

## REVIEW ARTICLE OPEN

# Quantum information density scaling and qubit operation time constraints of CMOS silicon-based quantum computer architectures

Davide Rotta<sup>1,2,3</sup>, Fabio Sebastiano<sup>4</sup>, Edoardo Charbon<sup>4</sup> and Enrico Prati<sup>1</sup>

Even the quantum simulation of an apparently simple molecule such as  $\text{Fe}_2\text{S}_2$  requires a considerable number of qubits of the order of  $10^6$ , while more complex molecules such as alanine ( $\text{C}_3\text{H}_7\text{NO}_2$ ) require about a hundred times more. In order to assess such a multimillion scale of identical qubits and control lines, the silicon platform seems to be one of the most indicated routes as it naturally provides, together with qubit functionalities, the capability of nanometric, serial, and industrial-quality fabrication. The scaling trend of microelectronic devices predicting that computing power would double every 2 years, known as Moore's law, according to the new slope set after the 32-nm node of 2009, suggests that the technology roadmap will achieve the 3-nm manufacturability limit proposed by Kelly around 2020. Today, circuital quantum information processing architectures are predicted to take advantage from the scalability ensured by silicon technology. However, the maximum amount of quantum information per unit surface that can be stored in silicon-based qubits and the consequent space constraints on qubit operations have never been addressed so far. This represents one of the key parameters toward the implementation of quantum error correction for fault-tolerant quantum information processing and its dependence on the features of the technology node. The maximum quantum information per unit surface virtually storable and controllable in the compact exchange-only silicon double quantum dot qubit architecture is expressed as a function of the complementary metal–oxide–semiconductor technology node, so the size scale optimizing both physical qubit operation time and quantum error correction requirements is assessed by reviewing the physical and technological constraints. According to the requirements imposed by the quantum error correction method and the constraints given by the typical strength of the exchange coupling, we determine the workable operation frequency range of a silicon complementary metal–oxide–semiconductor quantum processor to be within 1 and 100 GHz. Such constraint limits the feasibility of fault-tolerant quantum information processing with complementary metal–oxide–semiconductor technology only to the most advanced nodes. The compatibility with classical complementary metal–oxide–semiconductor control circuitry is discussed, focusing on the cryogenic complementary metal–oxide–semiconductor operation required to bring the classical controller as close as possible to the quantum processor and to enable interfacing thousands of qubits on the same chip via time-division, frequency-division, and space-division multiplexing. The operation time range prospected for cryogenic control electronics is found to be compatible with the operation time expected for qubits. By combining the forecast of the development of scaled technology nodes with operation time and classical circuitry constraints, we derive a maximum quantum information density for logical qubits of 2.8 and 4 Mqb/cm<sup>2</sup> for the 10 and 7-nm technology nodes, respectively, for the Steane code. The density is one and two orders of magnitude less for surface codes and for concatenated codes, respectively. Such values provide a benchmark for the development of fault-tolerant quantum algorithms by circuital quantum information based on silicon platforms and a guideline for other technologies in general.

*npj Quantum Information* (2017)3:26; doi:10.1038/s41534-017-0023-5

## INTRODUCTION

The<sup>1–4</sup> efforts toward large-scale quantum information processing (QIP) for practical applications have been boosted after two key milestones: the advent of quantum algorithms<sup>5, 6</sup> and the invention of quantum error correction (QEC) codes,<sup>7–12</sup> including the implementation of error correction in a fault-tolerant manner with concatenated<sup>13–16</sup> and topological codes.<sup>17–21</sup> Such methods call for a very large number of physical qubits, a requirement that has triggered the quest for a silicon platforms suitable for

manufacturing large arrays of devices. In addition, the mature silicon nanoelectronics platform may play a role in providing the extensive classical circuitry that can meet the speed and power specifications required to manipulate and readout such large qubit arrays. Finally, the protection of quantum coherence when quantum states are shuttled across multiple interdevice distances represents a difficulty that may be overcome by exploiting the nanometric scale size of the most aggressive technology nodes. To this extent, the maximum amount of quantum information

<sup>1</sup>Istituto di Fotonica e Nanotecnologie, Piazza Leonardo da Vinci 321 - 20133 Milano, Italy; <sup>2</sup>Scuola Superiore Sant'Anna, InPhoTeC, Via G. Moruzzi 1, I-56124 Pisa, Italy; <sup>3</sup>CNIT, Photonic Networks National Lab, Via G. Moruzzi 1, I-56124 Pisa, Italy and <sup>4</sup>Faculty of Electrical Engineering, Delft University of Technology, Mekelweg 4, Delft 2628CD, The Netherlands

Correspondence: Enrico Prati (enrico.prati@cnr.it)

Received: 27 October 2016 Revised: 26 April 2017 Accepted: 11 May 2017

Published online: 26 June 2017

density per unit surface is not a mere curiosity (Note: The maximum amount of information contained in a specific system is related to entropy, according to the holographic principle.<sup>22</sup> Such information content corresponds to the number of degrees of freedom (the system entropy) that is  $(2L_p)^{-2}$ , where  $L_p$  is the Planck constant, leading to the huge value of  $10^{65}$  qubits per square centimeter),<sup>22, 23</sup> but it represents a key technological boundary since the size of quantum computer chips will be reasonably limited to few  $\text{mm}^2$  for packaging and refrigeration considerations.

Silicon platforms for quantum computing hold the promise to achieve high scalability, testability, and reliability levels;<sup>24, 25</sup> however, a straightforward integration of multiple silicon spin qubits with complementary metal–oxide–semiconductor (CMOS) control electronics is not straightforward and a complete architecture for such an interface still lacks. In particular, the quantification of the scalability potential for this interface, its node-dependent performance, and the impact on the implementation of algorithms and QEC pose precondition issues have never been addressed systematically. Cryogenic CMOS circuits for readout of quantum states in semiconductor quantum dots (QDs) have been developed in the past.<sup>26–28</sup> CMOS-based QDs have been demonstrated,<sup>29, 30</sup> while small arrays of qubits have been employed for elementary operations in GaAs<sup>31–33</sup> and non-CMOS silicon QDs.<sup>34–36</sup> By reviewing the state-of-the-art of circuitual QIP and of its control electronics in silicon, we analyze the constraints binding the assessment of a fully CMOS approach that combines classical electronics with quantum circuits on the same substrate operating at cryogenic temperatures. To achieve such a goal, for technical reasons discussed later, we select the exchange-only silicon QD qubit<sup>37, 38</sup> based on three electrons in a double QD. In fact, this qubit is the most compact to be implemented with all-electrical control in semiconductor technology, and despite the important development required to scale up this architecture its features make it worthwhile to conduct a careful consideration of its advantages and drawbacks. Furthermore, the discussion can be relatively easily adapted to singlet-triplet (S-T) spin qubits in double QDs, at the cost of increasing the complexity of the system by adding magnetic field gradients. Let us first define a few elementary building blocks for the proposed architecture to compose scalable quantum circuits with classical readout and control. Inspired by the International Technology Roadmap of Semiconductors (ITRS), we determine the technology-dependent area of the key devices generating universal quantum circuits. The area of logical qubits is consequently calculated for two QEC architectures, namely, Steane code [7,1,3] and surface code, as to obtain the scaling law of quantum information density as a function of the technology node for both. The implications on the realization of scalable quantum algorithms, as well as the integration with classical control circuitry, are then discussed with a special attention to the matching of the operation time of the qubits with the control electronics. A general layered architecture includes physical, virtual, QEC, logical and application layers, respectively.<sup>1</sup> A silicon CMOS substrate may provide a physical realization of the physical layer of the qubits, the virtual layer by integration of classical control circuits for dynamical decoupling compensation sequences and readout, and the QEC layer. Cryogenic operation of the classical control circuitry is required to enable co-integration with the quantum processor, thus enabling a simplified interconnection scheme to a number of qubits large enough for fault-tolerant quantum processing. Such cryogenic time-division, frequency-division, and space-division multiplexing can significantly reduce the number of interconnects and power dissipation, by operating in a frequency range between 1 and 10 GHz, which is compatible with the operation time range imposed by physical and architectural constraints of logical qubits based on exchange-only three spin qubits in silicon double QDs, the latter being around 1–100 GHz.

## GEOMETRICAL CONSTRAINTS OF THE HYBRID QUBIT ARCHITECTURE

Logical qubits operated by a substrate-independent application layer are based on the fault-tolerant QEC circuit implementing classical operations for the manipulation of virtual qubits and virtual gates. In turn, the virtual layer is based on the measurement and manipulation of physical qubits, implemented throughout our review by electron (hole) spin in CMOS silicon QDs. In this section, we discuss how the virtual layer is assessed from the physical layout by keeping the space resources at the maximum compactness.

Several virtual qubits have been proposed for QIP in Si.<sup>39</sup> In single spin physical qubits, the two spin states of an electron in a QD are Zeeman-split by an external magnetic field and manipulated by means of microwave pulses.<sup>32, 40</sup> A S-T virtual qubit architecture is based on two electron spins in a double QD and it does not require microwave pulses.<sup>31, 34</sup> Fast operation is achieved through electrostatic control, provided that a strong magnetic field gradient is built across the double QD. Both architectures were implemented in silicon, demonstrating coherent qubit rotations at GHz frequencies, which is much faster than the coherence time predicted for such QD systems ( $T_2^* \sim \mu\text{s}$ ).<sup>37</sup> The integration of micron-sized resonant microwave antennas and micromagnets for every single qubit may be a major issue against the large-scale integration of single spin and S-T qubits, respectively, as state-of-the-art microwave line integration for qubit control ranges around 100 microns length scale of minimum feature size.<sup>41</sup> Inhomogeneity in the magnetic field gradient and microwave interaction with nearest qubits may introduce strong variability in the qubit functional properties and consequently high error rates at the virtual layer. On the contrary, an architecture featuring all-electrical manipulation could take the best of the CMOS technology in terms of scalability and natural integration of the classical control circuitry. This requirement is met by using three electron spin qubits, whose dynamics is governed only by their exchange interaction. Originally, an architecture of this kind was proposed in systems of triple QDs, where the complete control of qubit states is obtained by tuning the inter-dot exchange couplings.<sup>33, 42</sup> Later, a more compact version of the exchange-only triple electron spin virtual qubits has been proposed<sup>37, 43</sup> and implemented<sup>36</sup> by using only two QDs, one of which is doubly occupied. Both architecture are extensively reviewed in ref. 39. In contrast to previous architectures, only a small magnetic field is required during the virtual qubit initialization, provided externally.<sup>37, 42</sup> Therefore, one may take this qubit as the most compact one with all-electrical control.

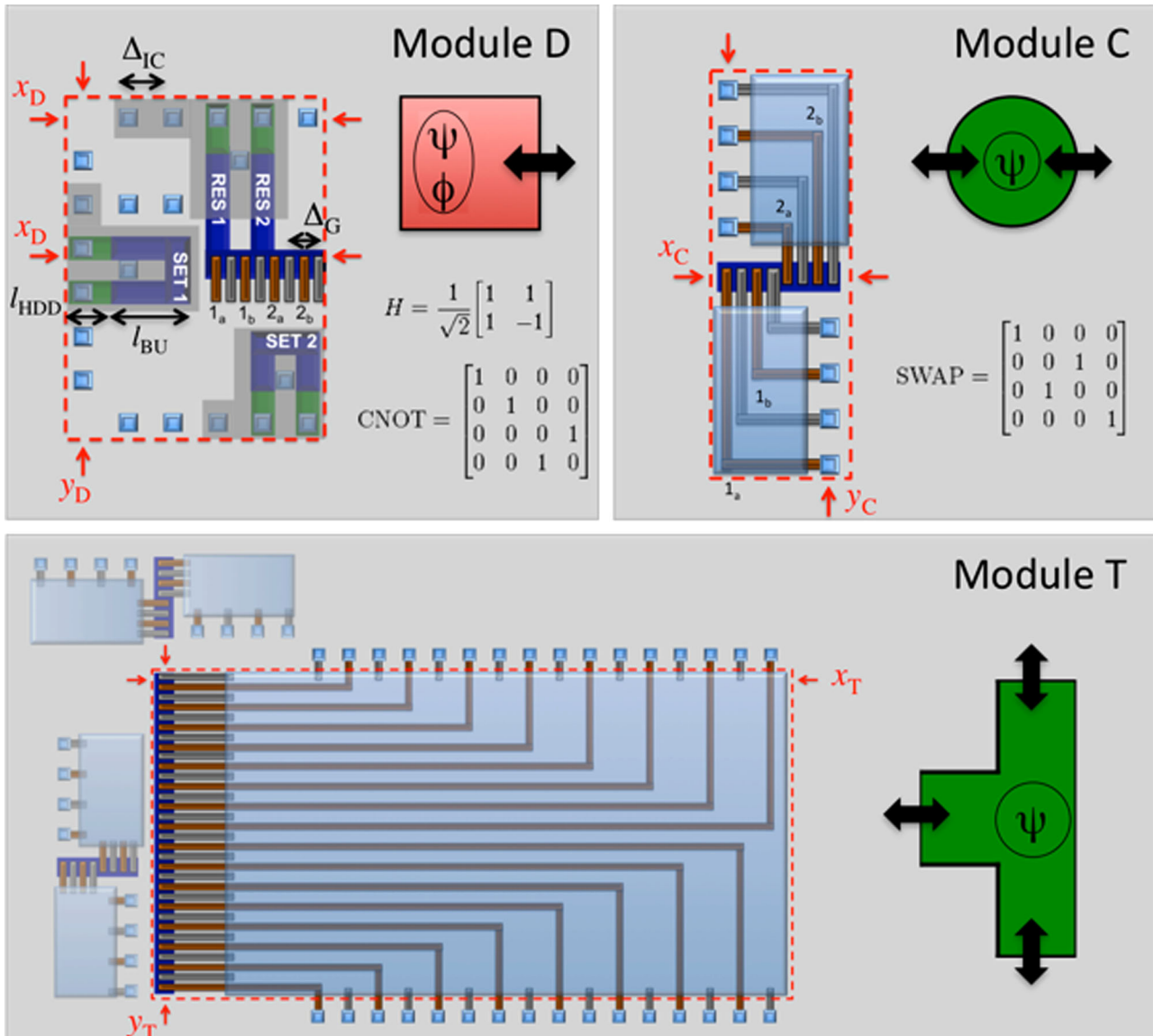
The exchange-only double quantum dot (DQD) qubit, sometimes referred to as a hybrid qubit for its intermediate nature between charge and spin qubits, employs three electron spins in a semiconductor DQD. The logical basis is defined in the spin subspace with total spin  $S = 1/2$  and vertical component  $S_z = -1/2$ . The logical states employed in the computation are:

$$|0\rangle \equiv |S\rangle|\downarrow\rangle, \quad |1\rangle \equiv \frac{\sqrt{1}}{3}|T_0\rangle|\downarrow\rangle - \frac{\sqrt{2}}{3}|T_{-}\rangle|\uparrow\rangle, \quad (1)$$

Where  $|S\rangle$  and  $|T\rangle$  denote the S-T states of the doubly occupied dot, and  $|\uparrow\rangle$  and  $|\downarrow\rangle$  indicate the spin orientation of the single electron in the second dot. Such a virtual qubit is set in the  $|0\rangle$  state by operations at the physical layer by first polarizing the uncoupled spin in a magnetic field that is briefly switched on. Although no magnetic field has been used in single qubit experiments,<sup>36</sup> this is required when dealing with more qubits in order to set a common reference for the spin orientation in different qubits. At operating temperatures of about 100 mK, a magnetic field intensity of 1.5 T is sufficient to impose the orientation of a single spin by means of spin-dependent tunneling from a reservoir to the QD.<sup>44</sup> We also note that, although not

necessary for the manipulation, a small magnetic field could even be beneficial in mitigating the effect of the low-frequency magnetic field fluctuations.<sup>45</sup> The time load for this operation corresponds to the electron tunneling time, which can be set exceedingly small by lowering the tunnel barrier. This technique leads to an initialization of the singly occupied dot with fidelity currently of 95%, which is limited by thermal noise.<sup>35, 46</sup> The doubly occupied dot is then initialized by driving the system in a configuration, where singlet is the ground state. The manipulation of the virtual qubit is performed by tuning the inter-dot effective exchange interaction. The effective Hamiltonian driving the

system can be written in term of spin–spin interactions only.<sup>47</sup> Such interactions are finely tuned by controlling the QD potential and the inter-dot electrostatic barrier, allowing fast and all-electrical manipulation of the qubit state. The first experimental work demonstrated coherent qubit rotations at GHz frequencies with fidelity of about 90%,<sup>36</sup> albeit much better performances are theoretically predicted with optimized fabrication and pulse sequences.<sup>43, 45, 48</sup> The final state is read out by means of charge sensing after collapsing the DQD system in either the (2,1) or the (1,2) charge state, corresponding to the  $|0\rangle$  and the  $|1\rangle$  logic state, respectively. Two-qubit gates could be performed similarly as in



**Fig. 1** The three physical building blocks constituting the virtual layer of the qubits based on exchange-only DQD qubits for the implementation of the Steane code. Schematic device mask is provided for each module together with its symbolic representation utilized in Fig. 2. Greek letters denote virtual qubits storable inside the module, whereas arrows indicate the available quantum connections to neighboring virtual qubits. *Top left*—Data virtual qubit (**d**) for one-qubit and two-qubit virtual gates, e.g., the Hadamard (**h**) and the CNOT. Labeled plunger gates (orange) control the chemical potentials of the four QDs defining the two qubits. Barrier gates (gray) control the inter-QD tunnel coupling. Shaded areas indicate the charge sensors (SET) and the electronic reservoirs (RES) of virtual qubits 1 and 2. Black arrows denote the critical feature sizes defined in the main text. Module sizes are calculated along the directions indicated by the red arrows. *Top right*—Chain module (**c**) for quantum communication through SWAP gates. **c** Bottom—T module (**t**). The module supports quantum communication along the silicon nanowire direction (in this case vertically) analogously to module (**c**). In addition, a horizontal virtual qubit chain can be connected on the left of such device to build up a T-shaped crossing with a vertical channel. Blue areas denote the space for metal interconnections between the active area over the Si nanowire and vias at the module boundary. Only plunger gates interconnections (orange) are depicted for clarity. Metal interconnections are disposed so that the space for surrounding (**c**) modules (depicted shaded on the left and top of **t** module) is taken into account

other qubit architectures by exploiting either the capacitive or the exchange coupling between two adjacent qubits<sup>45, 49–51</sup> and so providing a universal set of virtual gates<sup>38, 52</sup>. The exchange-only qubit has been realized in epitaxial Si–SiGe heterostructures and Rabi oscillations have been observed,<sup>36</sup> with a  $T_2^* = 2$  ns at nominal electron temperature of 150 mK and controlled at picosecond timescale. In principle, a complementary system based on holes<sup>53, 54</sup> instead of electrons is also possible by using CMOS technology.<sup>55, 56</sup> Holes in silicon carry the potential advantages of a considerably smaller hyperfine coupling with nuclei that cause decoherence,<sup>57, 58</sup> and lack of the valley degeneracy that both causes interference phenomena<sup>59</sup> and interleaved complicated valley sub-orbitals as happens for low-electron filling.<sup>60</sup> Therefore, most of the following discussion holds for both electron and hole double QDs. On the contrary, a different reasoning applies to donor impurities in silicon.<sup>61, 62</sup> In brief, single electron spin donor qubits exhibit very long coherence time of the order of seconds,<sup>44, 63</sup> and the effectiveness of atomic resolution lithography based on scanning tunneling microscope<sup>64</sup> is progressively approaching serial implantation.<sup>65</sup> The less accurate single ion implantation<sup>66–68</sup> method could achieve sufficient precision for some architecture such as a surface code implementation based on a two-dimensional array of distant donors,<sup>69</sup> which tolerates deviation of up to 11 nm from the ideal lattice position. However, the complexity of the serial design of devices involving either individual donors with single spins for qubits with microwave control or pair of donors with two or three electrons bound to donors controlled by gates that depends on a relatively high number of currently unaddressed assumptions including yield of implantation and activation of all the donor sites, control of inter-donor distance at single lattice precision, global or individual microwave control, S-T and exchange-only three spin qubits, CMOS mask design on top of silicon overgrowth on the donor layer. The surface code implementation proposal based on a spin-probe controlled two-dimensional array of donors<sup>69</sup> considers a distance of about 400 nm between neighboring donors, which represents an intermediate scale between QDs and superconductor qubits. Regardless of the chosen physical implementation, besides the desirable improvements concerning the single qubit specifications to achieve fault-tolerant fidelity for one-qubit and two-qubit gates, the practical demonstration of two-qubit devices and more complex circuits involves additional issues<sup>39, 45</sup>: the need for long range quantum communication, QEC, and the demanding requirements related to the massive integration of classical control electronics at the quantum chip level. In the following, we focus on the double QD design for hybrid qubits, which could also be naturally adapted to charge qubits<sup>51</sup> and S-T architectures<sup>43</sup> if the issue of adding a local magnetic field gradient can be addressed with no cost in terms of additional space requirements.

As one is interested in the large-scale fabrication of quantum circuits based on a silicon platform, the CMOS implementation of the hybrid qubit architecture is considered, toward the identification of the scaling law of computational power per unit surface area as a function of the technology node. Universal QIP can be addressed by different approaches. First, we consider the universal set of single virtual qubit rotations and virtual controlled NOT (CNOT) logic ports<sup>38</sup> for a Steane code [7,1,3], so we adopt the definition of the three physical building blocks,<sup>70</sup> namely, a data qubit (D) capable of one-qubit and two-qubit logic gates and two types of communication qubits, i.e., the chain module (C) and the T module (T). A schematic representation of these devices is reported in Fig. 1.

The module D incorporates two qubits (i.e., four QDs) and the corresponding individual electronic reservoir and single electron transistor (SET) for independent initialization and readout. To assess the one-qubit and two-qubit gates mechanisms, as well as measurement readout at the physical layer, the system is

equipped with a reservoir consisting of high electron density region with a quasi-continuous density of states that is controlled by an accumulation gate to provide electrons required for qubit manipulation during the initialization procedure. The SET is used as a single charge sensor to monitor the qubit charge state during readout. The coherent transfer of quantum information between distant qubits is a very challenging task, which could in principle be assessed through the sequential repetition of SWAP logic gates between adjacent qubits, as proposed in ref. 70. The module C is specifically designed for quantum communication following this procedure, with no need for initialization and readout. To this purpose, the use of different techniques, such as the coherent transfer by adiabatic passage (CTAP) and teleportation, will be also considered later, as well as the impact of electron loss during the qubit transportation. Finally, the module T is a modified version of the latter to connect perpendicular quantum communication lines and create two-dimensional arrays of qubits. Each device is developed within the same design rules for a given technology node and instantiated as a conventional component in the design of large arrays of identical qubits independently accessible by classical electronics (for more extensive review on the interface between the physical qubits and the classical electronic layer, see ref. 71). The physical size of the three building blocks constituting the virtual layer is then calculated as a function of the main critical sizes such as the pitch of metal interconnections and the width of Si islands. For example, module C height is constrained by the silicon wire width and by the four vias on each side. The sizes of the other devices are determined analogously in terms of the critical pitches. The expressions reported in Table 1 are defined only by the device layout, while being totally independent of the technology node.  $\Delta_G$  and  $\Delta_{IG}$  are the contacted gate pitch and the interconnect pitch, respectively,  $w_{Si}$  is the width of the silicon wires hosting the dots,  $l_{HDD}$  is the length of the highly doped drain, and  $l_{SU}$  is an undoped silicon buffer to avoid unintentional doping of the device active region.

As an alternative option, we examine the implementation of surface codes based on nearest neighbor qubit lattices, and having marked differences from the device layout envisaged for the Steane code. Surface codes appear to be promising candidates for the implementation of fault-tolerant quantum circuits with error thresholds in the  $10^{-2}$  range. Notably, a possible layout was proposed for the implementation of both QEC and leakage correction protocols by surface codes based on S-T qubits realized in semiconductor double QDs.<sup>72</sup> By exploiting the formal and geometrical analogy between the S-T qubit and the hybrid qubit here discussed (both based on semiconductor double QDs hosting electron spins), it is possible to evaluate the size of the consequent logical qubit for surface codes<sup>20</sup> as the two cases may be treated with no fundamental difference (David Di Vincenzo and Hendrik Bluhm, private communication), with the exception of the highest operation speed for the hybrid qubit, which is the case considered in this review.

A layout for the implementation of the surface code on a nearest neighbor lattice of physical qubits is depicted in Fig. 2 (fourth column), in analogy with ref. 72. The asymmetry of the hybrid qubit does not constitute a difficulty as the four hybrid qubits can be arranged in an alternated configuration block (top of fourth column of Fig. 2). The surface code consists of a two-dimensional lattice of data (*red*) and ancillae (*green* and *yellow*) qubits to implement Z-stabilizer and X-stabilizer for QEC by classical operations. The physical qubit area is estimated accordingly as the area of such block divided by 4, i.e.,  $(8 \Delta_G)^2/4$ . Nearest neighbor connections impose stringent proximity of the physical qubits, which from one side lead to a minimum footprint, but on the other raises potential difficulties in their control to address the physical operations to achieve an operating virtual layer. The SET-based charge sensors proposed for the Steane code may be replaced by rf-reflectometry sensors fully integrated within the

**Table 1.** Physical size of the three building blocks for QIP, of a logical qubit and of a qubyte as a function of the main critical sizes independent of the technological node

Device	x	y
D	$x_D = \max(6\Delta_{IC}, 4\Delta_G + l_{HDD} + l_{BU} + \Delta_{IC})$	$y_D = 8\Delta_{IC}$
C	$x_C = 4\Delta_G$	$y_C = 8\Delta_{IC} + w_{Si}$
T	$x_T = 20\Delta_{IC} + l_{BU} + w_{Si}$	$y_T = 14\Delta_G$
Qubit	$x_{qb} = 20y_T$	$y_{qb} = 2x_T - w_{Si}$
Qubyte	$x_{QB} = x_C \cdot [x_T/x_C + 1] + 40y_T + w_{Si}$	$y_{QB} = 7x_C \cdot [\max(x_T, x_D)/x_C + 1] + 2x_T + 7y_T$

The size of the elementary components, namely the module D, C and T, are expressed by considering the physical components along the vertical and horizontal pathways between the red arrows. The sizes of logical qubit and qubyte are calculated on the basis of the physical qubit arrangement proposed in Ref. [71].

control gates.<sup>73, 74</sup> It is worth mentioning that such compact layout comes at the expense of considerably higher complexity at the classical control layer, discussed later. The classical circuitry architecture is supposed to be able to deliver rf pulses of arbitrary shapes for qubit manipulation as well as readout, dealing with serious issues of high-density routing and cross-talk behavior.

While in the first experimental works on single qubit devices this has not been a blocking point, the complexity increases in a large-scale computer. The replication of the same wiring strategy is not a viable solution for two reasons: on the one hand, the extremely dense topology of physical qubit arrays generates serious routing problems for the interconnections on the quantum chip; on the other hand, interfacing billions of control lines with classical electronics on a different substrate or package is not realistic with current and foreseeable back-end technologies.

It is widespread opinion that the classical electronics will be split into several stages between the quantum chip and room temperature electronics (see Fig. 3a) to optimize the system thermal management. A classical integrated circuit directly attached on the quantum chip by conventional flip-chip technology may provide the low-level interface with the quantum layer. An interposer with superconducting through silicon vias could be effective in limiting the thermal load arising from classical electronics. Nonetheless, the partial integration of control electronics within the quantum chip level is an option that should be considered to alleviate routing issues at the quantum chip level, to make the interface with higher-level classical electronics easier, and to improve its performance by reducing the distance from the qubits.<sup>71</sup>

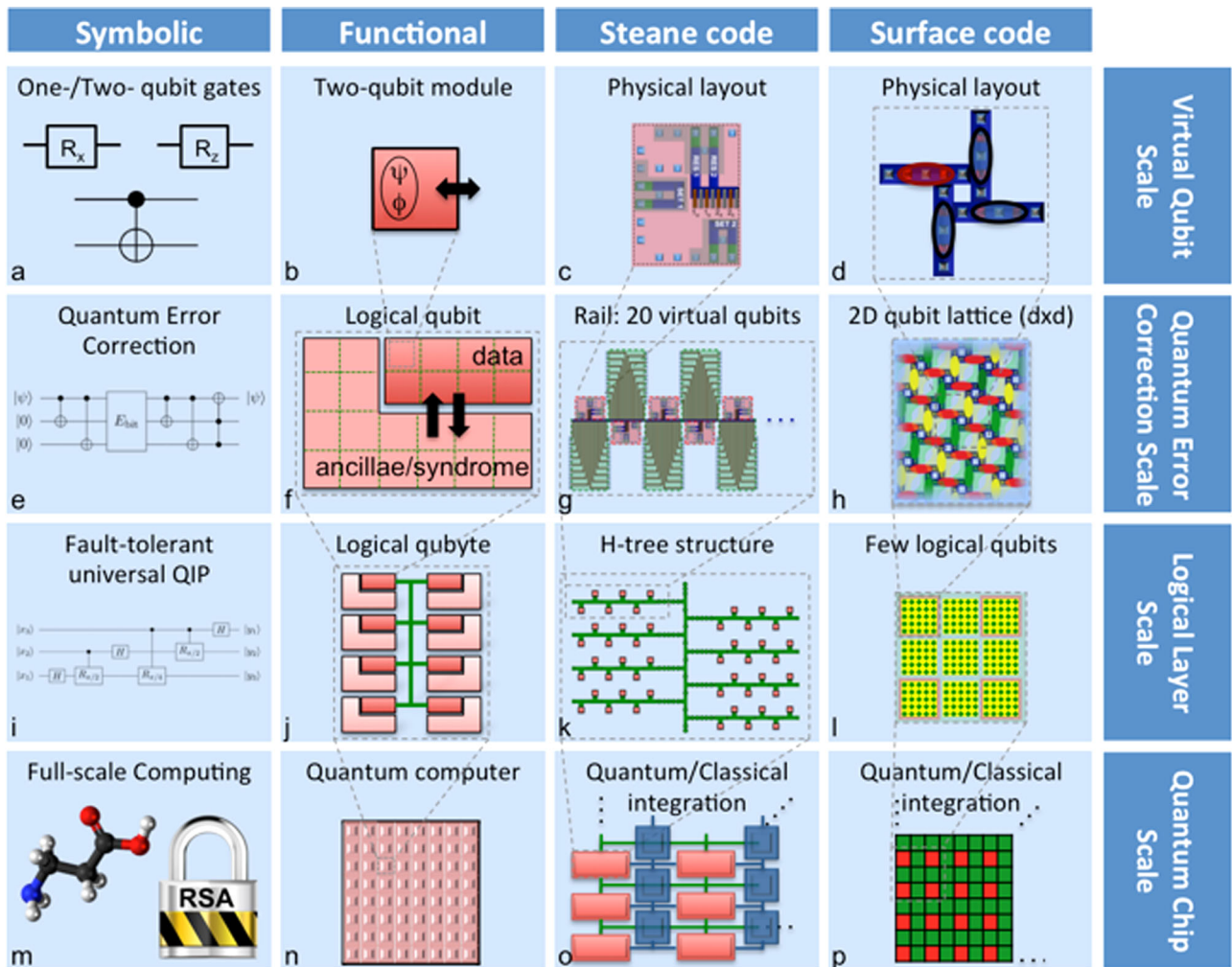
Starting from both the three building blocks D, C, and T for the Steane code [7,1,3], and the double dot qubit for the surface code, respectively, logical qubits and circuits for implementation of quantum algorithms can be designed. Figure 2 shows the symbolic notation, functional and physical representations of QIP at different scales of integration of the two architectures, from the virtual qubit to QEC circuit, to logical gate and application layer fault-tolerant quantum chip scale, respectively.

By following the typical arrangements of virtual qubits proposed for scaling to a logical qubit and as a further step to H-tree structures for concatenated codes,<sup>77, 78</sup> the first level of integration (second row in Fig. 2) for the Steane code is achieved by connecting a relatively small number (~20) of virtual qubits (device D) by means of quantum channels (modules C and T) to create the physical background for fault-tolerant quantum computing. The corresponding physical device is sketched at the third column by means of *green blocks* (communication modules, namely, C and T) and *red boxes* (data qubit D) following the symbolic representation defined in Fig. 1. This structure defines the smallest system of virtual qubits connected to a line of bidirectional quantum communication with the scope to store quantum information and correct potential errors. A number of seven virtual qubits is needed to define a logical qubit according

to the Steane's code, and about 20 virtual qubits suffice to correct potential X and Z errors on a logical qubit according to most quantum codes, e.g., the Shor's and the Steane's ones.<sup>9</sup> As long as the error rate is lower than the code error threshold, arbitrary errors occurring at the individual physical qubits can be detected and corrected in the framework of a logical qubit by employing few additional qubits as ancillae. We note that previous theoretical works identified several strategies to obtain gate fidelity approaching 99.999%.<sup>43, 70</sup> This may be good enough to enable QEC by means of the Steane code, which could deal with an error threshold in the range  $10^{-6}$ – $10^{-4}$  depending on the geometry.<sup>9, 71</sup> So, if the error rate is a factor of  $x$  better than threshold, encoding yields a final error rate approximately a factor of  $x^2$  below threshold. Further improvement can be achieved similarly by concatenation.<sup>9</sup> At any rate, we remark that only a deep understanding of the noise model applicable to the hybrid qubit will definitely set the relevant specifications for gate fidelity to be compatible with a specific QEC scheme.<sup>9, 70</sup> Such logical qubit therefore enables QEC and it is the functional building block for error-correcting quantum memory and fault-tolerant QIP. Differently, the implementation of surface code in the fourth column of Fig. 2 exploits the physical qubits in a two-dimensional nearest neighbor array so that the data virtual qubits (*red ovals*) are interleaved with ancillae for the X (*yellow*) and Z (*green*) stabilizer measurements.

The next length scale (third row) requires interconnections between multiple logical qubits to perform small-scale algorithms (in the first column, the Quantum Fourier Transform is given as an example) in a fault-tolerant manner. At such level, in order to estimate the effective area occupied by a single logical qubit by including the space needed to connect the qubits in some arrangements, for example by an H-tree structure,<sup>77, 78</sup> it is useful for the Steane code to *conventionally* define the logical qubyte or quantum byte by eight connected logical qubits, as a reminiscence of classical information processing. The qubyte allows calculating the minimum effective area of a single logical qubit by dividing by 8 its area. Such operation is not needed for the surface code architecture, which employs contiguous logical qubits. An example of operation carried at this level is provided by Quantum Fourier Transform, a key block of more complex algorithms like Shor's factoring. Finally, many logical qubit blocks are connected (fourth row) to allow control at the application layer and to implement quantum algorithms, e.g., Grover's, Shor's, and quantum simulation algorithms applied to problems where classical computation is unpractical.<sup>79</sup>

The physical size function of a logical qubit and of a logical qubyte, conventionally defined above to calculate the maximum density of the logical qubits by including interconnections, are reported in last two rows in Table 1. About the footprint of classical control circuits, the Steane code layout carries some unused area, which reaches about 23% in the logical qubit and



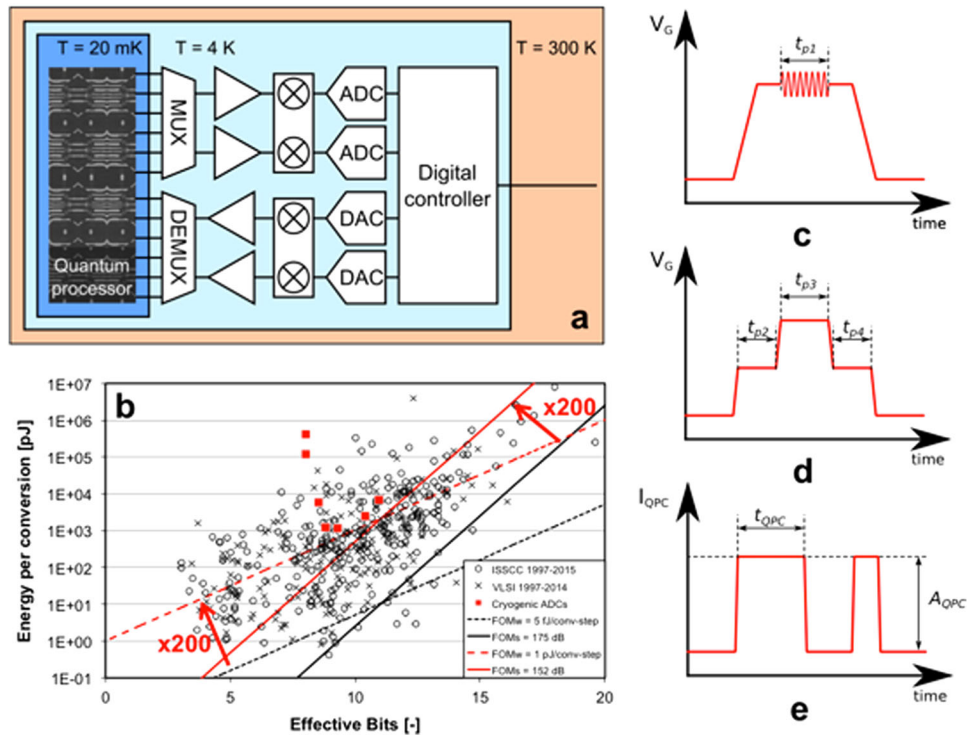
**Fig. 2** Representation of QIP at different scales of integration according to the different layers of implementation. The first column represents the different scales of integration with symbolic diagrams. The second one is a functional representation of the blocks required to execute the corresponding operations. The third column provides the physical layer implementation by the Steane code based on the modules **d**, **c**, and **t**, while the fourth column by the surface code. Each oval indicates a double QD qubit. In panel **h**, the representation of the 2D nearest neighbor array shows the virtual data qubits in *red* for the surface code and ancillae for the **x** (*yellow*) and **z** (*green*) stabilizer measurements. The four rows represent four scales of integration: the virtual qubits by their physical layout, the QEC scale for logical qubits, the logical layer scale, and finally the quantum chip scale enabling quantum algorithms such as Shor's factoring and quantum simulations. The logical layer scale assesses fault-tolerant computation from few logical qubits and it has been used to calculate the area of a single logical qubit of the Steane code by including interconnections

43% in the qubyte mask, that could be employed for classical electronics and low-level interconnections.

The surface code layout is more compact and so much more challenging to this regard, since even elementary circuits cannot fit into the small distance between physical qubits (the physical qubit pitch is of the same size order of the qubit itself as well as of a typical CMOS transistor). While increasing this feature size would be detrimental for the surface code operation and performance, a viable alternative could be to conceive a quantum processor architecture based on separated blocks of physical qubits interleaved by blocks of classical control circuits, as proposed in ref. 71. Each block, corresponding to one or few logical qubits, is connected to the neighbors by long-range quantum communication channels. Part of these blocks could be also reserved as ancillae factories to the high-throughput generation and purification of high-fidelity ancillae and cat states, which accounts for most of the resources involved in Shor's algorithm.

However, the operation of two-qubit gates within this topology and the effectiveness of such long-range couplers is still to be carefully evaluated in terms of fidelity and of required transfer bandwidth: the coherent transfer of a realistic logical qubit ( $d \approx 20-30$ ) means the coherent transfer of  $d^2 \approx 400-900$  physical qubits over few tens of  $\mu\text{m}$  on such a single track.

Anyway, the footprint of integrating classical control circuits obviously depends on the complexity of the functionality that is required within the quantum layer and this is a key point that the next-generation quantum engineer will have to deal with. This falls well beyond the scope of this review: we only note that although we did not estimate the footprint of control circuits, we obtain an estimate of the *maximum* density, i.e., a higher bound of quantum information. We will show how the evaluation of such figure of merit across different technology nodes, together with practical considerations about the hybrid qubit architecture, can give useful indications to the development of quantum computers.



**Fig. 3** Classical circuitry for qubit control and electrical signals typically used to perform operation and read out of three electron spin qubits.<sup>36, 75</sup> **a** Generic fault-tolerant interrogation correction loop, comprising both an analog front-end and a digital back-end. **b** Energy efficiency of state-of-the-art room-temperature CMOS ADCs and cryogenic CMOS ADCs.<sup>76</sup> The energy spent by an ADC for a single conversion, i.e.,  $E_C = P/f_s$ , where  $P$  is the power consumption and  $f_s$  the sampling frequency, strongly depends on its resolution, i.e., the number of effective bits  $N$ . Consequently, ADC energy efficiency is quantified by the Schreier Figure-of-Merit ( $FOM_s = 10 \log_{10} [2N f_s / (2P)]$ ) for high-resolution thermal-noise-limited ADCs and by the Walden FOM ( $FOM_w = P/f_s / 2N$ ) for low-resolution ADCs. In terms of those FOMs, the energy efficiency of existing cryogenic ADCs is 200x worse than state-of-the-art room temperature ADCs ( $FOM_s = 175$  dB,  $FOM_w = 5$  fJ/conversion step), due to the use of past node technologies and the unavailability of reliable cryogenic models for CMOS devices. **c, d** Voltage applied to the qubit gates to perform single-qubit rotations; typical waveform parameters:  $t_{p1} \sim 5\text{--}20$  ns,  $t_{p1}, t_{p2}, t_{p3} \sim 100\text{--}500$  ps, microwave burst frequency  $\sim 10\text{--}12$  GHz, pulse rise time (10–90%)  $\sim 80$  ps. **e** Typical current waveform read out at quantum point contact (QPC) to detect presence of electron in neighboring QD; typical waveform parameters:  $t_{QPC} \sim 10$  ns to 100  $\mu$ s,  $A_{QPC} \sim 200$  pA; QPC resistance  $\sim 25$  k $\Omega$

### TECHNOLOGY-DEPENDENT SCALING IN THE QUANTUM REALM

We turn now to the evaluation of the physical sizes of such integrated quantum circuits in various technology nodes from 7 to 65 nm. The mask layouts shown in Fig. 1 represent devices involving hybrid properties that cannot be univocally ascribed to any existing classical device, which rely on a fabrication precision currently unaddressed. Table 2 reports the minimum feature sizes of the above quantum devices as a function of the technology node, starting from 65 nm down to the 7 nm node by following the miniaturization from left to right.<sup>80–84</sup> We associate the contacted gate pitch  $\Delta_G$  to double the gate length of a state-of-the-art MOSFET. Forecasts of the forthcoming nodes at 10 and 7 nm were taken from the ITRS estimation of the uncontacted polysilicon pitch in flash memory. The definition of vias sets stringent requirements related to the alignment to higher-level interconnects. Therefore, the interconnect pitch  $\Delta_{IC}$  is estimated from the pitch of the first level of metal interconnects in microprocessor units.

The geometrical shape of QDs is maintained across the different technology nodes by applying an equivalent scaling of the width of the silicon islands  $w_{Si}$  and of the metal lines  $\Delta_G$ , corresponding to the lateral sizes of the QDs. We set  $l_{HDD} = \Delta_{IC}$  to provide reasonably wide doped regions for SET ohmic contacts and electron reservoirs. Finally, the undoped buffer length must be much larger than the lateral straggling range due to the dopant implantation process. Low-energy implantation results in shallow

distribution of dopants with a reduced lateral straggling of about 5–10 nm,<sup>67</sup> pointing to the reasonable condition  $l_{SU} > 20$  nm. However, a large  $l_{SU}$  size is required at SET 1 in Fig. 1 to accommodate all the adjacent gates and interconnections.<sup>70</sup> As a consequence, we set  $l_{SU} = 2\Delta_{IC}$  to allow a realistic routing of the metal wires. We note that in all the technology nodes this size is significantly larger than the dopant scattering length. Therefore, electrically active dopants in the active region of the device are excluded.

Table 3 reports the physical sizes of the three building blocks used for the Steane code, of the logical qubit and the qubyte at different technology nodes. Figure 4a compares the scaling trend of the area of a typical classical reference device, namely, SRAM 6T cell with the physical qubits, namely, the D module and the single hybrid qubit for the Steane and the surface codes, respectively. As we are interested in a possible link between classical and quantum technologies, we focus on the product of the two critical pitches, namely, the connected gate and interconnect pitch, as a relevant figure of merit to compare size scaling of classical and quantum devices. The area of the 6T-SRAM cell scales according to Moore’s law to  $\sim 17\Delta_G \cdot \Delta_{IC}$ . On the other way around, the quantum D module requires from 80 to 90 times  $\Delta_G \cdot \Delta_{IC}$ , corresponding to roughly five times the area of SRAM. As a result, the physical qubit area follows a characteristic scaling rule driven by its specific design that, however, turns out to be primarily dependent on the parameter  $\Delta_G \cdot \Delta_{IC}$  and has consequently many analogies with the classical device scaling law. For example, the change of the slope

**Table 2.** Size (nm) of the critical features at different technological nodes and corresponding references from IEDM conferences and ITRS documents

Technology node	65 nm	45 nm	32 nm	22 nm	14 nm	10 nm	7 nm
$\Delta_G$ (nm)	140	100	60	52	40	30	26
$\Delta_{IC}$ (nm)	220	160	112	90	70	64	50
$w_{Si}$ (nm)	140	100	60	52	40	30	26
$l_{BU}$ (nm)	440	320	224	180	140	128	100
$l_{HDD}$ (nm)	220	160	112	90	70	64	50
Reference	80	81	82	83	84	85	85

**Table 3.** Size ( $x$  and  $y$ ) of all the base (D,C,T) and composed (qubit  $qb$  and qubyte  $QB$ ) devices, area  $A$  of the D module, of the logical qubit and of the qubyte, and density of quantum information  $\delta_{QI}$ 

Technology node	65 nm	45 nm	32 nm	22 nm	14 nm	10 nm	7 nm
$x_D$ (nm)	1440	1040	688	568	440	384	304
$y_D$ (nm)	1760	1280	896	720	560	512	400
$x_C$ (nm)	560	400	240	208	160	120	104
$y_C$ (nm)	1900	1380	956	772	600	542	426
$x_T$ (nm)	4980	3620	2524	2032	1580	1438	1126
$y_T$ (nm)	1960	1400	840	728	560	420	364
$x_{qb}$ (nm)	39,200	28,000	16,800	14,560	11,200	8400	7280
$y_{qb}$ (nm)	9820	7140	4988	4012	3120	2846	2226
$x_{QB}$ (nm)	83,580	60,100	36,300	31,252	24,040	18,270	15,730
$y_{QB}$ (nm)	58,960	45,040	29,408	23,720	18,280	15,896	12,808
$A_D$ ( $\mu\text{m}^2$ )	2.5344	1.3312	0.616448	0.40896	0.2464	0.196608	0.1216
$A_{qb}$ ( $\mu\text{m}^2$ )	384.944	199.92	83.7984	58.41472	34.944	23.9064	16.20528
$A_{QB}$ ( $\mu\text{m}^2$ )	4927.877	2706.904	1067.510	741.297	439.451	290.420	201.470
$\delta_{QI}$ (Mqb/cm <sup>2</sup> )	0.162	0.296	0.749	1.079	1.820	2.755	3.971

The latter is expressed in units of logical qubits per unit area at the different technology nodes. The estimate for the 22 nm node reported in Ref. [71] has been refined on the basis of more detailed and quantitative considerations on the arrangement of the metal gates and interconnections in the device masks.

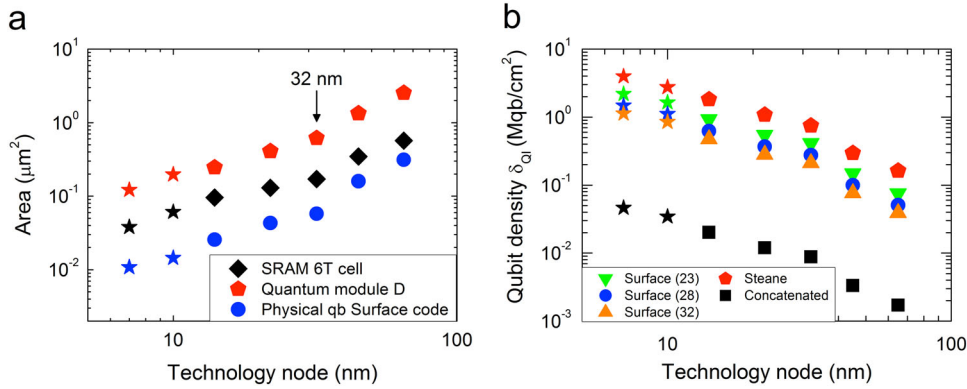
below 32 nm is mainly due to the slowdown of the transistor gate length reduction.

In the case of the Steane code, the resulting density of quantum information is defined  $\delta_{QI} = 8A_{QB}^{-1}$  in units of logical qubits per unit surface, and it is plotted in Fig. 4b as a function of the technology node. In order to express quantum computational power, the trend line refers to the technology node. As we are interested in determining the maximum quantum computational power storable in a solid-state manufacturable chip, we link the maximum amount of quantum information to the technology node. We anticipate that considerations on the operation time obey to additional constraints, which exclude the largest technology nodes from any practical applications. The quantum information density for the Steane code is of the order of Mqubits/cm<sup>2</sup>, but it decreases by two orders of magnitude if one shifts to concatenated codes. Figure 4b includes the physical footprint of logical qubits encoded by surface codes with several code distance values  $d$  for a comparison with Steane code and recursive coding, giving an intermediate value. Such  $d$  values correspond to the minimum resources needed to apply Shor's factorization algorithm to 128, 1024, and 8192 bit integer numbers. This data set covers a wide range of realistic problems including decrypting RSA systems that are actually considered safe with present and near-mid future technology of classical computing.

As noted above, the trend line reflects the change of the slope at the 32-nm node of the Moore's law, mainly associated with the change of the scaling law of the gate length with respect to the technology node. The two facts are only indirectly related, as the slowdown in the Moore's law depends on the effort capability of semiconductor industry and not on some intrinsic physical constraint. As a matter of fact, the introduction of new technologies and materials (strained silicon channel and high- $k$  gate stack, for example) has been even more important than bare geometrical scaling to maintain the equivalent scaling of device performances during the last two decades.<sup>85</sup> Moore's law steered semiconductor industry to the evolution from bulk planar transistors used in 65-nm node to silicon on insulator (SOI) devices up to the FinFET geometry adopted in the present 14-nm node.<sup>84</sup> Preliminary studies predict that new materials other than Si could be employed as the channel material and three-dimensional integration may become a cheaper alternative to continuing two-dimensional scaling to increment functional complexity of integrated circuits.<sup>85</sup>

In such a varying framework, some additional issues must be considered that are specific of ultra-scaled quantum technology and may generate significant deviations from the pathway foreseen for classical CMOS electronics. The main point is related to the complex design of qubit devices, which encompasses different building blocks, such as a single charge sensor and multiple QDs, with a large number of critical gates and





**Fig. 4** The quantum information density scaling law. **a** Comparison between the trend line of the area occupied by a reference SRAM classical cell and a CMOS physical qubit for both Steane (module **d**) and surface codes, respectively. The change in the trend at 32-nm node is due to different scaling law of the gate length below 32 nm. Stars refer to the forecasts of future nodes, namely, 10 and 7 nm, not yet available for industrial production. References for SRAM cells are the same reported in Table 2. **b** The quantum information density scaling law expressed as the maximum processable quantum information in terms of logical qubit number per surface unit as a function of the technology node. The concatenated code (black) is based on the Steane code [7,1,3] and lowers the quantum information density of two orders of magnitude with respect to the bare Steane code (red). The quantum information density is calculated for the surface codes at typical  $d$  values (green for  $d = 23$ , blue for  $d = 28$ , and orange for  $d = 32$ ). The stars correspond to forthcoming nodes, not yet realized and therefore subject to possible specification changes. The existing nodes and future ones (stars) are derived from IEDM and ITRS published specifications following the analysis in Table 1

components that are all fundamental for the device operation. To this extent, process variability issues are expected to be dramatically increasing, with respect to classical electronics, in view of large-scale implementation of CMOS qubits. As a consequence, the development of solid design tolerant to process variability must have a key role in the development of such devices. The device mask could be significantly made simpler by introducing an emerging single-charge sensing technique based on rf-reflectometry, where the charge sensor is fully integrated with the QD gate control as proposed for the surface code implementation.<sup>73, 74</sup> With this solution, the two SETs in the D module would be unnecessary for charge sensing as well as one of the electron reservoirs, with beneficial outcomes in terms of device area and complexity (the number of critical gates would reduce by a factor of 2). Another advantage would be much faster quantum state read-out, without the slow electron tunneling time limiting the procedure at μs timescales. On the other side, such technique will require intensive use of rf signals, with the potential drawbacks of a more challenging classical control circuitry and pronounced cross-talk. Although a large part of the technological challenges toward large-scale quantum computing are shared with the development of end-of-roadmap classical technology nodes, realistic qubit implementations, while being compatible with classical electronics manufacturability limits, will also presumably come to terms with the best compromise between stringent design rules imposed at the quantum level and limited cross-talk arising from classical control circuitry.

### OPERATION TIME CONSTRAINTS AND SOURCES OF DECOHERENCE

The figures of merit of a physical qubit, such as the energy spectrum of the QD and the intensity of the inter-dot effective interactions, are influenced by the different size of the devices achievable at the different technology nodes.<sup>86</sup> The vertical constraint on the silicon thickness ( $t_{Si} \ll w_{Si}$ ) is related to the transverse size  $w_{Si}$  and it may contribute to valley splitting so that a controlled valley states filling is obtained<sup>43, 60</sup> as preferable for the three spin exchange-only qubit.<sup>47</sup> Notably, such requirement is implicitly granted by ultra-thin body SOI technology to be employed for nodes beyond 14 nm. Valley splitting can be further increased up to ~1 meV by applying vertical electric field.<sup>87–89</sup> It is

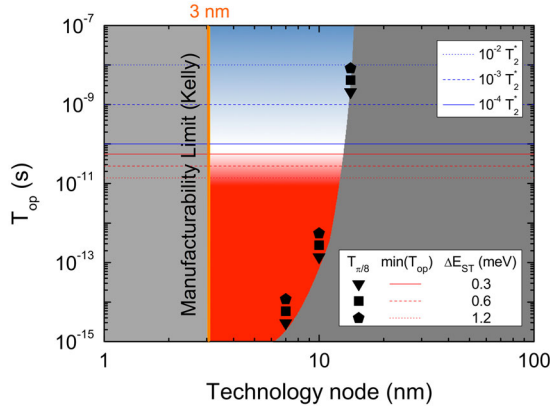
worth mentioning that the complementary realization of the CMOS qubits by hole spins would avoid such issues,<sup>55</sup> while a hypothetical realization by donor QDs<sup>63, 90</sup> would natively induce an adequate valley splitting.<sup>91</sup>

Virtual qubit manipulation is mediated by the effective exchange interaction  $J = t^2/\epsilon$ , where  $t$  is the tunnel coupling between the energy levels in the two QDs and  $\epsilon$  is the system detuning.<sup>37, 47</sup> During qubit operations,  $t$  and  $\epsilon$  are regulated by the inter-dot electrostatic barrier and by the electrochemical potential of the QDs, respectively, that are controlled through gate electrodes. Virtual gate operation frequency is therefore strongly dependent on the physical size of the QDs, since it is directly proportional to the maximum exchange interaction:

$$J_{max} = \frac{t^2}{\Delta E_{ST}}, \quad (2)$$

where  $\Delta E_{ST}$  is the S-T splitting.<sup>52</sup>

Tunnel coupling has an exponential dependence on the inter-dot distance, that in our case is equal to the metal gate pitch  $\Delta_G$ .<sup>52</sup>  $\Delta E_{ST}$  also depends on size and geometry of the QDs, though an additional fine tuning is possible due to a weak dependence on vertical electric field.<sup>92</sup> As a result, faster operations would be possible in principle at the ultimate technology nodes, with advantages in terms of fidelity. However, an upper bound to operational speed is set by the tunnel coupling being smaller than the single particle level spacing and the S-T energy splitting.<sup>43</sup> In order to identify the time operation window compatible with physical constraints as a function of the technology node, we consider the three following requirements: *adiabaticity*, *coherence*, and *node-dependent minimum operation time*. First, adiabaticity means that tunnel coupling  $t$  must be lower than half the S-T splitting  $t < \Delta E_{ST}/2$ .<sup>43</sup> Since the operation time is  $T_{op} \sim h/J_{max}$ , where  $h$  is the Planck constant and  $J_{max}$  is the maximum effective exchange interaction as defined in Eq. (2), then such condition reads  $T_{op} > \frac{4h}{\Delta E_{ST}}$ . Second, coherent qubit manipulation is obtained provided that the logic gate operation time is much smaller than the qubit dephasing time:  $\frac{T_{op}}{T_2} < \eta$  where  $\eta$  is the error threshold for QEC.<sup>9</sup> Third, logical operation time has also a lower bound imposed by Eq. (2) which is dominated by the exponential dependence of the tunnel coupling  $t$  on the technological node.<sup>70</sup> For example, for a  $\pi/8$  rotation gate, operation time is the



**Fig. 5** Physical constraints for QIP with Si hybrid qubits. Technological limits prohibit quantum computation in the domain of *gray* areas due to manufacturability limits on the left side (set by the vertical orange line at 3 nm) and by the maximum effective exchange interaction on the right side (*black scatters*). *Black data points* denote the minimum operation time for a  $\pi/8$  rotation that can be achieved at each technology node with the realistic values of  $\Delta E_{ST} = 0.3, 0.6,$  and  $1.2$  meV. *Red lines* indicate the minimum operation time imposed by adiabaticity at the same three values of S-T splitting of  $\Delta E_{ST}$ . *Blue lines* set the maximum operation time required by coherence requirement at three different error thresholds ( $\eta = 10^{-4}, 10^{-3},$  and  $10^{-2}$  errors per logic gate) characteristic of available coding techniques, namely, Steane code, concatenated/color codes, and surface codes, respectively. The ideal dephasing time  $T_2^* = 1 \mu\text{s}$  is taken according to theoretical predictions in ref. 36. The *white area* represents the optimal regime of operation for QIP with the Si qubit architecture here considered. Reasonable operating frequencies are in the order of 10 GHz, with bandwidth depending on the optimization of the QD parameters ( $\Delta E_{ST}, t$ ) and of the QEC scheme compatibility with the requirements of coherence and adiabaticity

following:<sup>38</sup>

$$T_{\pi/8} = 3.598 \frac{h}{J_{\max}} = 3.598 \frac{h \Delta E_{ST}}{t^2}. \quad (3)$$

In Fig. 5 we summarize the constraints related to the operating frequency of Si exchange-only qubits for realistic devices, including the technological limit of manufacturability at 3 nm (*orange line*). This limit was discussed and quantified by Kelly in 2011 by considering a number of arguments based on the intrinsic variability of top-down fabrication processes, on the unwanted electron tunneling and on the parasitic interferences, leading to a limit on the manufacturability at around 3 nm.<sup>3</sup> Furthermore, we highlight the physical bounds imposed by the requirements of adiabaticity, coherence, and node-dependent minimum operation time, which hamper quantum computing in the *blue, red,* and *gray* areas, respectively. The realistic operation range covered by *white areas* is centered at 10 GHz and, depending on the intensity of the exchange interaction and  $\eta$ , it spans one decade toward both lower and higher frequency, resulting therefore of about 1–100 GHz. It is a favorable circumstance that, as shown in the [Supplementary Information](#) and discussed later, such range covers the working frequency of most classical few qubit control circuits.

The main sources of decoherence are hyperfine interaction, electrostatic noise, and electron interaction with optical phonons.<sup>39, 43</sup> The first experimental demonstration of the exchange-only double QD qubit showed  $T_2^*$  of 10 ns with operation time of the order of 100 ps with Si-SiGe QD. More in detail, 5.2 GHz X-rotations and 11.5 GHz Z-rotations were coherently driven with fidelity of 85 and 94%, respectively.<sup>36</sup> Preliminary results were recently demonstrated for two-qubit gates in an analogous system.<sup>51</sup> Notably, promising results were obtained for single spin<sup>35</sup> and S-T<sup>34</sup> qubits in the same material, indicating significant improvement when spin-less material is employed to reduce spin

decoherence. Indeed, hyperfine interaction and magnetic field fluctuations, that are mainly responsible for decoherence in III-V semiconductors due to the high density of spin-carrying nuclei, are effectively inhibited by utilizing purified  $^{28}\text{Si}$ .<sup>25, 35, 46, 93</sup> Alternatively, hole spin qubits could be considered in the future<sup>53–56</sup> to reduce hyperfine interaction. Another significant noise component arises from the interface in the case of Si-SiO<sub>2</sub> QDs in particular. In fact, interface and oxide traps act as charge fluctuators that induce dephasing due to slow variations of QD potential and high-frequency random telegraph noise.<sup>94</sup> Such consideration is even more topical for etched SOI MOS nanostructures such those considered above, which confine electrons with a small number of electrostatic gates.<sup>95</sup> On the one hand, etched Si nanostructures provide enhanced electron confinement, leading to increased charging energy with respect to planar electrostatically defined QD, as well as large orbital and valley splitting.<sup>60</sup> On the other hand, increased surface to volume ratio will lead to a major impact of interface-related noise with respect to spin noise especially at the most aggressive technology nodes. In this regard, adequate device post-processing is effective in leading to very low density of defects of  $\sim 3 \times 10^{10} \text{ cm}^{-2}$ ,<sup>96</sup> corresponding to an average distance between defects of 60 nm. Furthermore, large sweet spots can be normally recognized in the energy diagram of the exchange-only qubit where robust qubit rotations can be performed with small impact from random fluctuations of the QD potential.<sup>36, 43</sup> We also note that, on scaling down to the most extreme technology nodes, a leading role in decoherence mechanism will be taken by electron-phonon interaction instead of charge noise.<sup>97</sup> Such phenomena must be effectively suppressed by limiting the material disorder and by cooling the system at temperatures of the order of 10 mK that are actually at hand in state-of-the-art dilution refrigerators.<sup>35, 46, 87, 88</sup> As a final remark, the reduction of disorder and of charge noise are apparently among the main challenges to reach the ideal coherence time of  $\sim 1 \mu\text{s}$ <sup>37</sup> enabling large-scale quantum computing within the stringent limits represented in Fig. 5.

## SCALING-UP QUANTUM TECHNOLOGY: QUANTUM COMMUNICATION AND ERROR CORRECTION

We now turn to quantum communication, which becomes a fundamental aspect in some large-scale fault-tolerant QIP architectures. The Steane code is compatible with coherent transfer of quantum information through different methods, namely, the SWAP chain protocol, the CTAP, and the qubit shuttle. The first method is based on the sequential repetition of SWAP gates between adjacent qubits. From previous estimates, for a 40-nm inter-dot distance, quantum information transfer between two adjacent data qubits (at a distance of 1  $\mu\text{m}$ ) can be carried out in 71.2 ns.<sup>70</sup> The SWAP operation speed is directly proportional to the maximum exchange interaction, that in turn has an exponential dependence on the inverse of the inter-dot distance  $\Delta_G$ .<sup>38</sup> CTAP is a viable alternative for long distances, since the time required for quantum communication has a sub-linear dependence on the distance between qubits.<sup>98, 99</sup> CTAP has an additional advantage over the SWAP chain: initial loading of qubits with three electrons is not required. Qubit shuttle, finally, has been proposed for quantum communication in large-scale single spin qubit architectures.<sup>77, 78</sup> According to the latter, quantum information is coherently transferred by moving the potential well, which confines the qubit along the communication channel.

The only condition that is determined by adiabatic motion is  $mv^2 \ll \Delta_G$ , where  $m$  is the electron mass and  $v$  is the speed of the flying qubit. Such requirement corresponds to a favorable upper bound of  $v \ll 10^4 \text{ m/s}$ , which is limited mainly by the speed of the classical control circuitry. Both SWAP chain and CTAP protocols cannot be used for the simultaneous transfer of many qubits

along the same quantum channel, but they can be employed for short-range quantum communication, e.g., during QEC. Conversely, qubit shuttle may be better exploited to transfer entire logical qubits over long distances, relying on the high operating speed. Alternatively, it is worth mentioning that surface code architecture would allow more relaxed requirements as they rely on neighboring qubit sites.

Teleportation may be another interesting option for massive long-range communication, as for the transfer of a whole encoded qubit to a different site. Moreover, teleportation was also proposed to inherently perform quantum gates at a distance by means of entanglement between far away qubits.<sup>77, 78</sup> Such methods for quantum communication are compatible with the proposed hybrid qubit implementation. However, the final choice of the method (or the methods) to be implemented will be necessary driven by considering the effective fidelity and bandwidth offered by each method to meet the requirements of a specific algorithm implementation.

Error threshold strongly depends on the specific code definition, but also on the complexity of the routines for quantum error detection and correction, involving several logic gate sequences and intensive transfer of qubits. Typical error thresholds range between  $10^{-6}$  and  $10^{-3}$  errors per logic gate, depending on the code properties, the qubit arrangement, and the physical implementation. An extensive review of quantum codes and QEC techniques can be found in ref. 9. The first experimental demonstration of a double QD exchange-only qubit reached a fidelity of the order of  $\sim 90\%$ ,<sup>36</sup> which is insufficient for QEC purposes. However, theoretical studies indicate that much lower error rates, approaching  $10^{-4}$ , can be achieved by improving device quality and by optimizing qubit parameters and control sequences.<sup>43, 48</sup> Analogous development will be reasonably required for two-qubit gates to obtain fault-tolerant fidelity.<sup>45, 47, 51</sup> In order to discuss the worst case applied to a CMOS implementation already discussed in literature, we consider both the Steane code and the surface code, which are potentially capable to deal with such specifications. The seven-qubit Steane code constitutes the fundamental building block for advanced coding techniques, such as recursive coding (which leads to an improvement of the gate fidelity on encoded qubits) and topological color codes (which lead to error thresholds of the order of  $10^{-3}$ ).<sup>100, 101</sup> Therefore, it helps to define a benchmark to estimate the minimum physical resources required to implement QEC with the exchange-only Si-QD qubit architecture. Furthermore, the area of a logical qubit for recursive coding is of the order of  $A_D(A_{qb}/A_D)^2$  as the Steane code logical qubits are—in this case—employed as intermediate building blocks to implement a higher level doubly encoded qubit. Circuit complexity and computational time will scale similarly. It is worth mentioning that among alternative QEC options, recursive coding and topological codes may lead to lower error rates on the encoded qubits, thus relaxing the constraints of the operation time,<sup>17, 18</sup> with the payoff that the quantum information density can be drastically reduced as shown in Fig. 4b.

Besides the protection against bit-flip and phase errors, another important aspect consists of the mitigation of the leakage errors, of charge noise disturbance and electron loss.

QEC only applies to errors within the logical subspace, therefore additional correction is needed to correct leakage errors. For surface code architecture, Preskill<sup>102</sup> described a gate sequence that detects leakage errors of a data qubit (D) by an ancilla qubit (A), only needed a the leakage detection unit (LDU). The method is based on measurements to detect if leakage has occurred, so the LDU inverts A if D has not leaked, while A remains unchanged if leakage has occurred. Later Mehl et al.<sup>72</sup> demonstrated a similar method called leakage reduction unit (LRU) for which the measurement process is not necessary to correct for leakage. In the case of leakage events, the definitions of D and A are then interchanged after the LRU. For Steane-code, a T-gate network has

been proposed to detect leakage errors (spin flips) and to replace them with errors within the logical subspace. In general, within the framework of a universal set of quantum logic ports, electron loss, similarly to spin flips, are mitigated by teleportation-based gates for which the data qubit is replaced with a new ancillary qubit.<sup>77, 78</sup>

The effects of the charge noise have been mitigated significantly by tuning the qubit energy dispersion, which is a function of the detuning between the two QDs.<sup>103</sup>

Blue lines in Fig. 5 indicate the physical bounds imposed by three representative values of error threshold spanning from  $10^{-4}$  (reachable by Steane code)<sup>104</sup> up to  $10^{-3}$  (concatenated and color codes)<sup>100, 101</sup> and  $10^{-2}$  (surface codes),<sup>20</sup> highlighting the importance of the coding technique in view of scalable quantum computing. To this purpose, Steane code approach features lower error threshold but it likely has an easier integration with the classical control circuitry. Alternatively, surface code-based implementation is simpler at the quantum layer and more robust against errors at the expense of very challenging demands in classical circuitry realization. In any case, a careful analysis will still be mandatory to individuate the best coding technique for a given implementation of QEC, compatibly with the nominal fidelity of logic gates and the time load required for quantum communication during QEC operations in CMOS technology.

Looking for a fair estimation of the physical resources for quantum computation, most of the space and time resources needed to run a quantum algorithm (e.g., Shor's factorization) is generally devoted to ancilla preparation for non-Clifford gates, such as phase gates, or Toffoli gates.<sup>20</sup> Purification up to reasonably low-error rates requires very large logical qubits, including few thousands of physical qubits. Surface codes with a minimum code distance  $d=23$  is necessary to deal with computational problems of practical interest such as the factorization of a 128 bit number.<sup>20</sup> The code distance is the minimum weight of a logical operator, i.e., the minimum number of physical bit flips to define a logical qubit. Lattice surgery technique allows universal and scalable computation within two-dimensional surface codes with physical resources of  $\approx 8d^2$  qubits per logical qubit.<sup>21</sup>

Notably, Table 4 shows that even at the largest scale problem here considered (8192 bit factorization) the quantum computer size could be limited to few mm, which is compatible with commercial microelectronic packaging on the one hand and with the use of state-of-the-art cryostats for efficient cooling on the other hand. With regard to the execution time, we also mention that it could be further reduced either by choosing optimized versions of the algorithm<sup>20</sup> or by improving the 100 ns readout time set in our calculation, e.g., by means of the faster rf-reflectometry readout scheme mentioned above.<sup>73, 74</sup>

## SCALABLE CLASSICAL CMOS CONTROL ELECTRONICS FOR FAULT-TOLERANT OPERATIONS

In double QD qubit, similarly to other embodiments, controlling and interrogating virtual qubits in a loop for QEC involves the generation of nanosecond scale pulses of amplitude-modulated voltage or current. Pulses are controlled in amplitude with resolution of at least 10 bits and duration of several tens of nanoseconds, with a time resolution better than 10 ns. Since the generation of these signals is done independently and in parallel for each virtual qubit to assess dynamical decoupling and compensation schemes on related virtual gates, it is necessary to implement concurrent QEC loops, each connected to a single virtual qubit, similar to refresh operation of a dynamic random access memory. As a classical electronic interface operated at room temperature involves problems such as linearity of interconnections with the number of qubits, linearly scaled thermal flux to the quantum system, and power dissipated to control each qubit before being attenuated of several orders of

**Table 4.** Minimum space occupied by the resources for running Shor's prime number factoring algorithm to numbers of different magnitudes

N bits	Data qubits	Surface code d	Physical qubits	Area quantum computer (mm <sup>2</sup> )			Time (h)
				14-nm node	10-nm node	7-nm node	
128	256	23	4.21E+08	10.8	6.1	4.6	0.01
256	512	26	5.48E+08	14.0	7.9	5.9	0.06
512	1024	28	6.77E+08	17.3	9.8	7.3	0.45
1024	2048	30	8.25E+08	21.1	11.9	8.9	3.58
2048	4096	32	8.79E+08	22.5	12.7	9.5	28.63
4098	8196	33	1.15E+09	29.5	16.6	12.5	229.06
8192	16,384	35	1.22E+09	31.1	17.5	13.2	1832.52

The surface area of the entire quantum computer is estimated for the last three technological nodes according to Fig. 3b. Running time is calculated on the basis of a surface code cycle time of 200 ns which is mainly limited by the measurement time [20, 90]. The implementation refers to  $2N$  logical qubits and  $40N^3$  sequential Toffoli gates to minimize occupied space at the cost of maximum computation time [20].

magnitude (often 60–100 dB), cryogenic multiplexing in CMOS is being developed.<sup>105–107</sup> A generic implementation of a fault-tolerant loop is shown in Fig. 3a. The performance observed at cryogenic temperature of the control electronics suggests that three types of multiplexing can be used based on time-division multiple access (TDMA), frequency-division multiple access, and space-division multiple access to significantly reduce the number of interconnects and reduce power dissipation to the cooling power of the refrigerator. The creation of TDMA multiplexers capable of operating at mK temperatures is required, while TDMA demultiplexers and the remainder of the loops can already operate at 1.6–4.2 K.<sup>105, 107</sup> The programming of the digital back-end and of the analog front-end can be done with high-speed serial lines, thus minimizing the number of interconnects connecting room temperature devices to cold circuits. In order to minimize the wiring requirements, a first layer of electronic control can be implemented as close as possible to the qubits in terms of temperature and/or physical distance, either on the same silicon substrate or via three-dimensional integration.<sup>71, 108</sup> This first layer would enable the aforementioned multiplexing schemes and, consequently, a simplified wiring toward higher electronic control layers, which can then be operated at higher temperature with relaxed power consumption and size constraints.

Cryogenic electronics has been used in several applications to improve electronics performance, e.g., by reducing the thermal noise in readout for high-energy or nuclear physics experiments,<sup>109</sup> or to serve in harsh environments, e.g., in space applications.<sup>110</sup> A few attempts have been made to interface quantum devices with cryogenic electronics to reduce the wiring toward room temperature,<sup>26–28, 105, 111, 112</sup> but those approaches have been limited to one or two quantum devices, thus not truly addressing the scalability issues of quantum computers. However, by relying on the progress of semiconductor technology, only CMOS technology can currently offer the integration of billions of transistors on a single chip, while ensuring low-power consumption, reliability, and functionality down to 30 mK.<sup>113</sup> Beside simple cryogenic CMOS amplifiers,<sup>26, 27, 114</sup> complex cryogenic analog circuits have been implemented in CMOS, including a full 400 MHz transceiver operating at 173 K<sup>115</sup> and several 4 K analog-to-digital converters (ADC), such as successive approximation register (see Note),<sup>116</sup> Sigma Delta,<sup>117</sup> and Flash ADCs.<sup>76</sup> However, such devices operate at a relatively high temperature ( $\gg 4$  K) or they show poor power efficiency. The energy efficiency of cryogenic CMOS ADCs is above 1 pJ/conversion step,<sup>76, 117</sup> which is 200 times worse than state-of-the-art CMOS ADCs operating at room temperature,<sup>118</sup> as shown in Fig. 3b. This would result in a power consumption over 200 mW for a 10 bit 200 MSa/s ADC, as required in the quantum-processor controller. The large gap to the state-of-the-art can be

attributed to the use of older technologies (feature size  $> 0.35 \mu\text{m}$ ) and/or to the adoption of conservative circuit topologies designed to be robust to cryogenic non-idealities even when accurate device models are unavailable. State-of-the-art performance at 4 K can be achieved by implementing aggressive circuit techniques, such as digitally assisted analog blocks and time-domain signal processing, in technology nodes beyond 40 nm CMOS and by developing relative accurate device models. Such approach, combined with an extensive multiplexing strategy, may allow meeting the objective of power dissipation in the order of 1 mW per qubit channel, thus enabling the operation of thousands of cryogenic concurrent fault-tolerant loops in existing refrigerators with cooling power in the order of 1 W at 4 K. The adoption of ultra-scaled CMOS technologies combined with the increase of the carrier mobility at 4 K may allow the operation of amplifiers, down-converter, and up-converter at few GHz. As last remark, we observe that the operation frequency of the cryogenic electronics is compatible with the timescale of the silicon qubits discussed above (see [Supplementary Material](#)).

#### THE FUTURE OF CMOS QIP

Many of the advances described above for QIP in silicon have been mirrored, either earlier or later, by similar advances in III–V semiconductors and superconductors. The latter offer the advantage that they exploit co-integration of the control electronics and more relaxed size constraints, respectively. The shorter coherence time and the large size provide also comparable disadvantages. Silicon platform is perhaps further ahead than III–V and superconductive qubits in areas where scalability and nanometric size can be used, but is behind in the control of either the electron or hole wavefunction. Furthermore, the qubit operation timescale is compatible with the GHz range operation frequency of CMOS cryogenic control electronics, so a fully integrated approach would be possible. The prospects for QIP devices realized in silicon may be bright in the long run because of the superior coherence time and scalability properties. It is also a great advantage that <sup>28</sup>Si lacks of spin-orbit interaction. As advances continue in modeling, manipulation, control and devices, the field of silicon CMOS qubits seems to be progressing to the point where the ultimate scaling of the transistor coincides with the most suitable embodiment for feasible massively parallel spin-based QIP. Such trend may drive the semiconductor device community to the end of the roadmap not in the classical sense, but by moving to the quantum domain represented by CMOS-based quantum computers.

## Data availability

Data sharing not applicable to this article as no data sets were generated or analyzed during the current study.

## ACKNOWLEDGEMENTS

The authors would like to thank D.P. DiVincenzo, H. Bluhm, and S. Mehl for useful discussions.

## AUTHOR CONTRIBUTIONS

D.R. and E.P. reviewed the silicon quantum computer architectures. F.S. and E.C. reviewed the silicon classical control electronics. All authors researched, collated, and wrote this paper.

## ADDITIONAL INFORMATION

**Supplementary Information** accompanies the paper on the *npj Quantum Information* website (doi:10.1038/s41534-017-0023-5).

**Competing interests:** The authors declare no competing financial interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## REFERENCES

- Jones, N. C. et al. Layered architecture for quantum computing. *Phys. Rev. X* **2**, 31007 (2012).
- Moore, G. E. Cramming more components onto integrated circuits. *Electronics* **38**, 114 (1965).
- Kelly, M. J. Intrinsic top-down unmanufacturability. *Nanotechnology* **22**, 245303 (2011).
- Prati, E. & Shinada, T. Atomic scale devices: Advancements and directions. in *IEEE Int. Electron Devices Meeting (IEDM)* pp. 1.2.1–1.2.4.
- Shor, P. W. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Sci. Stat. Comput.* **26**, 1484 (1997).
- Montanaro, A. Quantum algorithms: an overview. *npj Quantum Inf.* **2**, 15023 (2016).
- Shor, P. Scheme for reducing decoherence in quantum computer memory. *Phys. Rev. A* **52**, 2493–2496 (1995).
- Steane, A. Error correcting codes in quantum theory. *Phys. Rev. Lett.* **77**, 793–797 (1996).
- Devitt, S. J., Munro, W. J. & Nemoto, K. Quantum error correction for beginners. *Rep. Prog. Phys.* **76**, 76001 (2013).
- Calderbank, A. R. & Shor, P. W. Good quantum error-correcting codes exist. *Phys. Rev. A* **54**, 1098–1105 (1996).
- Laflamme, R., Miquel, C., Paz, J. P. & Zurek, W. H. Perfect quantum error correcting code. *Phys. Rev. Lett.* **77**, 198–201 (1996).
- Bennett, C. H., DiVincenzo, D. P., Smolin, J. A. & Wootters, W. K. Mixed-state entanglement and quantum error correction. *Phys. Rev. A* **54**, 3824–3851 (1996).
- Steane, A. M. Active stabilization, quantum computation and quantum state synthesis. *Phys. Rev. Lett.* **78**, 2252 (1997).
- Steane, A. M. Fast fault-tolerant filtering of quantum codewords. Preprint at <http://arXiv.org/quant-ph/0202036> (2002).
- DiVincenzo, D. P. & Aliferis, P. Effective fault-tolerant quantum computation with slow measurement. *Phys. Rev. Lett.* **98**, 20501 (2007).
- Knill, E. Quantum computing with realistically noisy devices. *Nature* **434**, 39–44 (2005).
- Bombin, H. & Martin-Delgado, M. A. Topological computation without braiding. *Phys. Rev. Lett.* **98**, 160502 (2007).
- Katzgraber, H. G., Bombin, H., Andrist, R. S. & Martin-Delgado, M. A. Topological color codes on Union Jack lattices: a stable implementation of the whole Clifford group. *Phys. Rev. A* **81**, 12319 (2010).
- Bombin, H. Clifford gates by code deformation. *New J. Phys.* **13**, 43005 (2011).
- Fowler, A. G., Mariantoni, M., Martinis, J. M. & Cleland, A. N. Surface codes: towards large-scale quantum computation. *Phys. Rev. A* **86**, 32324 (2012).
- Horsman, C., Fowler, A. G., Devitt, S. & Meter, R. Van Surface code quantum computing by lattice surgery. *New J. Phys.* **14**, 123011 (2012).
- Zizzi, P. A. Holography, quantum geometry, and quantum information theory. *Entropy* **2**, 39–69 (2000).
- Vedral, V. Information and physics. *Information* **3**, 219–223 (2012).
- Rotta, D. & Prati, E. in *Silicon Nanomaterials Sourcebook* (ed. Sattler, K. D.) (CRC Press Taylor and Francis Group, 2017).
- Morton, J. J. L., McCamey, D. R., Eriksson, M. A. & Lyon, S. A. Embracing the quantum limit in silicon computing. *Nature* **479**, 345–353 (2011).
- Ferrus, T. et al. Cryogenic instrumentation for fast current measurement in a silicon single electron transistor. *J. Appl. Phys.* **106**, 33705 (2009).
- Guagliardo, F., Ferrari, G. in *Single-Atom Nanoelectronics* (eds. Prati, E. & Shinada, T.) 187–210 (Panstanford, 2013).
- Clapera, P. et al. Design and cryogenic operation of a hybrid quantum-CMOS circuit. *Phys. Rev. Appl.* **4**, 44009 (2015).
- Ono, Y. et al. Fabrication method for IC-oriented Si single-electron transistors. *IEEE Trans. Electron Devices* **47**, 147–153 (2000).
- Takahashi, Y., Ono, Y., Fujiwara, A. & Inokawa, H. Silicon single-electron devices. *J. Phys. Condens. Matter* **14**, R995–R1033 (2002).
- Petta, J. R. et al. Coherent manipulation of coupled electron spins in semiconductor quantum dots. *Science* **309**, 2180–2184 (2005).
- Koppens, F. H. L. et al. Driven coherent oscillations of a single electron spin in a quantum dot. *Nature* **442**, 766–771 (2006).
- Medford, J. et al. Self-consistent measurement and state tomography of an exchange-only spin qubit. *Nat. Nanotechnol.* **8**, 654–659 (2013).
- Maune, B. M. et al. Coherent singlet-triplet oscillations in a silicon-based double quantum dot. *Nature* **481**, 344–347 (2012).
- Kawakami, E. et al. Electrical control of a long-lived spin qubit in a Si/SiGe quantum dot. *Nat. Nanotechnol.* **9**, 666–670 (2014).
- Kim, D. et al. Quantum control and process tomography of a semiconductor quantum dot hybrid qubit. *Nature* **511**, 70–74 (2014).
- Shi, Z. et al. Fast hybrid silicon double-quantum-dot qubit. *Phys. Rev. Lett.* **108**, 140503 (2012).
- Michielis, M. De, Ferraro, E., Fanciulli, M. & Prati, E. Universal set of quantum gates for double-dot exchange-only spin qubits with intradot coupling. *J. Phys. A Math. Theor.* **48**, 65304 (2015).
- Russ, M. & Burkard, G. Three-electron spin qubits. Preprint at <https://arXiv.org/abs/1611.09106> (2016).
- Pla, J. J. et al. A single-atom electron spin qubit in silicon. *Nature* **489**, 541–545 (2012).
- Brecht, T. et al. Multilayer microwave integrated quantum circuits for scalable quantum computing. *npj Quantum Inf.* **2**, 16002 (2016).
- DiVincenzo, D. P., Bacon, D., Kempe, J., Burkard, G. & Whaley, K. B. Universal quantum computation with the exchange interaction. *Nature* **408**, 339–342 (2000).
- Koh, T. S., Coppersmith, S. N. & Friesen, M. High-fidelity gates in quantum dot spin qubits. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 19695–19700 (2013).
- Morello, A. et al. Single-shot readout of an electron spin in silicon. *Nature* **467**, 687–691 (2010).
- Mehl, S. Two-qubit pulse gate for the three-electron double quantum dot qubit. *Phys. Rev. B* **91**, 35430 (2015).
- Veldhorst, M. et al. An addressable quantum dot qubit with fault-tolerant control-fidelity. *Nat. Nanotechnol.* **9**, 981–985 (2014).
- Ferraro, E., De Michielis, M., Mazzeo, G., Fanciulli, M. & Prati, E. Effective Hamiltonian for the hybrid double quantum dot qubit. *Quantum Inf. Process.* **13**, 1155–1173 (2014).
- Wong, C. H. High-fidelity ac gate operations of a three-electron double quantum dot qubit. *Phys. Rev. B* **93**, 35409 (2016).
- Veldhorst, M. et al. A two-qubit logic gate in silicon. *Nature* **526**, 410 (2015).
- Shulman, M. D. et al. Demonstration of entanglement of electrostatically coupled singlet-triplet qubits. *Science* **336**, 202–205 (2012).
- Ward, D. R. et al. State-conditional coherent charge qubit oscillations in a Si/SiGe quadruple quantum dot. *npj Quantum Inf.* **2**, 16032 (2016).
- Ferraro, E., De Michielis, M., Fanciulli, M. & Prati, E. Effective Hamiltonian for two interacting double-dot exchange-only qubits and their controlled-NOT operations. *Quantum Inf. Process.* **14**, 47–65 (2015).
- Li, R., Hudson, F. E., Dzurak, A. S. & Hamilton, A. R. Single hole transport in a silicon metal-oxide-semiconductor quantum dot. *Appl. Phys. Lett.* **103**, 163508 (2013).
- Spruijtenburg, P. C. et al. Single-hole tunneling through a two-dimensional hole gas in intrinsic silicon. *Appl. Phys. Lett.* **102**, 192105 (2013).
- Turchetti, M. et al. Tunable single hole regime of a silicon field effect transistor in standard CMOS technology. *Appl. Phys. Express* **9**, 14001 (2016).
- Maurand, R. et al. A CMOS silicon spin qubit. *Nat. Commun.* **7**, 13575 (2016).
- Keane, Z. K. et al. Resistively detected nuclear magnetic resonance in n- and p-type GaAs quantum point contacts. *Nano Lett.* **11**, 3147–3150 (2011).
- Gerardot, B. D. et al. Optical pumping of a single hole spin in a quantum dot. *Nature* **451**, 441–444 (2008).
- Koiller, B., Hu, X. & Das Sarma, S. Exchange in silicon-based quantum computer architecture. *Phys. Rev. Lett.* **88**, 27903 (2001).

60. De Michielis, M., Prati, E., Fanciulli, M., Fiori, G. & Iannaccone, G. Geometrical effects on valley-orbital filling patterns in silicon quantum dots for robust qubit implementation. *Appl. Phys. Express* **5**, 124001-1–124001-3 (2012).
61. Fuechsle, M. et al. A single-atom transistor. *Nat. Nanotechnol.* **7**, 242–246 (2012).
62. Mazzeo, G. et al. Charge dynamics of a single donor coupled to a few-electron quantum dot in silicon. *Appl. Phys. Lett.* **100**, 213107 (2012).
63. Prati, E. & Morello, A. in *Single-Atom Nanoelectronics* (eds. Prati, E. & Shinada, T.) (Panstanford, 2013).
64. Ruess, F. J. et al. Toward atomic-scale device fabrication in silicon using scanning probe microscopy. *Nano Lett.* **4**, 1969–1973 (2004).
65. Ballard, J. B. et al. Pattern transfer of hydrogen depassivation lithography patterns into silicon with atomically traceable placement and size control. *J. Vac. Sci. Technol. B, Nanotechnol. Microelectron. Mater. Process. Meas. Phenom.* **32**, 41804 (2014).
66. Prati, E., Kumagai, K., Hori, M. & Shinada, T. Band transport across a chain of dopant sites in silicon over micron distances and high temperatures. *Sci. Rep.* **6**, 19704 (2016).
67. van Donkelaar, J. et al. Single atom devices by ion implantation. *J. Phys. Condens. Matter* **27**, 154204 (2015).
68. Jamieson, D. N. et al. Deterministic doping. *Mater. Sci. Semicond. Process.* **62**, 23–30 (2017).
69. O’Gorman, J., Nickerson, N. H., Ross, P., Morton, J. J. & Benjamin, S. C. Erratum: a silicon-based surface code quantum computer. *npj Quantum Inf.* **2**, (16014 (2016)).
70. Rotta, D., De Michielis, M., Ferraro, E., Fanciulli, M. & Prati, E. Maximum density of quantum information in a scalable CMOS implementation of the hybrid qubit architecture. *Quantum Inf. Process.* **15**, 2253–2274 (2016).
71. Vandersypen, L. M. K. et al. Interfacing Spin Qubits in Quantum Dots and Donors—Hot, Dense and Coherent. arXiv:1612.05936v1 (2016).
72. Mehl, S., Bluhm, H. & DiVincenzo, D. P. Fault-tolerant quantum computation for singlet-triplet qubits with leakage errors. *Phys. Rev. B* **91**, 85419 (2015).
73. Gonzalez-Zalba, M. F., Barraud, S., Ferguson, A. J. & Betz, A. C. Probing the limits of gate-based charge sensing. *Nat. Commun.* **6**, 6084 (2015).
74. House, M. G. et al. Radio frequency measurements of tunnel couplings and singlet-triplet spin states in Si:P quantum dots. *Nat. Commun.* **6**, 8848 (2015).
75. Kim, D. et al. High-fidelity resonant gating of a silicon-based quantum dot hybrid qubit. *npj Quantum Inf.* **1**, 15004 (2015).
76. Creten, Y., Merken, P., Sansen, W., Mertens, R. P. & Van Hoof, C. An 8-bit flash analog-to-digital converter in standard CMOS technology functional from 4.2 K to 300 K. *IEEE J. Solid-State Circuits* **44**, 2019–2025 (2009).
77. Taylor, J. M. et al. Fault-tolerant architecture for quantum computation using electrically controlled semiconductor spins. *Nat. Phys.* **1**, 177–183 (2005).
78. Copsey, D. & Oskin, M. Toward a scalable, silicon-based quantum computing architecture. *IEEE J. Sel. Top. Quantum Electron.* **9**, 1552–1569 (2003).
79. DiVincenzo, D. P. Quantum computation. *Science* **270**, 255–261 (1995).
80. Bai, P. et al. in *IEDM Technical Digest*. IEEE International Electron Devices Meeting (ed. IEEE) 657–660 (IEEE, 2004).
81. Mistry, K. et al. in *2007 IEEE International Electron Devices Meeting* (ed. IEEE) 247–250 (IEEE, 2007).
82. Jan, C.-H. et al. in *2009 IEEE International Electron Devices Meeting (IEDM)* (ed. IEEE) 1–4 (IEEE, 2009).
83. Jan, C.-H. et al. in *2012 International Electron Devices Meeting* (ed. IEEE) 3.1.1–3.1.4 (IEEE, 2012).
84. Natarajan, S. et al. in *2014 IEEE International Electron Devices Meeting* (ed. IEEE) 3.7.1–3.7.3 (IEEE, 2014).
85. Allan, A. et al. *The International Technology Roadmap for Semiconductors, Executive Summary* (ed. ITRS) <http://www.itrs.net/in> (2013)
86. Zwanenburg, F. A. et al. Silicon quantum electronics. *Rev. Mod. Phys.* **85**, 961–1019 (2013).
87. Yang, C. H. et al. Spin-valley lifetimes in a silicon quantum dot with tunable valley splitting. *Nat. Commun.* **4**, 2069 (2013).
88. Hao, X., Ruskov, R., Xiao, M., Tahan, C. & Jiang, H. Electron spin resonance and spin-valley physics in a silicon double quantum dot. *Nat. Commun.* **5**, 3860 (2014).
89. Goswami, S. et al. Controllable valley splitting in silicon quantum devices. *Nat. Phys.* **3**, 41–45 (2006).
90. Hill, C. D. et al. A surface code quantum computer in silicon. *Sci. Adv.* **1**, e1500707–e1500707 (2015).
91. Lansbergen, G. P. et al. Gate-induced quantum-confinement transition of a single dopant atom in a silicon FinFET. *Nat. Phys.* **4**, 656–661 (2008).
92. Shi, Z. et al. Tunable singlet-triplet splitting in a few-electron Si/SiGe quantum dot. *Appl. Phys. Lett.* **99**, 1–4 (2011).
93. Tyryshkin, A. M. et al. Electron spin coherence exceeding seconds in high-purity silicon. *Nat. Mater.* **11**, 143–147 (2012).
94. Prati, E., Fanciulli, M., Calderoni, A., Ferrari, G. & Sampietro, M. Microwave irradiation effects on random telegraph signal in a MOSFET. *Phys. Lett. A* **370**, 491–493 (2007).
95. Pierre, M. et al. Compact silicon double and triple dots realized with only two gates. *Appl. Phys. Lett.* **95**, 242107 (2009).
96. Nordberg, E. et al. Enhancement-mode double-top-gated metal-oxide-semiconductor nanostructures with tunable lateral geometry. *Phys. Rev. B* **80**, 115331 (2009).
97. Gamble, J. K., Friesen, M., Coppersmith, S. N. & Hu, X. Two-electron dephasing in single Si and GaAs quantum dots. *Phys. Rev. B* **86**, 35302 (2012).
98. Greentree, A. D. & Koiller, B. Dark-state adiabatic passage with spin-one particles. *Phys. Rev. A* **90**, 12319 (2014).
99. Ferraro, E., De Michielis, M., Fanciulli, M. & Prati, E. Coherent tunneling by adiabatic passage of an exchange-only spin qubit in a double quantum dot chain. *Phys. Rev. B* **91**, 75435 (2015).
100. Landahl, A. J., Anderson, J. T. & Rice, P. R. Fault-tolerant quantum computing with color codes. Preprint at <http://arXiv:1108.5738> (2011).
101. Nigg, D. et al. Quantum computations on a topologically encoded qubit. *Science* **345**, 302–305 (2014).
102. Preskill, J. in *Introduction to Quantum Computation and Information* (eds. Di Lo, H.-K., Spiller, T. & Popescu, S.) 213–269 (World Scientific, 1998).
103. Thorgrimsson, B. et al. Mitigating the Effects of Charge Noise and Improving the Coherence of a Quantum Dot Hybrid Qubit. arXiv:1611.04945v2 (2016).
104. Gottesman, D. *Stabilizer Codes and Quantum Error Correction*. PhD Thesis, Caltech (1997).
105. Hornibrook, J. M. et al. Cryogenic control architecture for large-scale quantum computing. *Phys. Rev. Appl.* **3**, 24010 (2015).
106. Conway Lamb, I. D. et al. An FPGA-based instrumentation platform for use at deep cryogenic temperatures. *Rev. Sci. Instrum.* **87**, 14701 (2016).
107. Homulle, H. et al. A reconfigurable cryogenic platform for the classical control of quantum processors. *Rev. Sci. Instrum.* **88**, 45103 (2017).
108. Homulle, H. et al. in *Proceedings of the ACM International Conference on Computing Frontiers—CF ’16* 282–287 (ed. G. Palermo) (ACM Press, 2016).
109. Quaglia, R. et al. Silicon drift detectors and CUBE preamplifiers for high-resolution X-ray spectroscopy. *IEEE Trans. Nucl. Sci.* **62**, 221–227 (2015).
110. England, T. D. et al. A new approach to designing electronic systems for operation in extreme environments: part II—the SiGe remote electronics unit. *IEEE Aerosp. Electron. Syst. Mag.* **27**, 29–41 (2012).
111. Prager, A. A., George, H. C., Orlov, A. O. & Snider, G. L. Experimental demonstration of hybrid CMOS-single electron transistor circuits. *J. Vac. Sci. Technol. B, Nanotechnol. Microelectron. Mater. Process. Meas. Phenom.* **29**, 41004 (2011).
112. Das, K., Lehmann, T. & Dzurak, A. S. Sub-nanoampere one-shot single electron transistor readout electrometry below 10 kelvin. *IEEE Trans. Circuits Syst. I Regul. Pap.* **61**, 2816–2824 (2014).
113. Ekanayake, S. R., Lehmann, T., Dzurak, A. S., Clark, R. G. & Brawley, A. Characterization of SOS-CMOS FETs at low temperatures for the design of integrated circuits for quantum bit control and readout. *IEEE Trans. Electron Devices* **57**, 539–547 (2010).
114. Kleine, U., Bieger, J. & Seifert, H. A low-noise CMOS preamplifier operating at 4.2 K. *IEEE J. Solid-State Circuits* **29**, 921–926 (1994).
115. Kuhn, W. et al. A microtransceiver for UHF proximity links including mars surface-to-orbit applications. *Proc. IEEE* **95**, 2019–2044 (2007).
116. Zhao, H. & Liu, X. A low-power cryogenic analog to digital converter in standard CMOS technology. *Cryogenics* **55–56**, 79–83 (2013).
117. Okcan, B., Gielen, G. & Van Hoof, C. A third-order complementary metal-oxide-semiconductor sigma-delta modulator operating between 4.2 K and 300 K. *Rev. Sci. Instrum.* **83**, 24708 (2012).
118. Murmann, B. *ADC Performance Survey* <http://web.stanford.edu/murmann/adcsurvey.html> (1997–2015).



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.