Enhancing HPC Education through Virtual Cluster Administration and Benchmarking

Charles Ross Lindsey
Georgia Institute of Technology
clindsey8@gatech.edu

Jeffrey Valdez
Georgia Institute of Technology
valdez@gatech.edu

Aaron Jezghani
Georgia Institute of Technology
ajezghani3@gatech.edu

Will Powell
Georgia Institute of Technology
ajezghani3@gatech.edu

Richard Vuduc
Georgia Institute of Technology
ajezghani3@gatech.edu

ABSTRACT

The rapid advancement in high-performance computing (HPC) poses significant challenges for the HPC community. Current HPC training approaches often are too generic or too customized to local environments, limiting their applicability and impact. Often, these shortcomings are due to the limited accessibility, excessive cost, and specialized support necessary to provide HPC environments for teaching. To address these challenges, we introduce Virtual Cluster, a hardware-agnostic platform designed to provide an easy-to-configure, generalizable, and scalable approach to HPC system management for training and education in computational research alongside production system configurations. We implemented this platform in a virtually integrated project (VIP) course aimed at training undergraduates for HPC cluster building. Drawing from our experience from the VIP course, we advocate for the integration of more comprehensive educational and training approaches, such as HPC Virtual Cluster, to better support HPC.

KEYWORDS

HPC, virtual cluster, training, virtually integrated project

1 INTRODUCTION

Recent years have seen an acceleration in the development of novel hardware, software, and workflows in the field of High Performance Computing (HPC), and more generally for the broad spectrum of efforts under the blanket of Research Computing and Data (RCD). Fueled especially by the generative artificial intelligence (or generative AI) boom, vendors have halved their hardware release cycles [9], with each subsequent generation costing increasingly more [14] and requiring more electricity and heat dissipation to operate [7, 13]. Consequently, computing clusters across the spectrum of academia, industry, and government all face the mounting challenge of ever-increasing capital and operational costs, and thus necessitate capable researchers to efficiently utilize resources and administrators to optimally configure and support these systems.

https://doi.org/10.22369/issn.2153-4136/16/2/4

Naturally, computer science and engineering students have the strongest theoretical background to understand the nuances of HPC clusters, with core curriculum focused on systems architectures, networking, and parallel and distributed computing frameworks. The natural sciences have built upon a foundation of experimentation to find empirical validation for demonstrating theory as well utilizing HPC resources. For example, students can easily observe the principals of evolution in practice through nearly half a century of *E. Coli* cultures [10], or find validation for Einstein's Theory of General Relativity with the results from the LIGO collaboration [2]. However, current HPC training approaches at colleges and universities, particularly in undergraduate education, are lacking, either too generic, tailored to local environments and limiting their generalizability, or limited to graduate education. Part of the reasons for this have to do with cost for access to HPC environments, with, by necessity, require funding through research and provides little opportunity for training undergraduate students for future careers in research, ultimately hurting progress in the HPC research community. Although consumer-grade hardware is relatively affordable and thus accessible in a classroom setting, the specialized hardware in HPC clusters is much more costly, with entry-level costs for systems alone hitting $10k in 2009 [1], and support personnel and infrastructure tacking on considerably more.

These recent challenges leave HPC in a precarious state, especially with the onset of rapid architectural overturn in support of AI. End users looking to utilize HPC cluster resources to pursue research careers in science or engineering suffer because they lack expertise necessary and robustness in experience to navigate transitions into and throughout careers. HPC practitioners that stay on an academic path require time to adapt to the next system, which is problematic in the face of short evaluation cycles such as postdoctoral positions and pre-tenure evaluations. For example, a recent assessment of user activity on Georgia Tech research clusters found that since 2015, the average period of activity from first to last job submission is only 1.38 years (see Figure 1).

HPC practitioners that turn to industry may initially fare better due to robust onboarding and training into company workflows and teams, but, lacking foundational knowledge in HPC in their undergraduate education, may find themselves vulnerable to downsizing that has become more common with tech companies in recent years [12]. More urgent is the growing need for a workforce ready for the challenges of configuring, tuning, and maintaining HPC systems

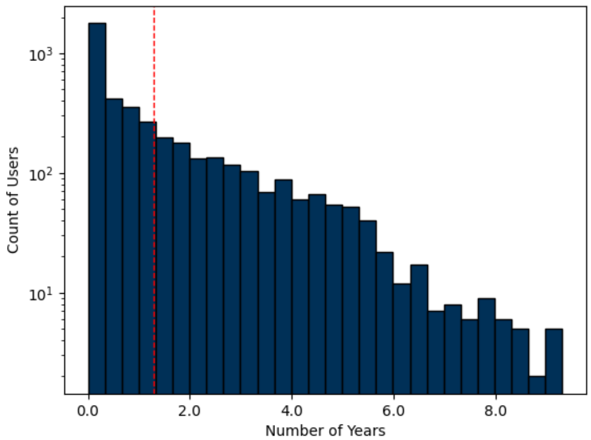

Figure 1: The average active period for users on Georgia Tech clusters was recently determined to be 1.38 years. Active period is defined as the time between first and last job submission, with the data spanning 2015 through present. The short period to engage researchers highlights the importance of effective training and education programs.

as part of their companies' Informational Technology (IT) infrastructure, which can affect whole organizations and their ability to execute computational research and development.

Consequently, this all points to a need for more comprehensive educational and training approaches in support of HPC. To address these challenges, we introduce **HPC Virtual Cluster**, a hardware-agnostic platform designed to provide an easy-to-configure, generalizable, and scalable approach to HPC system management for training and education in computational research alongside production system configurations.

2 VERTICALLY INTEGRATED PROJECTS

Since the late 1990s, the Vertically Integrated Projects (VIP) Program enhances core curriculum through research projects [8]. Ideally, VIP teams should be comprised of 10 or more students from multiple campus departments, and they should participate for three or more semesters to provide a spectrum of academic ranks and experience. Assessment is designed to facilitate these aspects of team composition, with a third of the students' final grades coming from documentation, teamwork, and accomplishments/contributions. Undergraduate students find the VIP program attractive because some colleges such as the College of Computing offer VIP in lieu of the traditional junior design course, while masters students can use it to satisfy elective credits. Currently, Georgia Tech hosts 97 VIP teams exploring topics such as technology-enhanced arts, AI-driven engineering, and Smart infrastructure based on expansive telemetry and automated intelligence.

2.1 Team Phoenix VIP

Originally, Team Phoenix was a participating team in the 2017 Supercomputing Student Cluster Competition (SCC). In Spring 2020, Team Phoenix was reborn as a VIP course under advisement by academic and research faculty from the College of Computing and the Partnership for an Advanced Computing Environment (PACE). Each semester, students progress through a series mini projects to further their knowledge of parallel and distributed computing on an HPC cluster, including resource management, building and configuring drivers and software, and benchmarking applications using various architectures and techniques. Students in the fall have submitted applications for the International Supercomputing High Performance Conference (ISC) SCC, while students in the spring have applied to the Supercomputing SCC and IndySCC competitions.

In addition to the assigned projects, students are exposed to the cluster design, procurement, and support processes. Vendors are brought in to discuss emerging hardware and platforms, and tours of the campus data centers are conducted to familiarize them with enterprise operations. A particular favorite among students is the "Build-a-Bear server dissection," where a production server is brought in and disassembled to foster an understanding of the interplay between hardware and software, as seen in the screen capture in Figure 2.

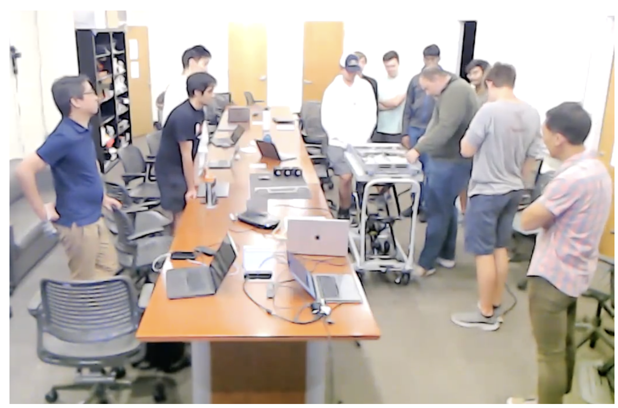

Figure 2: Screen capture of the hybrid class recording showing a class-favorite activity, "Build-a-Bear" Server Breakdown/Reassembly. Using a surplus V100 GPU server, students actively disassemble the server, review all of the components, and then reassemble the system.

2.2 The Missing Middle

Historically, the conversation of a "missing middle" in HPC largely centered on businesses or scientists who could have experienced a boon if given access to a cluster network [4]. The barrier to entry was insurmountable, unfortunately, for some potential users who *i*) could not acquire a supercomputer or purchase access, *ii*) lacked computing knowledge, and/or *iii*) failed to program a cluster to tackle their problem set [4]. In education, undergraduate students not only experience similar roadblocks but are faced with a lack of curriculum necessary to bridge the gap between student and researcher.

Additionally, Team Phoenix was designed to address the "missing middle" [5] in education by filling in the gap in curriculum between foundational undergraduate courses in programming, architectures, and networking and specialized graduate topics such as AI methodologies, HPC systems, and advanced architectures. As a team-based, project-focused course, students collaboratively work through the process of building, configuring, and running applications for a stronger foundation in topics like parallel and distributed computing, communications frameworks, and software engineering. The multi-semester approach and promotion to leadership roles within the team provides further reinforcement, and the subsequent discourse facilitates a richer understanding of the underlying computer science principles.

Students enrolled in Team Phoenix are provided access to multiple cluster systems, one of which is detailed in this report. Once given access to a cluster, undergraduates are imparted with knowledge pertaining to Linux command-line fundamentals for system administration, use of build systems and compilers, job scheduling, file systems configuration, and other topics in HPC. Lastly, our cohort is provided hands-on instruction with scientific and benchmarking applications. Having met the wants of the historical "missing middle," Team Phoenix provides a workforce-development opportunity through participation in cluster competitions which test undergraduates on cluster building/management and utilization.

2.3 Current State of Team Phoenix

Since Team Phoenix's inaugural year, enrollment has increased 144%, from a total of 9 students to 22 in the present term, as presented in Figure 3. Furthermore, Figure 4 shows that retention rate among students, remains high with roughly 64% completing at least two semesters (one academic year). By building and retaining undergraduates, Team Phoenix has begun to fulfill its mission of building the "missing middle" by advancing course alumni to industry positions in Wall Street, Sandia, and NVIDIA. Additionally, students who have completed our VIP course have entered undergraduate or graduate HPC research at Georgia Tech.

In reviewing the demographics of students as reported in Table 1, a few trends quickly become apparent. Despite the multidisciplinary focus of the VIP program, and particularly the ubiquitous nature of computational research across all fields of science, students enrolling in the VIP are almost exclusively CS or CSE majors. Furthermore, females comprise roughly 10% of the total enrolled students, compared to the average of 20% in CS and Engineering degree programs [3], while minority participation sits well below averages, with Black/African American and Hispanic enrollment amounting to 3% and 6% compared to 8% and 12%, respectively [11]. In particular, it would be beneficial to develop and implement strategies to more effectively recruit and promote diversity within the Team Phoenix VIP.

3 INSTRUCTIONAL CLUSTERS

There are multiple HPC clusters that are available at Georgia Tech which the students have access to as students in the VIP Team Phoenix course. All clusters listed here use similar job schedulers (Slurm) in a Linux environment, and can be accessed using a shell environment or Open On-Demand. Table 2 summarizes the available node counts and hardware on each of the clusters.

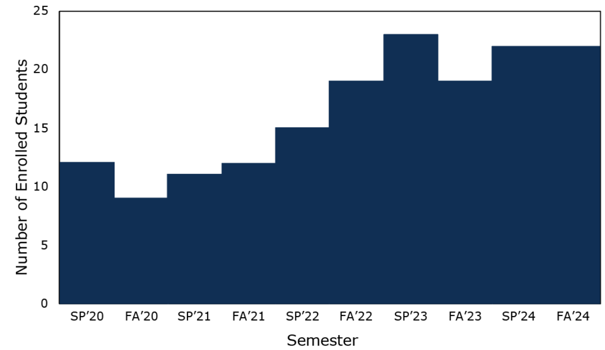

Figure 3: Team Phoenix VIP course enrollment by semester. Since its creation in the Spring 2020 semester, enrollment has increased 144% to 22 students in the current semester.

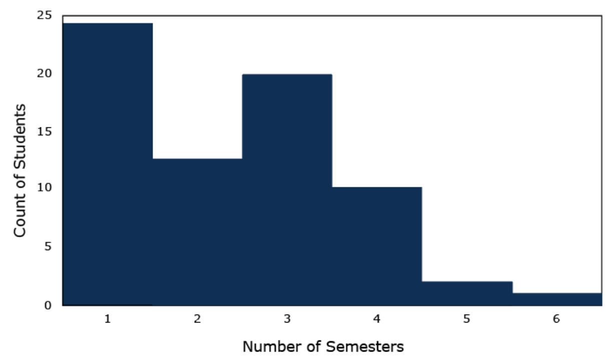

Figure 4: The count of students by numebr of semesters enrolled in the Team Phoenix VIP course. The vertically-integrated aspect of the VIP course hinges on students returning for successive semesters, and to date, Team Phoenix has seen roughly 64% come back for at least one additional semester.

3.1 PACE-Supported Clusters

The two clusters used by the VIP Team Phoenix course students include the Instructional Cluster Environment (ICE) and Phoenix clusters. The ICE cluster resources offer an educational environment that matches that of our research clusters and provides thousands of graduate and undergraduate students at Georgia Tech opportunities to gain first-hand experience with HPC. The Phoenix cluster is our research cluster.

For the VIP course, students typically use the ICE and Phoenix clusters to install and run examples in parallel computing, including running the LINPACK benchmark and parallel computing examples. However, using these clusters does require continual maintenance by the PACE team through Maintenance Periods and other updates,

Table 1: Breakdown of enrolled student demographics in Team Phoenix VIP.

Category	Number of Students			
Total Student Count	69			
Student-reported Gender	M: 62	F: 7		
Student-reported Race	Asian: 38 Two or more: 5	White: 31 Unknown: 2	Hispanic: 4	Black or AA: 2
Student Major	Comp Sci: 52 Elec Eng: 1	Comp Eng: 6 Comp Media: 2	Math: 4 Bio: 1	Mech Eng: 1 Elec Comp Eng: 1

and requires lectures and workshops throughout the semester for proper cluster usage.

3.2 CRNCH-Supported Cluster

The Rogues Gallery is a heterogeneous test bed of HPC servers maintained by the Center for Research into Novel Computing Hierarchies (CRNCH) for research in novel computing environments. For VIP Team Phoenix uses the Rogues Gallery for access to the Frozone cluster, a set of nodes which include two A100 GPUs. During the “Build-A-Bear” workshop, students actively disassemble and reassemble a GPU server, and afterword test HPC and GPU benchmarks on the cluster that the server is part of for the VIP course.

3.3 VIP Cluster

The Pho cluster is a test bed of HPC servers purchased using Georgia Tech’s technology fee grant for instructional courses to provide VIP Team Phoenix course students hands-on experience with setting up and cutomizing their own cluster environment. This system is a 9-node system that includes two GPU nodes, each with an A100 GPU; two high-memory nodes with 1.5 TB 2933 MHz DDR4 RAM; two Intel Optane nodes with 4x128 GB Optane Persistent Memory 100 series modules; two high-storage nodes with 12x4TB hard drives; and one general-purpose node.

4 BUILD-A-CLUSTER ASSIGNMENT

4.1 KVM Cluster Virtualization on the Pho Cluster

For the Build-A-Cluster assignment, we used Kernel-based Virtual Machines (KVMs), an open-source virtualization included with Linux, for cluster virtualization on the Pho cluster. This use of KVMs allows for flexible deployment of cluster nodes for periodic refreshes (i.e., changes between users between academic years and semesters), quick configuration updates for new projects or assignments during the semester, or quick resets of cluster environments in case of issues.

Each of the 9 nodes in the Pho cluster included common hardware, with differences in hardware between nodes to add additional capabilities to the cluster. All nodes in the cluster include the following hardware: dual-socket Intel Cascade Lake Gold 6226R (2.9 GHz clock speed, 16C/32T), 192 GB 2933 MHz DDR4 RAM, Cisco VIC 1387 with 40 GB Ethernet, and 960 GB SSD, mirrored SW RAID. The two storage nodes include an additional 12x4 TB hard disk drive for additional storage. The two high-memory nodes include 1.5 TB 2933 MHz DDR4 RAM instead of 192 GB. The two Optane nodes include an additional 4x128 GB Optane PMem 100 series memory. The two GPU nodes include an additional NVIDIA A100 GPU for each node. Centos Stream 8 was the Linux distribution installed on all of the nodes.

The KVM images built for HPC instruction were installed on each of the nodes and configured to utilize CPU, memory, and storage resources in the hardware, with some differences depending on the node type. Each of the KVMs was installed with Centos Stream 8, and configured to use all of the CPU resources (32 cores across dual-socket CPUs) and 50 GB of disk space. The memory used across the nodes varied depending on the node type (84 GB to 1360 TB). For this initial pilot, the GPU nodes were not configured to use the NVIDIA A100 GPUs but will be used in future projects.

For accessing the KVMs when running on the network, Mac addresses assigned on KVM setup were recorded so that they could be reused and re-mapped to a set virtual machine hostname and IP address on the network upon subsequent re-installations.

4.2 Alternatives

Compared to other solutions, such as the configurable cloud environment for large-scale HPC research, Chameleon Cloud [6], our KVM cluster virtualization has multiple advantages. First, our approach is on-demand and can be reconfigured as needed. Second, because KVMs are part of Linux, no special infrastructure is required to support it, and so any available hardware can be leveraged. A corollary to this detail is that as much as older hardware can be utilized, if newer hardware becomes available, it can also readily be supported. Thus, for Team Phoenix, or any other group looking to replicate this work, they can just as easily utilize surplus servers as well as the cutting edge systems supporting research.

Furthermore, the lightweight nature of the approach lends itself naturally to providing a secure sandbox in which students can build a cluster from scratch, without risk of irreversibly breaking the system. Ultimately, the intent is to run this setup atop a production cluster, which is discussed in greater detail in Section 6, to provide a sustainable solution that can accommodate the bursty nature of instructional cluster use while leveraging available cycles from critical infrastructure.

4.3 Build-A-Cluster Project in HPC Virtual Cluster

This Build-a-Cluster project that we assigned as the last project for the Team Phoenix VIP course for the Spring 2024 semester,

Table 2: A summary table of the clusters accessible to students in the Team Phoenix VIP.

Cluster	Node Count	CPU Cores	GPU Count
Phoenix (PACE)	1,407	36,104	V100 (120), RTX6000 (148), A100 (32), H100 (32), L40S (64)
ICE (PACE)	117	4,352	V100 (40), RTX6000 (8), A40 (4), A100 (8), H100 (160), L40S (32), MI210 (4)
Rogues Gallery (CRNCH)	61	2,772	A30 (8), A100 (17), H100 (1), H200 (6), MI210 (2), Max1100 (2)
Pho (COC)	9	288	A100 (2)

we sought to provide students with an HPC Virtual Cluster using KVMs on Linux servers in order to prepare for Supercomputing (or High-Performance Computing, HPC) student competitions. Our goal for this project was for students to get hands-on experience with:

- Working on hardware (or KVMs) on the HPC Virtual Cluster that would be similar to working on multi-node clusters used in the real world or student cluster competitions.
- Installing, configuring, and optimizing software typically used for essential infrastructure for an HPC cluster (Ansible, Slurm scheduler, compilers, MPI).
- Installing, configuring, and optimizing scientific and simulation software used for HPC research.

For this last Spring semester, we divided the class into two teams of six students to work on two separate HPC Virtual Clusters. For this initial pilot, an HPC Virtual Cluster was comprised of a 3-node cluster corresponding to 3 KVMs (i.e., pho-storage1-vm, pho-highmem1-vm, and pho-optane1-vm for Team 1). Each KVM was built from scratch, and storage mounted from an NFS partition on a corresponding bare-metal storage nodes. MAC address assigned from the initial creation of the KVM were noted for hostname and IP address assignment and for future re-installation of the KVM.

We divided this project into different sections, where we would start by 1. introducing the students to the HPC Virtual Clusters they would be working with, continue to software infrastructure setup and job scheduler installation on the 2. login/head and 3. compute nodes, and then move to 4. installing software, which included installing GCC, Open MPI, Intel oneAPI MKL, and HPL. Other optional steps that we wanted to include but realized we did not have time for in Spring 2024 included 5. installing Lua and Lmod to load software configurations and 6. installing Ansible and developing Ansible playbooks to repeat the initial setup and software installation steps after completing all of the previous steps one time manually.

5 ASSIGNMENT OUTCOMES

For the Spring 2024, both teams encountered challenges due to time constraints near the end of the semester, so we were not able to proceed past the “Initial Setup for Login/Head Node” and “Initial Setup for Compute Nodes” of the original project plan. However, we were not able to get to the “Install Additional Software” step as we intended with the original time frame. While we were not able to get to the point of installing software like HPL for LINPACK benchmarking, we did have other projects with HPL and Mr. Bayes earlier in the Spring 2024 semester. In this upcoming Fall 2024, we are starting at the beginning of the semester with an updated version of the Spring 2024 project that should include the optional steps, including installing additional software, utilizing Lua and Lmod to reload software configurations with ‘module,’ and utilizing Ansible for quick rebuilds of the virtual cluster.

6 FUTURE WORK

In the future, we have additional ideas we plan to implement to improve this virtual cluster approach to HPC instruction. The first idea we plan to implement is using Slurm on the bare-metal machines to launch job batch files that deploy the virtual clusters on-demand. Using the host cluster Slurm’s suspend & resume framework, which was originally intended for cloud node management, we can launch our minimal VM images or a full production machine as needed. In this way, we can operate parasitically atop the production cluster, providing a more sustainable solution to support Team Phoenix and other classes that may want to explore cluster building projects.

The second idea we plan to integrate is a section on file systems for HPC systems, and how selecting different approaches and optimizing file systems could improve HPC benchmark performance. We plan to add mini project explaining the current state-of-the-art for file systems in HPC clusters, with a workshop to illustrate some of the setup involved with our current file system in the virtual cluster. In the future, we could also look into what would be required to add newer parallel file system hardware to the Pho cluster that students could set up, configure, and test HPC benchmark performance.

The third idea would be to look at newer, more advanced hardware to utilize with this virtual cluster approach, perhaps even to optimize it for better deployment in instructional settings such as this. Currently, the Pho servers use Cisco remote system management that makes it difficult for us to reconfigure the bare-metal servers easily. With the aforementioned suspend & resume framework, we could look at using the novel hardware available in the Rogues Gallery, or even consider a larger scale project using ICE compute nodes.

7 CONCLUSION

A myriad of hurdles present a barrier for entry into HPC, a field crucial to research in academia and industry in areas that include natural sciences, engineering, and computing. These roadblocks include,

1. Introduction to HPC Virtual Cluster
2. Initial Setup for Login/Head Node
   (a) Mount network mountable storage system (already completed)
   (b) Set up user accounts
   (c) Set up Slurm (DB and controller on login/head node)
   (d) Turn off firewall
3. Initial Setup for Compute Nodes
   (a) Mount NFS storage system (already completed)
   (b) Set up user accounts
   (c) Set up Slurm clients (on login/head node and other 2 compute nodes)
   (d) Turn off firewall
4. Install Additional Software
   (a) Install GCC (compiler)
   (b) Install Open MPI (for MPI support)
   (c) Intel oneAPI MKL (for BLAS support)
   (d) Install HPL (for Linpack benchmark)
   (e) Install additional software used in student cluster competition (Optional)
5. (Optional) Install Lua and Lmod to use 'module' to load software configurations
6. (Optional) Install Ansible and develop Ansible playbooks to rebuild the cluster from scratch

Figure 5: HPC Virtual Cluster Project Sections

but are not limited to, *i*) access to HPC infrastructure, *ii*) maintenance support consisting of advanced and niche knowledge, and *iii*) costly hardware. To fill this knowledge and personnel gap within the HPC industry, higher education institutions should invest in the training and teaching of supercomputing concepts for their students. To overcome traditional and academic challenges prevalent in HPC, we believe the **HPC Virtual Cluster** provides a hardware-agnostic platform designed to provide an easy-to-configure, generalizable, and scalable approach to HPC system management for training and education.

REFERENCES

[1] 2009. Is HPC going to cost me a fortune? https://insidehpc.com/hpc-basic-training/is-hpc-going-to-cost-me-a-fortune

[2] B. P. Abbott and et al. 2016. Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.* 116 (Feb 2016), 061102. Issue 6. https://doi.org/10.1103/PhysRevLett.116.061102

[3] Sapna Cheryan, Allison Master, and Andrew Meltzoff. 2022. There Are Too Few Women in Computer Science and Engineering. https://www.scientificamerican.com/article/there-are-too-few-women-in-computer-science-and-engineering/

[4] Nicole Hemsoth. [n. d.]. Engaging the Missing Middle in HPC. https://www.hpcwire.com/2010/06/07/engaging_the_missing_middle_in_hpc/

[5] A. Jezghani, J. Young, W. Powell, R. Rahaman, and J. Coulter. 2023. Future Computing with the Rogues Gallery. In *2023 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*. IEEE Computer Society, Los Alamitos, CA, USA, 262–269. https://doi.org/10.1109/IPDPSW59300.2023.00051

[6] Kate Keahey, Jason Anderson, Zhuo Zhen, Pierre Riteau, Paul Ruth, Dan Stanzione, Mert Cevik, Jacob Colleran, Haryadi S. Gunawi, Cody Hammock, Joe Mambretti, Alexander Barnes, François Halbach, Alex Rocha, and Joe Stubbs. 2020. Lessons Learned from the Chameleon Testbed. In *Proceedings of the 2020 USENIX Annual Technical Conference (USENIX ATC '20)*. USENIX Association.

[7] Tobias Mann. [n. d.]. A closer look at Nvidia's 120kW DGX GB200 NVL72 rack system. https://www.theregister.com/2024/03/21/nvidia_dgx_gb200_nvk72/

[8] Stephen Marshall, Edward Coyle, James V Krogmeier, Randal T Abler, Amos Johnson, and Brian E Gilchrist. 2014. The vertically integrated projects (VIP) program: leveraging faculty research interests to transform undergraduate STEM education. In *Transforming Institutions: 21st Century Undergraduate STEM Education Conference.*

[9] Nvidia. 2023. NVIDIA Investor Presentation October 2023. https://investor.nvidia.com/events-and-presentations/presentations/presentation-details/2023/NVIDIA-Investor-Presentation-October-2023/default.aspx

[10] Elizabeth Pennisi. 2013. The Man Who Bottled Evolution. *Science* 342, 6160 (2013), 790–793. https://doi.org/10.1126/science.342.6160.790

[11] Parth Sarin. 2023. Levers for Improving Diversity in Computer Science. https://cset.georgetown.edu/article/levers-for-improving-diversity-in-computer-science/

[12] Emily Sayegh. 2024. The Great Tech Reset: Unpacking the Layoff Surge of 2024. https://www.forbes.com/sites/emilsayegh/2024/08/19/the-great-tech-reset-unpacking-the-layoff-surge-of-2024/

[13] Anton Shilov. 2024. Intel's 1500W TDP for Falcon Shores AI processor confirmed — next-gen AI chip consumes more power than Nvidia's B200. https://www.tomshardware.com/pc-components/gpus/intels-1500w-tdp-for-falcon-shores-ai-processor-confirmed-consumes-more-power-than-nvidias-b200

[14] Anton Shilov. 2024. Nvidia's next-gen Blackwell AI Superchips could cost up to $70,000 — fully-equipped server racks reportedly range up to $3,000,000 or more. https://www.tomshardware.com/pc-components/gpus/nvidias-next-gen-blackwell-ai-gpus-to-cost-up-to-dollar70000-fully-equipped-servers-range-up-to-dollar3000000-report