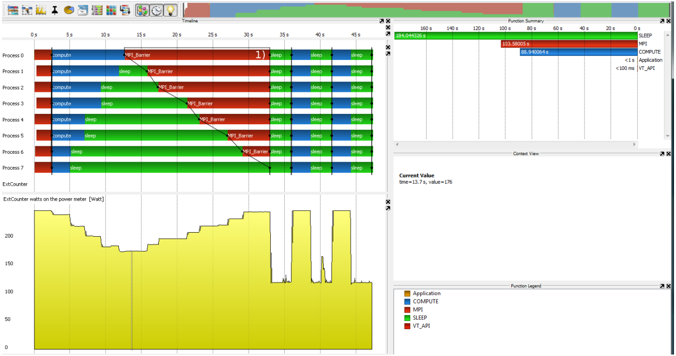
FORSCHUNGSZENTRUM JÜLICH

Scalasca Enhancements in the eeClust project

September 11, 2012 | Willi Homberg & Michael Knobloch

Motivation



MPI Busy-Waiting

Power consumption in phases of busy-waiting is very high due to constant CPU activity.

Outline

Wait-State Detection with Scalasca

Calculating Energy-Saving Potential

Examples

Conclusion

Outline

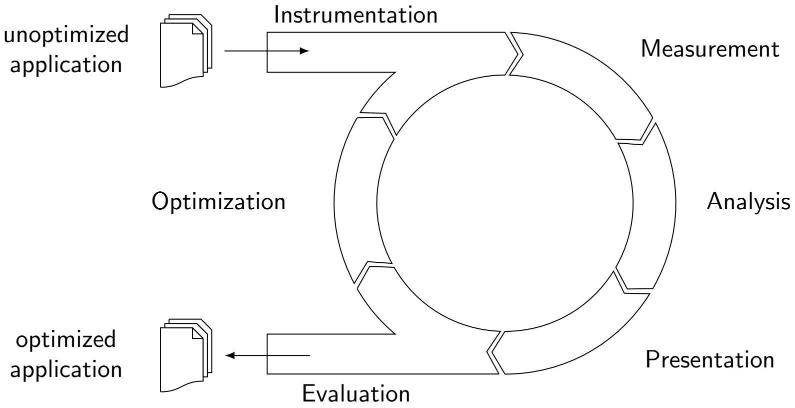
Wait-State Detection with Scalasca

Calculating Energy-Saving Potential

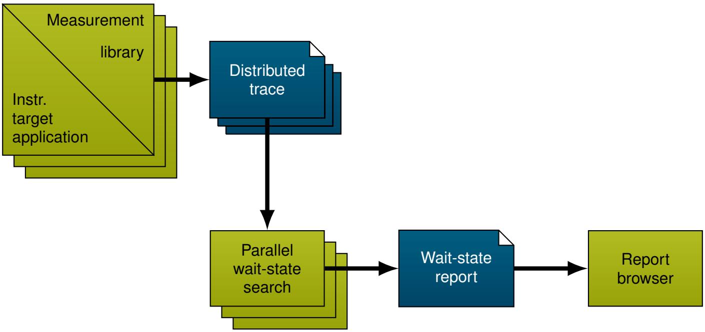
Examples

Conclusion

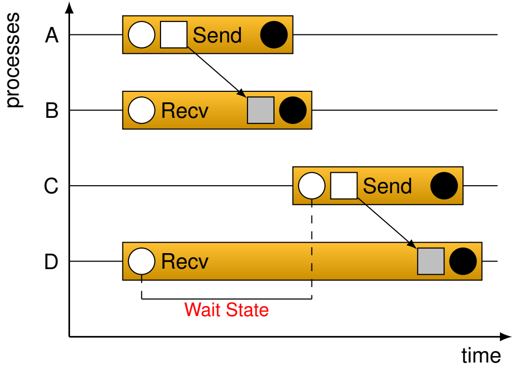
Performance Optimization Cycle



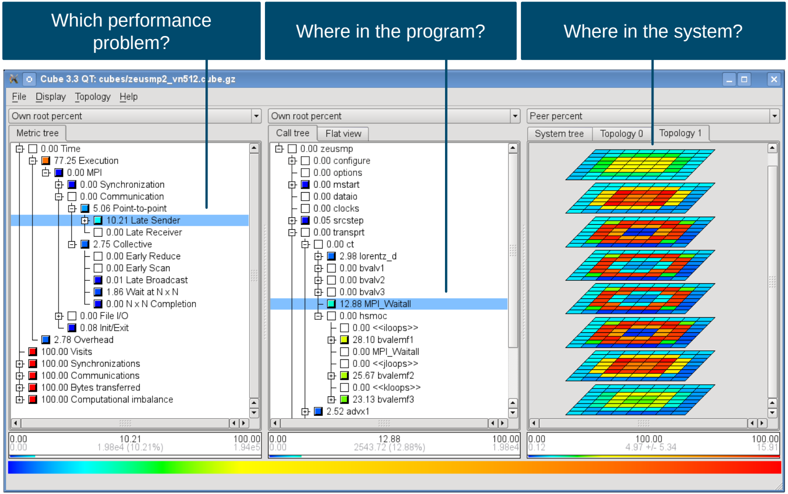
Scalasca Workflow



Wait-States



The Cube3 Analysis Report Browser



Outline

Wait-State Detection with Scalasca

Calculating Energy-Saving Potential

Examples

Conclusion

Calculating Energy-Saving Potential

Idle-Waiting

$$ESP = \max_{p \in PS}((t_w * A_{p_1}) - (t_w - t_{T_{p,p_1}}) * I_p + E_{T_{p,p_1}})$$

Busy-Waiting

$$ESP\_BW = \max_{p \in PS}((t_w * A_{p_1}) - (t_w - t_{T_{p,p_1}}) * A_p + E_{T_{p,p_1}})$$

PS – Set of power states
A_p – Active energy in P-State p
I_p – Idle energy in P-State p
t_w – Waiting time
$t_{T_{p_1,p_2}}$ – Transition time
$E_{T_{p,p_1}}$ – Transition energy

Outline

Wait-State Detection with Scalasca

Calculating Energy-Saving Potential

Examples

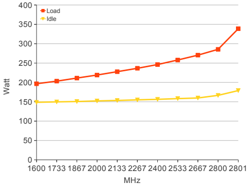
Conclusion

Test Systems – eeCluster

- 5 Nodes Intel Nehalem
- 5 Nodes AMD Opteron
- 2 LMG 450 Power Meter

Intel Nodes

- 2 Xeon X5560 (4 cores, SMT-2)
- 12 GB RAM
- Gigabit Ethernet



P-State	A_p (W)	I_p (W)	$t_{T_{p,p_1}}$ (s)	$E_{T_{p,p_1}}$ (J)
1 – 2800 MHz	35.68	20.81	0	0
2 – 2533 MHz	32.24	19.77	0.00001	0.1
3 – 2267 MHz	29.56	19.36	0.00002	0.2
4 – 1867 MHz	26.4	18.83	0.00003	0.4
5 – 1600 MHz	24.57	18.57	0.00004	0.8

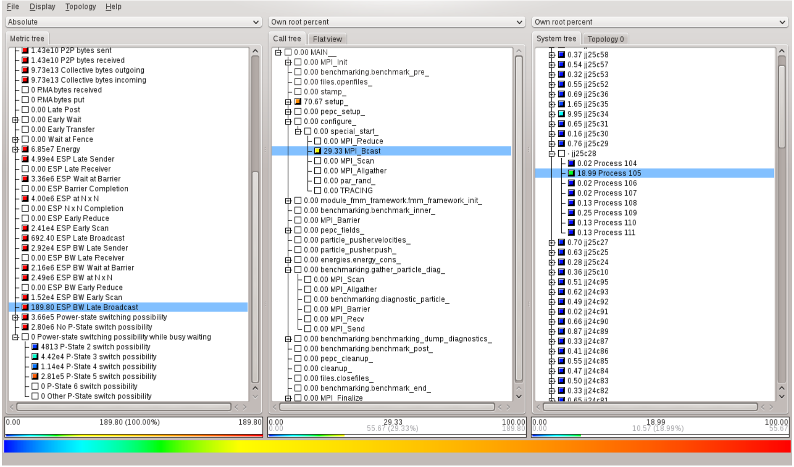
Test Systems – Juropa

- 2208 compute nodes
- 2 Xeon X5570 (4 cores, SMT-2)
- 24 GB RAM
- QDR Infiniband
- # 25 in Top500 (June 2011)
- 1.5 MW Power Consumption

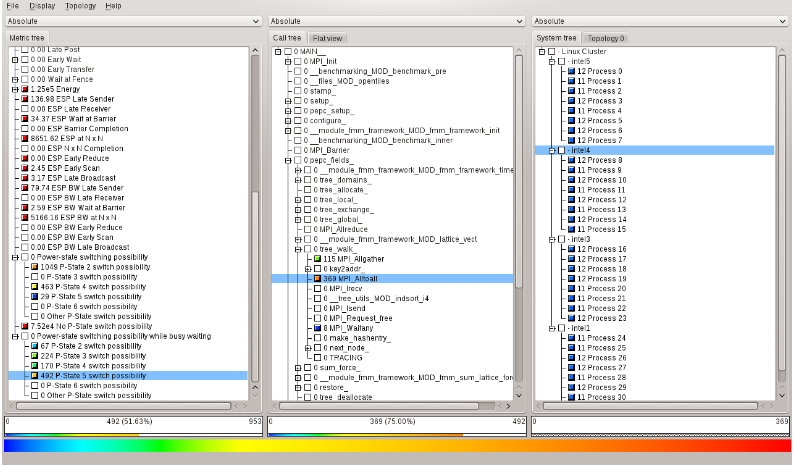
P-State	A_p (W)	I_p (W)	$t_{T_{p,p_1}}$ (s)	$E_{T_{p,p_1}}$ (J)
1	58.8	34.3	0	0
2	53.13	32.58	0.00001	0.1
3	48.72	31.91	0.00002	0.2
4	43.51	31.03	0.00003	0.4
5	40.48	30.6	0.00004	0.8

Values derived from Intel nodes of eeCluster

Juropa – 1024 processes



eeCluster Intel Nodes – 32 processes



Outline

Wait-State Detection with Scalasca

Calculating Energy-Saving Potential

Examples

Conclusion

Conclusion

- MPI Busy-Waiting consumes considerable amount of energy
- Scalasca detects wait-states in large-scale parallel programs
  - Indicate load-balancing problems
  - Cannot always be prevented
- We extended Scalasca to calculate the energy-saving potential in such wait-states and hint the optimal power-state

scalasca

www.scalasca.org
scalasca@fz-juelich.de

eeClust
Energy-Efficient Cluster Computing

www.eeclust.de

SPONSORED BY THE
Federal Ministry
of Education
and Research