Mitglied der Helmholtz-Gemeinschaft

JÜLICH
FORSCHUNGSZENTRUM

Enabling Applications for Accelerated Architectures

D. Pleiter | ADAC5 Workshop | Tokyo | 16 February 2018

Accelerated Architectures at JSC

GPU accelerated

System	Level of acceleration	Node architecture
JUDGE (… - 2016)	206/206	2x Xeon X5650 2x M2050/M2070
JURECA (2015 - 2020)	75/1872	2x Xeon E5-2680 v3 2x K80
JUWELS (2018 - ...)	~50 / ~2500	2x Xeon Platinum 8168 4x V100
JUPP (2015 - ...)	4/4	2x POWER8 2x K40
JURON (2016 - 2019)	18/18	2x POWER8+ 4x P100

FPGA accelerated

JUMAX (2017 - ...)	1/1	2x AMD 7601 EPYC 8x MAX5C

Industry Collaboration

NVIDIA Application Lab at Jülich

- Collaboration FZJ+NVIDIA
- Since 2012

POWER Acceleration and Design Center

- Collaboration FZJ+NVIDIA+IBM
- Since 2015

Approach

„Business model“

- Application owner: enabled, optimised application
- Lab: architectural insides and knowledge

Workflow template

- Anamnesis
  - Identification of performance potential
  - Exploration of programming models
  - Definition of expected outcomes
- Implementation
  - Lab
    - Definition of enabling strategy
    - POC implementations
    - Performance analysis
  - Application owner
    - Application modifications

Application B-CALM

B-CALM = Belgium-CAlifornia Light Machine

Base numerical method FDTD

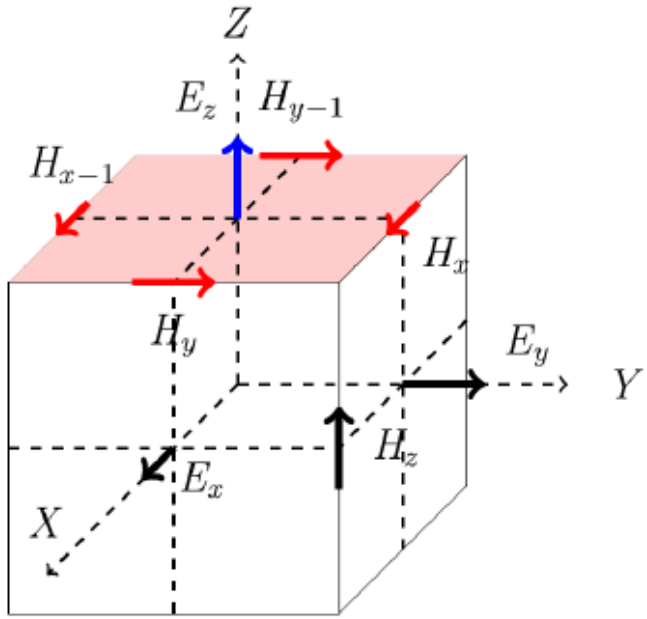
- Finite Difference Time Domain
- Method for electro-magnetic calculations

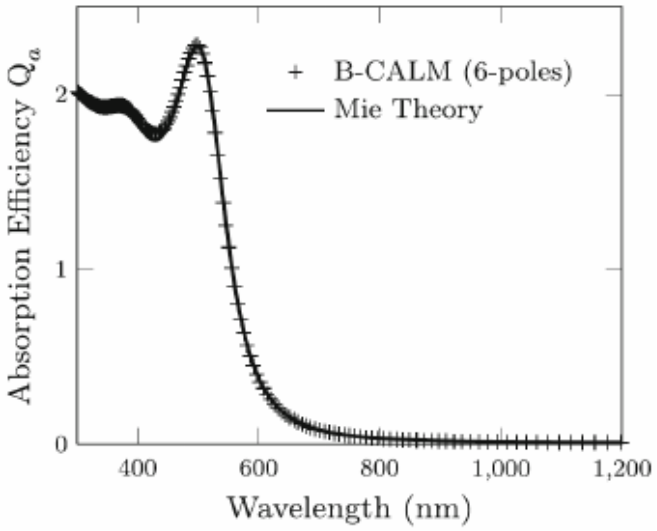
Example usage: Analysis of dispersive media

- Information technology
  Development of optical interconnects
- Energy technology
  Research on photo-electric cells

Focal area for application enablement

- Multi-GPU parallelisation
- I/O





[P. Wahl et al., 2012]

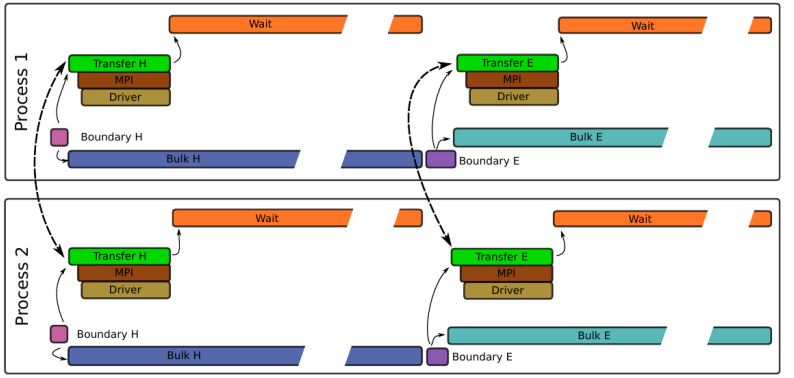
B-CALM Performance Modelling

Information exchange $I^k_{x,y}(W)$

- Information exchanged between storage device *x* and *y*
- Depends on problem size *W*

Semi-empirical performance modelling inspired approach

$$\Delta t = a + b\, I^k_{x,y}(W)$$



- Fit a and b to execution time measurements

Model ansatz for B-CALM

$$\Delta t = \Delta t_{\mathrm{bnd}} + \max\left(\Delta t_{\mathrm{bulk}}, \Delta t_{\mathrm{com}}\right)$$

Model Parameter Fitting / Model Use

doi:10.1109/HiPC.2015.24

Parameter fitting

- $P = 1$ and $P = 2$
- Different values $L_x = L_y$

Use: Parallelisation strategy

- Decision on domain decomposition

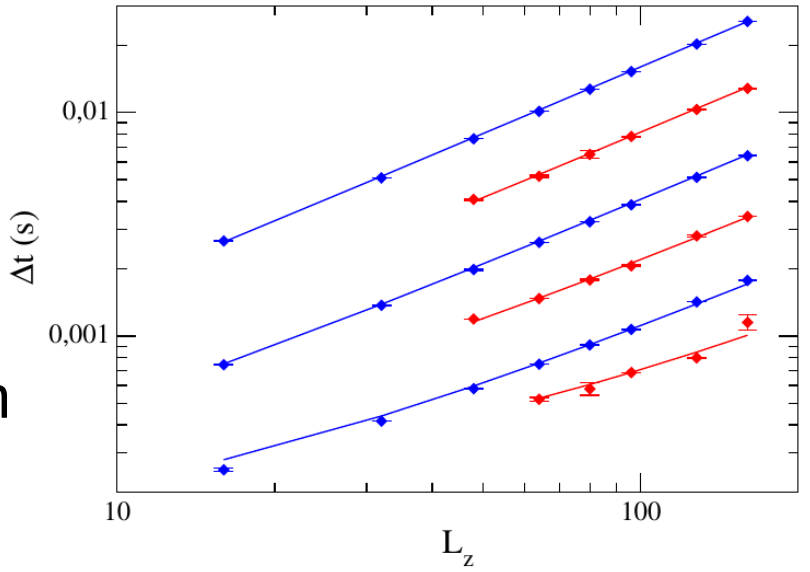
Use: Performance assessment

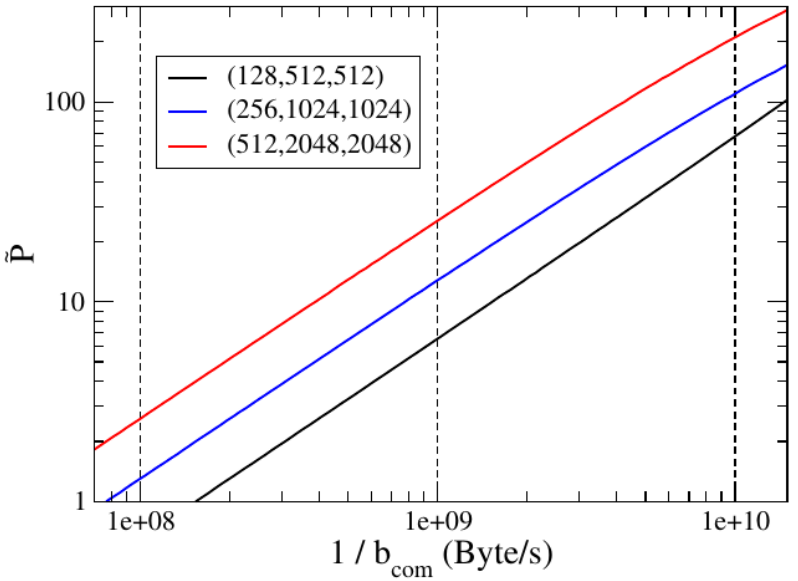
- Compare $1/b$ to theoretical bandwidth figures

Use: Hardware parameter requirements analysis

- For fixed problem size and given network link bandwidth: Determine number of nodes where

$$\Delta t_{\text{bulk}} = \Delta t_{\text{com}}$$





Application: KKRnano

Materials science application

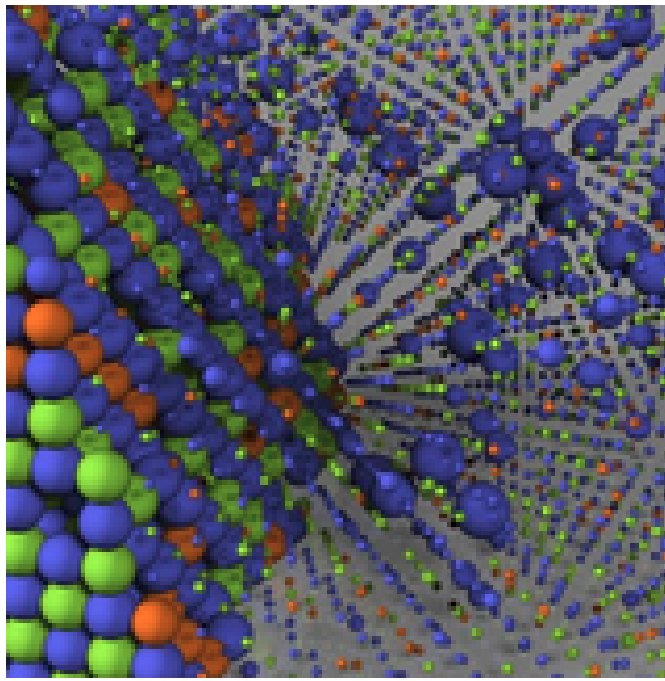
- High scalability due to truncation of long-range interactions
  → linear scaling in number of atoms

Performance characteristics

- Most time spent in iterative solver
- Dense matrix-matrix multiplications dominate performance (AI ≥ 4)

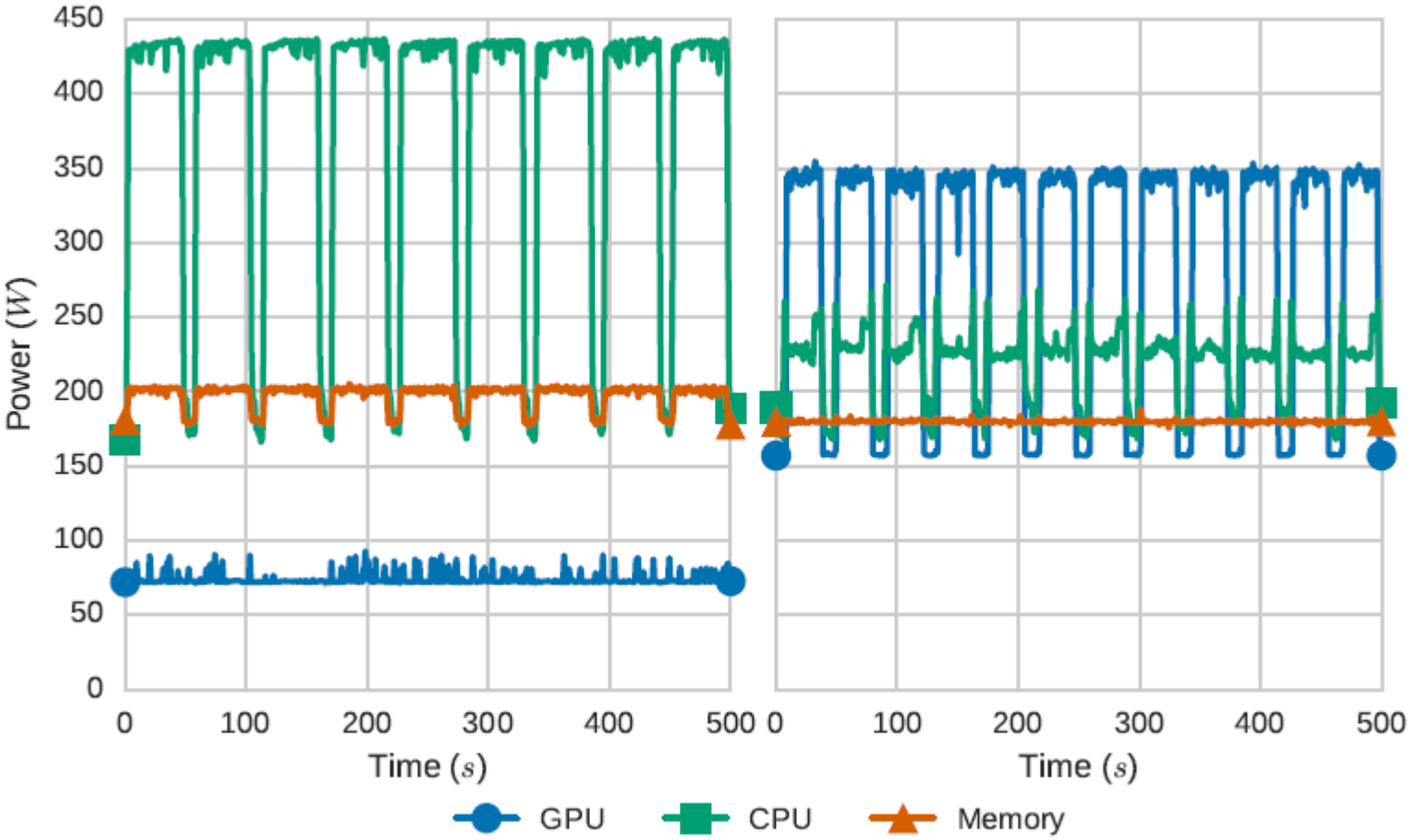
Implementation properties

- Fortran, MPI, OpenMP



Power/Energy Efficiency Evaluation

doi:10.1007/978-3-319-43659-3_6



POWER8: 2.5 nJ/Flop
K40: 1.95 nJ/Flop → 22% gain

New results from P. Baumeister:
~50% on P100 for b ≈ 30-50

Application: BQCD

High energy physics application

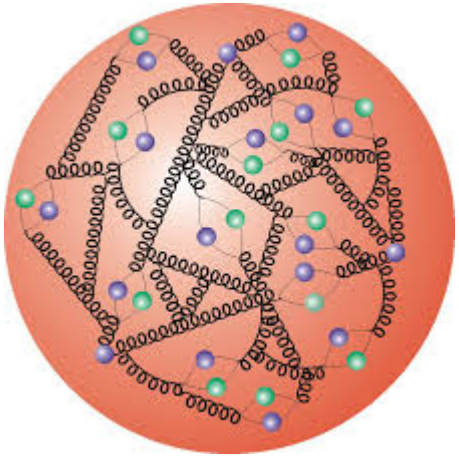
- Simulation of theory of strong interactions

Performance characteristics

- Most time spent in iterative linear solver (CG)
- Mainly of type regular grid (sparse matrix * vector)
- Low arithmetic intensity (0.45/0.9 for DP/SP)

Implementation properties

- Fortran90
- MPI + OpenMP



Maxeler Technology

Maxeler MAXJ

```
float A[N]; float B[N]; float C[N];
for (int i=0; i<N; i += 1)
  C[i] = A[i] + B[i];
```

```
DFEVar A = io.input(”A”, dfeFloat(8, 24));
DFEVar B = io.input(”B”, dfeFloat(8, 24));
DFEVar C = A + B;
io.output(”C” , C , dfeFloat(8, 24));
```

Application enablement challenge

- Data-flow paradigm
- Significantly reduced requirements for FPGA resource when reducing FP precision

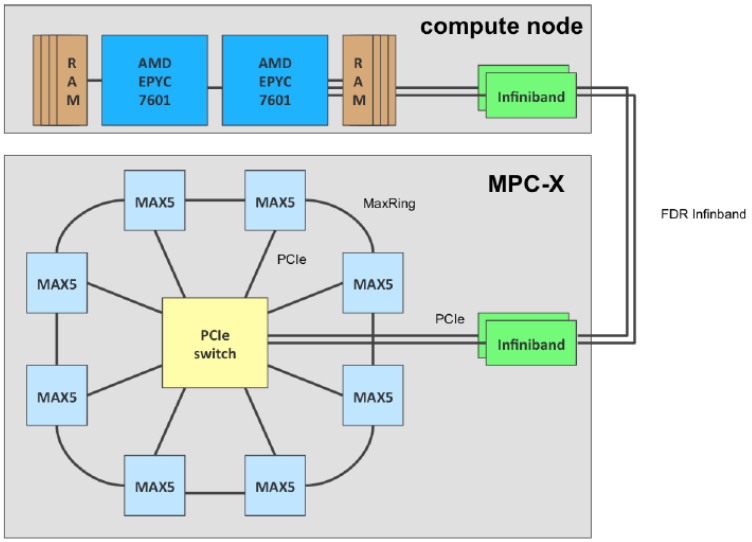
Likely enablement model: Domain Specific Libraries

- Community library similar to QUDA

Early results for BQCD and others

[Maxeler, 2017]

Application	Without DFE Acceleration		With DFE Acceleration	
	TTS [s]	ETS [Wh]	TTS [s]	ETS [Wh]
BQCD	8*507	562 (estimated)	~2500	406
NEMO	773	107 (estimated)	1099	160
Quantum Espresso	447	62	195	42

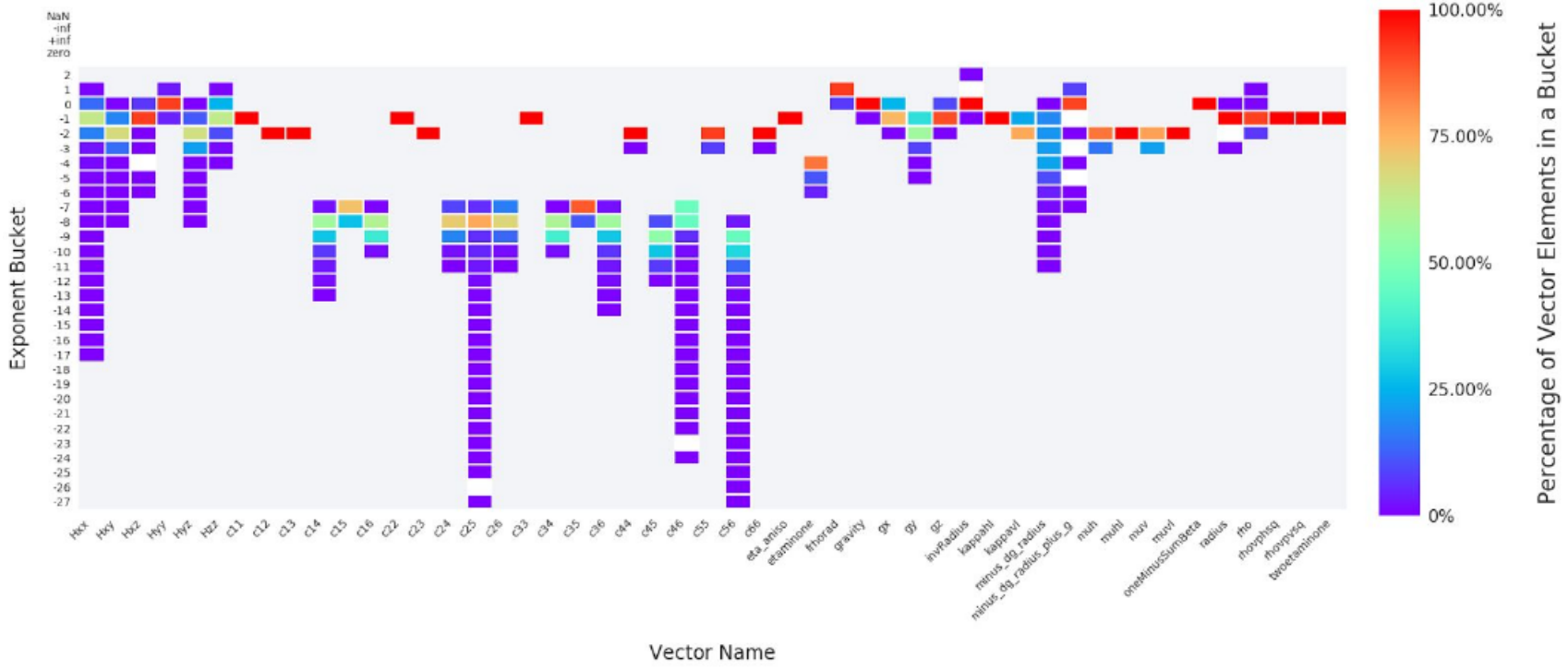


Maxeler‘s Value Profiling Library

Tool developed by Maxeler within PRACE-3IP PCP

- Floating-point operations can be implemented using different number of bits
- Profiling of the value range of variables can help guiding the decision

Example for SpecFEM3D



Summary

Enabling work currently more technology driven

- Future: More work through Simulation Labs

Proven enabling workflow for GPU accelerationn

- Users remain responsible for their application

New challenges for FPGAs

- Need to rethink applications and algorithms
  - Understand opportunities for reducing FP precision
  - Porting efforts may not pay-off for single application
- Need for community efforts