

oneAPI Code Together Podcast
Behind the Scenes: Arming Researchers with Visualization Know-How
Transcript

Brenda Christoffer: Welcome to Code Together, a discussion series exploring the possibilities of cross-architecture development with those who live it. I'm your host, Brenda Christoffer. Today our topic is scientific visualization. It's a process of representing raw scientific data as images, and these images improve scientists' interpretations of large data sets. They provide insights that may be overlooked by traditional statistical methods alone. Today our guest is Anne Bowen, she's a research engineering scientist at the University of Texas at Austin. She works at the [Texas Advanced Computing Center](#), known as TACC. TACC is one of the world's leading supercomputer centers and an Intel oneAPI Center of Excellence. Anne's background is in computational chemistry, which she uses to help scientists design scalable visualization technologies. Welcome Anne.

Anne Bowen: Hi.

Brenda: Our next guest is Donna Nemshick, she's a performance validation lead at Intel's Advanced Rendering and Visualization group. Donna leads performance analysis of the [\[Intel®\] oneAPI Rendering Toolkit](#) components on many types of Intel CPU and GPU platforms, from clients to huge HPC clusters. Donna, thanks so much for joining.

Donna Nemshick: Hi, Brenda. It's good to be here.

Brenda: So let's kick this off. Anne, could you share with us - how did you get started in scientific visualization?

Anne: Yeah, sure. So as an undergrad, I worked at the San Diego supercomputer center. And I had an internship analyzing reaction dynamics data and creating a visual analysis tool. And that really helped me see the power of scientific visualization, especially with the concept so abstract as chemistry. So I was really drawn to the power of using visual tools to visualize the abstract. And from there, I kind of branched to other areas of digitalization and really enjoyed being part of the research.

And after that, I went to pursue my degree in computational chemistry, where I really focused on looking at receptor surfaces and the properties at molecular surfaces and visualizing techniques for those. So after a completion of my degree, I really missed the research environment of a computational research center. And I wanted to go back to kind of working with diverse data sets. And I wanted to stay in visualization, so TACC for me was the perfect place to land.

TACC provides high performance computational resources for researchers across the nation. And that includes the development and support of tools and software for data analysis of visualization. So that includes the direct support and training of the researchers that use our resources, as well as the software and hardware stack. So I work one on one with researchers and also provide training. So it's kind of the best of all of my interests from my time as an undergrad and grad school. So I've really enjoyed my time at TACC. How about you, Donna? How did you get started in visualization?

Donna: I have an interesting path that got me to vis. So my 30-year career has been served well by my electric engineering and computer science degrees. So I've worked in three major segments of the industry. Being the telecom network switching industry, HPC and currently advanced rendering and visualization architecture. But how I serve in the vis role was key at doing embedded software development, which is the layer of software that's really embedded in the hardware. And I've always had a role in performance validation and optimization throughout all those segments of my career. And had some key initiatives going on with release and DevOps management.

So in vis with my primary role as performance validation lead, I really am getting into how fast do our software libraries work on the platforms being released at Intel, and any optimizations that can be had. So I have a goal to drive key initiatives in terms of performance, accessibility of software, and ease of use of our software. And let me not forget to mention I was a part-time adjunct professor at the Penn State University where I taught computer science. So this whole role of instructing folks and guiding folks and working in the software industry, no matter what the layers areas of the stack that I'm present in, it's just a whole lot of fun for me.

Anne: It must be really satisfying also to be part of making the tools easier to use. I remember as a graduate student, I was in the chemistry department and I was really the only person interested in computational tools or visualization. So I was kind of on my own, I didn't really know what libraries or what software to use. And I had to figure out a lot via trial and error. And a lot of time was spent trying to figure out what path to take. And then after I arrived at TACC and I really learned about all these tools, I probably could have saved a couple years from my PhD. I would've known about these things sooner. I really have a lot of sympathy for helping these researchers and helping guide them towards the right tool to use and also to help them become more aware of what tools are available with their data.

Donna: That's interesting you say that, because your role is you're enabling researchers to visualize their data sets and you've grown to be the expert at TACC in this regard. You've been doing this for quite a while, how has visualization of data evolved over your tenure in instructing researchers how to use the software?

Anne: Well I feel like now visualization is really an integral part of a researcher's analysis toolkit. I feel like it used to be a lot more in the past. Researchers would come to us and maybe ask for just a finale vis where they would want something for a publication or a presentation that was really impressive. And we're seeing more and more now that they want to use a software like Peer Review, which is a visualization software that we support at TACC. Every step of the way, they wanted to get a preliminary review of their data, they want to interact with their data and use it as a debugging and analysis tool. So that's one of the main ways that this software stack has changed, especially with training because we see that reflected in the people that are coming to our workshops and kind of what they want. They want to learn to use the tools themselves.

A good example of that. We had a collaboration with an experimental x-ray crystallographer at the University of Texas at El Paso. And he works with using experimental methods to determine the structure of biological molecules. So that ranges from rhinoviruses, which cause a common cold to really large marine viruses, which are very, not very well studied. So he determined the structure of this really large marine virus and wanted to get it up on his high resolution display and didn't even have any software that he could use to visualize it at all. So we worked with him and actually we worked directly with Intel at that point to come up with an application that he could use. And I think it was an earlier version of the [\[Intel®\] OSPRay](#) plug-in, it was a standalone viewer. And he used that and was able to successfully view his huge virus, which is actually called a girus on the display wall.

Fast forward to this last summer, we had a postdoc for in his lab that came wanting to interact with this girus and hoping to get our help. But now instead of having to have a custom solution, which was time intensive and required several people to help, he could just use [ParaView](#) and use that to load and interact with this virus. And also there's an application that's really popular with molecular scientists called [VMD](#) and he is also able to load it into VMD. So the OSPRay plug-in is now part of ParaView and VMD, so that was really nice. There was no development process. He was just able to look at his data directly.

Donna: Were you or your researchers able to get a chance to explore the oneAPI Rendering Toolkit yet, by chance?

Anne: No, not yet.

Donna: Oh, you'll find it really interesting. So the oneAPI Rendering Toolkit includes the [Intel®] [Embree](#), OSPRay, [Open VKL \[Volume Kernel Library\]](#), [OpenSWR](#), [OSPRay Studio](#). The components you're familiar with already. I know OSPRay is being integrated into ParaView, as you said. So what the oneAPI Rendering Toolkit does is it provides this solution that quickly enables visualization folks, the researchers, to bring up the software. How this works behind the scenes is, we pull the software from open source, the open-source component you're using. And we build binaries out of them. We add a bit of glue, as I'd like to call it, to support the ease of use. So when these get packaged up with the glue, you install them on your computer, your desktop, your server, an HPC cluster.

Donna: What happens is after the install, you're on one setup script and all your pads and libraries are set up automatically. So there's no more stressing over compatibilities of versions of source code to use or manual build processes. You really benefit tremendously from it. I have to ask though, so the software evolution has really impacted research. In your opinion, do you have any other exam of how research has just simply exploded and the scientists benefiting so much from what it is that the vis offers them for their research?

Anne: Yes. I think researchers are able to gain so much more insight from their data when they can interact with it. As I mentioned, River from UTEP wanted to be able to interact with this girus molecule. When researchers are able to interact with their data in real time, they can get so much more out of it than the old protocol of rendering an image, and then interacting with your data, and then rendering an image with minutes to hours in between that process.

Anne: For example, during the pandemic, I was working with an oceanographer on the hydrodynamics of Lake Champlain, which is a very cold lake in Vermont. And since it was COVID time, all of our interactions were over Zoom. And I was looking at his model of the lake data on [Frontera](#), and interacting with it in ParaView, because he also wasn't extremely technical, even though he's renowned in his field with oceanography. He did not want to learn to use ParaView and wasn't really familiar with the data, even though he was extremely familiar with the science.

So it was a really interesting collaboration, where I was interacting with the data in ParaView on Frontera, sharing my screen. And then he was commenting on what he was seeing, and what he expected to see and what he wanted to see. And then I was able to in real-time, change things like the different ISO surfaces and the cutting planes to help him explore the data. And that just simply wouldn't be possible if it weren't for the interactive frame rates and the speeds. So it was really a satisfying collaboration, even though it's kind of still new to do everything 100% on Zoom, but it worked pretty well, worked really well.

Donna: So performance is key in my role, as I say, and you're kind of alluding to that, right? So performance comes into play in various contexts, right? For example, CPU or GPU, core speeds, memory speeds, memory latency, capacity, fragmentation. And the compiled capabilities or source code, algorithm efficiencies, threading, parallel computation, the list goes on. But all of these impacting the rates at which data can be visualized and the visualization interacted with in real-time as you're mentioning.

So let me ask a rhetorical question, who instructs the instructor? You mentioned you have to learn all of these tools on your own and then provide instruction. So one of the areas I am driving is the completion of a program called the oneAPI Rendering Toolkit Certified Instructor Program. So what this is, is self-paced web-based program and you must pass a test in order to become officially certified. But the program will enable instructors like you to know and understand the features of the Rendering Toolkit software and how to use them. So there'll be various area of certification available. Rollout is targeted for first half of '22, but I'm thinking, would such a program benefit you Anne?

Anne: Yes, definitely. I don't remember if I explicitly mentioned it, but every summer we offer a visualization training workshop for researchers that lasts about a week. It would be really great to be able to enable our researchers to use those tools.

Donna: Fantastic, we'll have to get you involved into the program. So let me ask this now. What has been, say your biggest wow moment, in vis?

Anne: I don't know if I have a wow moment, but I've been really inspired. I've been most inspired I think when I've seen the scientists I've been working with kind of get new insights into their data. Whether it be a more holistic view, so maybe they had only looked at a small chunk of their data on their laptop, and then because they're using the resources at TACC, they're able to look at all of their data.

I think my favorite example, we have a researcher, he was a plasma physicist, had been working in that field for 50 years, many breakthrough discoveries in that time. And he had only ever worked with 2D slices of his data that he had created a new plot. So he reached out to us, and my colleague, Greg Foss, who also has an art background, made these stunning 3D visualizations of these plasma structures. And it was just amazing to hear the scientists talk about the new structures that he was seeing based on this 3D knowledge. And we actually were able to get the same structure into the [Microsoft HoloLens](#). So we actually have video footage of the scientist and his reaction as he was looking at this HoloLens in the room with him.

So after an entire career of analyzing the data in a 2D way, to be able to finally see it in 3D was amazing. And also it was just amazing to me how much he could do looking at the 2D slices of the data. I'll have to show them to you because it almost just looks like a graph. As opposed to when you see the whole thing in 3D and animated. It looks like a ghost or something, it looks like it a life of its own. So it's amazing.

Donna: So I bet you say wow, a lot like myself as well. My biggest wow moments. Well, I don't know, there's so many of them. Every time folks on the team send out some really cool vis that they've done. I'm just like, it's incredible, it's magical. For me in the performance area and just given the different various segments that I've been involved in. I kind of get really excited about the big data aspect of it. How big data sets can now be managed. And one of the really cool things for me was [\[Intel®\] Optane™ Memory](#). And I believe TACC has that integrated in their clusters. But what an incredible way to expand the footprint of your memory and pull in these terabyte size data sets. To me that was just incredible, how that gets pulled off. I'm sure you had to change to work on those clusters in TACC. It's just amazing, the amount of data is just amazing.

Brenda: Well, we're nearly out of time. I like to wrap up these interviews with, Anne, where do you see scientific visualization heading in the next five years?

Anne: I see visualization going more into the hands of the researchers. I've mentioned that just in the time that I've been in vis, I've seen a big increase in the number of scientists that want to use visualization as part of their data analysis toolkit. And I feel like that will continue in that direction where we'll see scientists being able to work with larger data sets using commodity software like ParaView, that's been enabled by the oneAPI and other things that are easy to use, easy enough that they can do it themselves.

I'm really looking forward to the software and the hardware improving with the rate of data production. I know the data sets are getting larger and larger, and I'm really excited to hear all the things that Donna has said about oneAPI and the pace of Intel's development. And I feel like the future is looking very bright that we will be able to keep pace with a deluge of big data that's coming in, and continue to help the researchers reach their goals.

Donna: Yeah. We look forward to continuing to work with TACC to enable advancements in their clusters from the hard and software perspectives.

Brenda: Donna, how about you, where do you see scientific visualization heading in the future?

Donna: Oh, I can take this off in many tangents, Brenda. I guess maybe the one I'll point out here is given the context of our software instruction and really getting the researchers up and going quickly. I think the industry specifications are playing a bigger role in vis. Namely the [Khronos ANARI spec](#) and the [oneAPI specification](#). These would truly enable software to be “written one time” and yet be able to work on a variety of different platforms. And I can just tie that into this conversation here with Anne. In the instructing of the researchers, is it really simplifies, not only the visualization software, but even the know-how around the software. So I think we're going to see a lot of movement there in the next few years.

Brenda: Anne, could you give us some resources, where can people learn more?

Anne: Yeah, so there's a web site, [Learn.TACC.UTexas.edu](#), which kind of summarizes our training products. And also just our main web site, [TACC.UTexas.edu](#) kind of tells more about TACC in general.

Brenda: And Donna, are there any resources from Intel that you'd like to point out?

Donna: Oh yes. [RenderingToolkit.org](#). If you go to that one spot, it is your landing zone for everything about the Rendering Toolkit. And will guide you to each component and deep dive into the technical details as well as get you all key links you need for downloading the software binaries and/or getting them from open source.

Brenda: Well, I want to thank you both. Anne, Donna, this has just been such an exciting topic, and bringing the visualization to reality for our scientists is just really incredible work. So thank you for all that you do. And I also want to thank all of our listeners for joining us today. Let's continue the conversation at [oneapi.com](#).

-#-