- Clusters injected in trace
- Phases vs. routines



GROMACS



SPECFEM3D



WRF

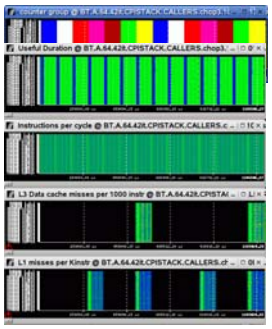
BSC tools @ PATC, Barcelona May 21-22 2012 12

HWC projection

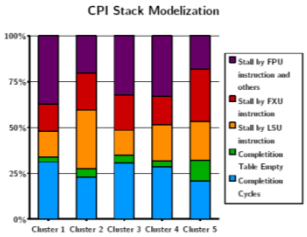
Full characterization

All HWCs and ratios between them
CPI stack model

From a single run



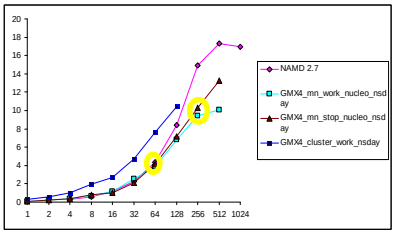
Cluster	1	2	3	4	5
%Time	54.88	17.96	16.90	6.44	1.42
Avg. Burst Dur. (ms)	1.02	0.78	13.14	2.50	1.11
IPC	1.02	0.65	0.89	0.91	0.53
MIPS	2231.8	1423.3	1966.5	2001.8	1163.0
MFLOPS	339.2	46.3	191.6	269.2	23.6
L1M/KInstr	0.92	1.53	1.19	1.17	2.88
L2M/KInstr	0.06	1.26	0.06	0.35	0.21
Mem.BW (MB/s)	16.79	218.47	13.87	85.77	29.76

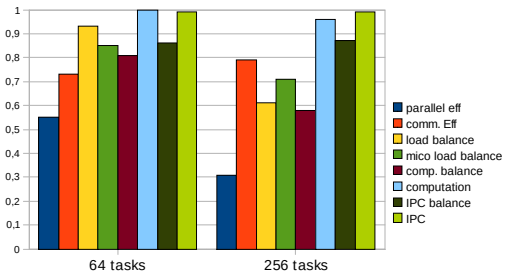


BSC tools @ PATC, Barcelona May 21-22 2012 13

Example: From quick analysis...

GROMACS analysis





Main scalability problems:
. Load balance
. Computation balance

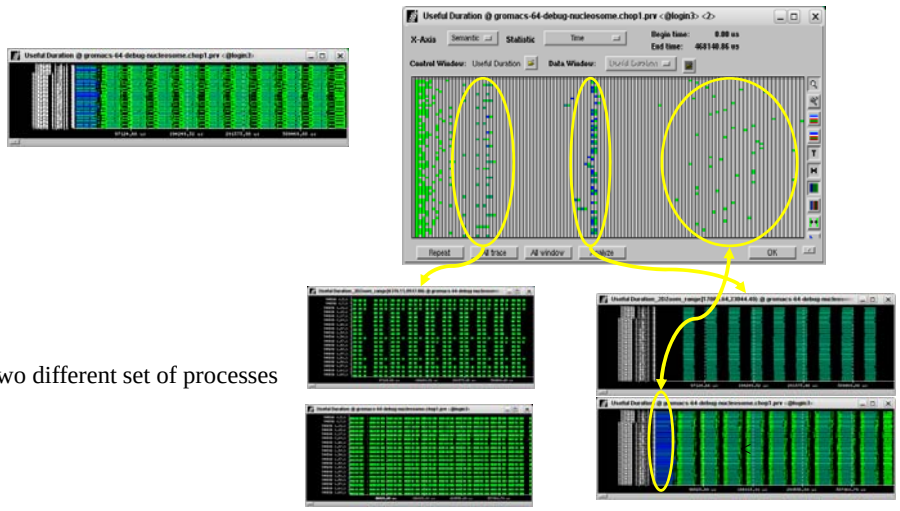
4% of code replication

64 tasks already poor efficiency!
. Communication problem

BSC tools @ PATC, Barcelona May 21-22 2012 14

Example: ... to details

64 tasks computation duration analysis

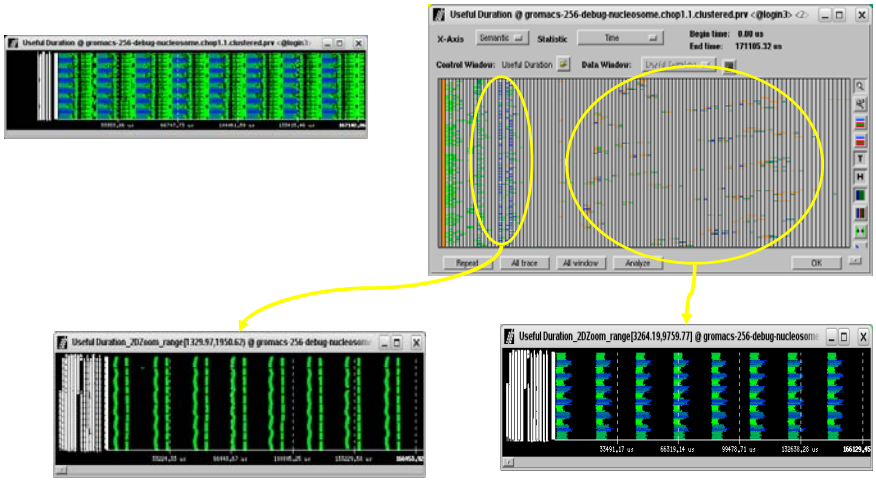


Two different set of processes

BSC tools @ PATC, Barcelona May 21-22 2012 15

Example: ... to details

256 tasks computation duration analysis

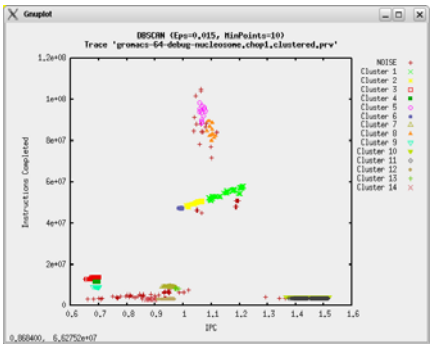


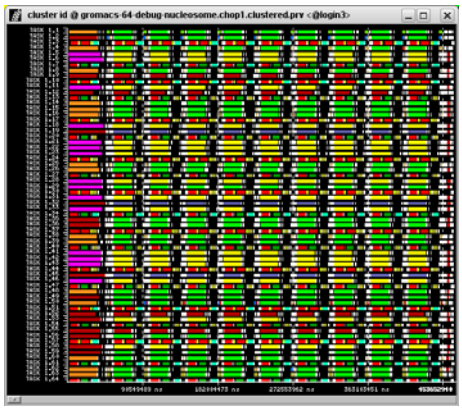
Two different set of processes unbalance increases

BSC tools @ PATC, Barcelona May 21-22 2012 16

Example: ... to details

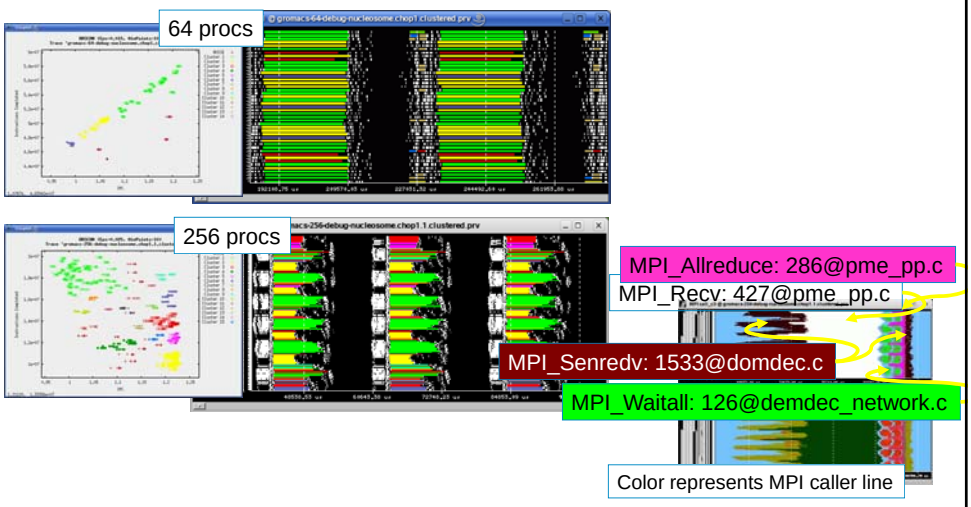
GROMACS structure, 64 tasks





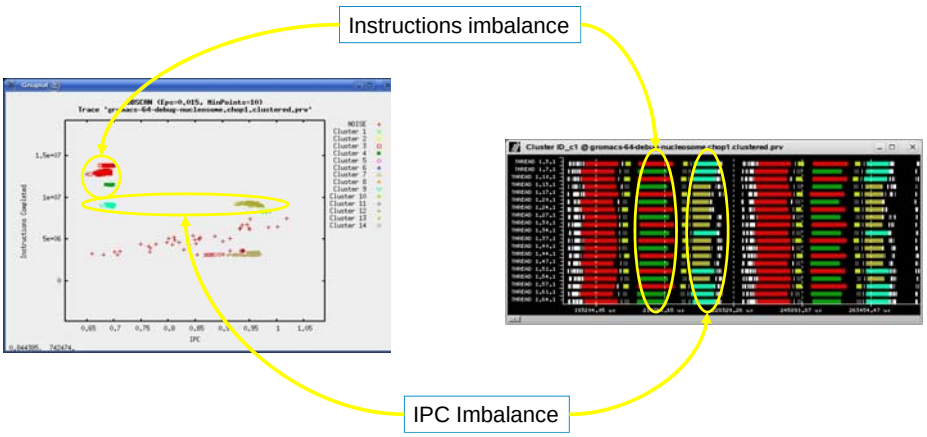
Example: ... to details

GROMACS PMEs balance / scalability



Example: ... to details

GROMACS FFTs balance

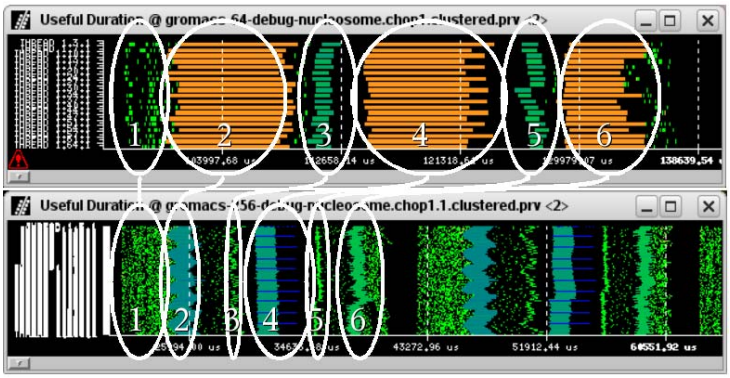


Example: ... to details

GROMACS: Comparing computation time for FFTs

64 tasks
(one iteration)

256 tasks
(2,2 iterations in the same time interval, 4 times more resources)



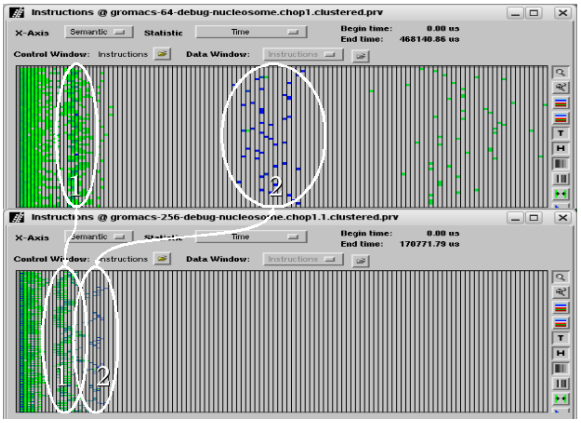
Zone 4 has a potentially structured imbalance with scalability problems (MPI_Alltoallv: 241@fftgrid.c)

Example: ... to details

GROMACS Computational load balance

64 tasks

256 tasks



Zone 1 does not scale and increases imbalance!

Folding: Combining instrumentation and sampling

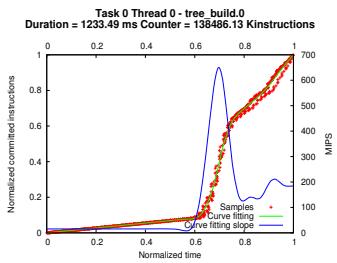
Folding samples – granularity / overhead

Application granularity vs. detailed granularity

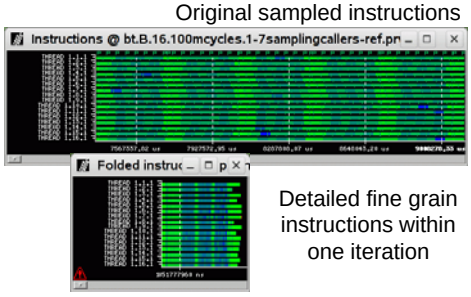
Samples: hardware counters + callstack

Folding: based on known structure: iterations, routines, clusters; project all the samples into one instance

Extremely detailed time evolution of hardware counts, rates and callstack with minimal overhead

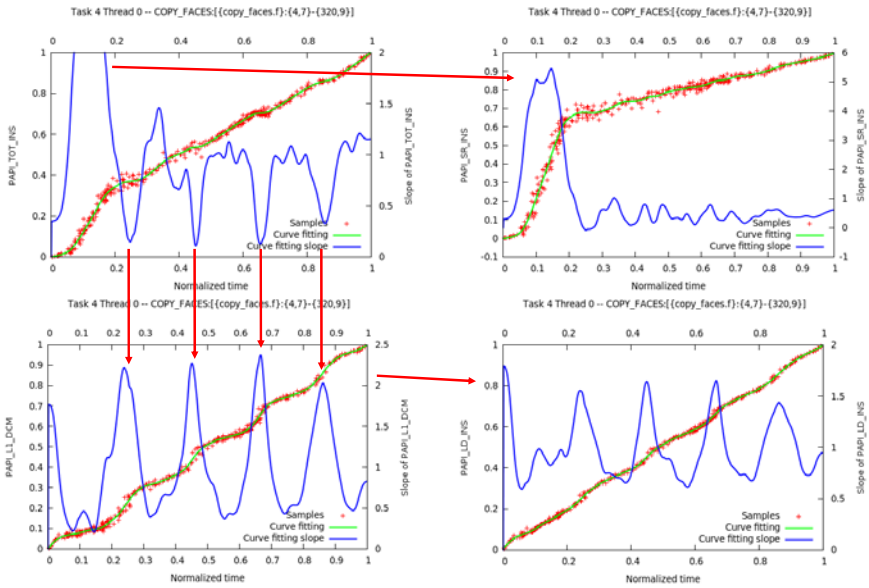


Original sampled instructions



Detailed fine grain instructions within one iteration

Folding samples – granularity / overhead

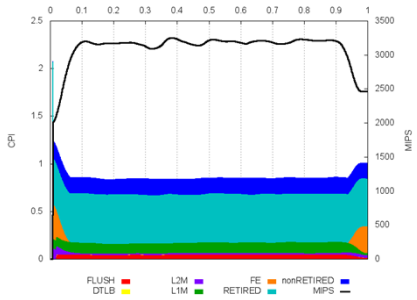


CPI Breakdown: PMEMD (2 tasks)

Core2 model

(core2duo t9400 @ 2.533 GHz)

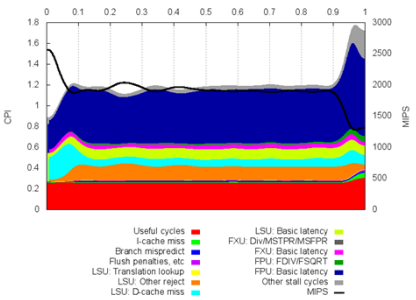
Intel Core2Duo CPI model evolution for pmemd.2tasks.x86.2.chop1.clustered.fused.extract.1.1.Cluster_1.0



Power5 model

(ppc970mp @ 2.3 GHz)

PowerPC model evolution for pmemd.2tasks.ppc.chop1.clustered.fused.extract.1.1.Cluster_1.0



FLUSH: Cycles stalled by branch mispredictions
DTLB: Cycles stalled by TLB misses
L2M: Cycles stalled by L2 misses
L1M: Cycles stalled by L1 misses
FE: Remaining stalled cycles
nonRETIRED: miss-speculated work? Instructions that entered the RS (Reservation Stations) but did not retire – should be counted in CPI?
RETIRED: useful work

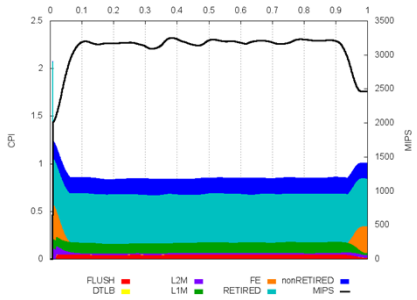
BSC tools @ PATC, Barcelona May 21-22 2012 25

CPI Breakdown: PMEMD (2 tasks)

Core2 model

(core2duo t9400 @ 2.533 GHz)

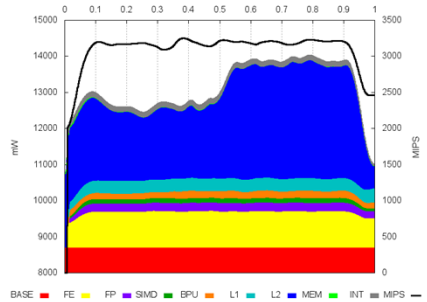
Intel Core2Duo CPI model evolution for pmemd.2tasks.x86.2.chop1.clustered.fused.extract.1.1.Cluster_1.0



Core2 power model

(core2duo t9400 @ 2.533 GHz)

Core2Duo Power consumption model evolution for pmemd.2tasks.x86.2.chop1.clustered.fused.extract.1.1.Cluste

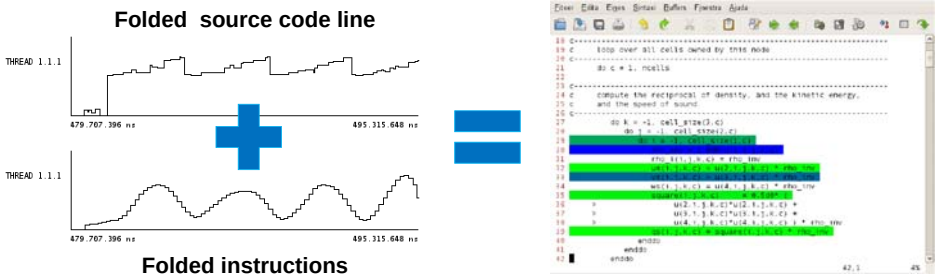


FLUSH: Cycles stalled by branch mispredictions
DTLB: Cycles stalled by TLB misses
L2M: Cycles stalled by L2 misses
L1M: Cycles stalled by L1 misses
FE: Remaining stalled cycles
nonRETIRED: miss-speculated work? Instructions that entered the RS (Reservation Stations) but did not retire – should be counted in CPI?
RETIRED: useful work

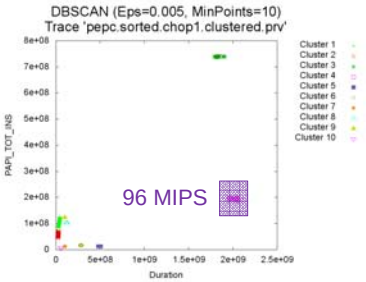
BASE: Static consumption
FE: Regular instruction consumption, front-end component
BPU: Branch Prediction Unit consumption
FP/SIMD/INT: Consumption for FP/SIMD/integer instructions
L1: L1 consumption
L2: L2 & Front-Side-Bus consumption
MEM: Memory consumption

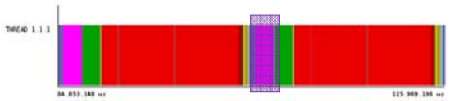
BSC tools @ PATC, Barcelona May 21-22 2012 26

Folding → profiles of rates

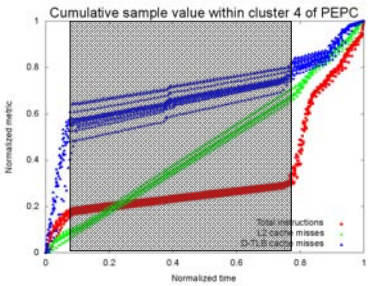


Example: PEPC





Performance metrics
16 MIPS
2.3 M L2 misses/s
0.1 M TLB misses/s



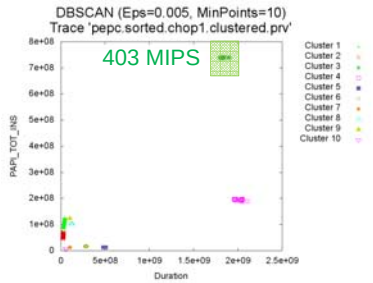

```
htable%node = 0
htable%key = 0
htable%link = -1
htable%leaves = 0
htable%childcode = 0
```

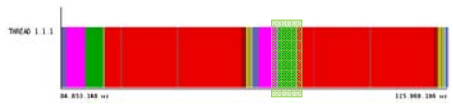
→

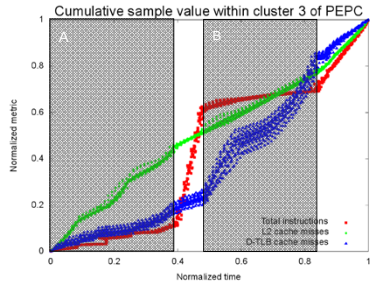
```
do i = 1, n
 htable(i)%node = 0
 htable(i)%key = 0
 htable(i)%link = -1
 htable(i)%leaves = 0
 htable(i)%childcode = 0
End do
```

Changes
-70% time
-18% instructions
-63% L2 misses
-78% TLB misses
253 MIPS (+163%)

Example: PEPC



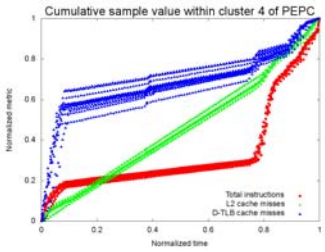


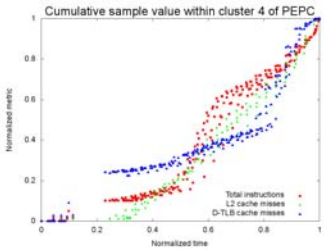


Performance metrics	
Region A	Region B
100 MIPS	80 MIPS
4 M L2 misses/s	2 M L2 misses/s
0.4 M TLB misses/s	1 M TLB misses/s

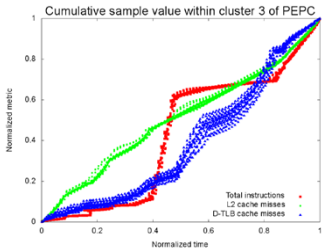
BSC tools @ PATC, Barcelona May 21-22 2012 29

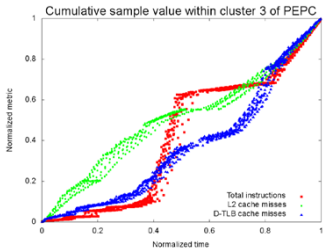
Example: PEPC (optimized)





Changes
-70% time
-18% instructions
-63% L2 misses
-78% TLB misses
253 MIPS (+163%)





Changes
-30% time
-1% instructions
-10% L2 misses
-32% TLB misses
544MIPS (+34%)

BSC tools @ PATC, Barcelona May 21-22 2012 30

INTELLECTUAL PROPERTY RIGHTS NOTICE:

- The User may only download, make and retain a **copy of the materials for his/her use for non-commercial** and research purposes.
- The User **may not commercially use the material**, unless has been granted prior written consent by the Licensor to do so; and **cannot remove, obscure or modify copyright notices**, text acknowledging or other means of identification or disclaimers as they appear.
- For further details, please contact **BSC-CNS** patc@bsc.es

PATC training, Barcelona, May 2012 ‹#›