SOCIETY OF PHYSICS STUDENTS
An organization of the American Institute of Physics

SPS Chapter Research Award Final Report

Project Title	Data Analysis and Accuracy: Small Supercomputer Versus a Dell Dinosaur
Name of School	University of Kentucky
SPS Chapter Number	3511
Total Amount Awarded	1157.20
Total Amount Expended	1057.20
Project Leader	Alex Blose, Dany Waller, Ben Kistler

Abstract

Computational methods provide powerful tools for solving problems in physics. We plan to build a mini-supercomputer from Raspberry Pis following the "Tiny Titan" build specifications available from Oak Ridge National Lab. Once built, the cluster can be used indefinitely for education, outreach, and community building.

Statement of Activity

The Tiny Titan project has aimed to introduce students to physics simulations, data analysis, and basic computer hardware. Roughly 25 students have been actively engaged in activities designed to bolster techniques in these areas. The first part of the project was understanding the need for parallel processing—what makes a computer running in parallel better for some programs than others—then transitioned to implementing introductory programming and data analysis as well as more challenging projects such as estimating pi simulating the Ising Model.

How does a supercomputer work? How is parallel processing used in physics? Why is this useful for data analysis? Physics research is already being shaped by the growing accessibility of parallel processing. Computational methods allow us to test theoretical models, and as these models increase in complexity, parallel processing is required to compute data in a realistic amount of time. We built the mini-supercomputer based on the "Tiny Titan" Raspberry Pi cluster and have constructed programs using parallel processing to demonstrate important physical concepts such as mathematical approximations, complex modeling, and interactive systems.

Overview of Award Activity

The Overview should be a detailed description of the work that was completed on the proposed project. Provide sufficient background information for a non-specialist to understand how your knowledge has advanced through the work that you did.

- The beginning of the project went relatively smoothly, installing software, setting up hardware, and running the first simulations, went according to the schedule. However, producing more complicated programs such as the Ising model and the estimation of pi have been more challenging. The code for the estimation of pi was straightforward for the regular computer version, and implementing MPI was a challenge, but we got it. We began coding the Ising model at the beginning of last summer, but patching all the bugs in the regular computer code took months. We are still developing the parallelized version of the code, as the delay from the regular code was so long and our team is still new to MPI, and the project will continue into the spring semester.
- It has been difficult to maintain consistent meetings throughout the fall semester, as some of our original cluster members, previous officers, dedicated seniors graduated. We plan to open a new SPS officer position to take charge of cluster operations, as we have found that balancing responsibilities between multiple existing officers is not sufficient for when we introduce even more projects. The goal of this position would be to organize big group meetings and help everyone through sticking points in programming on the cluster. We have found that large group meetings are not the best for developing a foundation for programming, so we are planning to divide less experienced programmers into groups of 2-3 under more experienced programmers next semester. This ensures that the new programmers get the necessary attention and doesn't isolate people who are new to programming.
- There is a group of more experienced sophomores who will be engaging in independent research projects for course credit next semester under the guidance of Dr. Brown. Some of our members have proposed projects to our chapter adviser and condensed matter/computational physicist Dr. Kaul for next semester as well using the cluster. Another member in our chapter has proposed a project for servers, including a messaging server and down the line possibly an application in robotics.

Description of Research - Methods, Design, and Procedures

Beginning:
The first step was assembling each raspberry pi and loading each with the raspbian, based on linux, operating system. The remainder of the software necessary came integrated into the pis. Each pi needed to be individually booted and tested. Everyone in the team learned how to access each node from the master unit and pass commands between each pi. This is when we realized there was a problem with the setup and eventually discovered a missing file from the setup. Once the file was installed, the pis communicated properly and the small supercomputer was alive!

Setup:
The nine raspberry Pi B+ nodes were connected to the ethernet cable switch. The master node, pi number one, was also connected to the monitor/television, mouse, and keyboard.

Water simulation:

We installed the fluid dynamics simulation from the ORNL Tiny Titan github with some slight modifications to the program, as some of the commands were out of date in the installation instructions--the command was meant for an older version of the software, but once the correct command was found, the water simulation worked.
We observed the behavior of the simulation under conditions which would strain the computational power. In this, we were looking for whether the simulation could continue to operate under these conditions depending on the number of nodes used. We also observed the smoothness of the simulation against the number of nodes. In all simulations, increasing the number of nodes increased or maintained the smoothness of the simulation.

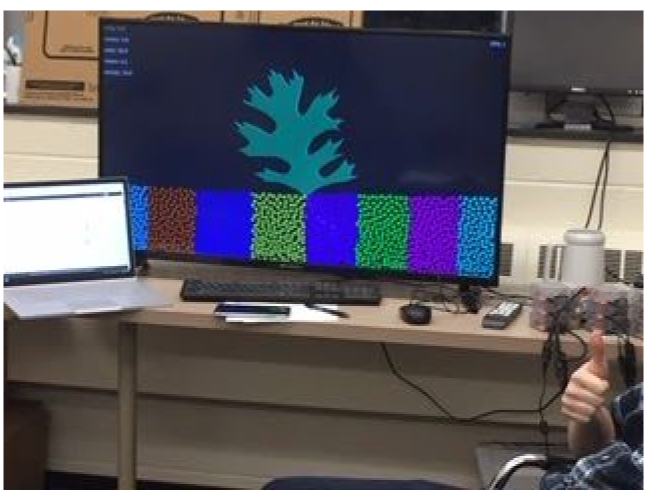
For the next programs, we have been working in c++ and parallelizing with MPI. The running time of the commands were measured using the "time" command in the terminal.

Estimation of pi:
How can we improve our calculation of pi? We can increase accuracy and decrease time to compute. We calculate pi with a monte carlo simulation that approximates the area of a square with a circle inside by generating random points inside. Taking the ratio of the area of the circle (area πr^2) inside to the area of the square (r^2) yields an approximate value for pi--details of the calculation and error are below. By using enough points to guess the area inside the circle, on the order of 10^3, we found that as we increased the number of nodes, our estimation of pi became more accurate. If we fixed the number of points that we guessed in the simulation, then we found that with a stop condition that the increased number of nodes would yield a faster and more accurate result. Details are described below.

Ising Model:
Given a lattice of a fixed size, we can generate the spin at each point on the lattice. From here, we can introduce defects in the lattice and measure properties such as size of the lattice, temperature, specific heat, and energy of magnetization.

Discussion of Results

Water Simulation



As more nodes were added, the accuracy of the simulation increased. The water simulation works by approximating fluids as balls, so when the number of balls was increased, the fluid looked more realistic. The best way to aid the simulation with more balls was to add more nodes to simulate a section of the balls. Each stripe of colored balls is simulated by a different node, and when the number of nodes increases, the simulation works visibly smoother. When the number of balls was increased when run with a low number of nodes, the water appeared choppy and the simulation looked much less smoothly.

Viscosity, gravity, density, pressure, and elasticity are the variables that were available to be changed by the simulation. These variables could be adjusted to create concrete situations such as standard conditions of water in a room or less normal situations such as highly viscous, no gravity, high pressure, and low elasticity liquid. In any situation, the more nodes added gave a more accurate and efficient simulation.

Estimation of Pi:

This simulation was used as an intermediate level simulation—for people who had acquired enough programming knowledge to be comfortable learning about MPI, message passing interface. MPI is what allows each node of a cluster to communicate, and this can even be done within a regular computer with its cores. This program took random points inside a square with a circle inscribed inside it and took the ratio of the number found inside the square to the number of random points. Think of it as taking the ratio of the area of a circle (pi * r^2) to the area of a square (r^2). This ratio, dividing the area of the circle by the area of the square, yields pi. With more points taken inside, we can calculate more precise values for pi. Specifically, for every N points, we can find pi up to an error of sqrt(N).

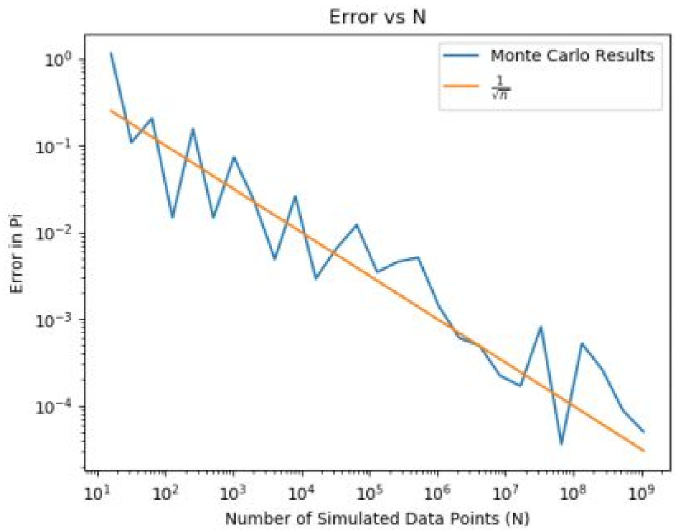
This plot shows the relationship between the number of simulated points and the error in pi fit to the sqrt(N). This is expected because the error^2 = 1/N * $\sum_{i=1}^{N} (\mu - x_i)$ is mostly dependent on the 1/N term where μ is the mean = $\sum_{i=1}^{N} x_i$.

Notice	Number of Nodes (Pis)	Running Time	Value of Pi
	2	0.456	3.139224
Time increase (not accuracy)	5	1.027	3.139224
Accuracy and time increase	7	1.034	3.142576
Accuracy and time increase	9	1.35	3.141217

Above are the results of calculating pi to the same number of decimal places with the same number of points N. We found that running the simulation without a stop condition would not necessarily decrease the time, as we were attempting to mix accuracy and time conditions.

N	Result	Time 1 node (seconds)	Time 9 nodes (seconds)
10^3	3.1	0.348	0.454
10^6	3.14	0.619	0.572
10^9	3.141	298.219	78.535

This program was made to stop when the simulation produced the last correct digit of pi. This shows that increasing the number of nodes at high enough values of N decreases the necessary running time. For small values of N, it is expected that the running time for one node will be lower, as the processor power of one node in this situation is more efficient than assigning different tasks to multiple nodes. However, as we increase N, we see that the time does decrease significantly.



Above are the results of the accuracy as mentioned above. We ran the simulation with different numbers of data points and found that the error fit well to 1/sqrt(N), as discussed above.

Ising Model

MPI code in progress!

Dissemination of Results

This section should describe how you have disseminated your results. This might include poster presentations, research talks, papers submitted for publishing, etc. You may attach copies of products to this report.

- Independent research projects will begin next semester. At the end of these research classes, there is an associated ten page paper. These papers are often submitted to undergraduate research journals and conferences when completed.
- Our department decided to not host any open lab days this semester, but there will be one next semester.
- Taking the cluster outside has also been an issue, as the individual cases are not durable enough to withstand dropping, so one of our members has volunteered to create a large case using a 3d printer on campus. In addition, this case could make the setup weatherproof, as we were not able to bring the cluster to some events because of the weather--freshman orientation for Kweek is traditionally outside, but the rain prevented us from bringing the cluster out. This case would allow us to safely transport and store the cluster outside.

Bibliography

- Bevington, P. R., & Robinson, D. K. (2010). Data Reduction and Error Analysis for the Physical Sciences. Boston: McGraw-Hill.
- Schroeder, D. V. (2000). An Introduction to Thermal Physics. United States: Robin J. Heyden. Simon, S. H. (2017).
- The Oxford Solid State Basics. Oxford: Oxford Univ. Press. Press, W. H. (2002).
- Numerical Recipes in C. Cambridge: Cambridge Univ. Press.
- Tiny Titan Build Guide. (2015). Retrieved from http://tinytitan.github.io/downloads/TinyTitanBuildGuide.pdf

Impact Assessment:

How the Project Influenced your Chapter

This has been a great experience for our SPS! It's been great to see all of the connections that have been made because of this project. It's been an excellent learning experience for everyone, older and more experienced programmers a chance to lead projects and mentor younger physics majors. Some advice would be to give the project time to code and make sure that there are more than just seniors working on the project. In addition, it is important to make sure that everyone is feeling productive and comfortable working in the group. It is also better for people who are new to programming to split into smaller groups for more individualized projects and learning foundations. The cluster has also been a great tool for recruitment, and it has helped to foster a more inclusive and welcoming environment in SPS. One important lesson has been scheduling consistent meetings throughout the semester, as college students are much more free in the beginning of the semester and typically busy after the midterms. The most surprising part of the project was seeing how many people were interested in the cluster, it has been a great chance to meet new professors, graduate students, and classmates and see how everyone interacts! This has been a great experience for our department and will greatly change how our chapter members experience physics at UK!

This has given our chapter a chance to connect with itself, bringing together more senior physics majors with younger majors. We have also bolstered our connection with graduate students and professors since beginning the project. Because more faculty and graduate students are interested in working with the cluster, more sophomores and freshmen have been more willing to join both the project and SPS.

In addition to physics majors, we have been coordinating with the math and computer science departments to also coordinate more with the math and computer science departments for more projects to run on the cluster.

The best advice we have is effective and productive time management. Have a schedule and stick to it! If you set aside at least one meeting per week, a lot can get done in that time if there is a good plan. In addition, try to get everyone up to speed with what you are doing. Because we took on more programming intensive projects, it takes careful planning to make sure everyone is involved and working at the right level. For example, we had people without programming experience join the project, so we started them with basic programming techniques and aimed for "hello world" programs in MPI. People with more experience had goals that corresponded to their levels of programming but also gave exposure to more advanced techniques to people without as much programming experience--it means a lot when a senior who as done research every semester and summer for the past three years talks about tricks to someone new to programming, and those are the kinds of interactions that happened here. The group work structure was another key dynamic, but it was more efficient when we switched to small groups, especially for people who were newer to programming.

Key Metrics and Reflection

How many students from your SPS chapter were involved in the research, and in what capacity?	Approximately 25 were active members in the project, and 4 people are taking on projects as independent research next semester
Was the amount of money you received from SPS sufficient to carry out the activities outlined in your proposal? Could you have used additional funding? If yes, how much would you have liked? How would the additional funding have augmented your activity?	Yes, we had sufficient funding for the project. If we had more funding, we could have purchased more pis, but the computational power associated with adding one or two additional pis to the cluster is much smaller than the first nine pis.

Do you anticipate continuing or expanding on this research project in the future? If yes, please explain.	Yes! This project has sparked undergraduate research projects which will be carried out as classes next semester! These projects include more physics simulations and a server. Additional individual projects outside of research have been planned as well.
If you were to do your project again, what would you do differently?	It would have been more efficient to document every project as we went for ease of progress reports. In addition, it would have been better to have an SPS officer position for the cluster to manage all of the projects happening at once.

Press Coverage (if applicable)

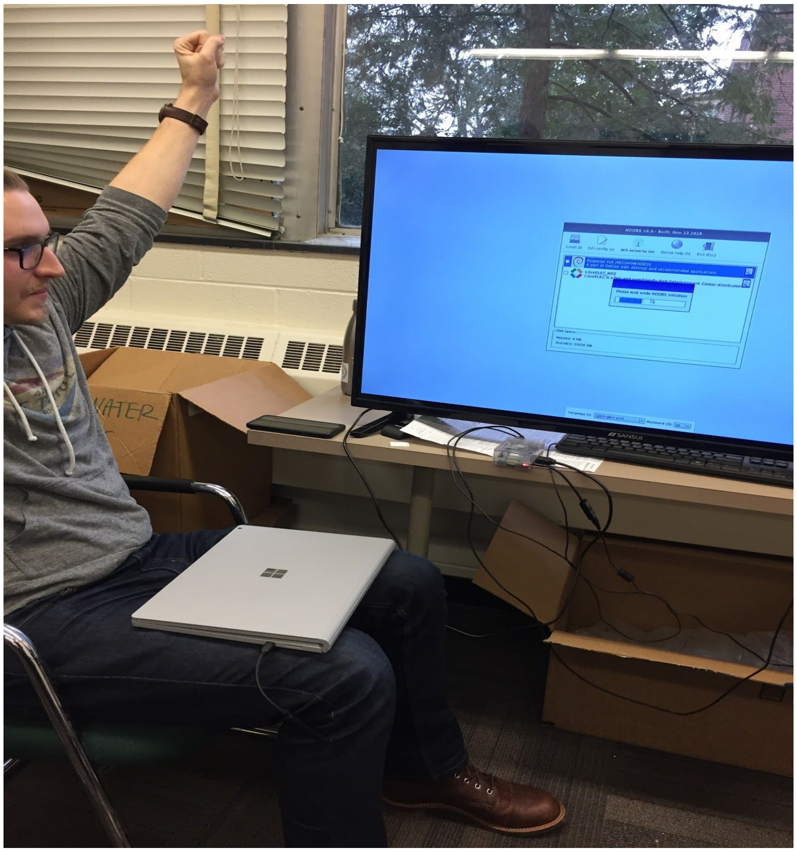
https://pa.as.uky.edu/sps
https://uknow.uky.edu/student-and-academic-life/uk-society-physics-students-recognized-nationally-excellence
"UK at the Half." radio show during the UK v. Ole Miss March 5th 98.1/WBUL-FM and 630/WLAP-AM

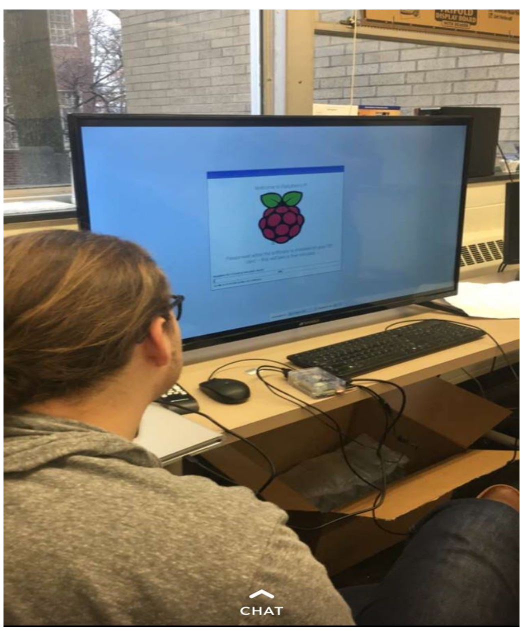
Item Number	Item Description	Notes	Website	Projected Cost	Actual Cost per Item	Number of items	Total Cost
1	How to Book for Raspberry Pi Projects (For reference and future projects)		https://www.amazon.com/Programming-Raspberry-Pi-Second-Getting/dp/1259587401/ref=tmm_pap_swatch_0?_encoding=UTF8&qid=1544809417&sr=8-13	8.20	12.41	1	12.41
2	CanaKit Raspberry Pi 3 B+ (B Plus) with Premium Clear Case and 2.5A Power Supply (Recommended model of Pi for the cluster from the Tiny Titan setup guide)		https://www.amazon.com/CanaKit-Raspberry-Premium-Clear-Supply/dp/B07BC7BMHY/ref=sr_1_4?ie=UTF8&qid=1544807720&sr=8-4&keywords=canakit+raspberry+pi+3+kit+with+clear+case+and+2.5a+power+supply	550.00	55.00	10	550.00
3	Micro SD (For storage for each node of the cluster)	64 GB	https://www.amazon.com/Samsung-MicroSDXC-Adapter-MB-ME64GA-AM/dp/B06XX29S9Q/ref=sr_1_6_acs_osp_osp18-83cd6d80-3e_1?s=pc&ie=UTF8&qid=1544807836&sr=1-6-acs&keywords=micro+sd+card&tag=thewire06oa-20&ascsubtag=83cd6d80-3eee-4fa8-9cd3-2ba36159cc59&linkCode=oas&cv_ct_id=amzn1.osp.83cd6d80-3eee-4fa8-9cd3-2ba36159cc59&cv_ct_pg=search&cv_ct_wn=osp-search&pd_rd_w=jYg9N&pd_rd_r=98569637-30ba-41fc-9362-678	100.00	10.99	10	109.90

			1947526be&pf_rd_r=W4GBZC2HKFZ3XHF5ET14&pf_rd_p=7f6b8bb9-631f-46f6-b8ad-496a9af123d5&pd_rd_wg=1bMuf&creativeASIN=B06XX29S9Q&pd_rd_r=98569637-30ba-41fc-9362-6781947526be&pd_rd_i=B06XX29S9Q&pd_rd_w=jYg9N&pf_rd_r=W4GBZC2HKFZ3XHF5ET14&pd_rd_wg=1bMuf&pf_rd_p=7f6b8bb9-631f-46f6-b8ad-496a9af123d5				
4	Ethernet Switch (To connect each node of the cluster together)	with 16 ports, do we need two switches?	https://www.amazon.com/TP-Link-Ethernet-Unmanaged-Lifetime-TL-SG116/dp/B07GR9S6FN/ref=sr_1_4?s=pc&ie=UTF8&qid=1544808304&sr=1-4&keywords=ethernet+switch&refinements=p_n_feature_four_browse-bin%3A5662321011%2Cp_36%3A1253505011	120.00	55.00	2	110.00
5	Ethernet Cables X 10 (Wires to connect each node of the cluster)	2 x 5packs 5ft cables	https://www.amazon.com/Cable-Matters-5-Color-Snagless-Ethernet/dp/B00E5I7VJG/ref=sr_1_5?s=pc&ie=UTF8&qid=1544808659&sr=1-5&keywords=ethernet%2Bcable&refinements=p_n_feature_keywords_five_browse-bin%3A7800924011&th=1	45.00	10.99	2	21.98
6	Xbox Controller (For controlling the fluid dynamics simulation)	AmazonBasic has good reviews+limited warranty	https://www.amazon.com/AmazonBasics-Xbox-One-Wired-Controller-Black/dp/B07CZTVHY8/ref=sr_1_14?s=electronics&ie=UTF8&qid=1544811085&sr=1-14&keywords=xbox%2Bcontroller&th=	17.00	19.99	2	39.98

			1				
7	Display (For optimal viewing and showcasing)	TV (refurbished to save costs)	https://www.amazon.com/Televisions-Definition-Widescreen-Monitor-Display/dp/B07FBFFDJF/ref=sr_1_5?s=tv&ie=UTF8&qid=1544811308&sr=1-5&keywords=tv%2Bhdmi&th=1	200.00	189.98	1	189.98
	Shipping costs		From amazon all shipping was knocked out	0			-
							1,034.25

Activity Photos





- First operating system has been booted up

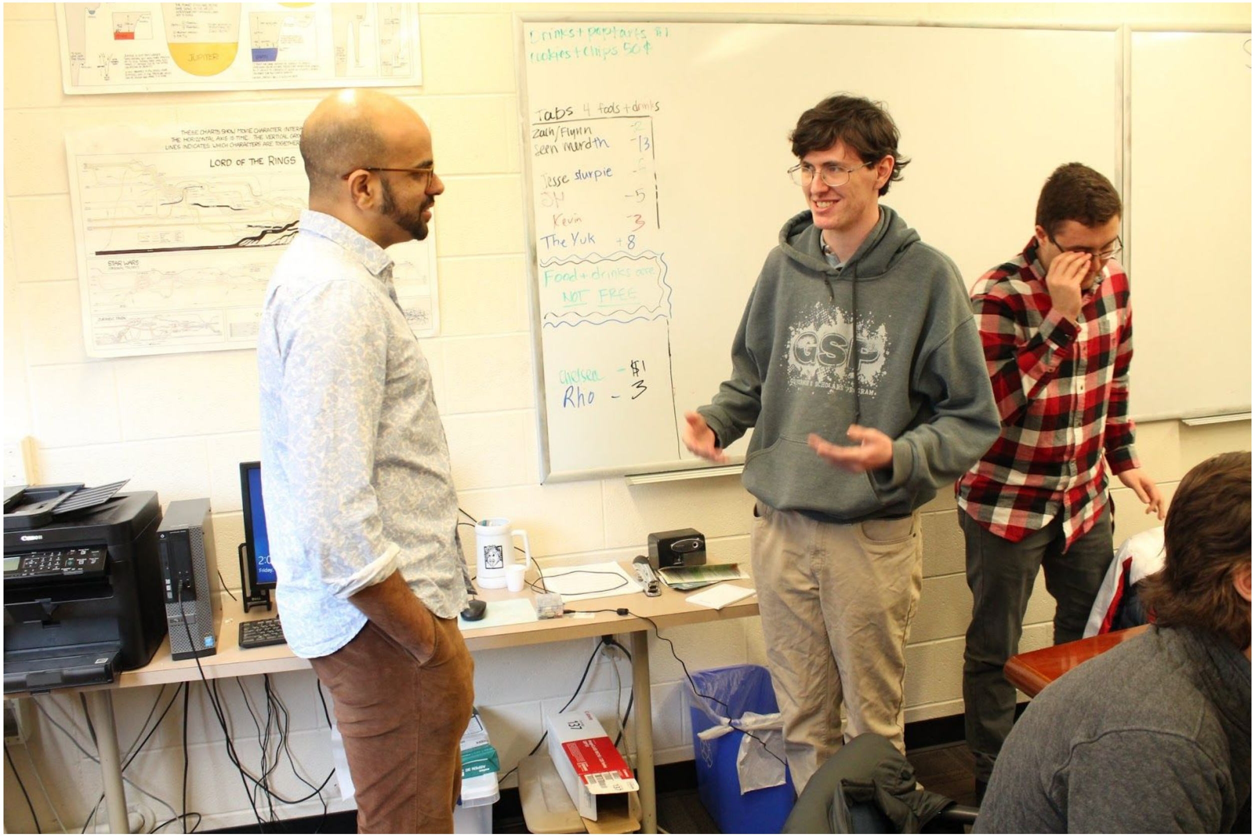
- Setting up the TV

- Cosmic lunch on Parallel Processing by David Mathews, a graduate student at UKY

- 

- Unboxing

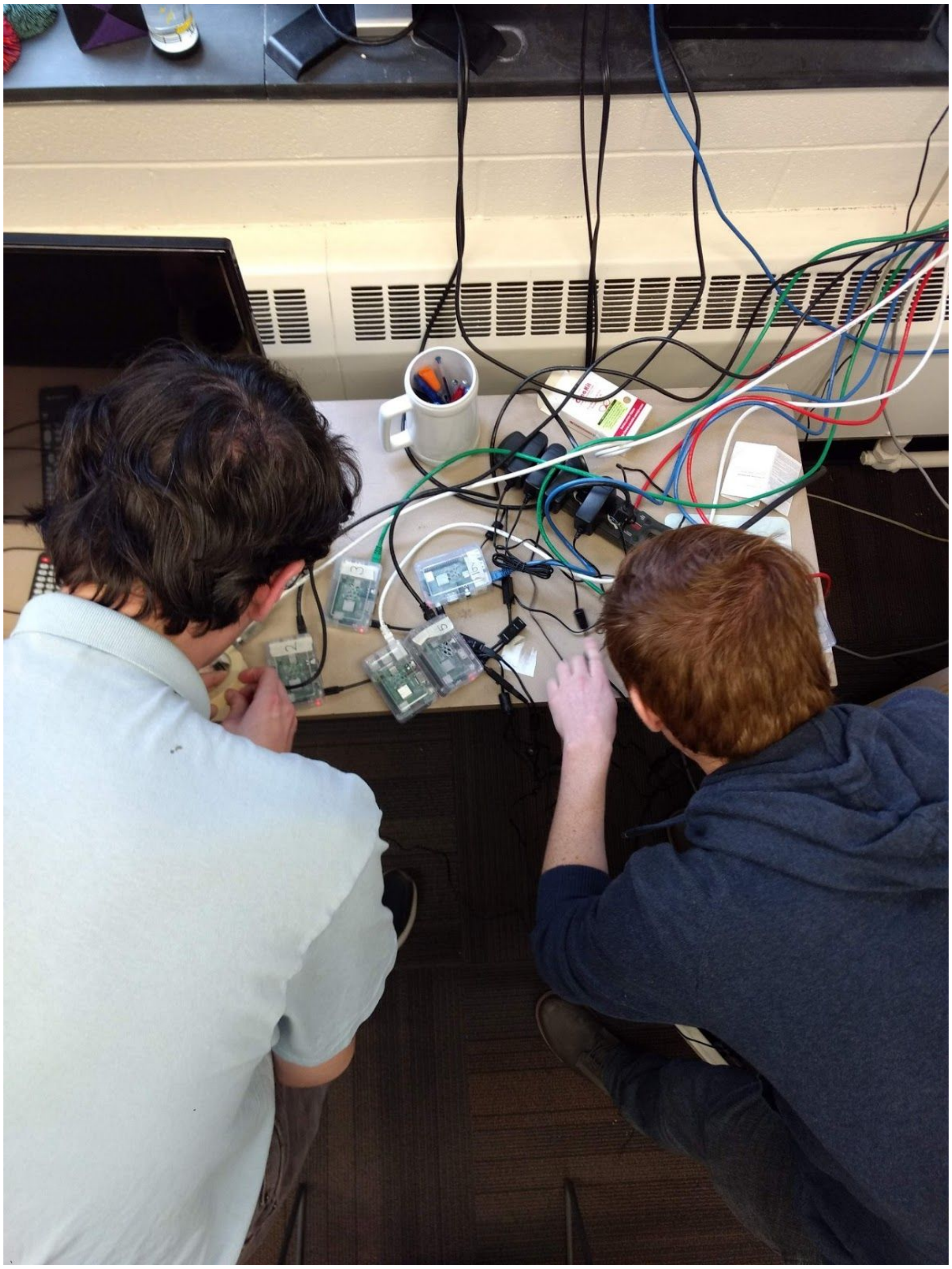
- Setting up the Pis



- Talking to Dr. Kaul about possible projects on the cluster

- Setting up the TV

- First signs of life from the cluster, booting up the operating system



- Cable management