The Engineering Challenges in Quantum Computing

C. G. Almudever*, L. Lao*, X. Fu*, N. Khammassi*,I. Ashraf*, D. Iorga*, S. Varsamopoulos*,
C. Eichler†, A. Wallraff†, L.Geck**, A. Kruth**, J. Knoch††, H. Bluhm‡, K Bertels*

* Computer Engineering Lab and QuTech, Delft University of Technology
† Quantum Device Lab, ETH Zurich
‡ JARA-FIT Institute for Quantum Information, RWTH Aachen University and Forschungszentrum Jülich
††Institute of Semiconductor Electronics and JARA-FIT, RWTH Aachen University
**Central Institute of Engineering, Electronics and Analytics, ZEA-2: Electronic Systems, Forschungszentrum Jülich

Abstract—**Quantum computers may revolutionize the field of computation by solving some complex problems that are intractable even for the most powerful current supercomputers. This paper first introduces the basic concepts of quantum computing and describes what the required layers are for building a quantum system. Thereafter, it discusses the different engineering challenges when building a quantum computer ranging from the core qubit technology, the control electronics, to the microarchitecture for the execution of quantum circuits and efficient quantum error correction. We conclude by discussing some compiler and programming issues relative to quantum algorithms.**

I. Introduction

A quantum computer holds the promise to solve efficiently some classes of computational problems that are intractable for a classical computer by using quantum algorithms that exploit fundamental quantum phenomena such are superposition and entanglement [1]. The most famous example is the factorisation of large numbers using Shor's algorithm, which is exponentially faster than its best classical counterpart By running this algorithm on a quantum computer we could factorize, for instance a 2000-bit number, in a bit more than 1 day as compared to the 100 years a data center bigger than Germany would need [2]. However such a quantum computer would require around millions or even billions of physical quantum bits or *qubits* [3], [4]. That large number of qubits mainly comes from the need to deal with the *fragility* of the quantum technology and to make quantum systems robust against errors. Qubits suffer from *decoherence* meaning that the information stored in the qubits is lost due to the interaction with the environment, leading to gate error rates of around $\sim 10^{-2}$. However, quantum systems can be protected and recovered from such errors by using *quantum error correction* (QEC) and *fault tolerant (FT) computations* if the gate error rates are below a certain threshold [5]. These procedures will be essential for any quantum computer but they will also dramatically increase the amount of qubits required for computation. Building a full scale quantum computer is therefore directly influenced by the above observations and basically fall apart in two challenges: to increase the fidelity of quantum operations and to scale the control infrastructure of large numbers of qubits

in the range of millions. These challenges are discussed in the remainder of this paper.

The paper is organized as follows. We begin by introducing the basics of quantum computing in Section II. Section III shows an overall quantum system architecture. Section IV, gives an overview of the different implementations of quantum processors. Section V, discusses the requirements and challenges of the control electronics. Sections VI and VII, describe a possible architecture for the quantum accelerator and the compiler infrastructure, respectively. Section VII concludes the paper.

II. Background on quantum computing

A. Qubits: superposition and entanglement

The elementary unit of information in quantum computers is no longer a single bit but a qubit. A classical bit has two mutually exclusive states, 0 or 1, and can only be in one state at any point in time. However, a qubit can be in any of the basis states, $|0\rangle$ and $|1\rangle$, but also in a superposition of both. Mathematically this is described as a linear combination of $|0\rangle$ and $|1\rangle$: $|\psi\rangle = \alpha |0\rangle + \beta |1\rangle$ where $\alpha, \beta \in \mathbb{C}$ are called probability amplitudes and satisfy $|\alpha|^2 + |\beta|^2 = 1$. The action of measuring the qubit in the computational basis will project the state of the single qubit onto one of the basis states $|0\rangle$ or $|1\rangle$ with probabilities $|\alpha|^2$ and $|\beta|^2$, respectively.

A classical system composed by n bits can be described by 2^n possible states that represent values from 0 to $2^n - 1$. Such a system of n bits can only store and process one of the 2^n possible states at a time. In quantum computing, n qubits can be combined in a way that the resulting new state is a superposition of all possible 2^n states, described by $|\psi\rangle = \alpha_0 |0\cdots00\rangle + \alpha_1 |0\cdots01\rangle + \cdots + \alpha_{2^n-1} |1\ldots11\rangle$, where $\alpha_i \in \mathbb{C}$, $\sum |\alpha_i|^2 = 1$. An example of a 2-qubit state is $|\psi\rangle = \frac{1}{2}(|00\rangle + |01\rangle + |10\rangle - |11\rangle)$. *Entanglement* is a special case of superposition of multiple qubits in which the combined qubit state cannot be decomposed into a product of individual states. The key of quantum computing is that by having these *n*-qubits superposed states, one can store 2^n different states and operate on all of them at the same time. This is the essence of the exponential speed up that quantum computers can offer.

Gate	Identity	Pauli-X	Pauli-Y	Pauli-Z	Hadammard	S	T	CNOT
Symbol	—$\boxed{I}$—	—$\boxed{X}$—	—$\boxed{Y}$—	—$\boxed{Z}$—	—$\boxed{H}$—	—$\boxed{S}$—	—$\boxed{T}$—	
Matrix	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$	$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

Table I: Quantum gate symbols and matrices.

B. Quantum gates

Quantum algorithms can be described by a *quantum circuit* when the circuit model is adopted as computational model. A quantum circuit is composed by qubits and gates operating on those qubits. All quantum gates are reversible, unitary operations and are represented by $2^n \times 2^n$ unitary matrices, where n is the number of qubits they act on. The most commonly used single-qubit and two-qubit quantum gates and their corresponding matrix representation are shown in Table I. Note that the controlled-NOT or CNOT gate is a 2-qubit gate that performs an X operation on the target qubit (bottom line) when the control qubit (top line with the black dot) is $|1\rangle$, and otherwise it does not change. In other words, the target qubit is flipped only if the control qubit is $|1\rangle$.

Another essential operation in quantum circuits is *measurement*. There are two important properties of this measurement: 1) the act of measuring collapses the quantum state to the state corresponding to the measurement result. 2) this implies that with a single quantum measurement one cannot 'read' the superposed qubit state. For instance, when measuring the qubit state $|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ in the computational basis one can only get two possible measurement outcomes: '0' (+1) or '1' (-1) both with probabilities $1/2$, that will leave the qubit in the (post-measurement) state $|\psi'\rangle = |0\rangle$ and $|\psi'\rangle = |1\rangle$, respectively. Therefore, a quantum state cannot be measured directly without losing the quantum state and thus the stored information.

Finally, it is worth noting that in quantum computing there also exist *universal sets of gates* and any quantum operation can be approximately implemented by a finite sequence of gates from such a set. One example of such a universal gate set is $\{H, T, CNOT\}$ [1].

C. Quantum error correction and FT computation

As we mentioned in the introduction, current quantum technologies are error prone. Qubits suffer from decoherence, meaning that they loose their information through the interaction with the environment. This decoherence can be measured in two ways. The first is called amplitude damping or T_1 and is the time it takes before a qubit goes from the excited state to the ground state due to dynamic coupling. The T_2 or phase damping refers to the time a qubit can be kept in a superposition state. For superconducting qubits, the T_1 and T_2 are around 30 μs and 60 μs, respectively [6]. This and other sources of errors directly impact the fidelity of the quantum gates that currently show error rates around

10^{-2}. These numbers are far from the $10^{-12} - 10^{-15}$ error rates required for running large quantum algorithms [2]. It is therefore inconceivable to think about building a quantum computer without using quantum error correction and fault tolerant mechanisms.

In QEC, quantum information is protected by: 1) encoding a single *logical qubit* into several physical imperfect qubits using a specific quantum error correction code (QECC) and 2) by continuously 'monitoring' the system to detect and recover from possible errors [7]. Encoding is performed by entangling several *data qubits* (qubits where information is stored), whereas errors are detected by doing parity check measurements, also called error syndrome measurements (ESM). ESM allow to perform parity checks between several data qubits without directly measuring them and thus preserving the qubit states. To this purpose, some 'helper' qubits, called *ancilla qubits* are needed. By measuring the ancilla qubits, the continuous quantum errors will be discretised into bit-flip errors (X) and phase-flip errors (Z). In addition, by looking at the measurement result of those ancilla qubits (+1 or -1), called error syndromes, what kind of error(s) and involved qubit(s) can be identified. The error identification process, called *error decoding* is handled by classical electronics.

One of the most popular quantum error correction code is *surface code* (SC) [8] because of its simple 2D structure with only nearest-neighbour (NN) interactions that perfectly fits with most of the quantum technologies and its high error threshold rate ($\sim 1\%$). This error threshold means that for physical error rates below 1% one can achieve better performance by increasing the code distance[1] [2]. In surface code, qubits are arranged in a regular 2D lattice (Figure 1(a)). The array comprises two kinds of qubits, data qubits, in which the quantum information is stored, and ancilla qubits that are used to detect errors. Note that there are two types of ancilla qubits, Z (in green) and X (in red) that are used to detect bit-flip and phase-flip errors, respectively. As is depicted in Figure 1(a), each ancilla qubit interacts with four data qubits. This configuration allows the surface code to do parity checks between four data qubits. In SC, errors are found by repeating ESMs over the entire lattice [2]. Every time that a ESM is performed, a set of +1's and -1's that points to all possible errors in the lattice is obtained. These measurement outcomes are then forwarded to a classical

---

[1]The code distance is defined as the minimum number of physical operations required to perform a logical operation.
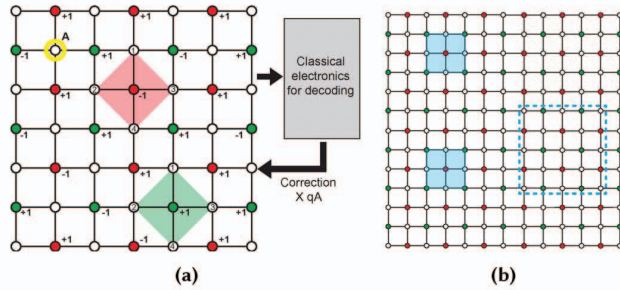
Figure 1: (a) 2D lattice implementation of the surface code. Open circles represent data qubits„ Z ancilla in green and X ancilla in red. (b) Two kinds of logical qubits in SC, planar-based and defect-based.

Figure 2: The quantum system view.

processor where this information is processed allowing to decode such errors (see Figure 1(a)). To this purpose, classical graph algorithms such as the Edmond's matching algorithm (also called Blossom) are used [9]. Once the errors have been identified, the corresponding corrections are applied. For instance, in Figure 1(a), the decoder indicates that a bit-flip error happened in data qubit A (marked with a yellow circle) and then an X gate is applied to it. We can conclude by saying that QEC prevents from accumulating errors in the encoded state during computation by detecting and correcting them.

In surface code, there are two possible ways of encoding a single logical qubit, planar-based and defect-based SC. In planar-based SC, a sub-lattice is used for encoding a single logical qubit (dashed square in Figure 1(b)), whereas in defect-based SC a logical qubit is created by stop measuring a pair of ancilla qubits during the ESM cycles. For instance, in Figure 1(b) the red ancilla qubits inside the blue squares will not perform the ESM. The number of physical qubits required per logical qubit depends on the distance of the QEC; the higher the distance, the more qubits are needed and the more robust the system will be. The distance is set based on the physical error rate of the quantum technology and the reliability required by the algorithm.

It is important to mention that in order for QEC to be effective at correcting errors, encoded gates (gates applied on the encoded qubits) must be designed to be fault-tolerant, implying that small errors do not cause significant information loss. The FT implementation of some gates in surface code such as the T or S gates is very expensive in terms of qubit resources and time because special high-fidelity ancilla states need to be created using a probabilistic procedure called state distillation [10]. In [2], they calculate what would be the size of a quantum computer for factoring a 2000-bit number in a bit more than 1 day using Shor's algorithm. Assuming surface code as QEC code and an error rate in the physical qubits of 10^{-3}, a quantum computer would require about 220×10^6 physical qubits; of which 20×10^6 would be for computation (4000 logical qubits, each composed by 3600 physical qubits) and 200×10^6 would be for generating and purifying the special ancilla states.
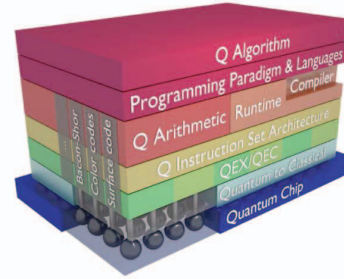
III. Quantum system overview

In [11], we propose an overall architecture for a FT and universal quantum computer (Figure 2). This layered stack defines all layers that need to be developed when building a quantum computer, going from a high-level description of a quantum algorithm to the actual physical operations on the quantum processor. Quantum algorithms [12] are described by high-level quantum programming languages [13]–[15]. Such algorithm description is agnostic to the faulty quantum hardware and assumes that both qubits and quantum operations are perfect. In the compilation layer, quantum algorithms are converted to their FT version based on a specific quantum error correction code such as surface code [2] or color codes [16] and compiled into a series of instructions that belong to the quantum instruction set architecture (QISA). The quantum execution (QEX) and quantum error correction (QEC) layer is responsible for the execution of quantum operations and for the detection and correction of errors [11]. It is in these layers, where quantum instructions are translated into the actual pulses that are sent through the classical to quantum interface to the quantum processor.

From this picture, one can realise that building a quantum computer involves more than building quantum devices. Whereas physicists are mostly working on the quantum chip layer trying to improve the coherence of the qubits and the fidelity of the gates, as well as to increase the number of qubits that can be controlled and entangled, computer and electronic engineers are responsible for the development of the infrastructure required for building such a quantum system. In the rest of the paper, we will discuss what are the main issues that both physicists and engineers are facing nowadays that include: i) improvement and scalability of quantum technology, ii) classical control electronics at (possibly) cryogenic temperatures and iii) the creation of heterogeneous quantum computer architecture and compiler infrastructure.

IV. Quantum Chip

Many implementations of quantum information processors share a set of common goals. Improving coherence properties of qubits and simultaneously enhancing single and two-qubit gate fidelities, at least beyond the fault tolerant threshold, is a goal pursued throughout. Within the next few years, demonstrations of logical qubits with performance

beyond the one of the constituent physical qubits enabled by quantum error correction is expected in a number of implementations. To operate systems of many physical qubits in an extensible fashion, scalable classical control electronics and tune-up routines for large scale quantum systems are to be realized, as discussed below in more detail. Systems with the best performance are preparing for demonstrating operations between error corrected logical qubits on a similar time scale. In five to ten years, demonstrations of quantum algorithms operating on logical qubits in a universal quantum computer are envisaged. At the same time functional quantum interfaces for short, medium and long distance communication between quantum computing modules will be needed for building larger networks. To date the following platforms are among the most promising ones to serve as basis for building quantum hardware.

Trapped ions – Ion trap quantum computing typically operates on a qubit register formed by a linear string of ions confined in a Paul trap [17]–[19]. Each physical qubit is based on two internal levels of a single ion; these levels are either defined within a Zeeman or hyperfine manifold or correspond to a forbidden optical transition. Single-qubit operations use microwave or laser fields, while two-qubit operations in most experiments employ laser fields. Quantum algorithms have been performed on strings of up to seven ions confined in a linear trap [20]. Longer chains of up to 20 ions and 2D crystals of up to ∼300 ions have been trapped and used for quantum state engineering or quantum simulation [21], [22]. Individual qubits can be initialized with a preparation error below ∼10^{-3}, are controlled with gate errors of ∼10^{-6} [23], and read out with an error of ∼10^{-4}. Two-qubit gates have errors of ∼10^{-3} [24] and are typically realized with optical schemes (Cirac-Zoller, Molmer-Sorensen, conditional phase gate). Alternatives based on microwave fields are also investigated. The conversion from stationary to flying qubits has been demonstrated [25], as well as the transfer of quantum information over short distances by physically transporting ions across a microchip. Scalability remains the most significant challenge in ion systems for which well-defined approaches based on micro-fabricated traps and photonic interconnects need to be developed. Various fabrication techniques and electrode configurations are investigated to scale the trap architecture. Micro-fabricated 2D RF-trap arrays have already been successfully demonstrated [18]. A difficulty encountered in miniaturized ion traps is the marked growth of the electric-field noise in the vicinity of trap surfaces causing unwanted motional heating. This issue has been addressed by operating at cryogenic temperatures, or by applying an in-situ cleaning of the surface of the trap; both approaches provide a sizeable reduction of the electric-field noise. However, an understanding of the physical mechanisms responsible for this noise is still lacking [26].

Superconducting circuits – Quantum computation with superconducting circuits exploits the intrinsic coherence of the superconducting phase and the Josephson effect as a re-source of dissipationless non-linearity. Qubits are realized as resonant microwave circuits, embedding a Josephson tunnel junction, of which the two lowest energy levels are used as an effective quantum bit [27], [28]. Superconducting qubits are fabricated with thin-film technology, are probed and controlled with microwave frequency radiation and can be strongly coupled to each other inductively or capacitively [29], [30]. Superconducting resonators and cavities provide opportunities for coupling widely different types of qubits in hybrid devices, including atoms, ions and impurity spins in quantum dots, crystals, and microtraps. Industry interest in superconducting quantum computing has sharply risen in recent years illustrating the potential of this technology. Quantum processors with 4-9 qubits have been demonstrated [31], [32]. Basic quantum error correction protocols, quantum algorithms and simulations have been realized. Universal gate operations are performed with fidelities in excess of 99.9% for single qubits and 99.5% for two-qubit gates [33]. The use of parametric amplification routinely enables single-shot, non-demolition qubit measurements with fidelities exceeding 99%. The bandwidth of parametric amplifiers has been extended from tens of MHz to several GHz, greatly facilitating scalability of quantum measurements. The coherence times of qubits are constantly increasing. At the same time, fast classical control electronics, as required for real-time feedback, are rapidly advancing [34], [35]. Designing and fabricating large scale superconducting circuits avoiding spurious cross-coupling while addressing all circuit elements in multiplexed structures is challenging. Microfabricated superconducting qubits are also sensitive to imperfections in their fabrication limiting yield and reproducibility of device parameters. Both aspects require optimization of design and production processes. Operation of devices below 50 mK requires refrigeration technology which is expected to be a realizable also for larger-scale systems beyond a few hundred qubits.

Electronic semiconductor qubits – In semiconductor host materials single electrons can be either trapped by isolated donor atoms or confined and controlled using gate-defined potentials. The spin degree of freedom in these systems is considered the most promising qubit representation due to its long coherence time [36], [37]. These devices can be measured and controlled fully electrically much like transistors in today's digital electronics and also their fabrication exploits the same technologies as the semiconductor industry. Recently, group IV materials such as silicon and germanium have attracted increasing attention, as they offer longer spin coherence times than GaAs systems. Overall, the wide set of semiconductor materials available offers a range of tunable parameters, such as high-spin-orbit coupling for faster manipulation (InAs), or low nuclear spin concentrations for longer spin coherence times (Si, SiGe) [38]. Quantum dot circuits with up to five quantum dots have been controllably loaded with electrons and the scale-up of qubit arrays along a 1D array is proceeding well [39]. Single spin qubits can be controlled both by electrical or magnetic driving fields.

Single qubit gates have fidelities in excess of 99%, spin states are initialized with 99.9% fidelity, and single shot readout of up to three qubits was demonstrated with an average fidelity of 97%. Coherence times as long as T2 (T2*) = 500 (0.2) ms have been measured in isotopically enriched 28Si. Coherent exchange coupling between two spins in a double dot has been demonstrated as well as the coherent interaction of two double-dot spin states by exploiting capacitive coupling [40]. Nuclear spin states are studied as even longer lived quantum memories and single electron shuttling on surface acoustic waves has been demonstrated [41]. Despite the outstanding performance of single spin qubits, one of the main challenges remains the development and improvement of high fidelity two-qubit gates, particularly for donor spins. Poor qubit uniformity and background disorder currently must be compensated for by tuning of gate voltages. In addition, the presence of low frequency charge noise requires constant retuning. Developing an understanding of the microscopic origin of low and high frequency charge noise will be crucial in identifying mitigation strategies.

Impurity spins in solids – Atomic and molecular spins in solids such as color centres, rare earth ions, deep donors, and molecular magnets, employ both the electron and nuclear spin degrees of freedom as qubits with long coherence times. Control of these systems is typically achieved by combining techniques from liquid state NMR with optical manipulation. The specific advantages of these systems includes long coherence times and access to highly advanced methods for precise manipulation of quantum states. Furthermore, certain spin systems are shielded well enough from their environments such that room temperature operation seems feasible. The most advanced quantum computing experiments have so far been realized with nitrogen vacancy centers in diamond [42], [43]. The efficient initialization and single shot spin readout are achieved with optical control, while single qubit gates employ microwave fields. Two-qubit gates between multiple spins are based either on magnetic dipolar interactions or on long distance optical coupling [44]. Multipartite entanglement, quantum teleportation over long distances, quantum error correction, and the implementation of elementary quantum algorithms has been demonstrated [45]. Despite some recent progress, nano-positioning and the creation yield of defects is still a major challenge, even though there exist schemes for which nano-positioning is not crucial. The creation of defect center arrays and their incorporation into photonic structures is expected to offer a viable path towards scalability, to increase the collection efficiency of light emitted from the defect and to improve coherence times.

V. Classical control electronics

Overview and thermal budget: In current experiments involving up to a handful of qubits, extensive equipment such as GS/s arbitrary waveform generators (AWGs), microwave sources and digitizers are used for qubit control and readout. In the layered framework discussed in section III, their functionality roughly corresponds to the lower part of the QEX layer, i.e. the translation of logic commands into analog wave-forms and the quantum-classical interface responsible for acquisition and processing of readout results.

These instruments are normally located at room temperatures and consume on the order of 1 kW per qubit, while most solid state qubits are operated in dilution refrigerators at temperatures between 20 and 100 mK, where the available cooling power is on the order of 1 mW. Even with a fairly compact PCI-type form factor for a single qubit controller, extending this approach to millions of qubits would fill thousands of racks. Hence, purpose-built, highly integrated solutions are desirable to reduce the cost and complexity of future quantum computers. In addition, locating all control electronics at room temperature poses a rather fundamental connectivity challenge. For example, 10^6 coaxial cables with a cross-section with 1 mm^2 each correspond to a total cross section if 1 m^2, which would impose an unacceptable heat load on the cryogenic end. For comparison, today's largest dilution refrigerators accommodate at most a few hundred high frequency lines. Furthermore, such a cable assembly would be extremely difficult to connect to the likely much smaller qubit chip. Multiplexing can somewhat alleviate the situation, but will eventually be limited to a relatively small number of signals per cable due to frequency crowding and throughput considerations. Hence, it is very attractive to place low level control electronics in the immediate vicinity of the qubits, thus pursuing a fully integrated approach. The key advantage is that microfabricated interconnects between the qubit and control layer can be used.

The limited cooling power available at low temperature is clearly a major challenge for this approach. Current cryogen-free dilution refrigerators supply up to a few mW at 100 mK, less than 1 W at 1 K, and a few W at 4 K. Using a Helium liquifier plant, it should be possible to deliver at least 100 W around 2 K. For 10^6 (10^9) qubits, one thus arrives at a few nW to 100μW (few pW to 100 nW) per qubit, depending on the operating temperature of the control electronics, $T_{control}$. Even in the most optimistic scenario but aiming at 10^9 qubits, one arrives at an ultra-low power budget of 100 nW per qubit. The choice of $T_{control}$ will depend on two factors: the operating temperature of the qubits, T_{qubit}, which could potentially be increased from the current level, and the ability to sustain a temperature difference between the qubit and control layers. These factors will need dedicated research to understand. Thermally isolating chip-to-chip or through silicon via connections might be possible if leveraging the low thermal conductivities of non-crystalline insulators and superconducting wiring below about 20 % of the superconducting critical temperature, which can be around 9 K for high quality Nb or NbTi.

Requirements for qubit control and degree of locality: Having argued that it is desirable to implement fairly advanced local control circuits, we now discuss what functionality it should provide. As current error correction codes only require a few distinct operations on each qubit,

a local control circuit associated with each qubit could have a few (say 4 bit) digital inputs determining the operation to be performed in each clock cycle, and a one bit output for readout results. Performing only the translation from digital single qubit commands to analog pulses with local electronics would still result in immense data transfer rates. For example, 10^8 qubits running at a clock speed of 10 MHz and requiring half a byte of instruction per cycle require a data transfer rate of 500 TB/s, which corresponds to 250'000 PCIe 4.0 lanes. Using global control broadcasts, which rely on the fact that the surface code often requires the same operation on many qubits, would still leave a factor eight lower data traffic for transmitting error syndromes (1 bit per cycle for every second qubit). Hence, it would be highly advantageous to implement at least part of the QEC layer locally as well. Most error syndromes would then be processed and acted upon locally. However, the efficient implementation of error decoding algorithms on dedicated hardware is largely an open problem. In the following, we will focus on the capabilities required for the QEX layer. A key factor are the type of signals to be applied to the qubits, which depend on the type of qubits. The following requirements are representative.

DC Voltages – Quantum dot based spin qubits typically need on the order of five DC bias voltages per qubit. Their level is of order 1 V and needs to be adjusted with a resolution of microvolts to millivolts. Stability to within a microvolt is crucial for many type of qubits (e.g. when using exchange coupling) to avoid charge dephasing. High frequency noise is even more detrimental. Low power bias sources will be facilitated by the fact that the qubits draw essentially no DC current.

Baseband control – Some operations such as exchange-based spin qubit manipulation or tuning gateable Josephson junction qubits ("gatemons") require baseband AC control with a bandwidth ranging from DC to up to a few hundred MHz. The AWGs commonly used to generate these pulses have a sample rate of order 1 GS/s. In many cases, an adequate gate performance should be achievable with a somewhat lower sample rate. Typical control amplitudes are a few mV. As an example for the acceptable noise level, we mention 0.2 nV/$\sqrt{\text{Hz}}$ at high frequencies and 1 μV rms low-frequency noise, which provides adequate single-qubit gate fidelities for two-electron spin qubits [46]. Again, it is useful to note that qubits typically represent a purely capacitive load that will likely be dominated by interconnects.

DC or baseband current bias or control – Some superconducting qubits use fluxes for bias or baseband control. These require currents of up to about 1 mA at zero resistance. These seem hard to generate efficiently with semiconductor electronics and would favor superconducting logic.

Microwave control – The most widespread control approach are microwave driven Rabi oscillations. For superconducting qubits, frequencies of 6 to 12 GHz at power levels of order -100 dBm are common. For spin qubits, carrier frequencies of 20 GHz and above are often used. For driving them

electrically, signal amplitudes of about 1 mV are common. The modulation bandwidth requirement for these microwave bursts is similar as for baseband control.

In addition to the above analog specifications, the degree of independent tunability of the qubit control pulses has to be considered. Each single qubit gate has three unitary (i.e. systematic) error generators, whereas a two-qubit gate has 15. The need to fully eliminate systematic errors by nulling the effect of each of these generators (unless small enough by construction) sets the number of tuning parameters per gate. There are two extreme ways to accommodate this requirement.

Control pulses for each gate and each qubit could be tuned separately, so that at least three parameters have to be set for each single qubit gate. These can vary from gate to gate and thus have to be retrieved rapidly by the control pulse generators or conditioners whenever a specific gate is applied. This approach puts modest demand on qubit homogeneity, but requires relatively complex control electronics. A simple example would be Rabi control with separately programmed I and Q amplitudes and a variable qubit level splitting. Alternatively, if the variation between qubits is small enough or can be tuned away with gate independent control parameters, the same pulse can be applied to all qubits to achieve a certain gate. The role of control electronics could then be limited to mere switching of externally supplied pulses [47]. An example would be Rabi control where tuning the resonance frequency and coupling strength of each qubit could simultaneously eliminate errors on all gates. So far, rather little effort has been made to address the homogeneity requirements associated with the second approach for microfabricated solid state qubits, so that it would be advantageous if qubit controllers provide some degree of gate-dependent tunability of control pulses. The detailed optimization of the tradeoffs between qubit and controller specifications along with hardware-adapted control approaches are another topic requiring substantial further research.

Viability and challenges of ultra low power cryo-electronics: While meeting the above power constraint seems daunting at first, we argue that it should be physically possible if one fully exploits the circumstances of the task. First, the small loads from the qubits do not require power-hungry output drivers if local circuits are used. Moreover, the small signal amplitudes and low operating temperatures allow a reduction of the supply voltage, possibly to values as small as 10 mV.

A key factor is that the ideal switching behavior of a conventional transistor scales with the electron temperature T_e. This relation is typically expressed in terms of the subthreshold swing $S = \log(10)k_B T_e/e$ via the relation $I_{SD} \propto \exp^{-e(V_{th}-V_G)/(k_B T_e)}$, where I_{SD} is the source-drain current, V_{th} the threshold voltage and V_G the gate voltage. At T_e = 2 K, one obtains S = 0.4 mV/dec instead of S = 60 mV/dec at 300 K. As S determines the transfer characteristics and noise margins of logic circuits, the supply voltage V_{dd}
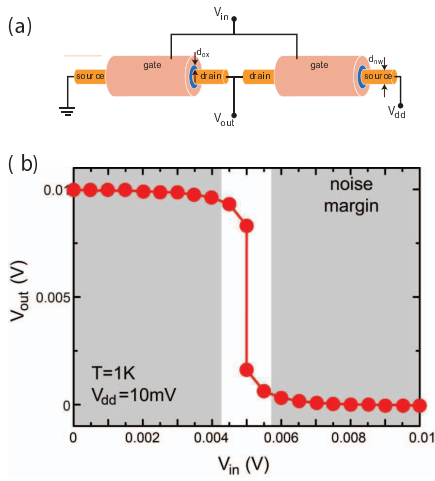
(a)

(b)

Figure 3: Non-equilibrium Greens function simulation of an inverter circuit based on idealized, disorder free nanowire transistors as shown in (a). (b) Inverter response at V_{dd} = 10 mV and T = 1 K.

can be scaled down accordingly, which will lead to a drastic reduction of the dynamic power consumption per transistor given by $P = CV_{dd}^2 f$, where C is the total switched capacitance (including gate and wiring capacitances) and f the switching frequency. For V_{dd} = 10 mV, f = 300 MHz, and C = 1 fF, which is reasonable for digital circuitry with moderate fan out and micron-scale interconnects, one obtains a value of only 30 pW per active transistor. For analog circuits, the minimal capacitances are set by the noise requirements as discussed below. Fig. 3 shows a simulation of an inverter based on the non-equilibrium Green's function formalism. The inverter consists of idealized, disorder free nanowire transistors exhibiting ballistic transport at T_e = 1 K and V_{dd} = 10 mV. The extremely sharp transition confirms that such ultra-small drive voltages are in principle viable.

While advanced CMOS technologies have been shown to exhibit little performance degradation at 4 K [48], [49], one typically finds that S saturates at a factor of a few below the room temperature value. Special exploratory transistor designs have revealed values as low as 4 and 8 mV/dec [50], [51]. This saturation may be a result of disorder, tunneling through the barrier, or self-heating. We see no fundamental reason or physical constraint that would prevent reaching much smaller values with transistor designs that are optimized specifically for low temperature operation. Such an optimization would also involve an adjustment of the turn on voltages to the targeted value of V_{dd}.

Rather old work has already demonstrated the operation of integrated circuits at V_{dd} = 27 mV at 77 K, however with a substantial loss of speed [52]. This loss of switching speed at low V_{dd} resulting from the linear scaling of the carrier density with gate voltage is a second concern of the low power approach. Helpful factors in this respect are that lower speeds than current state of the art room temperature circuits
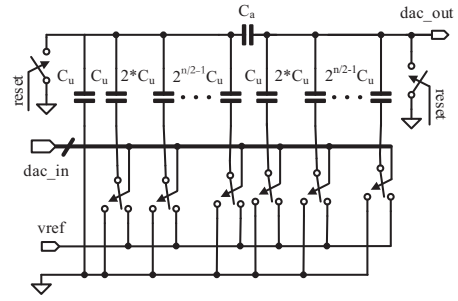

Figure 4: Capacitive voltage divider DAC implementation. The reset switch serves to occasionally reset charge accumulation due to leakage. The attenuation capacitance C_a has to be chosen as $C_a = \frac{2^{n/2}}{2^{n/2}-1} C_u$.

can be sufficient (e.g., for a DAC running at 300 MS/s) and that the mobility substantially increases at low T due to phonon freezeout. A hundredfold reduction in V_{dd} compared to current technology could thus be compensated with a tenfold reduction in clock speed and a tenfold increase in mobility.

In conclusion, a targeted technology optimization driven by a detailed physical understanding of what limits device performance at low temperature could provide a viable pathway to ultra-low-power electronics that are compatible with the tight power budget associated with cryogenic operation. An according adaptation of foundry processes will likely be rather costly, but the same can be expected for the reliable mass production of qubit devices.

Example for ultra-low-power circuit concept: As a concrete example for an ultra-low power circuit design, we discuss a digital-to-analog converter (DAC) based on charge division that could be used for both DC bias and AC control. The circuit diagram is depicted in Figure 4. The advantage of the approach is that there is only a dynamic power consumption in contrast to resistor based designs with static power consumption. Another factor leading to a low power dissipation is the omission of an operational amplifier as output buffer, which is feasible because the DAC only drives the purely capacitive loads of the qubit electrode, possibly in parallel with a storage capacitor. The error resulting from this output load is not a concern because it is constant and calculable.

The power consumption of the DAC in Figure 4 is determined by the capacitance and the sample rate. The total output capacitance of a n bit DAC, C_{out}, which is relevant for the Johnson noise level, is then given by $C_{out} = 2^{n/2}C_u$. The variation of the output voltage between conversion cycles due to Johnson noise is given by $\delta V_n^2 = k_B T_e / C_{out}$.

For a 10 bit DC-bias DAC with 1 V output range, 1 mV resolution, δV_n = 1 μV and T_e = 1 K, one obtains a lower bound of C_{out} = 15 pF. For a hypothetical refresh rate of f_{sample} = 500 Hz, which depends on difficult-to-predict leakage currents, this corresponds to a power dissipation of order $f_{sample}C_{out}V^2$ = 8 nW.

The second use case is the generation of baseband AC control signals with a resolution of 8 bit, a maximum voltage swing of 4 mV and f_{sample} = 300 MS/s. At a target noise level of 0.2 nV/$\sqrt{\text{Hz}}$, the rms output noise within the DAC bandwidth is 3.5 μV, leading to C_{out} = 1.1 pF and a power dissipation of 5 nW. The above calculations show that there are good chances for achieving digital to analog conversion with an acceptable power dissipation. Further reductions are possible with advanced measures such as not resetting the capacitor array with every clock cycle or using intermediate voltages for resets.

Figure 5: Decoding time for different error probabilities

VI. A HETEROGENEOUS QUANTUM COMPUTER ARCHITECTURE

Next, we discuss the architecture that exploits the low-level control electronics to execute the application binaries. We discuss the different topics that need to be addressed.

System Architecture: Even though a quantum computer is a universal Turing machine, it will, at least for the foreseeable future, most likely exhibit exponential speedup only in a limited number of application domains and most likely only for a specific set of routines. It is therefore reasonable to view quantum processors as dedicated hardware accelerators controlled by conventional (super)computers on which the applications run and that call the specific quantum routine. This implies that the overall system architecture will have to allow the execution of pure classical logic as well as of the quantum instructions on the quantum accelerator.

Instruction Set Architecture: The instructions that the accelerator can execute belong to the instruction set and it is not clear yet what a complete instruction set should look like but at least it should contain one of the universal gate sets as described earlier in this paper. However, one might envision that a coarser granularity is required or needed if certain routines, such as period finding, are very frequently used. It would pay off to provide architectural support and include it in the ISA. These instructions can be described using a quantum assembly language such as QASM which is evidently tightly connected to the ISA.

Micro-architecture for QEX and QEC: The next step is to define the micro-architecture that implements the instruction set. One example of a micro-architecture is given in [11] where attention is paid to defining the technology independent and dependent functions. Examples of the first category could be the use of instruction caches which may be needed when based on intermediate results certain parts of a circuit do not need to be executed. Addressing units are required to keep track of what physical qubits have been allocated to what operation and what composes a logical qubit. Pauli Frames could be useful to keep track of the errors that have occurred [53]. Another choice is to adopt a (horizontal) microcoded approach which potentially has two benefits: the first is that more complex instructions that can not be supported by the hardware can be added to the ISA and which are then emulated through simpler instructions. The second is that as quantum technology is maturing, more efficient low level sequences can be found and supported,
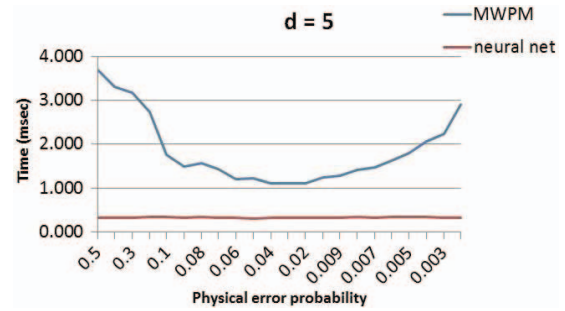
requiring only a change at the microcode level and at the microarchitectural level and further up.

Decoherence times: One of the key challenges at the architecture level is directly related to the problem of decoherence which was defined earlier in this paper. As stated, for superconducting qubits, the T_1 is 30 μs and the T_2 is 60 μs [6]. The challenge lies in the fact that only a small fraction of those times can be used for computation purposes given time budgets for quantum gates in the nanosecond regime. This is not challenging as far as the gate times are concerned, which are for superconducting qubits, 20ns for single qubit gates and 40ns for two qubit gates. However, the most time consuming operations are error syndrome measurement, taking up to 800ns and the decoding of the measurements in order to identify potential errors. The decoding is usually done using the minimum weight perfect matching (MWPM) algorithm such as Edmond's algorithm whose runtime will grow quadratically with the number of qubits and the number of errors that can occur [9]. As shown in 5 and for SC distance 5, the MWPM requires several milliseconds on a high-end server which is several orders of magnitude above the available time budget and varies in function of the error probabilities. One alternative is to use neural networks for constant time decoding.

Qubit plane organisation: A final component of the quantum accelerator is the organisation of the qubit plane. One constraint that current technology imposes is nearest-neighbour proximity for multi-qubit gates. This implies that qubit states will have to be routed in case a multi-qubit operation is to be performed on non-adjacent qubits. This requirement imposes the need for routing logic to compute the preferably shortest path for the qubit movement. This routing must be fault tolerant as certain qubits can be defect or certain regions can be heating up too much and impose less transport being routed in that region. Figure 6 shows the heat map of two quantum algorithms provided by ScaffCC [14], binary welded tree (upper left corner) and square root (center) being executed on a 2D lattice of 15×15 qubits. It shows what positions in the qubit plane are used more intensively than others. In the long term and assuming larger qubit planes, the spatial clustering of qubit movement can also allow for multi-threaded execution where circuits are assigned a certain region in the qubit plane and the runtime support ensures that paths are constrained to the
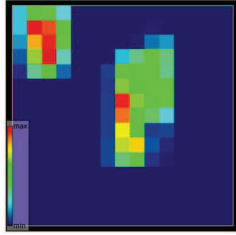
Figure 6: Heat map of quantum circuits

allocated qubit region. For efficiency reasons, the qubit plane can contain dedicated communication regions through which the routing is channeled resulting in what one could call a network-on-quantum-chip (NoQC). Finally, qubits can be moved by using SWAP operations, which exchange the state between two qubits, for short distances but when longer distances need to be travelled on the qubit plane, quantum mechanisms such as teleportation can be used and implemented in so-called quantum repeaters [54].

VII. Compilers for Quantum Programming

A number of programming languages and compilers exist in which quantum algorithms can be written, such as Quipper [13], Scaffold [14] and $LiQUi| >$ [15]. These compilers all generate a variant of QASM of which details can be found in [11], [15], [55], [56]. They provide the following functionality:

Synthesis of quantum circuits: Necessary classical computations such as an adder are still required and need to be implemented in a reversible way [57]. Any reversible quantum circuit still has to be optimally decomposed in a series of quantum gates, belonging to a universal gate set, for which a FT implementation is feasible and supported by the underlying quantum technology.

FT translation of circuits: In this compiler pass, given a quantum error correction code, logical qubits are encoded into several physical qubits and QASM instructions for performing FT operations on such encoded qubits are generated. The QEC choice will impact the number of physical qubits required per logical qubit for achieving the same level of protection as well as the number of physical instructions and cycles for performing a FT operation. Most of the papers discussing the cost of FT computations focus on concatenated codes such as Steane code and few investigated surface code.

Mapping of quantum circuits: When targeting a real quantum processor, the mapping of circuits is an important topic [58], [59]. The circuit description of the algorithms does not usually consider a physical location of the qubits and assumes that any kind of interaction between qubits is possible. However, qubits need to be placed on a specific physical qubit layout that will limit the possible interactions between them, leading to an increase of the circuit latency. It is therefore important to optimize the mapping process that includes the following:

Scheduling of operations – The parallelism of current quantum algorithms is pretty limited but applying classical scheduling methods and techniques, the inherent parallelism
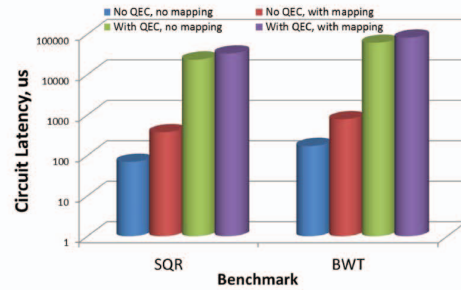

Figure 7: Latency for SQR and BWT quantum algorithms.

of the logical qubits can be exploited. Depending on the QEC choice, different constraints apply to the scheduling problem. For instance, in defect-based SC qubits single-control multi-target CNOT gates are possible whereas planar-based surface only supports single-control single-target CNOT gates. Furthermore, other limitations such as the number of available frequencies to control the qubits can also affect the scheduling process and restrict the parallelism.

Placement and routing of qubits – As mentioned before, most of the current quantum technologies are pursuing a 2D array of qubits with only NN-interactions. This means that 2-qubit (physical) operations are only possible between adjacent qubits. It also impacts the placement of logical qubits. For instance, a CNOT between two planar-based SC qubits can theoretically be performed transversally- i.e. applying pairwise CNOT gates to each pair of data qubits in the sub-lattices. However, it is not possible to implement such a transversal gate in a 2D array requiring techniques such as lattice surgery [60] where planar-based SC qubits still need to be placed next to each other. Finally, not all qubits can be placed in the necessary adjacent positions. Therefore, some of them will have to be moved or 'routed' for which the compiler will insert a *MOVE* operation which will be handled at runtime by the routing logic.

Figure 7 shows the latency of the square root (SQR) and the binary welded tree (BWT) algorithms for execution on superconducting qubits [6], [61], [62]. We analyse different scenarios: with and without QEC (planar SC distance 3) and with and without mapping into a 2D qubit lattice with only NN interactions. Figure 7 shows that the use of QEC will increase the circuit latency by 2 orders of magnitude (blue bar vs. green bar). In addition, the mapping of these quantum algorithms into a 2D lattice will result in an expected increase of the latency by $5\times$ (no QEC) and $1.5\times$ (with QEC) due to the insertion of SWAP operations.

VIII. Conclusion

In this paper, we presented the engineering challenges that need to be addressed when building a quantum computer at a scale at which the full potential can be harvested. At the level of physical qubits, enhancing coherence properties of qubits and gate fidelities as well as optimizing the design and production processes for scaling up quantum processors are

key issues. At the control level, implementing a scalable and ultra low power control circuits at cryogenic temperatures are important enablers. Finally, the system level assumes the definition of a microarchitecture that provides architectural support for the execution of quantum algorithms. System tools such as compilers and runtime support for routing need to address the optimisation requirements for efficient execution of large scale quantum circuits.

Acknowledgment

References

[1] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information.* Cambridge university press, 2010.

[2] A. G. Fowler et al. Surface codes: Towards practical large-scale quantum computation. *Phys. Rev. A.*, 86(3):032324, 2012.

[3] R. Van Meter and C. Horsman. A blueprint for building a quantum computer. *Commun. ACM*, 56(10):84–93, 2013.

[4] N. C. Jones et al. Layered architecture for quantum computing. *arXiv:1010.5022v3*, 2012.

[5] D. P. DiVincenzo and P. W. Shor. Fault-tolerant error correction with efficient quantum codes. *Phys. Rev. Lett.*, 77(15):3260, 1996.

[6] C. C. Bultink et al. Active resonator reset in the nonlinear dispersive regime of circuit qed. *arXiv:1604.00916*, 2016.

[7] S. J. Devitt et al. Quantum error correction for beginners. *Reports on Progress in Physics*, 76(7):076001, 2013.

[8] S. B. Bravyi and A. Y. Kitaev. Quantum codes on a lattice with boundary. *arXiv:9811052*, 1998.

[9] V. Kolmogorov. Blossom v: a new implementation of a minimum cost perfect matching algorithm. *Math. Prog. Comp.*, pages 43–67, 2009.

[10] N. Isailovic et al. Running a quantum circuit at the speed of data. In *ACM SIGARCH, pp. 177-188*, 2008.

[11] X. Fu et al. A heterogeneous quantum computer architecture. In *CF*, pages 323–330, 2016.

[12] J. Stephen. Quantum algorithm zoo. *Updated quantum algorithm list available at http://math. nist. gov/quantum/zoo*, 2011.

[13] A. S. Green et al. Quipper: a scalable quantum programming language. In *ACM SIGPLAN*, volume 48, pages 333–342, 2013.

[14] Ali J. Abhari et al. Scaffold: Quantum programming language. Technical report, DTIC Document, 2012.

[15] D. Wecker et al. Liqui|>: A software design architecture and domain-specific language for quantum computing. *arXiv:1402.4467*, 2014.

[16] H. Bombín. Dimensional jump in quantum error correction. *New Journal of Physics*, 18(4):043038, 2016.

[17] R. Blatt and D. Wineland. Entangled states of trapped atomic ions. *Nature*, 453(7198):1008–1015, 2008.

[18] J. P. Home et al. Complete methods set for scalable ion trap quantum information processing. *Science*, 325(5945):1227–1230, 2009.

[19] C. Monroe and J. Kim. Scaling the ion trap quantum processor. *Science*, 339:1164, 2013.

[20] D. Nigg et al. Quantum computations on a topologically encoded qubit. *Science*, 345(6194):302–305, 2014.

[21] K. Kim et al. Quantum simulation of frustrated Ising spins with trapped ions. *Nature*, 465(7298):590–593, 2010.

[22] T. Monz et al. 14-qubit entanglement: Creation and coherence. *Phys. Rev. Lett.*, 106:130506, 2011.

[23] T. P. Harty et al. High-fidelity preparation, gates, memory, and readout of a trapped-ion quantum bit. *Phys. Rev. Lett.*, 113:220501, 2014.

[24] C. J. Ballance et al. High-fidelity quantum logic gates using trapped-ion hyperfine qubits. *Phys. Rev. Lett.*, 117:060504, 2016.

[25] A. Stute et al. Quantum-state transfer from an ion to a photon. *Nat. Photon.*, 7(3):219–222, 2013.

[26] M. Brownnutt et al. Ion-trap measurements of electric-field noise near surfaces. *Rev. Mod. Phys.*, 87:1419–1482, 2015.

[27] Y. Nakamura et al. Coherent control of macroscopic quantum states in a single-Cooper-pair box. *Nature*, 398(6730):786–788, 1999.

[28] M. Devoret and R. J. Schoelkopf. Superconducting circuits for quantum information: An outlook. *Science*, 339:1169–1174, 2013.

[29] A. Wallraff et al. Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics. *Nature*, 431:162–167, 2004.

[30] I. Chiorescu et al. Coherent dynamics of a flux qubit coupled to a harmonic oscillator. *Nature*, 431(7005):159–162, 2004.

[31] J. Kelly et al. State preservation by repetitive error detection in a superconducting quantum circuit. *Nature*, 519:66, 2015.

[32] D. Ristè et al. Detecting bit-flip errors in a logical qubit using stabilizer measurements. *Nat Commun*, 6:6983, 2015.

[33] R. Barends et al. Superconducting quantum circuits at the surface code threshold for fault tolerance. *Nature*, 508(7497):500–503, 2014.

[34] D. Ristè et al. Feedback control of a solid-state qubit using high-fidelity projective measurement. *Phys. Rev. Lett.*, 109:240502, 2012.

[35] R. Vijay et al. Stabilizing Rabi oscillations in a superconducting qubit using quantum feedback. *Nature*, 490(7418):77–80, 2012.

[36] D. Loss et al. Quantum computation with quantum dots. *Phys. Rev. A*, 57(1):120–, 1998.

[37] B. E. Kane. A silicon-based nuclear spin quantum computer. *Nature*, 393(6681):133–137, 1998.

[38] J. J. L. Morton et al. Embracing the quantum limit in silicon computing. *Nature*, 479(7373):345–353, 2011.

[39] D. M. Zajac et al. Scalable gate architecture for densely packed semiconductor spin qubits. *ArXiv e-prints*, 2016.

[40] J. R. Petta et al. Coherent manipulation of coupled electron spins in semiconductor quantum dots. *Science*, 309:2180, 2005.

[41] M. Kindermann. Nanoelectronics: Shuttle transport for single electrons. *Nat Nano*, 6(11):691–692, 2011.

[42] F. Jelezko et al. Observation of coherent oscillations in a single electron spin. *Phys. Rev. Lett.*, 92:076401, 2004.

[43] R. Hanson and D. D. Awschalom. Coherent manipulation of single spins in semiconductors. *Nature*, 453(7198):1043–1049, 2008.

[44] W. Pfaff et al. Unconditional quantum teleportation between distant solid-state quantum bits. *Science*, 345(6196):532–535, 2014.

[45] G. Waldherr et al. Quantum error correction in a solid-state hybrid spin register. *Nature*, 506(7487):204–207, 2014.

[46] P. Cerfontaine et al. High-fidelity single-qubit gates for two-electron spin qubits in gaas. *Phys. Rev. Lett.*, 113:150501, 2014.

[47] J. M. Hornibrook et al. Cryogenic control architecture for large-scale quantum computing. *Phys. Rev. A.*, 3(2):024010, 2015.

[48] I. D. C. Lamb et al. An FPGA-based instrumentation platform for use at deep cryogenic temperatures. *Rev. Sci. Instrum.*, 87(1), 2016.

[49] Harald H. et al. A reconfigurable cryogenic platform for the classical control of scalable quantum computers. 2016.

[50] S. Habicht et al. Electrical characterization of Omega-gated uniaxial tensile strained Si nanowire-array metal-oxide-semiconductor field effect transistors with < 100 >- and < 110 > channel orientations. *Thin Solid Films*, 520(8):3332–3336, 2012.

[51] B. Roche et al. A tunable, dual mode field-effect or single electron transistor. *Applied Physics Letters*, 100(3), 2012.

[52] J. B. Burr. Cryogenic ultra low power cmos. In *ISLPED, pp. 82-83*, 1995.

[53] Emanuel Knill. Scalable quantum computing in the presence of large detected-error rates. *Physical Review A*, 71(4):042322, 2005.

[54] D. Copsey et al. Toward a scalable, silicon-based quantum computing architecture. *IEEE J. STQE*, 9(6):1552–1569, 2003.

[55] Ali J. Abhari et al. Scaffcc: Scalable compilation and analysis of quantum programs. *Parallel Computing*, 45:2–17, 2015.

[56] T. Häner et al. A software methodology for compiling quantum programs. *arXiv:1604.01401*, 2016.

[57] N. Abdessaied and R. Drechsler. Reversible and quantum circuits.

[58] C.C. Lin et al. Paqcs: Physical design-aware fault-tolerant quantum circuit synthesis. *IEEE Trans. on VLSI Systems*, 23(7):1221–1234, 2015.

[59] M. J. Dousti and M. Pedram. Minimizing the latency of quantum circuits during mapping to the ion-trap circuit fabric. In *DATE*, 2012.

[60] C. Horsman et al. Surface code quantum computing by lattice surgery. *New Journal of Physics*, 14(12):123011, 2012.

[61] S. Asaad et al. Independent, extensible control of same-frequency supercond. qubits by selective broadcasting. *arXiv:1508.06676*, 2015.

[62] D. Ristè et al. Detecting bit-flip errors in a logical qubit using stabilizer measurements. *Nature communications*, 6, 2015.