Opinions Libres
le blog d'Olivier Ezratty

Decode Quantum with Chris Langer from Quantinuum

Welcome to the 80th episode of the Decode Quantum podcast series. We are this time with Chris Langer from Quantinuum. This is one of the leading companies using trapped ions for the creation of quantum computers. Unfortunately, Fanny Bouton was not with me in this podcast, moderating a panel somewhere in Paris simultaneously with this recording.

Our guest Chris Langer is a Fellow at Quantinuum. He is their chair of the Technical Board and an advisor to the President and COO. Chris has a PhD in physics obtained in 2006 at the University of Colorado on trapped ions quantum computing under the supervision of David Wineland, Nobel laureate in physics in 2012, who was a **guest of this very podcast** in **July 2024**. Chris is a key contributor to the development of Quantinuum's hardware. Before joining Quantinuum (which was then initially Honeywell Quantum Systems) in 2016, Chris worked for 10 years at Lockheed Martin on various projects in the aerospace domain.

Here is an edited and illustrated transcript of our discussion.

Olivier Ezratty: how did you land in the quantum space?

Chris Langer: After graduating college, I was looking for where to do research in graduate school. And as an undergraduate leaving college, often you don't really know what field you want to go into. And I toured a handful of different universities to where to go to graduate school. I visited David Wineland's lab, and it was mesmerizing and caught my attention. I called Chris Monroe who was working for David Wineland and said, "*I really like what you're doing here. It's very exciting. I want to come and work for you. And if you will let me in, then I will come to the University of Colorado. And if not, I'll probably go somewhere else.*" He kindly said, "*let me look into your your application.*" Shortly thereafter, he called me back and said, "*yes, we'd love to have you come.*" That was the beginning for me. I started working for Chris Monroe and Dave Wineland starting from day one, when I went to graduate school. That's how I got into quantum. And then I just worked on it during my PhD with some amazing scientists, David Wineland and Chris Monroe and other students and postdocs in the group. It was a fascinating time. That was a great time in grad school. At Lockheed Martin, I pursued many things. And eventually at Honeywell Quantum Systems, they decided in 2016 that they really needed to invest in this technology. They recruited me and others. That was the shadow time when Honeywell Quantum Solutions was starting. We had a large investment sponsored by the CEO. For several years, we started developing hardware in a somewhat secret and quiet manner. In around 2019-2020, we released our first product and the activity became public. That's how the business at Honeywell got started. We merged with Cambridge Quantum Computing in 2021 and became a full-stack company. The merger doubled the company size. It was a great combination because CQC (Cambridge Quantum Computing) had a strong expertise on compilers, algorithms and software. They were doing projects with a variety of different customers, developing algorithms for their specific use cases and needs. And we were focused on the hardware. We were building machines based on trapped ions, and we had ideas on how to scale and how to get better.

Olivier Ezratty: Do you remember the title of your thesis.

Chris Langer: It was a "high-fidelity quantum memory". I was trapping beryllium ions at the time, and we showed a coherence time of 15 seconds for a single beryllium qubit.

High Fidelity Quantum Information Processing with Trapped Ions by Christopher Langer, 2006 (239 pages).

Olivier Ezratty: I presume you wrote a lot of papers back then, because usually when you do a PhD, you're going to do a lot of research.

Chris Langer: I was co-author on probably about 30 papers at the time. The group was very active during that time. My time there was 1999 to 2006. It was an exciting time to be there because 1999 was when our group demonstrated the first two-qubit gate using the Mölmer-Sorensen technique, that almost all trapped ions quantum computing groups use now to perform their gates in one form or another. Launching that gate really did enable us at the time to dive into doing some neat small algorithms demonstrations. We did quantum teleportation, quantum dense coding, a quantum Fourier transform, a lot of these kind of key demonstrations the

NIST group that I was a part of at the time, created in the early 2000s. It's an exciting time to be there.

Olivier Ezratty: What's happening when your PhD supervisor becomes a Nobel laureate? Isn't that crazy to some extent?

Chris Langer: It was, it's very exciting. I was very excited for David and I think it was well deserved. He was an icon, not just in quantum computing, but also in atomic physics and in the late seventies he was working with trapped electrons and the G minus two (g-2) experiments. David Wineland was the first to implement ions laser cooling in 1978. Cold atoms laser cooling came a little bit later. He's a great scientist, and I really enjoyed working for him.

Olivier Ezratty: Do you remember when you did your PhD, what were the kind of plans and forecasts for creating a functional quantum computer? I remember there was a report from 2004 about that, which was pretty visionary. I mean, when you look at the challenges, it's more or less the same as the ones we have right now. So how do you look at the way things have changed in 20 years.

A Quantum Information Science and Technology Roadmap Part 1: Quantum Computation, Advanced Research and Development Activity (ARDA), April 2004 (268 pages).

LA-UR-04-1778

A Quantum Information Science and Technology Roadmap

Part 1: Quantum Computation

Report of the
Quantum Information Science and Technology
Experts Panel

"… it seems that the laws of physics present no barrier to reducing the size of computers until bits are the size of atoms, and quantum behavior holds sway."

Richard P. Feynman (1985)

Disclaimer:
The opinions expressed in this document are those of the Technology Experts Panel members and are subject to change. They should not to be taken to indicate in any way an official position of U.S. Government sponsors of this research.

April 2, 2004
Version 2.0

ARDA

This document is available electronically at: http://qist.lanl.gov

Chris Langer: They've changed quite a bit. I think one of the key ways that they've changed was about a decade ago, I think, industry really got excited. Let's go back to the history. In the mid-'80s, Feynman says "*I wish I had a quantum computer to simulate chemistry*". The mid-'90s, and then it just kind of stayed there. In the mid-'90s, Peter Shor developed his integer factoring algorithm. That was really the genesis of a lot of

experimental groups trying to actually go build one quantum computer. That was the time when the ion groups from Chris Monroe and Dave Wineland in 1995 did the first Cirac-Zoller gate. I. Cirac and P. Zoller came up with a neat quantum gate scheme. In 1995, Chris Monroe and David Wineland demonstrated it. That kind of kickstarted the effort to try to build a quantum computer in academic settings. I think the academic setting was 95 to about 2015. Around that time, IBM got a really strong presence in the 2010s. They wanted their quantum ecosystem to be out there. They developed Qiskit. They started putting machines out there and really building a large ecosystem. And a lot of people started using those machines. And then a decade later, I think now machines are finally getting good enough that people start thinking about error correcting them and scaling them and finding useful applications. And, of course, the application. search has been happening really since day one.

Olivier Ezratty: My question was more related to the speed of progress because some people feel like there's some kind of acceleration right now. It's 20 years in a row. Do you feel there's such an acceleration? I don't know how you measure speed and acceleration in that space, but the unit is not clear.

Chris Langer: I do. And I think the enabler for that is demonstrations of error correction at the break-even or better level. We realize that if you don't have good primitive operations like high fidelity gates, then you can't error correct operations. Building larger machines of poor quality gates and poor quality qubits just doesn't work. 2024 was a breakthrough year for us and others as well, where we were able to show beyond break-even quantum error correction. This is where you finally get the acceleration. Once you can get beyond break-even, then you can error correct with better codes and suppress the error even further. And now you're in a position to scale. And I think that's where the acceleration comes in.

Olivier Ezratty: I asked David Wineland in July 2024 the question on how do you grade the challenges between pure scientific and fundamental science on one hand and practical and engineering challenges to develop those large-scale utility-scale quantum computers on the other hand? Where are we in between?

Chris Langer: I think we're in the technology development era. The science is well understood. And I think we've been there for a while. We just need better components and better technologies in order to scale. The fundamental science is not limiting, at least in the ion space.

Olivier Ezratty: Can we discover new phenomenons when we scale? We don't know how large-scale entanglement is going to work. We may discover new sources of noise. Do you think that things are entirely settled there?

Chris Langer: I think there is some unknown. And I know from a practical perspective, we often think about what's the technology that's limiting us. And we do at the small-scale limit, right, just a couple of qubits and gates, we have a very good model of the physics of how that works. But you are right. You could run into new things that won't be discovered until you get to those lower noise floor. I think one example of this is Google just had this recent quantum error correction experiment. It's a great experiment. a great paper. They did use a

repetition code to go to very high distance and see the floor. They saw this 10^{-10} logical error floor. They speculated that the cause of it was cosmic rays. That's a great example, I think, of what you're talking about. You're finding new fundamental physics because you broke past some technological barrier. Those things are still out in front of us. But I'm optimistic that if and when we find them, we'll overcome them and continue to go down.

Olivier Ezratty: Let's come back to Quantinuum, at the stage of when it was HQS, eight years ago. You mentioned that there was the kind of push from the CEO of the company. Why was that? I mean, why did the CEO was willing to invest in quantum? It's very rare in the high-tech space.

Chris Langer: It is rare, and it was beautiful. The main reason was because he just saw some really great value in the future. Back then, and still today, it's a high risk, high reward bet, where if you are successful, you can solve some amazing problems that can really, really change the world. This is a few years ago, but one of the key use case people were pursuing was about simulating nitrogenase (FeMoCo). This is effectively trying to find new processes to make fertilizer, and fertilizer construction or manufacturing takes up about 3% of the world's natural gas supply. These are some statistics that a lot of people have been talking about. And that's so if you can be successful in that, you get enormous savings and global impact. So, the CEO wanted to invest. It was some executives and some physicists who approached him and said, hey, we want to do this. And would you invest in us to do this? And the CEO said, "*I like it so much, I want you to come back and give me an even more aggressive plan*". And so they did that. And that was nice to have a sponsor and a vision like that. He wanted us to think forward 10 years. And then coming back with a plan that would achieve your goals in 10 years. That was David M. Cote. I give him a lot of credit for having that vision and sponsoring us to do this great work.

Olivier Ezratty: Honeywell is a bit special in the quantum world because when you look at the large companies that invest in that space, it's mostly IT companies like IBM, Google, Microsoft, and Amazon. Honeywell is more an engineering company. It's a very different business structure. And this is the only case I know of, at least in the U.S.

Chris Langer: There were good reasons for it, too. We had good alignment with the technology that Honeywell was experts at and the quantum space. And you might think, oh, but Honeywell didn't have expertise in quantum computing at the time. But what they did have was expertise in controls. They had expertise in chip manufacturing. And so even at that 2016 timeframe, Honeywell had already been making ion-trapped chips for other academic partners. They were already well-positioned to make some of the key building blocks, both the ion traps and the control systems, to be successful in building a quantum computer based on ions.

Olivier Ezratty: it makes a lot of sense. It means that it was kind of an extension of the existing business to some respect. You are kind of, as we say, you were an enabling technology vendor, which kind of moved up the stack.

Chris Langer: That's right. From an academic perspective, academics know a lot about how to make good qubits and the fundamental noise sources, but they really can't get their hands on sophisticated, custom-made chips, and they have to rely on partners to do that, and Honeywell was a partner for that. Honeywell effectively had some good discriminating and enabling technology.

Olivier Ezratty: We had a couple of guests in the trapped ions business. We had **Sebastian Weidt from Universal Quantum**, and we had the guys who founded **Crystal Quantum Computing** in France a while ago. We still have to come back on what makes ions special for computing. Just go on the basics, and then we're going to dig into some details on the technology. So why ions?
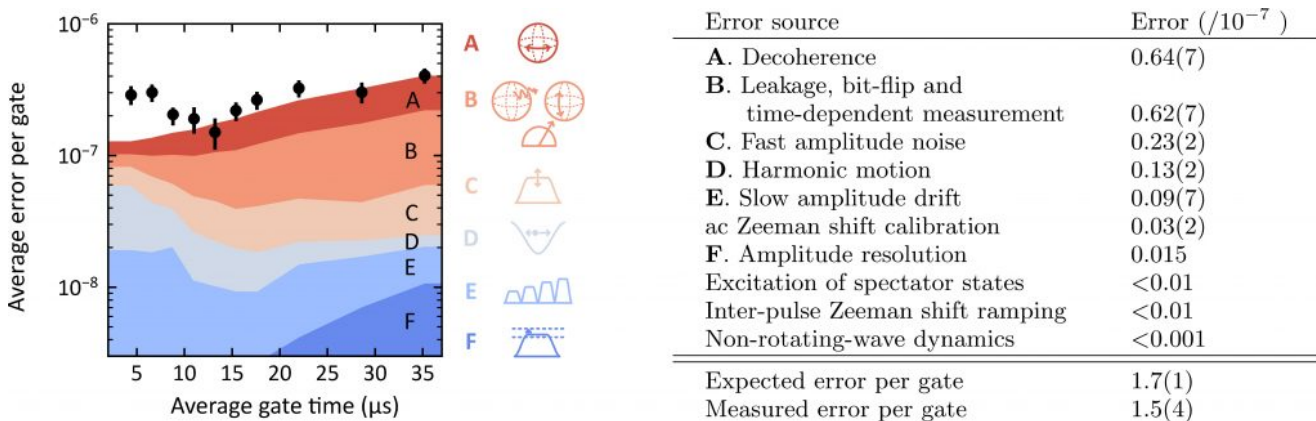
Chris Langer: What ions bring to the table is that they're strong with really high-fidelity operations. There's a history for this. Ions and atoms alike have been used as atomic clocks for decades, and they're really good atomic clocks because their parameters, like their states, are insensitive to the environment to a high degree. Therefore, they make good precision timing sources. It just so happens that a really good clock is a really good qubit. The properties that you're exploiting for clocks just directly translate over into good qubits. The physics of the interactions that you need to perform these quantum gates are very well understood. They've been studied probably for about 100 years.

You can say if I measure some error on a quantum gate, for example, you can model why. You might say, "Oh,

I think maybe my laser beam has too much noise on it, and that's causing my error." And you can model that and predict it, then go measure the noise on the laser and verify that your models are correct, and then you know where to work. So you say, "Okay, I'm going to reduce the noise on my laser." This iterative process is what researchers do and have been doing for decades. And they just continue to get improvements in the clock

space. I mean, clocks, several even over a decade or two ago, were hitting accuracy at one part in 10^{18}. This is unheard of. There's no other physical system where you can get that level of precision. I remember at the time, because David Wineland had a clock group as well, and we ate lunch with these people and talked all the time. They were doing some modeling, and double-precision numbers on the computer weren't good enough to capture the precision of what they were trying to do. And so they had to use higher-precision numerical libraries just to do the modeling. This is unheard of in other fields of physics. That's the key reason why ions make good qubits.

Olivier Ezratty: I saw this kind of chart in a recent paper from Oxford University. They realized recently, with a couple of people from Japan, seven nines single-gate fidelity on one ion. It was really amazing. They were showing all the sources of noise and what they could improve, and so on. So do you think it's special to ions? I would like to have this kind of chart for every qubit modality. I don't see that for superconducting qubits because it's more complicated. You have resonators and Jaynes–Cummings models behind that. Is that the reason, or is there some other engineering reason which explains why you don't have the same discipline in other modalities?

Single-qubit gates with errors at the 10^{-7} level by M. C. Smith, A. D. Leu, K. Miyanishi, M. F. Gely, and D. M. Lucas, Osaka University and University of Oxford, arXiv, December 2024 (17 pages).

Error source	Error ($/10^{-7}$)
A. Decoherence	0.64(7)
B. Leakage, bit-flip and time-dependent measurement	0.62(7)
C. Fast amplitude noise	0.23(2)
D. Harmonic motion	0.13(2)
E. Slow amplitude drift	0.09(7)
ac Zeeman shift calibration	0.03(2)
F. Amplitude resolution	0.015
Excitation of spectator states	<0.01
Inter-pulse Zeeman shift ramping	<0.01
Non-rotating-wave dynamics	<0.001
Expected error per gate	1.7(1)
Measured error per gate	1.5(4)

Chris Langer: I think other modalities have a similar level of discipline in terms of doing precision measurements, but the physical system just has more perturbations. In solid state, when you embed a quantum particle in the solid state, it has this environment around it, which is Avogadro's number of atoms right next to it. And there are impurities in those systems, and those all cause noise. That's, I think, the main reason why you can't get this level of precision. Neutral atoms have similar potential, because they're also isolated from the environment. And when we hold ions, we hold them in a vacuum. There is no solid-state environment nearby. There is still an environment, magnetic fields, for example, but it's engineered, and the noise sources are far away. They're not a nanometer away or a tenth of a nanometer away. They're hundreds of microns away. As a consequence, you get this natural isolation from the environment.

There is one key distinction between ions and atoms. In an ion, it's a charged particle. You've ripped off an electron and it's positively charged, so you hold it with electric fields that couple to that charge. That charge is really highly decoupled from the internal states that you actually use to form the qubit. In neutral atoms, it's different. To hold on to a neutral atom, because it's not charged, you have to use other forces like dipole forces from a laser beam, which has to perturb the internal states just to keep the atom in place. That's a level of

perturbation that can affect the gates and memory, which is different than ions. And that's one advantage ions have.

Olivier Ezratty: Even though we use different wavelengths—when you control cold atoms, you have at least six to seven different wavelengths for the tweezers, then for the gates, the single-qubit gates, the Rydberg state gates, and readout. But this is not enough to isolate the way you control the atoms? That's what you would say?

Chris Langer: They have really good techniques and tricks, and this is why they've been really, really successful. These 10^{-18} clocks were with neutral atoms as well. You don't want to say they don't have ways to get around it. One particular technique, though it's a bit technical, is using what's called a magic wavelength, so the perturbations that apply to the lower state also apply to the upper state in the same way, effectively decoupling the system.

Olivier Ezratty: That may explain the difference we still have right now in capabilities, which is much better in ions as far as I know. In lab records or in industry vendor records, there's at least an order of magnitude of difference.

Chris Langer: This is true. You can chart this over three decades. Fidelities across different technologies have all made improvements, but ions have consistently been about a factor of ten lower in errors. Our commercial system today hits three nines or better, and we guarantee it, running 24/7. Neutral-atom systems come in around 99.5% and 99.6% in some cases.

Olivier Ezratty: It's not far, but it's exponential. Just moving from 99.5% to 99.8% or 99.84%, I think you have on 56 qubits, seems small, but it's a big difference.

Chris Langer: Yes, yes.

Olivier Ezratty: At least when you want to deal with the depth of your circuit. So we all know that ions still have their own challenges. Maybe it would be interesting to discuss that. You've got gate times, scaling, the circuits. I'd like to get some insight from you on how you manage the way to drive the ions. Because when I look at the circuits, I have right here the racetrack H2 chip you have. You use everything—lasers, microwave, RF signals—everything in one place. That's a long list of questions I'd like to dig into. But these are the challenges. How do you cope with that?

Chris Langer: There are many subsystems that have to come together. Maybe we could start with how you actually hold on to the qubits themselves. This is the ion trap. The way we do this is we fabricate these traps using regular CMOS foundry processes, what they call the backside processes, where you lay down routing metal. We don't necessarily need, or haven't used yet, the CMOS portion where you dope the semiconductor. Rather, we just route metal, put electrodes patterned on top, and drive them with voltages. Those electrodes trap the ions above the surface.

You're effectively making a bucket for your ions or qubits right above the chip surface, and that bucket is defined by the voltages you apply to those electrodes. If you apply different voltages, that bucket moves, which is how we do transport. We slowly modify the voltages, and the ions move with the well. We have what we call transport primitives. We can move qubits left and right around the racetrack (in H2). We can also split two qubits apart if they're in one well into two separate wells. We have a merge operation, which is the reverse of the split. In H2 and H1, we have a swap operation, where two qubits in a line can be physically rotated to swap positions.

Because we hold the ions using their charge, the internal qubit states remain unaffected by these movements. We check that by preparing a superposition state and performing splits, merges, swaps, and then measuring.

The error is no worse than if the qubit just sat idle. So, transport doesn't induce extra errors beyond the usual memory error.

Once the trap is working, with good testing procedures, we can check that box and focus on the next piece: basic quantum operations. First, do we have good quantum memory? We prepare a superposition state, measure its coherence, and compare it with noise models. We find states that are insensitive to magnetic fields, making them robust to environmental noise, just like in atomic clocks. Then we move on to single-qubit and two-qubit gates. Most of the noise sources are technical and well-understood. There is a fundamental source of error called spontaneous emission. One thing we do differently than some ion groups is that we use lasers for our gates. Other groups sometimes use microwaves, but lasers are more common.

So that's where we stand. Because the physics is well understood, we can measure and model each source of noise, then mitigate it step by step.

Olivier Ezratty: But you use lasers for cooling as well and for qubit readout. How do you guide the laser beam on the ions? Do you use AODs, AOMs, SLMs like in cold-atom systems, or is it something different? So lasers, AOMs to control the light, to focus on the right atoms or ions. Because it's also based on lasers and fluorescence. By the way, is it a CCD or CMOS imaging detector that is being used?

I was wondering about that because I worked in AI before working in quantum. So I was wondering if you do any image recognition or if you just count photons. You have a fixed place where you measure the photons, so presumably you don't do image recognition for the qubit state.

Chris Langer: We use lasers for cooling as well. In our case, all the operations that we perform use lasers, minus transport, which does not. But everything else that acts on the qubit itself, including loading, uses lasers.

To load the trap, you've got this ion trap chip with electrodes on it, and you need to get ions in there. We start with a cold neutral atom source and flow those neutral atoms across the loading zone. Then we hit them with a laser beam that photo-ionizes each atom, stripping off an electron and leaving a positively charged ion. Once it's charged, the trap's electric fields grab onto it and hold it in place. At that point, it's usually hot, so we must laser-cool it with a separate laser. As you noted, we use lasers for loading, cooling, gates, and readout.

Regarding how we direct laser beams onto the ions, in today's systems we have optical setups outside the vacuum chamber. The beams go through a window and across the chip, parallel to its surface. It's a complicated arrangement because of the many different colors and beams we need. We do use AOMs all the time. They're great modulators. They can switch the beam quickly, and they give frequency and phase control. They're common in both ion and neutral-atom experiments.

For readout, the goal is to distinguish between two states. Even if the qubit is in a superposition, it collapses into one of these states upon measurement. We use a method called state-dependent resonance fluorescence. One of the states, when excited by a laser beam, scatters many photons in a cycle of excitation and decay. We collect these photons with a lens, send them to a detector, and if the ion is in this "bright" state, we see a flood of photons. If it's in the "dark" state, the laser is off-resonance and we see very few photons. For detectors, we currently use photomultiplier tubes, but you can also use avalanche photodiodes or CMOS cameras. Cameras are slower to read out, but they become attractive if you have many ions to measure in parallel.

We don't do image recognition; it's more like "bucket" detection. If you have a big array, you might match each qubit's position to a pixel or group of pixels. A bright spot means a qubit in the "one" state; a dark spot means "zero.".

Olivier Ezratty: the ion-trap vendor landscape is a bit particular because you're probably the only ones who

communicate a lot on the so-called SPAM metric—state preparation and measurement. Is that because you've got good numbers, or is there some other reason you focus on it? I remember when you started four years ago, you highlighted SPAM a lot. I'm curious about that.

Chris Langer: we talk about it because it's important for customers to know all the relevant numbers for predicting circuit-level fidelity and overall performance. It's also about keeping ourselves honest: if some error is 10 times bigger than others, we must fix it. Typically, one error source stands out. In our system, two-qubit gate errors are the largest. In some other platforms, measurement errors dominate. From a theory standpoint, you can sometimes mitigate measurement errors if your two-qubit gates are excellent, by measuring the same qubit multiple times with ancilla qubits via CNOT gates and voting on the result. But that only works if the gate fidelity is high enough. That's one reason we emphasize two-qubit gate error.

Olivier Ezratty: Could we say that SPAM was of moderate interest in the NISQ regime, but becomes more interesting when you implement quantum error correction? You need to do a lot of syndrome measurements. So if you want to have as fewer errors as possible, you need to do syndrome measurements.

Chris Langer: In the NISQ era, all of your fidelities are poor. And the error correction is great milestone because it really requires, multiple things to come together all at the same time and work well. You need to have high-fidelity gates. You need high-fidelity measurement. And to exploit quantum error correction in a fault-tolerant context, you need to perform mid-circuit measurement. You have to perform feed-forward and teleportation.

So you need to get everything working. And I think the groups that are working on building quantum computers work on their hardest thing at a time. They try to perfect that, and then something else pops up and is a little bit harder to work on, and then they perfect that, and eventually the whole system gets better. I expect that all the groups that are working on these problems will work on these measurement problems. But right now, it is true that our measurements are good, and you are right. Not all the companies tell you what their measurement error is.

Olivier Ezratty: you use dual species with barium and ytterbium. Can you explain why? I don't know many companies would do that with ions, even with cold atoms. There are some plans with cold atoms to use two species for various reasons.

Chris Langer: We do have two species. The reason is called sympathetic cooling that is needed to get good gates. The gates themselves are affected by a noise source based on the starting temperature of the atoms or the ions for that matter. You laser cool the ions first and then you perform your gates. If you're doing a "hero" experiment and you're only doing one gate, well, you can just cool the qubits themselves, and you don't need a second species. And so a lot of academic groups will do this, but if you want to perform a quantum computation and you want to perform gate after gate, after gate, after. gate. And even if you're moving qubits in the middle, well, they're going to heat up. They heat up for several reasons. One kind of more fundamental reason is there's just natural noise in the environment. Even if you were to leave the ions in place, they will heat up slowly. We work to mitigate this. That's one of the reasons we cool our traps to 20 to 40 Kelvin or so. But the act of moving them around, you don't always move them around perfectly. You can imagine this is like a particle on a spring, and you're going to yank on the spring and move it from left to right. If you don't slow it down perfectly, well, it's going to be oscillating a little bit. You must remove that energy. And the sympathetic cooling process is the way to do that.

You could cool the qubit itself but doing that destroys the quantum information. Instead, we co-trap a different species in H1/H2 that's barium. We cool barium atoms with a different laser wavelength than how we control ytterbium qubits. The barium laser is completely transparent to ytterbium; ytterbium doesn't even see the laser.

Because the ions are charged, it's like there's a spring between the barium and the ytterbium, and if you can just damp the motion on the barium all the ytterbium will damp along with it. So that's the sympathetic cooling process. The reason you need is because you need to be able to perform gate after gate. You want to be able to cool the ytterbium without destroying the quantum information. This is critical. This is a key element that you have to have in ion trap-based quantum systems. If groups don't have it yet, then they need to put it in.

Olivier Ezratty: It looks like with cold atoms, dual species may be used for QNDs (non-demolition measurement), more than cooling. Let's come back to noise. There are two sources of noise I've heard about. I'm not sure whether they are important or not. One is laser phase noise, and the other one is voltage noise to move the ions. How is that important? How are you mitigating that noise?
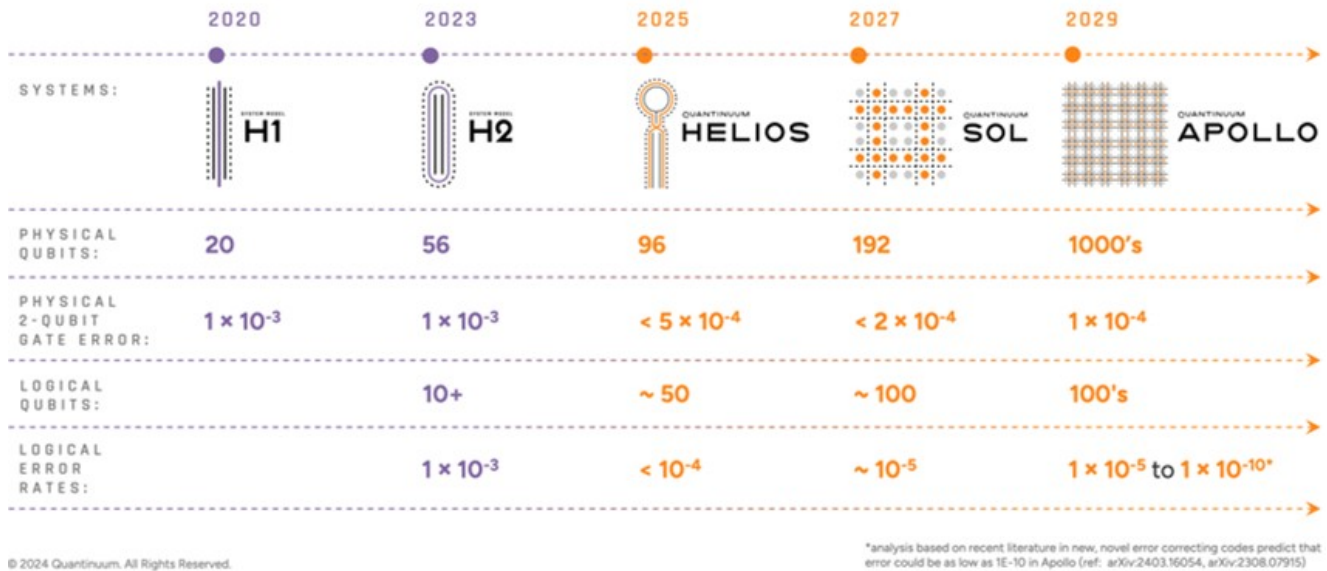
Chris Langer: Yes, it is important. Let's take laser phase noise. You have two laser beams that are going to run over your pair of qubits, and you're going to create interference patterns. These interference patterns are not fixed in time. Lasers use very slightly different frequencies and so you kind of have this beat note. We need to stabilize that in some level. Fortunately for the ions, the way we do the gates is that interference pattern only needs to be stable over the duration of the gate. It doesn't have to be stable between gates. You can kind of have a slightly different starting point of that interference pattern on gate one versus gate two doesn't have to be slightly, it could be completely different actually could be off by 1.4 ?, for example, and the gate will still function just fine. But it does need to be stable over that short period of the gate. Now, fortunately, the stability of laser beams is good enough to achieve that. Now, voltage noise is actually very similar.

You can imagine if you have like the phase of the laser, kind of it's noisy and the ion is fixed perfectly in space. That looks like a kind of a jitter of the phase profile that's kind of going over the ions in an unwanted way. Well, imagine now you move to the frame of the ion, and you have voltage noise. That voltage noise is going to create forces on the ion and it's going to move the ion. In the frame of the ion, if it is kind of jittering around in physical space, it looks like the laser phases are jittering around over the ion.
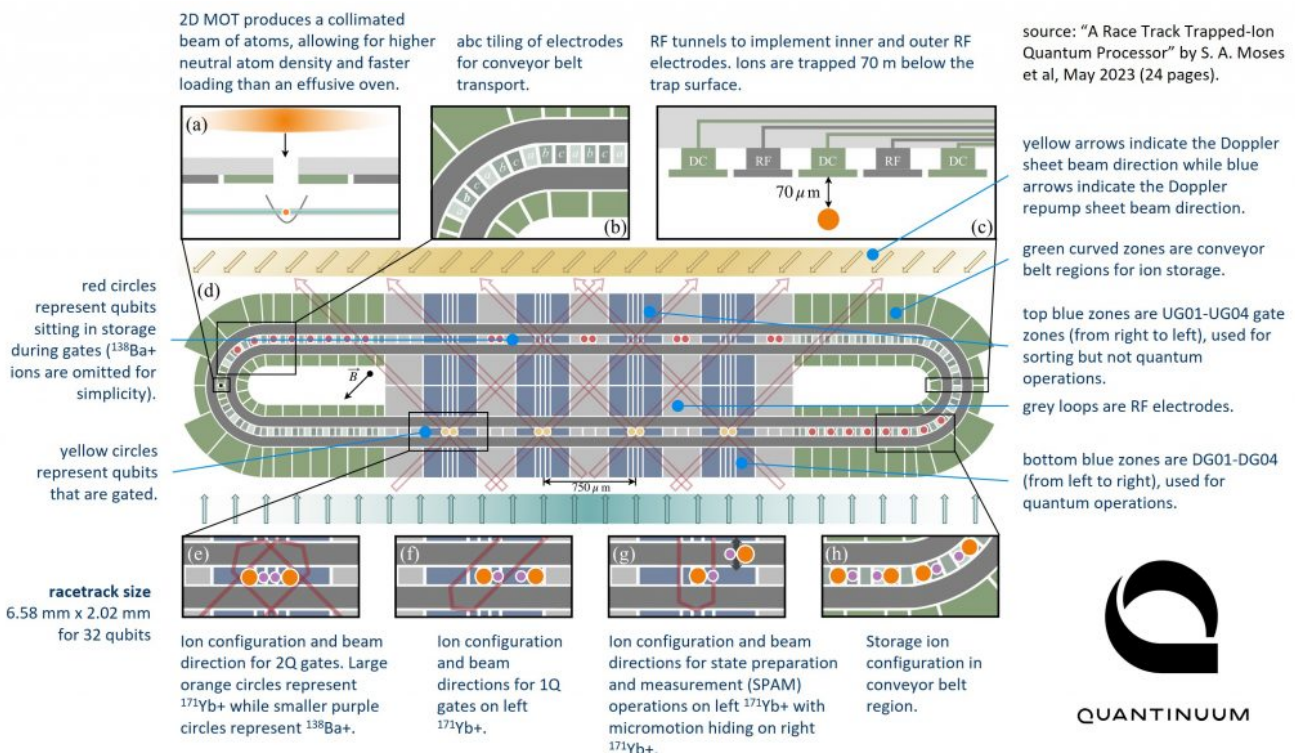
Quantinuum H2 racetrack ion trap. Source: **A Race Track Trapped-Ion Quantum Processor** *by S. A. Moses et al, May 2023 (24 pages). Legends added by Olivier Ezratty in 2023.*

Olivier Ezratty: let's move Quantinuum's roadmap. I'd like to understand the rationale between the various kinds of circuits you have. Initially, you had a 1D array of ions back in 2020 when you released the H1 system. Then you went with the racetrack two years ago with up to 56 qubits that you have right now in H2. Soon you're going to have the Helios version which looks like a small chromosome, and then a 2D architecture. Can you explain this path? Why did you choose the racetrack architecture? Why the Helios architecture? And I understand the 2D one, which is the way to scale.

	2020	2023	2025	2027	2029
SYSTEMS:	H1	H2	QUANTINUUM HELIOS	QUANTINUUM SOL	QUANTINUUM APOLLO
PHYSICAL QUBITS:	20	56	96	192	1000's
PHYSICAL 2-QUBIT GATE ERROR:	1×10^{-3}	1×10^{-3}	$< 5 \times 10^{-4}$	$< 2 \times 10^{-4}$	1×10^{-4}
LOGICAL QUBITS:		10+	~ 50	~ 100	100's
LOGICAL ERROR RATES:		1×10^{-3}	$< 10^{-4}$	~ 10^{-5}	1×10^{-5} to 1×10^{-10}*

© 2024 Quantinuum. All Rights Reserved.

*analysis based on recent literature in new, novel error correcting codes predict that error could be as low as 1E-10 in Apollo (ref: arXiv:2403.16054, arXiv:2308.07915)

Quantinuum roadmap announced in 2024.

Chris Langer: The high-level answer is really crawl, walk, run. Realize that the 2D architecture is our launching point for scaling, and that's the future for our company. And so the question is, why didn't we do that first? And the answer is because we wanted to do something simpler first. The linear trap is the simplest thing you can do, and that's the H1 system. That's why we chose H1. Maybe a good question really is, why did we choose H2, and why did we choose Helios? And H2 is just a larger qubit variant of H1.

It's still a one-dimensional architecture, but it loops around on itself. You can see some benefits there. How fast would it take you to transport an ion from one place to another? You can kind of go clockwise around the circle versus counterclockwise. The H2 system allowed us to put more qubits in there. One key element of technology that was inserted into H2 was the idea of broadcasting the voltages to electrodes. And so in H1, if you were to count up how many electrodes do you need to hold one zone, one qubit, it's about 10 or 20. And so if you say, I'm going to scale up to a million qubits, that means I have 10 or 20 million electrodes. Controlling all of this individually would be very, very complicated. We want to overcome that with broadcasting. The idea is to make an array of common zones and then take electrodes of common type across the zones and tie them together and control them together with a common voltage source. We do this in H2. Around the corners, we tie electrodes, every third electrode, together. And this reduces the voltage sources that we need. The main objective of H2 was to get more qubits. Their gate zones were kind of laser beam compatible with H1. Like the locations of them and the spacing of them were compatible. It looks very similar.

Olivier Ezratty : It looks very similar. It's funny because you don't have any switch when you have a racetrack. It's a bit like when you're a kid and you have your racetrack with small cars or trains. Yeah. It starts to become difficult when you have a switch. connecting different lines. So that's what happens in the future. So that's where it's going to be more difficult, I presume.

Chris Langer: So thinking of putting ourselves in the place of H2, we know we want to go to two dimensions. Two dimensions has that switch you're talking about. We call it a junction. It's like a crisscross, a two street crossings, for example, with the stoplight in the middle. Helios puts one of those junctions in place. The Helios layout has this storage ring. It looks like a circle on one side coming together at a junction.

And then you have two linear streets coming out of that junction. And those linear streets are backward compatible with the laser beam geometry of both H2 and H1. We can kind of leverage common subsystems like the laser beams, etc. But we leveraged the junction. One thing that occurred, one thing that we were able to develop and prove out prior to the Helios decision to launch was we did do a lot of junction transport. We had two really good experiments in the junction transport in our test beds.

The first that we showed what that we could take both barium and ytterbium together in a common well and transport it through a junction and saw no heating. And that was the key element for us that says, okay, we can now launch a commercial product with a junction in it. And let's, again, let's do crawl, walk, run. Let's do one junction, this commercial system, rather than a hundred, which is going to be our, our Sol product, the two-dimensional product that's coming out in a couple of years.

So Helios has one junction and it's been a beautiful success. And because we have this tight storage ring, we have higher density storage. It operates probably two to three times faster than H2, for twice the number of qubits.

Olivier Ezratty: You have not officially announced any characterization on Helios as far as I know.

Chris Langer: We have announced a little bit. We have a public roadmap now. We have a public roadmap with some specs on it, with 96 qubits. It will have a $5{\times}10^{-4}$ error two qubit gate errors. And it's releasing this year in 2025. So those are the things you can look forward to.

Olivier Ezratty; It's amazing in the roadmaps of most vendors this year, most of them plan to both increase the number of qubits and the fidelities. You plan to have a four nine fidelity in four years, 2029, with thousands of qubits. It's based on everything you discussed? I mean, how do you control the noise and how do you mitigate every source of noise? But it's a challenge to do both.

Chris Langer: It is a challenge. I think all the companies are very similar here where the first big objective is kind of suppress the noise and the second is to scale. And then you want to kind of merge these two a little bit. Again, common across all the companies. I think everybody sees a practical floor to their two-qubit gate error.

That floor that we just think it's maybe diminishing returns to try to like push past that. For us, that floor is 10^{-4}.

A lot of the previous roadmaps for some companies were 10^{-3}, but I think people are starting to lean forward a little bit and try to get lower. I think we have high confidence that we'll be able to get there. We've always been able to hit the highest fidelities in the market.

Olivier Ezratty: when looking at such roadmaps, as an engineer, I always wonder about the cycle. How long does it take to design a new circuit, to manufacture it, to characterize it, and so on? How far is it a problem? I mean, this cycle, and how can you speed up the rate of progress? Because it's slow. It's slower even for silicon qubits. It's in between with superconducting qubits. Okay, so it's CMOS, but you still have to design and manufacture it. So it takes time.

Chris Langer: It does. And we have these conversations in-house all the time. How do we go faster? You try to identify the key critical elements in your schedule and figure out creative ways to solve these problems. And these are active things that we're looking at all the time. Our commercial release schedule is one new system every two years, which does seem reasonable. People get on these machines and use them and develop

applications. And our customers have been very happy. Of course, we do want to go faster, and we'll continue to make those levels of improvements.

Olivier Ezratty: It works if you don't have to iterate. Because, OK, if you have a new circuit every two years, it's OK. But if you have to iterate within the circuit generation, you lose a lot of time. It depends on the modality, I presume. But if in your case, you've got a good CMOS right away and just you have to test it, it's fine.

Chris Langer: What you're describing is kind of a serial development path. That's very easy to think of, but we have parallel paths where we're already designing our 2029 system where you have deep concept designs and analysis for systems that come after that and different groups pursuing all of these things for us, we have, we've identified specific technology that needs to be matured. And we have dedicated teams and testbeds going to mature those things. And we estimate when you can on-ramp them into particular architectures and things like that.

Olivier Ezratty: How far do you think you can scale in a monolithic way? I mean, on a single circuit, I saw in the roadmap that Apollo in about five years would have thousands of physical qubits. There's a kind of limit somewhere. How far can you go?

Chris Langer: So Apollo is thousands of qubits, like you mentioned, that is a single chip design. We actually expect we can probably push that at least another order of magnitude. Okay. So tens of thousands of qubits on a single chip, that is certainly possible. And we actually don't see any fundamental reason why you can't actually scale to a full wafer size. If you do the math, you actually find that with our qubit densities, we can get millions of qubits on a single wafer. If you could fabricate and the techniques are there, and this is mature, right? You can do die stitching and stamp out a design that actually executes on a single wafer. The issue that we would worry about in that case is yield. Our logical approach is to first scale the single die to a large reasonable size. And we think we have at least an order of magnitude to go past Apollo, so tens of thousands of qubits. At that point, we're going to use tiled chip arrays.

Once you do that, you transport across the boundary, the idea is you take two chips and they're adjacent maybe there's a small gap of 10 microns or something like that, no problem, air gap, vacuum gap. The electric fields don't see this gap. They are continuous in space, and you can take that well, that bucket that you created with these electric fields, and just move it right across that boundary. This has been demonstrated already, not by us, but by some academic groups. This is very feasible. That's our approach. We're going to tile the plane with these kind of tens of thousands of qubit chips.

Olivier Ezratty: I presume you count also on the high qubit fidelities to be in a position to generate a large number of logic qubits given fidelities with a single monolithic architecture.

Chris Langer: Even in Apollo with thousands of qubits, we expect to get 100 logical qubits out of that device, and it's a single chip. So you can imagine if you use the same code and you had 10 times more qubits, you'd get 1,000 logical qubits in that particular device. Now, we see that the QEC space is still in development, and different codes have different pros and cons. One thing that we are finding, however, in this space is that there's a humongous benefit to having all-to-all connectivity. So if you can connect any qubit with any other qubit, every code that you can conceive of is possible to execute on our particular device. It just so happens that high-connectivity codes have higher encoding rates, have higher distances, enable you to do many, many other things. For example, you can perform transversal gates. You can do single-shot error correction. These are things that having high-connectivity devices enables you to do.

In this Apollo device, when we say we have a hundred logical qubits, it is a universal physical machine with

thousands of qubits. Any quantum error correction code you can conceive of, if you come up with even higher rate code, maybe you get a few hundred logical qubits on that particular device. Or maybe you have 80 logical qubits at 10^{-10} infidelity. And then of course scaling that Apollo chip by a factor of 10 you get 10 times more logical qubits for the same code. And then of course, tiling it, you really can go to millions or well beyond that.

Olivier Ezratty: Are the savings on QEC in circuit depths offsetting the slow gate speed, compared to other qubit modalities? I presume the answer is yes, but how far? Is it just balanced or not?

Chris Langer: I think that's actually helping a lot. One of the key tools in our toolbox that we're really, really realizing right now is really this idea of space-time trade-offs. The idea is if you have more qubits, you can do things faster. This is not a new concept. This has been around in classical computing for a really long time. You consider that even a couple decades ago, like, clock speeds in classical computing kind of peaked out at the few gigahertz level, and people were worried that Moore's Law was going to slow down or end.

But really what happened was instead of scaling faster, these compute nodes would scale out. They would add more cores and able to get more and more computational power. And the same is true in quantum computing. You can trade qubits for time. I'll give you a couple examples. One is an easy example to understand. This is like naive parallelism. A lot of… quantum circuits that you want to run, you want to run multiple shots, right? If you want to run 100 circuit shots, let's say this is like executing the same circuit 100 times and performing statistics on the output to get your final answer. If you have 100 times more qubits, you could run 100 instances at the same time. That's naïve parallelism. At the algorithm layer, there are tricks you can do to widen the circuit to shorten its depth. This is another technique that you can do. I would say for the hardware development team, myself included, we actually try to perform these very similar techniques at the very, very physical level in order to trade qubits for time. But one more technique I'll add for you, I just think is really useful, is the idea of single shot error correction. Let's first examine how long does it take, to perform an error correction cycle using the most common code known that people talk about, the surface code.

In the surface code, you perform a depth four circuit. You perform four CNOTs. You initialize, perform four CNOTs and a measurement to extract one syndrome, to extract the syndromes, one round of syndrome extraction. That must be repeated about 10 times. The code distance level of times in order to kind of overcome measurement error. And so right away, if you have a distance, a distance 9 or 13 code, you may be repeating this thing 30 to 50 times. If you can perform single shot error correction, that's an immediate savings to you. And the all-off connectivity of our machines enables you to do that. So, space-time trade-offs, single shot error correction and several other techniques in our toolbox can improve speed.

Olivier Ezratty: 2024 was a busy year for most vendors on the logical qubit space. You, you demonstrated a lot of advancements, particularly with the H2 system, but there was one caveat there, because I think Craig Gidney was the one who pinpointed that on X, I think that you are using a lot of post-selection for running this high fidelity logic qubits, which is supposed to not scale. I presume it's an intermediate step. How will you get rid of that post-selection or post-rejection?

Chris Langer: Post-selection used alone does not scale, but post-selection can be a valuable tool to kind of go above and beyond to a modest extent beyond a regular error correcting code. To scale, you do need higher distance codes. You need to go beyond distance-2. And some of the work that we did with Microsoft as well use distance-4 codes. The benefit here is you can error correct one fault with distance-3 or higher, and then after, but you can actually look for two faults. And if you find two faults, you can now discard. And the probability of two faults is lower than the probability of one fault. And if you just had a distance three code, eventually you're going to have undetectable errors, undetectable two faults that you wouldn't even know occurred and therefore you'd run this computation and think you had the right answer, but you actually had the wrong answer. So I think post-selection can be used to give you some modest wins after you've done the error

correction. That's one direct angle that we see. And the second is, it is acceptable to do some post-selection when you're preparing resource states.

Olivier Ezratty: Particularly when you do magic state distillation as well.

Chris Langer: That's right. You say, okay, I need to prepare these resource states. It could be a magic state. It could be a Bell state. It could be other states that you need for a variety of purposes, error correction is one. You just need to prepare that state. And if you have a, if you prepare and then verify that you actually have it without destroying it, then it goes into the bank, the bank of useful states I can draw on later, right? I have inventory of useful known verified states. And so you, now, you, now you think of it as architecturally, I have a factory, I have a factory that's going to make these states. This factory is going to fail sometimes. And, but I'm going to know that I failed. I'm going to reject those states.

Olivier Ezratty: It's decoupled from the circuit itself. That's maybe the reason why.

Chris Langer: It's exactly right. Because now if your factory took a really long time to make these states, then if you say I'm going to post select on success, well, my post selection just might be so low, like one out of 10,000 that now that doesn't look like it scales, but it just so happens that a lot of these factory circuits aren't very long. Like they're, they're, they're not that big at all. You can actually just tolerate a little bit of discard rate to prepare those states and put them in the factory.

Olivier Ezratty: you use a very interesting benchmark that was not very popular a while ago, which is a GHZ state. These are ways to characterize larger entangled states. Can you explain why you guys do that and how it's valuing the capability to create those large entangled states? And that was one of the recent papers, I think, last fall with I don't know how many logical qubits you did, 12 or more.

Chris Langer: creating GHZ states has been a goal and a metric for quantum computing and quantum physics for decades. The first two qubit gate creates a two qubit GHZ state, also known as a Bell state. And it's a measure of how much entanglement you'd have. You're trying to create this very large, entangled state. The number of qubits that are entangled is a great metric, and people track that. And we just recently entangled 50 logical qubits with a a logical fidelity estimated to be between 98% and 99%. We use an efficient error detecting code called the Iceberg code. We did this in the H2 device, and it created a 50-qubit logical GHZ state.

Olivier Ezratty: It's funny because 75% looks low, but for a GHZ state of that size, it's very high.

Chris Langer: I do believe this is the largest GHZ state that has been made, and I do believe, the fidelity is very high for such a large state, yes. Previously, we had done 28.

Olivier Ezratty: I don't think we have much time left to talk about software. It's another thing that you guys do at Quantinuum. Yeah. I met recently with Bob Coeke in New York. There's another, on NLP and LLMs. Of course, it's early stage algorithm development, but it's another interesting path. You work also on chemical simulations, with the guys from CQC. This is another story within the company because you seem to try, like any full-stack company, to look at the case studies which would be appealing for your customers. Any comment on that on a global basis?

Chris Langer: We're trying to look at algorithms in all the different spaces. One of the most obvious ones is chemistry and materials. We have multiple groups looking at that. These could be some of the best candidates for early use cases. In fact, on the chemistry and material side, we're actually just recently finding that maybe some of the requirements on the machine are actually getting a little bit softer. This is a great benefit. We had a recent paper come out on the dilution of error.

Dilution of error in digital Hamiltonian simulation by Etienne Granet, and Henrik Dreyer, arXiv, September 2024 (17 pages).

The concept here is if you're trying to simulate physics and the physics is all local and you want to measure observables, which are local observables, you actually have some tolerance to error and that you have nice error mitigation strategies to kind of recover the answer, even though you could have several errors. I'd encourage your listeners to look further into that. That's some opportunistic things that we're looking at. But we also look at algorithms across all the other spaces. We look at algorithms and optimization. We're looking at how can AI be used to benefit quantum computing and how can you use AI and quantum computing together to get better answers and things like that. We have a whole segment of our company actually dedicated to exploring these areas.

Olivier Ezratty: Do you think that trapped ion QPUs will be highly competitive in the energy consumption space?

Chris Langer: I do actually. I think this is true in ions. It's also true in all the quantum computers because the key benefit here is if you can, is that the quantum computers can perform a calculation that's going to be exponentially hard for a classical computer to do. So one demonstration that we recently did was called random circuit sampling. We did this with our 56 qubit H2 device earlier this year, about April 2024. One of the metrics that we actually found was the energy consumption for a classical computer to simulate the exact same result was about 30,000 times the energy consumption of our machine. It is a good entry point, actually, to show quantum advantage. And if you can't do random circuit sampling, then I think you'll struggle to perform quantum advantage in other applications.

Olivier Ezratty: You have to do that on a real circuit with real data entry and real usefulness. It may be different. When you compare an emulation with an actual circuit, it's not fair for the classical side, because you may do something differently in the classical side to do the other stuff you do on the QPU.

Chris Langer: I agree with you on your philosophy that if you want to compare a quantum computation to a classical computation for the same problem, you have to compare your quantum computation to the best possible method on the classical side to perform the same computation. In a random circuit sampling, the classical methods that are being pursued are the best possible. I think sampling is a hard problem to try to somehow spoof the statistics of the quantum computer's output. That's a hard problem. I believe that the classical comparisons that both we have done and Google has done on that problem are good. And I think the community can have confidence that when we say we have 30,000 less energy consumption, that we really do have quantum advantage in the RCS problem. And similarly, when Google says it's $>10^{20}$ years or whatever it is, that people can have confidence that they have quantum advantage and beating that problem.

Olivier Ezratty: Just to summarize, the company is very optimistic because you plan to reach a couple of hundreds of logical qubits by a couple of years, which is a short time frame in the quantum world. So it's very good news for users, I would say.

Chris Langer: Yes indeed!

Olivier Ezratty: thank you very much. We were with Chris Langer from Quantinuum. It was the 80th episode of Decode Quantum. We started that back in October 2019. No, it was in March 2020, the first episode. So it's nearly five years now. So thank you very much, Chris.

*PS: Fanny Bouton and I have been hosting the **Decode Quantum podcast series** since 2020. We do this pro-bono, without an economic model. This is not our main activity. Fanny Bouton and I are active in the ecosystem*

in several ways: she is the "quantum lead" at OVHcloud and I am an author, teacher (EPITA, CentraleSupelec, ENS Paris Saclay, etc.), trainer, independent researcher, technical expert with various organizations (Bpifrance, the ANR, the French Academy of Technologies, etc.) and also a cofounder of the Quantum Energy Initiative.