Design of Parallel and High-Performance Computing
Fall 2013

Lecture: Introduction

Instructor: Torsten Hoefler & Markus Püschel
TA: Timo Schneider
Goals of this lecture

- Motivate you!
- What is parallel computing?
  - And why do we need it?
- What is high-performance computing?
  - What’s a Supercomputer and why do we care?
- Basic overview of
  - Programming models
    - Some examples
  - Architectures
    - Some case-studies
- Provide context for coming lectures
Let us assume ...

... you were to build a machine like this ...

... we know how each part works
  - There are just many of them!
  - Question: How many calculations per second are needed to emulate a brain?

Source: wikipedia
Exponential Growth of Computing
Twentieth through twenty first century

Logarithmic Plot

Science
Researchers Simulate Mouse Brain on Computer

Michael Hoffman [Email] - April 30, 2007 5:57 PM

Can we do this today?
Growth in Supercomputer Power

Blue Waters, ~13 PF (2012)

Tianhe-2, ~55 PF (2013)

1 Exaflop! ~2022?

Blue Waters, ~13 PF (2012)

Source: www.singularity.com
Human Brain – No Problem!

- ... not so fast, we need to understand how to program those machines ...
Human Brain – No Problem!

Simulating 1 second of human brain activity takes 82,944 processors

By Ryan Whitwam on August 5, 2013 at 1:34 pm | 21 Comments

The brain is a deviously complex biological computing device that even the fastest supercomputers in the world fail to emulate. Well, that’s not entirely true anymore. Researchers at the Okinawa Institute of Technology Graduate University in Japan and Forschungszentrum Jülich in Germany have managed to simulate a single second of human brain activity in a very, very powerful computer.

Source: extremetech.com
Other problem areas: Scientific Computing

- Most natural sciences are simulation driven and are moving towards simulation
  - Theoretical physics (solving the Schrödinger equation, QCD)
  - Biology (Gene sequencing)
  - Chemistry (Material science)
  - Astronomy (Colliding black holes)
  - Medicine (Protein folding for drug discovery)
  - Meteorology (Storm/Tornado prediction)
  - Geology (Oil reservoir management, oil exploration)
  - and many more ... (even Pringles uses HPC)
Other problem areas: Commercial Computing

- Databases, data mining, search
  - Amazon, Facebook, Google

- Transaction processing
  - Visa, Mastercard

- Decision support
  - Stock markets, Wall Street, Military applications

- Parallelism in high-end systems and back-ends
  - Often throughput-oriented
  - Used equipment varies from COTS (Google) to high-end redundant mainframes (banks)
Other problem areas: Industrial Computing

- Aeronautics (airflow, engine, structural mechanics, electromagnetism)
- Automotive (crash, combustion, airflow)
- Computer-aided design (CAD)
- Pharmaceuticals (molecular modeling, protein folding, drug design)
- Petroleum (Reservoir analysis)
- Visualization (all of the above, movies, 3d)
What can faster computers do for us?

- **Solving bigger problems than we could solve before!**
  - E.g., Gene sequencing and search, simulation of whole cells, mathematics of the brain, ...
  - The size of the problem grows with the machine power
    -> *Weak Scaling*

- **Solve small problems faster!**
  - E.g., large (combinatorial) searches, mechanical simulations (aircrafts, cars, weapons, ...)
  - The machine power grows with constant problem size
    -> *Strong Scaling*
High-Performance Computing (HPC)

- a.k.a. “Supercomputing”
- Question: define “Supercomputer”!
High-Performance Computing (HPC)

- a.k.a. “Supercomputing”

- **Question: define “Supercomputer”!**
  - “A supercomputer is a computer at the frontline of contemporary processing capacity--particularly speed of calculation.” (Wikipedia)
  - Usually quite expensive ($s and kWh) and big (space)

- **HPC is a quickly growing niche market**
  - Not all “supercomputers”, wide base
  - Important enough for vendors to specialize
  - Very important in research settings (up to 40% of university spending)

  “Goodyear Puts the Rubber to the Road with High Performance Computing”
  “High Performance Computing Helps Create New Treatment For Stroke Victims”
  “Procter & Gamble: Supercomputers and the Secret Life of Coffee”
  “Motorola: Driving the Cellular Revolution With the Help of High Performance Computing”
  “Microsoft: Delivering High Performance Computing to the Masses”
The Top500 List

- **A benchmark, solve Ax=b**
  - As fast as possible! → as big as possible 😊
  - Reflects **some** applications, not all, not even many
  - Very good historic data!

- **Speed comparison for computing centers, states, countries, nations, continents 😞**
  - Politicized (sometimes good, sometimes bad)
  - Yet, fun to watch
<table>
<thead>
<tr>
<th>Rank</th>
<th>Site</th>
<th>System</th>
<th>Cores</th>
<th>Core (Tflop/s)</th>
<th>Rpeak (Tflop/s)</th>
<th>Power (kW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>National University of Defense Technology</td>
<td>Tianhe-2 (MilkyWay-2) - TH-IVB-FEP Cluster, Intel</td>
<td>312000</td>
<td>33862.7</td>
<td>54902.4</td>
<td>17808</td>
</tr>
<tr>
<td></td>
<td>China</td>
<td>Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Xeon Phi 31S1P NUDT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>DOE/SC/Oak Ridge National Laboratory</td>
<td>Titan - Cray XK7 , Opteron 6274 16C 2.200GHz, Cray</td>
<td>560640</td>
<td>17590.0</td>
<td>27112.5</td>
<td>8209</td>
</tr>
<tr>
<td></td>
<td>United States</td>
<td>Gemini interconnect, NVIDIA K20x Cray Inc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>DOE/NNSA/LLNL</td>
<td>Sequoia - BlueGene/Q, Power BQC 16C 1.60 GHz,</td>
<td>1572864</td>
<td>17173.2</td>
<td>20132.7</td>
<td>7890</td>
</tr>
<tr>
<td></td>
<td>United States</td>
<td>Custom IBM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>RIKEN Advanced Institute for Computational</td>
<td>K computer, SPARC64 VIII fx 2.00GHz, Tofu</td>
<td>705024</td>
<td>10510.0</td>
<td>11280.4</td>
<td>12660</td>
</tr>
<tr>
<td></td>
<td>Science (AICS)</td>
<td>interconnect</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Japan</td>
<td>Fujitsu</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>DOE/SC/Argonne National Laboratory</td>
<td>Mira - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom</td>
<td>786432</td>
<td>8586.6</td>
<td>10066.3</td>
<td>3945</td>
</tr>
<tr>
<td></td>
<td>United States</td>
<td>IBM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Texas Advanced Computing Center/Univ. of</td>
<td>Stampede - PowerEdge C8220, Xeon E5-2680 8C</td>
<td>462462</td>
<td>5168.1</td>
<td>8520.1</td>
<td>4510</td>
</tr>
<tr>
<td></td>
<td>Texas</td>
<td>2.7000GHz, Infiniband FDR, Intel Xeon Phi SE10P</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>United States</td>
<td>Dell</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Forschungszentrum Juelich (FZJ)</td>
<td>Swiss Scientific Computing Center (CSCS)</td>
<td>42</td>
<td>35340</td>
<td>626.9</td>
<td>745.5</td>
</tr>
<tr>
<td></td>
<td>Germany</td>
<td>Piz Daint - Cray XC30, Xeon E5-2670 8C 2.600GHz,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Aries Interconnect</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cray Inc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>DOE/NNSA/LLNL</td>
<td>Vulcan - BlueGene/Q, Power BQC 16C 1.600GHz,</td>
<td>393216</td>
<td>4293.3</td>
<td>5033.2</td>
<td>1972</td>
</tr>
<tr>
<td></td>
<td>United States</td>
<td>Custom Interconnect</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Leibniz Rechenzentrum</td>
<td>SuperMUC - iDataPlex DX360M4, Xeon E5-2680 8C</td>
<td>147456</td>
<td>2897.0</td>
<td>3185.1</td>
<td>3423</td>
</tr>
<tr>
<td></td>
<td>Germany</td>
<td>2.70GHz, Infiniband FDR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Piz Daint @ CSCS
March 19, 2013

Swiss 'GPU Supercomputer' Will Be Fastest in Europe

Tiffany Trader

The NVIDIA GPU Technology Conference is in full-swing today in San Jose, Calif. The annual event kicked off this morning with a keynote from NVIDIA CEO Jen-Hsun Huang, who revealed that the Swiss National Supercomputing Center (CSCS) is building Europe's fastest GPU-accelerated supercomputer, an extension of a Cray system that was announced last year.

As Cray Vice President, Storage & Data Management Barry Bolding told HPCwire, this will be the first Cray supercomputer equipped with Intel Xeon processors and NVIDIA GPUs.

CSCS is part of ETH Zurich, one of the top universities in the world and the alma mater of Albert Einstein. The supercomputing center installed phase one of its shiny new Cray XC30 back in December 2012.
Blue Waters in 2009

Imagine you’re designing a $500 M supercomputer, and all you have is:

This is why you need to understand performance expectations well!
Blue Waters in 2012
History and Trends

Source: Jack Dongarra

Single GPU/MIC Card

162 PFlop/s
17.6 PFlop/s
76.5 TFlop/s

6-8 years

My Laptop (70 Gflop/s)

My iPad2 & iPhone 4s (1.02 Gflop/s)
High-Performance Computing grows quickly

- Computers are used to automate many tasks
- Still growing exponentially
  - New uses discovered continuously

IDC, 2007: “The overall HPC server market grew by 15.5 percent in 2007 to reach $11.6 billion [...] while the same kinds of boxes that go into HPC machinery but are used for general purpose computing, rose by only 3.6 percent to $54.4”

IDC, 2009: “expects the HPC technical server market to grow at a healthy 7% to 8% yearly rate to reach revenues of $13.4 billion by 2015.”

“The non-HPC portion of the server market was actually down 20.5 per cent, to $34.6bn”
How to increase the compute power?

Clock Speed:

Power Density (W/cm²)

Source: Intel®
How to increase the compute power?

Not an option anymore!

Clock Speed:

Source: Intel®
Microprocessor Transistor Counts 1971-2011 & Moore’s Law

The graph shows the growth in transistor count every two years from 1971 to 2011, highlighting the exponential growth described by Moore’s Law. The curve indicates that the number of transistors doubles every two years, reflecting technological advancements in semiconductor manufacturing.

So how to invest the transistors?

- **Architectural innovations**
  - Branch prediction, Tomasulo logic/rename register, speculative execution, ...
  - Help only so much 😞

- **What else?**
  - Simplification is beneficial, less transistors per CPU, more CPUs, e.g., Cell B.E., GPUs, MIC
  - We call this “cores” these days
  - Also, more intelligent devices or higher bandwidths (e.g., DMA controller, intelligent NICs)
Towards the age of massive parallelism

- Everything goes parallel
  - Desktop computers get more cores
    2, 4, 8, soon dozens, hundreds?
  - Supercomputers get more PEs (cores, nodes)
    > 3 million today
    > 50 million on the horizon
    ➢ 1 billion in a couple of years (after 2020)

- Parallel Computing is inevitable!

**Parallel vs. Concurrent computing**

Concurrent activities *may* be executed in parallel

Example:
A1 starts at T1, ends at T2; A2 starts at T3, ends at T4
Intervals (T1,T2) and (T3,T4) may overlap!

Parallel activities:
A1 is executed *while* A2 is running
Usually requires separate resources!
Goals of this lecture

- Motivate you!
- What is parallel computing?
  - And why do we need it?
- What is high-performance computing?
  - What’s a Supercomputer and why do we care?
- Basic overview of
  - Programming models
    - Some examples
  - Architectures
    - Some case-studies
- Provide context for coming lectures
# Granularity and Resources

<table>
<thead>
<tr>
<th>Activities</th>
<th>Parallel Resource</th>
</tr>
</thead>
<tbody>
<tr>
<td>▪ Micro-code instruction</td>
<td>▪ Instruction-level parallelism</td>
</tr>
<tr>
<td>▪ Machine-code instruction (complex or simple)</td>
<td>▪ Pipelining</td>
</tr>
<tr>
<td>▪ Sequence of machine-code instructions:</td>
<td>▪ VLIW</td>
</tr>
<tr>
<td>&quot;Blocks&quot;</td>
<td>▪ Superscalar</td>
</tr>
<tr>
<td>&quot;Loops&quot;</td>
<td>▪ SIMD operations</td>
</tr>
<tr>
<td>&quot;Loop nests&quot;</td>
<td>▪ Vector operations</td>
</tr>
<tr>
<td>&quot;Functions&quot;</td>
<td>▪ Instruction sequences</td>
</tr>
<tr>
<td>&quot;Function sequences&quot;</td>
<td>▪ Multiprocessors</td>
</tr>
</tbody>
</table>

- Multicores
- Multithreading
## Resources and Programming

<table>
<thead>
<tr>
<th>Parallel Resource</th>
<th>Programming</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instruction-level parallelism</td>
<td>Compiler</td>
</tr>
<tr>
<td>Pipelining</td>
<td>(inline assembly)</td>
</tr>
<tr>
<td>VLIW</td>
<td>Hardware scheduling</td>
</tr>
<tr>
<td>Superscalar</td>
<td>Compiler (inline assembly)</td>
</tr>
<tr>
<td>SIMD operations</td>
<td>Libraries</td>
</tr>
<tr>
<td>Vector operations</td>
<td>Compilers (very limited)</td>
</tr>
<tr>
<td>Instruction sequences</td>
<td>Expert programmers</td>
</tr>
<tr>
<td>Multiprocessors</td>
<td>Parallel languages</td>
</tr>
<tr>
<td>Multicores</td>
<td>Parallel libraries</td>
</tr>
<tr>
<td>Multithreading</td>
<td>Hints</td>
</tr>
</tbody>
</table>

- Instruction-level parallelism
- Pipelining
- VLIW
- Superscalar
- SIMD operations
- Vector operations
- Instruction sequences
- Multiprocessors
- Multicores
- Multithreading
- Compiler
- (inline assembly)
- Hardware scheduling
- Compiler (inline assembly)
- Libraries
- Compilers (very limited)
- Expert programmers
  - Parallel languages
  - Parallel libraries
  - Hints
Historic Architecture Examples

- **Systolic Array**
  - Data-stream driven (data counters)
  - Multiple streams for parallelism
  - Specialized for applications (reconfigurable)

- **Dataflow Architectures**
  - No program counter, execute instructions when all input arguments are available
  - Fine-grained, high overheads

Example: compute $f = (a+b) \times (c+d)$
**Von Neumann Architecture**

- **Program counter → Inherently serial!**
  Retrospectively define parallelism in instructions and data

<table>
<thead>
<tr>
<th>SISD</th>
<th>SIMD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Serial Computer (nearly extinct)</td>
<td>Vector Machines or Extensions (very common)</td>
</tr>
<tr>
<td>MISD</td>
<td>MIMD</td>
</tr>
<tr>
<td>Redundant Execution (fault tolerance)</td>
<td>Multicore (ubiquitous)</td>
</tr>
</tbody>
</table>
Parallel Architectures 101

- Today’s laptops
- Today’s servers
- Yesterday’s clusters
- Today’s clusters
- ... and mixtures of those
Programming Models

- **Shared Memory Programming (SM/UMA)**
  - Shared address space
  - Implicit communication
  - Hardware for cache-coherent remote memory access
  - Cache-coherent Non Uniform Memory Access (cc NUMA)

- **(Partitioned) Global Address Space (PGAS)**
  - Remote Memory Access
  - Remote vs. local memory (cf. ncc-NUMA)

- **Distributed Memory Programming (DM)**
  - Explicit communication (typically messages)
  - Message Passing
Shared Memory Machines

- Two historical architectures:
  - “Mainframe” – all-to-all connection between memory, I/O and PEs
    - Often used if PE is the most expensive part
    - Bandwidth scales with P
    - PE Cost scales with P, Question: what about network cost?

Source: IBM
Shared Memory Machines

- **Two historical architectures:**
  - “Mainframe” – all-to-all connection between memory, I/O and PEs
    - *Often used if PE is the most expensive part*
    - Bandwidth scales with P
    - *PE Cost scales with P, Question: what about network cost?*
    - **Answer:** Cost can be cut with multistage connections (butterfly)
  - “Minicomputer” – bus-based connection
    - *All traditional SMP systems*
    - High latency, low bandwidth *(cache is important)*
    - Tricky to achieve highest performance *(contention)*
    - Low cost, extensible

Source: IBM
Shared Memory Machine Abstractions

- Any PE can access all memory
  - Any I/O can access all memory (maybe limited)

- OS (resource management) can run on any PE
  - Can run multiple threads in shared memory
  - Used since 40+ years

- Communication through shared memory
  - Load/store commands to memory controller
  - Communication is implicit
  - Requires coordination

- Coordination through shared memory
  - Complex topic
  - Memory models
Shared Memory Machine Programming

- **Threads or processes**
  - Communication through memory

- **Synchronization through memory or OS objects**
  - Lock/mutex (protect critical region)
  - Semaphore (generalization of mutex (binary sem.))
  - Barrier (synchronize a group of activities)
  - Atomic Operations (CAS, Fetch-and-add)
  - Transactional Memory (execute regions atomically)

- **Practical Models:**
  - Posix threads
  - MPI-3
  - OpenMP
  - Others: Java Threads, Qthreads, ...
An SMM Example: Compute Pi

- Using Gregory-Leibnitz Series:

\[ 4 \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \]

- Iterations of sum can be computed in parallel
- Needs to sum all contributions at the end

Source: mathworld.wolfram.com
int main(int argc, char *argv[])
{
    // definitions ...
    thread_arr = (pthread_t*)malloc(nthreads * sizeof(pthread_t));
    resultarr = (double*)malloc(nthreads * sizeof(double));

    for (i=0; i<nthreads; ++i) {
        int ret = pthread_create(&thread_arr[i], NULL,
                                compute_pi, (void*) i);
    }
    for (i=0; i<nthreads; ++i) {
        pthread_join(thread_arr[i], NULL);
    }
    pi = 0;
    for (i=0; i<nthreads; ++i) pi += resultarr[i];

    printf("pi is approximately %.16f, Error is %.16f\n", pi, fabs(pi - PI25DT));
}

int n=10000;
double *resultarr;
int nthreads;

void *compute_pi(void *data) {
    int i, j;
    int myid = (int)(long)data;
    double mypi, h, x, sum;

    for (j=0; j<n; ++j) {
        h = 1.0 / (double) n;
        sum = 0.0;
        for (i = myid + 1; i <= n; i += nthreads) {
            x = h * ((double) i - 0.5);
            sum += (4.0 / (1.0 + x*x));
        }
        mypi = h * sum;
    }
    resultarr[myid] = mypi;
}
Additional comments on SMM

- OpenMP would allow to implement this example much simpler (but has other issues)

- Transparent shared memory has some issues in practice:
  - False sharing (e.g., resultarr[])
  - Race conditions (complex mutual exclusion protocols)
  - Little tool support (debuggers need some work)

- Achieving performance is harder than it seems!
Distributed Memory Machine Programming

- Explicit communication between PEs
  - Message passing or channels

- Only local memory access, no direct access to remote memory
  - No shared resources (well, the network)

- Programming model: Message Passing (MPI, PVM)
  - Communication through messages or group operations (broadcast, reduce, etc.)
  - Synchronization through messages (sometimes unwanted side effect) or group operations (barrier)
  - Typically supports message matching and communication contexts
Send specifies buffer to be transmitted

Recv specifies buffer to receive into

Implies copy operation between named PEs

Optional tag matching

Pair-wise synchronization (cf. happens before)
int main( int argc, char *argv[] ) {  
    // definitions  
    MPI_Init(&argc,&argv);  
    MPI_Comm_size(MPI_COMM_WORLD, &numprocs);  
    MPI_Comm_rank(MPI_COMM_WORLD, &myid);  

    double t = -MPI_Wtime();  
    for (j=0; j<n; ++j) {  
        h = 1.0 / (double) n;  
        sum = 0.0;  
        for (i = myid + 1; i <= n; i += numprocs) {  
            x = h * ((double)i - 0.5);  
            sum += (4.0 / (1.0 + x*x));  
        }  
        mypi = h * sum;  
        MPI_Reduce(&mypi, &pi, 1, MPI_DOUBLE, MPI_SUM, 0, MPI_COMM_WORLD);  
    }  
    t+=MPI_Wtime();  

    if (!myid) {  
        printf("pi is approximately %.16f, Error is %.16f\n", pi, fabs(pi - PI25DT));  
        printf("time: %f\n", t);  
    }  

    MPI_Finalize();  
}
DMM Example: PGAS

- **Partitioned Global Address Space**
  - Shared memory emulation for DMM
    - *Usually non-coherent*
  - “Distributed Shared Memory”
    - *Usually coherent*

- **Simplifies shared access to distributed data**
  - Has similar problems as SMM programming
  - Sometimes lacks performance transparency
    - *Local vs. remote accesses*

- **Examples:**
  - UPC, CAF, Titanium, X10, ...
How to Tame the Beast?

- **How to program large machines?**
- **No single approach, PMs are not converging yet**
  - MPI, PGAS, OpenMP, Hybrid (MPI+OpenMP, MPI+MPI, MPI+PGAS?), ...
- **Architectures converge**
  - General purpose nodes connected by general purpose or specialized networks
  - Small scale often uses commodity networks
  - Specialized networks become necessary at scale
- **Even worse: accelerators (not covered in this class, yet)**
Practical SMM Programming: Pthreads

Covered in example, small set of functions for thread creation and management.
Practical SMM Programming:

- Fork-join model

Types of constructs:

1. Fork-join model

2. Parallel region

3. Task constructs:
   - Fork
   - Join
   - Do (for loop)
   - Sections
   - Single

Source: OpenMP.org

Source: Blaise Barney, LLNL
#include <omp.h>

main () {
    int var1, var2, var3;
    // Serial code

    // Beginning of parallel section. Fork a team of threads. Specify variable scoping
    #pragma omp parallel private(var1, var2) shared(var3)
    {
        // Parallel section executed by all threads
        // Other OpenMP directives
        // Run-time Library calls
        // All threads join master thread and disband
    }
    // Resume serial code
}
Practical PGAS Programming: UPC

- PGAS extension to the C99 language

- Many helper library functions
  - Collective and remote allocation
  - Collective operations

- Complex consistency model
Practical DMM Programming: MPI-1

Collection of 1D address spaces

Helper Functions

- broadcast
- scatter
- gather
- reduction

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>(0,0)</td>
<td>(0,1)</td>
<td>(0,2)</td>
<td>(0,3)</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>(1,0)</td>
<td>(1,1)</td>
<td>(1,2)</td>
<td>(1,3)</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>(2,0)</td>
<td>(2,1)</td>
<td>(2,2)</td>
<td>(2,3)</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>(3,0)</td>
<td>(3,1)</td>
<td>(3,2)</td>
<td>(3,3)</td>
</tr>
</tbody>
</table>

many more (>600 total)

Source: Blaise Barney, LLNL
Complete Six Function MPI-1 Example

```c
#include <mpi.h>

int main(int argc, char **argv) {
    int myrank, sbuf=23, rbuf=32;
    MPI_Init(&argc, &argv);

    /* Find out my identity in the default communicator */
    MPI_Comm_rank(MPI_COMM_WORLD, &myrank);

    if (myrank == 0) {
        MPI_Send(&sbuf, 1, MPI_INT, rank, 99, MPI_COMM_WORLD);
    } else {
        MPI_Recv(&rbuf, MPI_DOUBLE, 0, 99, MPI_COMM_WORLD, &status);
        printf("received: %i\n", rbuf);
    }

    MPI_Finalize();
}
```
MPI-2/3: Greatly enhanced functionality

- Support for shared memory in SMM domains

- Support for Remote Memory Access Programming
  - Direct use of RDMA
  - Essentially PGAS

- Enhanced support for message passing communication
  - Scalable topologies
  - More nonblocking features
  - ... many more
Accelerator example: CUDA

Hierarchy of Threads

Simple Architecture

Kepler Block Diagram
- 8 SMX
- 1536 CUDA Cores
- 8 Geometry Units
- 4 Raster Units
- 128 Texture Units
- 32 ROP units
- 256-bit GDDR5

Complex Memory Model

Source: NVIDIA
Accelerator example: CUDA

Host Code

```c
#define N 10
int main( void ) {
  int a[N], b[N], c[N];
  int *dev_a, *dev_b, *dev_c;
  // allocate the memory on the GPU
  cudaMalloc( (void**)&dev_a, N * sizeof(int) );
  cudaMalloc( (void**)&dev_b, N * sizeof(int) );
  cudaMalloc( (void**)&dev_c, N * sizeof(int) );
  // fill the arrays 'a' and 'b' on the CPU
  for (int i=0; i<N; i++) { a[i] = -i; b[i] = i * i;  }
  // copy the arrays 'a' and 'b' to the GPU
  cudaMemcpy( dev_a, a, N * sizeof(int), cudaMemcpyHostToDevice );
  cudaMemcpy( dev_b, b, N * sizeof(int), cudaMemcpyHostToDevice );
  add<<<N,1>>>( dev_a, dev_b, dev_c );
  // copy the array 'c' back from the GPU to the CPU
  cudaMemcpy( c, dev_c, N * sizeof(int), cudaMemcpyDeviceToHost );
  // free the memory allocated on the GPU
  cudaFree( dev_a ); cudaFree( dev_b ); cudaFree( dev_c );
}
```

The Kernel

```c
__global__ void add( int *a, int *b, int *c ) {
  int tid = blockIdx.x;
  // handle the data at this index
  if (tid < N)
    c[tid] = a[tid] + b[tid];
}
```
OpenACC / OpenMP 4.0

- Aims to simplify GPU programming
- Compiler support
  - Annotations!

```c
#define N 10
int main( void ) {
  int a[N], b[N], c[N];
  #pragma acc kernels
  for (int i = 0; i < N; ++i)
    c[i] = a[i] + b[i];
}
```
More programming models/frameworks

- **Not covered:**
  - SMM: Intel Cilk / Cilk Plus, Intel TBB, ...
  - Directives: OpenHMPP, PVM, ...
  - PGAS: Coarray Fortran (Fortran 2008), ...
  - HPCS: IBM X10, Fortress, Chapel, ...
  - Accelerator: OpenCL, C++AMP, ...

- **This class will not describe any model in more detail!**
  - There are too many and they will change quickly (only MPI made it >15 yrs)

- **No consensus, but fundamental questions remain:**
  - Data movement
  - Synchronization
  - Memory Models
  - Algorithmics
  - Foundations
Goals of this lecture

- Motivate you!

- What is parallel computing?
  - And why do we need it?

- What is high-performance computing?
  - What’s a Supercomputer and why do we care?

- Basic overview of
  - Programming models
    - Some examples
  - Architectures
    - Some case-studies

- Provide context for coming lectures
DPHPC Lecture

- You will most likely not have access to the largest machines
  - But our desktop/laptop will be a “large machine” soon
  - HPC is often seen as “Formula 1” of computing (architecture experiments)

- DPHPC will teach you concepts!
  - Enable to understand and use all parallel architectures
  - From a quad-core mobile phone to the largest machine on the planet!
    - MCAPI vs. MPI – same concepts, different syntax
  - No particular language (but you should pick/learn one for your project!)
    - Parallelism is the future:
DPHPC Overview

- locality
  - caches
  - memory hierarchy
- parallelism
  - vector ISA
  - shared memory
  - distributed memory
- cache coherency
- memory models
  - locks
  - lock free
  - wait free
  - linearizability
- distributed algorithms
- group communications

models
- Amdahl's and Gustafson's law
- memory $\alpha - \beta$
- PRAM
- LogP
- I/O complexity
- balance principles I
- Little's Law
- balance principles II
- scheduling