

TACC and UCSD: Tackling COVID-19 with HPC

Transcript from June 25, 2020
Customer Spotlight Webinar

Mary Killelea: Hi there. Welcome everyone. Thank you for joining us for Intel's customer spotlight series. This series highlights innovative, industry-leading companies that are undergoing digital transformation, have tackled business and technology challenges, and created new opportunities using Intel data centric technologies and platforms. Today, we are excited to feature the groundbreaking research that is happening at TACC, with the National Science Foundation funded Frontera supercomputer.

Today's host is Tim Crawford. Tim is a strategic CIO advisor that works with enterprise organizations. Tim, I'm going to turn it over to you now to kick off today's conversation.

Tim Crawford: Sounds good. Thanks, Mary, for the introduction, and welcome to everyone joining us online.

We've got about 25 minutes of content to get through, and then at the end, we'll shift gears into a 15-minute Q&A session.

I'm joined today by Dr. Dan Stanzione, and Dr. Rommie Amaro. Dan, Rommie, thank you so much for taking part in the conversation today.

Tim Crawford: So, let's start off by setting a foundation for those folks that are attending and listening to this, maybe we could start by talking about the Texas Advanced Computing Center, talking about TACC.

And then we'll shift, Rommie, to you and talk a little bit about UC San Diego and the work that you're doing as part of your group and your lab. So, Dan, do you want to kick us off there?

Dr. Dan Stanzione: Sure. So, I'm the Director at the Texas Advanced Computing Center, and we are part of the University of Texas at Austin. But we are funded by the National Science Foundation to provide advanced computing resources, large-scale supercomputing and data resources, and people to help use them to advance science in society to users all around the country and around the world. So, we're really sort of a national resource for this.

And the machine we're talking about today, Frontera, we've built with Intel and Dell, and it's one of the largest supercomputers in the world. And of course, right now, we're using a lot of it to support several tens of projects in response to the COVID-19 work. And the largest of those is the work we're doing with Rommie's lab. So, I can turn it over to Rommie here.

Tim Crawford: Great. Rommie?

Dr. Rommie Amaro: Great. Yes. So, I am a Professor of Chemistry and Biochemistry at UC San Diego. And basically, what my lab is doing is developing these different atomic models.

So, we're modeling the virus and other systems, biological systems mainly at the atomic level, and we're using resources like TACC to run what we call all-atom molecular dynamics simulations, which are really just the computational and numerical representation of these biological systems, and how they move in time.

Tim Crawford: Wow, that's a mouthful, and I'm looking forward to digging into that. We've had a chance to talk a little bit about the work that you're doing, and I'm really excited to share that with our audience.

Maybe before we delve into that, you can talk a little bit about the relationship between TACC and UCSD. This is not the first project that you've jointly worked on. If you could touch on that for just a minute.

Dr. Rommie Amaro: You bet. I'm not sure if Dan wants to take a stab at that, but I can say, we've been using—we and many other labs, have been using TACC to investigate various types of scientific questions, for many years, I guess for over 10 years now.

And so, we have worked together I think quite closely on projects, for example in cancer, in environments of chemistry is another area where we have a lot of activity, and also with virus research. Before SARS-CoV-2, we worked quite extensively on influenza. I'm not sure if Dan wants to add anything.

Dr. Dan Stanzione: Yes. Let me just say that I think there's the sort of public perception that when something like COVID-19 happens, all of us were doing other things, and sort of leap into action. And these things happen instantly, and we can create dramatic results.

And we can create dramatic results, but it's the results of a long period of investments in building up relationships. You know, as I mentioned, Rommie's lab has been the largest user of our systems, and was one of the first in response to COVID. And the reason for that is, we have worked together a ton before, and these relationships help a lot.

She knows how, and everyone in her lab knows how, to make great use of the systems that we provide and how they work and what the software environment is like. We know how to support the codes, that she uses and runs at scale, and we've worked together before.

And so when this started, you know, we had a quick conversation and set this up and got going, but, you know, if we'd never worked together before, the systems weren't already in place, if the people in Rommie's lab hadn't been working for years and years to get ready for this kind of an event, we could never do this kind of response quickly, and then the way that sort of society demands that it's quickly.

So, that long term relationship and previous collaboration, is a huge part of getting this work done.

Tim Crawford: That's great and appreciate the foundation there. You know, a lot of folks will talk about high performance computing and research, and this is actually becoming more than just a lab issue. This is something that even enterprises are starting to move into and solving these big problems.

But when we talk about solving these big problems, we talk about solving them at a clinical or genome level. What I want to do is maybe start to shift to Rommie a bit and talk about the importance about understanding these types of problems in an atomic level.

So maybe you could talk about the work you're doing, and why the understanding at the atomic level is so critical.

Dr. Rommie Amaro: You bet. So, I should start by saying, you know, you mentioned that you weren't used to hearing about the clinical level studies and genomic level studies. The atomic level really ties very closely to both of those, you know, but it's just a different aspect of the same problem.

So, what we are doing, as I mentioned is, we're making these—sort of very highly detailed models, like you were saying, at the atomic level. So, we're representing the virus, for example SARS-CoV-2 virus. We are starting with all different types of experimental data, or a number of different types of experimental data, that tell us what the virus looks like.

And then we are building these in silico representations, and we are exploring the function of these molecules in great detail. And, you know, why we care about that is because at the end of the day, therapeutics, like, any sort of drug molecule that you might take once you've already contracted SARS-CoV-2,—and you're trying to cut short the duration of disease. Or for example, for people who are trying to design vaccines.

So, we hear a lot of talk about neutralizing antibodies. This will be the molecules that ultimately hopefully we all get injected with so that we can go back out and, you know, hug people again. Being able to design these and to understand how they work, requires that we understand the virus at the level of how the molecules move and how the atoms move.

And one of the things that's really unique I think about the computing in this space, and particularly, you know, with TACC is that lab experiments can't tell us everything we need to know.

By using these types of computational approaches we can actually get a much better view of what these system actually look like, and what we're dealing with so that we can better prepare and design therapeutics.

Tim Crawford: You know, and one of the things, Rommie, that we talked about on—as part of our lead up to this webinar—is how you're starting to understand the shape of the molecules or of the virus in this particular case, and how you're running different simulations as part of that. Can you touch on how this kind of applies to the TACC system, to Frontera?

Dr. Rommie Amaro: Right. Okay. So the virus, and viruses in general, let me just give a little bit of biological background first, and then I'll try to answer your question. So, viruses in general have evolved

this really interesting way of trying to sort of evade the human immune system, and that they basically cloak themselves in sort of a shield of sugar, or what we call the glycan, sort of a glycan shield.

And again, this is something that we know experimentally is there, but it's very difficult experimentally to actually get a view of. And so, what we've done is again, build these models of these glycan shields, which gives us really the first glimpse of what this sort of protective coating on the virus looks like.

And again, why that's important is because the shield is not like a perfect sort of shield of armor. There are holes in the shield, and we can go after these holes using therapies like the neutralizing antibodies or drugs.

And the thing is, is that these molecules are really big. And so machines like CAT—like Frontera in particular that have available to us, you know, all of the nodes and all the compute power, becomes really critical to being able to interrogate these systems with the speed and the precision that we need in order to get this done in a timeframe that we can turn it around and actually give this data to vaccine developers and clinicians so that they can use it, you know, in this ongoing battle that we have with COVID-19.

Tim Crawford: That's great. And I want to kind of dig maybe a little further into Frontera. And Dan, I'm going to shift to you a little bit. As we talk about Frontera, maybe you could give us some perspective on what Frontera is. I know you touched on it briefly as part of your introduction, but maybe you could delve a little further into Frontera for our audience.

Dr. Dan Stanzione: Sure. So, Frontera is a machine that the National Science Foundation funded at UT to address sort of these leadership computing class challenges, right, the biggest computing problems that the world faces. The machine is an about a year old now. It debuted at number five in the world and held on to that ranking through a couple of updates of the list that ranks supercomputers.

And so actually just one of those updates happened just Monday, and we've fallen to number eight in the world as new systems have come in online in Japan and in Italy just this week. But it's still among the largest machines. It's still the largest machine at any university on the planet.

It has a little more than 8,000 individual servers that make up the machine, and about 450,000 Intel processing cores, and then a very tightly coupled network to allow folks like Rommie to use hundreds of thousands of those on a single problem at one time.

And so, we actually support a range of different machines at different scales at TACC, and support many thousands of researchers around the country, but we reserve Frontera for the very largest projects. And we only have about 705 projects on the machine at a given time, so that each user can get enough time and big enough run to tackle some of these large problems.

So, in the ecosystem of machines that we have for university researchers, it's the one reserved for the very largest users. Since March when the pandemic response really got underway computationally, about 30% of the time on the machine for the last few months, has been dedicated in one way or another to COVID research.

You know, Rommie mentioned the sort of atomic level work that's going on, and that's been the largest computational user, is her project, and several others with similar goals.

And then we've also done those other things that you mentioned, working at the genomic level with a couple of dozen projects, and then a couple of dozen more, sort of modeling people in their interactions, either the epidemiology, the way the virus expands around the country, and things like tracking large-scale cellphone data to look at how people are moving and interacting.

There was a great piece just in the New York Times today that visualizes how all those cases spread that one of our researchers had some input on.

Tim Crawford: That's great. And you've talked about how Frontera is being used for some of the COVID projects, COVID related projects, compared to other solutions. There are many other solutions out there. Can you touch on that just briefly?

Dr. Dan Stanzione: Other solutions or other problems, Tim?

Tim Crawford: Other solutions. So, for example, when you think about like the HPC Consortium as an example, there are other solutions that are part of that consortium that are coming together to solve these big problems.

Dr. Dan Stanzione: Sure. Well, there's a number of architectures and other computing approaches that people use. So there is now a National COVID HPC Consortium that's been around for about 90 days, as of the time we're reporting this so—that is organized by the Office of Science and Technology Policy at the White House, that's coordinating across now about 15 large-scale computing providers, including Amazon, Microsoft, and Google as cloud providers, DOE labs, a few universities, and now some other resources around the world, UK, and Japan and a center in Switzerland, are all part of the consortium now.

And collectively, the consortium is supporting about 70 projects that are related to research on COVID-19. And so those do go to different sites, depending on what the computing needs are. So we have focused on a sort of CPU centric large-scale computer that we think is the most general purpose, but there are some more special purpose architectures, some things that are more GPU focused, and then there are some that are just big collections of more loosely coupled problems that you can run on cloud servers.

Most of those are when we have, you know, very large collections of small data analysis problems. So, you know, Rommie's problems for instance that use thousands of nodes at a time, wouldn't work very well in that setting, but there are plenty of other projects that do. So, we do, I think, have the most among those providers.

We're supporting about a dozen that have been assigned through the consortium, but there are some again, more specialized architectures and cloud providers all chipping in to try and help, you know, at least do the computational part of tackling the challenge.

Tim Crawford: That's great. So, I wanted to bring Intel into the mix, into the conversation, and talk about where Intel fits into Frontera. You touched on the 450,000 cores, Intel cores, but I know that programming and relationships play a role in this. Dan, maybe you could start off by sharing your perspective on how relationships and programming fit into your relationship with Intel.

Dr. Dan Stanzione: Sure. And, you know, I mentioned earlier about relationships with the researchers. And we at TACC sort of sit in the middle, between the end users, the scientists who are doing the work, and the vendors like Intel, who are providing the technology to do this work.

And so, you know, that set of relationships has also been critically important to us. So again, Frontera is built, integrated by Dell, but built around Intel technologies, including, you know, the Xeons [Intel® Xeon® processors], the latest Cascade Lake Xeons [Intel® Xeon® Scalable processors] that we used to build the system.

And we have some of the Optane [Intel® Optane™ technology] and the DIMM technologies for large memory nodes embedded. But it's far from our first machine with Intel, and in fact, it's sort of a linear follow on to our Stampede and Stampede2 supercomputers that we also built with Intel.

And part of that really make these things work, right, one part is having the chips that work, but another huge part of it is the software and firmware and tuning that goes with it. So, we've work very closely with Intel engineers to tune the message passing libraries that underlay the molecular dynamics codes that Rommie and others use, to make sure that we not only have these big machines, but they're really tuned for science.

And we worked with Intel to tune the applications to work well on them. And yes, that's been also something that goes on for decades, but has intensified in recent months around these specific problems.

Tim Crawford: Sure. And Rommie, I want to bring you in to this conversation a little bit. How much does that consistency play a role in the work that you're doing?

Dr. Rommie Amaro: Oh, I'd say it's really important. You know, and as Dan mentioned, especially for this particular problem, for COVID-19, you know, we needed to get up and going really quickly.

And, you know, having already had that sense of longstanding development of the code on these systems and the relationship already with TACC and the whole team there, it allowed us to get going—I would say at least one full month earlier than we would have, which, you know, in times like now, is actually really critical.

So, I think it's been really important for like timeliness of response, especially for this particular problem.

Tim Crawford: That's great to hear. And when you talk about the different kinds of projects, Rommie, that you're doing, you know, we've kind of centered on talking about COVID because that's kind of front and center in everyone's life right now, both professionally and personally.

But what are—maybe give some folks some perspective on other types of projects that your team and your group has been doing at UCSD.

Dr. Rommie Amaro: I can mention a couple. So, we work in sort of two major areas, one being disease related, right. There, we have a number of exciting projects, especially trying to develop anti-cancer therapeutics.

So that's been something that has been successful, and what we've been able to do there is actually use TACC, the earlier system actually, Stampede, which I think Dan also was involved with—where the tone of level models that were simulating, were able to find new drug pockets that people had never seen before, and those were now advancing drug molecules that sort of target this whole new modality of class of action to cancer.

And then another area that we're really excited about, and I think it could potentially be our next pandemic, is related to climate change and trying to understand how aerosols or these small sort of suspended particulates, particularly of ocean sea spray, how they sort of control chemistry and participate in various chemical reactions, as they float up through the atmosphere and do things like seed cloud formation or, you know, make it rain for example.

So, we're trying to understand these really small details that ultimately have a very—a much larger scale effect. But, it all sort of starts at trying to understand really the smallest steps that we can at the atomic level.

Tim Crawford: It just amazes me that you're solving these massive critical issues at the most atomic level, both figuratively and literally.

So where are TACC and UCSD and Rommie and your lab headed from here? And Dan, let me start with you and talk about TACC, and then we'll shift to Rommie and talk about her group.

Dr. Dan Stanzione: Sure. So, there's always a couple of things going on at TACC as we look towards future systems and sort of growing. And one is, well, really, we have sort of three threads.

One is operations, right? You know, how do we support folks like Rommie and many other researchers across all different fields of science that use our machines most effectively?

How is the software stack changing with the increase in artificial intelligence methods that we see working their way into more and more of the scientific workflows, the increase in very cheap digital data that's coming down the pipe that we have to deal with in large quantities for things like, oh, everything from autonomous vehicles to environmental monitoring to, you know, the Large Hadron Collider room particle physics.

So, we're sort of working operationally on how we can support users and involve the stack, the software stack to support them, but we're also always planning the future systems, right? So even now, although it's still several years away, Frontera is only a year old, but Stampede is now several years old, we're looking at the systems that will follow it, and the sort of next generation of technology, and then how

that's going to meet what the needs of science are going to be in future years.

So we're planning [inaudible] onto Frontera about four years from now, that should be 10x, both the capability of that machine, both in terms of being able to solve single problems faster, in the case of very large problems like Rommie's, but also to handle 10 times as much scientific work at the same time.

Tim Crawford: Wow, that's impressive. Rommie, how about you?

Dr. Rommie Amaro: Yes. Well, for us, we basically are going to continue to expand the complexity of the systems that we're studying, and also the size of the system. So, you know, the first study that we've done with SARS-CoV-2 for example, was just looking at the so-called spike protein, which is sort of the main infection machinery that it has. But what we're doing already now, we're trying to move towards simulating the whole SARS-CoV-2 virus, and then how it basically associates with the host cells.

So, we will sort of continue to expand the scale, the size of the actual problem that we are seeking to solve on Frontera and its successor.

Tim Crawford: That's incredible. So maybe a question on that. What is the constraint for understanding these big problems? Is it computing? Is it brain trust? Is it time? What tends to be the constraint when we talk about solving some of these big problems?

Dr. Rommie Amaro: I mean, it's a bit of all of that. I mean, one, you know, our models are driven by experimental data. And so one limitation or sort of something that we continually need to update, is as experiments, let's get more information about the landscape of the problem, so details about the virus or details about what it's interacting with in the human system, we have to go back and update our models to, you know, continue to make sure they reflect what we know as accurately as possible. And then, you know, we solve them—we try to solve it again.

So, I guess the brain trust part of it is like, what is the data that is feeding into our models to help make it more accurate? But at the same time, compute—having big compute solutions—we are a very hungry group of researchers in that sense, and we will continue to expand the types of questions we ask as these compute solutions grow and get bigger, you know.

So, we will grow to fit the size of the machines, I think. I don't know if Dan has a different perspective on it.

Dr. Dan Stanzione: No, that's a great answer, Rommie. And I was going to answer Tim's question with a yes, because it's all those things and it's really sort of an iterative process, right, what we see is, as we increase the resolution and the amount of physics that go into the computational models, it does sort of two things.

One is, helps you resolve any discrepancies with observations in the actual universe as to how things happen as those models get better. But it also leads you to new insights, right? So I mean, thinking in more of a sort of astronomy, astrophysics thing, you know, the studies going on of dark matter and dark

energy and what are these forces in the universe, are the result of discrepancies between model and observation, right, that leads you to new theory. It comes to that, right?

So, at some sense, especially when we're dealing with things that are very small, you know, working with atoms and fundamental nature materials, or things that are very large, like galaxies spinning around each other or colliding black holes, experiments are tough.

So, computation is how we replace very expensive or impossible experiments to gain new insights into that. And as the models get better and better, it makes the scientists ask new questions. And so, there's an iterative sort of process that happens, and, we keep providing these new tools and new exciting things keep happening with folks like Rommie, who can, move on to that next question.

Tim Crawford: That's great. And as we kind of shift from this section to Q&A, I just want to close this one thing that I know you've mentioned in the past about urgent computing. And so—and how it compares to the (HPC). And I think—I couldn't recall whether, Dan, it was you that had said that or Rommie had said that, but maybe you can just incredibly briefly talk about what you mean by that.

Dr. Dan Stanzione: Yes. So, you know, high performance computing is something that we do all the time and have a huge need for. It's sort of driven by the need to do everything from advancing basic science, you know, the study of those black holes kinds of things, or the fundamental mechanisms of science, to fast engineering, right, building better automobiles, self-driving cars, and there's a lot of time to solution pressure to do fast iterations to enhance the manufacturing process.

But we also have this new sub-branch of high-performance computing and what I would call urgent computing. And that is, around these sorts of natural disasters and other societal challenges, right? Today it's COVID. Two years ago if we were having this conversation, we would have been talking about the computational response to Hurricane Harvey.

It's summertime, which means right now we're doing a ton of tornado forecasting. We do a lot of earthquake modeling, you know, in a world with more severe storms and severe climate events, people—and going into regions that are flood prone, or just as the world gets more crowded, you know, these disasters have bigger and bigger impact.

We need sort of real-time computing response to that, and that drives a lot of demand that would otherwise go towards basic research and changing how we sort of operate these big machines. So, it's sort of a new driver for us and why we need to build very large and very responsive systems.

Tim Crawford: Yes. I mean, so one of the takeaways that I had from our conversation leading up to this webinar, has been that the work you're doing is not just academic. The work that both of you are doing is very mindful and very critical to humans and to the earth. And so, I think that's incredibly important.

Let's shift gears to Q&A. we've got a couple of questions from our audience. The first one is, how do you take your molecular structure and movement knowledge, and turn it into an algorithmic model to run on HPC?

Dr. Rommie Amaro: That's a great question, right. So, you know it used to be that sort of the field of chemistry was really just theoretical and people would, you know, write the theories in pen and pencils and so forth, right? But then we have all of these different ways of numerically encoding these theories and, you know, solving them now on these architectures.

And so, there's a few really important pieces that go into this particular type of computational modeling. So the first is that we—as I mentioned, we are describing the system at the atomic level. So, we have—we define basically a mathematical equation that tells us what the interactions should be between the different atoms and the types of atoms in our system.

And then we had this thing called a force field that has been developed over many decades that sort of are parameters essentially that this equation uses. And then all essentially, we're doing is we're using these great machines like Frontera to integrate Newton's equation of motion over time.

And that gives us essentially this dynamical propagation of the atomic movement, you know, sort of in their real biological context. So, I don't know if that's like too technical, but that's essentially—you sort of asked a technical question. So that's essentially how we do it. That's basically how we do it.

Tim Crawford: Great. So, our next question, and I think this one also, Rommie, is probably best for you to start with. I wonder if you could talk a bit about some of the inputs to the model. And is there a point when the model will be “done”?

Dr. Rommie Amaro: Great. That's another great question. So, inputs to the model. As I mentioned, it's various—for us, it's various types of experimental data. So very important for this, we're building three-dimensional representations of these biological themes or, you know, the viral components in the host cells.

So, we use structural data. Like cryo electron microscopy data, X-ray crystallography data, tomography. So, these are all ways that experimentalists have of acquiring structural—information about structural datasets in biology. We also then sort of merge that data with what we call a glycomic.

So basically, it's like the understanding what sort of this glycan shield looks like. They can do that using various mass spectrometry approaches. So, we take glycomic data, genomics data, lipidomics.

So, there's all this information that we have from experiments, and we use many of those to give us sort of this initial boundary condition, if you will, or sort of what the system looks like all put together. And then, you know, we bring it to life, you know, through stimulation.

Tim Crawford: Great.

Dr. Dan Stanzione: Yes. If I could expand on that a little, Tim. Yes. There's always another question to ask, right? And these models are always just sort of, taking our imperfect understanding of the universe and making it just a little bit better. So in some sense, that work is never done, but I think what a lot of people maybe fail to appreciate if they're not in the middle of the process, is that just because it's not done, doesn't mean we don't get enough information to be actionable in an impactful way, right?

And so, a lot of this information, like the structure of the S-protein that Rommie was referring to, is good enough already that it's being used upstream by other researchers. You know, we're working with a team at the University of Chicago that has a pipeline that based on what we know about the structure, are there compounds that we can say absolutely will not work as a drug, right?

And so—and they've gone through and thrown out millions of possibilities, right, and leaving just a few dozen that they've handed off to the medical chemists to try and synthesize and test, right, which speeds up this whole development of therapeutics and vaccines dramatically, right?

So, their search phase of what we have to test in new clinical trials, goes from millions to thousands based on what we know now. So, there is always more science to do. And, you know, I could give examples in a dozen other fields where similar things happen, right.

But just because it's not done, doesn't mean what we've done so far isn't useful and actionable. And, you know, dramatic responses can come out—so, that's sort of the process, right? Rommie and team, they run simulations. They do a bunch of analysis. They publish their results.

Then they have more work to do, and they move on to, you know, doing a more detailed model and answering the question in an even better way, or answering new questions but at the same time, that information does have impact and it gets used. You know, and we're already to the point where I think it's had a huge impact on our quest for therapeutics and vaccines for this.

And so, there's always more to do. The answer can always be better, but it's useful already. And I think making that distinction is important.

Tim Crawford: If we had more computers, more people, more scientists, we could do a lot more to answer those questions.

Dr. Dan Stanzione: Absolutely.

Tim Crawford: So, we've got a couple more questions and only a few minutes to maybe cover them. Let me shift gears into more of a lightning round and see how we can quickly get through these. So, my first one is, how long do simulations typically take?

Dr. Rommie Amaro: Oh, okay. That depends on the size of the system, and also the question that one wants to ask. But in general, I mean for us—and do you mean world-clock time? I suppose you do, or do you mean real-life time?

Tim Crawford: Yes. That's a great question. It's another question within a question.

Dr. Rommie Amaro: Commonwealth question. Yes. Okay. So, for us, for this particular system that we've, used most recently Frontera for, we're talking about simulating on a biological timescale, multiple microseconds. So maybe about 10 microseconds, and that took us about two months on the Frontera supercomputer, roughly.

Dr. Dan Stanzione: Yes. It was a huge range. You know, we have experiments and fortunately, we have the ability to sort of, what we call checkpoints, but start and stop simulations. So, they get sort of 48 hours on the machine, and then somebody else gets to run. But yes, but one experiment, we've had projects take up to a year and a half just to finish so a single run.

Tim Crawford: Wow. Okay, next question. How does the convergence of HPC and AI, impact our ability to solve for grand problems such as the impact of COVID, climate change, and natural disasters?

Dr. Dan Stanzione: Yes. Let me take a stab at that. Really, it's the convergence of our sort of traditional scientific numerical methods and AI, because the underlying computing is actually fairly similar. There are differences and when we're training neural networks versus during simulations, we can use reduced precision, and there are some, you know, fundamental computing differences, but it's—and so there's a sort of I think notion of convergence of HPC and AI.

We run AI workloads on HPC platforms. And the answer in short is yes, and we're making some tweaks around that. But the fast answer in science is, we're incorporating AI methodologies, particularly the notion of what we call surrogate models, which are sort of statistical based models inferred from data, as opposed to sort of models based on physical principles derived from first principles to accelerate the search base.

And there are so many open questions on how to use AI for science. One is in how you validate the answers and verify that the results are correct. But where we're seeing it most effectively—and even in this drug pipeline work, right, as you use the AI and the surrogate model, which is trained on, maybe you run the output of 1,000 previous simulations, you reduce your search base, right?

So, you take 10 million possible answers for how you either—whether we're talking about have a compound bind to COVID, or the shape of an airfoil on an airplane. Instead of having a million possibilities, you let the AI take you down to five or 10 candidates, and then you only run the deep physics on those. So, it is a technique that we can use to accelerate things greatly.

Tim Crawford: That's great.

Dr. Rommie Amaro: Yes, that's true. And if I can just add something just really quick. I mean, one of the things that we've definitely seen is that—like in the space of drug discovery—is that it's really at this intersection that this person was asking, the intersection of these physics based models, together with AI, is particularly powerful for making the model more predictive in a meaningful way.

Tim Crawford: That's great. And so, unfortunately, we are out of time. I first want to thank Dr. Dan Stanzione and Dr. Rommie Amaro. Thank you so much for sharing your insights for this webinar.

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.