HYPERION RESEARCH

HPC User Forum Update

Interviews with Steering Committee Members: Jack Collins, Frederick National Laboratory for Cancer Research

Steve Conway and Thomas Gerard
June 2020

IN THIS UPDATE

After the global pandemic forced Hyperion Research to cancel the April 2020 HPC User Forum planned for Princeton, New Jersey, we decided to reach out to the HPC community in another way – by publishing a series of interviews with members of the HPC User Forum Steering Committee. Our hope is that these seasoned leaders' perspectives on HPC's past, present and future will be interesting and beneficial to others. To conduct the interviews, Hyperion Research engaged Rich Brueckner (1962-2020), president of insideHPC Media. We welcome comments and questions addressed to Steve Conway, sconway@hyperionres.com or Earl Joseph, ejoseph@hyperionres.com.

This interview is with Jack Collins. Dr. Collins is the director of the Advanced Biomedical Computing Center at the Frederick National Laboratory for Cancer Research. Dr. Collins' research focuses on biomedical computing applications pertaining to cancer. His research group develops and applies high-performance algorithms to solve data-intensive computational biology problems in the areas of genomic analysis, pattern recognition in proteomics and imaging, molecular modeling, and systems biology. He received his BA degree in mathematics from Park College with minors in physics, chemistry, and economics. His Ph.D. is in physical chemistry (theoretical chemistry) from the University of Nebraska at Lincoln. Prior to joining NCI, Dr. Collins worked at SRI International (1985 to 1989) as a research associate and the Molecular Research Institute (1989 to 1992) as director of Computational Biology. Dr. Collins has over 70 peer-reviewed scientific publications and 3 patents. He has given dozens of invited talks and Webinars at international and national meetings and conferences and serves on numerous review panels and advisory committees.

The HPC User Forum was established in 1999 to promote the health of the global HPC industry and address issues of common concern to users. More than 75 HPC User Forum meetings have been held in the Americas, Europe and the Asia-Pacific region since the organization's founding in 2000.

JACK COLLINS INTERVIEWED BY RICH BRUECKNER, insideHPC

Brueckner: Hi, I'm Rich with insideHPC. I have Dr. Jack Collins with me today. Jack, welcome.

Collins: Thank you, Rich.

Brueckner: Well, Jack, let's start at the beginning. I've known you for a long time, but how did you get involved with HPC?

Collins: In graduate school I was in quantum chemistry. Calculating molecules at that time with non-orthogonal valence bond methods, there was a lot of calculations involved. We were using IBM computers and my input/output device for storing all of my integrals was nine-track tape. Not necessarily the fastest in the world. I got a chance to work on a [Control Data Corporation, or CDC] Cyber 205. Some of the folks there had friends at CDC, so I ended up trying to use the first Cyber 205 Fortran compiler and calling them three times a day and them saying, "now try this, I think we fixed that bug."

From there, I then got to use a Cray when they started to become available more generally, and when the compiler worked, and it was so easy to use -- that was it. Instead of waiting forever, you want to graduate, and you want to get gone, so people started using Crays. The NSF supercomputer centers were coming online, and I used virtually any one of them I could get a hold of because time-to-solution counted. From then on, I've been doing computational work and that's sort of it. It was a necessity and a desire to get done a lot faster and get a real paying job.

Brueckner: Okay, so, where do you go from there after that first experience with the Cray?

Collins: I went to Stanford Research Institute and we worked on, again, still, quantum chemistry. But then we were starting to work on, instead of small molecules, we were getting into drug molecules and then I started modelling enzymes and proteins and trying to predict their reactivity and how drugs interacted with the protein. And so, at that point I needed a bigger and bigger computer and faster and faster and, so that's when I started using the NSF resources. First at Pittsburgh and then San Diego, and then anybody that wouldn't throw me off. I would use anything that we could get. That's essentially how I ended up here at NCI because they bought a Cray for biological usage and we knew some of the people that were on the advisory panel and committee here. I was invited to be able to use it and I would fly out and spend a week, day and night, 24 hours-a-day almost, to keep the thing running and then take my tapes and go back and analyze the data. That's how I got here at the NCI and into HPC and the whole supercomputing thing, and then I just stayed here for a long time.

Brueckner: You must've seen a lot of change in computing. What were the biggest changes in HPC?

Collins: The miniaturization of all of the silicon, and when I was able to buy a system with GPUs in it, s As a matter of fact, we built one in the early 2000s with water-cooled GPUs in it to do molecular docking, and when we could do docking and types of dynamics on a relatively small system as fast as we could do it on the supercomputer, that's when I knew that there were some major changes happening. I'd been playing with FPGAs and we could get great speed-up but the programming paradigm for that was so difficult and the debugging and everything else. It just wasn't something that you could maintain and keep going for code that was going to be changing quite a bit. That was a big deal.

[Collins showed his smartphone to Brueckner.] This is like 2,000 Crays speed-wise and you've got enough memory. Today, we've got terabytes of memory that we can have access to. The first tapes and drives that we had were in the megabyte ranges, right? So, the amount of data that we have now, the amount of processing power, I don't even have to go to a supercomputer. When I go to a supercomputer now, we're talking about using 700-800 GPUs at a time. The amount of computing is just amazing but the amount of data that we collect now is also just amazing. That convergence of machine learning, data, supercomputing, everything together, that's really changing what we're doing now and that change has really exploded over the last several years.

Brueckner: So, where do you think HPC is headed?

Collins: That's an interesting question. The answer to that is really, how do you define HPC? If HPC is the next thing out there over and above what is available to the general population, I think it's going to get more heterogeneous and I think it's going to get more specialized and it's going to be not a single Cray that kind of fits everybody's niche, but it's going to be a bunch of different systems all hooked together.

If you mean HPC as in really high-performance-computing, the fact that I can buy a 4-GPU machine, stick it under my desk, and have enough power to run it and not burn down my house is kind of an amazing thing. I think that it's really going to become completely ubiquitous, everything that we have. Especially as machine learning, AI and deep learning get more prevalent, everywhere you're going to have these engines, everywhere. We're playing with these Nvidia Nanos and Jetsons and things to plug in and you're just going to have HPC everywhere. It's going to come to the point where, like our children today, they don't know that cell-phones weren't around. They're just never going to know that HPC or all the computing you could ever want wasn't available to them. I think it's just going to be everywhere. As I said, I think there are two different definitions. I think it just comes down to which way you want to define it.

Brueckner: Jack, are there any computing trends in HPC that have you particularly excited or concerned?

Collins: Well, every new technology has two edges to that sword. The ability to compute and what we're doing now, where we are integrating massive amounts of genomic information with imaging information and proteomic information for human health, is incredible. We're taking advantage of a lot of AI methodologies, the ability to slice and dice all sorts of imaging modalities and visualize that and change it and morph it. At the same time I can see, online, it's almost impossible to tell what's the difference between real and what's been hacked and morphed into something that never occurred but it certainly looks real on the computer. I see what people can do with that same computing technology, that blurring between reality and made-up. I think our ability to take advantage of advanced algorithms, bring in all of the data, merge it in a way and just tweak it enough. We have enough compute, we have enough algorithmic knowledge, so blur the line between what's real and what's not, what's the truth and what's not. It's going to become very blurry. It's already becoming blurry.

When all this computing power is used to discover the truth and help society's problems, whether it's, in my case, health, or whether it's in global warming, climate change, the economy, inequality, any kind of a thing where anybody would be looking at that, to be able to use it for the good makes it a really powerful tool. That tool can also can be used to bring people, essentially, to the dark side and not know what the truth is and bring distrust. We have to have an ethical way of doing that. People talk about ethical AI, and there's also ethical HPC. There's ethical whatever-it-is, if it's a technology. Those

are things that I'm concerned about. It's really hard. "Seeing is believing," is what people used to say. I can hack a video and make it look like anything in an afternoon. That's potentially societal altering. We have to be very careful with that.

Brueckner: Is there anything else you wanted to share with the audience today before we go?

Collins: I would say that if you have any interest in computational aspects or how computing is being used and applied to change the world and solve some really hard problems, the High Performance Computing (HPC) User Forum is a great place to go and talk to people and share what you know, and if you've got new knowledge please come and share with the rest of us. I think that it's important for people to get together and share, whether it's going to be virtual for a while or whether it's going to be together. That's one of the places where you build that network, you learn from people, and you can actually change things. A lot of the things I know today came about from either interaction I had at the HPC User Forum or people giving a talk or whatever. I think it's a really great environment. Supercomputing is huge. There is a lot of stuff going on but it's huge. HPC User Forum is where you can sit down and talk to people.

Brueckner: I agree. And I'm really looking forward to the next time and seeing you at one of these meetings in person once again.

Collins: Me too.

About Hyperion Research, LLC

Hyperion Research provides data-driven research, analysis and recommendations for technologies, applications, and markets in high performance computing and emerging technology areas to help organizations worldwide make effective decisions and seize growth opportunities. Research includes market sizing and forecasting, share tracking, segmentation, technology and related trend analysis, and both user & vendor analysis for multi-user technical server technology used for HPC and HPDA (high performance data analysis). We provide thought leadership and practical guidance for users, vendors and other members of the HPC community by focusing on key market and technology trends across government, industry, commerce, and academia.

Headquarters

365 Summit Avenue
St. Paul, MN 55102
USA
612.812.5798
www.HyperionResearch.com and www.hpcuserforum.com
