

# **INTRODUCTION TO GPU COMPUTING FOR STATISTICIANS**

Will Landau, Matt Simpson, Prof. Jarad Niemi

# WHY GPU COMPUTING?

The whole point is to use new hardware (the GPU) in order to make long, repetitive calculations run faster.

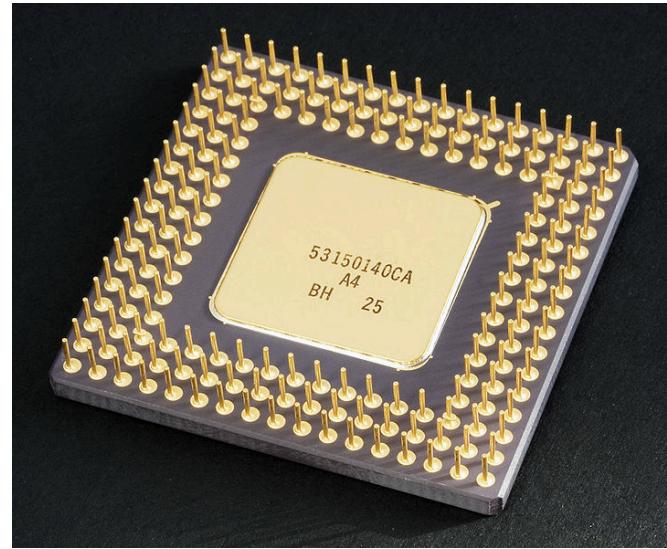
# COMPUTER PROCESSORS

The computer processor is the "brain" of a computer.  
It's a computer chip that is responsible for *actually doing* calculations and performing executive functions.

# THE CPU

**CPU** = “Central Processing Unit”

- The kind of processor you would find in a regular computer.
- Designed for general purpose computing.
- Does parallel computing, but not very fast.



# THE GPU

**GPU** = “Graphics Processing Unit”

- The kind of processor that you would find in a graphics card or video card.
- Originally designed to speed up graphics throughput in video games, not to do general purpose computing.
- Performs massively parallel computing, able to run orders of magnitude more threads at a time than a CPU.
- Higher memory bandwidth than the CPU.



# GPUS WITH CUDA: COMPUTE UNIFIED DEVICE ARCHITECTURE

- First released by NVIDIA in 2007
- Supports CUDA C, an extension of C for programs that can run on GPUs and CPUs simultaneously.

CUDA systems have the data crunching power of the GPU and the versatility of the CPU.

## WE HAVE CUDA SYSTEMS!

- impact1.stat.iastate.edu (up and running)
- impact2.stat.iastate.edu (coming soon)
- impact3.stat.iastate.edu (coming soon)
- impact4.stat.iastate.edu (coming soon)

# SPECS OF IMPACT1

- linux
- no GUI or remote desktop capabilities yet (use the command line for now)
- Four CUDA-capable Tesla M2070 GPUs, each with:
  - 448 cores.
  - CUDA Driver and Runtime Version 4.1

Enter:

```
cd /usr/local/NVIDIA_GPU_Computing_SDK/C/bin/linux/release
```

and then:

```
./deviceQuery
```

in the command line while logged into impact1 for more details.

# LOGGING INTO IMPACT1

1. Connect to the internet and open your favorite command line utility:  
Terminal in Mac OS X, Command Prompt in Windows, etc.
2. Type in:

```
ssh -p 323 your_ISU_ID@impact1.stat.iastate.edu
```

and press enter.

For me, a login looks like this:

```
~> ssh -p 323 landau@impact1.stat.iastate.edu
Last login: Mon May 28 11:37:06 2012 from landau.student.iastate.edu
KRB5CCNAME: Undefined variable.
[landau@impact1 ~]$
```

**MAKE SURE TO USE PORT 323!**

Contact me (at landau@iastate.edu or in person) if you'd like help with:

- Command line tools for logging in.
- Easy ways to transfer files between impact1 and your local machine.
- SSH key setup: for logging in from your personal machine without having to type your password.
- Setting up a shortcut command for logging in so that you don't have to type in all of “ssh -p 323 your\_ISU\_ID@impact1.stat.iastate.edu” every time you log in.

For questions about using command line tools or the linux file system in general, contact me or see:

- <http://www.makeuseof.com/tag/an-introduction-to-the-linux-command-line/>
- [http://www.freesoftwaremagazine.com/articles/command\\_line\\_intro](http://www.freesoftwaremagazine.com/articles/command_line_intro)
- <http://tldp.org/LDP/intro-linux/html/>
- [http://tldp.org/LDP/intro-linux/html/sect\\_03\\_01.html](http://tldp.org/LDP/intro-linux/html/sect_03_01.html)
- <http://dhavalv.wordpress.com/2007/10/17/quick-introduction-to-linux-filesystem-fhs/>
- [http://linux.die.net/Intro-Linux/chap\\_03.html](http://linux.die.net/Intro-Linux/chap_03.html)
- [http://linux.about.com/od/itl\\_guide/a/gdeitl28t02.htm](http://linux.about.com/od/itl_guide/a/gdeitl28t02.htm)

# IMPORTANT DIRECTORIES ON IMPACT1

- `/home/your_ISU_ID`

Your private home folder on the department's linux repository (also connects with linux10 and linux11). Code and data in here are stored remotely on the linux repository but used locally with the hardware in impact1.

- `/Cyfiles/your_ISU_ID`

Your private Cyfiles folder. Code and data in here are stored remotely on the university's Cyfiles system but run locally on impact1.

- `/tmp`

Everything in here is stored locally on impact1. To ensure fast computation, put your huge data set here. That way, your program doesn't have to stream lots of data through a network. **WARNING:** `/tmp` automatically empties periodically.

- **/usr/local/NVIDIA\_GPU\_Computing\_SDK**

Contains example code for those of you who want to learn CUDA C.