Season 2, Episode 5:

Advancing Computing Performance and Workforce Diversity

SPEAKERS
Sarah Webb, Valerie Taylor

Sarah Webb 00:00
I'm your host Sarah Webb, and this is Science in Parallel, a podcast about people and projects in computational science. In this episode, you'll hear from Valerie Taylor, the director of the mathematics and computer science division at Argonne National Laboratory. Before Argonne, Valerie completed her Ph.D. in electrical engineering and computer science at the University of California, Berkeley.

She spent more than 25 years working in academia at Northwestern and Texas A&M. Valerie's work at Argonne involves both management and research. She supervises teams working on numerical methods, AI, data management, storage, new computer architectures, and more. Her research explores computing methods that optimize power usage and application efficiency. And she's the principal investigator on Threadwork, a new DOE-funded project on microelectronics.

Valerie has invested significant time and effort supporting scientists from communities that are underrepresented in computing. Much of that has been through CMD-IT ("command-it") a nonprofit organization that she cofounded and now runs. We talked about how she got her start soldering circuits, the challenge of today's complex computer architectures and the sustained commitment needed to build a more diverse, equal and inclusive computing workforce.

Sarah Webb 01:32
Welcome, Valerie, it's great to have you on Science in Parallel.

Valerie Taylor 01:35
Thank you. It's a pleasure to be here, Sarah,

Sarah Webb 01:38
Can you give me a little bit of background and how you in particular got interested in these particular problems, and what have been the challenges as you've been working on those problems?

Valerie Taylor 01:47
My interest in computing started in high school. This was back during the time when we had punch cards. I took Fortran and it was great in terms of the functionality and what you were able to do with the

©2022 Science in Parallel/Krell Institute

computer through the programming. This continued on in terms of undergrad, and when I got to grad school, I started to really enjoy work related to parallelism and looking at ways that you can actually efficiently utilize the parallelism that's available. Because I began to see that as you had, let's say, 20 processors, it's not always the case that you would be 20 times faster. So it became the curiosity: Why not? Why can't we get to 20? Sometimes we will be just 10 times faster. And I'm like, "No, I want to get to 20." So I think it was that curiosity, to actually want to see, you know, the efficient use of the resources that really sparked my interest. And my work in grad school was around sparse matrix computations and doing that efficiently.

Valerie Taylor 03:04
After grad school, when I started at Northwestern, I had a collaboration with Argonne. The collaboration at Argonne really excited me about continuing to work with performance analysis and with efficiency, but really focused on the science and scientific applications and understanding what is going on when the application executes on the system. And I think that aspect allows me to combine the interest in computer architecture, also understanding the software stack, and then having knowledge and working in detail with the application. I think the performance analysis allows you to combine the interest, hardware, software and the application.

Valerie Taylor 03:55
So with all of this complexity that we're seeing that's available in high performance computing systems, and that heterogeneity, such that now when you look at performance, you're not just looking at performance of a CPU, you're looking at the performance across the system, and taking into account now that complexity of what does a node look like how are the nodes configured? So it's an exciting time. And then you also want the power efficiency. You can also take that another step with the power efficiency to say, now what about the impact on the environment? What's the carbon emission? And what's that impact? So looking at the full ecosystem. I think it's an exciting time to be working in the area of high performance computing.

Sarah Webb 04:49
Valerie and I spoke before Frontier, Oak Ridge National Laboratories new supercomputer, crossed the exascale barrier in May. Like so many others, she was excited about the possibilities for science areas such as climate modeling, materials design, and quantum information sciences. She also talked about the challenges of heterogeneity, not just CPUs and graphics processing units, or GPUs, but the integration of new technologies such as quantum processors, and specialized chips that accelerate AI applications.

Valerie Taylor 05:21
Now, we're seeing the systems for example, Frontier and also Aurora, the exascale systems, will have CPUs and GPUs. Another aspect is coupling those systems with AI accelerators. Then, I think, is also coupling those systems with advances that will occur in the quantum information science space. So you can have quantum processing units also connected with the system in the long term. There's also, for example, the ecosystem of seeing, you have sensors out in the field, you can do some computing, and more than likely you're looking at doing AI at the edge, where you can do the data analysis at the edge. So is not where you're having always to send so much data to an HPC system or cloud system.

Valerie Taylor 06:22
However, while you're doing that analysis at the edge, there's still some simulation that you want to do with that data. So you may be doing some data collection, have that coupled with an HPC system as well. So now you're looking at these complex workflows that may involve, let's say, a self-driving lab, where you still may have a human in the loop. But that lab in terms of the experiments that are being done, it may be based on the simulation. And that may mean that you conduct one set of experiments, get characterization, that's fed to simulation. And that simulation's results: it's probably utilizing AI in terms of possibly surrogate models with the simulation. Those results feed back to, let's say, robotics in the lab, this actually identifying the next set of experiments to be done.

Valerie Taylor 07:24
Now you need a software stack that actually supports that. So you're getting complexity, you're also including the network that's connecting the system. When you want to look at energy efficiency, it's across all of these aspects. Now, let's add on one other piece, as we look at where we need more energy efficiency, can we start to do work at the materials level? And that work at the materials level may be a decade out, but you're starting to do research at the material level, the material science level, materials design level that is based upon application needs. So now your codesign with respect to all the way down to the materials, looking at what you need with the applications. When you look at different axes of systems, system software application, and even going further down in the design stack to the materials. You just go, oh, there's a lot of opportunity much work to be done various levels of complexity. And it just provides a very exciting timeframe.

Sarah Webb 08:42
Valerie and her colleagues are working on these questions. She's leading a Department of Energy-funded project called Threadwork to design new microelectronics for the chips, devices and science instruments of the future. This research could help with new neuromorphic devices. Those chips mimic the way that the brain signals. And the terahertz interconnect: devices that can bridge between optical and electrical signals. Some of the materials work will support the development of memtransistors, devices that combine transistors with programmable resistive elements.

Valerie Taylor 09:18
So with the microelectronics project, it was just recently funded by the Department of Energy. And they funded 10 projects in total across the 17 labs, our project in which we call it Threadwork. And the Threadwork is really focused on that codesign aspects of going from materials, looking at its impact on devices, which leads to compute systems, and then looking at the impact on applications. So we look at that codesign framework as really having that opportunity for pairwise interaction, where it doesn't have to go through your different levels of a system, devices, the compute system, the software stack, and the application stack.

Valerie Taylor 10:10
So we have applications staff and researchers talking with materials researchers. So it's not just each adjacent level talking to each other. We want pairwise interactions such that, for example, it involves researchers at Argonne, also Northwestern, that's doing work in terms of neuromorphic devices. Also,

terahertz interconnect, as well as looking at our application that we're looking at has to do with detectors. And that's in the area of high energy physics and nuclear physics, in terms of their detectors, that really, the high energy physics and nuclear physics in terms of the community, they look 10 years out of what's needed with their detectors.

Valerie Taylor 11:03
So working with that community in terms of what's needed with the detectors, the amount of data that's going to be collected, also, there's going to be power requirements, which get to energy efficiency. So what are the computations that's going to be done? Can we utilize neuromorphic computing to do some of the data analysis? So looking at those type of applications with the detectors, and having that discussion now at the materials level. And that is to say, at the materials level, we're looking at particular type of devices related with memtransistors, what becomes the impact now and looking at the detectors? And you're looking at that with that pairwise interaction, so you're looking at the type of computations, how these devices can be used. What is that relationship? The same is occurring with terahertz interconnects. And that is we can have terahertz interconnects, but those last few centimeters that you're looking at going from electrical to optical conversion, that often can slow things down and can require quite a bit of area. And so that's where we're looking at utilizing plasmonics with respect to the terahertz interconnect.

Valerie Taylor 12:31
And so what does that mean in terms of being able to now interconnect and go chip to chip? And what does that impact in terms of now looking at the detectors? You're looking at all of these aspects. So with the detectors, what are some of the anomalies that they have to take into consideration in terms of their detectors? Can we design robust algorithms to take advantage of that? So we also have work being done with algorithms, where you're looking at now the pairwise connection between algorithms and applications and the pairwise connection between algorithms and the devices and the materials as well. So our project is really that codesign aspect and looking at it too, with the pairwise codesign and not just going between adjacent levels.

Sarah Webb 13:27
Valerie and I also talked about the people and experiences that shaped her interest in computer science and her career path.

Valerie Taylor 13:34
If I go back to childhood, I would say it was really from my father. So my father is an electrical engineer. And he started a communications company. Initially he worked with IIT Research Institute or IITRI. But then he and some of his friends started a company with defense contracts of developing that communication equipment. They opened up shop and built boards, did the design. So I grew up in an environment where we always had tools. I grew up building circuit boards. And so I understood, even before going to college, how to read a schematic and how to build a circuit board because that's what my father did. Even now, you know, if you go to my home, I have my tools, and I do have some circuit boards at home. But it was the norm.

Valerie Taylor 14:35

Someone asked me: what's a smell that reminds you of childhood? And I go: solder. In high school. I worked with Sonicraft on the assembly line, and I got to the point where I could solder some things and didn't burn the wire. When I was an undergrad at Purdue, faculty encouraged me to go to grad school. I knew I wanted to get a masters after finishing the bachelor's. I was like, oh, there's more to know. And I'm curious. My senior year, I was what they called a dualie: I was finishing undergrad while starting grad school. And then I had one faculty member at Purdue: Aren't you going for a PhD? And I was like, no. Why not? You should definitely go for a Ph.D. Then I had my research advisor also was a strong encouragement to get the Ph.D. I then went on to Berkeley and completed the Ph.D., and it was a very supportive environment. It was very few Blacks getting a Ph.D, and Berkeley worked hard to make certain we had a community. And it was in large part to a person, Sheila Humphreys, that would come by and say: What do you need? How are you doing? That was really good. They were very few Blacks. So you can imagine, there were even fewer Black women, that part can be very isolating, but it was still good in terms of the community. And the fact that we all stay in contact, and reach out to each other often to congratulate each other.

Sarah Webb 15:52
That sort of network and that sort of support is just so fantastic. Well, I wanted to ask you about your policy and service work supporting diversity, equity and inclusion, because I know that's been a big thing that you've been involved with. Can you tell me where that started for you?

Valerie Taylor 16:28
it started at home because my father was very active in the community. And he would always say, you should always take time to give back to community. Just as someone helped you, you need to help others. So now, a group of us in 2011, we started a Center for Minorities, and People with Disabilities, and we call it "Command-It." CMD, computing command-- when you had command line arguments- so command it CMD dash IT. And the reason for starting this organization was that we recognized there was a lot of work that was being done in terms of gender diversity, and that was very much needed. But when it came to cultural diversity, there wasn't as much being done. Plus, we looked at the community and said, Okay, there's work being done with Blacks/African Americans, work being done with Hispanics/LatinX also working with Native Americans/Indigenous people, and then people with disabilities.

Valerie Taylor 17:42
Each community has unique characteristics, but the commonality has to do with culture. Understanding culture requires intentionality. Because with gender, we all have women in our families, everybody. [laughter] So it's a bit easier to understand gender, because more than likely, a young woman is coming in your family to say, you're not going to believe what happened to me today and how I was discouraged when I said I wanted to do math and science. However, it's not where you can say that about culture. Not everyone has a Black child coming home, a Hispanic child, Native American child. So what does it mean to really understand culture? And that was the reason for looking across these four communities to say when we look at each individual community the numbers are small, but can we do some things that go across the four communities that really highlight the importance of being intentional about culture.

Sarah Webb 19:02
In 2018, Valerie co wrote an article in *IEEE Computer* about women and underrepresented minorities in computing, including steps that organizations and individuals can take to increase diversity and inclusion. I've included a link to that article in the show notes. Here's one assessment: "A sea change is happening now. And we are excited about the potential for transformation. Having said that change requires persistence and a sustainable commitment. It is still too early to know whether these changes will stick." I asked Valerie to reflect on where we are in 2022. Where are we making progress? And where do we need to continue to push that commitment?

Valerie Taylor 19:50
My perspective is based upon my experience. I think a key year is actually 2020, especially with the death of George Floyd and the racial unrest in the US sparked a lot of dialogue. And the circumstances around George Floyd: the fact that we were in a pandemic, when everybody was home, and they could view the video, and then you had different protests and marches. And everybody came together. And then we saw with the justice system that the white cop was found guilty. I bring that up because I think that's a key point in which you had a lot of dialogue occurring in companies and universities, where at universities, you had the Black students, the Hispanic students, the Native American students, all the marginalized communities coming together to say: We want change.

Valerie Taylor 21:08
And now we're in 2022. And the question that comes up, and this has to do with the comment about persistence and having sustainable change. Oftentimes, change comes about because you have an individual who is pushing or a group that is pushing things forward. And what happens when that group, or that individual, is no longer pushing things forward for a variety of reasons? And I think that's the question that we need to ask now. Let's say the person that's pushing things forward, they retire. Is it that is all of a sudden forgotten? Are there policies put in place? Do people just go back to normal of doing nothing? In which case, people value and will rise up to say, why did that stop? We need that program. And so it needs to be everybody engaged in valuing that diversity, not just a few.

Valerie Taylor 22:19
So I worry a great deal about the persistence, sustainability, and really having that strong voice to bring about a change. Or you can have it where, you know, people get so tired of going against the wall, and then become silent, which is even worse. And it's very interesting because right now, a lot of people attributed to the pandemic, of people moving, you're hearing a lot about people moving jobs. But I often wonder about the demographics of people who are leaving positions. And is it all because the pandemic has made it easier for people to leave? Or is it with the heightened discussion around, you know, racial injustice and systemic racism That people just say, I'm fed up. I want to leave. So it would be good to look at things at a national scale of what's going on, especially in terms of tech and identify what are some trends disaggregating data, and disaggregating data to from an intersectional approach.

Valerie Taylor 23:48
I was involved with the National Academy of Science, Engineering and Medicine, where there's a Committee on Women in Science, Engineering and Medicine (CWISM) that just published a report on Transforming Trajectories of Women of Color in Tech. And it really advocates for disaggregating data

from an intersectional lens. And that's so important, because if you have programs about women, if you don't disaggregate the data, you have no idea if those programs are effective for your women of color. It's so important to disaggregate data and not just have it as this monolithic view. Even among Blacks, did it work well for Black men? What about Black women? And, and I think you also have to look at gender and gender fluidity.

Valerie Taylor 24:51
I think we're at a point where we're talking about AI/ML because we have this flood of data. We have the computing resources to do the analysis. We're also at a point where we're looking at racial injustice, systemic racism. And so not that we want to take AI and apply it to racial injustice. But we have a lot of data that we can leverage from AI and ML to do analysis and disaggregate the data and to look at these trends, to better understand what's going on with different communities.

It's those implicit things that if you walk in the room, if you're the only woman, the only Black, do they assume that you're there to serve lunch? Or do they assume that you're part of the group? People are having the discussions, and there's awareness. But there's a lot more to be done.

Sarah Webb 26:00
When you're talking with early career researchers, what is the key piece of advice that you usually try to pass along?

Valerie Taylor 26:09
The one aspect is to be curious and ask questions about things. Often, innovation results from questioning why something is done a particular way, or why things occurred in that way. So take the opportunity to be curious. And also, to know that your voice matters. Take the opportunity to bring your full self to what you do because your voice and your experiences matter.

Sarah Webb 26:47
To learn more about Valerie Taylor, the Threadwork project and CMD-IT, please check out this episode's show notes at scienceinparallel.org. An article based on this interview also appears online in ASCR Discovery.

Sarah Webb 27:03
Science in Parallel is produced by the Krell Institute and highlights computational science with a particular focus on work by fellows and alumni of the DOE Computational Science Graduate Fellowship program. Krell manages this program for the U.S, Department of Energy. Our music is by Steve O'Reilly. This episode was produced by Sarah Webb with editing and sound mixing by Tess Hanson.