The Coherence Times

A Quarterly Readout

Accelerating scientific research

2 State Preparation
14 Gates and Operations
18 Observations
21 Phase Kickback
26 Error Mitigation

Our quarterly leadership update

A note from Jay

Hi all, and welcome to the Quarter 4 2024 edition of the Coherence Times: Accelerating scientific research.

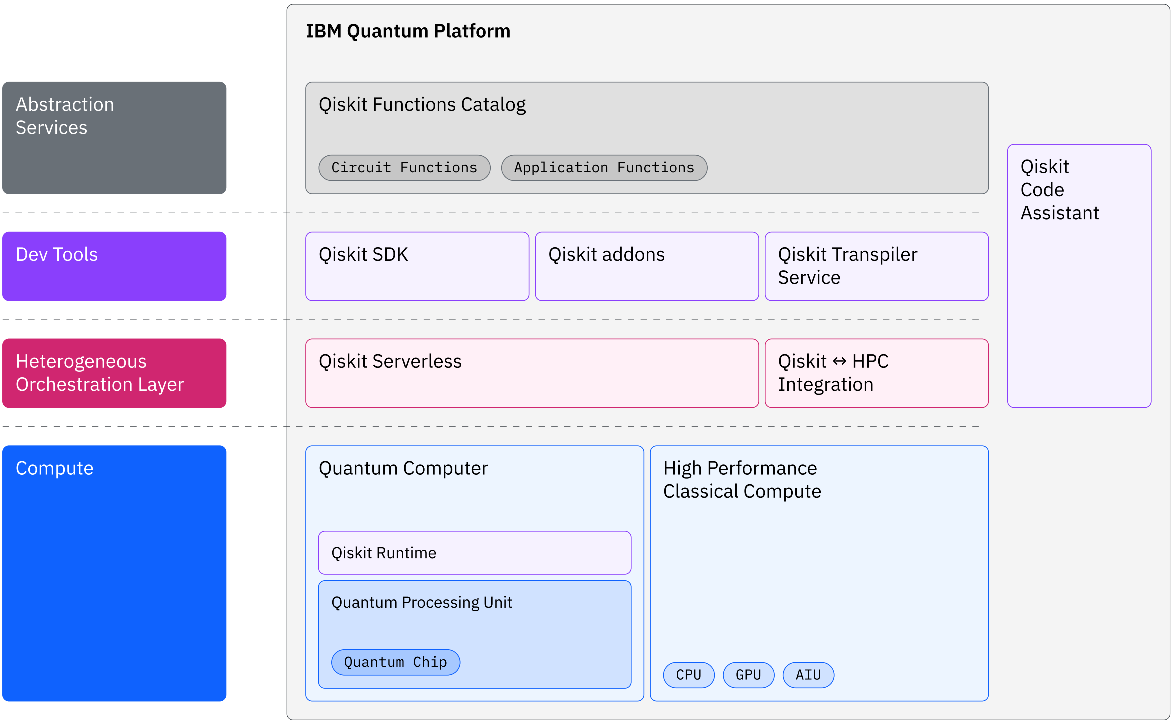
In November, at our first-ever Quantum Developer Conference (QDC), we delivered a State of the Union address reporting results two years in the making: we announced that we'd successfully and accurately run circuits with 5,000 gates. This is the culmination of the 100x100 challenge that we issued in 2022—100 layers of two-qubit gates acting on 100 qubits is equivalent to 5,000 two-qubit gates. We chose this number because it sits firmly within the regime of circuits beyond the ability of exact classical simulation of the wavefunction, a regime we refer to as quantum utility.

This milestone represents something much bigger than just the improvements we've made to our hardware. Together, our hardware and software now provide users the performance and ease of use required to accelerate scientific discovery with quantum computers. At the State of the Union, we showed that startup partners QEDMA and Algorithmiq are doing research approaching the 5k scale using IBM Quantum hardware and software. And new Qiskit tools—specifically Qiskit addons—have extended research developed at RIKEN over the course of a year, such that our partners at Cleveland Clinic Foundation and Lockheed Martin were able to complete equivalent work in a matter of months.

The following week, we delivered an even bolder message during my invited talk at the Supercomputing 2024 conference in Atlanta. We told the crowd that **quantum advantage will happen within the next two years, but only if the HPC and quantum communities work together**. We also offered a glimpse into the future of quantum: the first-ever demonstrations of quantum-centric supercomputing with our partners at RIKEN, showcasing the use of the supercomputer Fugaku to assist our Heron processors, and RPI, running a Qiskit addon as part of a heterogeneous compute workflow.

Now that we've set a timeline for quantum advantage, it's more important than ever that we work together to build the best hardware, middleware, software, and services, while also encouraging clients to use IBM Quantum tools for their own discovery. But it's your hard work that has gotten us this far, and it's thanks to all of you that I feel confident in our prospects for success.

Our quarterly leadership update continued



At our 2024 State of the Union, we introduced this diagram, showing how our hardware, middleware, and software tools work together. At the bottom layer, you can see how we define a QPU (a quantum chip plus the required control electronics), plus a quantum computer (a QPU plus Qiskit Runtime). Moving up the stack, you can see middleware tools like Qiskit Serverless and HPC integration. We then offer tools for developers to write, optimize, and run workflows on these resources, plus abstraction services so users can begin exploring quantum without hand-tuning circuits.

Defining quantum advantage

Let’s take a moment to clearly define what we mean when we say “quantum advantage.” At its core, quantum advantage refers to an information-processing task being performed more efficiently, cost-effectively, or accurately using quantum devices than is known to be possible with classical computers alone. On the surface, this definition is straightforward and likely familiar to many. However, there are two important nuances that often go unaddressed.

The first has to do with ensuring our solutions are accurate. Accuracy is essential for any comparison of computational advantage between any two methods for some computing task. Otherwise, we may find ourselves claiming that an efficient method that returns inaccurate solutions has advantage over a less efficient method that returns accurate solutions. The challenge is that, in practice, how we ensure trust in the outputs of a real and noisy quantum device is far from obvious. This is especially true for problems where classical computers struggle, and we only have answers from the quantum computer. In some cases, we may be able to directly check the accuracy of the quantum computer’s outputs, but other problems exist where ensuring the accuracy of results requires establishing trust in the quantum computation itself.

The second is that quantum advantage is not something that will happen at a singular moment in time, nor is it a definitive “claim” one should make. Why? The reason lies at the heart of scientific inquiry itself. The philosophy of science teaches us that scientific claims are not provable statements, but rather are hypotheses subject to falsification. Through repeated attempts to challenge and disprove a hypothesis, we may gain confidence in its truth — but absolute certainty remains impossible.

This same principle applies to quantum advantage. When we reach the point where we believe a quantum device may have demonstrated an advantage, this belief will be a hypothesis. We must then allow classical computing to respond — and it should not be underestimated! If a hypothesized quantum advantage is later eclipsed by classical computing, it does not represent failure. On the contrary, it is a testament to the scientific method. In fact, this sort of back-and-forth is healthy and should be encouraged, for it will lead to improvements in both classical and quantum computing. We have already seen evidence of this following the publication of our quantum utility paper in 2023. As you may recall, we were very careful **not** to claim a quantum advantage with that paper.

We fully expect an intense period of back-and-forth competition, during which continued improvement to both classical methods and classical computers is all but guaranteed. Such advances should not be seen as setbacks for quantum computing! After all, today’s classical computers are truly extraordinary machines, arguably the most powerful tools ever created for scientific exploration. However, we also believe that the time is fast approaching when quantum computing will eclipse classical computing — not for all tasks, but for certain tasks.

Hypotheses of quantum advantage will stand so long — and only so long — as classical computing fails to measure up. The challenge before us is to work to extend how long this takes, continually pushing classical computing response times (CCRT) to infinity while using them to gauge success.

Aiming for a quantum advantage

with the utility of our hardware

We've stated that quantum advantage will happen within the next two years — by which we mean that the process described in the preceding article will begin in earnest — but only if the HPC and quantum communities work together. This statement isn't made out of idle speculation.

We have always maintained that quantum advantage will be approached in two steps: first, by demonstrating the ability of existing devices to perform accurate computations at a scale beyond the scope of brute-force classical simulation (quantum utility); and second, by finding quantum solutions to problems that derive an advantage from quantum devices. Following the demonstration of step one, it is time to focus on step two.

As we stated in our definition of quantum advantage, we need the result of computations to be accurate. To make this more concrete, let's take a moment to look at the kinds of problems that we expect quantum computers will be able to solve. We classify them in two distinct buckets, each having pros and cons.

In one bucket, we have **sampling** algorithms that produce probability distributions from quantum circuits. According to our current understanding of computational complexity, these algorithms are well poised to solve problems that are difficult for classical computers — such as integer factorization through Shor's algorithm — but the apparent need for error correction represents a major challenge. In the case of integer factorization, it is straightforward to check whether a purported solution is accurate. We see a similar property in quantum circuits running the quantum approximate optimization algorithm (QAOA) for a combinatorial problem, or running sample-based quantum diagonalization (SQD) for a chemistry or materials science problem, in the sense that the solutions they provide can be classically checked. However, we expect the journey to quantum advantage for optimization methods like QAOA will be a fairly long one, since currently known classical optimization methods are comparatively strong. Newer methods like SQD seem to hold more promise for near-term quantum advantage.

Yet another task that falls into the sampling category is random circuit sampling (RCS), which was first proposed by Scott Aaronson[1]. RCS offers potentially good candidates for quantum advantage — but verifying the correctness of solutions to these problems is classically intractable at scale. Moreover, noisy implementations of random circuits have already been shown to break some of the complexity hardness assumptions that point toward an advantage[2]. A natural way to establish trust in such computations is to ensure that the method of producing the samples is free (or nearly free) of errors. To date, the only way we know to achieve this is through error correction, which is why several experimental demonstrations [3] claiming quantum advantage have, in our view, not achieved this result. These demonstrations fail to establish trust in the accuracy of the computation, which negates a fair comparison between quantum and classical methods.

New theory work[4] has proposed a variant of RCS based on random peaked circuits. Such circuits are designed to output a particular bit string with a non-negligible probability while appearing to be random in all other respects. Sampling problems based on random peaked circuits may lead to an efficiently verifiable quantum advantage, due to the fact that peakedness can be easily recognized by examining sampled bitstrings. However, recent work by IBM has shown that peakedness can also simplify classical simulation in

Aiming for a quantum advantage with the utility of our hardware continued

certain cases [5]. In addition, classical algorithms may be able to recognize peakedness directly without simulating the circuit. Overall, given the lack of efficient verification and unclear practical value, we do not believe that random circuit sampling is a fully satisfactory route toward quantum advantage.

In the second bucket, we have algorithms for **calculating expectation values of observables**, which is an essential task for understanding quantum systems. We now have tools, such as rigorous error mitigation methods, that provide noise-free estimates of expectation values on noisy devices with error bars [6] even in the absence of error correction. For shallow circuits, we can trust these calculations because the error bars only depend on the device noise and are independent of the actual circuit and observables [7]. The classical difficulty of these calculations depends on the specific circuit and observable; however, running such experiments on quantum hardware offers practical insights that are otherwise challenging to obtain classically.

But there's a snag: our method for calculating observables with noisy quantum computers, which we call error mitigation, has exponential cost. The exponential comes from **a metric called γ**, which is a property of learned noise in error mitigation. The time it takes to extract answers using error mitigation is proportional to γ raised to an exponent enumerating the scale of the problem.

What makes us feel confident about finding near-term quantum advantage with these observable calculations? Well, even though the cost of error mitigation is exponential, it is an extremely slow-growing exponential when error rates are low. It is important to point out that there are classical simulation methods that can simulate a large class of noisy circuits and observables in polynomial time[8]. However, if one takes a closer look, we see that these polynomials grow extremely quickly, so much so that the classical method becomes computationally intractable even for a small number of qubits. Moreover, the scaling gets worse and worse as our noise gets lower. A direct comparison for our system sizes shows that the slow exponential growth of our error mitigation methods is feasible in practice, while the classical polynomial method is not. It is therefore more important to compare our quantum computation to other classical simulation methods that don't rely on the noise in the system—such as tensor network methods, for example—to determine how challenging the simulation is classically.

That leaves us with today's hardware. In recent years, improvements in each new processor and revision have helped us bring γ down significantly. This progress has yielded a substantial improvement in problem runtimes, which ultimately allows us to run larger circuits. Hardware improvements come from advances such as our new tunable couplers, as well as recently introduced tools that mitigate the effects of a noise source called two-level systems and allow for the calculation of accurate observables from increasingly long circuits. We project further improvements in γ as we progress along our roadmap. As we drive our error rates down, we go further into the utility regime, where ever-improving noisy quantum computers with error mitigation offer our scientists tools to chase after advantage by running longer and longer circuits. By adding in the power of classical HPC, we also get new possibilities in terms of what we can do with error mitigation. We've already seen this in research conducted this year using tensor-network error mitigation (TEM) and multi-product formulas (MPF) for error mitigation of expectation values. With these advances, we feel confident that we'll soon reach a point where quantum methods will outperform existing classical-only methods.

[1] arXiv:1612.05903
[2] PRX Quantum 5, 030317 (2024)
[3] Nature. 2019 Oct;574(7779):505-10, Nature (https://www.nature.com/) volume 634, pages 328–333 (2024), arXiv:2406.02501
[4] arXiv:2404.14493
[5] arXiv:2309.08405
[6] Phys. Rev. Lett. 119, 180509 (2017)
[7] Nat. phys. 2023 Aug;19(8):1116-21.
[8] arXiv:2407.12768

Our vision for quantum-centric supercomputing

Today, we have supercomputers for solving problems beyond the limit of our PCs. But what about problems beyond the limit of those supercomputers? Well, for at least some of those problems, we hope that quantum computers will provide solutions. In that context, we expect that quantum processing units will become components in broader quantum-centric supercomputing architectures, serving as accelerators for CPUs working in a heterogeneous compute model.

We are building our hardware, software, and services with this future in mind. We have already presented our vision for the hardware piece of the puzzle with IBM Quantum System Two, the building block of quantum-centric supercomputing. IBM Quantum System Two anticipates multiple QPUs running quantum circuits in parallel with distributed classical computers. On the software side, we are pushing the performance of Qiskit so that it has the feature set and speed required to run alongside today's lightning-fast exascale supercomputers.

However, this isn't just a vision of the future, we are already seeing the results of CPUs aiding today's noisy quantum computers to tackle problems beyond the ability of either alone.

This year, we debuted a method called sample-based quantum diagonalization (SQD), a chemistry simulation technique for estimating the ground state energy of physical systems—and a first demonstration of the potential power of quantum-centric supercomputing. SQD uses the quantum computer to extract a noisy distribution of bitstrings, representing possible electronic configurations of a molecule. Then, it runs an iterative correction entirely on a classical computer to cluster these bitstrings around the correct electron configuration. This is an example of a sampling approach embedded into a larger workflow.

For a large enough noisy quantum computer with low enough error rates, this technique has the potential to perform more efficiently than classical approximation methods. RIKEN, Cleveland Clinic Foundation, and Lockheed Martin have already published promising results featuring this method. Meanwhile, RPI has implemented the first-ever demonstration of a live quantum-centric supercompter, scheduling jobs between the AiMOS supercomputer and on-site IBM Quantum System One.

This has always been our vision, but today, it's finally becoming a reality. What's more, we believe the supercomputing community is taking notice, as evidenced by Jay Gambetta's invitation to share this vision at Supercomputing 2024. So, let's keep working to realize a future where QPUs are an innate piece of any high-performance computer.

Accelerating research with Qiskit addons

This year we introduced **Qiskit addons**, a collection of research capabilities developed as software tools to help researchers design new algorithms at the utility scale. Qiskit addons are designed to be modular, meaning that users can integrate them into their code without having to restructure their pipelines, while also leveraging advanced capabilities demonstrated in previous research.

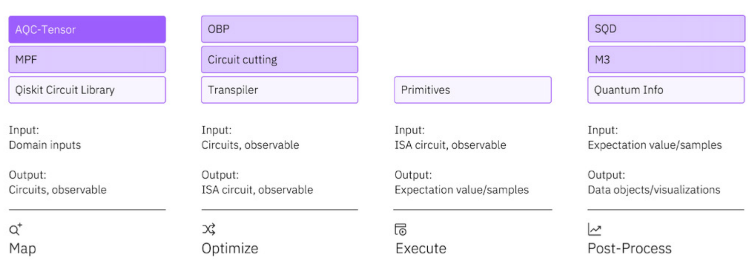
By stacking, modifying, and rearranging these capabilities on top of the foundation provided by the Qiskit SDK, users can more easily and quickly scale up their workloads. The first set of Qiskit addons we released include multi-product formulas (**MPF**), approximate quantum compilation (**AQC-Tensor**), operator backpropagation (**OBP**), and sample-based quantum diagonalization (**SQD**).

Qiskit addons are meant to be discovery accelerators—driving new research or contributing to the formation of new Qiskit Functions. Addons are power tools that are configurable to suit a specific task in a workflow, such as optimizing a circuit or post-processing samples.

Qiskit Functions, by contrast, are designed as abstracted services that wrap multiple steps of a quantum workflow into a high-level interface that you can use without in-depth knowledge of the internal components. This makes Qiskit Functions perfect for domain scientists who would like advanced research capabilities packaged in a way that manages hardware performance or enhances workflows in an application context—with no need for the user to interact with anything under the hood. Learn more in **this session** from the IBM Quantum Developer Conference.

For researchers looking to incorporate a Qiskit addon into an application function they are developing, we are creating function templates. These are starter guides that provide an example of a realistic workflow along with methods to set domain-level inputs and outputs, as well as approaches for evaluating the progress of a workload through logging or a dry run mode. Coming soon is a function template for Hamiltonian simulation using the AQC-Tensor Qiskit addon to map the problem description to a reduced-depth circuit for execution on hardware.

We're already seeing signs of how Qiskit addons are accelerating our clients' research. For example, SQD came out of research code from a paper we published with RIKEN and University of Colorado Boulder, a paper that took over a year of research. Using the SQD Qiskit addon, Cleveland Clinic was able to use the same method to **get results** showing the first quantum simulation of hydrocarbon following only 2.5 months of research—a nearly 5x improvement in turnaround time. We look forward to seeing how the broader community uses these power tools to unlock new utility-scale algorithms.



What is resource management?

Today, classical computing facilities are capable of running tremendously challenging tasks in the form of complex workflows thanks to workload management systems, which handle resource allocation and job scheduling. Workload management systems break a computational workflow into a series of tasks that are scheduled in a queue and assigned to computational nodes.

Realizing our vision for quantum-centric supercomputing requires us to bring quantum into the picture of workload management. This year, we set a goal to make quantum a first-class citizen in high-performance computing ecosystems, so that our users can more efficiently execute tasks based on the resource requirements of their hybrid quantum-classical workflows.

We're designing our system architectures to more easily integrate into existing high-performance computing environments. Our new hybrid resource management model facilitates tighter resource coupling, which creates the potential to run high throughput parallel workloads. It also enables smart resource utilization by separating compute into classical, typed quantum, and hybrid models. To integrate a new resource into our compute infrastructure, we have to expose resource control capabilities.

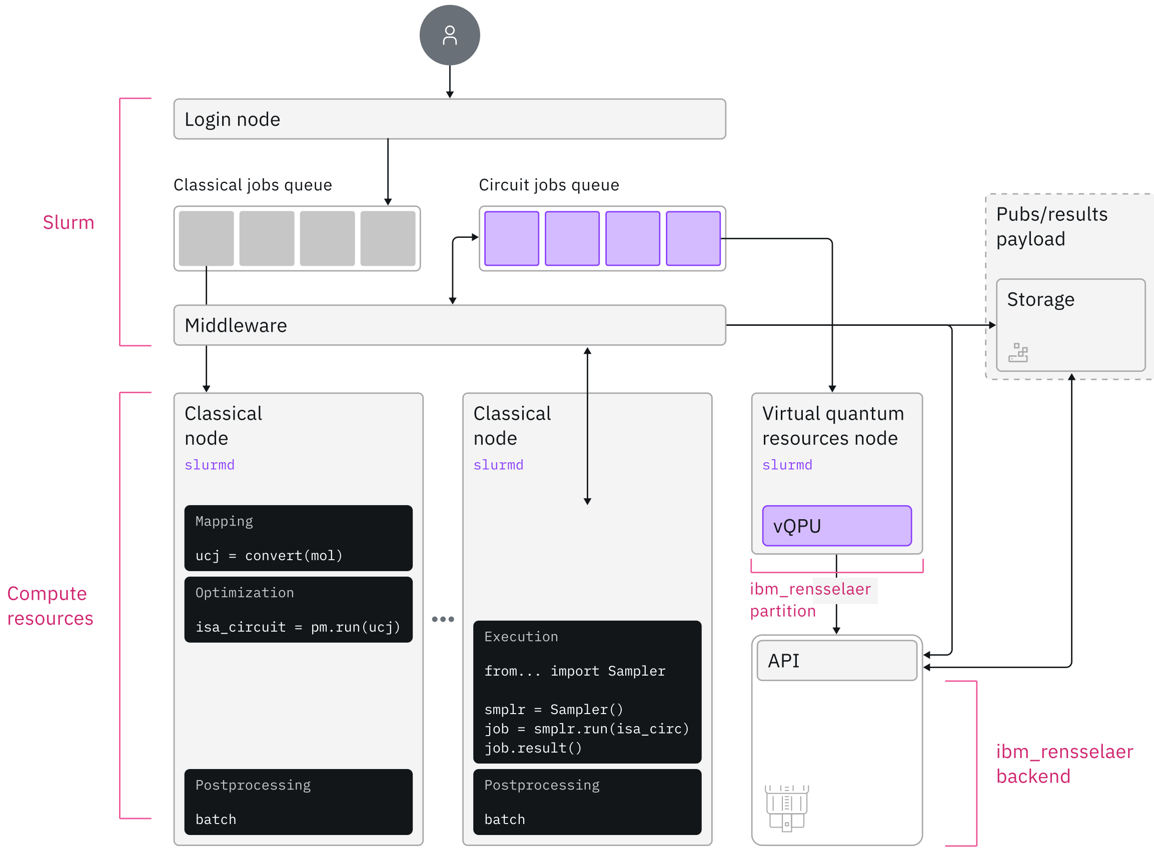
The first method we've used to achieve this is called **quantum cloud bursting**. In this model, the workload manager offloads tasks to IBM Quantum Platform via the channel's API. The second approach is a new, lower-level quantum system control API called **Direct Access**. This mode allows users to control the system state directly and leverage in-system parallelism to drive performance. To facilitate a seamless transition to workload management systems for Qiskit users, we developed a provider that uses the workload management system to execute Qiskit primitives.

We've been working with our partners at Rensselaer Polytechnic Institute (RPI) to establish the first quantum-centric supercomputing facility within a university setting. We connected RPI's Artificial Intelligence Multiprocessing Optimized System (AiMOS), a supercomputer with 504 IBM Power9® processors optimized for AI applications, and IBM Quantum System One into a singular computational environment managed by the Slurm Workload Manager.

For our first demonstration, we chose to run the sample-based quantum diagonalization (SQD) workflow, which uses quantum and distributed classical computing together in an attempt to produce good approximate solutions for chemistry problems beyond the reach of exact diagonalization methods. We wrote mapping, optimization, execution, and post-processing as job scripts and submission scripts for Slurm. This demonstration was the first truly heterogenous workflow in a fully realized quantum-centric supercomputing facility within a university setting.

Now, we have a path forward for bringing quantum into an HPC infrastructure without disrupting existing HPC users, who are critical to helping us realize a future of computing that combines CPUs + GPUs + QPUs to drive scientific discovery.

What is resource management? Continued →



A workflow diagram showing how we can use Slurm to run programs that incorporate both classical and quantum compute resources — a true demonstration of quantum-centric supercomputing.

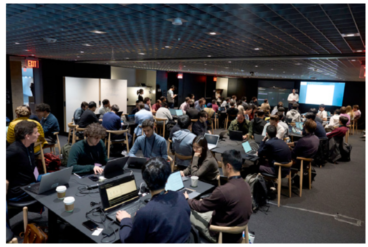
QDC recap

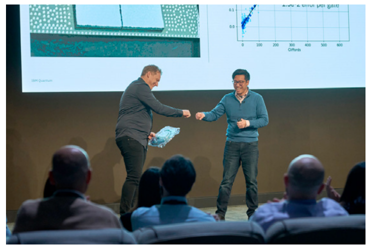
Last month, we hosted the inaugural IBM Quantum Developer Conference (QDC) at the Thomas J. Watson Research Center in Yorktown Heights, New York. There, 170 external quantum developers gathered to preview forthcoming updates to the IBM Quantum Roadmap and to get hands-on experience with state-of-the-art IBM Quantum software tools.

The agenda began with our annual State of the Union address, where we detailed updates to our development and innovation roadmaps, and where we announced that we'd successfully completed the 100×100 challenge by getting accurate results from a circuit with 5,000 gates. Three days of morning seminars followed. Day 1 covered tools for writing quantum circuits at the utility scale. Day 2 covered tools for efficient transpilation—i.e., running Qiskit code efficiently on quantum hardware. Day 3 covered tools for squeezing more accuracy out of quantum code.

Afternoon activities featured challenges targeted at both new and experienced users, as well as office hours, an activation village staffed by our startup partners, and lab tours. There were multiple opportunities for networking with researchers, engineers, other IBM experts, and the broader quantum community.

Overall, the event emphasized IBM's commitment to advancing quantum computing technology and fostering a collaborative ecosystem for developers and researchers alike. For those interested, all recorded sessions and materials are now available to the public via **IBM Quantum Learning**.





Scenes from the IBM Quantum Developer Conference

How we met our 100×100 commitment

We have met our commitment from the 100×100 Challenge: At the first-ever IBM Quantum™ Developer Conference (QDC) in November, IBM debuted a quantum computer capable of running accurate calculations employing circuits with 5,000 two-qubit gates—in other words, 100 qubit width by 100 gate deep circuits.

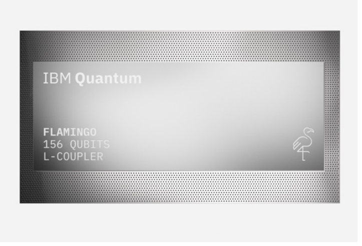
We achieved 100×100 with an IBM Quantum System Two powered by the second revision of the IBM Quantum Heron quantum processor. This upgraded Heron chip features 156 qubits in a heavy-hex layout, and preserves the tunable coupler architecture Heron r1 introduced in 2023 to suppress crosstalk. We also added new two-level system mitigation to help reduce the impact of an important source of noise. Two-level systems are essentially disturbances to the qubits interacting with the material surrounding them.

Over the course of the past year, we've made updates to the quantum system software stack, further optimized data movement, and introduced our latest generation runtime. We also introduced parametric compiling, so you only have to compile iterative circuits once if parameters are the only things changing between iterations. Thanks to these updates, we were able to announce speeds of over 150,000 circuit layer operations per second (CLOPS) at QDC. As we have continued to tune and roll out our gen3-runtime engine in the following weeks, we have already surpassed that benchmark and are now seeing machines with 180k to 230k CLOPS.

Three Herons are now available with gen3 for clients to try: torino, fez, and marrakesh. More are on the way. We will continue to tune gen3 as we move it into general production, so expect the universe of the possible with Heron r2 and Qiskit to continue to expand.

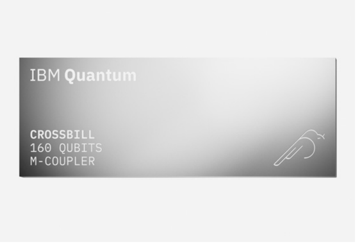
Five facts about the new interconnected QPUs

There are limits to the number of qubits that can be packed reliably onto a single quantum chip. When the scale grows too large, the problem of wiring all those qubits in a practical way also becomes challenging. In 2022, IBM proposed a solution to this scaling problem: QPUs composed of multiple individual chips, linked together with long- and short-range quantum interconnects. In 2024, we realized these technology demonstrations on our innovation roadmap with IBM Quantum Flamingo and IBM Quantum Crossbill. Here are five key facts about each of these interconnected QPUs.



IBM Quantum Flamingo

- For the first time, we demonstrate long-distance l-couplers.
- Flamingo brings together two Heron-style chips with four connectors measuring up to a meter long.
- We have demonstrated excitation transfer between qubits on two distant chips leading towards CNOT gates.
- We have previously measured CNOT operations at 3.5% errors per gate on a prior test device.
- A scaled-up Flamingo (with up to 3x the number of interconnected chips, and technology to go up to 7x over the long term) is on track for late 2025.



IBM Quantum Crossbill

- This is the first demonstration of edge-to-edge m-couplers.
- Crossbill comprises three connected Heron-style chips on a common interposer with direct integrated connectors.
- A total of 8 interchip m-coupler connections link the three Herons.
- There are 548 couplers across the device.
- The full Crossbill device includes >1,000 quantum elements with just 1/5 the area of a full Condor circuit board.

Google magic state cultivation paper

What it is: In this paper, Google researchers prepared low-distance magic states using Clifford measurements and a substantial amount of post-selection. The result is a complete T state preparation protocol with lower space-time volume.

Why it matters: The paper presents a solid optimization (and rebranding) of existing magic state preparation methods, where these methods are required for full-scale quantum computing. Escape from the low-distance codes into well-protected codes has been implicit in these types of protocols, but this is the first work to carefully design and evaluate this important step as part of an end-to-end analysis of the protocol.

Quantinuum development roadmap

What it is: Quantinuum announced accelerated roadmap projections at this year's Quantum World Congress. The roadmap states that Quantinuum plans to scale to ~1000 physical qubits and 1e-4 error rates by the end of the decade. They intend to encode qubits into codes with a roughly 2:1 physical to logical ratio and achieve logical error rates of 1e-5.

Why it matters: We are eager to see Quantinuum's progress on this front and agree that they are doing good work. However, while the stated plan to increase scale is consistent with the trajectory of the work in their published papers, achieving lower logical error rates is a different problem that they do not address. They indicate a target range of 1e-10 logical error rate and cite two theoretical papers, including one from IBM, but there are no further details about the proposed architecture that will take them from the 1e-5 scale to the 1e-10 scale.

E.ON case study

What it is: IBM and E.ON published a joint study on the potential for quantum computing to help refine the prediction and modeling techniques used to optimize critical energy routing for the 40 million customers E.ON serves across Europe.

Why it matters: Efficiently routing critical power supplies around the world is a highly complex process that involves sophisticated risk management. Global shifts to a wide range of renewable energy sources add further complexity as providers attempt to accurately anticipate dynamic global energy consumption patterns. E.ON is one of the world's largest investor-owned electric utility companies, and they continue to be highly engaged partners of IBM Quantum.

EU data center

What it is: Europe's first IBM Quantum Data Center is up and running, so users can execute workloads on quantum computers based in Europe. The European Quantum Data Center features two quantum computers with the 127-qubit IBM Quantum Eagle chip, *ibm_strasbourg* and *ibm_brussels*. A third based on the 156-qubit IBM Quantum Heron chip, called *ibm_aachen*, will join the fleet soon.

Why it matters: An IBM Quantum Data Center on European soil is a commitment from IBM to the European quantum computing community, and one that addresses some of our European partners' constraints in terms of processing proximity and regulatory compliance. With this new facility, it is easier for IBM and our European partners to work together, explore near-term business value, and build long-term partnerships.

Jay at SC24

What it is: Dr. Jay Gambetta delivered a Keynote address and outlined a vision for the future of quantum-centric supercomputing in his first appearance at the International Conference for High Performance Computing, Networking, Storage, and Analysis (SC24).

Why it matters: An invited talk on quantum computing demonstrates our continued leadership in the field, and makes clear that the HPC community is taking an interest in the subject. Jay used this important forum to publicly predict that quantum advantage is achievable within the next two years if the high-performance computing and quantum communities work together. The keynote offered an opportunity to share our vision for when and where we expect to find quantum advantage with a huge audience of industry insiders, peers, and competitors.

A new Multiverse offering for the Qiskit Functions Catalog

What it is: Multiverse Computing has joined the group of collaborators working with IBM to expand the recently announced Qiskit Functions Catalog. Multiverse's Singularity is a machine learning classification tool that uses quantum machine learning to develop supervised learning techniques for the research community.

Why it matters: This new third-party function expands the availability of quantum machine learning techniques for members of the IBM Quantum Network. All premium users can now request a free trial. We hope continued collaboration of this sort and the continued growth of the Qiskit Functions Catalog will accelerate quantum application development across the community.

CVaR paper

What it is: A new paper demonstrates that calculating an important function called the conditional value at risk (CVaR) is robust against the noise inherent to our current quantum computers. CVaR tells an investor the average amount of money they can expect to lose when markets slide, but is useful as a loss function for other algorithms.

Why it matters: This work is interesting because it offers noise-free information about the outputs of a quantum computer that we can then use in our quantum optimization algorithms. Where other error mitigation methods like PEC and ZNE are computationally expensive and provide exact expectation values, CVaR is computationally cheap and provides boundaries on expectation values, rather than an exact output. It is an important result that will further motivate quantum optimization algorithm discovery.

LOCC paper in Nature

What it is: A new paper from IBM researchers demonstrates novel techniques to establish a real-time classical link between two 127-qubit Eagle devices.

Why it matters: As IBM has worked to scale superconducting quantum processors beyond the 100-qubit and 1,000-qubit barriers, it has become increasingly clear that a multi-QPU approach may be needed to help quantum computing achieve its true potential. This paper represents an essential milestone on the journey toward that goal. It is the first demonstration of two connected QPUs working together to execute a quantum circuit beyond the capabilities of either alone. To get there, the researchers behind the paper made important advances in circuit cutting, error mitigation for dynamic circuits, and simulation methods for quantum states with periodic boundary conditions—which appear widely in nature and are useful for error correction and other quantum computing applications.

Optimization white paper

What it is: A new white paper from representatives of the Quantum Optimization Working Group offers an updated look at the prospects and challenges for quantum computing in the field. The publication includes proposals from IBM, MIT, Los Alamos National Lab, HSBC, and other industry leaders on ways the community can work together to accelerate research.

Why it matters: Optimization problems are everywhere, and they are among the hardest we ask our classical computers to tackle. The working group's white paper provides the first ever timeline for how and where we expect quantum resources to help find inroads into these infamously challenging problem spaces.

Goodbye Zapata

What it is: Zapata is no more. The quantum computing company turned generative-AI venture announced its closure just a year after its first public offering.

Why it matters: Clarity of value proposition is important. Muddled focus risks a lack of market confidence. Investors ultimately perceived a company unable to compete with major players in the generative AI space, and similarly too lightweight to offer meaningful value as a quantum computing firm.

Quantum Networking

Scaling error mitigation with Algorithmiq's TEM Qiskit Function

As part of the newly released Qiskit Functions Catalog, Algorithmiq has made available an out-of-the-box error mitigation solution called Tensor-Network Error Mitigation (TEM). The TEM Qiskit Function is a hybrid quantum-classical algorithm that enables users to compute the expectation values of observables while mitigating noise from the quantum hardware. By offloading the burden of mitigation onto classical compute resources, TEM helps users perform error mitigation with less QPU runtime and increased accuracy—critical ingredients for scaling to problems in the utility regime. [1, 2]

As Matteo Rossi, CTO and co-founder of Algorithmiq, describes it: "When you scale to circuits with 50+ qubits and many layers, it can become very heavy in terms of extra QPU runtime—potentially from hours to days." TEM uses tensor networks together with a proprietary measurement technique based on informationally complete measurements to combat this exponential overhead, even allowing multiple observables to be estimated with the same run on a quantum computer.

TEM follows a quantum-centric supercomputing paradigm that leverages the synergies between quantum computing and tensor networks, a classical computational tool first developed to describe quantum systems in an approximate way. Using the QPU to prepare states that are hard to reproduce on classical systems and then using the CPU and/or GPU to perform tensor network computations to post-process the results, TEM reduces the quantum runtime overhead by up to 1000x compared to probabilistic error cancellation (PEC) or zero noise extrapolation with probabilistic error amplification (ZNE-PEA). [2] TEM also uses the same interface as the Estimator primitive in Qiskit Runtime, making it easy to integrate into Qiskit workloads while also providing customization opportunities to enhance performance for particular use cases.

Applications using TEM in areas such as condensed matter physics, chemistry, and optimization are

[1] S. Filippov, M. Leahy, M. A. C. Rossi, and G. García-Pérez, **arXiv:2307.11740** (2023).

[2] S. N. Filippov, S. Maniscalco, and G. García-Pérez, **arXiv:2403.13542** (2024).

Quantum Networking continued

Scaling error mitigation with Algorithmiq's TEM Qiskit Function

already beginning to show promise to scale into regimes where analytical solutions do not exist. In November, Algorithmiq published results in collaboration with IBM and Trinity College Dublin of a utility-scale demonstration with circuits of up to 91 qubits and 91 layers of two-qubit gates—the equivalent of over 4,000 two-qubit gates. [3]

For this research, the team used different parameters to simulate the chaotic dynamics of dual unitary circuits representing a kicked Ising model, an important use case in condensed matter physics. They were able not only to benchmark the results of the quantum computer against analytical results to verify the output, but also to press into the utility regime and obtain accurate results beyond the scope of brute-force numerical simulations.

"The next step is to push all of this even further," says Matteo. "Since we ran this on the Eagle processor, we look forward to using Heron and hopefully exceeding the 5,000-gate milestone that was recently reached by IBM."

Algorithmiq is also partnering with Cleveland Clinic to use TEM for simulating biologically relevant systems to improve photodynamic therapy. They anticipate additional industrially relevant applications will emerge as TEM enables new users to try error mitigated workflows at a scale of 50+ qubits.

IBM Quantum Premium Plan users can access a free trial of TEM and request a license through the **Qiskit Functions Catalog**.

[3] L. E. Fischer et al., **arXiv:2411.00765** (2024).

Q&A:
Sergiy Zhuk

How would you describe your role?

I am a senior research scientist and manager of the quantum computing team at the IBM Research Lab in Dublin. There is a lot of collaboration and brainstorming, and I don't always know where my work will lead.

How did you end up in this role?

My background is in applied math, namely optimization for dynamical systems. I joined IBM Quantum from Exploratory Math Sciences (XMS) where I worked on parameter and state estimation for dynamical systems—from Markov decision processes to Navier-Stokes equations. The key challenges I faced in my work were nonlinearity and curse of dimensionality.

Back then, I got excited to explore if quantum computers could help with resolving at least one of those challenges. I was lucky to get excellent mentors in IBM Quantum, which made my transition a natural thing to do.

Can you tell us about your work with BasQ?

In September of 2023, my team began a collaboration with DIPC @ BasQ modeling time crystals on Eagle and Heron QPUs using multi-product formulas (MPFs). It started as a purely exploratory project, primarily concerned with convergence analysis and uncertainty quantification, and then, as a result of team effort, it culminated in a powerful computational algorithm for time evolution: dynamic MPF (now a Qiskit addon). We use this now to explore phase diagrams of time crystal models with our BasQ partners.

This work required a diverse team with expertise in Trotter-based time evolution, tensor networks, device error mitigation, and condensed matter physics, as well as a considerable amount of technical coordination. This has led to two papers, and now we are close to finished analyzing a final set of experiments on 145 qubits on *ibm_fez*.

What do you do outside of work?

I spend time with my family. I have two young children. I love riding a road bike, and I do a bit of mountain biking with my son in Wicklow National Park. We hike all together on a sunny day.

Phase Kickback

Quantum computers have entered a new phase in their history, one in which they are finally beginning to demonstrate their usefulness for scientific exploration. As we move further into this “utility” era, it is vital that the quantum research and developer communities devise new quantum algorithms that take full advantage of our increasingly powerful quantum hardware and software stacks. The quantum algorithms we develop today may very well be the first to yield practical demonstrations of quantum advantage.

So, how do we develop new quantum algorithms? One potentially fruitful approach is to look for useful heuristics—i.e., shortcuts, educated guesses, or general “rules of thumb” that enable both classical and quantum algorithms to quickly generate high-quality solutions to problems that may be intractable or impractical for exhaustive, brute-force methods.

Humans have used heuristics to solve problems since the dawn of time. We use them every time we make a “common sense” judgment or a prediction based on some prior experience. The history of heuristic algorithms in computer science goes just as deep, stretching back to the earliest days of computation.

Today, heuristic algorithms make up many of the most effective and widely studied algorithms in both classical and quantum computing, and there’s good reason to believe that heuristics will continue to be important well into the future. In this article, we take a look at some seminal moments in classical and quantum heuristics to get a sense of how heuristic methods have evolved, and where they may be headed next.

Some of the earliest heuristic algorithms took the form of “genetic algorithms,” which use the logic of evolution to identify useful solutions to search and optimization problems. A genetic algorithm might frame potential solutions to a problem as members of a population, and evaluate those solutions using a fitness function that selects for the best or “fittest” members. It creates new members of the population through evolutionary concepts like “mutation” and “crossover,” gradually culling lesser solutions through multiple rounds—or “generations”—of analysis. **The earliest explorations of genetic algorithms began in the 1950s, with one of the more formal demonstrations being published by physicist Richard M. Friedberg in 1958.**

While some researchers spent the late 1950s investigating the value of specific heuristic methods, others were looking for ways to generalize them into machine learning (ML) algorithms capable of developing useful heuristics of their own. **One of the most influential examples of these efforts was published in 1961 by MIT researcher Marvin Minsky, who was among the first to develop a formal approach to the ML method known as reinforcement learning (RL).** An RL algorithm is one that teaches itself how to perform some task based on a training system where good solutions are “rewarded” and bad solutions are penalized. In many cases, this system leads to the RL algorithm developing its own heuristic functions. RL algorithms are used widely in computing today, including in quantum computing, where it is used for tasks like circuit synthesis and for quantum code generation models like the one that powers Qiskit Code Assistant.

Phase Kickback

Another important development in classical heuristic algorithms arrived in 1968 with the publication of the A* (pronounced “A-star”) algorithm. A* is a search algorithm that finds an optimal path between some starting node and some goal node in a graph when given a suitable heuristic. It does this in part by using this problem-specific heuristic to estimate the cost of traveling from any given node to the ultimate goal node. This allows it to quickly prune entire regions of the solution space by discarding partial solutions that cannot lead to the optimal path, ultimately speeding up the overall computation. Even today, the A* algorithm continues to be widely influential in search and optimization algorithm design.

In 1982, computer scientist Judea Pearl proposed his belief propagation algorithm, an efficient heuristic algorithm originally designed to make predictions or inferences in graphical models. It does this by passing messages between connected pieces of data, such as nodes in a network, to update what each piece "believes" is true about the relationships between the data. Researchers eventually realized that this message-passing strategy could be generalized beyond graphical models, serving as a useful heuristic for tasks like error correction. Today, it is used widely in image processing, recommendation systems, and both classical and quantum error-correcting codes to make predictions and solve problems where direct answers are hard to compute.

A year later, **in 1983, IBM researchers in Yorktown Heights published their simulated annealing algorithm**, a physics-inspired optimization method that would go on to become widely influential in quantum heuristic algorithm design. All combinatorial optimization algorithms are designed to maximize some objective function that characterizes the quality of potential solutions, but physics-inspired methods like simulated annealing often reframe this as an energy minimization problem, an equivalent formulation—and one widely useful in quantum computation—where each solution is assigned an energy and the goal is to find the lowest energy.

Inspired by the cooling and solidification of metals, simulated annealing takes an initial trial solution and uses a method known as Metropolis sampling to generate new solution configurations corresponding to the Boltzmann distribution, a fundamental probability distribution in physics. Metropolis sampling accepts or rejects new solution configurations based on (1) the energy difference between the new proposed solution and the initial guess, and (2) the temperature, a parameter that determines the probability that the algorithm will accept a new higher-energy solution.

Simulated annealing efficiently searches the solution space by gradually decreasing the temperature as it runs, allowing it to escape local energy traps in the solution space, while also making it increasingly likely that new solutions will yield lower energies as the runtime progresses. This method would be recast in quantum terms by one of the very first quantum heuristic algorithms, **the adiabatic quantum optimization algorithm published in 2001 by researchers at MIT and Northeastern University.**

Phase Kickback

A decade later, in 2011, then-Ph.D. student Kristan Temme would serve as lead author on a paper demonstrating a quantum version of the Metropolis sampling method previously used in the 1983 simulated annealing paper. This posed a unique challenge as classical Metropolis sampling presumes that you can take some trial solution state, evolve it to a different state, and if necessary, go back to a copy of the original state. That's easy to do with a classical state, but quantum states cannot be copied—that's the "no cloning theorem" from quantum mechanics. In their 2011 paper, Temme and his colleagues overcome this problem by developing a method that determines whether to accept or reject a new state while revealing as little information about it as possible, inspired by an idea from Marriott and Watrous's result on QMA amplification.

In 2014, we get perhaps two of the most important and most widely studied quantum heuristic algorithms in the history of the field—the quantum approximate optimization algorithm (QAOA) and the variational quantum eigensolver (VQE). QAOA, first demonstrated by the lead authors of the 2001 adiabatic quantum optimization algorithm, is in many ways a refinement of their 2001 effort. In adiabatic quantum optimization, you vary between two Hamiltonians in a fixed way, with the heuristic being that if you vary too fast, the approach will not work. The QAOA algorithm allows more freedom. The VQE algorithm is also in some ways a descendent and refinement of the adiabatic quantum optimization algorithm that allows for more freedom than its predecessor. In the VQE method, we attempt to find the eigenvalue of a matrix by first running an "ansatz" circuit—essentially a guess at the target state—and then running that circuit iteratively and using a classical optimizer to continuously refine its parameters until it converges on the desired solution.

In 2016, researchers from Lawrence Berkeley National Laboratory and UC Berkeley demonstrated a now-widely used heuristic method known as quantum subspace expansion. This method combines quantum and classical computation to solve problems where we know the desired solution is hiding in some extremely large solution space. We use the quantum computer to take random samples from the solution space, and then use those samples to restrict the search space to some useful subspace. From there we use a classical computer to find the optimal solution. The method is considered a heuristic because a randomly selected subspace may not contain the true global minimum for the entire solution space, but generally provides at least a very good one.

Just this past year, in 2024, researchers at Google, Stanford, MIT, and Caltech demonstrate a new method known as decoded quantum interferometry (DQI). DQI is a novel approach to solving optimization problems that uses a quantum Fourier transform to arrange quantum amplitude interference constructively on solutions for which the energy is small, thereby enhancing the probability of obtaining good solutions. It leverages the fact that the Fourier spectrum (the optimization landscape after the quantum Fourier transform) possesses certain desirable properties such as sparsity. This opens up the possibility of exploiting novel structures that go beyond those explored by quantum algorithms so far.

Phase Kickback

Seminal works

1958
A Learning Machine: Part I ↗, Friedberg, R. M.

1961
Steps Toward Artificial Intelligence ↗, Minsky, M.

1968
A Formal Basis for the Heuristic Determination of Minimum Cost Paths ↗, Hart, P., et al.

1982
Reverend Bayes on Inference Engines: A Distributed Hierarchical Approach ↗, Pearl, J.

1983
Optimization by Simulated Annealing ↗, Kirkpatrick, S., et al.

2001
A Quantum Adiabatic Evolution Algorithm Applied to Random Instances of an NP-Complete Problem ↗, Farhi, E., et al.

2011
Quantum Metropolis sampling ↗, Temme, K., et al.

Phase Kickback

Seminal works continued

2014

A Quantum Approximate Optimization Algorithm, Farhi, E., et al.

2014

A Variational Eigenvalue Solver on a Photonic Quantum Processor, Peruzzo, A., et al.

2016

Hybrid Quantum-Classical Hierarchy for Mitigation of Decoherence and Determination of Excited States, McClean, J. R., et al.

2024

Optimization by Decoded Quantum Interferometry, Jordan, S., et al.

Upcoming Events

For more information on upcoming events, contact **qiskit.events@us.ibm.com**.

23-24 January
High Energy Physics Working Group Meetup
Geneva, Switzerland

4-5 February
Opening Ceremony of the International Year of Quantum Science and Technology (IYQ)
Paris, France

24-28 February
Quantum Information Processing (QIP) Conference
North Carolina, US

10-13 March
SCAsia 25
Singapore

11 March
Convergence Quantum Forum
Massachusetts, US

16-21 March
APS Global Physics Summit
California, US

23-27 March
ACS Spring
California, US