Project Seminar on Energy-Aware Computing in Heterogeneous Data Centers

Introduction: Motivation, Organization, Hardware, Project Topics

Max Plauth, *Sven Köhler*, Lukas Wenzel, Felix Eberhardt, and Andreas Polze

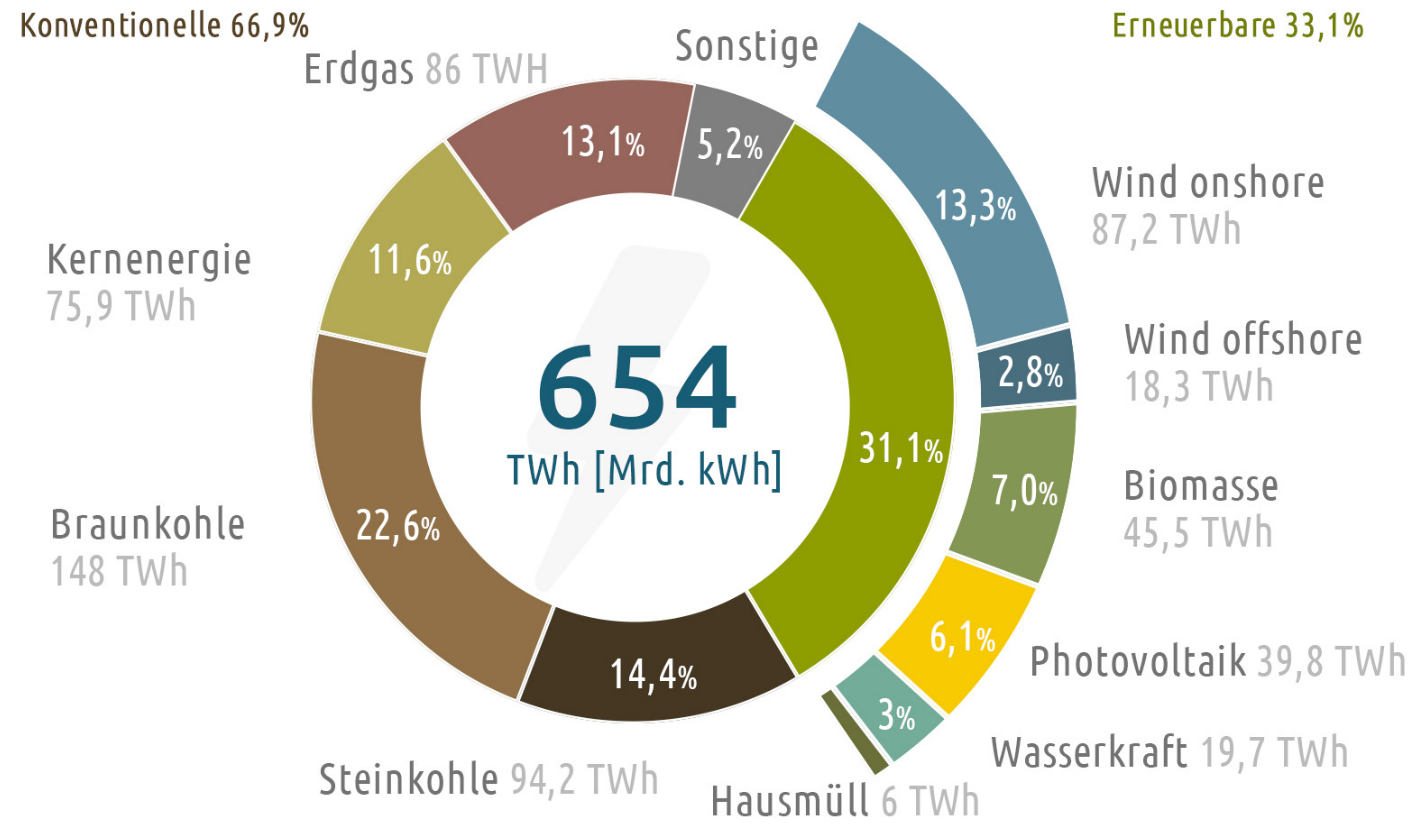
Operating Systems and Middleware Group

1

Motivation

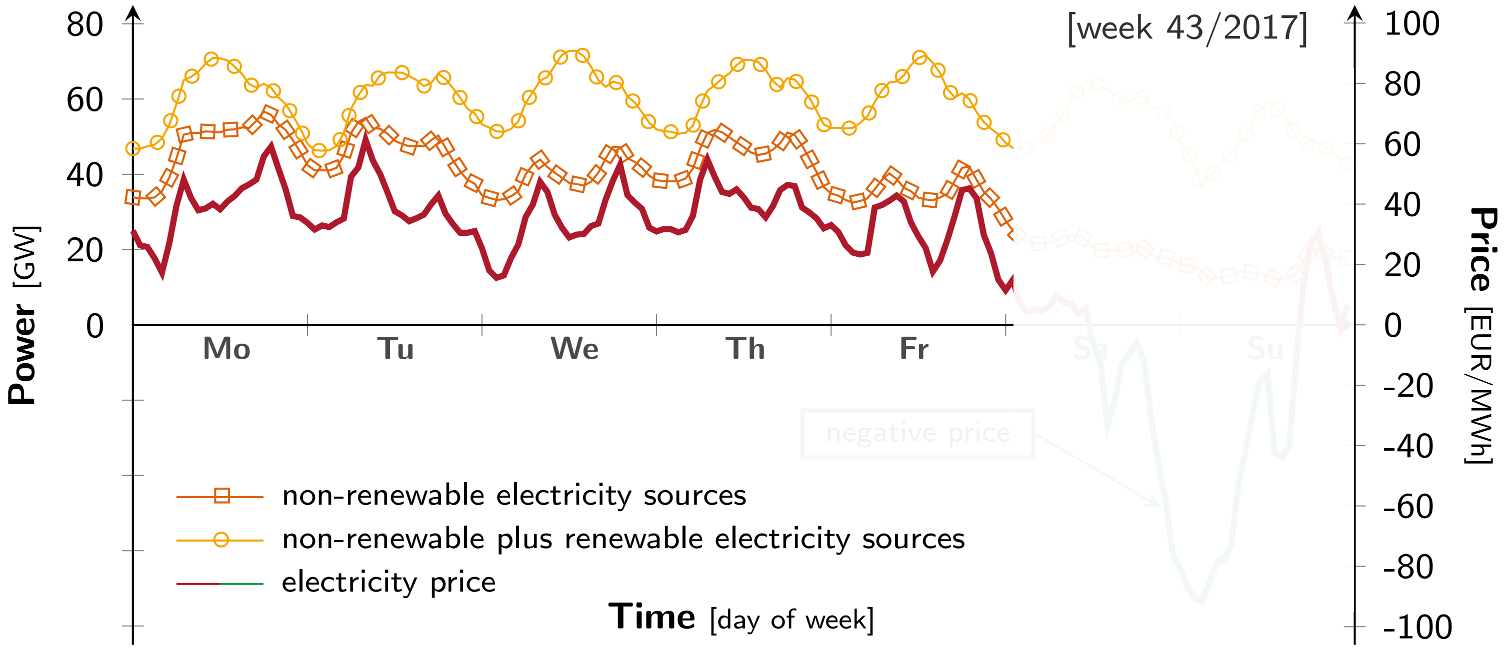
DER STROMMIX IN DEUTSCHLAND 2017 [BRUTTO]

Anteil der Energieträger an der Bruttostromerzeugung in Deutschland

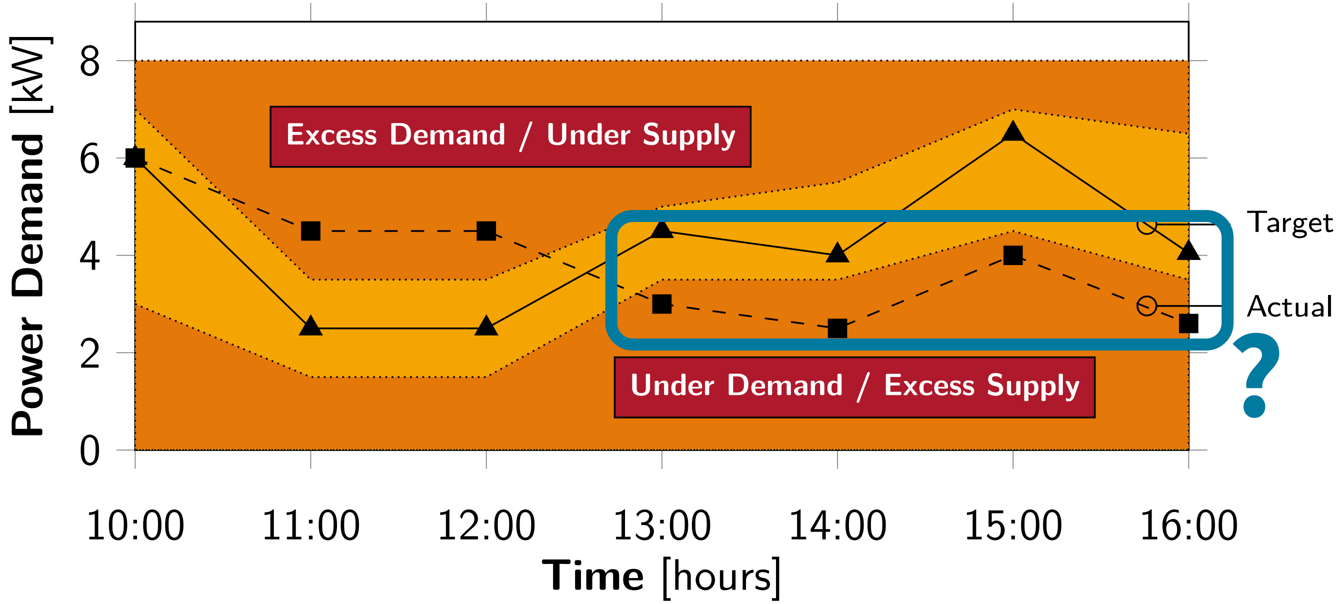


Daten: AGEB e.V. 2018

1. Motivation
Energy Price Over Time



Fluctuating energy supply leads to highly variable energy prices



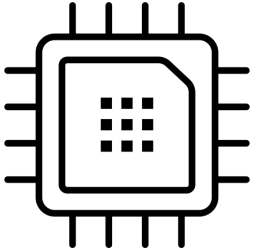
Variable power-target and penalty zones must be considered in computing laboratory operation

How to control power consumption of computing lab operation?

1. **Work another time:** Defer/pick jobs with matching energy profile (popular research in embedded systems, requires known and deferrable jobs)
2. **Work less/more:** Use additional hardware components (commonly used to scale depending on operation requirements)
3. **Work elsewhere:** Use other hardware components

This project seminar

1. Motivation
Different Hardware Classes



CPU

Intel

AMD

Power

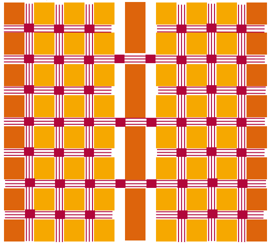
ARM

GPU

AMD

NVIDIA

Mali

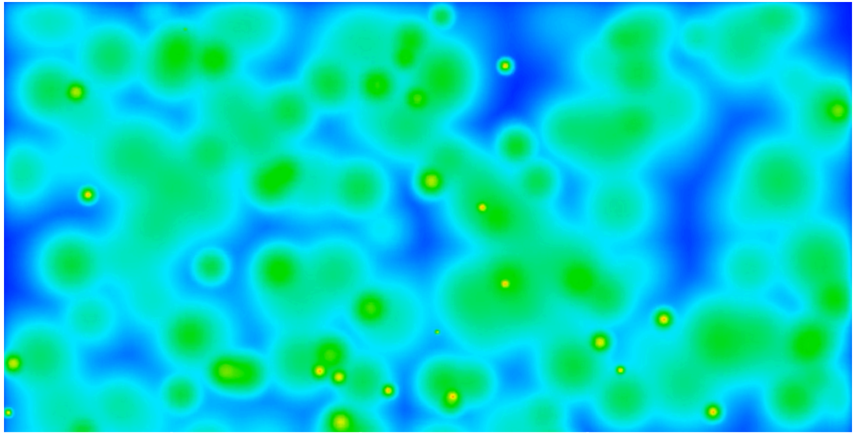


FPGA

Zynq SOC

Xilinx via CAPI

1. Motivation
Recap: Parallel Programming Assignments ST 2019



Device	Problem Size (MCells)	Cell Type	Throughput (MCells/s)	Power Consumption (W)	Power Efficiency (MCells/Ws)
NVidia Tesla K20Xm	256	float (32bit)	1127,88	60	18,80
Intel E5-2630 v4	256	float (32bit)	1435,48	85	16,89
Intel E5-2630 v4	256	char (8bit)	1420,29	85	16,71
Xilinx XCKU060	1024	char (8bit)	2209,35	9,5	232,56

Goal of This Project Seminar

In this project seminar we want you to learn:

1. How to target different hardware platforms

2. How to assess the current power consumption of your algorithm

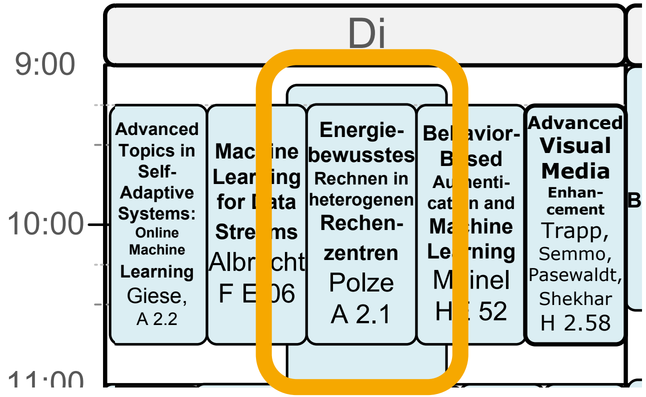
3. Finding the preferred hardware platform for your given algorithm class

2

Organization

2. Organization
Time Slots

- **Project Seminar** (= *Not a lecture*)
- **6 ECTS** credits (4 SWS)
- **Each Tuesday** (start of the semester):
  - Announcements
  - Heterogeneous Programming Course
  - Profiling Course
  - Exchange with fellow students
- No regular Tuesday meetings later in the semester.
  Use the time to work on your projects :)



2. Organization
Your project focus

Each team (1-3 persons) can choose to focus on:

mastering 2+ hardware platforms in depth

or

optimizing 2+ algorithm classes on multiple platforms

or

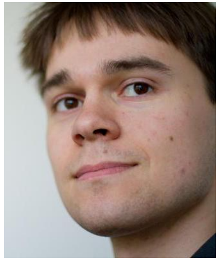
developing new profiling tools

speak to your project mentor

2. Organization
Mentoring Team (PhD Students)

- Max Plauth
  - GPU Accelerators
  - Virtualization of Accelerators
  - Bandwidth Compression

- Felix Eberhardt
  - NUMA Topologies
  - Dynamic Scaling
  - Profiling



- Sven Köhler
  - On-Core Accelerators
  - Programming Models for Accelerators
  - Hardware Performance Counters

- Lukas Wenzel
  - FPGA Accelerators
  - Near-Storage Computing
  - FPGA-OS integration

2. Organization
Grading Process

Requirements for each team to successfully pass this course:

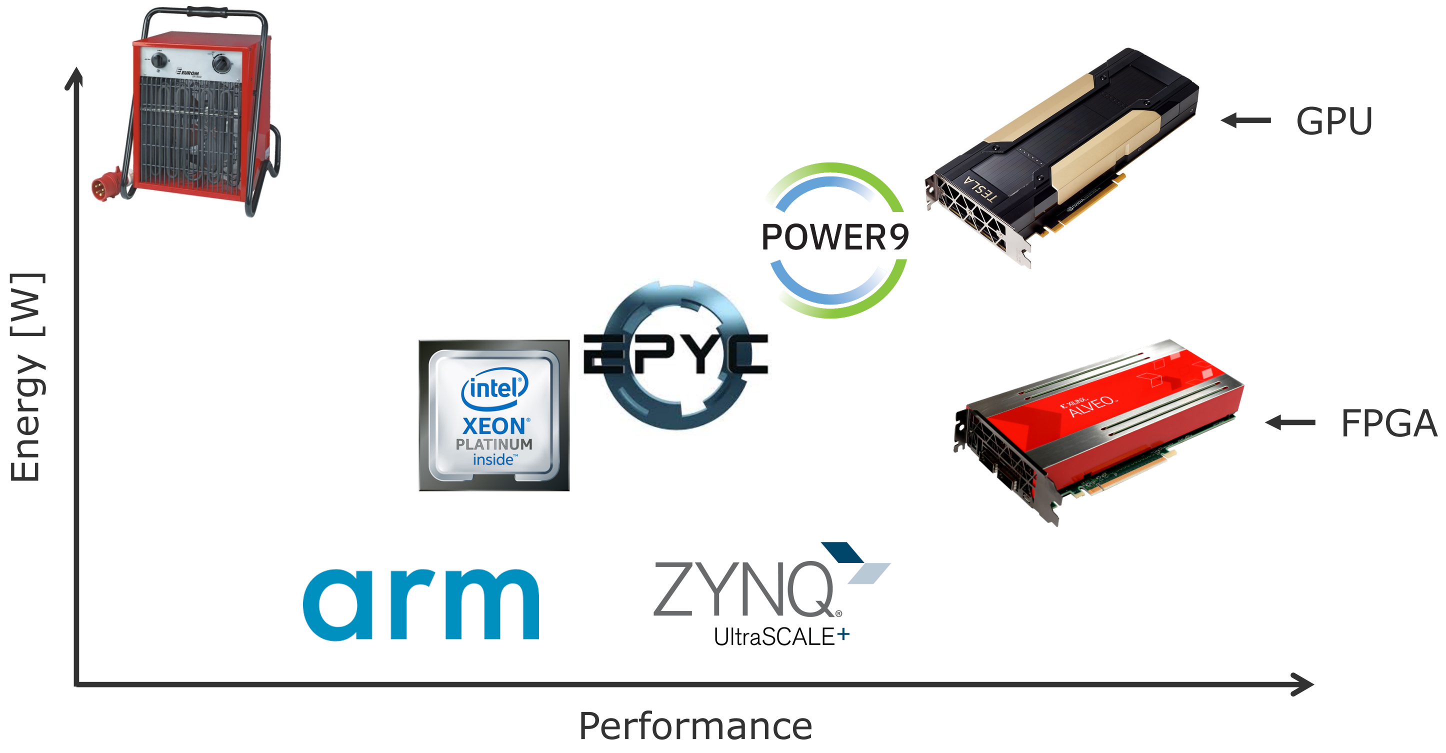
- Meet regularly with your project mentor
- 10 minute presentation in November
- 30 minute presentation at end of semester
- 5-10 pages technical report

- **Grading**: 50 % project, 50 % presentations and report

3

Potential Hardware Targets

3. Potential Hardware Targets
Energy vs. Performance

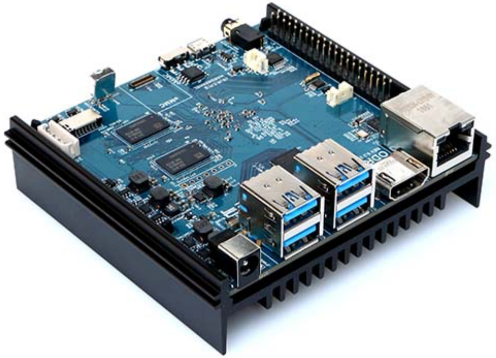


Energy-Aware Computing Seminar

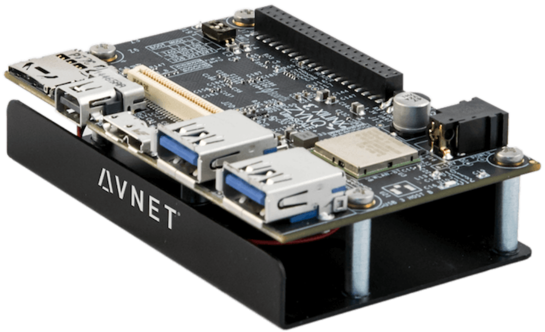
Max Plauth

3. Potential Hardware Targets
Low-Power Systems

- Odroid N2
  - 4x ARM A73 / 2x ARM A53 CPUs
    - ISA: aarch64
  - 1x Mali-G52 GPU
  - ~6.5 Watt under load



- Ultra96
  - 4x ARM Cortex-A53 CPU
    - ISA: aarch64
  - 1x Xilinx UltraScale+ FPGA
  - ~5.5 Watt under load



3. Potential Hardware Targets
Mid-Power Systems

- HPE m710p Moonshot Cartridge (up to 10 machines available)
  - Intel Xeon E3-1284Lv4
    (4 Cores / 8 Threads)
    - ISA: x86_64
  - 1x Iris Pro P6300 GPU
  - RoCE-enabled NIC
  - ~30 Watt under load

3. Potential Hardware Targets: High-Power Systems (CPU-only)

- HPE ProLiant DL580 Gen9
  - 4x Intel Xeon E3-1284Lv4 (18 Cores / 36 Threads)
  - >1kW Watt under load

3. Potential Hardware Targets
High-Power Systems (with Accelerators)

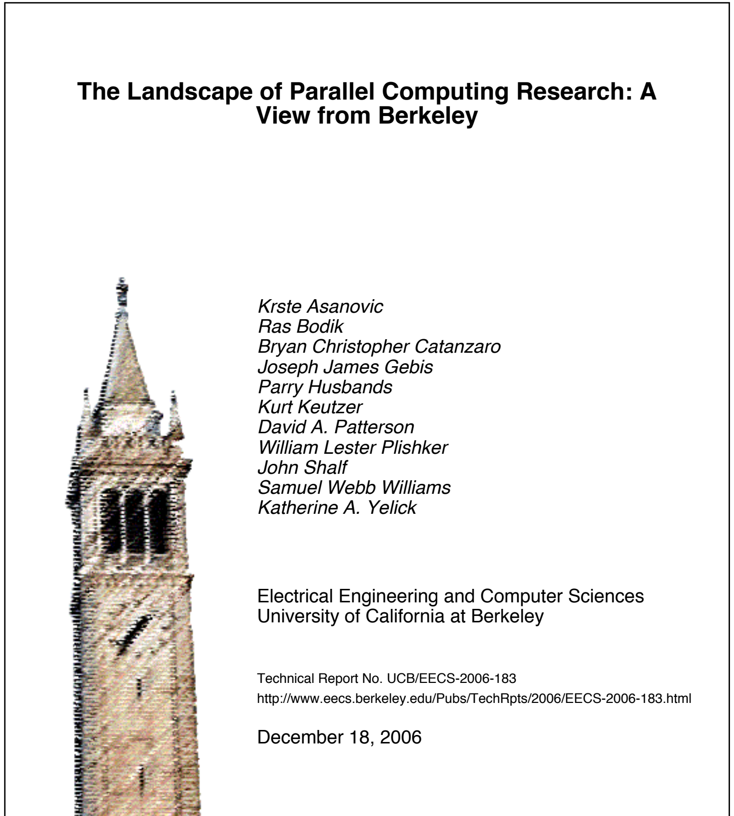
- HPE ProLiant DL380 Gen9
  - 2x Intel Xeon E5-2630v4 (10C/20T)
    - ISA: x86_64
  - 2x NVIDIA Tesla K20Xm GPU
  - >1kW under load

- IBM Power System S824L
  - 2x IBM POWER8 (10C/80T)
    - ISA: ppc64le
  - 1x NVIDIA Tesla K80 GPU
  - 1x Nallatech 250S FPGA
  - >1kW under load

3. Potential Hardware Targets
High-Power Systems („Accelerator-only“)

- NVIDIA DGX-1
  - 2x Intel Xeon E5-2698v4 (20C/40T)
    - ISA: x86_64
  - 8x NVIDIA Tesla V100 GPU
  - ~3.5 kW under load

4 Potential Algorithms



The Landscape of Parallel Computing Research: A View from Berkeley

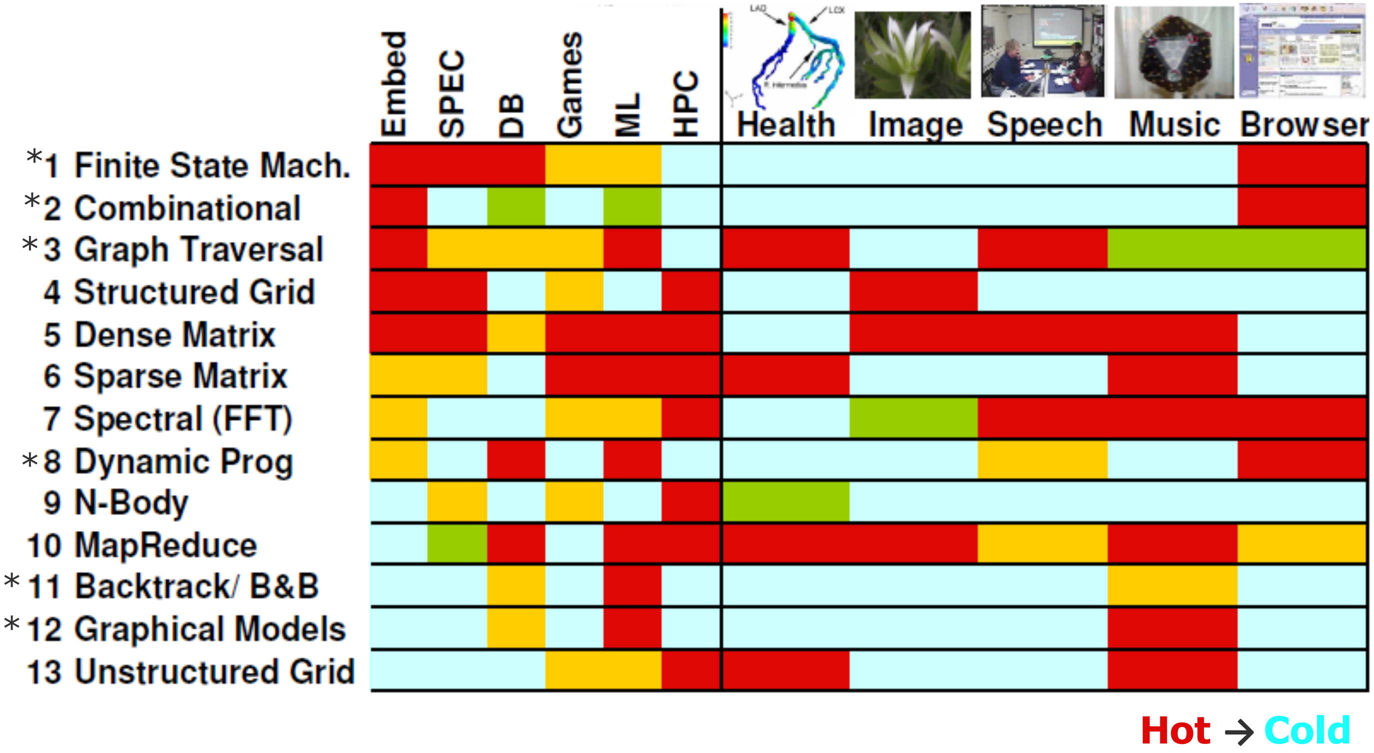
Krste Asanovic
Ras Bodik
Bryan Christopher Catanzaro
Joseph James Gebis
Parry Husbands
Kurt Keutzer
David A. Patterson
William Lester Plishker
John Shalf
Samuel Webb Williams
Katherine A. Yelick

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2006-183
http://www.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-183.html

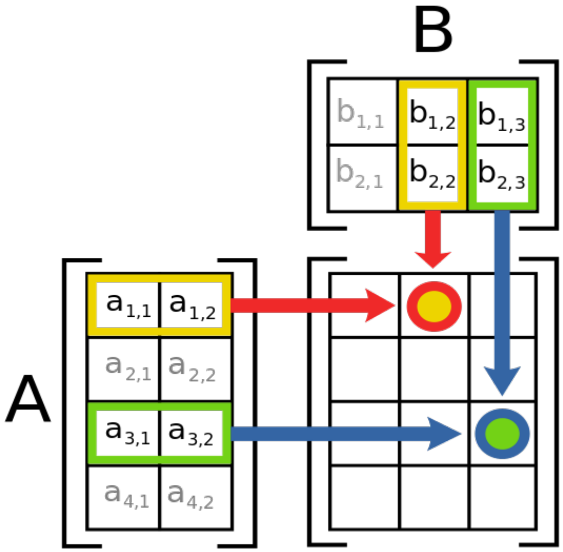
December 18, 2006

4. Potential Algorithms
How Berkley Dwarfs Relate to Compelling Application Fields



* added later

Dwarf 1: Dense Linear Algebra



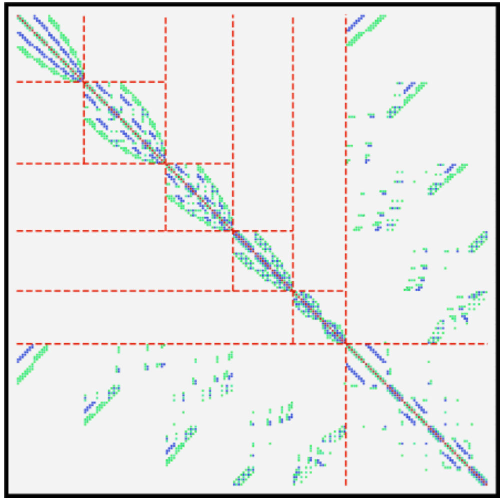
Classic vector and matrix operations

```
do i=1,n
  do j=1,n
    do k=1,n
      a(i,j) = a(i,j) + b(i,k)*c(k,j)
```

Frequent operation in computer graphics and as training step in machine learning

C = AxB

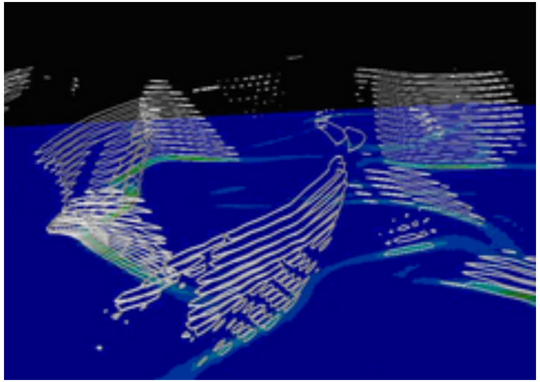
Dwarf 2: Sparse Linear Algebra

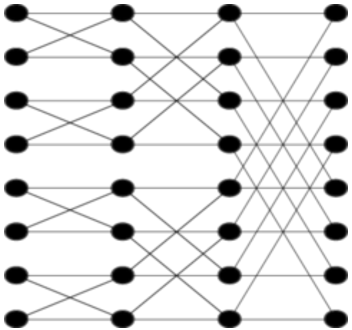


Operations on a sparse matrix (lots of zeros)

```
do i=1,n
  do j=row_start(i),row_start(i+1)-1
    y(i) = y(i) + val(j)*x(col_index(j))
```

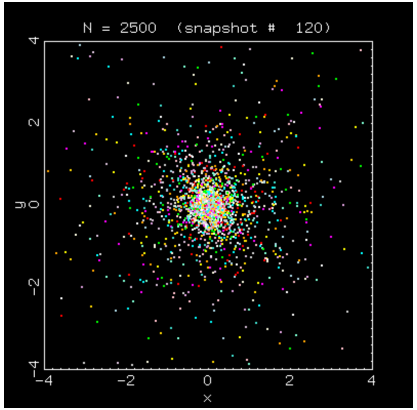
Complex data-dependency structure
Common in e.g. in graph problems.

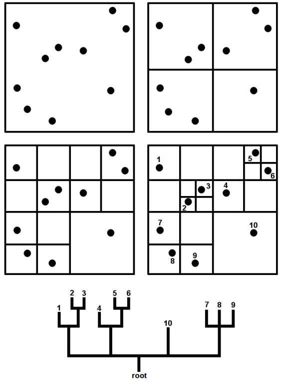




Data is converted into other domains, which means multiple stages with inter-depended data access patterns.

Common ML data preparation step, or used in signal processing.

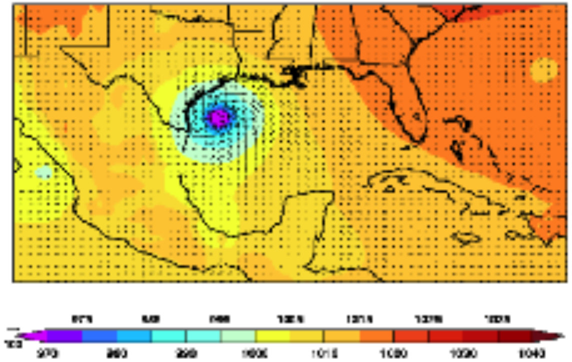


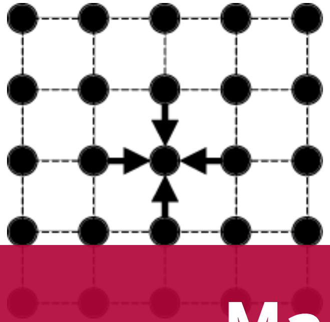


Calculations on interactions between
Many discrete points.

Large number of independent calculations
in a time step, followed by wide communication.

Dwarf 5: Structured Grid





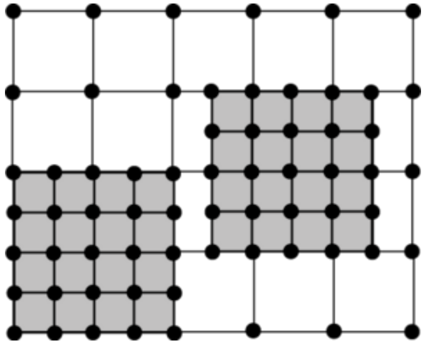
Data as a regular multidimensional grid: access is regular and statically determinable (strided).

Computation is sequence of grid updates (all points are updated using values from a small neighborhood).

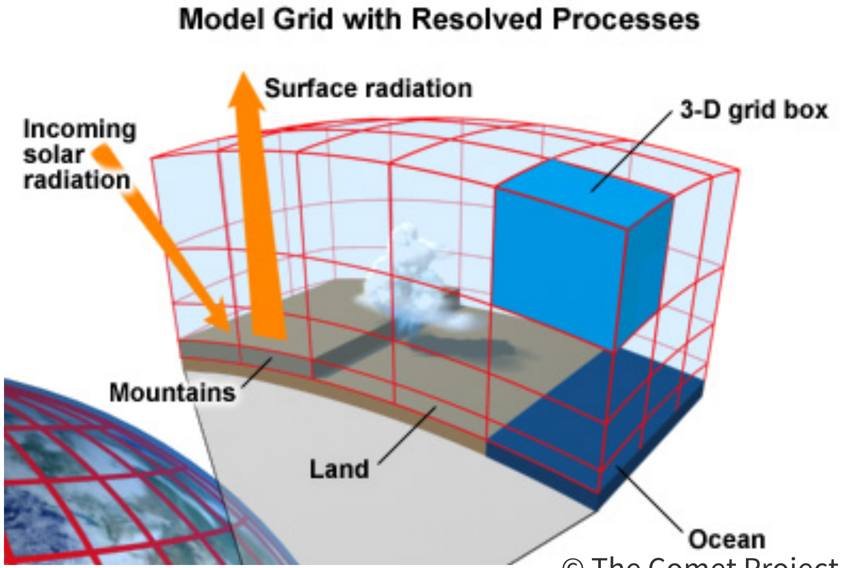
Typical Application: Weather simulations

Mandatory warm-up for all teams

Dwarf 5 Variant: Adaptive Mesh Refinement



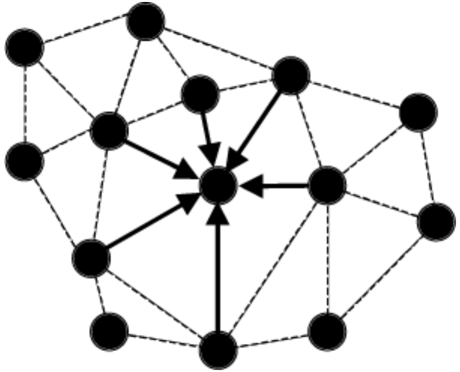
Overlaying higher-resolution grids across areas of interest. Requires complex indexing and difficult communication across nodes.



Example:

Modular Ocean Model

4. Potential Algorithms
Dwarf 6: Unstructured Grid



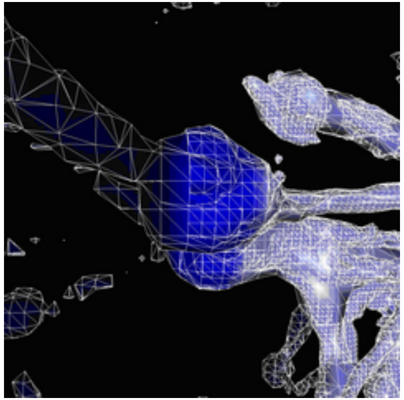
Elements update neighbors in irregular mesh/grid with static or dynamic structure

Problematic data distribution and access requirements, usually indirection by tables.

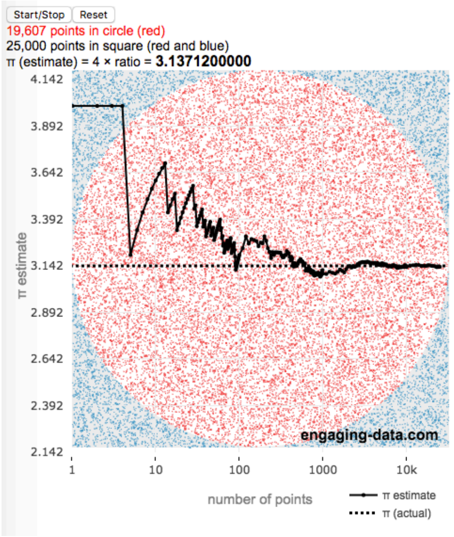
$$A'[B[C[i]]] = f(A[B[C[i+1]]] + A[B[C[i+2]]] + A[B[C[i+3]]])$$

Modelling domain (e.g. physics engine)

- Mesh represents surface or volume
- Entities are points, edges, faces, volumes, …
- Applying tension, temperature, pressure



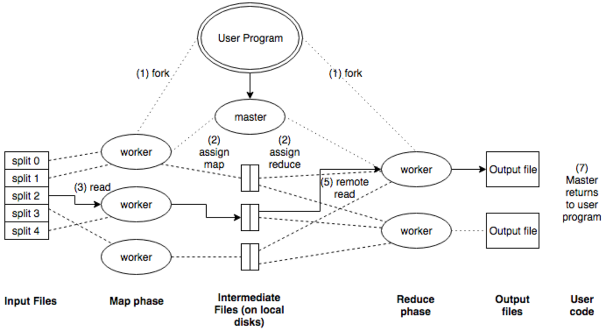
4. Potential Algorithms
Dwarf 7: MapReduce (= “Monte Carlo”)



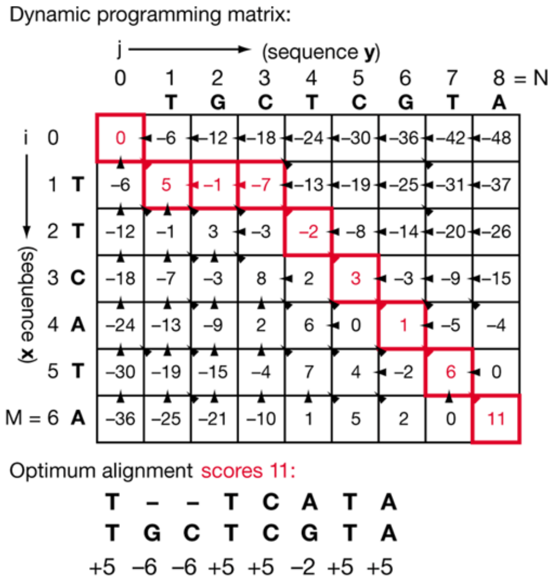
Repeated independent execution of a function (e.g. RNG, map function), results aggregated.

Examples:

Monte Carlo Pi, BOINC (SETI@home), Optimization Protein Structure Prediction



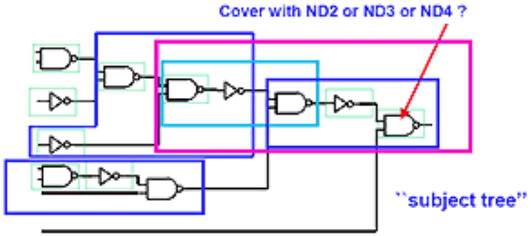
4. Potential Algorithms
Dwarf 10(*): Dynamic Programming



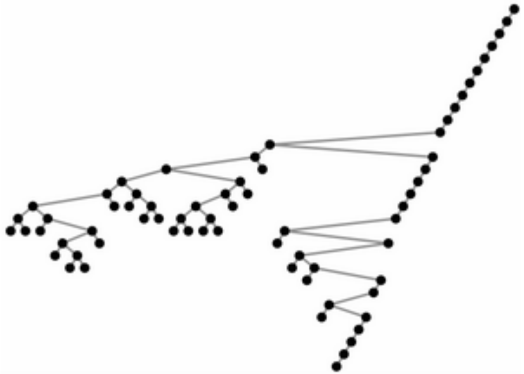
Compute optimal solutions by combining optimal, yet simpler overlapping subproblem solutions (typically use a table to avoid recomputation)

Examples:

circuit design, DNA sequence matching (Needleman–Wunsch), Viterbi, Knapsack, …

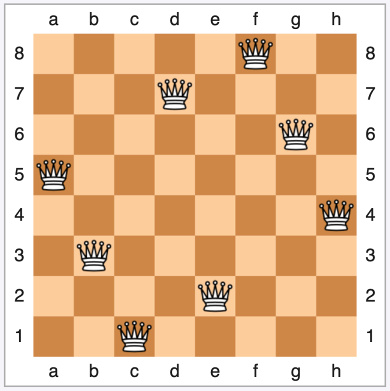


4. Potential Algorithms
Dwarf 11(*): Branch-and-Bound



Global optimization problem in very large search space:

- Branches into subdivisions
- Rules out infeasible regions to optimize execution time and energy consumption



Examples:

Integer Linear Programming, Boolean Satisfiability, Combinatorial Optimization, Traveling Salesman, Constraint Programming, N-Queens

AND

end

1. Motivation
Vision: Hardware Agnostic Job Format