Parallel Computing Notebook
Leanne Chan
April 2017

Flipping Coins in Parallel
Model Website

Q: If each worker can flip one coin per time step, how many time steps does it take the serial worker to flip two coins?
A: It takes 2 time steps for the serial worker to flip two coins.

Q: How many time steps does it take two parallel workers to flip the two coins?
A: It takes 1 time step for two parallel workers to flip the two coins.

Q: Set the work size to 16. How many time steps does it take the serial worker to flip the coins?
A: It takes 16 time steps for the serial worker to flip the coins.

Q: How many time steps does it take the parallel workers to flip the coins?
A: It takes 8 time steps for the parallel workers to flip the coins.

Q: Set the number of parallel workers to 4. How many time steps does it take the parallel workers to flip the coins now?
A: It takes 4 time steps for the parallel workers to flip the coins.

Q: Set the number of parallel workers to 8. How many time steps does it take the parallel workers to flip the coins now?
A: It takes 2 time steps for the parallel workers to flip the coins.

Q: From what you've seen, what is one reason why it would be a good idea to use parallel workers instead of a serial worker?
A: It would be a good idea because parallel workers can get the same amount of work done in a shorter amount of time (SOONER).

Q: Decrease the number of parallel workers to 2. Decrease the max time to 2. In 2 time steps, how many coins can be flipped by 2 parallel workers compared to one serial worker?
A: 4 coins can be flipped by 2 parallel workers compared to the 2 coins that can be flipped by one serial worker.

Q: Increase the number of parallel workers to 4. In 2 time steps, how many coins can be flipped by 4 parallel workers compared to one serial worker?
A: 8 coins can be flipped by 4 parallel workers compared to the 2 coins that can be flipped by one serial worker.

Q: Increase the number of parallel workers to 8. In 2 time steps, how many coins can be flipped by 8 parallel workers compared to one serial worker?
A: 16 coins can be flipped by 8 parallel workers compared to the 2 coins that can be flipped by one serial worker.

Q: From what you've seen, what is another reason why it would be a good idea to use parallel workers instead of a serial worker?
A: It would be a good idea because parallel workers can get more work done in the same amount of time (MORE: TIME LIMIT).

Q: Increase the max time to 16 time steps. Decrease the number of parallel workers to 2. Decrease the max worker memory to 2 coins. If each worker can only hold 2 coins in memory, what is the maximum number of coins that can be flipped by 2 parallel workers compared to 1 serial worker?
A: 4 coins can be flipped by 2 parallel workers compared to 2 coins flipped by 1 serial worker.

Q: Increase the number of parallel workers to 8. If each worker can only hold 2 coins in memory, what is the maximum number of coins that can be flipped by 8 parallel workers compared to 1 serial worker?
A: 16 coins can be flipped by 8 parallel workers compared to 2 coins flipped by 1 parallel worker.

Q: From what you've seen, what is another reason why it would be a good idea to use parallel workers instead of a serial worker?
A: It is a good idea because parallel workers can get more work can be done given limited memory (MORE: MEMORY LIMIT).

<u>Human Parallel Computer - Data Parallelism through Forest Fire Simulations</u>
My number: 8
Total number of students: 22
My probability: .364
Percentages: 5.53%, .34%, 1.38%, 1.38%, 12.45%
Iteration counts: 7, 1, 3, 3, 12
Average percentage: 4.216%
Averge # of iterations: 5.2

Q: What were some of the **tasks** we did in this exercise? What were they, and who did them?
A: The first task was to choose a number. Each worker did this by coming up and taking a sticky note. Another task was to find the average percentage and average iterations. Each worker did them according to their specific number.

Q: What kinds of **data** did we work with in this exercise?
A: We worked with numbers, percentages, decimals, averages, probabilities, and graphs in this exercise.

Q: In which steps was there **communication** or **message passing** during this exercise (mark these steps)?
A: Initially taking a sticky note from the front and returning the sticky note with each worker's personal data were the steps of communication during this exercise.

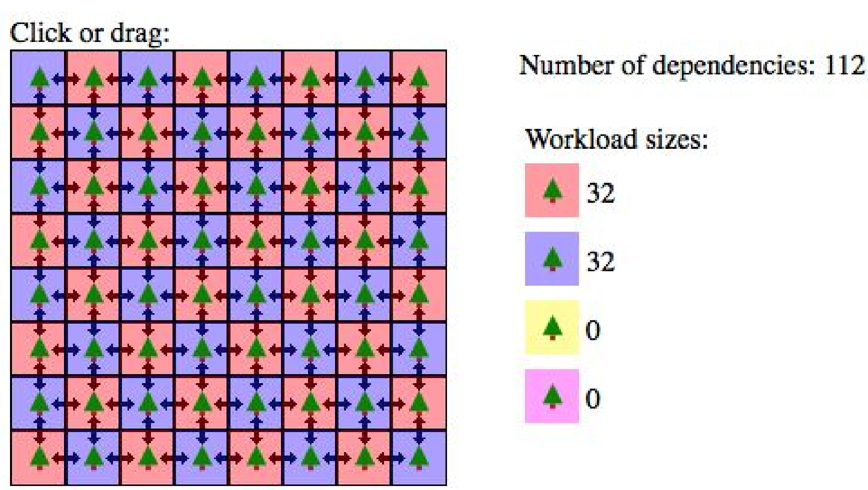
Q: In what ways could this exercise have been **optimized** so it could take less time?
A: This exercise could have been optimized if the communication was more efficient, possibly by each worker putting their own data onto the data points rather than having the instructor do so.

Q: How could we have run this exercise using two instructors instead of one?
A: We could have run this exercise using two instructors by having one instructor explaining while the other one was passing things out. In addition, one instructor could have read off the numbers while the other types them into the computer to make the exercise more efficient.

Q: In what ways did we simulate a **parallel computer** in this exercise?
A: We simulated a parallel computer because each of the students simulated a parallel worker and we all responded to a single worker, which was the instructor.

<u>Domain Decomposition</u>
[Model Website]

Images:



This image shows the maximum number of dependencies. It uses only two different colours.

This image shows the minimum number of dependencies with an even spread of the workload among the 4 colours.

Q: If we assumed each color is assigned to a researcher in a real forest, and each researcher is studying how a fire spreads through the forest, why do we call them **dependencies**? Why do we call them **workload sizes**?
A: We call them dependencies because each researcher has to depend on the trees of another researcher. We call them workload sizes because that is the amount of work in trees that a given researcher has.

Q: If we assumed each color is assigned to a computer running a simulation for that part of the forest, why do we call them **dependencies**? Why do we call them **workload sizes**?
A: We call them dependencies because one computer has to depend on another computer to know what to do in the simulation. We call them workload sizes because it is the amount of work in trees that a computer has to run.

Q: Why would we want to minimize the dependencies?
A: We would want to minimize dependencies because it leaves less room for error when communicating from one computer to another.

Q: What are some reasons we might want to give more work to one of the colors/researchers/computers?
A: We might want to give more work to one of the colours/researchers/computers because one might be more capable/experienced at doing so. Another reason would be if one of them has more time than the others to do the work.

Parallel Recipes

My serial recipe: How to Make a Peanut Butter and Jelly Sandwich

Materials:

1. Peanut Butter
2. Jelly
3. Knife
4. Plate

5. Bread
6. Napkin

Instructions:

1. Open up your bag of bread by twisting the tie which is holding the bag closed with your fingers. Twist it apart until it is no longer holding the bag closed.
2. Reach into the bag with your hand and pick up 2 slices of bread. Take the bread out and place it on your plate, side by side.
3. Open up your jar of peanut butter by taking the lid and twisting it counterclockwise with your hand. Twist it until it is loose, and then take the lid off.
4. Take your knife in your hand, holding it on the side that is not sharp, and dip it into the jar of peanut butter until the tip of the knife is covered in peanut butter.
5. Remove the knife and spread it across the face of one of your slices of bread.
6. Repeat steps 5 and 6 until the entire face of a single slice of bread is covered with peanut butter.
7. Clean the knife off the sharp part of the knife with your napkin. Wipe off all of the peanut butter.
8. Repeat steps 4, 5, 6, and 7 with the jelly jar and the other piece of bread.
9. Put the 2 pieces of bread together. Make sure the sides with the spreading are facing each other.
10. Enjoy!

Dependencies:

- You have to have a surface to place the materials on.
- You have to have the materials to do the recipe.
- You have to have kept the plate on the surface to place the bread on it.

My parallel recipe

Materials:

1. Peanut Butter
2. Jelly
3. 2 Knives
4. Plate
5. Bread

Instructions:

1. Person 1 will open up the bag of bread by twisting the tie which is holding the bag closed with their fingers. They will twist it apart until it is no longer holding the bag closed. Person 2 will open up the jar of peanut butter by taking the lid and twisting it counterclockwise with their hand. They will twist it until it is loose, and then take the lid off. Person 3 will open up the jar of jelly by taking the lid and twisting it counterclockwise with their hand. They will twist it until it is loose, and then take the lid off.
2. Person 1 will reach into the bag with their hand and pick up 2 slices of bread. They will take the bread out and place it on your plate, side by side. Person 2 will take a knife in their hand, holding it on the side that is not sharp, and dip it into the jar of peanut butter until the tip of the knife is covered in peanut butter. Person 3 will take a separate knife in

their hand, holding it on the side that is not sharp, and dip it into the jar of jelly until the tip of the knife is covered in jelly.
3. Person 2 will remove the knife from the peanut butter jar and spread it across the face of one of Person 1's slices of bread. Person 3 will remove the knife from the jelly jar and spread it across the face of the other slice of Person 1's slices of bread.
4. Person 1 will put the 2 pieces of bread together. They will make sure the sides with the spreading are facing each other.

Q: In what ways was your **parallel** recipe different than your **serial** (non-parallel) recipe?
A: The 2 recipes differed in that there were more steps in the serial recipe. They also differed in that the materials were slightly different. The parallel recipe needed 2 knives, but it didn't need a napkin.

Q: In what ways was your **parallel** recipe the same as your **serial** (non-parallel) recipe?
A: The 2 recipes were similar in that they both had essentially the same steps, just done by more than one person.

Q: In what ways was your parallel recipe more efficient? In what ways was it less efficient?
A: The parallel recipe was more efficient because it took less steps. It was less efficient because there would be more dependencies where each person would have to make sure that another person had done the step.

Q: Did anything need to change about the resources/materials/ingredients/tools in your recipe when you went from serial to parallel?
A: Yes, one more knife was required for the parallel recipe. In addition, the parallel recipe did not need a napkin.

Q: In what ways do you think this activity relates to computing and parallel computing?
A: This activity relates to computing and parallel computing because it shows how parallel computing is efficient in some ways due to how there are more than 1 person/computer doing the task, however it could also be less efficient because there are more dependencies where each person/computer has to rely on another one and make sure that there were no errors.

<u>Going Shopping</u>
Solutions:
1. Give each of them an equal amount of money and each of them can buy whatever they want.
2. The instructor can go in and buy the same thing for each child.
3. The instructor can take orders from each of the children, and then he can go in and buy everything.

4. Each child chooses what they want and they bring it to the checkout where the instructor will go and pay for it.
5. The instructor brings the children through the aisles and they each take whatever they want off the shelves together.
6. Each child buys ingredients for a sandwich, and the instructor buys bread. They each make their own sandwiches.
7. The instructor drives each of the children to their homes individually to get their lunches from home.
8. Order pizza/take out.
9. Call the parents and ask them to bring the children's lunches.

What if's:
1. What if the van can't fit all the children?
2. What if the instructors can't drive?
3. What is the van is broken?
4. What if the van is out of gas?
5. What if the store is closed?
6. What if the store doesn't have what the children want?
7. What if the children have allergies/food restrictions?
8. What if a child gets lost?
9. What if the van gets into an accident?
10. What if the van gets pulled over?
11. What if they have no money?

Q: Where is there inherent **parallelism** in your solutions?
A: There is inherent parallelism in the solutions where each of the kids goes and buys their own food.

Q: Where are there **dependencies** between tasks in your solution?
A: A dependency was you have to drive to the store before people could get food. Another dependency is the instructor can't get the food until he gets the children's orders. Most of the solution depend on the store having food.

Q: Where is there **communication** in your solution?
A: There was communication from the children to the instructor when he takes the orders. There is also communication about how much money each child gets.

Parallelism in Nature

1. **Model link:** http://www.shodor.org/interactivate/activities/ABetterFire/
   Data: wind control, forest tree density, size of forest, % of trees burned
   Tasks: regrow forest, burn forest, select place to start fire, spread fire
   Parallelizable data: the size of the forest, wind control, and forest tree density can be changed at the same time

Parallelizable tasks: multiple trees can decide if they are going to spread fire in the time step at once

2. **Model link:** https://www.agentcubesonline.com/project/313425
   Data: row/column of agent, agent age, agent hunger, counter
   Tasks: agent moving, agent eating, agent dying, grow up (baby to adult)

Parallelizable data: multiple row/columns can change at once, multiple agent ages can change at once, multiple agent hungers can die at once, multiple counters can change at once

Parallelizable tasks: multiple agents can move at once, multiple agents can eat at once, multiple agents can die at once, multiple agents can grow up at once

Q: What patterns do you notice in the types of data and tasks that can be parallel?
A: I noticed that the data and tasks that can be parallel often involve more than one agent. In addition, data and tasks where multiple things occur in one timestep can be parallel.

Q: What patterns do you notice in the types of data and tasks that cannot be parallel?
A: I noticed that the type of data that cannot be parallel often involves only one agent or a single thing occurs in one timestep.

Careers in High Performance Computing

Career: Architect
How HPC can be used in that career:

I was looking for a traditional design architect, but many of the sources were about a "supercomputer architect". An HPC architect plans, designs, documents, maintains, and develops solutions for a company's core software products. Essentially, it seems as if they are the ones that develop and build the supercomputers, such as Blue Waters, to be as efficient as possible. I could not find any information on traditional architecture using HPC, but they could most likely use parallel computing to digitally process their designs and blueprints.

Sources:

- http://old.adaptivecomputing.com/about/careers/pos_frame.php?posid=49
- https://en.wikipedia.org/wiki/Supercomputer_architecture

The World's "Fastest" Supercomputers

Q: When was the most recent Top500 list published?
A: November 2016

Q: What is the name of the fastest supercomputer in the world according to the most recent list?
A: Sunway TaihuLight

Q: Where is that supercomputer located?
A: National Supercomputer Center
Wuxi, China

Q: How many **cores** does it have?

A: 10,649,600

Q: How much **peak performance (RPEAK)** does it have?
A: 125,435.9 TFLOP/S

Q: How many of the Top500 sites in the top 10 are located in the United States?
A: 5

Q: If the **Blue Waters** supercomputer was capable of a **peak performance** of **13,000 TFLOP/S** when it came on-line in 2012, where would it be listed in the November 2012 list?
A: 3rd

Q: Why doesn't Blue Waters appear on that list?
A: Blue Waters doesn't appear on that list because they chose to not be in it. This is because they did not like the way Top500 ranks the computers. The list focuses on Linpack, but doesn't consider storage capacity, bandwidth, and memory capacity. This causes a problem because some centers will only focus on increasing the peak performance and nothing else, which could limit the types of applications it could use. In addition, Top500 essentially just lists which computers get the most funding.
Source: http://www.ncsa.illinois.edu/news/stories/TOP500problem/

Q: What are **cores**?
A: Cores are the processing unit which receives instructions and performs calculations based on those instructions (CPU).

Q: What does **TFLOP/S** stand for?
A: TFLOP/s stand for teraflops per second, or trillion floating point operations per second.

Q: What does **Linpack** measure?
A: Linpack measures a system's floating point computing power.

Q: What would be some different ways to rank supercomputers?
A: Supercomputers can be ranked by the number of cores, Rmax, peak performance, or power. They could also be ranked by the number of projects completed on it.
LittleFe
Where the name comes from: The name comes from the slang term for supercomputing, which is "Big Iron", and the chemical name for iron, which is Fe.
Components:

- Motherboards (6)
- Cores (2 per CPU)
- Hard Drive (1)
- Network (Ethernet)
- RAM or Memory (1-4 per Motherboard)

- Power Supply
- Cooling System (Fan, Heat Sink)
- CPU
- Case
- Ethernet Switch

Blue Waters demo

YouTube video

Q: What are the advantages to using a remote supercomputer as compared to a local supercomputer like LittleFe?
A: Some of the advantages to using a remote supercomputer are that you can access it from many different locations, and it is capable of doing much more since it is more powerful.

Q: What are the disadvantages?
A: Some of the disadvantages are that it may be slower since it is being connected to from farther away, and it is larger and harder to transport.

Parallel Computing: Terminology and Examples

Slides

Shopping for Your Own Supercomputer

- **Part:** Motherboards (6)
  Cost: $130.00 * 6 = $780
  Link:
  https://www.newegg.com/Product/Product.aspx?Item=N82E16813157746&cm_re=intel_motherboard-_-13-157-746-_-Product

- **Part:** Memory (24)
  Cost: $40 * 24 = $960
  Link:
  https://www.newegg.com/Product/Product.aspx?Item=9SIA7253RY0185&cm_re=low_voltage_4gb_memory-_-0RM-002P-006G8-_-Product

- **Part:** Cooling
  Cost: $120
  Link: https://www.newegg.com/Product/Product.aspx?Item=N82E16835146042

- **Part:** Ethernet Cable (6)
  Cost: $4.50 * 6 = $27
  Link:
  https://www.newegg.com/Product/Product.aspx?Item=N82E16812119169&cm_re=ethernet-_-12-119-169-_-Product

- **Part:** Hard drive
  Cost: $120
  Link:
  https://www.newegg.com/Product/Product.aspx?Item=N82E16822178865&cm_re=hard_drive_external-_-22-178-865-_-Product

- **Part:** Ethernet Switch
  Cost: $33
  Link:
  https://www.newegg.com/Product/Product.aspx?Item=12K-008X-00026&cm_re=ethernet_switch-_-12K-008X-00026-_-Product

- **Part:** Power Supply
  Cost: $190
  Link:
  https://www.newegg.com/Product/Product.aspx?Item=N82E16817438013&ignorebbr=1

- **Part:** Case
  Cost: $128
  Link:
  https://www.newegg.com/Product/Product.aspx?Item=N82E16811139085&ignorebbr=1

- **Part:** Router
  Cost: $36
  Link:
  https://www.newegg.com/Product/Product.aspx?Item=N82E16833124190&cm_re=router-_-33-124-190-_-Product

- **Total Cost:** $2394

Q: What are the most important parts of a supercomputer?
A: The most important parts of a supercomputer are the motherboard, CPU, RAM, and cooling system.

Q: What assumptions did you make when you did your shopping? Which of these assumptions were false?
A: I assumed that all of the parts are working and brand new. Many of the parts on the website may be used. I also assumed that the more expensive parts would be higher quality. This was false in certain situations because some of the more expensive ones had no reviews and no pictures.

Daqri demo

Q: What are some of the ways you can envision augmented reality being used for science?
A:

Q: What specific scientific examples can you think of that would benefit from augmented reality?
A: