

Opinions Libres

le blog d'Olivier Ezratty

Sorting out the confusion on quantum computing utility

Understanding quantum computing's pace of progress is not an easy task. Recent months have complicated it with contradictory messages. CEOs like Jensen Huang (Nvidia) and Mark Zuckerberg (Meta) estimated that useful quantum computers were decades away, while others like Satya Nadella (Microsoft) and Sundar Pichai (Google) cast it as being just years from now. Some companies even stated that quantum computers would soon run LLMs more efficiently, with a lower energy footprint than classical HPCs running Nvidia GPUs.

Your newsfeed is now full of imperatives: "the advent of useful quantum computers is not a matter of if, but when", "this is the year to get ready" (whatever the year), "if you don't jump in now, you're late" (striking fear), such and such technology announcement "is a breakthrough" (Google Willow in December 2024, Microsoft Majorana-1 in February 2025), "we are getting closer to scalable quantum computers" (hopefully, not farther), "quantum computing progress is accelerating" (how do measure progress speed?), Google Willow or China's Zuchongzhi quantum computers are already computing zillion "times faster than the largest classical supercomputer", and so on!

The stock market's ups and downs shake the market cap of Rigetti, D-Wave and IonQ. Here, trust is not based on science but on buzz and beliefs, often only grounded on company's investor decks and quarterly results. This is the characteristic of a "technology hype", with its overpromises. These are fueled by research and industry vendor funding mechanisms, technology sovereignty issues, analysts and service companies' opportunism and media buzz, which I described in an **arXiv paper** in 2022.

End-user organizations and policymakers are also confused. You need a cold head to analyze this frenzy. You're better off being embedded in the quantum ecosystem to know its stakes, but also, not being too much conflicted from the financial standpoint. Evaluating the quantum computing state of the art requires a broad scientific and technological background. If you are a quantum physicist, you may be too specialized, not knowing much about quantum algorithms. On the other hand, if you are a quantum software developer with not much knowledge of the hardware, you may not grasp the physics challenges to build a large scale quantum computer. And you are better off with having some knowledge of the state of the art in classical computing (GPUs, tensor networks, machine learning, etc.) against which quantum solutions are compared.

So, here I am, with a couple of others, trying to bridge these gaps and to provide some understanding of the situation. To ease its reading, this paper is structured as a FAQ, addressing the following questions:

- How to define utility-grade quantum computing?
- What quantum utility has been achieved so far?
- What is the difference between a use case and a use case?
- Why are the first quantum utility use cases about simulating condensed matter physics at 0K and not

solving your business problem?

- What are the challenges of building utility-grade quantum computers?
- What is the current TRL of quantum computers?
- Are NISQ computers useful for the industry?
- What is it to be pessimistic or optimistic?
- How to grade the related scientific vs technology uncertainty?
- Is the pace of progress of quantum computing accelerating?
- What is special about quantum computing with regard to predictions?
- How can we compare today's situation and uncertainties with quantum computing to the ones we had 30 years ago or so in other domains?
- Will quantum computers run LLMs more efficiently than classical computers?
- How about the if/when mantra?
- Should I wait, or should I go?
- How to interpret the quantum stocks' ups and downs?

It's a bit like if you had 16 short papers for the price of one, given it's free. If you cry very loudly, I can turn this into a series of 16 TikTok videos. Or maybe not... Still, you have a one liner blue response to all these questions. And then, elaborations!

How to define utility-grade quantum computing?

It depends on who you are as a user and what you are looking for!

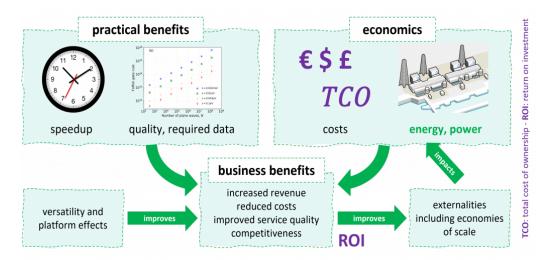
<u>Who you are</u>: a quantum physicist doing a Fermi-Hubbard simulation of some ferromagnetic material (**example**), a pharmaceutical company willing to simulate large molecular dynamics to design some new drug, a chemical company designing a new battery, a financial institution in search of better portfolio optimization techniques or a transportation company willing to optimize its operations?

What benefit are you looking for: a quantum speedup, providing improved results like better chemical accuracy for some chemical simulation, a better solution with heuristics used to find approximate solutions of optimization problems, do something that is entirely impossible with classical computers, do stuff cheaper, with cheaper hardware or less energy? The definition of a "quantum computing advantage" is nowadays becoming multifactorial.

<u>Compared to what?</u> The notion of usefulness is also not absolute but relative to classical computing. It can be against your own IT legacy or with best-in-class up-to-date classical solutions. Nvidia GPUs, deep learning techniques and classical tensor networks (MPS, DMRG) are continuously moving the goalpost for quantum computing to yield some quantum advantage. One interesting example is AlphaFold 3 from **Google DeepMind**, which does stuff initially thought only possible with future quantum computers in the 3D modeling predictions of complex organic molecules.

At what cost? The value of anything new comes from computing its return on investment in comparison with competing solutions. With current quantum computers, this is a rarely debated aspect. Nobody really elaborates on the current and future prices of quantum computers. In the FTQC regime, it will become an important one, which will depend on the affordability of these systems. Hint: utility grade quantum computers supporting the thousand logical qubits and teraquops needed to run industry-grade applications may cost between \$10M (which is acceptable) to several \$1Bs (which is not), depending on the technology.

This is like in quantum physics, it's all about the measurement context!



Variation of a slide from my QEI Workshop presentation showing how practical benefits are connected to the solution's economics and business benefits (video, slides).

One effect described in the above chart is the virtuous cycle that can come from platform and network effects coupled with the versatility of quantum computing systems. These could enlarge the quantum computing market and contribute to developing economies of scale, in turn driving prices down and extending its reach. This is the ultimate definition of "utility grade".

What quantum utility has been achieved so far?

So far, it has been very narrowly obtained with some limited quantum material simulations.

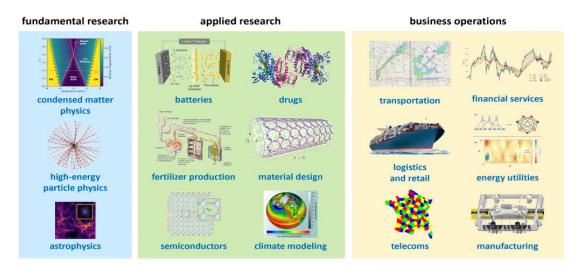
Consider IBM's communication on "quantum utility" which started in June 2023 with the publication of a **Nature paper** and subsequent scientific papers related to the use of their most recent quantum computer (Eagle with 127 qubits, and, later, Heron with 133 or 156 qubits), or more recently, D-Wave with its 4,400 qubit **Advantage 2 quantum annealer**. These "utility cases" were about simulating some kicked Ising model for IBM, and the quench dynamics of two-, three- and infinite-dimensional spin glasses for D-Wave. This is about simulating the magnetism of materials at around 0K, a temperature found only in physics labs, at a scale supposedly unreachable by classical computers! These are interesting physics problems but far from any industry relevant problem. On top of that, both in both the IBM and D-Wave cases, subsequent papers argued that the quantum solution could be implemented classically, using tensor network methods, noticeably from the Flatiron Institute in New York (vs IBM and vs D-Wave). This spurs endless debates between the vendors and these classical computing specialists (example from D-Wave's CEO).

Right now, we have in-store quantum computers operating in gate-based mode like IBM's, or using some analog processes like D-Wave quantum annealers or Pasqal/QuEra cold-atom-based quantum analog simulators.

Quantum computing utility faces different challenges in all these cases due to algorithms resource

requirements, theorical vs practical speedups, and (un)availability of fault-tolerant quantum computers of sufficient size. I ordered these who/what scenarios by order of potential appearance, spanning from now to 20 years.

- **Fermi-Hubbard simulation** of some ferromagnetic material is possible right now in the "quantum utility regime" with IBM QPUs, not far from a quantum advantage regime. It can also be tested at small scale with current analog quantum computing based on Ising models embedding. This is a typical materials simulation case study. Others are dealing with simulating **high-energy physics**.
- Large molecular dynamics simulations require fault-tolerant quantum computers (FTQC) and thousands of logical qubits. This is for the long term, at least over 10 years.
- Portfolio optimization and risk analysis with financial institutions. Goldman Sachs and JPMorgan Chase made various resource estimations for these use cases and they are also in the large FTQC regime, requiring thousands of logical qubits.
- Cryptanalysis. Any company fearing the advent of quantum computers able to break RSA-based classical cybersecurity. Based on current estimates, breaking such keys requires several thousand logical qubits which are also far away in time, as well explained in the German BSI's updated 2025 report. But they don't delve much into the QPU interconnect challenges, let alone practical things like cabling, signals multiplexing, energetics, cryoelectronics, correlated errors correction and the like.
- Operations optimizations for transportation, logistics and retail companies are less likely since it is currently hard to find quantum algorithms with provable theoretical, deterministic, and practical speedups in heuristic-based combinatorial optimizations and at production scale. But D-Wave folks may argue about that. In January 2025, SAP's CEO said that quantum computing was a near-term game changer, using the example of some supply chain optimization case study coming from D-Wave, which is not scientifically documented. Another recent case was put forward with "better result" from Pfizer.



A chart positioning various use cases for quantum computing, from fundamental research to applied research and business operations. Current NISQ solutions seem to fare well in the first category. Future FTQC algorithms will probably bring some exponential speedups and improved results in the second category. The third one based on optimization and machine learning algorithms is more challenging at this point. Source: slides.

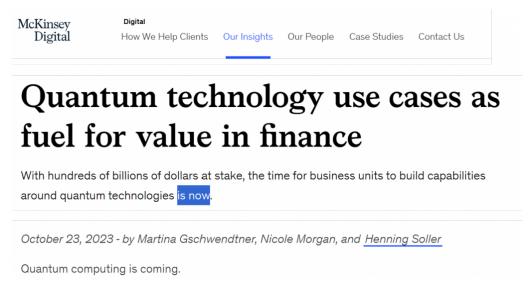
You now understand why CEOs and commenters don't look at the same target. Their wording and definitions are imprecise. BSI is focusing on cryptanalysis, so, as a threshold, breaking an RSA-2048 key using Shor integer factoring, while Jensen Huang talked about the fuzzy notion of useful quantum computers, probably thinking about doing stuff that is out of reach of existing and future classical systems. And when D-Wave touts quantum supremacy, it's about implementing some physical simulation that happens to be useless for all companies. As a result, the irrelevance and standard deviation of all these CEO predictions are pretty high. But claiming that "the utility era is here" is clearly an overpromise, particularly, with mentioning the capability to solve some "real-world chemistry problem" with... two logical qubits!

What is the difference between a use case and... a use case?

There is some double-speak here indeed!

Crafted in 1982, the notion of vaporware described a software or hardware product that was announced but released much later or even never released. One famous vaporware was the first version of Microsoft Windows that was announced in 1983 but released only in 1985. In the quantum world, this would be like releasing a product at the speed of light.

Is quantum computing a modern version of vaporware? No! We have interesting algorithms and software that could provide some potential benefit. They "just" lack the quantum hardware to be executed and deployed. This explains why we have so many application "use cases" which are not real "use cases". Most quantum computing use cases are small-scale prototype algorithms that are tested either in emulation mode or with NISQ quantum computers, often in a regime using so few qubits that they could be emulated on a simple laptop. In that situation, a laptop or simple server would be a more efficient solution with regards to computing time and solution cost. These use cases are frequently presented as "industry use cases", which is ambiguous. It may mean "for the industry", but not "in production". Unfortunately, most of these use cases do not describe well the conditions to obtain some quantum advantage (problem size, number and quality of qubits, etc.). It even happens with scientific use cases like this **recent one** that is using a mere 16-qubit QPU.



Use cases in analysts parlance, source: McKinsey. In 2023!

These "use cases" can start to become interesting when they reach a regime where a classical quantum code emulation is impossible, even using tensor networks compression, when no equivalent classical solution can deliver a similar result in a reasonable time and when it is using real-life data sets. So far, this seems to be not yet possible for industry applications. When it happens, it sits in the "applied research" zone and not in the "industry utility" zone. It seems we'll have to wait for FTQC QPUs to enable real-life industry use cases.

For more insights on this, I published a framework to analyze use cases back in October 2023.

Why are the first quantum utility use cases about simulating condensed matter physics at 0K and not solving your business problem?

One reason is that this kind of physics is close to the physics of quantum computers.

Simulating spin-glass models is easier with D-Wave annealers because they are themselves implementing spin models (*below*). Gate-based computers like the ones from IBM are also adapted to the implementation of spin models due among other things to their qubit layout. This is inline with the famous Richard Feynman's motivation to build a quantum computer in 1981, laid out in the 1982 paper: **Simulating Physics with Computers** (22 pages).

Science RESEARCH ARTICLES

Cite as: A.D. King et al., Science 10.1126/science.ado6285 (2025).

Beyond-classical computation in quantum simulation

Andrew D. King¹¹*, Alberto Nocera², Marek M. Rams³, Jacek Dziarmaga³, Roeland Wiersema⁴¸⁵, William Bernoudy¹, Jack Raymond¹, Nitin Kaushal², Niclas Heinsdorf²¸⁶, Richard Harris¹, Kelly Boothby¹, Fabio Altomare¹, Mohsen Asad¹, Andrew J. Berkley¹, Martin Boschnak¹, Kevin Chern¹, Holly Christiani¹, Samantha Cibere¹, Jake Connor¹, Martin H. Dehn¹, Rahul Deshpande¹, Sara Ejtemaee¹, Pau Farre¹, Kelsey Hamer¹, Emile Hoskinson¹, Shuiyuan Huang¹, Mark W. Johnson¹, Samuel Kortas¹, Eric Ladizinsky¹, Trevor Lanting¹, Tony Lai¹, Ryan Li¹, Allison J. R. MacDonald¹, Gaelen Marsden¹, Catherine C. McGeoch¹, Reza Molavi¹, Travis Oh¹, Richard Neufeld¹, Mana Norouzpour¹, Joel Pasvolsky¹, Patrick Poitras¹, Gabriel Poulin-Lamarre¹, Thomas Prescott¹, Mauricio Reis², Chris Rich¹, Mohammad Samani¹, Benjamin Sheldan¹, Anatoly Smirnov¹, Edward Sterpka², Berta Trullas Clavera¹, Nicholas Tsai¹, Mark Volkmann¹, Alexander M. Whiticar¹, Jed D. Whittaker¹, Warren Wilkinson¹, Jason Yao¹, T.J. Yi¹, Anders W. Sandvik², Gonzalo Alvarez⁵, Roger G. Melko⁵¸⁶, Juan Carrasquilla⁴,5,10, Marcel Franz², Mohammad H. Amin¹,118

¹D-Wave Quantum Inc., Burnaby, British Columbia, Canada. ²Department of Physics and Astronomy and Quantum Matter Institute, University of British Columbia, Vancouver, British Columbia, Canada. ³Jagjellonian University, Institute of Theoretical Physics, Kraków, Poland. ⁴Vector Institute, MaRS Centre, Toronto, Ontario, Canada. ⁵Department of Physics and Astronomy, University of Waterloo, Ontario, Canada. ⁶Max Pianck Institute food Istate Research, Stuttgart, Germany. ⁷Department of Physics, Boston University, Boston, MA, USA. ⁶Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA. ⁶Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada. ¹⁰Institute for Theoretical Physics, ETH Zürich, Switzerland. ¹⁰Department of Physics, Simon Fraser University, Burnaby, British Columbia, Canada.

*Corresponding author. Email: aking@dwavesys.com (A.D.K.); amin@dwavesys.com (M.H.A.)

Quantum computers hold the promise of solving certain problems that lie beyond the reach of conventional computers. Establishing this capability, especially for impactful and meaningful problems, remains a central challenge. Here we show that superconducting quantum annealing processors can rapidly generate samples in close agreement with solutions of the Schrödinger equation. We demonstrate area-law scaling of entanglement in the model quench dynamics of two, three- and infinite-dimensional spin glasses, supporting the observed stretched-exponential scaling of effort for matrix-product-state approaches. We show that several leading approximate methods based on tensor networks and neural networks cannot achieve the same accuracy as the quantum annealer within a reasonable timeframe. Thus quantum annealers can answer questions of practical importance that may remain out of reach for classical computation.

But chemical simulations are about simulating so-called fermionic models, when both gate-based and analog quantum computers are implementing bosonic models. As a result, computational problems that are not natively formulated as spin problems or bosonic problems need some conversion, which incurs a significant classical and quantum computing overhead. It is the same with solving combinatorial optimization problems. They can be mapped on quantum annealers or quantum analog simulators, but with some significant "embedding" overhead.

All this explains why other problems require larger quantum computers, with our current knowledge of quantum algorithms. Hopefully, quantum hardware and algorithms are making progress and folks are still hopeful to find some Holy Grail there.

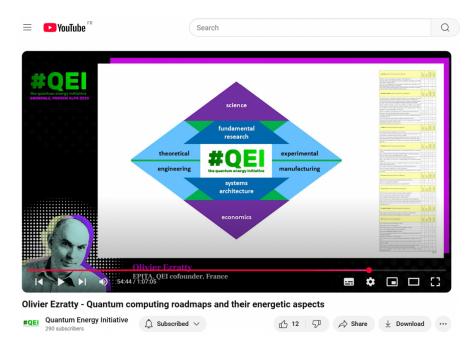
What are the challenges of building utility-grade quantum computers?

There are many such challenges from fundamental research to engineering, and even economics.

As a cofounder of the **Quantum Energy Initiative** with an engineering background and perspective, I am looking at the challenges ahead to build utility-grade quantum computers, particularly in the FTQC regime. Understanding, evaluating and optimizing the energetics of these systems must be based on a holistic scientific and engineering perspective, up to looking at the economics of quantum computers. This complements a

bottom-up approach that is focused on determining the fundamental bounds of qubits operations, belonging to the fields of quantum energetics and quantum thermodynamics.

In the **second QEI Workshop** held in Grenoble in January 2025, I had a chance to provide a view on current quantum computing roadmaps and their energetics aspects (**video**, **slides**).



In this talk, I inventoried some challenges and connected them to dozens of scientific and engineering questions:

- Finding quantum algorithms bringing actual practical speedups over their classical counterparts, and not just theoretical ones. And preferably in a generic way, meaning, for large classes of problems. This is a particularly difficult task for solving combinatorial and decision problems. You can forget about Grover's algorithm which supports only some quadratic speedup, which is not sufficient in practice. See why here and there.
- Considering all classical software computing costs of any quantum solution, like all the circuit preparation (for a QPE-based chemical simulation in a FTQC regime) or classical optimizer preparation (for a variational VQE-based quantum simulation in an existing NISQ system), but also classical error syndrome detection (for FTQC) and large code compilation (for FTQC). By the way, all quantum algorithms are "hybrid". Whether they are variational in NISQ or not in FTQC, they always have a significant classical part. I highlighted here the fact that quantum code compilation is a variable cost against the use case data. Given large circuits compilation and optimization may be costly, it may prevent quantum computing from being used on business applications with fast duty cycles. These costs, for a starter, can be expressed either in computing time or energy spent.
- Assessing the cost of QPU interconnect which will be necessary for all qubit modalities, even the ones which will squeeze the largest number of physical qubits in monolithic QPUs. In the case of Google, their plan is to create a 10K physical qubit chip. In order to support a "teraqops" chemical simulation (teraquantum-operations), a single chip may support only a single logical qubit!
- Creating higher fidelity qubits at larger scale, which remains an overarching huge challenge. This has its load of engineering questions related to the cost of electronics, cabling, cryogenics, and lasers, which

depend on the qubit type.

- Optimizing large algorithms computing times, with either speeding it up at the software/compiler level, or with having faster gates/readout/error correction, or at last, with parallelizing several circuit shots on multiple similar QPUs, provided they are affordable enough. I nicknamed this the EFP framework, for "efficient-faster-parallel". Within this framework, I equated software to the total energy cost within the perspective of what to do to reduce computing time for solving a given problem. The total solution cost is proportional to energy costs, which can be optimized at the hardware level. You can then optimize computing time with three complementary means: reducing the total classical and quantum software cost, accelerating quantum gates, readout and error correction, and/or parallelize circuit shots. You can also reduce the number of circuit shots by reducing your outcome precision needs and by improving software.
- **Defining some energy footprint acceptability threshold** for large FTQC systems, with a reasoning that could be applicable to quantum computer overall economics and pricing. Based on current estimates, utility-grade QPUs supporting thousands of logical qubits may need a power ranging from less than a MW to several hundred MW. This power drain and economic question should soon drive many rationales, discussions and comparisons in the development of FTQC quantum computers.

I should add here various other scalability challenges like handling data preparation efficiently, particularly for solving linear equations and quantum machine learning tasks. Then, we are still in search of some potential qRAM, a memory type that is nearly mandatory to execute oracle-based algorithms.

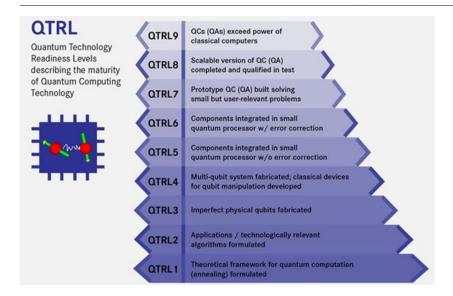
Then, each qubit type (or "modality") has its own scalability challenges. They can be very different. For example, solid-state qubits like superconducting and silicon-spin qubits are fighting against decoherence, noise and crosstalk whereas photonic qubits are hampered by low photonic sources and detector efficiencies and losses along the paths in photonic integrated circuits and optical fibers. These are very different scientific and engineering challenges.

What is the current TRL of quantum computers?

It depends on how you define it!

A TRL, aka, technology readiness level, is a metric from 1 to 9 defined by NASA in the mid-1970s. It characterizes the level of maturity of a given technology or product. 1 stands for an idea in the head of a researcher and 9 for a widely deployed technology like your smartphone or laptop computer. In between are physical experiments, prototype products and the first products being deployed. The TRL of a product can also be characterized by the readiness of its related ecosystem (training, skills, applications, documented deployed case studies).

Let's consider the quantum TRL scale created by Kristel Michielsen in 2018, as shown below.



According to this scale, quantum annealer's TRL sits around 7 while gate-based systems are at around 5 given they are not yet commercially implementing quantum error correction. You can still buy a gate-based QPU with over a hundred physical qubit, like Cleveland Clinic who acquired a System One from IBM and installed it in Spring 2023, You can use them in the cloud through AWS, Microsoft, IBM or D-Wave online services. You can learn quantum programming and test your code. Even D-Wave has an (undocumented) SLA (service level agreement) should you use their online QPUs for production. There are already tens of thousands of quantum developers worldwide thanks to these systems being available online. You can test a 127-qubit IBM Eaglegeneration QPU for free.

But the TRL is also currently quite low when considering the usefulness of quantum computers with regards to their added value and cost. Let's say you test an online quantum computer with 20 operational noisy qubits. The system probably costs several \$Ms and its hourly usage is charged at about \$2K. But \$2K is the price of a laptop where you could run your quantum code in a free open source emulator. And it would even run faster than on the quantum computer. For a higher number of qubits on a current noisy quantum computer, the situation is blurred depending on your use case as explained when responding to the previous question.

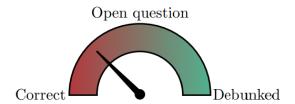
So, technically, quantum computers (QPU) TRL sits at around 5-8 depending on the case, but their actual value is still low compared to legacy classical systems.

Are NISQ computers useful for the industry?

For prototypes and physics research but not yet for industry production cases.

NISQ computers are the current noisy qubit quantum computers. Those you can buy, rent or test today. The current wisdom is that, while they have some scientific value, their industry value is questionable. This is **John Preskill's view**. I described NISQ challenges in a **2023 paper**, and things have not significantly changed since then.

The main NISQ recent advances come from some improvements in qubit numbers and fidelities, with both superconducting and ions qubits, and from the use of quantum error mitigation, a technique promoted among others by IBM. Some use cases are presented, like with the **interesting work** from Algorithmiq, a software startup based in Finland.



Verdict 6. While proving exponential speedups for commercially valuable end-to-end applications remains mathematically challenging, it is reasonable to expect that quantum computers will eventually deliver meaningful speedups in practical problems.

A broad industry and academic team led by Sabrina Maniscalco, who is the CEO of Algorithmiq, published a paper willing to fight some myths on NISQ: Myths around quantum computation before full fault tolerance: What no-go theorems rule out and what they don't by Zoltán Zimborás, Fernando G. S. L. Brandão, Elica Kyoseva, Ivano Tavernelli, Sabrina Maniscalco et al, arXiv, January 2025 (11 pages). It tried to show that the power and use cases of NISQ with quantum error mitigation could be expanded thanks to the upcoming improvements of qubit fidelities. However, the paper leaves the question open of whether NISQ systems will deliver some exponential or practical speedups. Another open question is lingering about the potential known and unknown limits of analog quantum computing.

What is it to be pessimistic or optimistic?

It is a mood, not a scientific approach.

First, you need to understand where we are and to identify the scientific and engineering challenges ahead. Some are workable, others are really hard. It never necessarily means that "it's impossible". It is up to scientists and engineers to find solutions. You then need to appreciate how long it could take, understanding for example the lengths of the manufacturing-testing-experimenting cycles that dominate in this industry. These cycles are long!

Usually, the more you know a topic, the less optimistic you may become because you understand the sheer weight of the challenges ahead. It can lead to **sheer pessimism**. Listening to educated pessimists should be used to challenge the optimists. But you can also be excited by these challenges and by human creativity. That's my situation.

I try to avoid flirting with optimism or pessimism, which are just two opposite moods. I stick to the scientific state of the art and remaining challenges. My own belief is that by doing so, you are neither pessimistic nor optimistic, you are just neutral with some optimism bias even though the pure optimists will perceive you as being a pessimist.

Home / Physics / Quantum Physics









(T) OCTOBER 22, 2024



Optimism meets skepticism at **Quantum Summit in Chicago**

by Robert McCoppin, Chicago Tribune

So, pessimism or skepticism? Source.

When you identify challenges in the quantum computing space, and there are many, most of them are currently being addressed in multiple ways by scientists and industry vendors. What can make you optimistic is that multiple technologies are being developed, like with the types of qubits or even with all the varieties of quantum algorithms around.

How to grade the related scientific vs technology uncertainty?

It's a mix of both.

A common wisdom with many physicists, including the Nobel in physics laureates I met in Lindau in July 2024 is that the challenges ahead are mainly about engineering and technology developments. According to them, the theory is accepted and proven. But some scientists, like Alain Aspect, also think that creating large scale quantum computers may require significant time, a bit like the LIGO gravitational waves detection experiment which took about 25 years from design to completion (from 1989 to 2015).

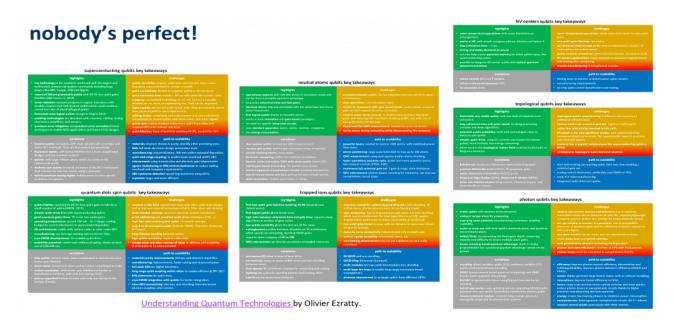
I do not entirely agree with that focus on engineering. While for sure, engineering and technology developments are inescapable routes to building fault-tolerant quantum computers, there are still some scientific uncertainties to unfold. A bit like the whole history of physics, experimenting new artificial settings will uncover new mysteries which may require building new theories, a bit like Zeeman's effect discovered in 1896 could be explained only about 30 years later with electron spins.

Here, we don't know yet what we will discover when creating large-scale physical qubit systems. Will they showcase high-fidelity entanglement states at scale? How will we characterize it? Will we reach the famous Heisenberg cut, when macroscopic systems stop behaving quantumly with coherence and behave classically without it? Will we face the effects of the Lieb-Robinson bound that conditions the speed of correlation propagation in large scale entangled systems? It may affect large QPUs, particularly those which are implemented using quantum interconnect techniques. The confusion may come from the difference between theory and science in general. Quantum physics theory may be somewhat settled, but fundamental research and experimental research is still an open field.

Another huge quantum computing scientific uncertainty lies with algorithms. We don't have that many quantum algorithms and even fewer quantum algorithms with provable exponential speedups, or practical speedups with reasonable times. There is still a lot of fundamental research to do there. It is likewise with quantum error correction. How about new sources of correlated errors that we'll discover with scalable systems and for which current error correction codes are not well adapted? In the technology realm, we don't know yet if the improvement of chips manufacturing processes will yield better quality qubits. We need to test it. At last, how about various "unknown unknowns"?

I'd say quantum computing is becoming a broad scientific experimental endeavor, mixing theoretical and experimental advances, one helping the other. We need these experiments to validate the theory and/or to extend it. We also need to accept failures and try various technology avenues. That's why it doesn't yet make much sense to launch a Manhattan like project on a specific qubit modality. It explains for example why DARPA is funding about 11 different such avenues as part of its **Underexplored Systems for Utility-Scale Quantum Computing (US2QC)** and **Quantum Benchmark Initiative** programs.

What is the cost of trying these various avenues? Is it acceptable? It is currently mainly supported by a mix of public and private investments with a high proportion of the former. It will go on like this if the strategic stakes are perceived as being as high as they are today.



A bird's eye view on the realizations (green), challenges (orange/red), variations (grey) and paths to scalability (blue) for seven qubit types: superconducting, silicon, topological, cavities, cold atoms and trapped ions. They are all different! Source: Understanding Quantum Technologies 2024.

Is the pace of progress of quantum computing accelerating?

It is advancing at a steady pace, but not really accelerating.

The perception of acceleration comes from industry vendor hyperbolic communication. Recent news pushed by large companies like Google and Microsoft are driving this impression. Microsoft oversold its topological qubit, that are supposedly around the corner. They are not! Too many scientific or industry communications are about "getting closer to scalable quantum computing" (example 1, example 2, example 3).

I don't believe there is some acceleration. There is progress for sure, but no acceleration. First and foremost, how do you measure acceleration and speed of progress? What is the comparable? What is the meter and unit?

One way to proceed is to compare what scientific and technological advances have been achieved vs past plans. When I started working in the quantum computing space in 2018, I tracked the following announcements: Rigetti was **planning** to release a 128 qubit QPU in 2019, IonQ was **mentioning** that it had controlled 160 qubits on which it could handle operations on 79 qubits, and Intel was **showcasing** a silicon chip with the capacity to host 1,500 qubits in a chip thinner than a human hair. In 2019, Google had a 53 qubit chip (Sycamore) and IBM, a 65 qubit chip (Hummingbird r1).



Where are we now, 6 to 7 years later? Rigetti's best chip is Ankaa-3 with 84 qubits. IonQ is at 36 qubits with its **Forte system** while Intel has a vague 12-qubit chip named Tunnel Falls released in 2023, that is not integrated in a full QPU system. Google reached 105 qubits with its Willow chip announced between August and December 2024 and IBM has a 156-qubit chip. Meanwhile, these qubits fidelities have somewhat improved but not in a stellar way. IBM's 433 Osprey and 1,121 Condor qubit chips which were released in late 2022 and 2023 were even discontinued in favor of smaller chips (Heron 156 qubits) due to their lack of tunable couplers. With these couplers, IBM's Heron is faring well, at about 99.7% two-qubit gate fidelities, not far from what is currently achieved with trapped-ions (99.84% at Quantinuum with its H2-1 system).

Still, all this doesn't look like being exponential. It proved more challenging to increase the number of qubits of decent quality at intermediate scales of about a hundred physical qubits. You can see the trend in the updated scatter plot below. The blue dots are about superconducting qubits and the green ones with trapped ions. Cold atoms in orange seem to scale well but have a hard time exceeding two-qubit gate fidelities of 99.5%. Ions achieve that, faring at >99.8% but are still struggling with scaling. Superconducting qubits are in-between, with 99.7% two-qubit gate fidelities at between 105 and 156 physical qubits.



In other qubit modalities, progress looked faster, like with cold atoms, first on large settings like Pasqal with the control of 828 atoms in 2024, and Caltech with 6,100 atoms, and then in gate-based mode (QuEra, starting in December 2023 with 48 basic logical qubits, Atom Computing with 24 logical qubits in November 2024).

Photonic qubits are either locked at low qubit count in the non-scalable KLM paradigm (Quandela, Orca Computing, ...) or are bound to implement FTQC scalable architecture in the more or less distant future (again with Quandela, with PsiQuantum, and Xanadu).

Since 2018, we've also seen the emergence of new breeds of qubits implementing some autonomous error correction, like the cat-qubits from Alice&Bob and AWS, or the GKP qubits from Nord Quantique. These are encouraging new technology developments that may reduce the physical per logical qubits number overhead.

There are several technical limitations to the # of qubits per monolithic quantum processors. It leads their designers to envision using various quantum interconnect solutions (microwaves, optical photons, transduction), which will bring a wealth of new challenges, from physics to quantum code compiling and optimization. We thus have to be very careful when extrapolating physical qubit numbers, particularly toward million qubit ranges.

While there was significant progress in the field of quantum error correction, these advances were not well interpreted. For example, Google Willow's logical qubit only corrects errors in a quantum memory, but not single or two qubit logical gate errors. It is thus far from being "fault-tolerant". Likewise, the logical qubits created by QuEra and Quantinuum (with Microsoft) implement logical gates, but with the non-scalable trick of "post-selection". It has the flavor of quantum error mitigation implemented in post-processing manner more than a fault-tolerant scheme adapted to running large quantum circuits.

As I described in my Moore's paper in 2023, scaling QPUs won't probably follow any Moore's law like with CMOS semiconductors because the scalability challenges are highly multidimensional. It's not just a matter of shrinking transistors. I simplify things here since advances in CMOS were more complicated than that. Also, the test & learn cycles with various quantum computing technologies are relatively slow and don't follow any Moore's law.

However, we may underappreciate some interesting technology developments happening in the background that will pay off later: on cryogenic control electronics, on high-density cabling, control signals multiplexing, high-power and low phase noise lasers, and with high-power cryogenics.



The Coyote and Road Runner. A famous Warner cartoon known by boomers. Image source. Now, who's who here? The Coyote may represent quantum computers while road runner is classical computing. And maybe, someday, it will be the other way around.

What is special about quantum computing with regard to predictions?

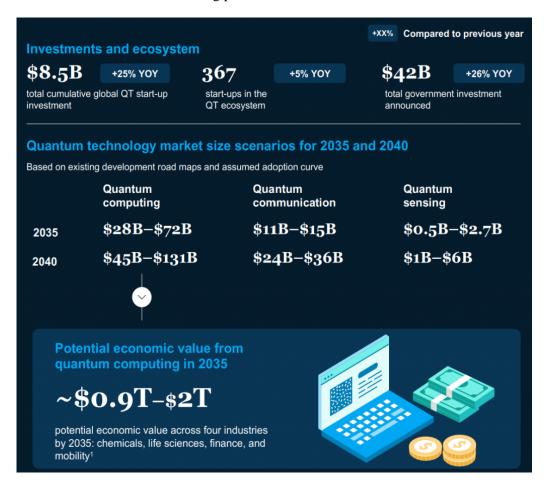
There is still a lot of scientific uncertainty, including on large-scale experiments.

In the not-so-distant past, predictions abounded about how fast a technology revolution would be adopted. The cloud, mobile computing, Internet of things, and more recently the metaverse. Artificial intelligence made its comeback in two periods, one between 2012 and 2022 with machine learning and deep learning around data science concepts, and since 2022, with the advent of LLM-based chatbots.

In these domains, the technology was more or less mature, but it had proven use cases that were beyond the state of the art. For example, mobile computing and cloud computing brought usability and easy access to multi-format communication on the go. In other domains like with IoT and metaverses, the generic user value was neither obvious nor universal.

The uncertainties in these various domains were however not so much about the underlying science and technology, than on other factors: costs, ease of use, user value, physiological acceptability in the case of the metaverse, social impact and so on.

With quantum computing, we are exposed to a flurry of market predictions including how much value it could generate, around \$2T as soon as 2035 according to **McKinsey** before even the technology is ready. Even if it was ready in 2025, it would take more than 10 years to yield such value. You'd have to deploy it, to create algorithms, to run it, to try different options to create a digital twin of a new drug or chemical compound candidate. Then, you'd have to test it in the real world, either undertaking multi-year clinical tests or launching the buildout of new manufacturing plants. And so on. This takes time!



McKinsey quantum monitor as of April 2024. Notice the \$42B total government investment which is wrong due to bad estimated of China's investments as well explained in The Quantum Panic – The U.S. wants to be prepared for whatever quantum technologies bring, but is it time to rightsize the threat from China? by Rachel Cheung, The Wire China, February 2025.

How can we compare today's situation and uncertainties with quantum computing to the ones we had 30 years ago or so in other domains?

Carefully...! The mix of scientific, engineering, economic and societal uncertainties is quite variable across domains.

There were not many doubts back then about the progress ahead with raw computing power, memory, storage

and networking capabilities growth. Moore's law was kind of predictable, at least until the mid-2000s.

Telecom infrastructures were key for developing the Internet and for smartphones and the cloud. There were some infrastructure built out delays for fiber installation around the world. However, this was not due to uncertainty in scientific or technology. It was more an economic issue. Also, many Internet developments benefited from the technology but were accelerated for societal reasons, like with the advent of the web 2.0 around 2004. There are some parallels to draw here with FTQC QPUs. At some point, they will become heavyweight infrastructure programs requiring significant funding, beyond classical research funding in the academic and startup world.

In our analysis, we are often affected by a "survival bias". We use the example of successes as proof of our reasoning, forgetting past failures. They don't resonate in memories since they didn't affect us much. In the mid-1980s, there was some buzz on GaAs electronics, that was supposed to supplant CMOS. It didn't, mostly for engineering and economic reasons. Nanomaterials and graphene technologies had similar troubles, ranging from scientific and technology overpromises and societal backlash.

Classical analog computing was promoted in the 1960s to 1970s but failed, being overcome by digital computing. It was not generic enough and didn't scale well due to noise. Quantum computers are also analog to some extent, of course, with analog quantum systems ala Pasqal and D-Wave, but also with NISQ systems implementing arbitrary rotation gates. One key difference is that FTQC mixes analog and discrete operation, which is to enable quantum error correction, a feat that was not accessible to classical analog computers. Also, analog computing didn't benefit from the same attention and public investments as we have right now with quantum computing. There was no equivalent of Shor's algorithm to drive government interest.



The EAI 8800 Analog computing system in 1986. Source.

By the way, where are we with the promises of the Hyperloop, promoted by a certain Elon Musk in a 2013 white paper? Several ambitious infrastructure projects were supposed to be launched around the world. A couple startups were created. Twelve years later, it's not really in the radar. The idea is plagued with a mix of engineering problems, safety problems, capacity and economic limitations and huge infrastructure costs. How about drones flying around to transport passengers from airports to cities? They can indeed fly. But with limited range, air trafic control challenges, economic challenges again, and so on.

Now, let's consider the Apollo program. It took only 8 years from JFK's speech in May 1961 to Congress to having men landing on the Moon in July 1969. Yuri Gagarine had been sent to space a couple weeks earlier, in

April. JFK confirmed the goal in **another speech** in September 1962, listing some technical challenges. Weeks from the start of the Cuban missiles crisis! Of course, the Moon quest project was fueled by huge investments but it was mostly about solving technology challenges. In 1961, engineers already knew how to launch a rocket and send a human into space. Newton's laws were already vindicated. But there was the Cold War between the USA-USSR rivalry.

Quantum computing seems much different and more challenging than all the above examples from the scientific and engineering perspective. There are superposed stacks of uncertainties from securing large scale entanglement to designing useful algorithm, and with relatively long test and learn cycles. And we have not yet started the real economical discussion about the price of these systems and their return on investment. From the scientific and engineering standpoint, quantum computing seems closer to nuclear fusion. Both are linked to the limits of physics and experimentation. The advantage of quantum computing is that it is way less capex intensive than nuclear fusion. At least at this point.

Will quantum computers run LLMs more efficiently than classical computers?

Probably not, even in the long term! But quantum artificial intelligence has some interesting prospects.

Some quantum computing companies like IonQ and Quantinuum are flirting with the idea that quantum computers may someday run LLMs more efficiently than classical computers. They could even be more energy efficient.

IonQ explores power of quantum computing in LLM use cases

Jul. 08, 2024 8:54 AM ET \mid **lonQ, Inc. (IONQ) Stock** \mid By: Brandon Evans, SA News Editor \mid 24 Comments

What investors got from some IonQ presentation in 2024. Source.

Unfortunately, quantum computers will probably not help mitigate the growth of energy consumption linked to consumer LLM based-generative AI. Data loading is quantum computers' Achilles' heel and will remain so for a long time given they will unlikely benefit from some hardware Moore's law in relation to their clock speed. Given the size of both the training data (in petabytes and tera-tokens encoded in highly multidimensional vectors) and models (in tera-parameters), they won't be loaded at a reasonable timescale on any quantum computer for training, even with some classical data ingestion and preparation. Variable computing and energetic costs are mostly in inferences which are run nearly real-time. Any quantum algorithms running in a quantum advantage regime will probably not be real-time. We can also use some simple logic here. An inference would require one quantum computer and there are hundreds of millions of users of these LLM chatbots. Unless some clever algorithm enables multiple inferences to run on single QPUs, this can't scale well.

The path to quantum transformers by David Wakeham, Xanadu, April 2024, provides sketchy classical and quantum computing requirements to train LLMs. It shows that even with very optimistic assumptions on the clock speed of quantum computers (in GHz), and with several tens of thousands of logical qubits, quantum computers would be faster than classical computers for the embedding part of LLM only after a couple decades, but this task could be run classically and only kernel operations handled quantumly (linear operations using block-encode matrices). Quantum linear algebra is all you need for Transformer architectures by Naixu Guo, Zhan Yu, Matthew Choi, Aman Agrawal, Kouhei Nakaji, Alán Aspuru-Guzik, and Patrick Rebentrost, arXiv, February-May 2024 (31 pages) states that "embedding large data into quantum computers is difficult in

the absence of the availability of functioning quantum RAMs".

So maybe, we should invent the notion of qSLM for quantum small language models!

However, the broader field of quantum artificial intelligence, not including LLMs for consumer chatbots, is a good exploration path, particularly when there is a need to use smaller data for training and to obtain better accuracy in inferences. Quantum machine learning can potentially be used for both training and inferences. There are still many technology interdependencies here: on data preparation and on the development of qRAM and on the connection between classical and quantum components of an AI solution.

One thing is sure: classical AI already helps develop quantum technologies, in quantum error mitigation, quantum error correction at the error syndrome detection level, qubit calibration, and even circuit design (using custom Transformers).

How about the if/when mantra?

It is overly simplistic. There are still many ifs.

Analysts and others are frequently parroting the "if not when" mantra, saying that the question is not if scalable quantum computers will appear but when. As shown above, there are still many "ifs" which explain a lot of the uncertainty on the "when". Otherwise, we would have an infinite certainty on the "when". It is a bit like if "if" and "when" were complementary observables in Niels Bohr's parlance. And due to Heisenberg indeterminacy principle, you can't be precise on both observables. So if you dare want to be a bit more precise on the "when", you need to be less accurate on the "if".

Should I wait, or should I go?

Go! Learn, evaluate, test.

That's the typical **question** for end-user companies. Some analysts are playing with FUD (fear, uncertainty and doubt) to drive customers on the quantum computing bandwagon: human skills and quantum computing resources will be scarce, there will be a significant first-mover advantage, and your competitiveness will highly depend on your adoption of quantum technologies.

I don't believe any of these assertions to be true or verifiable. If quantum computers work at scale and benefit customers, hardware vendors will manufacture them and sell them. In the history of technology, even for large systems like supercomputers, there's never been a real scarcity problem. And if there was one, it would mean that scalable quantum computers are way too large and costly. Then, human skills can follow-on. We already have a good stream of young people being trained. Finally, your competitiveness may only marginally depend on quantum computing.

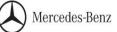






















matmut 🕸

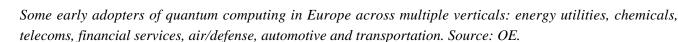












Still, I would advise customers to invest in quantum computing. Am I contradicting myself? No, I'm just basing my rationale on different arguments. One is that testing quantum computing is not expensive and is part of any techno-screening of any large company. You don't need to acquire a \$40M QPU. You can rent it for cheap in the cloud. You can even learn quantum coding with free open source emulators.

Second, learning quantum computing is an excellent way to revisit your existing applications portfolio. It helps you identify pain points in scalability or output quality, revisiting the algorithms you are using. At some point, it will help you modernize your IT, and not necessarily with quantum computing. You may land in using GPUs and/or tensor network "quantum inspired" classical applications.

Learning about quantum computing can help you upgrade your algorithms teams. It will push them to develop better mathematical skills, like in the financial services vertical. It will also help customers deal with the aggressive push from industry vendors and to decipher the continuous news stream. In just one week, you'll have thousand of talks from quantum physicists at the APS March Meeting held in Anaheim, and the Nvidia Quantum Computing Day in San Francisco, hosting a dozen quantum computing vendors. Get ready for a flurry of announcements!

Of course, another minimal task to launch is the audit and upgrade of the company cybersecurity infrastructure. Whether you like it or not, you'll have to fold in deploying post-quantum cryptography.

So, learn quantum computing at small scale, educate your own technical people, try small things, be ready to adopt the technology should quantum computing scale well on a reasonable time horizon.

How to interpret the quantum stocks' ups and downs?

Irrational exuberance or despair...!

In Summer 2024, Rigetti and D-Wave risked being delisted from the Nasdaq due to their stock becoming for some time a "penny stock", valued at less than a US dollar. But in October 2024, they escaped this dangerous zone. Then, Google's Willow PR blitz in December 2024 drove the stock to sky highs. It happened also with IonQ's stock, which was not endangered. In January 2025, Jensen Huang's declaration created a sharp drop in value for all these stocks and they rebounded relatively fast afterwards, then fell down again, and are up again since the beginning of March 2025.

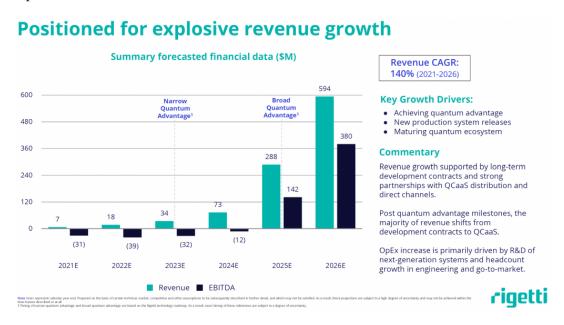
What was the rationale here? You may wonder. Google Willow advancements had no connection at all with IonQ's roadmap. Rigetti is a direct technology competitor of Google, but lagging behind it. And D-Wave? It's been struggling for a while at generating some decent revenue. And the company is now over 25-year old. It seems to be a rare case of a company losing money for such a long time, although it became public only in 2022. These stock values are driven by a mix of volatile human sentiments and machine-based automated triggers.



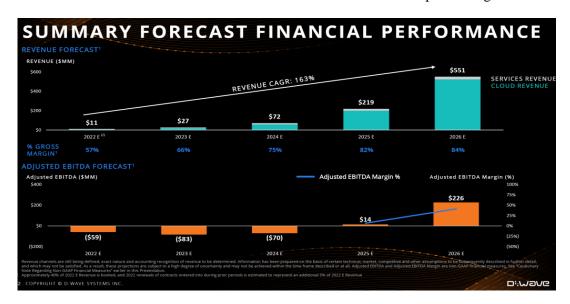
Chart showing Rigetti, IonQ and D-Wave stock value evolution over one sliding year (source).

Now let's have some fun and look at the roadmaps, past revenue forecasts and actual results for these traded companies, which all went public using the shady SPAC funding vehicle.

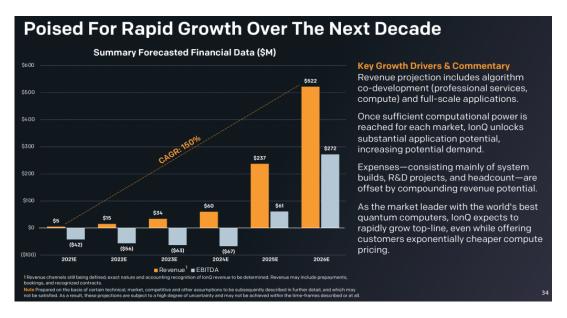
Rigetti's 2021 roadmap announced that they would obtain a "broad quantum advantage" in 2025. How many use cases do they have with their 84-qubit QPUs? Not many, besides some QEC experiment with Riverlane announced in October 2024. Their qubit quality is too bad to do useful NISQ stuff, even with using quantum error mitigation. Their revenue projection was \$73M for 2024 and \$288M for 2025. In March 2025, they reported a revenue of \$10.7 for 2024, while 2023 was at \$12M. And a loss of \$201M!



D-Wave's 2022 investor presentation was forecasting a revenue of \$72M in 2024 and \$219 for 2025. Actual revenue for 2024 was \$8.8M while 2023 was at \$8.7M. Not stellar exponential growth either!



In its 2021 investor presentation, IonQ planned to reach 64 "algorithmic qubits" by 2025 with a 16:1 error-correction encoding, meaning a thousand physical qubits if I count well. And 2024 revenues of \$60M. The actual was not that far, with \$43M, although, mostly thanks to US federal commitments from AFRL. And they currently have only 36 operational qubits. The wake-up call for investors is hard, as shown in Kerrisdal Capital's recent short-selling and **well-documented report!** Now, they just acquired IDQ with a \$250M stock swap with SK Telecom. The rationale? Revenue diversification and patent portfolio consolidation!



That's not a pretty picture from the investor standpoint. We must recognize that all startups have the same problem. They make inflated sales forecasts to attract investors and then usually fail to deliver it. The big difference here is that this is not linked to some sales or competition shortcomings, or lack of some viral effects like in web consumer space, but because the technology is not yet there to capture the target market.

Still, the market value of these companies totals over \$10B (\$3.15B for D-Wave, \$2.75B for Rigetti, and \$5.41B for IonQ). I have observed how IonQ investors on X were disappointed by the company's quarterly results. Seemingly, they didn't talk much with scientists and technology specialists to forge an educated opinion about the company. All investors are misled by a superposition of over-expectations: on the potential business value of quantum computers (culprits: McKinsey and BCG), on the related revenue opportunity (culprits: same folks), on the technology roadmap (culprits: most vendors and analysts, including Google and Microsoft), and

on the difficulty of creating useful quantum computers. In comparison, IBM looks like a sane player, being more careful with its promises and roadmaps.

Now, these three traded companies are not the best-in-class QPU vendors around. There are many other vendors who are not yet traded. Some are advancing their R&D quietly, avoid overselling their hardware, and adopt a long-term ploy. They benefit from a mix of private and public funding, including public procurements like with the EuroHPC program in the European Union.

Conclusion

The quantum scene is becoming the theater of double-speak with utility, use case, industry-grade, acceleration, if, and when, where the present and the future collide in the public discourse.

Thanks for reading this paper until the end. What key question did I miss out?

PS: thanks to Kate Prebble, Philippe Grangier, Cécile Perrault, Marie Campo, Axel Ferrazzini, and Michel Kurek who gave me feedback on this paper content.

Cet article a été publié le 19 mars 2025 et édité en PDF le 19 mars 2025. (cc) Olivier Ezratty – "Opinions Libres" – https://www.oezratty.net