

Rashmi Mohan: This is ACM ByteCast, a podcast series from the Association for Computing Machinery, the world's largest educational and scientific computing society. We talk to researchers, practitioners, and innovators who are at the intersection of computing research and practice. They share their experience, the lessons they've learned and their own visions for the future of computing. I am your host Rashmi Mohan. Reeling from a global pandemic and its impact on our healthcare systems, we have often paused to wonder how we can improve outcomes for patients with the plethora of technology solutions at our disposal.

The marriage of biological sciences and high performance computing is one such discipline that is integral to this area. Our next guest is a pioneer in this field. Amanda Randles is the Alfred Winborne and Victoria Stover Mordecai, assistant professor of biomedical sciences at Duke University. She's been the recipient of the National Science Foundation Career Award, Grace Murray Hopper Award, and was selected as one of the 10 researchers to work on the Aurora Exascale Supercomputer. Her visionary work in simulating blood flow through the human body in a system called Harvey, led her to be featured in the MIT tech reviews, Innovators Under 35 list. Amanda, welcome to ACM ByteCast.

Amanda Randles: Thank you very much.

Rashmi Mohan: Amanda, I'd love to start with the question that I ask all my guests. And it's always fascinating to hear the journey that people undertake as they get into the field of interest that they have. If you could please introduce yourself and talk about what you currently do and also give us some insight into what brought you into this field of work?

Amanda Randles: Sure. So I'm in biomedical engineering at Duke University now, and our research program is really focused on building large scale, personalized blood flow simulations. So I'm really passionate about using the biggest supercomputers in the world and figuring out how we can leverage them to answer questions we wouldn't otherwise be able to answer. And I've probably been working in this space about maybe 10 or 15 years at this point. And I would say what really started, even just getting involved in computing in general, I grew up in Michigan, I went to a math and science center at high school where half of the day people from around the district got together and we would take our science classes.

And we took computer programming in high school, and I had just a phenomenal computer programming teacher back as a freshman in high school, that really got me excited about programming. And through that program, I was exposed to really strong science, the connection of how you can do computational modeling, how you can use the programming side. I was part of the first robotics team in high school, and really just, all of those experiences got me excited about, how do we use technology and what can we do with it? When I got to Duke, I was really interested in physics and taking a physics-based approach.

So I started actually majoring, at the time I started with political science and was thinking about, it was time when they were doing all of the genome and trying to think about international law associated with genetics. So I was really interested in the mix of policy with science. But I started out in political science and physics, and I realized how much I missed coding. And I really enjoyed the nights, we'd stay up all night, coding problems, and I really liked the hands-on coding side of things. As a freshman in college, some of my friends were taking coding classes and I kept trying to help them with their homework and just kept realizing how much I actually really missed it.

So I ended up double majoring in physics and computer science. And I think everything I've done throughout my career has really been at this kind of intersection between different disciplines of how do we take the physics-based model? How do we couple that with the computer science side? And I've always really been interested in how we apply that to biomedical questions. So really taking this physics-based approach and answering different bio-based questions. So in undergrad, I worked in a few different genetics labs and I was really involved with some bioinformatics projects, and I would spend some time at the bench sometime doing research with the computational side.

And kind of slowly realized I was much more excited about the days I was spending on the computational side and tried to spend less and less time at the bench. I'm not as strong on the experimental piece. And I guess just through that, I started getting more and more involved on the computer science. And when I graduated from Duke, I was offered a position at IBM to go work on their Blue Gene supercomputer. And at the time I hadn't really had any exposure to high performance computing. I didn't really know what a parallel computer was. I didn't really have any experience at that stage.

But when I joined IBM, I was really lucky to get to work on one of the... This is right at the height of the Blue Gene L transition to Blue Gene P. Whereas they were creating the biggest supercomputer in the world, and it was really exciting to get to play with that kind of technology and get to work with that. And one of the jobs I had was trying to get applications to run faster and port different biological applications to the supercomputer and try to identify how do we optimize them and make them more efficient on these systems.

So I got to learn a lot about computational modeling and how you tune a code for these large scale systems and really just the power of these large systems and how you could run simulations, so much of a larger scale than you otherwise could have. And through my time at IBM, I started to get very excited about the applications and what we could do on the biomedical side, which led me to go back to graduate school, because I wanted to shift away from being the one building the supercomputer, to being the one who was really using the supercomputer and really setting up those questions and driving the questions and the research agenda.

Which is really what motivated me to go back to graduate school, where I did get the PhD in applied physics, really focused on the physics-based model and how we develop these physics-based models. But the entire time kind of addressing biomedical questions along the way. So it's always been an interest in this marriage of the computing side with the physics and the biological application.

Rashmi Mohan: That's a phenomenal journey. I mean, and also it sounds like you identified that intersection of those three areas early on in your career. And what I also found fascinating was that you started in industry and you actually chose to go back to academia post that once you discovered that you wanted to be the person using the supercomputer rather than building it. But I'm sure the exposure that you had in building the computer itself gave you some ideas of its capabilities and its limitations. Did that play into your decisions at all? And how has that assisted you in your journey so far?

Amanda Randles: Yeah, I think I was incredibly lucky to be a part of the Blue Gene team. Just in general, it was just such a unique time period in parallel computing, because that was right when the US was taking over as number one in super computing and they had just created of this Blue Gene system that was being put at Lawrence Livermore National Lab. I remember having to go in early in the morning, just had to have the opportunity to boot 64,000 processors and how exciting that was before people in California got up to be able to just turn on the system and do that kind of thing. It was a really great atmosphere where we had phenomenal people who were working on all aspects.

We had, the control systems team was amazing, the NPI team was amazing. We had really strong teams, but it also was a system that was kind of... We were working on the system, developing the system. So if a code wasn't scaling or a code didn't work on the system, you didn't know if it was because you had potted it incorrectly, or if it was because we hadn't finished, there was something wrong with the hardware because it was still being developed or there was something wrong with the underlying software stack. There are so many places where that could be going wrong. And it was because everything was just in flux. So it really taught me a lot about how to debug programs and how to identify problems and how to work with people from both the hardware and the software stack and try to find, how do you really sort out where the problem is and how to build a piece of software that's actually going to scale well, and all of that.

And that played in, it was a really good experience from that end of like what do you need to think about and how... Don't just think about developing a software or a code base. I think a lot of people, you don't take into account the hardware piece and how much memory is there? How do we make it fit on the system? How do we optimize it for the system? So when I first started developing Harvey, I came at that from really building it to be tuned to scale and be efficient from a parallel computing standpoint, with all of that in mind. And I

think that's really paid off and been very helpful. And then having the industry experience before going to graduate school was also... I knew why I was going to graduate school.

I knew what I was interested in. I was very direct. I think it made me much more efficient when I was in graduate school and it helped me come in and hit the ground running, of why was I there? I had a full salary, I had a house, I had a car, everything as someone working in software industry and leaving a nice salary from IBM to go back to graduate school, there's a reason you're choosing to do that and you know why you are there. And I think it gave me a lot more purpose of, I knew what I was trying to accomplish and why I was there. And I think that was very helpful.

Rashmi Mohan: Yeah. I think that's an excellent point that you bring up because I've often heard from friends, that the journey of graduate school, especially, and when you're going into research, a large part of your time is really trying to understand, what is it that you're trying to solve. Right. Even to get to that core question that you want to answer is a large part of that journey. And it sounds like you had a little bit of a head start because you were very keen on... You knew what that transition was like for you and why you were going back.

Amanda Randles: Right.

Rashmi Mohan: But speaking of Harvey, I mean, it's hard read about your phenomenal work and not mention Harvey. Could you tell us a little bit more about it? How did it come about? Where did that idea even germinate for you?

Amanda Randles: So while I was in graduate school, I had joined Tim Kaxiras lab at Harvard. And his research is really, he's a very strong multi-scale physicist and he uses those kind of techniques in a wide variety of areas. And I think, again, I've been very fortunate in the exposure that I've had, where he had a project going, where he had just started a blood flow project probably the year before I joined the lab. He had strong collaborators that were the founders of the methods we were working with, just absolutely amazing people that were spending time in the lab while we were there.

So they were trying to get this blood flow project kind of up and going, and there was room for a graduate student to join it. So I joined the work on a code called [Murphy 00:11:08]. And through that, I had a lot of exposure and I really learned about fluid dynamics because I had never taken a fluid dynamics class. And I learned about the problem of what do we need to know about? How do we build? What are the questions we can answer with the fluid dynamics code base, and why do we need this? And because we were in Boston, we had strong clinical collaborators at the hospitals nearby, to really understand what is the potential power of personalized fluid dynamics.

And in that, we kind of hit the end of the road of what we could do with the original code and kind of hit the limit of, how do we scale the code? How do we run larger simulations? And so I went back through and started writing Harvey from scratch as a graduate student, based on some of the other work that we had been doing and used some of the same fluid dynamics algorithms, and based on the same methods, but really tried to write Harvey from scratch, with the plan to parallelize it and try to make it scale as well as it could on these large supercomputers, which is part of... William Harvey is the person who discovered that blood circulates through the body, so we named it after Harvey and it was while I was at Harvard.

So that all made sense. Harvey kind of came out that. And in the beginning, Harvey was just... It's a code where you would take in the 3D geometry that you would get from medical imaging. So from a CT scan or from an MRI scan, we use commercial software to segment that medical imaging data. So by segmenting it, you're pulling out a 3D geometry. So you're getting that 3D geometry representing your coronary arteries. Harvey takes that as your input, so you have 3D geometry of the vessels and runs a fluid simulation to identify the velocity, the pressure, different quantities, like shear stress. And those are quantities that are associated with the development and the progression of heart disease.

So this gives you an opportunity to non-invasively, so you're taking a CT image of the patient, you pull out that 3D geometry without ever having to do anything invasive, and then we can get at quantities like the velocity of their flow, the pressure, the shear stress, and get diagnostic metrics or predictive metrics of what could happen to that patient. And in the beginning, we were looking at, in Harvey, was the bulk fluid parameters. We've built that up from there now we've added fluid structure interactions. So we have millions of red blood cells that can be modeled at the same time. We're looking at how cancer cells move through the body, looking at the adhesion of cancer cells to the wall. So really built the complexity of this code up over time, and really try to scale that as we go along.

Rashmi Mohan: I mean, there's so much to unpack there, Amanda. But I'll start with a very naive or simple question. So if I have Harvey and I actually have CT scan information and MRI information, as a doctor, could I just plug that into Harvey and get results where I can actually determine what my treatment should be for the patient in front of me? Does it work at that level of... I don't know, I mean, at that scale?

Amanda Randles: Yeah. So we're trying to make it as user-friendly as possible. And I guess I should say one positive thing for listeners here that are more on the software engineering side is, it has been very important, and I guess you've said, is, one of the takeaways from working at IBM was bringing strong software engineering practices to the development of Harvey, where we are using things like version control, we're using unit tests, we have continuous integration. So just one

comment of like, by having strong software engineering practices in place, that has been critical to making Harvey be able to grow and be successful and even be usable in any way in the clinic and make sure if someone's working on the cancer project, they don't accidentally break something from the heart project or something like that.

But with that, we do have a series of ways to interact with the data. From a clinician's standpoint, there's a wide range of projects, some that are much more translatable into the clinic and some that are much more basic research. The most straightforward way for clinicians to interact at this stage, there are questions like, should they, or should they not place a stent in the coronary artery of a patient? And current the gold standard for that is fractional flow reserve. So it's a quantity that's really just a pressure gradient across and narrowing in the vessel. And if that pressure gradient is above or below 0.8, that will determine if the doctor should or should not place a stent.

There are currently FDA-approved software packages available, where you can take the CT scan of the patient, you run a fluid simulation, and you use that simulation to calculate something like the fractional flow reserve. And you now have a metric that you've gotten noninvasively, of the patient, and you can determine, should you, or should you not stent the patient based purely on this fluid dynamic simulation. So questions like that are ways where the doctors can interact with these types of simulations much more easily. To run a simulation directly in Harvey, you do need to do that first step of... The critical piece is creating that geometry that's going into Harvey.

So we use commercial software to segment the geometry from a CT scan. We have other software packages and collaborators we work with to create 3D reconstructions from angiography. But there is that manual step of, how do we extract the 3D geometry from the medical images? But once you have that, that would be the input into Harvey. And we did have a really great student that worked with us the year between his undergrad and... He entered an MD-PhD program. And in between, he created a tool that is a user interface for Harvey that allows you to take that geometry. And you can actually modify the geometry in any way that you would want. So if you want to do some surgical treatment planning, so identify, if I placed a stent in this vessel, what would happen?

So you change to the geometry by placing a stent or removing the narrowing in the vessel. And then it'll re-mesh that geometry and then automatically launch it into Harvey. So the user doesn't necessarily have to use the command line programming, know how to use a large scale supercomputer, and that would all happen in the background. And then when the simulation is completed, it will retrieve the results and allow you to visualize the flow field. So there are easier ways for users or clinicians to interact with the data. And there are fun features where that can... He did build this so it could work in virtual reality. So you can

actually immerse yourself in the blood vessels and do this using an Oculus Rift or a zSpace or any other kind of semi-immersive or immersive experiences as well.

Rashmi Mohan: Wow. That's amazing. One of the things that you mentioned, and so I'm just trying to understand this a little bit better. Right?

Amanda Randles: Right.

Rashmi Mohan: If I were... And I know I'm know that there are two parts to it now I'd probably ask both my questions. As a clinician, if I were to use the system, even in the state that as I understand that there's much more to it, that you have to work on in terms of actually developing the geometry, what kind of compute power do I need in order to be able to run this?

Amanda Randles: Right. So it really varies based on what question you want to ask. We're in an exciting time period. We're working very closely with different cloud providers like Microsoft and Amazon to try to see, how do we actually port Harvey into the cloud and make this more usable? Because that's the direction we see this going in the future. The way this would actually work in a clinic is how is using cloud-based resources. We're not expecting every hospital to own their own supercomputer. It's how do we make this tractable? So a lot of our work is, at the basic science side we use the large-scale super computers, but we also want to push it and see, if we put absolutely everything into the model on these big... When we're using the production runs on the big super computers, what can we take out? And how do we minimize that? And what is required?

How patient specific does the model have to be? Do you actually have to model the red blood cells? For the question you're asking, what needs to be included? And then we want to minimize that so that it can be more tractable to run these simulations. Depending on the question, it can be something that you could easily run in the cloud, on even a few GPUs and you could run it for under an hour and be able to calculate something like a fractional flow reserve. We were the first team to run with the full arterial network in the body. And that was from the head to the toe, everything one millimeter and above.

That took the entire 1.6 million processor supercomputer at Lawrence Livermore Lab, if you wanted to run that at cellular resolution. So it's kind of that spectrum of, are you trying to run an extremely large region? Are you looking at... Do you actually need all of the red blood cells? What we hope to put in practice in the clinic would be more, identifying when can you use bulk fluid simulations and likely moving things into the cloud and running on that and making it much more tractable and affordable for clinicians and hospitals to actually use.

Rashmi Mohan: Got it. Thank you for that clarification. My second question was, so obviously there is also, I mean, I was obviously talking about at the clinician's environment, but there is a large part of your work which is really dealing with

getting the large volume of maybe medical imaging data that is available out there to do more research and understand like you were talking about, understand some of these medical problems in more depth and possibly use that research to drive better solutions, right. Whether that's on the medical side or on the computing side. What are some of the biggest challenges you're seeing with that side of the world, both interesting computing problems that you're trying to solve or scale challenges that you're running into? I'd love to hear more about that.

Amanda Randles:

Yeah. I think the area of personalized simulation, it's incredibly exciting right now. There are so many opportunities for people to jump in and help in so many different fields, which is, it's really ripe for other people to jump in, which is great. We have many challenges and we need a lot of assistance. So there are questions, we're at a great time point where we're seeing the explosion of machine learning, we're seeing the explosion of AI. We've seen a lot of fantastic work, really applying machine learning to questions like segmentation of getting out that geometry from the medical imaging. All that's becoming very mature at this stage. We're really at the beginning of seeing the connection of machine learning with physics-based models.

George Karniadakis' work at brown is, he's doing a lot of physics-inspired neural networks. And we're starting to see that connection of the physics laws with machine learning to see what we can predict. But with that, you need to have large scale populations, you need to have a lot of medical imaging data, we need to have a lot of simulations completed. If we want to bring in anything from the fluid simulation standpoint, we have one fluid simulation, when we talk of the full body model that took the entire, one of the world's largest supercomputers, and we took the entire supercomputer to be able to do that, it would be difficult to run a huge parameter study and create 500 different patients, or run the entire simulation.

So it's trying to figure out, how do we balance that? How do we actually make so much data? And then some of the issues that we're running into on a daily basis, even with one simulation, the work that we're doing with Aurora, the upcoming Exascale scale system, we want to be able to... We're running the simulation so we can analyze the data, visualize the data, and actually understand what we're doing in the simulation. Each supercomputer has several petabytes of memory across the entire supercomputer. Our simulations are using that entire supercomputer for every single timestep, and we are running millions and millions of timesteps.

So you can say that we are creating, every single timestep, we're creating five petabytes of data. It is not tractable to just download all of that data for the entire simulation and do post-talk analysis. It's really becoming necessary to have in situ visualization techniques, in situ analytic techniques, even in situ machine learning training options. We really need to be able to analyze and work with the data while it's still on the system. The crux of the project we have

going with Aurora is trying to set up and investigate some of these in situ options, because we're really hitting the point where it's just not tractable to download that much data and do any kind of post-talk analysis.

That's a huge part of it. We had one project recently where it was, we ran something for maybe four or five timesteps, just a very small subset. And we still had about 400 to 500 gigabytes of data that we were trying to just visualize to see, where did the cancer cell, where was it placed? How did this get set up? And just trying to... How do we actually just get one image and snapshot to see what's happen with 400 to 500 gigabytes of data? So I think the amount of data we're creating is exciting on the front of... We're creating a lot of data that can feed into these new methods, like the machine learning side, but it also brings a lot of challenges for how do we actually deal with it and what do we do with all this data? How do we access it? How do we actually get it off the supercomputer in a way that is usable and tractable?

Aside from actually trying to build large scale population databases with whatever disease area you're trying to look into. We've been very focused on trying to augment patient data with also synthetic databases. Meaning taking one coronary geometry and modifying the radius of the vessels in there slightly and creating 15 different versions of that patient with slight modifications to the geometry, and having a large scale synthetic database that really allows us to pinpoint the effect of changing the radius or changing the angle between two vessels. And what does that do to the resulting fluid dynamics?

So we can do all of that. We require large scale computing and we need to be able to create large scale synthetic databases that are inspired by the patients, but allow us to really identify and pinpoint different factors. We've also been very interested in how do you combine questions like design of experiments, with machine learning and fluid dynamics results, where when we're looking from the clinician's perspective, they want to have a full view of what's happening with the patients. You can imagine, it's not just how the patient is sitting there at rest and what their pressure gradients going to be. It's are they going to be at risk if they start running? Are they going to be at risk if they start running at high altitudes?

What if they go to Denver and they're running? And what if they're pregnant? What if they're... There are so many different factors that could play in and could change someone's hemodynamics, that you can't just measure in the doctor's office. And that's the power of having these personalized flow simulations, that we can actually simulate each of these cases, but again, it's not really tractable. There are an unlimited number of different examples of someone running in cold in the winter. There are many different scenarios you can imagine. And we can't do an individual simulation for each of those situations for each patient that comes in.

So we had one large project trying to look at the design of experiment approach of, can we identify the minimal number of simulations we would need to run to be able to couple that with machine learning, to then predict what would happen in all of these other use cases? So trying to find intelligent ways to make the most use of our super compute hours and get the most out them and really couple it with these new metrics and these new algorithms to try to make that go as far as possible.

Rashmi Mohan: As you were talking about machine learning and the amazing set of challenges that this field has to offer, but the machine learning piece was especially interesting to me only because, we talk so much about using machine learning carefully and biases introduced into the data that could impact what your outcomes could be. What do you see as things that you have to be particularly careful about when using these large volumes of data and making decisions on the basis of that? And also what does validation and testing look like in this field? Because I'm thinking, if you think about bias, I mean, you get an advertisement wrong on the internet, you can probably bear that loss, right. But the field that you're working in, there's very little room for error. So how does one mitigate risk?

Amanda Randles: That's a really great question. It's definitely been an issue that's really plaguing this area. Because you see, there have been countless papers already where it's bias, you didn't really appreciate that was in the data, makes transfer of that method to other data sets or other hospitals, they're not able to replicate the same experience or the same results. And oftentimes comes from unidentified biases and things like that. We need to really... I think that's where the focus on these... The shift to interpretable machine learning methods. We don't want to view this as a black box. We need to understand why it's making the decisions. Why are the recommendations being made? What's happening under the hood? And what's going on there? It's incredibly important.

From our end, with the validation piece, we've built up... It's been very important to understand what are the critical components and what could be... Where do you have to validate? Aside from doing any kind of machine learning piece, just, how do you actually validate these flow models? Because anytime you see a physics-based simulation and you should be questioning, "Why should I believe this? And is it accurate?" And we started working with a team at Arizona State where they would take that 3D mesh file that we got from the segmentation, and they would 3D-print the mesh file.

So we'd have a phantom or a 3D printout that represented someone's aorta or their coronary arteries. And then they would run a controlled flow experiment within that aorta, and within that phantom. And the nice part there is they knew the viscosity of the fluid, the inflow, everything is a very controlled setup. And they're able to actually measure using particle image velocimetry, the entire flow field, not just one measurement in one time point of what was happening with the fluid within the phantom. And we were able to compare the

measured velocity flow fields with what we're getting out of the simulation to make sure that we were getting the right results and we're actually capturing vortices that may appear in really complex flow profiles and make sure that we are getting that correctly in our models.

And for us, we really, we started that, that was one of the first major projects we tried to do once Harvey was up and running because the first question was is, why should you actually believe us? And we started with the aorta because it has such a high Reynolds number, which means it's closer to turbulent flow. The flow in the aorta, the aorta's one of the largest vessels and it has very high velocity flow. So it's very chaotic. And the thinking was, "If we can get the aorta correct, then everything else will be much easier." So we started with the aorta and tried to compare that. But we did have to think through, every time we start looking at a different vessel, we start looking at different disease areas. We start over and we do another 3D Phantom, we print it out.

We have coronary phantoms, we have femoral artery phantoms. We've now started working with a team at Lawrence Livermore Lab where they're bioprinting microvasculatures and doing the same in that space to make sure that we're actually getting the right flow profiles in the geometries and the length scales and the flow regimes that are relevant to each of these different disease areas. And then beyond the preclinical work, we need to compare against results in the body and see, what are we missing? There are many assumptions that go into the phantom. So having direct comparison with what we see in the body is incredibly important.

We've just finished up a large clinical study, where we had 200 patients at two different sites, where we looked at this quantity that I mentioned at the fractional flow reserve, where they actually put a wire in the patient, they invasively measured the fractional flow reserve over the stenosis, and we're comparing the measurement that they got from the guide-wire to what we have in our simulation. So we're able to really ensure that we're getting the right results and matching what is happening in the body. And I think finding ways to validate the model in that setting, we have, from the cell models we work with in-vitro experiments.

We do a lot of work trying to replicate optical tweezers, or how you're pulling red blood cells and make sure we get the right mechanical behavior, how they're moving through constrictions, if we get the right passage time. And we do a lot of work replicating known experiments in our simulations to ensure that we're modeling everything accurately. And it's very important to take that into account and make sure you have the right validations in place for whatever research question you're trying to ask.

Rashmi Mohan: Yeah. No, thank you. That was very insightful. I was reading a little bit about the idea of digital twins used in medicine and how that is helping the world of personalized medicine. Right?

Amanda Randles: Right.

Rashmi Mohan: And then I think the point that you brought up about actually the 3D printing, the parts that you're actually working with to ensure that you're able to validate it. That makes perfect sense that you actually want to... You have a digital twin, you use that to test out your theories, but then you also validate it using this the phantom organ or piece that you're testing out. Does this mean that this whole area is now becoming more affordable, Amanda, for us to... I mean, is this something that will scale from a cost perspective as well?

Amanda Randles: I think so. It's definitely becoming more affordable and I mean, we're also piggybacking on the side of, we are seeing an explosion of HPC technology in the cloud. It's getting easier and cheaper to use these options in the cloud. Some of our code, we have looked at options... We are running on GPUs, we're running on CPUs. We've also started looking at running on arm processors, which are often much cheaper to run on in the cloud. There's that trade off of depending what question you're asking, your time to solution versus the cost side of it. And we have a lot more flexibility to see, can we use slightly slower processors that might be cheaper and might only take an extra half an hour? Do you need it to run in that half an hour, or is it okay if it takes 90 minutes?

And depending on the question, we have some flexibility on that end. But as computing gets cheaper and things like the cloud and processors are getting faster, we're kind of run that whole wave still and we're able to make this a bit more accessible to other researchers as well. And all of that has been a really good progress for this kind of work.

Rashmi Mohan: Got it. Yeah. One unique part of your career that I observed was how embedded you are with various external lab or consortiums or industry partners. Why is this so important to you?

Amanda Randles: Right. Yeah. Our work is very interdisciplinary. I think it's very unique in our lab, at least we're... You might have a call with a hardware vendor one week and then be in the imaging lab, or the cath lab at the hospital the next week. We have to understand the full spectrum and work very closely with the right people. And we are trying to be on the cutting edge. We're trying to use the brand new supercomputers with brand new features that we don't really know. They're making modifications to the hardware all the time, how do we best exploit that? So we work very closely with the hardware vendors to understand, how do we write our code to run as efficiently as possible on your system?

So we work very closely with the Intels, with Microsoft, with IBM on that side of it. And the students in the lab really, they can talk across that whole spectrum, which I think is pretty unique and pretty great from their end. And then every single project we work on has a clinician or a researcher. It has a domain expert or the user invested in it. So I think a lot of times it's easy for the physicists or the computer scientists to come up with questions that we think are interesting

and are exciting. But then when you go and tell the clinician about it, they're like, "Yeah, but that would never work in practice. We would never use this."

We have regular interaction with the clinicians to make sure the questions that we're asking are actually useful to them and that the tools we're developing for them will be something that could work in the clinic. When we're looking at this virtual reality tool, we've conducted a lot of user studies in the cath lab to try to understand, how would you actually use this? How do we make this usable for the clinicians? What are the best ways to do that and make it not just some kind of kitschy tool we have? Something that could actually be usable in the clinic.

My postdoc was at Lawrence Livermore National Lab, and we still work very closely with the teams at Lawrence Livermore because they're fantastic in the... They're in the middle where they're physicists and PhDs and physics that are really, they're very used to working with the large scale computers. So from the application standpoint, they're some of the best people in the world for knowing, how do you get a physics-based model to scale on these big supercomputers? So we're working with the applications experts, the hardware vendors, and all the way through to the domain user and trying to make sure we're touching on each of those pieces and pushing that along. And being able to communicate across those boundaries has been very important.

Rashmi Mohan: Yeah. It sounds like the absolute right thing to do, right? I mean, definitely having the air of your customer, if you will, because they're the ones who are actually going to be using the products that you put out, is so critical. And it's also, I think, a great way for you to get feedback and continuously improve on solutions that you were building.

Amanda Randles: Yeah, it's very important because there are just basic things of... I'm trying to think of an example, yeah, just basic ideas of, we're trying to identify a tool that we think is helpful, or we say, if you could get this bifurcation angle, or you could put in a stent in this location. Having the doctor come back and be like, "That's great, but stents only come in these four sizes." These are the restrictions that you may not have thought about. Or, "There is actually a vessel in the way, we wouldn't be able to do that." Having them kind of keeping us honest and making sure points that we may not realize that come into practice when you're actually in the operating room or the time that you have to spend.

There are just questions in part that as a computer scientist or as a physicist, we don't really realize about the real day-to-day of how they can use these models. So having them involved has been, I think, incredibly useful. There are definitely questions that have come up that we find interesting and then the doctors help tune that and say, "Well, actually, this would be incredibly helpful for us." And we come to the middle and try to make sure we're asking questions that are actually useful.

Rashmi Mohan: Yeah. Makes a lot of sense. I also heard about your recent work during the pandemic around ventilator use, right, and helping physicians determine the optimal sharing of ventilators. So a couple of questions I had around that is, one, if you could tell us a little bit more about what was the work that you did there? But also, how easy is it for you to pivot into addressing an imminent challenge. This is not something maybe that you would've anticipated prior to the pandemic. So how easy is it for you to use the work that you've spent years building expertise on to address a problem that's right in front of you, that might be slightly, I don't know, maybe applied in a different way?

Amanda Randles: Right. Yeah. So the ventilator work was actually, it was very interesting in that sense of... Yeah. I guess just for some background, I'm sure everyone's pretty aware. Early on in the pandemic, one of the first issues was the ventilator shortage. And we were seeing that in the news every day. And the question really came out, that we had a significant ventilator shortage in the US and worldwide, right. And one of the teams at Duke on the medical side, they jumped in immediately and were trying to find ways to increase ventilator capacity. So their hypothesis was, you could take a ventilator and if we could split it between two patients or even more, we immediately have a way of at least doubling our ventilator capacity and making the ventilators that we currently have go further.

So they developed a device that was tuneable, it was a 3D printed device that would allow you to split a ventilator between two patients. And the interesting piece was it was any two patients. So they didn't have to be matched in terms of weight or lung compliance or any other feature of the patient, which also was what had been going on the state of the art. So they're able to expand it so that any two patients that came up, you could tune this tool and use it to split the ventilators. They came to us because we had worked with them in the past on fluid dynamics projects related to the heart and blood flow simulations.

And they were asking if we could develop a way to tell the doctors, how do you actually tune it when you have two different patients? We have this new information, how do I take two brand new patients and identify the optimal setting for this ventilator splitter? And it was an interesting project of... I think we were probably, they came to us within the first week of when we had locked down in North Carolina and everyone had stopped coming into work. I think we were all sitting there hoping... You really wanted to contribute and wanted to find some way of helping, but I couldn't see at the time, how would computational fluid dynamics help the pandemic?

You just wished you could be helpful. And so I think everyone in my lab was very excited when this opportunity came up and they jumped right in. So we had a large team that really started meeting every day, and we had a phone call each day to go over, how are things going? What hurdles did you run into? How do we help you? But we actually started out having two different teams working on projects. So we had one that was trying to pivot immediately with Harvey and

see, do we need a 3D model? How do we use a 3D model? And I won't go into the technical details, but just from the high level, this is very complicated because there's a fundamental difference of, we're modeling a fluid, a liquid and looking at the air flow is actually very different how you're going to do the computational modeling of that.

So the first question's, can we even pivot Harvey into an airflow model? And then the other team was set up of, ignore Harvey, could we start from scratch and create a very simple one-dimensional, lumped parameter model to represent the ventilator and the lung capacity and everything? And in the meantime, the team from the medical school was creating a benchtop model and getting the data that we would need to validate against. They were using mechanical lungs, completing the experiments and getting that data. So we had full team meetings every day to check in if, okay, if we had a model working, what data would we need to validate and make sure we had all that running in parallel?

And we were very fortunate, the low-hanging fruit of having the new 1D model worked incredibly well to match the data. So we were able to shift focus to that. But it was interesting in that, it was completely brand new model, we had never worked on before and people were able to draw from their experience and knowledge of the fluid dynamics to build this model, but it was actually a brand new model to us and something new within two weeks, pulled together, validated, and had ready to go. And then the other piece of that project, that was just, and really worth noting and very interesting was, you can imagine there are many different patients that, if you have many different patients, someone could be a hundred pounds, someone could be 120 pounds.

There are many different patient parameters that could come in, that kind of explodes very quickly of the different simulations you might want to run. And to submit this for FDA emergency-use approval and have everything available for the doctor, so they weren't waiting on any simulations. We wanted to have everything simulated before ahead of time. So we had the results for every possible combination available, and that meant running millions of simulations. So we actually submitted a proposal to the HPC consortium for COVID asking for compute time. And we put the pie in the sky idea of like, "It would be amazing if we could run all of these simulations."

So we asked for a million compute hours. We're like, "This would be great." And we assumed when we submitted it, they were going to get some subset of that and have to cut it back and run only a subset of the simulations. We submitted the proposal on a Monday and by Thursday, we were paired with Microsoft to work on the Azure cloud setting. And by the following Monday, we had actually completed 800,000 compute hours and run every simulation that we wanted to do and were able to go through the entire parameter sweep and had that completed within a week of submitting the proposal itself.

So I think just in general, it was a really great example of bringing an interdisciplinary team together to really accomplish something in a time constraint where we really needed to get it forward. Within three weeks we went from never having worked with this model before to developing the model, validating the model and running the entire simulation. So everything was kind of ready. There was a company in the area that donated their time to create a mobile phone app so that the doctors could access the information and work with the data on that end. And it all just really came together very quickly.

And I think no one slept the weekend that they were running all the simulations, but it was great that everyone was willing to put in that kind of time and effort to really make this happen.

Rashmi Mohan: It was so heartening to hear that kind of partnership and the amazing things that comes out from that kind of commitment. And getting these, is like you mentioned, the diverse teams together to solve a really crucial problem. So thank you for sharing that. This has been an absolutely captivating conversation for me, Amanda. For our final bite, I'd love to know, what are you most excited about in the field of biomedical engineering or the work that you're doing maybe over the next five years?

Amanda Randles: Yeah, I'm really excited to see, I think at this stage, we've validated the models in a lot of different spaces. We've started collecting a lot of data and we're kind of ripe to both have this marriage that we've talked about with the machine learning, of how do we move this further? But then how do we move into spaces we weren't able to go into before? A lot of the work in our lab is focusing on looking at cancer metastasis and trying to understand the interaction with circulating tumor cells when they come off the primary tumor, they're moving and might make a metastatic site.

And we're only now able to complete simulations that are able to track a cancer cell on the order of centimeters or longer distances that are really necessary to be able to understand where a cancer cell may be going. And in the similar vein, for the cardiology projects, we're now able to run simulations where we can look at not just a heartbeat, but trying to understand, can we predict longer time periods? Can we look at the grand challenge of what happens over the course of even a day? But it would be amazing if you can go to weeks or months.

And I think coupling with these ideas of physics-inspired neural networks and other ways of tying into the machine learning, we're able to look at much larger spatial domains and much larger temporal domains that are really allowing us to ask questions we couldn't ask before. So I'm excited to see, as we shift to using systems like Aurora and we have so much power that we haven't been able to have before and tie that in with these new algorithmic changes, I think we're able to see, what is it about a certain cancer cell that causes it to move in

different directions? Can we really understand those underlying mechanics driving metastasis, driving disease development?

And how do we really use that to create new treatment options, new potential to really address, how do we help address different disease areas and improve outcomes for these patients? And I think we're really moving beyond the validation stage and getting into, how do we start expanding the use of personalized flow simulations for different disease areas outside of just coronary heart disease and conventional cardiovascular work.

Rashmi Mohan: Got it. Yeah. I'm super excited to continue to follow your work and learn more about what you do. Thank you for taking the time to speak with us at ACM ByteCast.

Amanda Randles: Thank you very much for having me.

Rashmi Mohan: ACM ByteCast is a production of the Association for Computing Machinery's Practitioners Board. To learn more about ACM and its activities, visit acm.org. For more information about this and other episodes, please visit our website at learning.acm.org/bytecast, that's, learning dot ACM dot org slash B-Y-T-E-C-A-S-T.