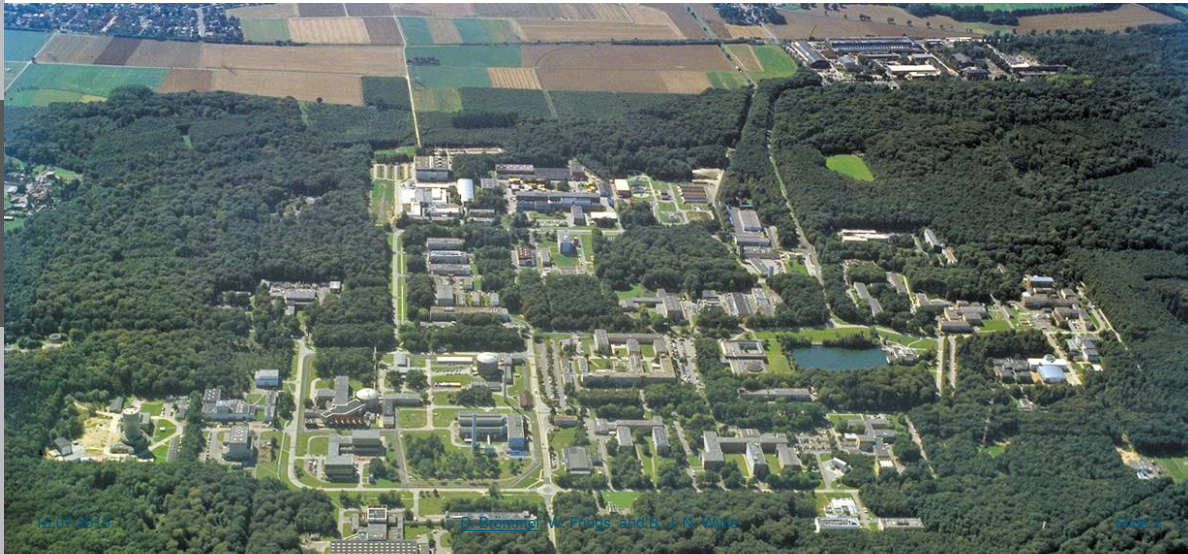
JÜLICH
FORSCHUNGSZENTRUM

Extreme-Scaling Applications 24/7 on JUQUEEN

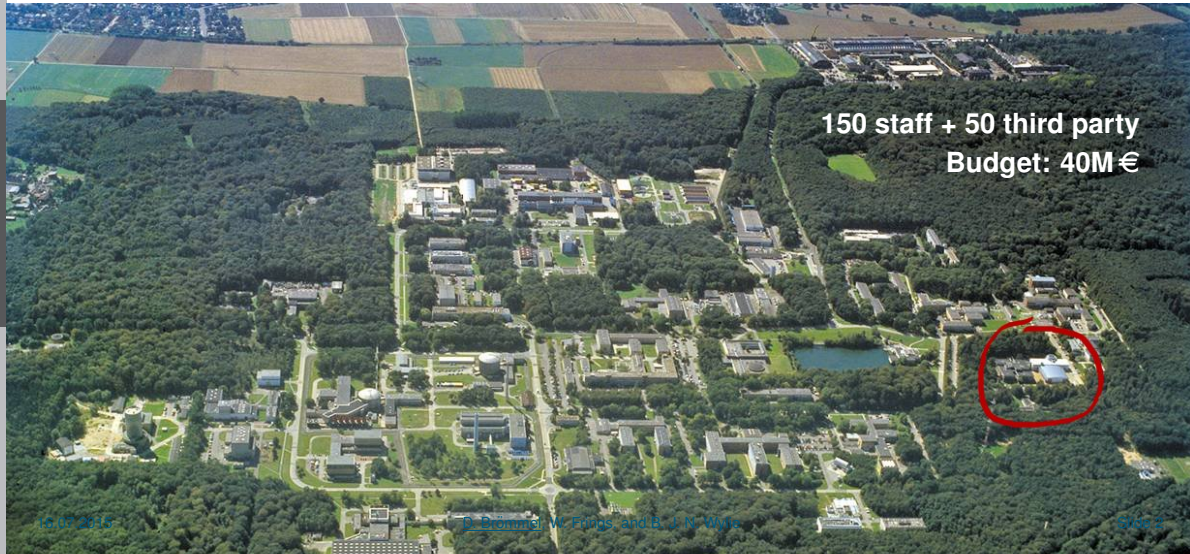
The High-Q Club and our workshops: Exascale enablers?

16.07.2015 | D. Brömmel, W. Frings, and B. J. N. Wylie | Jülich Supercomputing Centre | Forschungszentrum Jülich GmbH

Forschungszentrum Jülich and JSC

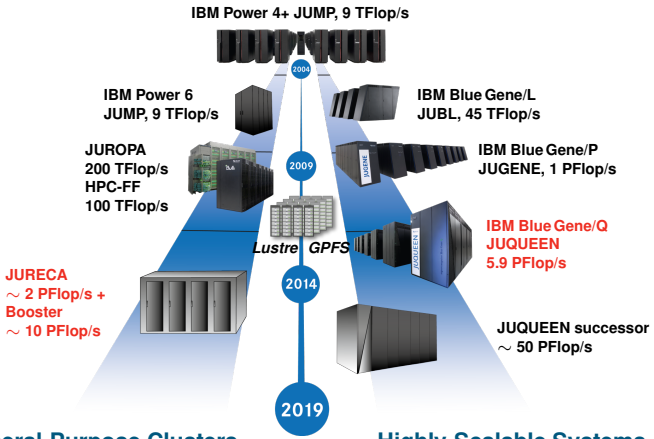


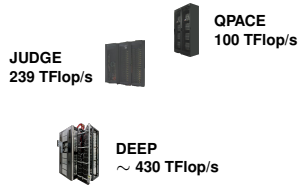
Forschungszentrum Jülich and JSC



Forschungszentrum Jülich and JSC

Computer systems at JSC

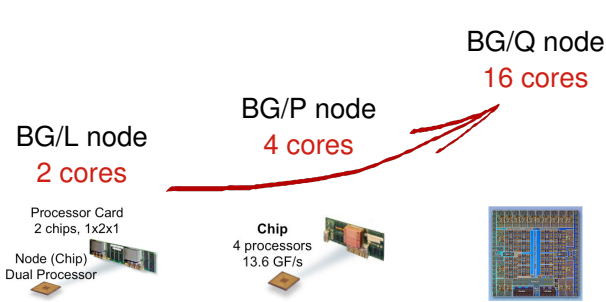


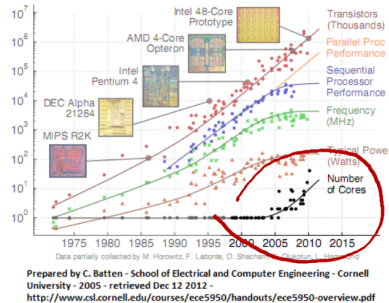


Motivation

The trend towards much higher core counts seems inevitable

$\longrightarrow$ Users need to adapt their programming strategies





Idea

Start a collection of codes to showcase running on all 28 racks of Blue Gene/Q at JSC, effectively using all 458 752 cores with up to 1.8M hardware threads

$\longrightarrow$ Promote the idea of exascale capability computing

$\longrightarrow$ Spark interest in tuning and scaling codes

Goal

- Encourage our users to try and reach exascale readiness
- Establish milestones in application development towards future systems
- Identify and understand bottlenecks in trying to reach millions of threads/processes and learn how to transition to exascale systems

Idea

Start a collecti… Blue Gene/Q at JSC, effecti…

Goal

- Enc…
- Estab… future systems
- Identify and … millions of threads/processes … scale systems

The High-Q Club

Current status of the High-Q Club

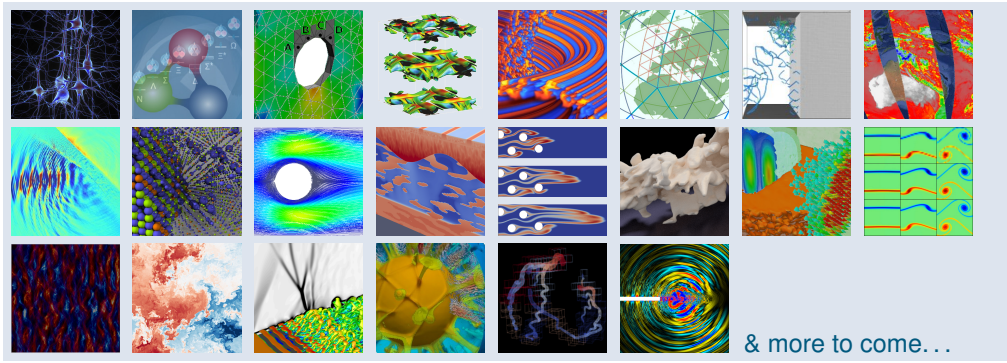
Diverse membership of 24 codes from fundamental physics, neuroscience, plasma physics, molecular dynamics, engineering and climate and earth science.

10 codes

7 codes

7 codes

Current status of the High-Q Club



CoreNeuron, dynQCD, FE2TI, FEMPAR, Gysela, ICON, IMD, JURASSIC, JuSPIC, KKRnano, MP2C, $\mu\varphi$ (muPhi), Musubi, NEST, OpenTBL, PEPC, PMG+PFASST, PP-Code, psOpen, SHOCK, Terra-Neo, waLBerla, ZFS

Becoming a member

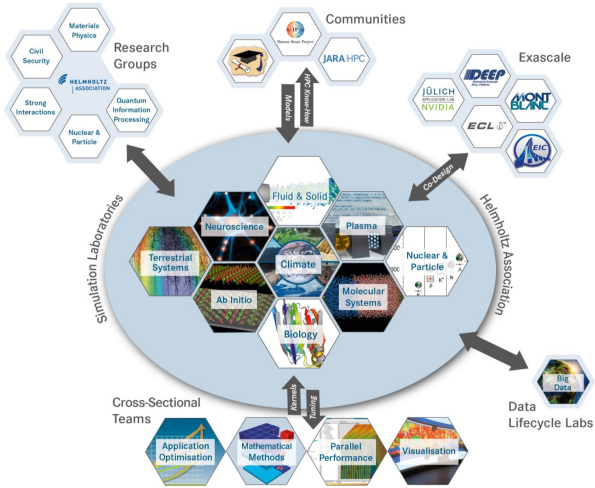
- Wide range of applications $\longrightarrow$ no common set of criteria
- Selection criteria are flexible (open for discussion!)
  - We try to collect as much information as possible (not all is made public)
  - Discussions with developers and within JSC
- Run a non-trivial example, ideally very close to production runs
- Submit evidence of strong and/or weak scalability to all available cores
- Preference on multi-threading (at least use HWTs)
- Include I/O if possible
- Possibly provide peak performance numbers

Related activities

Established support levels at JSC provide help scaling application codes.
This includes:

- Application support (initial contact point)
- Cross-sectional teams (Performance Analysis and Mathematical Modelling)
- Simulation Laboratories (part of Computational Science Division at JSC)

Related activities



A continuous
24/7 effort

Related activities

Established support levels at JSC provide help scaling application codes.
This includes:

- Application support (initial contact point)
- Cross-sectional teams (Performance Analysis and Mathematical Modelling)
- Simulation Laboratories (part of Computational Science Division at JSC)

In addition:
Workshops on Porting and Tuning on JUQUEEN and Extreme Scaling on JUQUEEN with dedicated or even exclusive access to the system and direct support during hands-on sessions.

Extreme Scaling Workshop on JUQUEEN

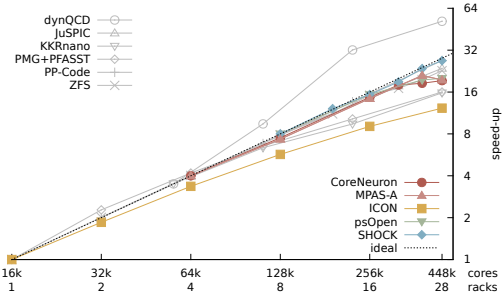
Extreme Scaling 24/7

- This latest edition of Extreme Scaling Workshops invited 7 applications teams and was extremely successful: all teams had their codes running on the full system within 24 hours.
- The workshop provided exclusive access to JUQUEEN with close support by JSC Simulation Laboratories for Climate Science, Fluids & Solids Engineering and Neuroscience assisted the code-teams, along with JSC Cross-sectional Teams, JUQUEEN and IBM technical support.
- 5 new codes entered the High-Q Club as a result.
- A detailed report with user contributions is available as technical report FZJ-JSC-IB-2015-01 http://juser.fz-juelich.de/record/188191

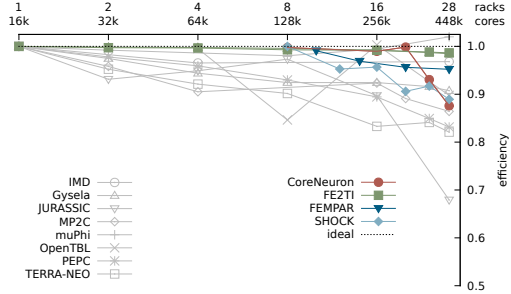
Extreme Scaling Workshop on JUQUEEN

Scaling results

Strong scaling

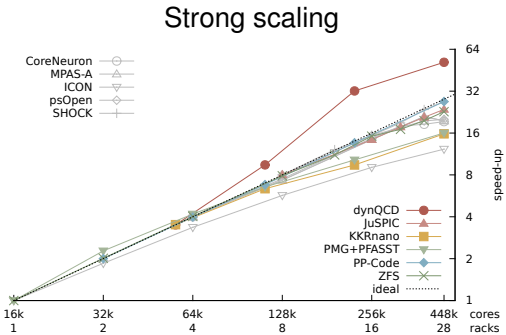


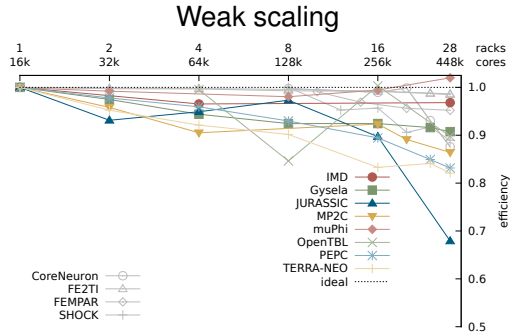
Weak scaling



Remaining High-Q codes

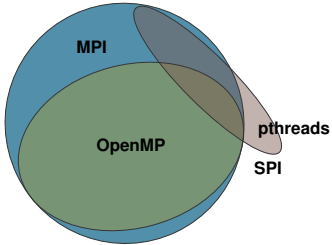
Scaling results



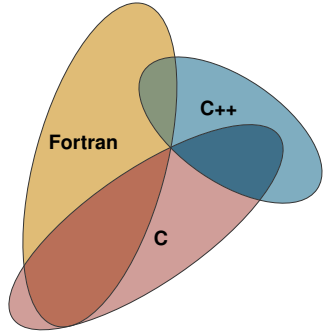


Statistics

Programming models

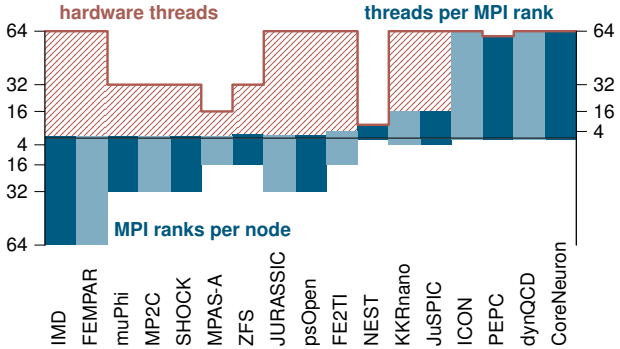


Programming languages



Venn diagrams with areas proportional to absolute numbers.

Node concurrency



Number of MPI ranks per node and threads per MPI rank

Either side of the diagram use all 64 hardware threads, purely with MPI or with threading

High-Q Club lessons – I

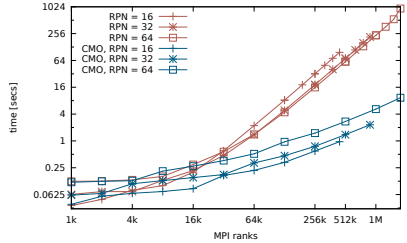
Wide range of HPC applications have demonstrated excellent scalability, generally with only modest tuning effort

- Over-subscription of cores delivers important efficiency benefits
  - Use vectorisation/SIMDization & libraries for node performance

High-Q Club lessons – I

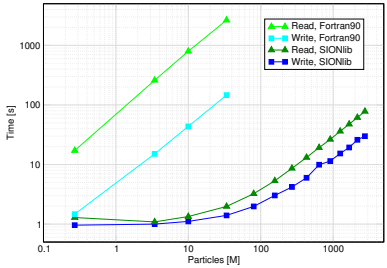
Wide range of HPC applications have demonstrated excellent scalability, generally with only modest tuning effort

- Over-subscription of cores delivers important efficiency benefits
  - Use vectorisation/SIMDization & libraries for node performance
- Standard languages and MPI+multi-threading are sufficient
  - MPI-only also possible but only 256MB available per rank
  - MPI communicator management gets increasingly costly

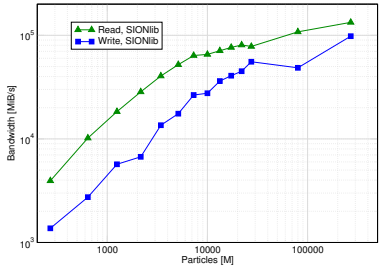


High-Q Club lessons – II

- File I/O remains the most common impediment to scalability
  - Effective solutions need to be employed, such as SIONlib
  - 11 (24) codes use parallel I/O, 5 (24) use SIONlib



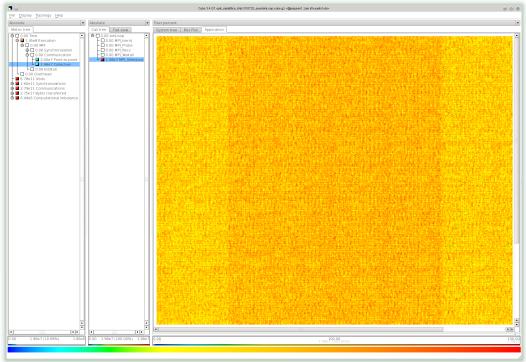
MP2C on one midplane



MP2C on 28 racks

High-Q Club lessons – III

- Scalable performance tools such as Scalasca can locate bottlenecks and identify opportunities for communication/synchronisation optimisation



A Scalasca profile on 1.3M MPI processes reveals imbalance in `MPI_Allreduce` call

High-Q Club discussion points

- Scaling on BG/Q also delivers benefits for other HPC computer systems
  - Enforce more disruptive changes?
- Membership criteria
  - Scaling to all cores? 90%?
  - Lower limit on efficiency?
- Metrics considered
  - Flops?
  - Limit user control and run ourselves?
- More focus on I/O or other features?
- Ranking of Codes?
- Can we draw conclusions for other systems?

Summary

- High-Q sparked interest: they now come to us
  $\longrightarrow$ currently 24 codes listed, 2 more accepted
- Hopefully enable our users to transition from peta to exascale
- Identified bottlenecks, solutions to common issues at hand
- So far no disruptive changes necessary or chosen

- Browse the High-Q Club webpages:
  `http://www.fz-juelich.de/ias/jsc/high-q-club`
- Download our technical report: FZJ-JSC-IB-2015-01
  `http://juser.fz-juelich.de/record/188191`