

CSC266/ECE206 Introduction to Parallel Computing using GPUs

Sreepathi Pai

University of Rochester

September 6, 2017

- 1 Organization
- 2 Performance Metrics
- 3 Program Optimization

- 1 Organization
- 2 Performance Metrics
- 3 Program Optimization

- Lectures: Dr. Sreepathi Pai
 - E-mail: sree@cs.rochester.edu
 - Office: Wegmans 3409
 - Office Hours: By appointment
- Labs: Dr. Alex Page
 - E-mail: alex.page@rochester.edu

- Class: CSB 523
 - M, W 1650–1805
- Course Website:
 - <https://cs.rochester.edu/~sree/fall-2017/csc-266>
- Blackboard:
 - Announcements, Assignments, etc.
- Piazza:
 - TBA

- No required textbook for the class
- Useful to have a book on architecture as a reference
 - But this is not a computer architecture class
- Links to manuals, papers, etc. will be provided
 - Feel free to search for them

Project Expectation

You will demonstrate your mastery of the course goals.
Specifically, for a program, you will:

- Identify parallelization opportunities
- Implement programs on the GPU
- Optimize programs on the GPU

- 1 Organization
- 2 Performance Metrics
- 3 Program Optimization

Metrics we're interested in

- Latency
 - Time units: $1\mu\text{s}$ or 1000000 cycles
 - Lower is better
- Throughput
 - Rate: FLOPS or Instructions per Cycle (IPC)
 - Higher is better
- Other interesting performance metrics
 - Power (Watt)
 - Energy (Joule)

Applications where latency is crucial

- Audio/Video
 - MP3 players
 - MPEG4 players
 - VoIP (e.g. Skype)
- Games
 - Multi-user gameplay
 - Responsiveness
- Servers
 - Search Engines
 - Web applications

Applications where throughput is crucial

- Audio/Video
 - MP3 encoders
 - MPEG4 encoders
- Games
 - Frame rate
- Scientific Applications
 - Molecular Dynamics
 - Finite-element Code
- Servers
 - Search Engines
 - Web applications

Better performance can open up new vistas



Google

Search bar



@gmail.com

Gmail

Navigation buttons: Home, Refresh, More

1-3 of 3 | Settings

Compose

- ☆ Gmail Team Stay more organised with Gmail's inbox - 15 Gmail's inbox puts you in control Inbox video Meet the inbox Gmail's inbox sorts 16 Jul
- ☆ Gmail Team The best of Gmail, wherever you are - 15 Get the official Gmail app The best features of Gmail are only available on your ph 16 Jul
- ☆ Gmail Team Three tips to get the most out of Gmail - 15 ... Tips to get 16 Jul

Inbox

Starred

Sent Mail

Drafts

More

9 GB (9%) of 16 GB used
Microsoft

©2014 Google - Terms & Privacy

Chat search bar



No recent chats
[Start a new one](#)



New Message

To: _____ Cc: Bcc: _____

Subject: _____

Message body area

Send | Attach | Insert | Link | Undo | Redo | Refresh | Delete | Print

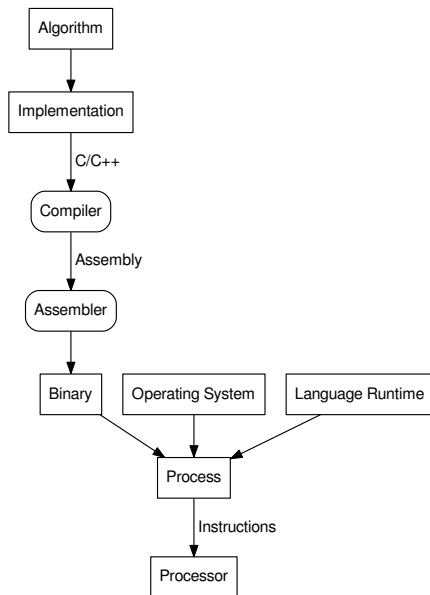
- 1 Organization
- 2 Performance Metrics
- 3 Program Optimization**

Principles of Optimization

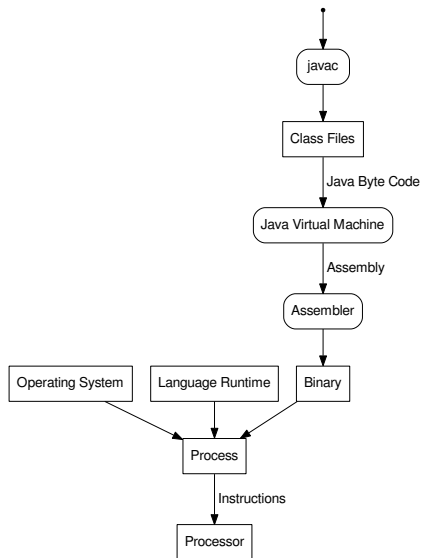
- Work less
- Work cheaply
- Work concurrently

applies to programs only

Layers



Layers for Java



Conclusion

- Two metrics of interest
 - Latency
 - Throughput
- Unit of work is the *instruction*
- Principles of optimization
 - Use fewer instructions
 - Use cheaper instructions
 - Concurrent instruction execution
- C/C++ chosen for
 - Fewer abstractions
 - Easier understanding

Acknowledgements

Images of Toy Story and GMail from Wikipedia