www.bsc.es

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

BSC Tools

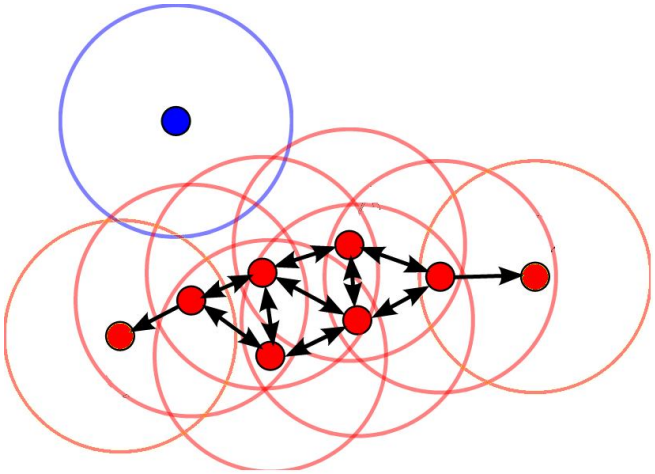
Detecting and analyzing application structure with clustering

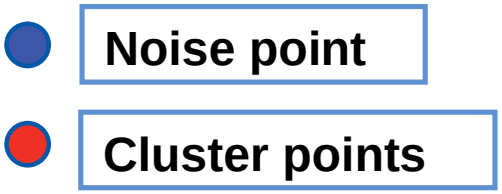
Judit Gimenez (judit@bsc.es)

Extreme-Scale Performance Tools
- SC12 workshop

Clustering

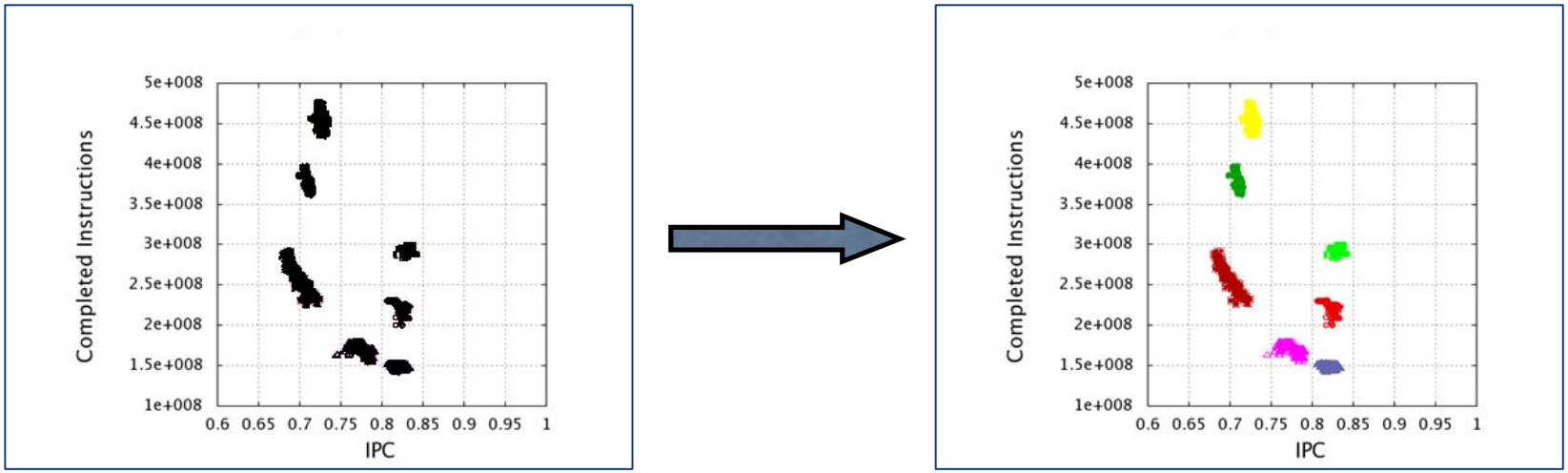
- Identification of computation structure
  - CPU burst = region between consecutive runtime calls
    - Described with performance hardware counters
    - Associated with call stack data
- Using DBSCAN density-cluster algorithm
  - Data not necessarily Gaussian
  - Two parameters: Epsilon (search radius), MinPoints





Clustering

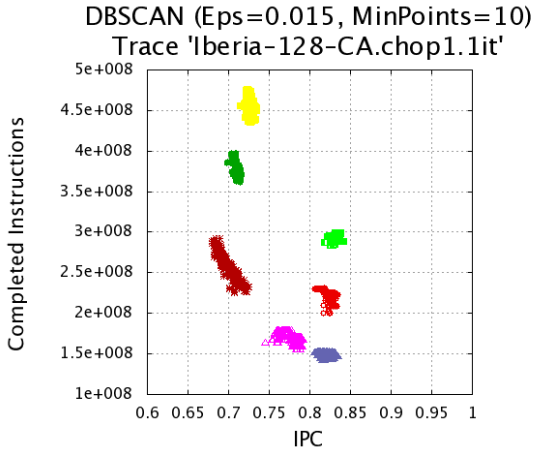
- Identification of computation structure
  - CPU burst = region between consecutive runtime calls
    - Described with performance hardware counters
    - Associated with call stack data
- Using DBSCAN density-cluster algorithm
  - Data not necessarily Gaussian
  - Two parameters: Epsilon (search radius), MinPoints



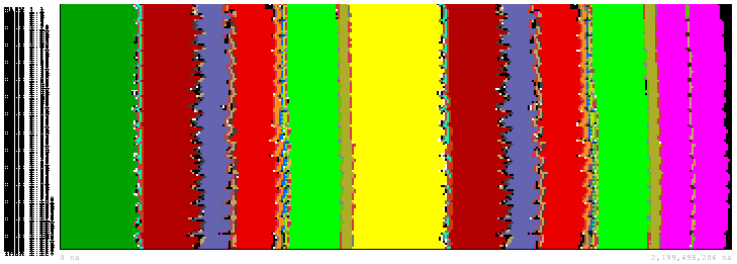
Automatic Detection of Parallel Applications Computation Phases. (IPDPS 2009)

Outputs

Scatter Plot of Clustering Metrics



Clusters Distribution Along Time



Cluster Statistics

Cluster	1	2	3	4	5	6
% Time	36.29	29.52	10.13	9.68	3.73	1.71
Avg. Burst Dur. (ms)	220.46	177.70	60.81	29.09	38.71	44.83
						.59
						8.24
L1M/kInstr	22.72	32.63	12.65	8.39	16.12	6.86
L2M/kInstr	0.59	1.23	1.08	0.61	1.23	1.73
Mem.BW (MB/s)	90.77	182.65	193.32	136.33	236.15	295.71

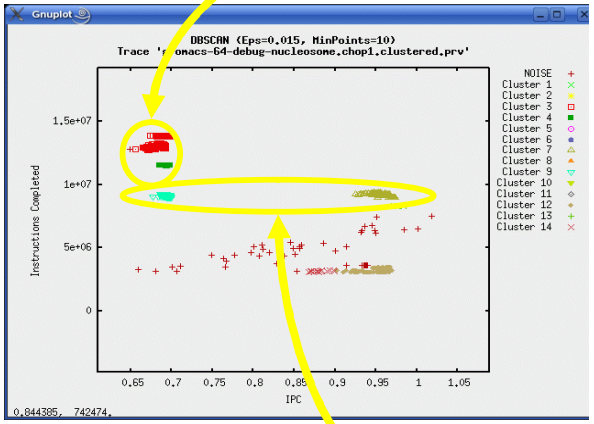
All counters in a single run

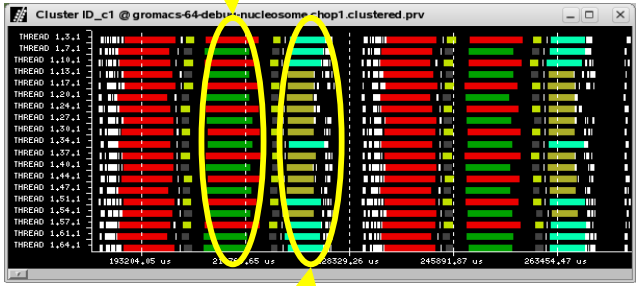
Code Linking

Cluster	Code Section
1	solve_nmm.f: [2037 - 2310]
2	solve_nmm.f: [1478 - 1782]
	solve_nmm.f: [2030 - 1782]
3	solve_nmm.f: [1241 - 1345]
4	solve_nmm.f: [2771 - 2865]
	solve_nmm.f: [2388 - 2489]
5	solve_nmm.f: [1478 - 1569]
6	solve_nmm.f: [1607 - 1633]

Using clusters to understand apps behavior (GROMACS)

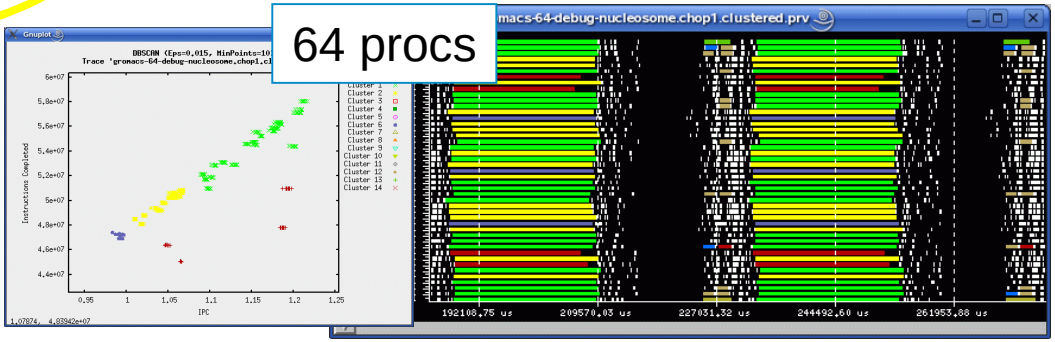
Instructions imbalance



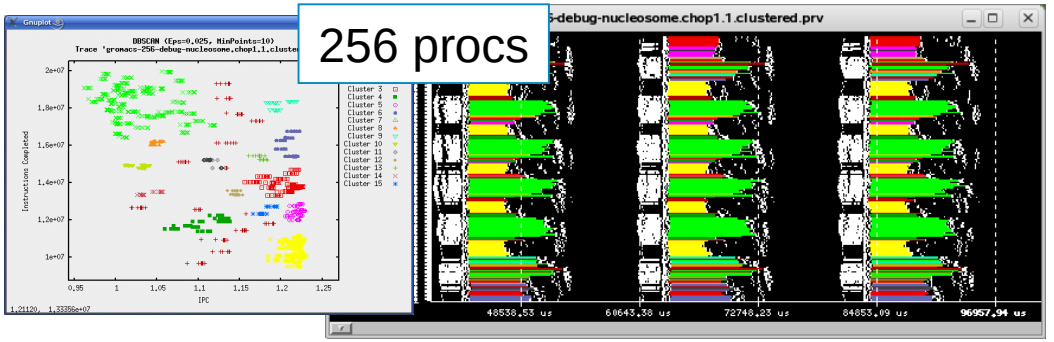


IPC Imbalance

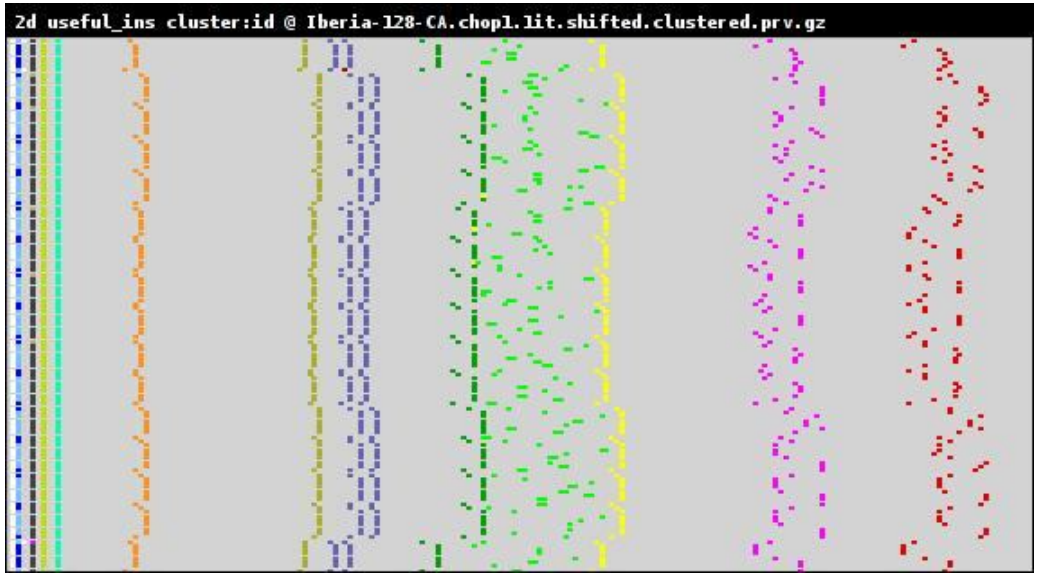
64 procs



256 procs

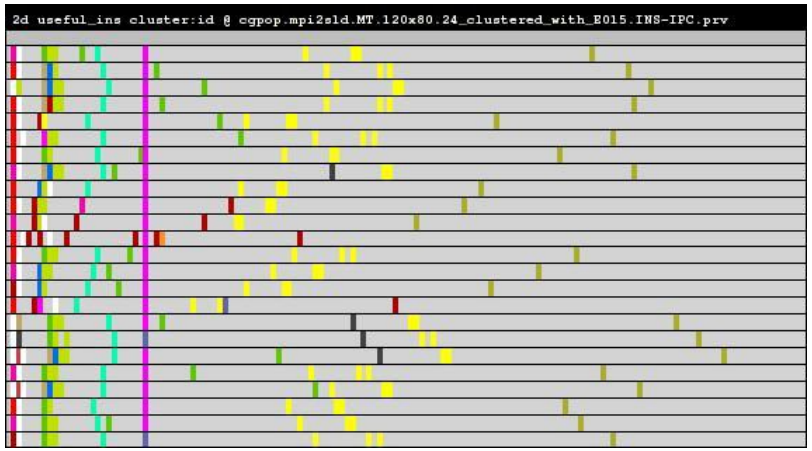


Identifying main code regions

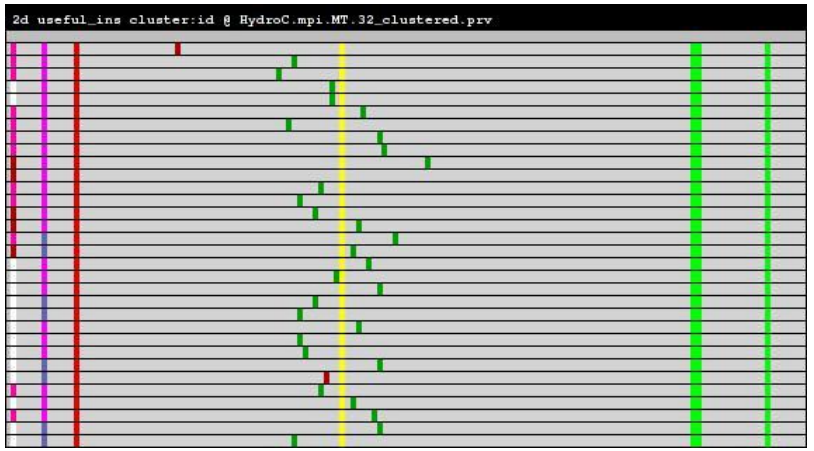


WRF

instr. vs. cluster

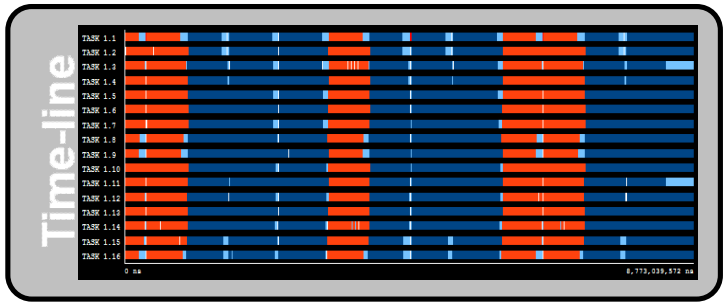


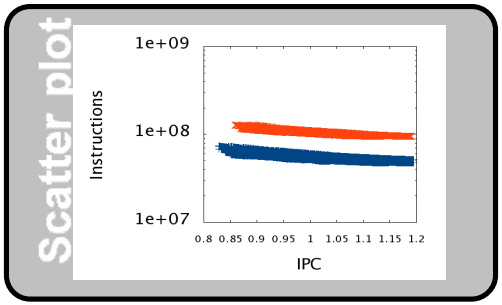
CG-POP

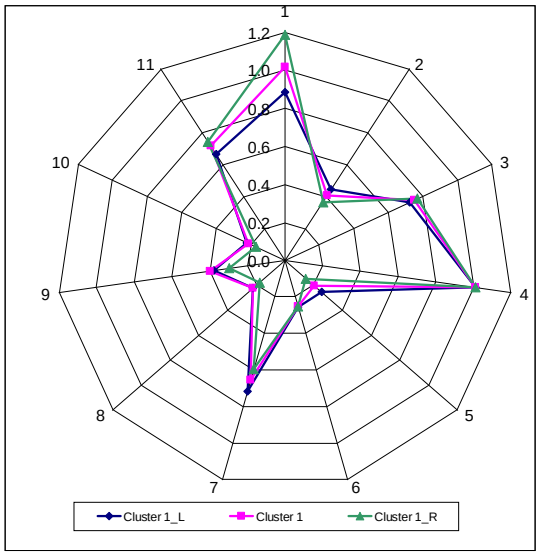


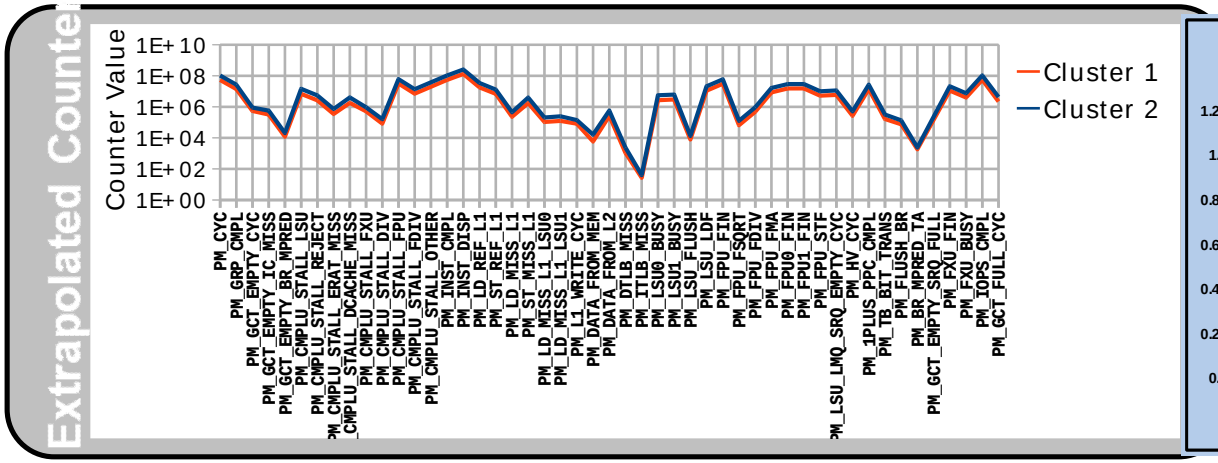
HydroC

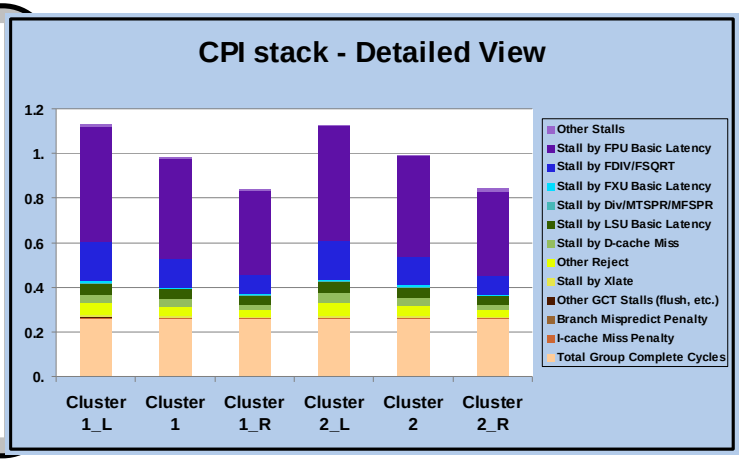
Hardware counters projection



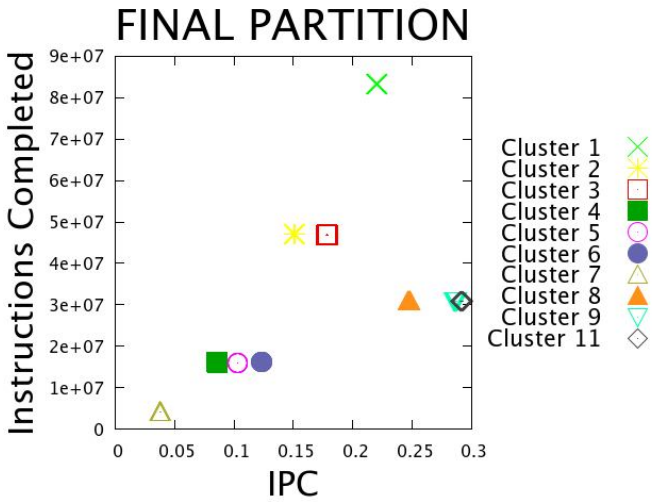


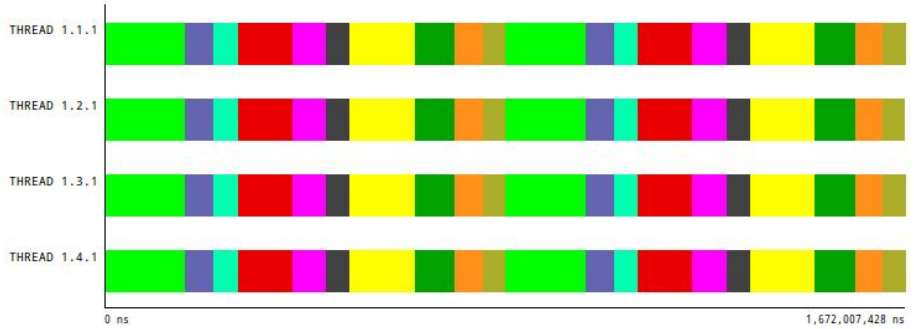


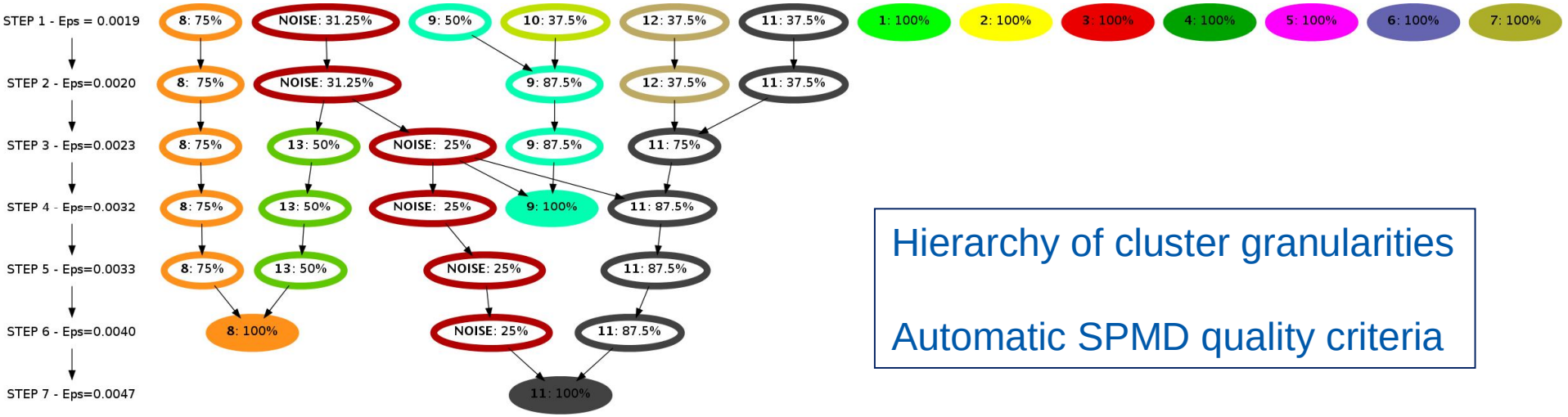




Quality driven clustering algorithms







Hierarchy of cluster granularities

Automatic SPMD quality criteria

Conclusions

- Performance analytics: Data analytics applied to raw performance data
  - From data to insight, huge room for research
- Clustering enables focusing the analysis and opens many different uses
  - Correlation of sampled data to generate instantaneous metric evolution
  - Separate speed factors per cluster on predictive simulations with Dimemas
  - Track application evolution using the clustering scatter plots

www.bsc.es/paraver