

Quantum computing for data science

Barry C. Sanders

Institute for Quantum Science and Technology, University of Calgary, Alberta T2N 1N4,
Canada

E-mail: sandersb@ucalgary.ca

Abstract. I provide a perspective on the development of quantum computing for data science, including a dive into state-of-the-art for both hardware and algorithms and the potential for quantum machine learning.

1. Introduction

Quantum information theory transforms the very foundations of information theory and computing [1, 2]. The pre-quantum (known as ‘classical’) scientific framework allows objective information to be labelled by integers such as bit strings (e.g., representing text characters by 7-bit strings according to the American Standard Code for Information Interchange, or ASCII), which is foundational to information theory. Processing information can be executed under the rules of Boolean logic, manifested as concatenations of one-bit operations such as NOT and two-bit operations such as NAND.

Quantum information changes the information game entirely by allowing coherent superpositions of informational states, which, following the quantum principle of complementarity, can be thought of as being both corpuscular (particle-like) and undular (wave-like). For example, a three-bit string 010 becomes quantumly a label for a quantum state $|010\rangle$ (element of Hilbert space), which could be manifested physically, say, by three electrons that are spin-down, spin-up and spin-down, with a spin-down state labelled $|0\rangle$ and a spin-up state labelled $|1\rangle$ in Dirac, or bra-ket, notation [3]. Superposing this three-electron state with its orthogonal complement would be $|101\rangle$. For state normalisation implicit throughout this paper, the superposition of these two states is $|010\rangle + |101\rangle$, which, in binary representation, is a superposition of the numbers 2 and 5.

These superpositions of information states can be processed quantumly, i.e., in a way that preserves coherence. Ideally, this superposition state can be transformed by an arbitrary unitary map (isometry on Hilbert space). Realistically, open-system effects such as noise and loss can impede performance, but almost-unitary maps (such as completely positive trace-preserving maps [1] that are close to being unitary) can suffice for useful quantum-information processing provided that the quantum version of error correction is employed in a fault-tolerant way [4].

An early motivation for quantum computing was for simulating physics, particularly for simulating quantum systems in a way that is natural for quantum descriptions [5], i.e., by using quantum computing. Since that original idea, remarkable quantum algorithms have arisen, where remarkable refers to delivering superior performance compared to classical algorithms, such as efficient computation meaning computational resources such as run-time and number of



bits or qubits, being quantum bits, scale no worse than polynomially with respect to the size of the problem quantified by how many bits are required for specifying the problem instance [1].

One example of quantum algorithms being advantageous is for efficiently solving number factorization, whose hardness is subexponential (almost exponential) using the best known classical algorithm [6]. Another example is a provably quadratically faster algorithm for function inversion (hence, unstructured search) [7]. Quantum annealing is speculated to enhance optimisation methods in some cases [8]. Building quantum computers is rapidly advancing, albeit small-scale and noisy, making quantum computing now viable (in a limited way) and commercial [9, 10, 11].

Quantum computing is arguably a disruptive technology, i.e., a technology that is both innovative and market-changing [12], if quantum computing eventually delivers a game-changing application such as for data science [12, 13]. Some big ‘ifs’ stand in the way of quantum computing being disruptive, though. My view is that quantum computing faces two huge ‘ifs’, i.e., challenges: (1) building a quantum computer and (2) making said quantum computer useful.

In a way, the latter is harder than the former. Scaling a quantum computer to large size (many qubits) and large depth (able to pass through many cycles of quantum logic gates, defined as unitary or near-unitary maps of quantum information states) is technically extremely challenging but conceivable. Figuring out, on the other hand, what to do with a quantum computer that is truly useful and transformative is, at best, challenging because of a limit to our imagination of what could be the next great quantum algorithm and, at worst, hopeless if no new great quantum algorithm exists. Thus, quantum computing is a wonderful, scary, risky adventure, which we need to navigate with our eyes wide open.

2. Nexus between quantum computing and data science

Now I elaborate on the nexus between quantum computing and data science, as any meaningful overlap between these two galvanising research fields is exciting. Data science [14] concerns the full pipeline from capturing and maintaining data to processing and analysing these data and, finally, communicating or executing data-driven action [15]. Computing is a key part of this pipeline, not only for exploiting quantum speedup [16] or superior accuracy [17] but also for the possibility of storing information in superposition in a quantum random access memory [18] as well as considering cyberattacks on data storage and processing.

Secure data storage could involve quantum cryptography [19], and quantum-secure use of servers could need to exploit methods such as blind quantum computing [20] or quantum homomorphic encryption [21, 22]. The full impact of quantum information processing on data science is not yet known and needs extensive study. Here, we focus here strictly on the impact of quantum computing per se on data science.

I mentioned a little about quantum computing being about superpositions of information states and processing by unitary or near-unitary maps. As quantum computing is rather subtle in its nature, let us now take a moment to appreciate how it works. I would like to illustrate how quantum computing works by considering a quantum simulation, executable on a quantum computer, of Schrödinger’s famous example of a cat whose existential state can be in a life-death superposition, so to speak [23]. Our purpose is not to philosophise here but rather to recast Schrödinger’s cat paradox [24] as quantum information processing thereby illustrating how quantum computing works.

Here, I cast the cat’s quantum state of being alive ($b = 0$) or dead ($b = 1$) as being represented by a quantum state $|b\rangle$. Similarly, we can consider a radioactive nucleus, that decays by emitting an alpha (α) particle. α decay can be regarded as the emission of an α particle, equivalent to a He^{2+} ion, by quantum tunnelling out of the remaining nucleus [25]. The undecayed nucleus can be regarded as being an electromagnetic potential trap for the α particle, with the nuclear interior being attractive and the nuclear boundary being repulsive. We represent the undecayed

nucleus by $|0\rangle$ and the decayed nucleus by $|1\rangle$. Left alone the initial nuclear state $|0\rangle$ evolves into a superposition such as $|\pm\rangle := |0\rangle \pm |1\rangle$ over one half-life with choice of \pm indicating the coherence phase between not having decayed and having decayed.

To make clear the connection with quantum information processing, Schrödinger's cat paradox is depicted in Fig. 2 as a cat in a box with poison gas triggered by an α -particle detection. The correlation between α decay and the cat dying is shown in Fig. 2, with the resultant entangled nuclear-cat state shown in Fig. 2 as well. We can write this entangled state as $|00\rangle + |11\rangle$. This state is extremely important in quantum information science and is often called a Bell state. Just as we have resources like qubits and gates, this resource is sometimes called an ebit, short for a pair of maximally entangled qubits.

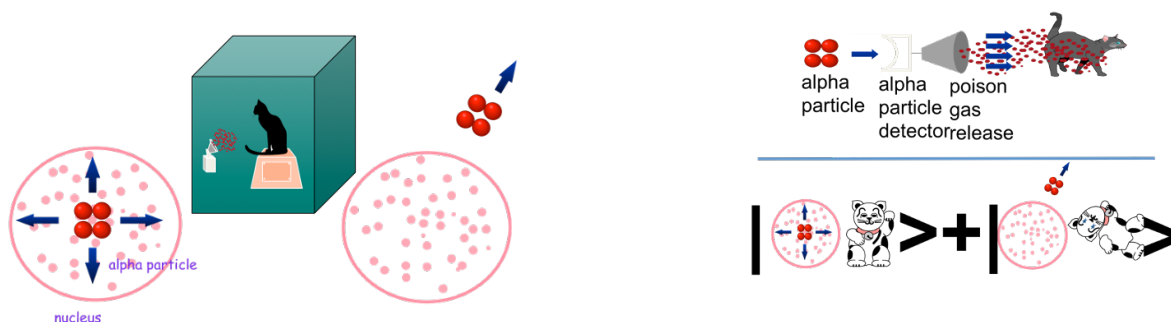


Figure 2. Above: the α particle is detected, which causes the release of a poison gas that kills the cat; in Dirac notation, the nuclear-cat entangled state is a superposition of the nucleus being undecayed and the cat being alive with the orthogonal state of the nucleus having decayed and the cat having died.

We can think of Schrödinger's cat paradox in terms of a quantum circuit. The input is initialised to $|00\rangle$ for an undecayed nucleus and a living cat. Then one waits for precisely one half-life to achieve $|0\rangle \mapsto |+\rangle$. If decay were a unitary process for one degree of freedom (not actually the case for nuclear decay but didactically convenient) then decay could be described by a quantum logic gate known as the Hadamard gate (H), whose unitary implies $H: |1\rangle \mapsto |-\rangle$ as well, which would mean that a decayed nucleus implausibly returns to its undecayed state.

The killing of the cat depicted in Fig. 2 is effecting the quantum logic gate known as CNOT, which stands for the controlled-not operation. The CNOT gate is equivalent to the XOR gate: flip the second qubit, i.e., apply the quantum NOT gate, if the first qubit is $|1\rangle$ but not if the first qubit is $|0\rangle$. Mathematically, for any two qubits in the computational basis, written as $|b\,b'\rangle$ for any bits b and b' , the mapping is $\text{CNOT} : |b\,b'\rangle \mapsto |b\,b' \oplus b\rangle$. The CNOT gate can be extended in an obvious way, as a NOT gate controlled by more than one qubit, such as the controlled-controlled-NOT, or Toffoli, gate $\text{CCNOT} : |b\,b'\,b''\rangle \mapsto |b\,b'\,b'b'' \oplus b\rangle$, which is valuable as a quantum version of repetition codes [26].

For Schrödinger's cat paradox, waiting a half-life means the first qubit is a superposition of $|0\rangle$ and $|1\rangle$, not specifically one of these two states, and the output is then the Bell state. Bell-state preparation, described by the procedure to manufacture the Schrödinger cat paradox, is a workhorse of some quantum computers, such as for ion-trap quantum computing [27].

The Hadamard gate plays a key role in quantum computing. An outer product of Hadamard gates maps a qubit string by $|00\cdots 0\rangle \mapsto |++\cdots +\rangle$, which is a superposition of all strings of qubits of given length. Some quantum algorithms commence this way, which is at the heart of

some references to quantum computing as parallelisation, but parallelisation is only part of the quantum computing story [2].

3. Building a quantum computer

Of course we would not be performing quantum computing using cats (although this terminology is used popularly for quantum computing [23], and the term “two-component Schrödinger cat states” is used for some type of superconducting quantum computing [28] although I prefer the term “entangled coherent state” for such states [29]) and instead resort to realisable quantum computers. I now list the current favourite platforms with examples of how the qubits are manifested [9, 10], and I sidestep the growing and intriguing areas of so-called continuous variable quantum computing [30, 31] and qudit quantum computing [32].

Photonic quantum computing was an early candidate and remains a strong contender. A single-rail qubit could be manifested as a superposition of no photon and one photon in a particular spatiotemporal mode; in contrast a dual-rail qubit is a single photon with the logical qubit inherent in the polarisation or which-path or which-time ($|early\rangle$ and $|late\rangle$, say) state of the photon. In magnetic resonance, the qubit can be the nuclear spin state, which can be controlled through hyperfine interactions with electronic qubits in the atom. Trapped ions can be superpositions of atomic excited and ground states, and controlled quantum logic gates can be achieved by exploiting collective motion of the ions leading to phononic sidebands of the atomic energy states. Neutral atomic qubits are also superpositions of ground and excited electronic states, with entangling two-qubit gates achievable by controlled collisions between atoms. Solid-state systems could be semiconductors, with the qubit inherent in the spin of a trapped electron, and superconducting quantum computing can involve quantized flux or superpositions of charge or using microwave photons in the resonator. A hybrid semiconductor-superconductor system could lead to topologically protected qubits.

Quantum computers can be realised according to different approach to processing quantum information. The gate ‘model’ is about making a universal primitive set of gates, such as H and CNOT discussed earlier, as well as another gate called T, which maps $|0\rangle \mapsto |0\rangle$ and $|1\rangle \mapsto e^{i\pi/4}|1\rangle$. Any multi-qubit unitary map can be decomposed efficiently into an efficiently scaling number of computer cycles, with each cycle comprising these single-qubit gates, including the identity gate, plus CNOT gates.

Measurement-based quantum computing, on the other hand, begins by entangling all qubits together and then carving out a circuit by measuring unwanted qubits in an appropriate basis. The remaining qubits are sequentially measured, with measurement basis determined by previous measurement outcomes; the final measurement outcomes provide the desired solution. Adiabatic quantum computing is about slow, controlled, continuous-time evolution of a Hamiltonian to realise approximately the target unitary map. Topological quantum computing involves braiding anyonic worldlines and could be realised via the fractional quantum Hall effect. This list describes universal quantum computers; purpose-built quantum computers such as quantum annealing—adiabatic or diabatic [33]—and boson sampling [34] enrich the set of quantum computers in development.

Quantum computer technology is quite impressive today, with quantum annealing reaching thousands of qubits and gate-based quantum computing having over a hundred qubits. Quantum computing primacy has arguably arrived [35, 36]. Certainly, quantum computing technology has surpassed what I imagined would be the status today, but much more is needed to make quantum computers genuinely useful.

4. Applied quantum algorithms for data science

A plethora of quantum algorithms has been developed [37]. Although factorisation is the most stunning of the quantum algorithms, near-term algorithms are the current focus. Two especially

enticing approaches to exploiting noisy intermediate-scale quantum (NISQ) computers [38] involve optimisation [39], which can use variational quantum algorithms [39, 40], and quantum machine learning [41]. In both cases, a quantum advantage remains an open question.

A business analysis of the potential for commercially viable quantum computing commercially is in combinatorial optimisation [42]. This analysis identifies three business verticals, which are niches that serve needs of a specific set of customers. They identify the verticals of financial portfolio management, computing lowest-energy configurations with applications to material design and to pharmaceuticals, and predicting rare failures for advanced manufacturing. The near-term gain would, one hopes, arise from exploiting quantum-inspired algorithms [43]. Intriguingly, even without a quantum advantage, a theoretical study suggests that a quantum economic advantage for companies developing quantum computers can emerge even if quantum computers do not offer a quantum computational advantage [44]. This argument for a quantum computational advantage, even if quantum computers do not pay off, is based on modeling a duopoly comprising a quantum computing company vs a classical computing company. Their model shows a “quantum economic advantage”, even without quantum computing advantage, partially through market creation.

Quantum machine learning aims to achieve quantum-enhanced machine learning, such as enhancing unsupervised, supervised, generative or transfer learning. Here, I summarise Wiebe’s list of key questions that a “quantum machine learner” should ask [45]. Wiebe relates quantum-enhanced machine learning specifically to certain potential advantages over using classical computing. He identifies four sought-after quantum enhancements to machine learning, namely, fewer computational steps for training or classifying, reducing sample complexity, generating a new data model, and using quantum optimisation for loss functions. As a second track of quantum machine learning research, Wiebe summarises how a quantum device could be natural for classifying or extracting features. To this end, he lists examples, namely, quantum Boltzmann machines and quantum neural networks, quantum algorithms such as for principal component analysis, and a quadratic time reduction for quantum-enhanced function inversion applied to data-training.

Consultants are involved in prognosticating commercial applications of quantum algorithms, and I share a little bit here of such endeavours. Gartner, Inc, anticipates various quantum computing potential applications [46]. They regard chemistry applications as first past the post, with such applications requiring somewhere around one hundred to two hundred qubits. As the number of qubits and quantum computational cycles increase, they forecast quantum optimisation and then quantum machine learning as being viable, followed by material science, with each of these application areas needing hundreds to thousands of qubits. These forecasts are in line with the business forecast discussed above [42]. Gartner’s analysis culminates with the expectation that, eventually, quantum computers will solve “unknown problems” requiring hundreds of thousands of qubits.

A November 2020 TechRepublic report summarises six experts’ predictions for 2021 [47] beginning with IBM predicting it will achieve 127 qubits, which indeed happened in 2021. The aforementioned Gartner, Inc, foresaw that cloud providers would incorporate quantum capability, which is so. KnowBe4, a company that focuses on security awareness training, predicted that quantum computing would break traditional public-key cryptography; did not happen, and I do not see this happening in the foreseeable future. Lux Research foreshadowed advances in optimising quantum hardware to reduce resource needs for running quantum algorithms. Finally, Forrester predicted that 2021 would be a trough of disillusionment, which is the third of five phases of the Gartner hype cycle [48, 49]. The hype cycle is a plot of maturity or visibility of an emerging technology as a function of time, and these five phases are the “technology trigger”, followed by the “peak of inflated expectations”, then the “trough of disillusionment”, subsequently the “slope of enlightenment” (waning interest with failure

to deliver with survival dependent on satisfying early adopters) and, finally, the “plateau of productivity”.

5. Conclusions and outlook

We have reviewed the idea and advantages of quantum computing, the connection with data science emphasising the computational side such as optimisation and machine learning, and took an excursion into what consulting firms are saying. Now we take stock of all this information to figure out where quantum computing is today and how it affects data science and its applications and verticals.

Clearly, quantum computers are a game changer for computing, at least at a conceptual level. The idea that hard problems can be solved efficiently on a quantum computer shows that quantum computing is more powerful, at least in one type of application. That quantum computing is *provably* quadratically superior, by some measure, at inverting functions, also points to quantum computing being superior in some way to classical computing.

On the other hand, quantum computers today are small and noisy, and, even if they were large and coherent, their application to data science entails only speculative advantages. Whether or not quantum computers are advantageous for data science might only be resolved by building quantum computers, testing algorithmic performance on such computers, and seeing whether an advantage has been found or not; this empirical approach is already ongoing in the field. Fortunately, quantum-inspired algorithms, which typically arise as ways to challenge a purported quantum advantage by solving the same problem as well or better on a classical computer, are giving beneficial computational results at present that arguably would not be happening if quantum computing research were not occurring.

Given the unknowns of quantum computing, including how much quantum computers can scale up and whether we will have good algorithms that exploit the potential of quantum computing for data science, an approach I call “quantum-aware, quantum-ready, quantum-active” is prudent. This approach means that, at this stage, we stakeholders in quantum computing, including data scientists, need to be aware of when and how quantum computing could change the game. We maintain awareness by testing how well existing quantum computers or simulators thereof perform on problems of interest. If and when quantum computing matters for the application at hand, it is time to become quantum-ready, which means train up on the technology and prepare for transitioning to quantum computing so that, when the time comes, we are ready to use it without delay. Finally, in the optimistic scenario that quantum computing has become a disruptive technology whose adoption is necessary, we have reached the quantum-active stage where data scientists need to be fully engaged in using quantum computing to solve problems. We need to remember that quantum computing is a high-risk high-reward venture and treat it accordingly.

Acknowledgments

This project has been funded by the Alberta Government and by NSERC. Thanks to S. L. Sanders who prepared the figures over two decades ago; those figures have been valuable for my presentations on Schrödinger’s cat ever since but have not been published before.

References

- [1] Nielsen M A and Chuang I L 2010 *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge MA: Cambridge University Press)
- [2] Aaronson S 2013 *Quantum Computing since Democritus* (Cambridge MA: Cambridge University Press)
- [3] Dirac P A M 1939 *Math. Proc. Cambridge Philos. Soc.* **35** 416–418
- [4] Gottesman D 2020 *Commun. ACM* **64** 105 ISSN 0001-0782
- [5] Feynman R P 1982 *Int. J. Theor. Phys.* **21** 467–488

- [6] Shor P W 1994 Algorithms for quantum computation: discrete logarithms and factoring *Proc. 35th Annual Symp. on Foundations of Computer Science* pp 124–134
- [7] Grover L K 1996 A fast quantum mechanical algorithm for database search *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing STOC '96* (New York, NY, USA: Association for Computing Machinery) p 212–219
- [8] Heim B, Rønnow T F, Isakov S V and Troyer M 2015 *Science* **348** 215–217
- [9] Sanders B C 2017 *How to Build a Quantum Computer* (Bristol: IOP Publishing)
- [10] Sanders B C 2020 Building a quantum computer (invited) *2020 ACM/IEEE International Workshop on System Level Interconnect Prediction (SLIP)* pp 1–4
- [11] MacQuarrie E, Simon C, Simmons S and Maine E 2020 *Nat. Rev. Phys.* **2** 596–598
- [12] Christensen C M 1997 *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail* The Management of Innovation and Change (Boston: Harvard Business Review Press)
- [13] Wu L, Wang D and Evans J A 2019 *Nature* **566** 378–382
- [14] Kotu V and Deshpande B 2019 *Data Science: Concepts and Practice* 2nd ed (Cambridge MA: Morgan Kaufmann)
- [15] Muller M, Lange I, Wang D, Piorkowski D, Tsay J, Liao Q V, Dugan C and Erickson T 2019 How data science workers work with data: Discovery, capture, curation, design, creation *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York: Association for Computing Machinery) pp 1–15
- [16] Rønnow T F, Wang Z, Job J, Boixo S, Isakov S V, Wecker D, Martinis J M, Lidar D A and Troyer M 2014 *Science* **345** 420–424
- [17] Willsch D, Willsch M, De Raedt H and Michielsen K 2020 *Computer Physics Communications* **248** 107006
- [18] Blencowe M 2010 *Nature* **468** 44–45
- [19] Pirandola S, Andersen U L, Banchi L, Berta M, Bunandar D, Colbeck R, Englund D, Gehring T, Lupo C, Ottaviani C, Pereira J L, Razavi M, Shaari J S, Tomamichel M, Usenko V C, Vallone G, Villoresi P and Wallden P 2020 *Adv. Opt. Photon.* **12** 1012–1236
- [20] Fitzsimons J F 2017 *npj Quantum Inf.* **3** 23
- [21] Broadbent A and Jeffery S 2015 Quantum homomorphic encryption for circuits of low t-gate complexity *Advances in Cryptology—CRYPTO 2015* ed Gennaro R and Robshaw M (Berlin: Springer) pp 609–629
- [22] Dulek Y, Schaffner C and Speelman F 2016 Quantum homomorphic encryption for polynomial-sized circuits *Advances in Cryptology—CRYPTO 2016* ed Robshaw M and Katz J (Berlin: Springer) pp 3–32
- [23] Gribbin J 2014 *Computing with Quantum Cats: From Colossus to Qubits* (Prometheus) ISBN 9781616149215
- [24] Schrödinger E 1935 *Naturwissenschaften* **23** 807–812
- [25] Gamow G 1928 *Zeitschrift für Physik* **51** 204–212
- [26] Bacon D 2006 *Phys. Rev. A* **73**(1) 012340
- [27] Mølmer K and Sørensen A 1999 *Phys. Rev. Lett.* **82**(9) 1835–1838
- [28] Mirrahimi M, Leghtas Z, Albert V V, Touzard S, Schoelkopf R J, Jiang L and Devoret M H 2014 *New Journal of Physics* **16** 045014
- [29] Sanders B C 2012 *J. Phys. A: Math. Theor.* **45** 244002
- [30] Lloyd S and Braunstein S L 1999 *Phys. Rev. Lett.* **82**(8) 1784–1787
- [31] Bartlett S D, Sanders B C, Braunstein S L and Nemoto K 2002 *Phys. Rev. Lett.* **88**(9) 097904
- [32] Wang Y, Hu Z, Sanders B C and Kais S 2020 *Front. Phys. (Lausanne)* **8** 589504
- [33] 2021 *Nat. Rev. Phys.* **3** 466–489
- [34] 2013 *Theory of Computing* **9** 143–252
- [35] Dalzell A M, Harrow A W, Koh D E and La Placa R L 2020 *Quantum* **4** 264
- [36] Sanders B C 2021 *Physics* **14** 147
- [37] Jordan S 2021 Quantum algorithm zoo URL <https://quantumalgorithmzoo.org>
- [38] Preskill J 2018 *Quantum* **2** 79
- [39] Moll N, Barkoutsos P, Bishop L S, Chow J M, Cross A, Egger D J, Filipp S, Fuhrer A, Gambetta J M, Ganzhorn M, Kandala A, Mezzacapo A, Müller P, Riess W, Salis G, Smolin J, Tavernelli I and Temme K 2018 *Quantum Science and Technology* **3** 030503 URL <https://doi.org/10.1088/2058-9565/aab822>
- [40] Cerezo M, Arrasmith A, Babbush R, Benjamin S C, Endo S, Fujii K, McClean J R, Mitarai K, Yuan X, Cincio L and Coles P J 2021 *Nat. Rev. Phys.* **3** 625–644
- [41] Schuld M, Sinayskiy I and Petruccione F 2015 *Contemp. Phys.* **56** 172–185
- [42] Bova F, Goldfarb A and Melko R G 2021 *EPJ Quantum Technology* **8** 2
- [43] Arrazola J M, Delgado A, Bardhan B R and Lloyd S 2020 *Quantum* **4** 307
- [44] Bova F, Goldfarb A and Melko R G 2022 Quantum economic advantage Working Paper 29724 National Bureau of Economic Research
- [45] Wiebe N 2020 *New J. Phys.* **20** 091001

- [46] Gartner 2019 The CIO's guide to quantum computing URL tinyurl.com/2s4dvjae
- [47] Greig J 2020 6 experts share quantum computing predictions for 2021 URL tinyurl.com/y79vyds9
- [48] Fenn J and Raskino M *Mastering the Hype Cycle: How to Choose the Right Innovation at the Right Time* (*Gartner*) (Boston: Harvard Business Review Press)
- [49] Steinert M and Leifer L 2010 Scrutinizing Gartner's hype cycle approach *Technology Management for Global Economic Growth (PICMET)*, *2010 Proceedings of PICMET '10* pp 1–13