VI-HPS

Virtual Institute – High Productivity Supercomputing

6 February 2013

Brian Wylie
Jülich Supercomputing Centre
b.wylie@fz-juelich.de

Virtual Institute – High Productivity Supercomputing

VI-HPS

Goal: Improve the quality and accelerate the development process of complex simulation codes running on highly-parallel computer systems

- Start-up funding (2006-2011) by Helmholtz Association of German Research Centres

- Activities
  - Development and integration of HPC programming tools
    - Correctness checking & performance analysis
  - Training workshops
  - Service
    - Support email lists
    - Application engagement
  - Academic workshops

www.vi-hps.org

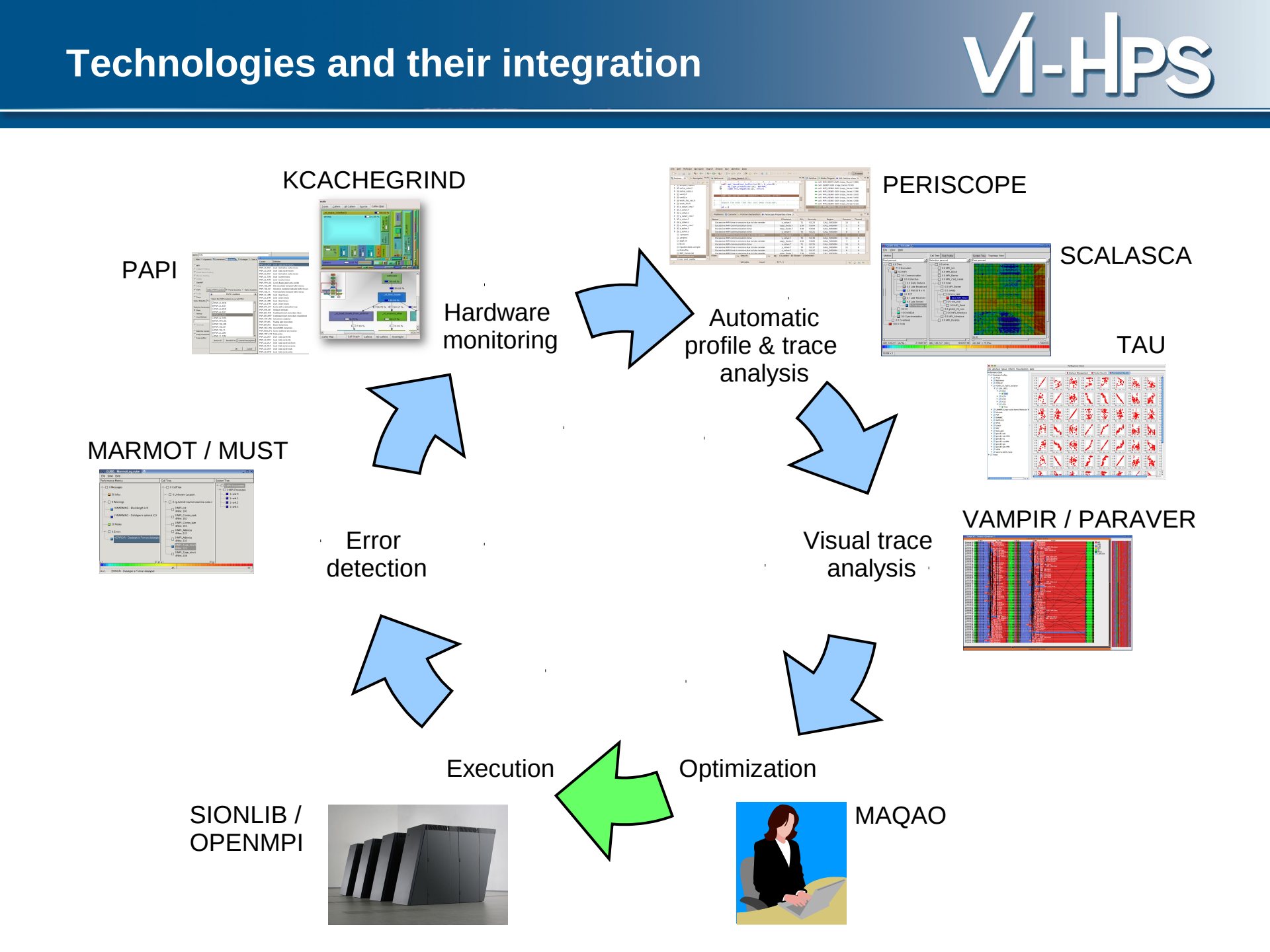
VI-HPS productivity tools

- Marmot/MUST
  - MPI correctness checking
- PAPI
  - Interfacing to hardware performance counters
- Periscope
  - Automatic analysis driven by on-line distributed search
- Scalasca
  - Large-scale parallel performance analysis
- TAU
  - Integrated parallel performance system
- Vampir/VampirTrace
  - Event tracing and graphical trace visualization & analysis
- Score-P
  - Common instrumentation & measurement infrastructure

VI-HPS productivity tools (cont.)

- KCachegrind
  - Callgraph-based cache analysis [x86 only]
- MAQAO
  - Assembly instrumentation & optimization [x86 only]
- mpiP/mpiPview
  - MPI profiling tool and analysis viewer
- ompP
  - OpenMP profiling tool
- OpenMPI
  - Memory checking
- Open|SpeedShop
  - Integrated parallel performance analysis environment
- Paraver/Extrae
  - Event tracing and graphical trace visualization & analysis

Technologies and their integration



VI-HPS Training & Tuning Workshops

- Goals
  - Give an overview of the programming tools suite
  - Explain the functionality of individual tools
  - Teach how to use the tools effectively
  - Offer hands-on experience and expert assistance using tools
  - Receive feedback from users to guide future development
- For best results, bring & analyse/tune your own code(s)!

- VI-HPS Tutorial series
  - SC'08, ICCS'09, SC'09, Cluster'10, SC'10, SC'11, EuroMPI'12
- VI-HPS Tuning Workshop series
  - 2008 (Aachen & Dresden), 2009 (Jülich & Bremen), 2010 (Garching & Amsterdam/NL), 2011 (Stuttgart & Aachen) 2012 (St-Quentin-en-Yvelines/F & Garching)

3rd VI-HPS Tuning Workshop (JSC)

- 16-20 Feb 2009, Jülich
  - *Jugene* IBM BlueGene/P + *JUMP* p5-575 AIX cluster + RWTH Intel Linux cluster + Sun Niagara2 + SGI Altix
  - Scalasca, Vampir, PAPI, Marmot
  - 28 participants
    - (D, BG, DK, UK)
  - mostly MPI codes, but several MPI+OpenMP
  - scaling runs analysed with up to 16k processes
  - participants greatly appreciated advance access to workshop systems to install and prepare their application codes

9th VI-HPS Tuning Workshop (UVSQ)

- 23-27 Apr 2012, St-Quentin/France
  - TGCC *Curie* Bullx x86_64 Linux cluster
  - Scalasca, Vampir, TAU, Periscope, Kcachegrind, MAQAO
  - + Likwid, MPC & VTune
  - 35 participants
    - (F, B, BG, E, I, IRL, N)

  - mostly MPI+OpenMP application codes, few MPI only
  - “The workshop was quite a boost for me in my work to optimize [our] main application, which is a MPI+OpenMP parallel code.”

Upcoming VI-HPS training events

- 11th VI-HPS Tuning Workshop (22-25 Apr 2013)
  - PRACE ATC event hosted by MdS, Saclay (S. Paris), France
  - using Intel x86_64 Linux cluster (and perhaps TGCC *Curie*)
  - Score-P, Scalasca, Vampir, TAU, MAQAO, ...
- Further events to be determined
  - (one-day) tutorials
    - with guided exercises usually using Live DVD
  - (multi-day) training workshops
    - with your own applications on actual HPC systems

Check www.vi-hps.org/training for announcements

- Contact us if you might be interested in hosting an event

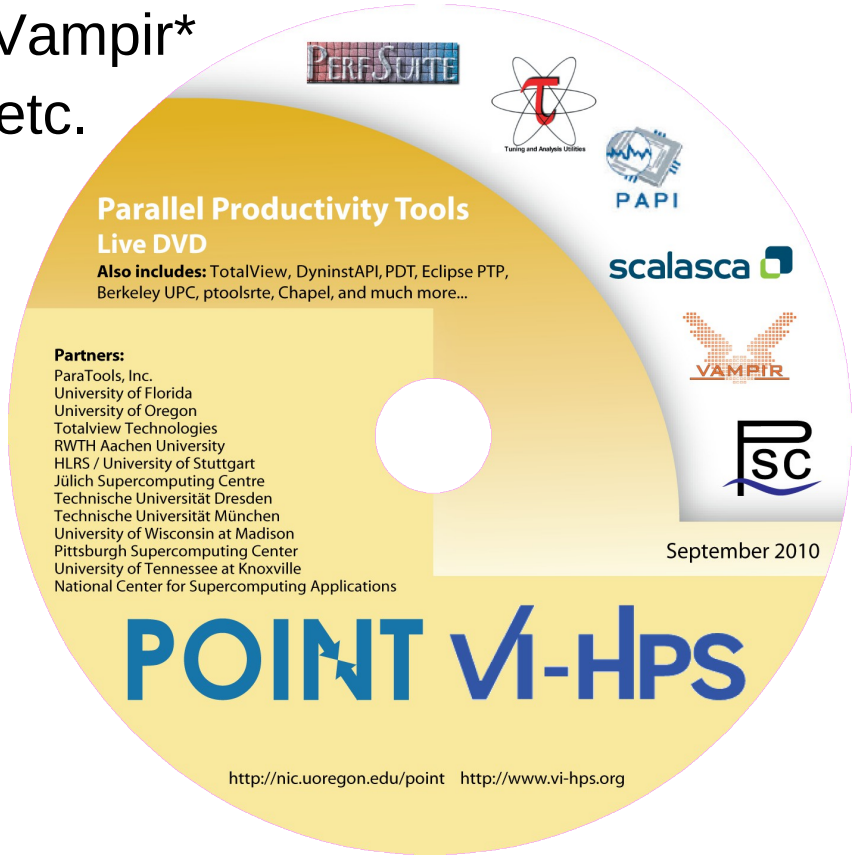
VI-HPS training events in planning

- 12th VI-HPS Tuning Workshop (Sept/Oct 2013)
  - PRACE ATC event hosted by JSC/GCS, Jülich, Germany
  - using *Juqueen* BlueGene/Q
  - Score-P, Scalasca, Vampir, TAU, Periscope, …
  - lots of time for coaching with participants' own applications

- TBD
  - 4 or 5 days?
  - introductory and/or more advanced level?
  - include debugging/correctness tools?
  - classroom workstations or participants' notebook computers?
  - only BG/Q or also include Linux cluster (e.g., *Juropa*)?
  - large-scale runs?

VI-HPS Live-ISO/DVD

- Bootable Linux installation ISO (on DVD or USB drive)
- Includes everything needed to try out our parallel tools on an x86-architecture notebook computer
  - VI-HPS tools: KCachegrind, Marmot, PAPI, Periscope, Scalasca, TAU, VT/Vampir*
  - Also: Eclipse/PTP, TotalView*, etc.
    - * time/capability-limited evaluation licences provided for commercial products
  - GCC (w/ OpenMP), OpenMPI
  - Manuals/User Guides
  - Tutorial exercises & examples
- Produced by U. Oregon PRL
  - Sameer Shende



VI-HPS Live-ISO/DVD

- ISO image approximately 4GB
  - distributed on DVD or USB drive
  - or download from VI-HPS website
- Boot directly from disk
  - enables hardware counter access and offers best performance
- Boot within virtual machine
  - faster boot time and can save/resume state, but no hardware counter access
- Boots into Linux environment
  - supports building and running provided MPI and/or OpenMP parallel application codes
  - and experimentation with VI-HPS (and other) tools

JÜLICH
FORSCHUNGSZENTRUM

BlueGene Extreme Scaling Workshops

2012-12-06 |

Wolfgang Frings, Bernd Mohr & **Brian J. N. Wylie**
Jülich Supercomputing Centre

b.wylie@fz-juelich.de

BlueGene Extreme Scaling Workshops

JÜLICH FORSCHUNGSZENTRUM

Workshops held by JSC in Jülich in 2009, 2010 & 2011

- organized by Wolfgang Frings & Bernd Mohr
- scaling assistance and tuning advice from JSC & IBM experts
- 3 day access to entire *Jugene* BlueGene/P to run scaling tests
  - on-the-fly daytime scheduling
- 8-10 teams of application developers
  - selected from open competitive call
  - demonstrate reasonable scaling on BG to 100,000+ cores
  - expected to execute within available compute node memory
  - I/O requirements not excessive
- report required from each team documenting achievements: successes and difficulties/limitations

2009 Jülich BlueGene/P Extreme Scaling Workshop

JÜLICH FORSCHUNGSZENTRUM

- 26-28 October 2009
- 10 participating teams from
  Harvard University, MIT, ANL,
  Rensselaer Polytechnic Institute,
  University of Edinburgh,
  Swiss Fed. Inst. of Technology,
  Instituto Superior Técnico (Lisbon)
  RZG MPG, DESY Zeuthen
  and JSC
- 398 jobs with 135.6 rack days (out of 169.3 rack days provided): 80% utilization during workshop!
- All but 1 team succeeded in executing their code on full machine (294,912 cores)
- FZJ-JSC-IB-2010-02: http://juser.fz-juelich.de/record/8924

2009 Scaling Workshop Applications

JÜLICH FORSCHUNGSZENTRUM

- QCD, EPCC, UK
- Gyrokinetic turbulence simulation, RZG/IPP, DE
- Neutron transport simulations, ANL, US
  - **Gordon Bell Finalist**
- Astrophysics (particle-in-cell), UTL, PT
- Simulation of coronary arteries, EPFL, CH
- Parallel Evolutionary Biology Suite, Harvard Uni., US
- Adaptive computational fluid dynamics, RPI, US
  - **Gordon Bell Finalist**
- Mesoscopic particle dynamics, JSC, DE
- QCD, Hungary/France/German HMC Collaboration
- QCD, DESY/Bonn Uni., DE

2010 Jülich BlueGene/P Extreme Scaling Workshop

JÜLICH FORSCHUNGSZENTRUM

- 22-24 March 2010
- 10 participating teams from Harvard University, ANL, CORIA (France), ETH Zurich, KIT, LRZ, ZIB, JSC, and the Universities of Chemnitz, Erlangen and Marburg
- 392 jobs with 138.7 rack days (out of 164 rack days provided): 84% utilization
- 6 teams succeeded in executing their code on full machine (including team that failed 2009)
- 3 teams could “only” scale to 64 racks (262,144 cores) due to the need to run with a power-of-two number of processes
- 1 team got stuck at 32 racks (131,072 cores) due to program bug
- FZJ-JSC-IB-2010-03: http://juser.fz-juelich.de/record/9600

2010 Scaling Workshop Applications

JÜLICH FORSCHUNGSZENTRUM

- Particle flow simulation, Erlangen Uni., DE
- Simulation of fluid flow and mass transport, Marburg Uni., DE
- Turbulent flows in complex geometries, CORIA, FR
- Spectral element code, ETH, CH / ANL, US
- Simulation of coronary arteries, Harvard/EPFL
  - **Gordon Bell Finalist**
  - **George Michael Memorial PhD Fellowship (A. Peters)**
- Parallel Fast Fourier Transform, Chemnitz Uni., DE
- QCD, ZIB/LRZ, DE
- Mesoscopic particle dynamics, JSC, DE
- Hydrodynamic turbulence, KIT, DE
- Large-scale density-functional calculations, JSC, DE

2011 Jülich BlueGene/P Extreme Scaling Workshop

JÜLICH FORSCHUNGSZENTRUM

- 14-16 February 2011
- 8 participating teams from KTH (SE), KAUST (SA), Princeton PPL, RZG MPG, University of Heidelberg, Mickiewicz University (PL), University College London (UK), Euskal Herriko Unibertsitatea (ES)

- Selected from 15 proposals
- 308 jobs with 122 rack days (out of 157 rack days provided): 77% utilization
- Teams succeeded in executing 11(!) codes on full machine
- FZJ-JSC-IB-2011-02: http://juser.fz-juelich.de/record/15866

2011 Scaling Workshop Applications

JÜLICH FORSCHUNGSZENTRUM

- 2 x Neuronal network simulations, KTH, SE
- Molecular dynamics (BBMD), KAUST, SA
- Gyrokinetic PIC simulations (GTC-P), Princeton PPL, US
- Eigenvalue solver (ELPA) +
  Ab-initio molecular simulation, RZG MPG, DE
- Molecular nanomagnets (QTM), Mickiewicz University, PL
- Lattice Boltzmann (LB3D), University College London, UK
- Time-dependent density functional theory (Octopus),
  Euskal Herriko Unibertsitatea, ES
- Algebraic Multigrid Solver (Muphi/DUNE-ISTL),
  University of Heidelberg, DE

Lessons from Scaling Workshops

JÜLICH FORSCHUNGSZENTRUM

Applications from a wide range of subject areas successfully scale to very large numbers of processes (O(300k))

- Weak scaling is easiest, but good strong scaling is possible
- Most employ MPI, often combined with OpenMP threading (and other hybrid parallelizations)
- Message-passing needs to be local (close neighbours) and appropriate placement (rank-reordering) may be necessary
- Asynchronous communication is generally beneficial
  - *Even with limited overlap of computation*
- Implicit collective communication synchronizations and inherent computation/communication imbalances grow to dominate at scale
- Effective parallel file I/O is critical

Each doubling of scale exposes a new bottleneck or bug!

Future Extreme Scaling Workshops

JÜLICH FORSCHUNGSZENTRUM

2013 BlueGene Extreme Scaling Workshop?

- October?
- require demonstrated scalability to 128k processes/threads and potential to run with 1 million and more?