PARLab Parallel Boot Camp

Short Course on Parallel Computing
August 19-21, 2009
parlab.eecs.berkeley.edu/2009bootcamp

Jim Demmel
EECS and Mathematics
University of California, Berkeley

Outline

- Goals
- Background and Motivation
  - ParLab, a new research center in Parallel Computing
  - The Designated Emphasis (DE) in Computational Science and Engineering (CSE)
  - CSE at Lawrence Berkeley National Lab
- Schedule and Instructors
- Logistics
- The Audience

Motivation (1/2)

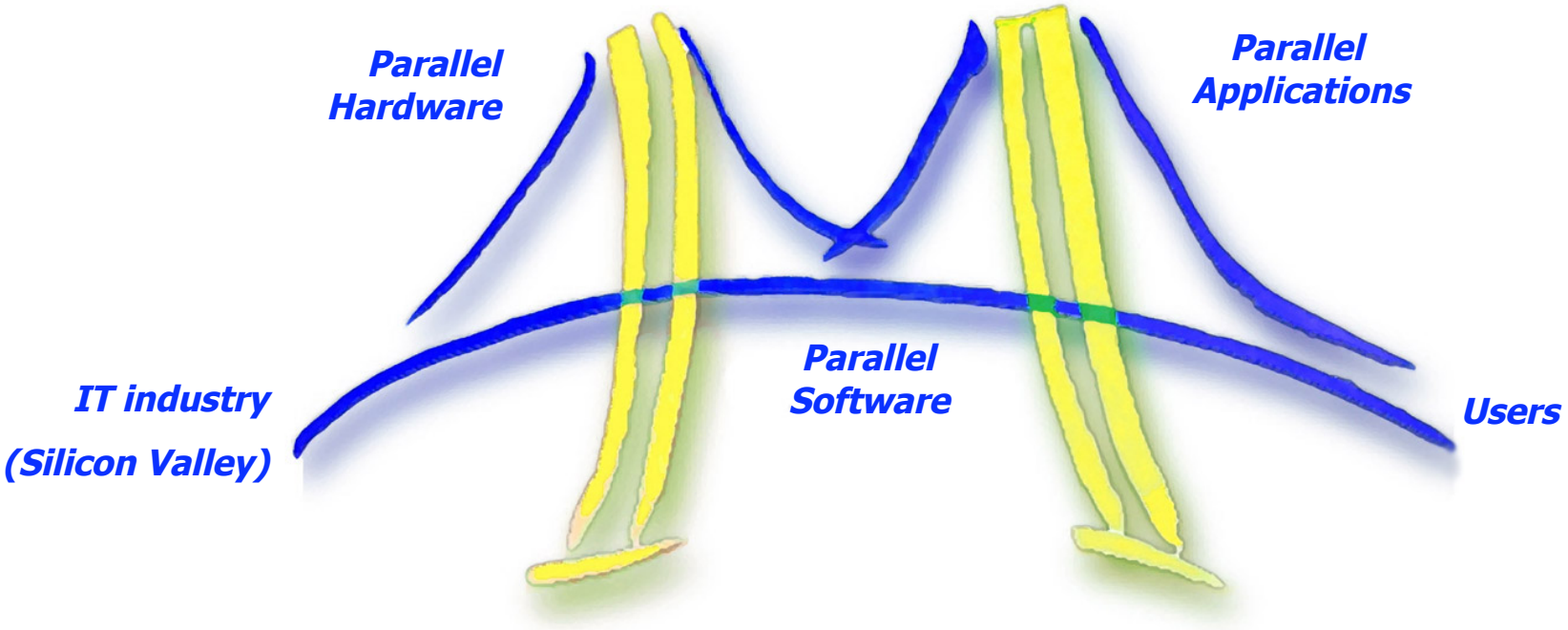
- Parallel Computing is becoming ubiquitous
  - Only way forward for computing industry (unless you don't care if your programs never run faster than in 2008)
  - Unfortunately, parallel programming is (still) harder than sequential programming
  - Until better (easier) programming tools come along, we need to train everyone in parallel programming
- So welcome!

Motivation (2/2)

- Recent events at UCB will provide support for many new activities to develop and use parallel computing
  - ParLab established – parlab.eecs.berkeley.edu
    - » Research center about "Multicore Revolution"
  - Designated Emphasis in Computational Science and Engineering (CSE) established – cse.berkeley.edu
    - » New graduate program with 120 faculty from 22 departments

Short Course Goals

- Teach the basics about parallelism
  - How to program, including hands-on lab
- Tools you can use now (simple and advanced)
- Tools we hope to build, and ongoing research



Berkeley ParLab Project

Krste Asanovic, Ras Bodik, Jim Demmel, Tony Keaveny, Kurt Keutzer, John Kubiatowicz, Edward Lee, Nelson Morgan, Dave Patterson, Koushik Sen, John Wawrzynek, David Wessel, and Kathy Yelick

7 Dwarfs of High Performance Computing (HPC)

Structured Grid
Dense Matrix
Sparse Matrix
Spectral (FFT)

Particle Methods

Unstructured Grid
Monte Carlo

7 Dwarfs – Are they enough?

-

Embed
SPEC
DB
Games
ML
CAD
HPC

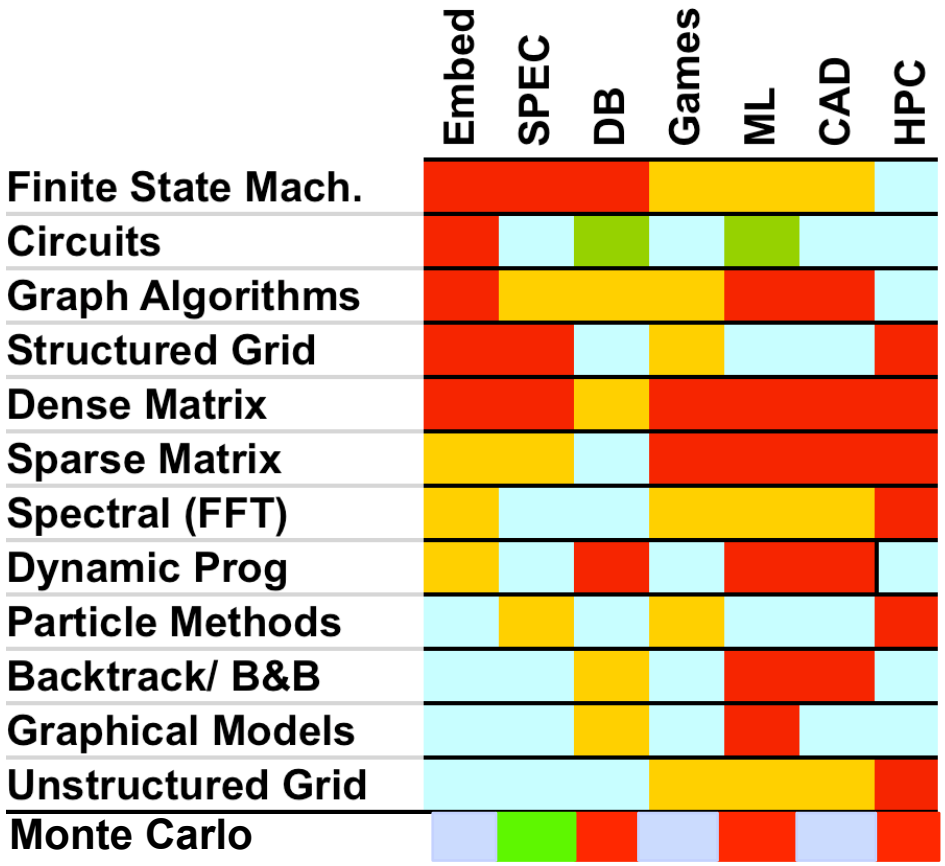
Structured Grid
Dense Matrix
Sparse Matrix
Spectral (FFT)

Particle Methods

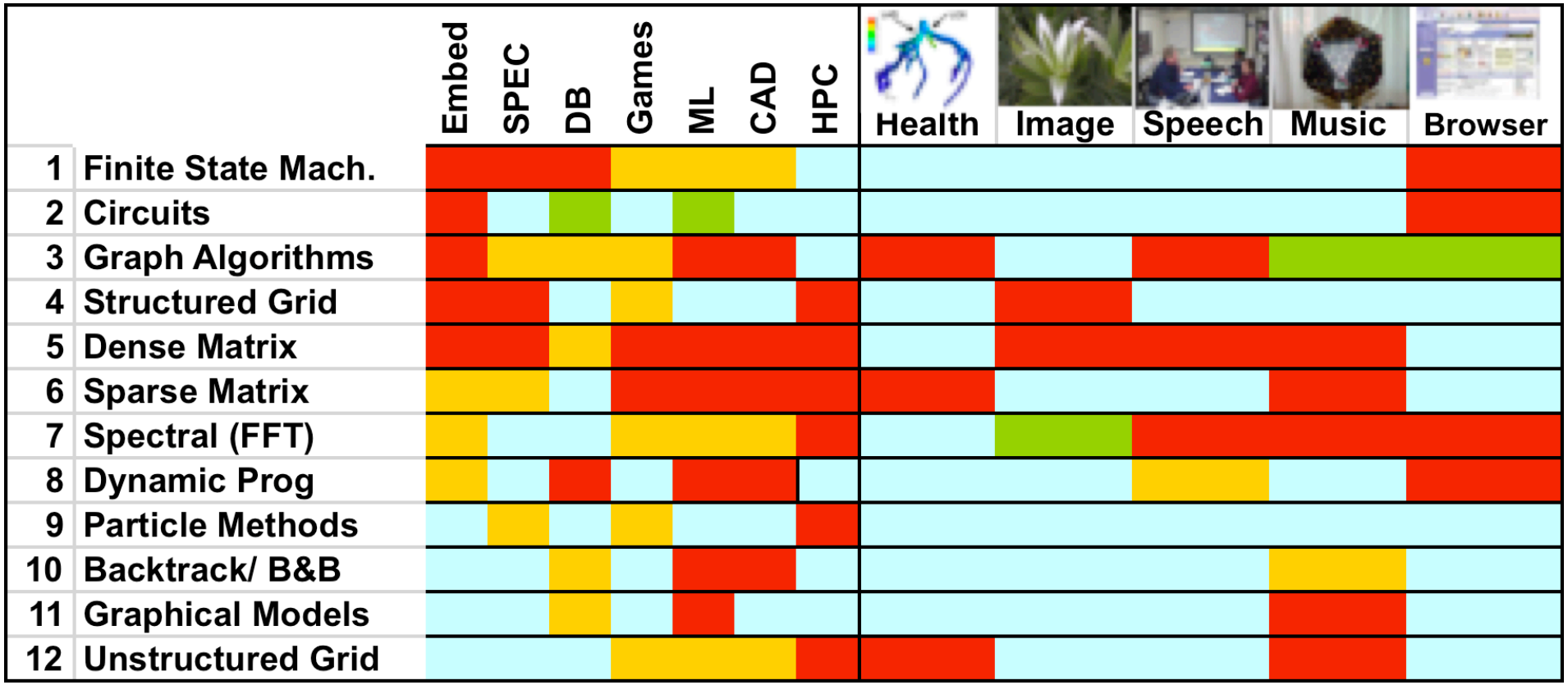
Unstructured Grid
Monte Carlo

13 Motifs (nee "Dwarf") of Parallel Computing

Popularity: (Red Hot / Blue Cool)



Motifs in ParLab Applications
(Red Hot / Blue Cool)



- What happened to Monte Carlo?

Programming Pattern Language 1.0 Keutzer& Mattson

Applications

Choose your high level structure – what is the structure of my application? Guided expansion

Pipe-and-filter
Agent and Repository
Process Control
Event based, implicit invocation

Choose your high level architecture - Guided decomposition

Task Decomposition ↔ Data Decomposition
Group Tasks Order groups data sharing data access

Model-view controller
Iterator
Map reduce
Layered systems
Arbitrary Static Task Graph

Identify the key computational patterns – what are my key computations? Guided instantiation

Graph Algorithms
Dynamic Programming
Dense Linear Algebra
Sparse Linear Algebra
Unstructured Grids
Structured Grids
Graphical models
Finite state machines
Backtrack Branch and Bound
N-Body methods
Circuits
Spectral Methods

Productivity Layer

Refine the structure - what concurrent approach do I use? Guided re-organization

Event Based
Divide and Conquer
Data Parallelism
Geometric Decomposition
Pipeline
Discrete Event
Task Parallelism
Graph algorithms
Digital Circuits

Efficiency Layer

Utilize Supporting Structures – how do I implement my concurrency? Guided mapping

Fork/Join
CSP
Distributed Array Shared -Data
Shared Queue
Shared Hash Table
Master/worker
Loop Parallelism
BSP

Implementation methods – what are the building blocks of parallel programming? Guided implementation

Thread Creation/destruction
Process/Creation/destruction
Message passing
Collective communication
Speculation
Transactional memory
Barriers
Mutex
Semaphores

Designated Emphasis (DE) in Computational Science and Engineering (CSE)

- Goals
- Participants (116 faculty from 22 departments – so far)
- How the DE works
- Resources and Opportunities
- Details at cse.berkeley.edu

Designated Emphasis (DE) in CSE

- New "graduate minor" – approved, starting July 1, 2008
- Motivation
  - Widespread need to train PhD students in large scale simulation, or analysis of large data sets
  - Opportunities for collaboration, across campus and at LBNL
- Graduate students participate by
  - Getting accepted into existing department/program
  - Taking CSE course requirements
  - Qualifying examination with CSE component
  - Thesis with CSE component
  - Receive "PhD in X with a DE in CSE"
  - Details at cse.berkeley.edu (soon!)

Participating Departments (1/2)

(# faculty by "primary affiliation", # courses)

- Astronomy (7,3)
- Bioengineering (3,1)
- Biostatistics (2,0)
- Chemical Engineering (6,0)
- Chemistry (8,1)
- Civil and Environmental Engineering (7,8)
- Earth and Planetary Science (6,3)
- EECS (19,14)
- IEOR (5,5)
- School of Information (1,0)

Participating Departments (2/2)

(# faculty by "primary affiliation", # courses)

- Integrative Biology (1,0)
- Materials Science and Engineering (2,1)
- Mathematics (15, 4)
- Mechanical Engineering (9, 6)
- Neuroscience (7,1)
- Nuclear Engineering (2,1)
- Physics (1,1)
- Political Science (2,0)
- Statistics (5, 11)
- New: Biostatistics, Public Health

Course Structure

- 3 kinds of students, course requirements
  - CS , Math, "Applications"
- Each kind of student has 3 course requirements in other two fields
  - Goal: enforce cross-disciplinary training
  - Non-CS & Non-Math students:
    - » 1 or 2 Math courses from list
    - » 1 or 2 EECS courses from list
    - » Other classes from Stat, IEOR
  - Math & CS students: substitute 1 or 2 courses from "applied" department for 1 or 2 inside

Example Course – CS267

- "Applications of Parallel Computing"
  - Long version of this short course!
  - see www.cs.berkeley.edu/~demmel/cs267_Spr09
- Taught every Spring, during Spr09 semester to:
  - UC Berkeley, UC Merced, UC Santa Cruz, UC Davis
  - All lectures on web (slides + video), freely available
- Google "parallel computing course" to get older version, with detailed text-book like notes on algorithms

A few sample CS267 Class Projects (all posters and video on web page)

- Content based image recognition
  - "Find me other pictures of the person in this picture"
- Faster molecular dynamics, applied to Alzheimer's Disease
- Better speech recognition through a faster "inference engine"
- Faster algorithms to tolerate errors in new genome sequencers
- Faster simulation of marine zooplankton population
- Sharing cell-phone bandwidth for faster transfers

Some CSE Resources

- Executive Director Masoud Nikravesh
  - nikravesh@cs.berkeley.edu
- Student Affairs Officer Pat Berumen
  - patbcoe@berkeley.edu
- Head Graduate Adviser Andy Packard
  - pack@me.berkeley.edu
- Money
  - From L&S, COE, Chem, VCR, MathWorks (course development)
  - CITRIS
    - » Research initiation funds
    - » Access to corporate partners
    - » GSI support, broadcast to other campuses
- Computing resources
  - Cloud computing, start up allocations from LBNL/NERSC, clusters
- LBNL …

NERSC
National Energy Research Scientific Computing Center

www.nersc.gov

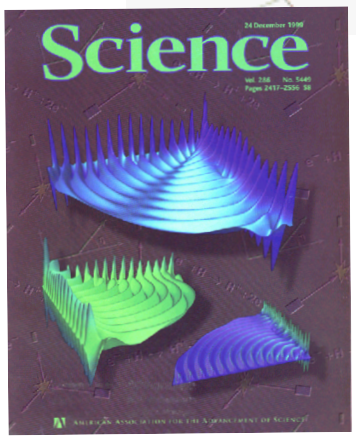
National Energy Research Scientific Computing Center

Serves the <u>entire</u> scientific community

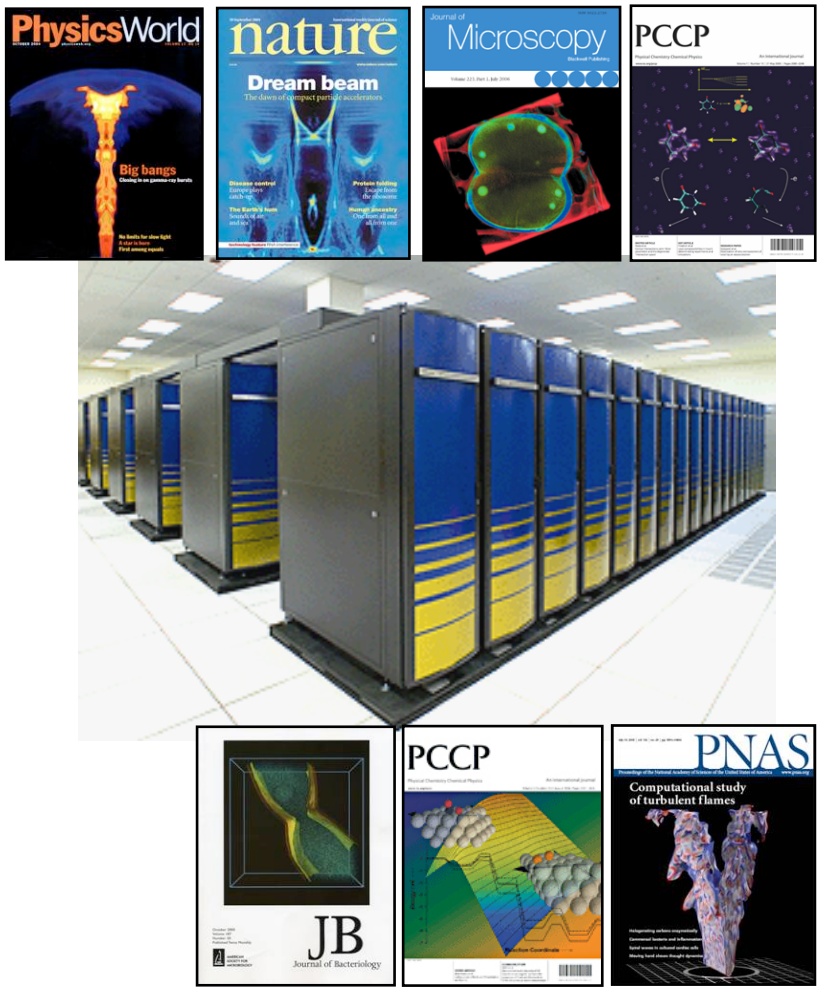
~2500 Users in

~250 projects

- Focus on large-scale computing



100 Tflops Cray XT-4 at NERSC



Cray XT-4 "Franklin"
19,344 compute cores
102 Tflop/sec peak
39 TB memory
350 TB usable disk space
50 PB storage archive

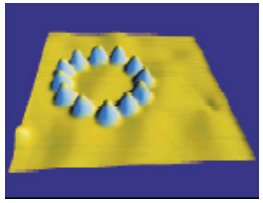
NERSC is enabling new science

CRD
computational
research division

Computational Research Division

crd.lbl.gov

Computational Science Mission

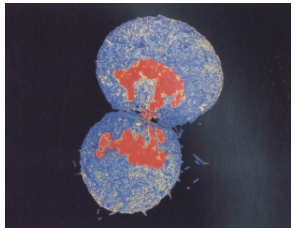


nano

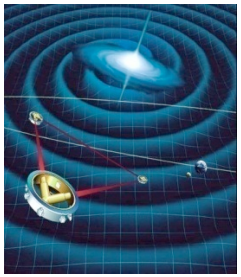
systems

biological

systems

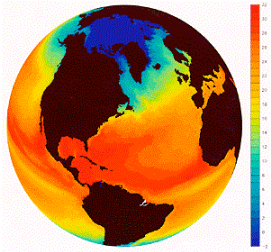


The Computational Research Division is engaged in computational science collaborations, creating tools and techniques that will enable computational modeling and simulation, and lead to new understanding in areas such as

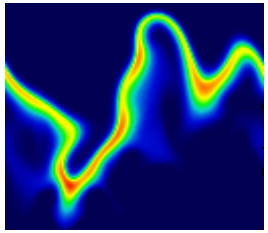


astrophysics

simulation



global climate



combustion

processes

Parallel Computing Short Course – offered by LBNL

- Aug 18-21 – now!
  - Sutardja Dai Hall (CITRIS Building)
  - acts.nersc.gov/events/Workshop2009/
- How to use selected computational tools developed for high performance computing
  - ScaLAPACK, PETSc, Hypre, Zoltan, GlobalArrays, …
- Feel free to visit (their web site)

Schedule and Instructors (1/3)

- Wednesday, Aug 19
  - 9-9:10 am – Introduction and Welcome
    - » Dave Patterson and Jim Demmel (UCB)
  - 9:30-12pm – Introduction to parallel computer architecture
    - » John Kubiatowicz (UCB)
  - 12-1:15pm – Lunch (see web page for suggested venues)

  - 1:15-2:15pm – Shared Memory Programming with PThreads
    - » Kathy Yelick (UCB & LBNL)
  - 2:15-3:15pm – Shared Memory Programming with OpenMP
    - » Tim Mattson (Intel)
  - 3:15-3:45pm – Break
  - 3:45-4:30pm – Shared Memory programming with TBB
    - » Michael Wrinn (Intel)
  - 4:30-4:45pm – Challenges in composing these models
    - » Heidi Pan (UCB)
  - 4:45-5:00pm – How to use Franklin
    - » Katie Antypas (LBNL)
  - 5-6:00pm – Hands-on Lab
    - » All Rooms in Soda Hall: 310, 320, 373, 380, 405, Wozniak Lounge (430)
  - 6-7:30pm – Party in Wozniak Lounge!

Schedule and Instructors (2/3)

- Thursday Aug 20
  - 8:45-9:45am – Sources of Parallelism and Locality in Simulation
    - » Jim Demmel (UCB)
  - 9:45-10:45am – Architecting Parallel Software Using Design Patterns
    - » Kurt Keutzer (UCB)
  - 10:45-11:15am – Break
  - 11:15-12:15pm – Data-Parallel Programming on Manycore Graphics Processors
    - » Bryan Catanzaro (UCB)
  - 12:15-1:30pm – Lunch
  - 1:30-2:30pm – OpenCL
    - » Tim Mattson (Intel)
  - 3-6pm – Hands-on Lab
    - » Same rooms as yesterday

Schedule and Instructors (3/3)

- Friday Aug 21
  - 8:45-10:45am – Computational Patterns of Parallel Programming
    - » Jim Demmel (UCB)
  - 10:45-11:15am – Break
  - 11:15-12:15pm – Building Parallel Applications
    - » Ras Bodik, Tony Keaveny, Nelson Morgan, David Wessel (UCB)
  - 12:15-1:30pm – Lunch

  - 1:30-2:30pm – Distributed Memory Programming in MPI and UPC
    - » Kathy Yelick (UCB & LBNL)
  - 2:30–3:30pm – Performance Analysis Tools
    - » Karl Fuerlinger (UCB)
  - 3:30-4pm - Break
  - 4-5pm – Cloud Computing
    - » Matei Zaharia (UCB)

Logistics

- Coffee
  - Available in 120A Bechtel, not allowed in lecture hall!
- Live webcast of lectures
  - mms://media.citris.berkeley.edu/parlab09
  - Email questions to parlab.bootcamp.2009@gmail.com
- Lecture Materials
  - Slides and archived video will be posted on bootcamp website
- Labs
  - Bring your own laptop
  - We supply wireless access, accounts on Franklin (Opteron Cluster)
    - » Account problems: ask your TA
      - Kaushik Datta (Head TA), Grey Ballard, Rajesh Nishtala, Razvan Carbunescu, Marghoob Mohiyuddin, Andrew Gearhart, Shoaib Kamil, Sarah Bird, Yunsup Lee
  - Lab assignment(s) posted at www.cs.berkeley.edu/~volkov/cs267.sp09
    - » Used in CS267, we'll try Assignment 2 in OpenMP and Pthreads

The Audience - you

- 184 on-site, 208 off-site registrants
  - 102 Students, 21 research staff/postdocs, 10 faculty, 18 LBNL
  - 21 ParLab affiliate members, 12 "other"
- How much do you know about parallel programming?
  - 60% novice, 36% somewhat knowledgeable, 6% expert
- What kind of parallel computing are you most interested in? (click all that apply)
  - 71% multicore, 53% distributed memory, 56% GPU, 48% Cloud, 64% mixed
- What are your goals?
  - Learn practical programming techniques 83%
  - Apply parallelism to particular applications 54%
  - Use parallelism in a business environment 29%
  - Use parallelism in an academic or research environment 63%
  - Build parallel programming tools 45%
  - Learn about open research questions in parallelism 57%

LET'S GET STARTED!