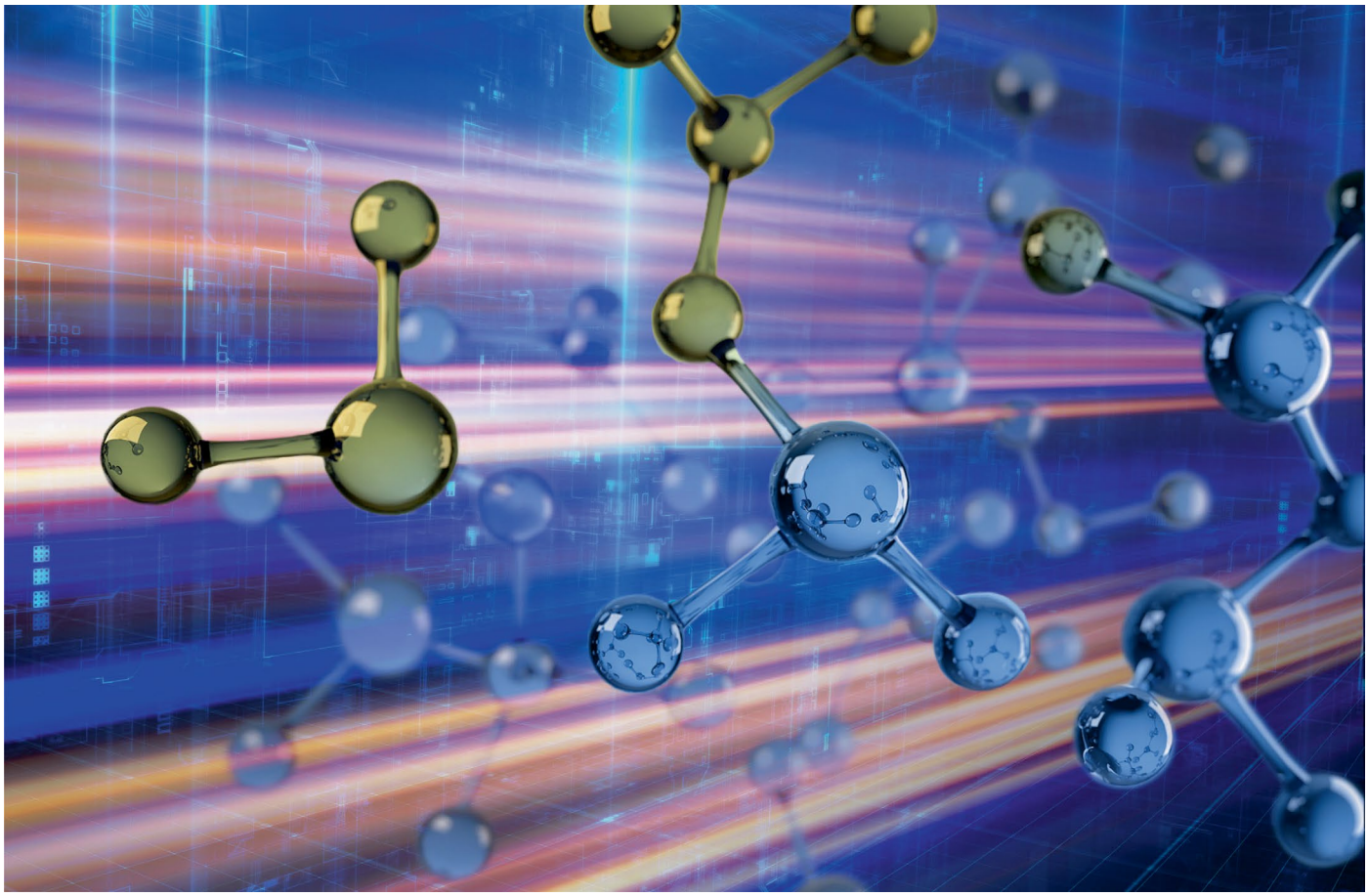
https://doi.org/10.1038/s41592-024-02202-z



Collaborate, disseminate, accelerate

Collaborations between researchers and companies can progress swimmingly and teams quickly validate findings and mature methods. All too often, things can't advance and the 'Valley of Death' looms. New ways to collaborate, underpinned by computational muscle, can help.

By Vivien Marx

René Descartes, in the seventeenth century, "was probably the last major thinker to believe that science could be conducted in splendid solitude," writes science historian Lorraine Daston, director emerita of the Max Planck Institute for the History of Science, in *Rivals: How Scientists Learned to Cooperate*[1]. Science knows competition and rivalry. Alongside such behaviors, collaboration between researchers has become commonplace. Given events such as the COVID-19 pandemic, she notes, "the scientific community has shown itself capable of consensus and concerted action."

Concerted action can mean collaborating with others across campus, in international consortia or with companies. In all cases, what emerges must be validated, matured and disseminated. Papers and repositories are ways to disseminate findings, methods and technology, but additional modes are emerging for quicker and broader approaches. In many instances, infrastructure needs to be scaled up. In the realm where academia and companies meet, the collaboration dynamic is changing. Some companies launch with businesses models supported by deep-pocketed funders who approach commercial activities and collaboration in different ways than was previously typical. Some companies get involved in academic research projects earlier. Government-funded public–private collaborations support this dynamic by giving a broader group of scientists new types of computational tools to collaborate. It means

Collaborations with companies are taking new routes for transferring findings, methods and technology from academia into the commercial realm.

wider access to high-performance computing environments to run complex software, as well as machine learning and artificial intelligence (AI) tools. Getting high-performance computing workflows to interact is a form of technological collaboration and can get just as tangled as collaboration between people.

Changed tech transfer

Traditionally, companies become university partners by licensing a method or technology from an academic lab and commercializing it. But gone are the old modes of technology transfer, says neuroscientist Stephen Susalka, CEO of AUTM. AUTM used to be an organization of tech transfer managers and has reinvented itself as a global organization focused on the different levels of technology transfer. Its members are in over 60 countries at companies as well as universities and national labs.

Back in, say, the 1980s, says Susalka, inventions at universities were "essentially written on the back of a napkin." Next, a patent application was filed and a tech transfer office reached out to companies about licensing it. These days, a tech transfer office will handle many more tasks to get a collaboration with a company going. They might finance a prototype and iterate on it with a researcher, which makes it easier for companies to assess an idea. Before joining AUTM, Susalka iterated on projects in this way when he was technology transfer manager at the University of Virginia and Wake Forest University. This approach addresses a trend he began seeing of companies seeking to lower the risk in their involvement with universities. Given the commercial de-risking tendencies, universities have to advance concepts emerging from their labs further than they previously had to.

When academics advance their idea beyond the needs of their own lab — and the idea can be a method, a technology, a drug target or other type of invention — they hunt for how to fund next steps. That can sometimes involve venture capital. But the invention can get stuck in what is colloquially called the 'Valley of Death' — a feared chasm between funding and commercial readiness of a method or technology. "That valley is expanding," says Susalka, "because companies are becoming more and more risk averse." To close that valley a bit, he says, startup companies take on ideas that others find too risky. Startup companies have become, he says, the bridge between universities and established companies. Gene editing is one of areas in science in which this approach is taking place.

Another trend in tech transfer, says Stephen Susalka, is one toward earlier commercial partnering and engagement in research.

Another trend in tech transfer, says Susalka, is one toward earlier commercial partnering and engagement in research. These interactions do not start with invention disclosure followed by evaluation of commercial potential, patenting or copyrighting. Instead, tech transfer offices open the door to companies in a project's early phases of a method or technology development.

Through this earlier engagement, companies learn about a project, can influence the research and might decide to sponsor it or license an aspect of it. Long before an invention has been created, a company can express interest.

In some collaborative projects, scientists from universities or national labs huddle over a basic research project to mature and scale it for many users as a product or commercial service. At whiteboarding sessions, a tech transfer person might be present. "You're trying to get the right people together, working on projects that have tangible outcomes," says Susalka. With such collaborations it's important, he says, to put in place a confidentiality agreement stating that what is discussed stays between project members.

Physical proximity of collaborators can support interactions, and many collaborations between researchers in academia, non-profit research institutes and companies "are driven by who's in town," he says. But he sees many long-distance collaborations, too. Whether they are near or far, as researchers work together, discussions will emerge about "whose invention is that?" Perhaps contributions come from two academic labs, which is when inter-institutional agreements come into play. They are becoming ever more common between scientists. Says Susalka, "the tech transfer office has their back there."

New business models

Some firms blend traits of public–private collaboration and translation into their structure. Altos Labs, a privately held company, focuses on cellular rejuvenation research. In its labs in the San Francisco Bay Area, San Diego and Cambridge, UK and in their interactions with external scientists, it pursues deep scientific questions. Altos seeks to combine "the best of academia and industry."

Privately funded Arena BioWorks strikes a similar note. With 50,000 square feet of lab space in Cambridge, Massachusetts, it has launched an independent biomedical research institute. According to a statement, its model is about accelerating basic research on biological mechanisms underpinning processes such as aging and challenges in immunology,

At Pacific Northwest National Laboratory Karl Mueller, Karol Kowalski and Christina Lomasney work with PNNL colleagues as well as company scientists in a project called TEC[4] to give the computational chemistry community access to high-performance computing resources.

oncology and neuroscience, and — under the same roof — translate findings into therapies.

At the World Economic Forum in Davos, Switzerland, investor and philanthropist Steve Pagliuca, former chair of Bain Capital, and executive chair of Arena, said in an interview with CNBC that the COVID-19 pandemic revealed how to accelerate drug development. As they bring together scientists and leverage AI, along with advances in computational hardware, he said, "we're trying to create the Bell Labs of biotech." The company is taking "best demonstrated practices" from university science and pharmaceutical companies and combining them in an entity akin to Bell Labs, he said. The reference is to what was originally called Bell Telephone Laboratories, a research and development laboratory. Among other inventions, its scientists and engineers developed the transistor, the laser, the charge-coupled device, information theory and much more.

As Arena BioWorks projects and companies emerge, the plan is to encourage them to collaborate. Scientists external to the company are being recruited, one of whom is Keith Joung, who has worked on gene-editing tools at Massachusetts General Hospital. In a statement, Arena CEO chemical biologist Stuart Schreiber says that the company's single source of funding "frees our scientists from the typical short-term cycles of grant and venture capital funding." Says University of Chicago researcher Yamuna Krishnan, "I love the sales pitch 'Bell Labs for biology'. Now they have to live up to that claim."

Steep scale-up

An essential factor when scaling up research to advance science and collaboration involves infrastructure access. Three months after the Biden administration's Executive Order on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, the US National Science Foundation; ten other collaborating government agencies, including the National Institutes of Health and Department of Energy (DOE); companies; and nonprofit and philanthropic organizations launched a two-year pilot project called the National Artificial Intelligence Research Resource (NAIRR) Pilot— to build a shared research infrastructure in AI.

A portal gives researchers access to resources from many companies, non-profits and government organizations. There are models and 'AI-ready datasets', software systems and analysis platforms. To use the platforms' offerings, users can submit proposals online. The criteria for deciding on projects includes aspects such as the readiness of a project for "near-term progress" and the feasibility of the proposed technical approach. NAIRR is currently only open to US-based researchers and educators. Vetted users gain access to a variety of computational resources from the projects' partners, such as Amazon Web Services, IBM, Microsoft, OpenAI and OpenMined, as well as resources from non-profits and government agencies such as the National Institutes of Health's ImmPort, an immunology data-sharing platform, and NASA's Earth Science AI training datasets. National Science Foundation director Sethuraman Panchanathan said in a press conference that NAIRR also addresses "the other AI," which includes accessible infrastructure, advancing ideas and accelerating innovation.

Sustainable ecosystem

Collaboration can flourish within a sustainable ecosystem. DOE has invested heavily into computational chemistry, which has led to a rich tool landscape, says Karol Kowalski, a physical chemist at the Pacific Northwest National Laboratory (PNNL). Computational chemistry describes the behavior of molecules, exploring the mechanisms behind reactivity, bond formation, bond-breaking processes and many other facets of molecular interactions. Its approaches are applicable to many molecules and started with hydrogen and trusted analog tools such as slide rules.

Each advance in computational technology applied to computational chemistry, from analog to digital to quantum computing, begins with benchmarking to the H_2 molecule, says Kowalski. This started, he says, with the collaboration of two scientists whose work has stood the test of time. In the 1920s, Walter Heitler and Fritz London collaborated on what has become known Heitler–London theory of valence bonds. They applied quantum mechanical methods to establish the valence bond theory for hydrogen.

Many scientists, not just those with access to DOE supercomputers, want to run complex open-source software efficiently, says Karol Kowalski.

As the field of computational chemistry has advanced, it's taken huge efforts to write and adapt code as each new computing tech chapter unfolds. "We are always busy because we have to harness the newest and the greatest technology," says Kowalski. Inspired by the latest and greatest computational technology, computational chemistry researchers are "always chasing the running target." New algorithms, software and computational workflows emerge, says Karl Mueller, a chemist who manages scientific mission development at PNNL and who was interviewed jointly with Kowalski. But "then, when you try to put those codes together, they don't necessarily work quite as well as you would like them to." Being able to use the latest tools and on a platform that is continually updated, says Mueller, would be more than just having a repository.

To make the computational chemistry ecosystem sustainable and keep the field vibrant, it must stay abreast of technology[2]. That means retaining people, advancing methods and technology, and assuring that computational tools and systems talk to one another in

CREDIT: A. STARR, PNNL

a broader way so the field itself and collaboration within it can advance.

As users and developers of tools tackle complex chemical challenges, they need to integrate parallelism into their workflows to undertake multiscale modeling of complex chemical landscapes. They might work on new types of fertilizers or biodegradable plastics, explore complex genomic phenomena, or seek ways to use X-ray free electron laser sources and other technologies to probe matter and explore intramolecular electron and proton motion on ultrafast time-scales.

To run predictive simulations on new and changing computational architectures, the research community must rise to the scaling challenges with a stable, sustainable programming environment that supports improving current methods and developing new ones for many researchers. In a collaborative project underway, scientists work on just that.

A team of 30 researchers is in a two-year collaborative project initiated by PNNL called Transferring Exascale Computational Chemistry to Cloud Computing Environment and Emerging Hardware Technologies (TEC^4). It involves scientists at PNNL, Microsoft and Micron, and it is one of a number of DOE-funded projects for scaling and accelerating projects with a view to technology transfer in chemistry and materials science for areas such as sustainability and energy security. The overarching $73-million project is called Accelerating Innovations in Emerging Technologies to support how basic research can deliver innovation and economic impact.

The plan is to complement high-performance DOE computing facilities with a cloud-based infrastructure for the wider scientific community and for broader collaborative projects. Many scientists, not just those with access to DOE supercomputers, want to run complex open-source software efficiently, says Kowalski, "which is really needed in modeling chemical processes and transformations." The project will enable exascale computational chemistry for many scientists to use and to empower collaboration. The United States has two exascale computing facilities: Frontier at Oak Ridge National Laboratory and Aurora at Argonne National Laboratory. El Capitan at Lawrence Livermore National Laboratory is set to be ready later in 2024. Microsoft Azure Quantum Elements is in the exascale computing class, too.

A desktop computer, says Kowalski, can do one trillion – 10^{12} – floating-point operations per second, or one teraflop. To perform math operations, floating point numbers are the way computers prefer numbers to be written. Exascale computing is 10^{18}, or one quintillion, floating point operations per second. "Imagine you have aligned two hundred, three hundred thousand desktops," says Kowalski, but they are neither connected nor synchronized. The project team works to orchestrate computing so that its cumulative power lets, in this instance, computational chemists solve problems more quickly. On its way to the desired calculations and simulations, the workflow needs to not get stuck.

Nathan Baker is part of the TEC^4 team. He is a former PNNL researcher who directs Microsoft's Azure Quantum Elements chemistry and materials science projects.

One foe is latency, which, in computer gaming, is due to wobbly internet connections. In high-performance computing, latency-based traffic jams can be due to slowed interactions between interconnected modules in a computing system. High-performance computing systems have many nodes, and data must travel with low latency, which means as seamlessly as possible, says Kowalski.

Exascale collaboration

Computational chemistry uses quantum mechanics to address chemistry problems, says Mueller. But when these problems are computationally scaled, the equations quickly become unwieldy. Scaling these calculations and working on complex predictive models brings greater chemical accuracy and a way to support both experimental efforts and experimental design.

One challenge, says Kowalski, is based on strong inter-electron correlation effects, which, for example, happen in catalysis. Scientists need accuracy to explore, for instance, a more energy-efficient way to make less environmentally problematic fertilizers. The accuracies, he says, deliver understanding and lend predictive power. To move toward the needed accuracies, one needs high-powered computing in which software and hardware themselves have collaborative configurations. It's computationally challenging to get to that power.

Data-intensive computing benefits when there are more CPUs. At some point, however, latency rears its ugly head and the exchange between processors, memory and storage hits a bottleneck. Says Tony Brewer, a senior system architect for scalable memory systems at Micron and who is also part of the TEC^4 team, when multiple servers need to share the same data, data must be copied from one server to another. It's time-consuming, inefficient and currently unavoidable. Micron is a large semiconductor company that makes dynamic random access memory (DRAM).

An emerging industry standard for interconnects is called Compute Express Link (CXL). It can enable memory sharing and disaggregated memory. Some existing CXL modules each have a capacity of 256 gigabytes and move data at 64 gigabytes per second. The TEC^4 team is still determining which class of Micron memory modules will work best for their needs. The team wants to let computational chemists be able to scale memory in line with workloads, says Brewer.

With new CXL based multi-host, shared memory systems, the data are shared such that multiple hosts, be they CPUs, GPUs or xPUs, can directly access a shared memory pool, "which can be dynamically scaled up and down as needed by the researchers," says Brewer. The CXL fabric has lower latency than Ethernet or InfiniBand, so processors stall less as they whip through trillions of calculations in computational chemistry.

Collaborative whiteboarding

Whiteboarding invariably means getting hands dirty from markers. In their collaboration sessions, says Kowalski, "the hands, they're really dirty." Collaborative whiteboarding lets the team explore how to best predict chemical properties of a material and work out the details of the best computational architecture, in which software tools talk to one another to help users better understand molecular behavior, he says. They discuss synthesis or fabrication and "how we use these kinds of tools to really make things in novel in unique ways," says PNNL chief commercial officer Christina Lomasney.

CREDIT: MICROSOFT

"I don't think any of us could actually resist whiteboarding," she says. It helps to discuss where this project can lead next. Lomasney is a physicist who joined PNNL after running a materials science company called Modumetal. Whiteboarding together stimulates creativity and is a way to assure "we understand the problems in the same way," says Kowalski. It's also how to look at an aspect of a problem from different perspectives "because you can come up with alternative solutions."

TEC^4 is focused on classic high-performance computing, says Nathan Baker, who is on the team and manages Microsoft's Azure Quantum Elements chemistry and materials science projects. Down the road, quantum computing might be involved, too. The team's emphasis is on getting the highest accuracy using the most scalable calculations to characterize chemical systems, he says.

The project dovetails with Micron's work to design memory systems for new workloads and environments because "sometimes yesterday's solutions are just not robust enough," says Brewer. In this case, it's about moving large scientific applications to the cloud. Cloud environments, he says, "have largely focused on areas other than the high-performance computing space."

In TEC^4, collaborative whiteboarding sessions happen, among other places, in Seattle at the Microsoft Azure Quantum Elements team hub, and sometimes Micron researchers are there. Micron is expanding its Boise, Idaho, campus for R&D and manufacturing of memory products. PNNL is in Richland, Washington, which is 200 miles from Seattle. Collaboration might happen in PNNL's Seattle office or on the University of Washington campus. Not all issues can be solved in virtual meetings, says Kowalski, which is why they set up the collaboration with nearby companies.

"Scientific collaboration is based on trust," says Nathan Baker, especially trust built over a long time.

Microsoft, says Baker, wants to support making computational chemistry run more smoothly and for more researchers. Along the way, the team is addressing the way software tools talk to one another in a complex workflow. The plan is to enable simulations that run in a cloud-based high-performance, even exascale, computing environment.

Academics as well as small- and medium-sized companies currently do not have access to this kind computing power, says Mueller. So this project will, the team hopes, put this group of people in a position to perform complex simulations and collaborate in a way they wish they could now do. As the project moves forward, the plan is to begin testing it with an early cadre of users.

Trust and incentive

Baker was a PNNL researcher and faculty member on the applied mathematics faculty at the University of Washington. In October 2022, he became product lead for chemistry and materials science at Microsoft Azure Quantum Elements. "I'm a chemist," he says. It's his training and "the field that's nearest and dearest to my heart."

He has worked closely with Kowalski and Mueller and found it tempting to work on high-performance computing and AI to solve hard problems in chemistry and materials science.

To expand access to tools, the TEC^4 team is building a platform so scientists who do not have access to a DOE supercomputer can run these tools in a high-performance computing environment, says Baker. With shared scientific instrumentation or facilities such as a supercomputer or a beamline, "there's often a wait time," which can be circumvented with this type of cloud computing.

From other collaborative projects between PNNL and Microsoft, "we have a lot of trust that was built up amongst partners," says Mueller. The software tools in TEC^4 are open- source tools the computational chemistry community knows. These tools are from many labs, and "we all know what each other's codes can do," says Mueller. "That's a much easier way to collaborate than if you're bringing together secret pieces."

As Baker explains, researchers at all career levels — postdoctoral fellows, early career scientists, mid-career and established scientists and programmers who work in chemistry, materials science and applied math — are involved in TEC^4. Another team member is trust. "Scientific collaboration is based on trust," says Baker, especially trust built over a long time. The working together is anchored in shared goals.

Right up front, the team had candid conversations about questions that can lead to tension if they crop up late. For example: who is going to be the author of a paper? And how will the group handle intellectual property? This track record of collaboration and setting up a framework for candid conversations, says Baker, "creates that safe space."

As a federally funded research and development center, says Lomasney, part of PNNL's statutory mission is technology transfer for the benefit of national security and the economy. For TEC^4, this translates to access to these resources for academics or researchers at small and medium-sized companies. They don't have to come to PNNL to use resources on site. This project has brought on a recalibration of how to collaborate with the scientific community. "We're now rethinking the way that that small and medium-sized enterprises can engage with the computing resources that are necessary for some of these really hard problems," she says.

At the outset of TEC^4, Lomasney assured the researchers about the mechanisms in place to protect ideas and aspects that might need copyright or patents. It's important to talk about such issues, and "by virtue of giving someone ownership of an idea, you open up this opportunity for collaboration," she says. Even for discussions at the whiteboard, an agreement was put in place that all agreed to.

Among the incentives for researchers to take part in this kind of project is the fact that it can raise the visibility of their software tools. Especially to early career researchers, this matters. Says Kowalski, "they get the visibility, they get citations, they can build their career on top of that."

Vivien Marx ✉
Nature Methods.
✉e-mail: v.marx@us.nature.com

Published online: 27 February 2024

References

1. Daston, L. *Rivals: How scientists Learned to Cooperate* (Columbia Global Reports, 2023).
2. Felice, R. D. et al. *J. Chem. Theor. Comput.* **19**, 7056–7076 (2023).