Rent’s rule and extensibility in quantum computing

D.P. Franke a,⁎, J.S. Clarke b, L.M.K. Vandersypen a,b, M. Veldhorst a

a QuTech and Kavli Institute of Nanoscience, Delft University of Technology, P.O. Box 5046, Delft 2600 GA, The Netherlands
b Components Research, Intel Corporation, 2501 NE Century Blvd, Hillsboro, OR 97124, United States

A R T I C L E   I N F O

Article history:
Received 8 June 2018
Revised 3 December 2018
Accepted 12 February 2019
Available online 14 February 2019

A B S T R A C T

Quantum computing is on the verge of a transition from fundamental research to practical applications. Yet, to make the step to large-scale quantum computation, an extensible qubit system has to be developed. In classical semiconductor technology, this was made possible by the invention of the integrated circuit, which allowed to interconnect large numbers of components without having to solder to each and every one of them. Similarly, we expect that the scaling of interconnections and control lines with the number of qubits will be a central bottleneck in creating large-scale quantum technology. Here, we define the quantum Rent exponent \( p \) to quantify the progress in overcoming this challenge at different levels throughout the quantum computing stack. We further discuss the concept of quantum extensibility as an indicator of a platform’s potential to reach the large quantum volume needed for universal quantum computing and review extensibility limits faced by different qubit implementations on the way towards truly large-scale qubit systems.

© 2019 Published by Elsevier B.V.

1. The tyranny of numbers

One of the most significant advances in the field of quantum computation has been the invention of quantum error correction (QEC) [1–3]. While quantum bits (qubits) are delicate systems, these algorithms can enable fault-tolerant quantum computation with sophisticated correction codes tolerating error rates of up to 1% [2]. Similar values are already achieved or within reach for experimentally observed qubit infidelities across a range of different platforms [4–11]. However, a trade-off between the tolerated error rates and the number of qubits has to be made. Quantum error correction can lead to an overhead between \( 10^3 \) and \( 10^4 \) physical qubits per logical qubit [2,12], such that millions or even billions of physical qubits will be required for practical applications. To host and control this daunting number of qubits, formidable requirements have to be met by different elements of the system, including interconnects, control electronics and quantum software. It is therefore essential to develop an extensible approach to the hardware and software throughout the full quantum computing stack.

Today, experimental qubit systems make use of one to a few control terminals per physical qubit (component), which means that the total number of control terminals, \( T \), scales linearly with the number of internal components, \( g \) [4–11]. This linear scaling implies an unimpaired increase of \( T \) for large \( g \), a situation reminiscent of the late 1950s, where engineers were working with electrical systems containing many component, each requiring soldering to numerous others. Ian Ross, president of Bell labs, stated: ‘As you built more and more complicated devices, like switching systems, like computers, you got into millions of devices and millions of interconnections. So what should you do?’ [13]. Jack Morton, vice president of device development at Bell Labs, referred to this situation as ‘the tyranny of numbers’ [14]. He believed a solution would be to search for devices that could perform multiple tasks, such that the total number of components could be reduced. The real breakthrough was made, among others, by Jack Kilby of Texas Instruments, and Robert Noyce of Fairchild Semiconductor, who improved the Integrated Circuit (IC) to an industrial level. Integrated circuits circumvented the tyranny of numbers and were faster, better, and cheaper.

2. Rent’s rule

An interesting trend between the number of \( T \) and \( g \) on an IC was observed in the 1960s by E.F Rent, IBM [15]. Landman and Russo described the correlation using the empirical formula

\[
T = tg^p
\]

(1)

which they called Rent’s rule [16], and which was later formally justified [17]. Intuitively, \( t \) refers to the number of connections required for each internal component \( g \). The Rent exponent \( p \) ac-
counts for the level of optimization, such that with no optimization \( p = 1 \), while, for example, the X86 series of Intel microprocessors have \( p = 0.36 \) [18]. Guided by the history of classical ICs, we envision that quantum systems, will experience a similar down-scaling in \( p \) due to similar motivations.

To exemplify the corresponding situation in few qubit experiments, a typical measurement setup for quantum dot spin qubits is shown schematically in the left part of Fig. 1. Here, the qubits are controlled by lithographically defined gates and a microwave delivery antenna which fan out to bond pads that are wire-bonded to a chip carrier. Then, the lines are filtered and are wired through the different stages of the dilution refrigerator that keeps the device at its millikelvin operating temperature. Each line is then individually connected to the outputs of low noise digital analogue converters (DAC) and arbitrary waveform generators (AWG) that are used to control the electronic potential landscape. Adding another qubit to the present device would require an additional two gates (corresponding to \( t = 2 \)) and two bonding wires, as well as two additional AWG and DAC channels. This linear scaling is described by an exponent \( p = 1 \) at all levels of the experiment, as indicated in the central column of Fig. 1.

The limitations posed by this scaling law as well as the possible solutions could be very different at different levels of the quantum computer stack. We therefore propose to define several scaling exponents \( p \). On the lowest level, \( p_g \) describes the number of gates per qubit. Here, typical limitations will be due to geometric restrictions and the limited number of gate layers. For example, at least \( \sqrt{3}/2 \) gate layers are necessary to directly address \( g \) qubits in a 2D array. In close analogy to Rent’s rule for IC terminals, \( p_{RT} \) describes the number of IO terminals of the chip. Clear limitations are given by the size of these terminals and the space on the chip and, as with classical processors, the number of connections will likely be limited to a few thousand. The third exponent \( p_{RT} \) then refers to the number of wires leaving the cryostat. Here, constraints will, for example, be posed by the geometry of the dilution refrigerator and the heat transport trough such wires. As each of the exponents includes the effect of optimization achieved on lower-lying levels,

\[
 p_g \geq p_{IO} \geq p_{RT}. \tag{2}
\]

At the current stage, the experimental qubit implementations across all platforms make use of a direct control of each qubit, corresponding to \( p_{RT} = 1 \). This straight-forward implementation at the few-qubits level provides maximum flexibility and control, such that the individual calibration of the unique qubits and adaptations to inhomogeneities or defects are possible. While this concept reduces the demands on the fabrication uniformity, it clearly will not be able to support the large numbers needed for practical error correction. Therefore, schemes of shared control lines have to be developed and implemented at different levels. Several concepts that have been suggested addressing these issues are summarized in the right part of Fig. 1. As proposed in Refs. [19–22], two-dimensional arrays and crossbar gating schemes can help to achieve notable improvements in \( p_g \). While, again, two gates are used to control a qubit (\( t = 2 \)), the exponent \( p \) can be around 0.5 here. The benefit for overcoming the interconnect bottleneck is substantial: with \( p \sim 0.5 \), one million qubits require not of order one million wires (infeasible) but one thousand wires (feasible). It will be a milestone if such architectures can be realized experimentally. As a promising way to further reduce the number of IO terminals and hence achieving \( p_{IO} \sim p_g \), cryogenic electronics that can implement on-chip control circuits are pursued [23,24]. To increase the available cooling power to a level compatible with the dissipation in such circuits, an increase in the qubit operating temperature could be a central step [25,26]. Furthermore, local microelectronics and logic circuits used to control error correction cycles or other feedback could be implemented to reduce latencies and trivial communication with room temperature equipment. As a result, \( p_{RT} < p_{IO} \). In Fig. 1 this is illustrated by the boxes labeled quantum instruction set architecture (QISA) or microarchitecture (MA).

It is also worth mentioning that such concepts for enhancing the scalability of the wiring will have a direct influence on the way the qubits are operated. While in case of a direct control the highest possible flexibility is maintained, a reduced number of control lines will likely result in an overhead. As suggested in the schematic quantum circuit in Fig. 1, algorithms will have to be compiled differently, since the parallel application of arbitrary pulses to different qubits will be constrained as a result of shared gates. In most cases, this will lead to slower qubit operation giving rise to the question whether this limitation will influence the overall capabilities of a quantum processor. Similarly, the operation of error correction schemes will become more challenging if such a limited control has to be considered. For the example of the crossbar structure proposed in Ref. [22], it has
been shown that surface code operation can indeed be implemented to create a logical qubit with a very low logical error rate even under the limits imposed by shared control [27]. Another particularity of such schemes that remains to be investigated is the influence of correlated error due to the shared control gates. While creating and connecting multiple logical qubits will be a challenging task with shared control, we are optimistic that the advantages of shared control schemes for the extensibility of the qubit devices outweigh the possible restrictions and that they will play a central role in the development of large-scale quantum processors.

3. Quantum extensibility

We want to broaden the above discussion to a more general view on the extensibility of proposed and realized qubit systems. In an attempt to define a metric that can describe the usefulness of a quantum chip by not just the number of qubits but also considering the quality of their implementation, the quantum volume \( V_Q \) was introduced by Bishop et al. [28,29]. This metric is a function of the number of qubits \( N \) and the circuit depth \( d \) of their operations. Here, \( d \) is given by the number of operations that can be performed before, on average, an error will occur. It describes to what extent the system can use entanglement and profit from a quantum speed-up [29]. The most straightforward definition of \( V_Q \) as the product of \( d \) and \( N \) already gives a useful metric, but it is mostly meaningless when either of the two factors are small. Therefore, the volume is defined as

\[
V_Q = \min(N, d)^2.
\]

In Fig. 2, color plots of the quantum volume as a function of \( N \) and \( d \) are shown using logarithmic axes. We discriminate four different regimes as given in the legend of Fig. 2. The quantum volume of the experimental qubit implementations today is still small and below the threshold where classical computers can still efficiently simulate the quantum system. This is indicated by a red background color. The yellow background describes quantum systems that might be too complex to be fully simulated by classical computer but are not yet powerful enough to harness the full potential of quantum computing. For these, the term quantum supremacy has been coined [30,31], which serves as a benchmark in the development of early stage quantum processors. Where exactly this line is drawn is still under active discussion [32,33], and specific problems that are designed to be hard on classical computers, but could be solved already by small scale quantum computers have been proposed [34]. From a different viewpoint, this class of devices is also referenced as Noisy Intermediate-Scale Quantum (NISQ) technology [35] and while there might be a limited range of applications, they are mostly considered an intermediate step towards more powerful systems. The two green regions correspond to a quantum volume large enough to allow for relevant quantum simulations [36,37] or even fault-tolerant universal quantum computing. Where these regions can be reached, it is widely believed that the impact on computing and many other disciplines in science will be revolutionary.

In Fig. 2(a), the quantum volume is shown for the area in \( N-d \) space that is expected to be covered for four fictional quantum platforms. Furthermore, the initial state as set by a state-of-the-art system in that platform is shown (\( A_1 \) etc.), as well as the maximum quantum volume region that can be achieved (\( C_{\text{max}} \) and \( D_{\text{max}} \)). In the example of platform A, a rather large number of qubits could be expected to be reached, while their fidelity faces stricter limits and will limit the development of a quantum processor. Platform B, on the other hand, is already in an initial state of better circuit depth, but the number of qubits that can be realized in this approach will restrict the maximum quantum volume. System C will be able to reach beyond quantum supremacy into the NISQ era of quantum applications, but it still lacks the ability to reach a quantum volume large enough for universal quantum computing. Only system D can be expected to reach the green regions of large quantum volume should be considered as a system with real potential for quantum computing.
Quantum computer architectures are envisioned as layered control stacks [38,39], illustrated in Fig. 2(b). We expect extensibility limits to occur at all layers of the stack. This ranges from the most fundamental level of the actual physical implementation of qubits that suffers from decoherence limiting the fidelity of qubit operations to more practical limits such as the classical computing power needed to analyze error syndrome measurements in quantum error correction. Other examples include the interconnect bottleneck discussed above, the available cooling power for low temperature operation, space on a chip, or timing issues due to delay in control lines. While these issues are sometimes addressed in a highly speculative way, for example by referring to future developments in fabrication, an honest and preferably quantitative way of giving extensibility limits would be highly beneficial to the field.

We therefore propose that, in addition to estimating the maximum quantum volume \(V_{Q,\text{max}}\), a system is described by its extensibility \(X_R\) with respect to a resource \(R\). In a generalization of the quantum version of Rent’s rule, we assume that the use of most resources can be described by a power law

\[
R(V_Q) = R_1 \left( \frac{V_Q}{V_Q^*} \right)^{\frac{1}{X_R}},
\]

where \(V_Q^*\) is the initial quantum volume of the system, \(R_1\) is the initial use of the resource \(R\), and \(X_R\) is the extensibility of the system with respect to \(R\). In words, the extensibility of a system describes at what expense its quantum volume can be increased. As would be expected, a large extensibility means that a larger quantum volume can be achieved with only a small increase in resources, while for small \(X_R\) only a large increase in \(R\) will allow to expand \(V_Q\). In the extreme case of exponential scaling of \(R\), we define \(X_R = 0\). Exponential scaling has in our definition thus zero extensibility, consistent with Feynmann’s original view on quantum simulation stating that the number of elements should not explode with the space-time volume of the physical system [40]. For direct comparison and the determination of the most relevant resources, (4) is best written in relation to the resource limit \(R_{\text{max}}\) of \(R\), such that

\[
\frac{R(V_Q)}{R_{\text{max}}} = R_1 \left( \frac{V_Q}{V_Q^*} \right)^{\frac{1}{X_R}},
\]

where we have defined \(r_1 = R_1/R_{\text{max}}\).

Both the initial use of resources expressed by \(r_1\) and the extensibility \(X_R\) are relevant to quantify the capability of a system to reach high quantum volumes. This becomes clear from Fig. 2(c), where the scaling of four fictional resources \(R_1, \ldots, R_4\) with \(V_Q\) is shown. Here, \(R_4\) has the highest \(r_1\) and is therefore closest to its resource limit at the initial state of low quantum volume. However, because of the high extensibility of the system with respect to \(R_4\) (\(X_{R4} \gg 1\)), this resource will not enforce a relevant limit to the development of the quantum volume. For quantum dot qubits, a similar behavior could be assumed for the effort going into the fabrication of the gate structures. While this effort is already quite high initially (high \(r_1\)), it is likely that once optimized the fabrication can rather easily be expanded to the creation of large numbers of qubits (high \(X\)). In contrast, a resource with low \(r_1\) can still limit the system’s development if the corresponding extensibility is low. This is the case for \(R_2\) in Fig. 2(c), which is far from its limit at the initial state, but grows quickly as the system is extended to larger \(V_Q\). It is therefore the first to reach the resource limit and will dictate the maximum quantum volume that can be realized. For the resource \(R_3\), on the other hand, \(r_1\) is so small that \(R_3\) is always far from limiting the development, even though \(X_{R3}\) is similar to \(X_{R4}\). An example for such a resource could be the area on a semiconductor chip occupied by quantum dot qubits. In Fig. 2(c), the resource that is most strongly limited at any point in the development is emphasized by a thick black line. The slope of this line in the double logarithmic plot defines a local extensibility which could possibly be used to classify the short-term development of a system. The overall extensibility \(X\) of the platform, however, is best described by the extensibility with respect to the critically limiting resource, or \(X = X_{R1}\) for the system described by Fig. 2(c).

For most resources, \(R_{\text{max}}\) and hence the relative initial state \(r_1\) are not sharply defined or can at least be bent with some effort. Therefore, the extensibility at this maximum quantum volume can give a valuable insight whether a system will be able to profit from such optimizations. This becomes clear, when (4) is solved for \(V_{Q,\text{max}}\) such that

\[
V_{Q,\text{max}} = \left( \frac{R_{\text{max}}}{R} \right)^{X_{R1}} V_Q^*,
\]

where \(R\) refers to the critically limiting resource. For a system with low extensibility, changes to \(R_{\text{max}}\) will barely influence \(V_{Q,\text{max}}\), while for a high \(X\), even a small change in \(R_{\text{max}}\) will translate to a large change in the achievable quantum volume \(V_{Q,\text{max}}\). It is also clear that even small changes of \(X_{R1}\) will have a large influence on \(V_{Q,\text{max}}\). Therefore, studying and understanding the extensibility graph of a system already at an early stage of the development is crucial for judging the promise of a particular quantum technology.

3.1. Quantum chip: Qubit platforms

Among the broad range of physical systems that are developed for quantum computing [41], the current prospects of reaching a large quantum volume vary significantly. We will therefore specifically address some of the most prominent approaches to the implementation of qubits.

The qubit platforms that bears most similarities with traditional semiconductor technology is that of semiconductor quantum dot qubits defined in silicon. In fact, the similarity of classical transistors and quantum dot qubits suggests that the fabrication of millions of such qubits will be feasible in the near future. Therefore, the question of the extensibility of the qubit control and the scaling of interconnects as described by the quantum version of Rent’s rule are of pressing relevance. Following early proposals of two-dimensional architectures [42], implementations of a shared control based on crossbar designs have been proposed for the related system of donor atoms [20] and using complementary metal-oxide semiconductor (CMOS) control elements [21]. However, these layouts assume fabrication technology that is far out of reach of that of today’s cutting edge semiconductor technologies. In contrast, a recent proposal where quantum dots are defined using shared gates arrange in a cross-bar architecture can be realized using today’s methods [22]. Still, as of today, only modest numbers of qubits are operated [43] and challenges in fabrication uniformity and control need to be overcome before a large-scale quantum chip becomes feasible.

Having seen remarkable improvements in the qubit properties [44], the platform that will likely be the first to reach the NISQ era of quantum applications is that of superconducting qubits. Devices with ∼50 qubits can already be fabricated [45,46] and concrete steps are taken to achieve the specific goal of reaching the quantum supremacy threshold [47,48]. The extensibility beyond the NISQ era is, however, less clear and the implementation of the large number of qubits that will be needed for meaningful quantum computation will pose new challenges. One of these is the physical size of the resonators (on the order of mm), which limits the number of qubits that can be fabricated on a wafer. Designs that address the extensibility limit that is posed by the limited number of high frequency connection to the quantum chip have been proposed. This includes a design for a surface code unit cell for both quantum hardware and control signals [49] as well as cavity grids [19].
Trapped atomic ions are arguably still the most advanced qubit platform today [50,51], but do face serious challenges in their extensibility. In contrast to the other two systems discussed above, they cannot directly be implemented using semiconductor fabrication technology, which could be argued to be the only technology that has been proven capable of the necessary large numbers of components. Therefore, approaches to implement ion traps on semiconductor chips have been realized [52] and are considered a possible route for scaling [53–55]. An advantage of the trapped ion approach is that ions in different traps can be entangled with each other via room temperature photonic links in the optical domain [54], allowing a modular approach that relases the interconnect bottleneck. For this to be practical entanglement generation rates have to be increased by order of magnitudes but if this can be realized, the advantage over monolithic quantum circuits is that not all control wires have to interface to a single substrate. Given that optical links allow well-separated modules, even a Rent exponent $p_R = p_{10} = 1$ may be acceptable, although higher in the stack economic considerations may still enforce $p$ substantially below 1. The higher operation temperature of ions, which can operate at room temperature and are typically only cooled by liquid nitrogen [55] relaxes other architecture restrictions faced by quantum dot qubits and in particular by superconducting qubits. Nevertheless, the physical size (amounting to about $100 \times 100$ $\mu$m$^2$ for $2 \times 10^5$ ions [55]) could make such implementations impractical. Many of the same considerations apply to qubits represented by spins bound to color centers in solids, such as nitrogen-vacancy centers in diamond [56–60].

As a final example, there have been proposals for creating quantum circuits based on topological qubits [61]. These could, when they can be realized in the future, profit from certain protected states [52]. Depending on the details of the implementation, the number of physical qubits could potentially be reduced compared to QEC concepts in other systems. This may relax requirements on the number of physical qubits, such that a lower extensibility limiting the fabrication of physical qubits in this platform may still allow for practical quantum computation.

3.2. Higher in the quantum computing stack

A central advantage of a systems view of a quantum computer is that higher levels can, to a certain extent, be developed independently of the physical qubit implementation [38,39]. This way, extensibility limits that occur here can already be addressed and specific solutions will likely be beneficial for most platforms. To avoid exploding numbers of off-chip connections, some parts of the control electronics can possibly be integrated with the qubit device. Even the minimum logic signals needed to apply the necessary gates for quantum error correction quickly lead to a bandwidth that will be challenging to realize [38], meaning that a basic part of the QEC logic would have to be integrated on-chip. Here, spatial and thermal budgets will play a central role. Other clear resource limits are given by the availability of the classical computing power and memory needed to process error syndrome measurements. Also, the resources for technically more difficult operations such as microwave control and other fast pulses put constraints on a scalable classical control. The parallel and routed application of electrical signals is therefore necessary and dictates the way quantum gates can be applied to the qubits [63–65].

In addition to such limitations that occur at higher levels of the stack, the efficiency of the QEC is directly related to the extensibility that is achieved at lower levels, such as the quantum chip. There, the use of resources is mostly connected to the number and fidelity of physical qubits. A platform that is capable of a particularly efficient form of QEC can therefore benefit from relaxed restrictions and advances made to these codes will be directly reflected in a higher extensibility at the lower levels of the stack. Furthermore, in many cases trade-offs between different resources will have to be made. One example is the concept of shared control discussed above. While the implementation of shared control can significantly improve the extensibility of quantum dot qubits with respect to the number of interconnects and chip terminals, this comes with restrictions to the parallel operation of qubits, which directly influences resources such as the number of operations within the qubit coherence time.

4. Discussion

Feynman argued in his seminal work *simulating physics with computers:* The rule of simulation that I would like to have is that the number of computer elements required to simulate a large physical system is only to be proportional to the space-time volume of the physical system. I don't want to have an explosion. That is, if you say I want to explain this much physics, I can do it exactly and I need a certain-sized computer. If doubling the volume of space and time means I'll need an exponentially larger computer, I consider that against the rules (I make up the rules, I'm allowed to do that) [40].

Here, we have tried to capture this vision by defining Rent exponents across a quantum accelerator stack and by broadening the discussion to include all resources needed for a future quantum computer. The effort required for a platform to reach a certain computing power can be revealed by quantum extensibility graphs. Where exactly the threshold to useful quantum computing is reached will depend on the development of efficient quantum algorithms and will therefore remain a subject of active research. Similarly, whether the quantum volume as defined above truly reflects the usefulness of a system could depend on the particular use case. In predictions for running Shor’s algorithm using quantum error correction on a large scale quantum computer, typically around $N \sim 5000$ logical qubits are used [2,66], suggesting $V_Q \sim 10^7$ as an order of magnitude for relevant computation. In any case, it is clear that the quantum volume $V_Q$ will have to grow by many orders of magnitude to get from the current state to a volume capable of useful quantum computation. It is therefore likely that for the critical resources only an extensibility $X_R > 1$ (corresponding to a sublinear scaling) can support such growth.

All platforms will face great challenges in achieving the high extensibility that will allow the development of large-scale quantum computation. For quantum dot qubits in silicon, these challenges bear many parallels to the development of classical integrated circuits and the relation of chip terminals to the number of components can hence be expected to be a central metric for extensibility in this platform. Similar metrics can likely be found in other platforms and should be identified to motivate and focus future research. Only if these critical extensibilities can be optimized and non-zero $X_R$ are achieved for all components in a quantum computer, Feynman’s vision of harnessing the computational power of an exponentially growing number of quantum states in polynomial time and space will become a reality.

The authors would like to thank W. Lawrie for thorough reading of the manuscript. M.V. acknowledges support by the Netherlands Organization of Scientific Research (NWO) VIDI program.

Supplementary material

Lieven Vandersypen received a M.Sc. in Mechanical Engineering from KU Leuven (1996) and a Ph.D. in Electrical Engineering from Stanford University (2001), for the first experimental realizations of quantum algorithms. He then moved to TU Delft for a postdoc and since 2007 as Antonie van Leeuwenhoek professor, pioneering electron spin qubits in semiconductor quantum dots. He co-founded QuTech in 2013, took a part-time position with Intel in 2015 and is co-director of the Kavli Institute of Nanoscience Delft since 2016. Prof. Vandersypen received the Nicholas Kurti European Science Prize (2008) and the IUPAP Young Scientist Prize for Semiconductor Physics (2008).

Menno Veldhorst is a team leader in fault-tolerant quantum computing and roadmap leader QuTech Academy at QuTech, TU Delft, The Netherlands. He holds a Ph.D. degree from the University of Twente, The Netherlands (2012), which he received cum laude for his research on coupling superconducting and topological systems. From 2012 to 2016 he carried out postdoctoral research at UNSW, Sydney, Australia, where he demonstrated universal quantum logic with silicon qubits. His current research interests include silicon and germanium quantum computing with semiconducting and superconducting elements. More information and contact info: https://qutech.nl/person/menno-veldhorst/.