Deliverable – D3.1 Overview of cryogenic integrated and improved quantum computer electronics

Project acronym	SEQUENCE
Project number	871764
Project title	Cryogenic 3D Nanoelectronics (Sense and Readout Electronics Cryogenically Integrated for QUantum ENhanced Computation and Evolving Communication)

Document Properties	
Nature of Document	Report
Work Package	WP3 - Cryogenic Circuits and 3D Systems
Task Leader	IBM - Cezar ZOTA
Authors	Jean-Baptiste DAVID, Christian ENZ, Arnulf LEUTHER, Fabian THOME, Cezar ZOTA
Version	1.0
Status of Document	Final Version
Due Date of deliverable	M9
Actual delivery date	M9
Dissemination Level	PU

Document history			
Version	Date	Author	Status – Reason for change
1.0	Sept 25 2020	JB DAVID	Final

Release approval			
Version	Date	Name and organisation	Role
1.1	200930	Lars-Erik Wernersson	Project Coordinator

The project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant agreement number 871764 (SEQUENCE)

Contents – Parenthesis indicates section author

1 Introduction (LETI)

Like Joseph C. Bardin from Google and University of Massachusetts exposed in the August 2020 IEEE Microwave Magazine, during the past decade, quantum computing (QC) has grown from the field of promising scientific papers to a point that seriously opens the opportunity to rethink computing as we know it [1.1]. Major companies around the world are now investing on the subject and produce breakout announcements[1.2][1.3][1.4], start-ups are flourishing [1.5], and large national research initiatives on quantum information are voted [1.6][1.7][1.8]. The SEQUENCE consortium and its associated funding demonstrate that European Union is looking at this mutation with great interest. Not only because of the thunderous forecast that QC would hack current security systems within five years[1.8][1.9], but also because of the promising applications in various fields such as chemistry, pharmacology, alternative materials, communication, energy management, and a lot of other human activities where mathematical formulation of the problem to be solved becomes intractable for classical computing.

However, even if the quantum advantage is already claimed by Google[1.4], from the today's prototypes embedding a handful of qubits, to an industrial machine working with a thousand or a million of qubits (like envisaged), the road still needs to be discovered and secured. "In reality, we're likely a decade or more away from a quantum computer that can solve useful problems", reported a journalist for MIT Technology Review this last August 2020 [1.10]. But on the way, pulled by these developments, technological research might unlock secondary markets which might grow in a shorter term.

"In terms of control electronics, wiring, and packaging, today's state-of the- art technology requires that each qubit is driven by dedicated on-chip lines that carry microwave, analog and dc signals. High-speed (1 gigasample per second) digital signals are carried from room temperature pulse generators to individual qubits through dedicated coaxial cables and coplanar wave interconnects. The cables interface to the package via connectors, and the package connects to the chip via wire bonds. Can these technologies be scaled to a 1-million-qubit computer? For the control system only, this would mean at least 3 million DACs and millions of coaxial cables running into a cryostat holding a chip package with millions of connectors. This is where digital-analog-microwave ASIC engineering comes in: the cost, size, thermal, assembly, reliability, and signal integrity concerns arising in a system for which hardware demands innovations." [1.1]

While many qubit architectures and strategies for scaling have been proposed, a completely worked out pathway to create qubit systems that can be expanded to a large-scale quantum processor yet has to be defined and a key step is the design of a scalable classical-quantum interface[1.11]. "With what technology? Executing a quantum application with potential for societal impact will require a formidable number of high-fidelity qubits, operating in concert with a control interface that passes signals between the classical- and quantum-domains of a quantum computer. At present, it is not explicitly known if, or how, this complex control interface can be realized at scale", said D. J. Reilly et al from Microsoft Sydney in 2019[1.12]

Let's tell it right now: we don't have pretention to pave the way of the quantum computing industrialization on the next 15 years. The aim of the present document is to draw a picture of the current research and future challenges that we need to overcome **in the field of cryoelectronics**, in order to develop functionalities that help scaling up from a few qubits to thousands of qubits. **We intend to establish a general vision for the physical layer of a superconducting-transmon or silicon-spin qubits Quantum Computer, especially by drawing the expected 3D integrated Si/III-V based architecture of such a computer, and therefore, its expected set of building blocks**.

More particularly, this document will *mainly cover*:

- Quantum computer physical implementation
- Silicon-based technologies for quantum computing

- Silicon-based circuits for quantum computing
- Superconducting materials for improved back-end and heterogeneous integration
- III/V will be mentioned as a key enabler technology for identified critical circuits in the physical implementation
- Superconducting logic will be mentioned as a reporting to IRDS, since IRDS consortium has widely covered this field

This document will *partially cover*:

- Cryogenic testing challenges & cooling management (i.e. cryostats)
- Quantum bits technologies

This document will *not cover*:

- Physical Qubits TO Logical Qubits layer involving hardware/software Error correction techniques
- Quantum-based algorithms and their potential consequences on specific physical architecture developments
- Quantum-based applications

References

[1.1] Joseph C. Bardin *et al*, "Quantum Computing: An Introduction For Microwave Engineers", IEEE Microwave Magazine, 08/2020
[1.2] M. Berboucha, "IBM's first commercial quantum computer," Forbes, Jan. 23, 2019. https://www.forbes.com/sites/meriameberboucha/2019/01/23/ibms-first-commercial-quantum-computer/#cf020a441b2a
[1.3] F. Lardinois, "Microsoft open-sources its quantum computing development tools," TechCrunch, San Francisco, May 6, 2019. https://techcrunch.com/2019/05/06/microsoft-open-sources-its-quantum-computing-development-tools/
[1.4] J. Hsu, "Google's quantum tech milestone excites scientists and spurs rivals," IEEE Spectrum, 2019. https://spectrum.ieee.org/tech-talk/computing/hardware/googles-quantum-tech-milestone-excites-scientists-and-spurs-rivals
[1.5] "5 Top Emerging Quantum Computing Startups", 07/2020, https://www.startus-insights.com/innovators-guide/5-top-emerging-quantum-computing-startups/
[1.6] French national report, released in January 2020: « Quantique, le virage technologique que la France ne ratera pas », https://forteza.fr/wp-content/uploads/2020/01/A5_Rapport-quantique-public-BD.pdf
[1.7] In June 2020, German chancellor Angela Merkel announced a 2 Billion innovation program for quantum technology including the investment in two quantum computers. Original announcement here: https://www.bundesfinanzministerium.de/Content/DE/Standardartikel/Themen/Schlaglichter/Konjunkturpaket/2020-06-03-eckpunktepapier.pdf
[1.8] https://ec.europa.eu/digital-single-market/en/policies/quantum-technologies-flagship
[1.9] "Google CEO says quantum computing could end encryption within five years", 04/2020: https://www.idquantique.com/quantum-computing-industry-review-q1-2020/
[1.10] MIT Technology Review, 08/2020: https://www.technologyreview.com/2020/08/03/1005891/search-for-quantum-proof-encryption-computing-nist/
[1.11] L. M. K. Vandersypen et al., "Interfacing spin qubits in quantum dots and donors—hot, dense, and coherent," npj Quantum Information, vol. 3, no. 1, Art. no. 1, Dec. 2017
[1.12] S. J. Pauka et al., "A Cryogenic Interface for Controlling Many Qubits," arXiv:1912.01299 [cond-mat, physics:physics, physics:quant-ph], Dec. 2019, Accessed: Feb. 12, 2020. [Online]. Available: http://arxiv.org/abs/1912.01299

2 Frame of analysis (LETI)

Should we discuss in terms of a final vision of a quantum computing architecture, or should we explore the possibilities for making as efficient cryogenic devices and circuits as possible, based on a generalized architecture? This question was convoked as an underlying reflection, since the demand for a roadmap on quantum computing appears to be so early regarding the state of the art.

- The short term vision is strongly contained in today's technology limitations (low temperature electronics knowledge, available building blocks, room temperature signal generation systems, cooling facilities), and based on a functional single or two-qubits line.
- The long term vision is towed by a general multiplexed and digital approach, according to up-to-date signal processing architectures for Radio Frequency communications. This vision intends to support the scaling of physical qubits number.
- The road in between is roughly imagined by virtue of parallelization, integration, and power efficiency criteria.

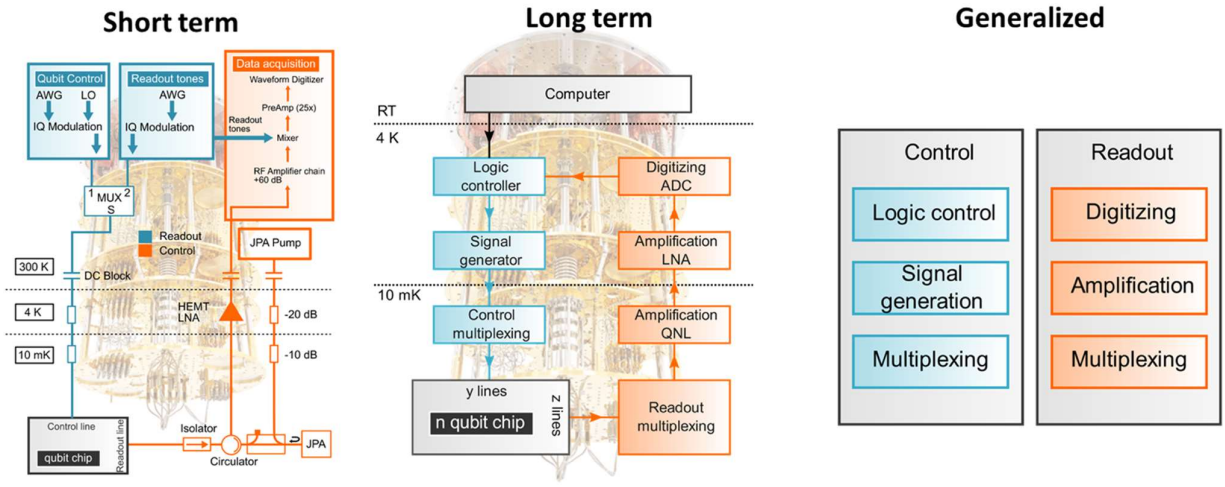


Figure 2-1: Short term (single Josephson Junction [transmon] qubit-based), long term (temperature partitioning may change), and generalized vision of a quantum architecture.

Compared to "classical" electronic systems, some completely new aspects tend to design the inner boundaries of such a QC architecture: physical dimensions of coolers and cooling efficiency, power budget versus temperature, qubit properties (coherence duration, control/readout frequency, fidelity) and qubits integration strategies, need of new materials like superconductors, and physical behavior of electronics at low temperature.

On one hand, we can't ignore the relevance of the short term vision, and we can't deny the value conferred in the long term vision. On the other hand, a precocious roadmap effort should not act as blinders on the presumed nature of our knowledge.

The limited fidelity of physical qubits directly affects the ability of a Quantum Computer to operate. Besides fidelity enhancement research, QC architecture assumes a high number of these physical qubits, in order to generate the equivalent of a few perfect logical qubits. We talk about a factor from thousands to millions, depending on qubits technology and QC implementation. We bet that future qubit's fidelity will increase with technology maturation, mitigating the required number of physical qubits to make a logical qubit.

Qubit's temperature of operation is also a field of research (which is related to the fidelity). By the way, cryostats that are more efficient might be developed, and would lead to reconsider temperature partitioning of the system, which is a main parameter in the architectural implementation.

Today's prototypes are made of an accumulation of single lines (short term in figure 2-1). It seems this approach is viable until about hundred physical qubits. But hard challenges might appear during the multiplexing approach, depending on the ability to maintain fidelity, and to manage power budget during the control/readout process.

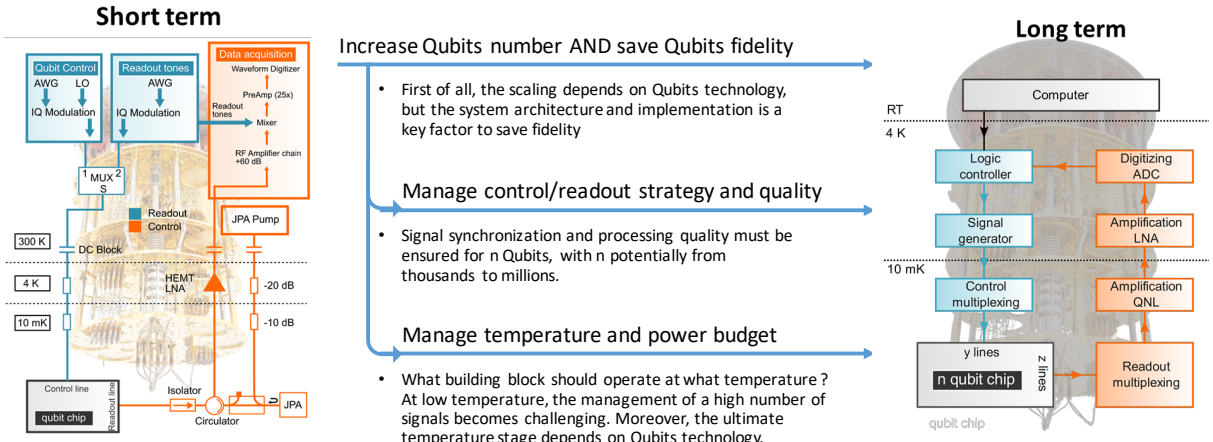


Figure 2-2: The number of qubits and their fidelity will drive the QC architecture

The control/readout separation is justified by the specific nature of qubits, which involves two different types of signal generation and processing. This partitioning might call back the transmit/receive partitioning of a communication system. However, we will see in this document how they are strongly different.

In a complementary approach, the traditional splitting for the development of a complex electronic system (Technology improvement – Component and circuit design – System implementation), which also corresponds to usual skills and market partitioning, is supposed to remain operant, as shown in the figure below.

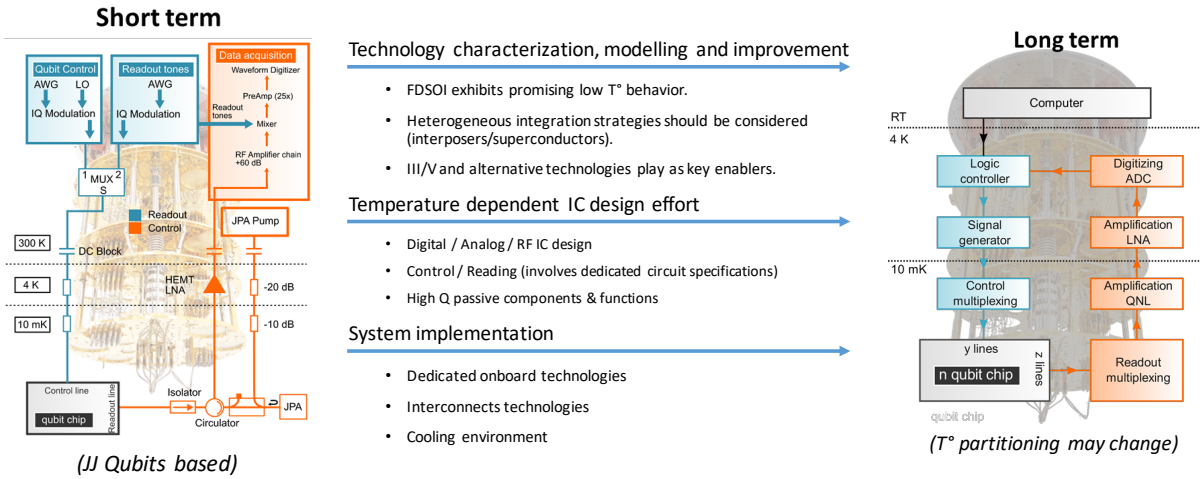


Figure 2-3: The number of qubits and their fidelity will drive the QC architecture

Nevertheless, by considering the two above figures, we see new contours of skills and markets emerging, mainly driven by thermal environment issues:

- Technology players will need to extend their design kit to a few Kelvins, where new physical behaviors require advanced modeling,
- Dedicated technologies will emerge either as add-on in standard process, or in 3D heterogeneous integration process, with thermal dissipation control at low temperature, and extensive utilization of superconducting materials,
- Digital, Analog, and RF designers will include the Ultra-Low Power budget and local heating effects as a main design parameter, assuming multi-physical simulations, and will develop innovative circuits under stringent specifications,
- The thermal budget link will drive the architecture partitioning through the cryostat stages,
- Effort for on-probe and in-board test facilities at low temperature will increase,
- New requirements for cryostats will appear, in term of form factor, temperature staging and efficiency.

In addition, that's not the least of the challenges: qubits technologies and models will be transferred from physicist's hands to industrial actors, requiring potential efforts in the formalism of quantum-electrical link within CAD environment.

3 About metrics (LETI)

In the expectation of a roadmap building, one should wait for something like a metric, that one could track over time. The previous chapter suggests that a natural metric for Quantum Computing implementation should be the number of qubits. Nevertheless, we suggest that this number might be too coarse to be effective. Indeed, control/readout management, and system implementation contribute to the more general dimension of fidelity, and therefore, the expected quantum computer performance.

3.1 Quantum Volume (LETI)

Beyond the technological implementation, IBM proposed the notion of "Quantum Volume" to evaluate the performance of a quantum computer [3.1][3.2][3.3][3.4]. This metric intends to “provide an overall sense of the quantum capabilities of a device for high-level comparisons” [3.1].

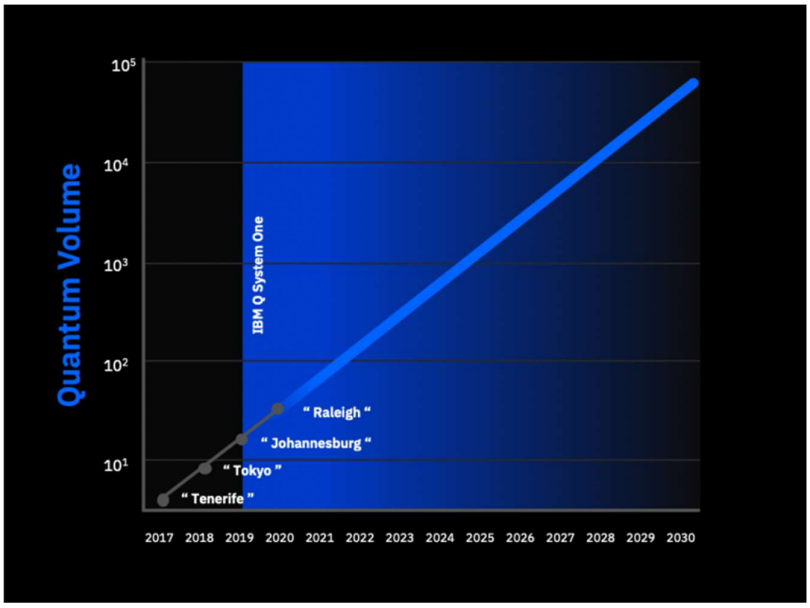


Figure 3-1: Quantum volume roadmap of IBM Q systems

The metric aggregates:

1. The number of physical qubits;
2. The number of gates that can be applied before errors make the device behave essentially classically;
3. The qubits connectivity graph of the device;
4. The number of operations that can be run in parallel.

Based on this metric, the authors declared in early 2020 that "**in order to achieve quantum advantage within the next decade, we will need to at least double the Quantum Volume of our quantum computing systems every year**."[3.3]

In place of Moore's law, will Quantum Volume operate for a QC roadmap? Recently, Honeywell used this metric to promote their ion-trap solution [3.5].

We can dispute this metric [3.6]. Anyway, even if the physical implementation is voluntarily hidden for abstraction reasons, it has the merit to aggregate various properties related to the QC physical layer:

- the second point is linked to the number of control/readout signals that can be applied before losing qubits coherence (this number increases with quantum algorithm complexity), so it's linked to the electronic speed and signal coupling mastery.
- The third point talks about the entanglement possibilities inside the qubits network. One tells that any quantum algorithm could be declined into a set of single and two-qubits gates [1.1], but the quantum algorithm complexity is linked to the level of entanglement inside the network. This is an active field of research at the qubit level, and might require extrinsic cryoelectronic developments [3.7].
- And the fourth point, involving parallel operations on qubits, has an impact on signal synchronization complexity, hence again, linked to the electronic accuracy.

Thus, the quantum computer performance is not only dependent on the physical qubits number.

Building a roadmap for the physical implementation of a quantum computer regarding these four dimensions, that might be a good way to guide our choices. Unfortunately, the conversion of cryoelectronic performance criteria into such overall system performance criteria cannot be done with obviousness.

This pitfall had already been formalized in the NEREID project, where a tentative roadmap for wireless communication with Figures of Merit (FoM) at the system level had been proposed in order to complete technology driven ITRS/IRDS initiative [3.8]. Considering a roadmap for integrated and improved quantum computer electronics, more classical FoM should be used. These FoM are correlated to the overall system performance, however details of this correlation are concealed. And they especially impact on the qubit fidelity.

Hence, **without being the only one, saving the qubit fidelity at the same time as increasing the qubits number** appears as a key point for quantum computer integration. And all technological properties susceptible to fidelity contribution should deserve our interest.

We will see along this document that some technological properties are carefully observed because they boost the system implementation possibilities. For example, the natural threshold voltage tuning of FDSOI, or the minimum noise figure of GaAs. At the design level, we will see that power consumption per qubit, signal sensitivity or phase noise are key parameters. And beyond this "hard" information, "soft" information must draw attention, like modelling effort evolution (including the implementation into a design kit), or test facilities improvements.

3.2 Rent's rule (LETI)

In the context of a brainstorming for a metric of quantum computing roadmap, it's interesting to report the paper from QuTech-Delft in Microprocessors and Microsystems 67, 2019 [3.9].

In order to implement a fault-tolerant quantum computer, the qubit's number is expected to grow from a few to a range of millions depending on required practical applications. Franke *et al* then recovered memory about the 'tyranny of numbers' that Bell labs encountered in the late 1950s, when engineers were working with electrical systems containing many components, each requiring soldering to numerous others. The issue was overcome by the Integrated Circuit industrialization.

In 1960s, E.F. Rent from IBM had observed the relation that exists between the number of internal components to connect in order to build a system, and the number of connections for each internal component. That was finally summarized in a formula:

$$T = tg^{p}$$

Where 't' refers to the connections required for each component 'g'. The *Rent* exponent 'p' accounts for the level of optimization, such that with no optimization, p=1.

In the referenced paper, QuTech defines the *quantum Rent* exponent 'p' to quantify the progress in overcoming the challenge of qubits interconnections at different levels throughout the quantum computing stack.

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

The authors go on the expectation to reduce the Rent exponent by playing on strategies to reduce the number of physical or virtual interconnects at each level of the quantum stack.

"With p~0.5, one million qubits require not of order one million wires (infeasible) but one thousand wires (feasible). It will be a milestone if such architectures can be realized experimentally", said Franke *et al*.

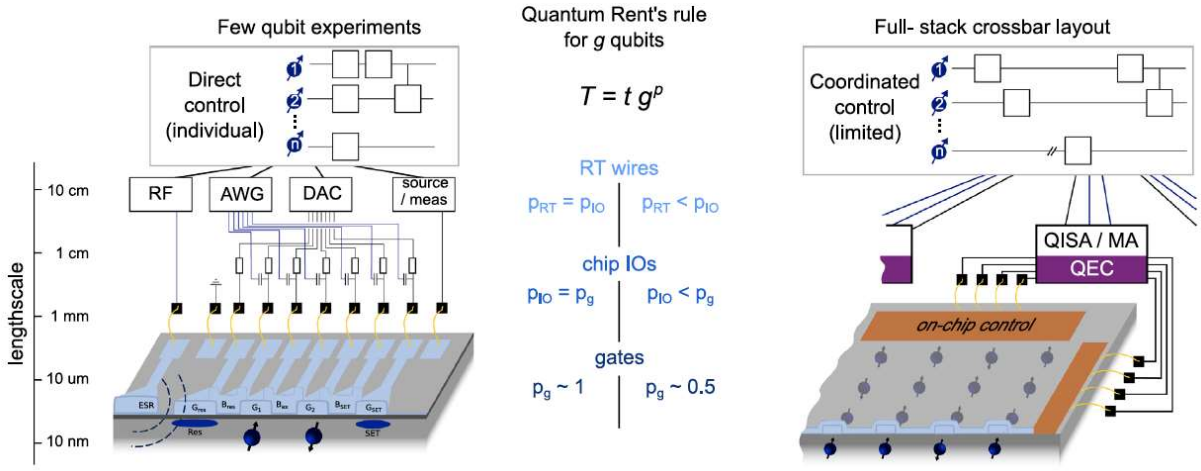


Figure 3-2: Comparison of the quantum Rent component p defined for different layers of a typical few qubit experiment and a possible optimized integration scheme for a spin qubit processor. The exponent p can be improved by solutions at different layers, such as a crossbar gating layout (reducing p_g), on-chip routing and multiplexing (p_{io}) or cryogenic logic for Quantum Error Correction cycles (p_{RT}). As an effect of shared control and the reduced bandwidth per qubit, limitations in the qubit control and the timing of gate sequences will occur as an effect of such optimizations.

State-of-the-art processors contain more than 10^9 transistors. Furthermore, the structure of these transistors is comparable to semiconductor-based qubits. However, an important difference is that conventional processor chips have only ≈10^3 input-output connections (IO's). This brings the transistor-to-IO ratio over 10^6.

In the absence of multiplexing or on-chip control logic, the limit for the qubit count is probably similar to the pin-limit of the package, which is currently around 10^3.[3.10]

Even if, like with Quantum Volume, it's not easy to establish a direct correlation between cryo-electronic developments and Rent exponent, it's quite intuitive that any enhancement leading to the reduction of interconnects number can be used as a metric to inform on achievements.

References

[3.1] Jay Gambetta *et al*, "Quantum Volume", in Semantic Scholar, March 4, 2017
[3.2] Jay Gambetta *et al*, "Validating quantum computers using randomized model circuits", in ArXiv, March 17, 2019
[3.3] Jay Gambetta *et al*, "Quantum Takes Flight: Moving from Laboratory Demonstrations to Building Systems", January 8, 2020 : https://www.ibm.com/blogs/research/2020/01/quantum-volume-32/
[3.4] https://newsroom.ibm.com/2020-08-20-IBM-Delivers-Its-Highest-Quantum-Volume-to-Date-Expanding-the-Computational-Power-of-its-IBM-Cloud-Accessible-Quantum-Computers
[3.5] https://www.honeywell.com/en-us/newsroom/news/2020/06/the-worlds-highest-performing-quantum-computer-is-here
[3.6] https://www.scottaaronson.com/blog/?p=4649
[3.7] Baptiste Jadot, "Coherent long-range transport of entangled electron spins", 02/2020: http://www.theses.fr/s199400
[3.8] https://www.nereid-h2020.eu/
[3.9] Franke *et al*, « Rent's Rule and Extensibility in Quantum Computing ». *Microprocessors and Microsystems* 2019
[3.10] L. M. K. Vandersypen et al., "Interfacing spin qubits in quantum dots and donors—hot, dense, and coherent," npj Quantum Information, vol. 3, no. 1, Art. no. 1, Dec. 2017

4 Short introduction on qubits technologies

4.1 Quantum bit (LETI)

In a quantum computer, physical building blocks that are used to register unit information, exhibit a quantum property. These building blocks are called quantum bits, or qubits. In a construction that is similar to classical computers, the qubit refers to a zero state, and a one state. However, what distinguishes a qubit from a classical bit is that it can be in a superposition of its zero and one states. By the way, it presumes that a quantum bit is a physical building block where **only two quantum states** are isolated. According to quantum mechanics, the mathematical representation of superposition of these two states is a 2D complex state vector of unit amplitude. A useful graphical representation of this complex state is the so-called Bloch sphere.

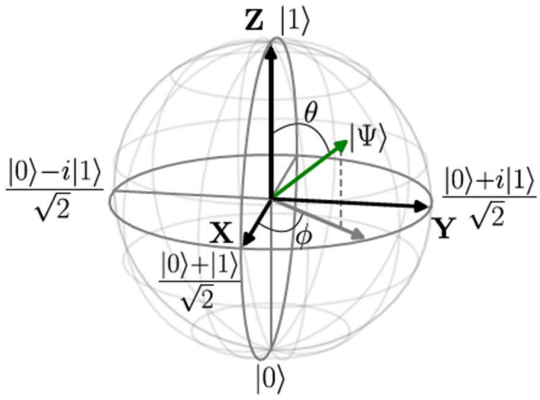


Figure 4-1: Bloch sphere representation of a two-level quantum state

What we call “qubit control” reflects to a physical way to apply a rotation to this complex vector around XYZ directions. It means the quantum register interacts with a precise quantity of energy (precise in frequency, phase, amplitude, and duration), and it modifies its quantum state in a predictable way.

What we call “qubit readout” reflects to a physical way to return the module of this complex vector, and then, either a zero state, or a one state, in a range of probability. This operation requires taking a small amount of energy from the quantum register, and it tends to produce quantum state decoherence.

It is not possible to isolate completely quantum systems from their environment. They interact in a random way, so that the quantum information is lost: this is called decoherence. It is convenient to separate decoherence in two processes: relaxation and dephasing. Relaxation corresponds to the dissipative process where the one state turns into the zero state, and is associated with a characteristic time T1. On the Bloch sphere, it corresponds to a path of the state along a meridian, i.e. a change of θ. On the contrary, dephasing corresponds to a change in the phase ϕ, a path along a parallel of the Bloch sphere. It is associated with a characteristic time T2. The decoherence of a qubit needs to be carefully evaluated, taking into account all the possible sources of noise. To make an efficient quantum bit, the decoherence times T1 and T2 need to be increased as much as possible while keeping the ability to perform fast rotations. Therefore the relevant figure of merit for single-qubits operations is the number of operations that can be performed before decoherence, that is $Q = f_{Rotations} \times T2$ (assuming $T2 << T1$)

In this section, we propose a short introduction on two physical implementations of a quantum bit: the superconducting transmon, and the Silicon spin-qubit.

4.2 Silicon Spin-qubits (LETI)

As we have seen, a quantum computer is composed of quantum bits, which are two-levels system, thus, any quantum two-levels systems can in principle be used for quantum information storage and manipulation. The spin qubits in semiconductor quantum dots were first studied in III-V heterostructures, notably in GaAs. Thanks to the bandgap mismatch between two III-V materials, which is responsible for a quantum well at the hetero-interface, these materials are good candidates for confining electrons.

In Silicon, thanks to the crystallographic organization, the conduction band can degenerate in so-called valleys, depending on Silicon isotope purity, dopants nature, and abrupt interfaces of specific nanostructures. Silicon is found to be a good host for electron spin qubits, thanks to large valleys splitting, and the low spin-orbit interaction, leading to a long coherence time.

Among various physical implementations of a quantum dot with Silicon, here, we expose a basic version of the technology that is under development in CEA-Leti. For more information on implementation strategies, please refer to [4.2].

The figure below shows the cross-section of one quantum dot, and the topology of a double quantum dot. This qubit architecture relies on the use of nanowires on Silicon-On-Insulator substrate, leading to a 2D structural confinement that makes single charge control possible.

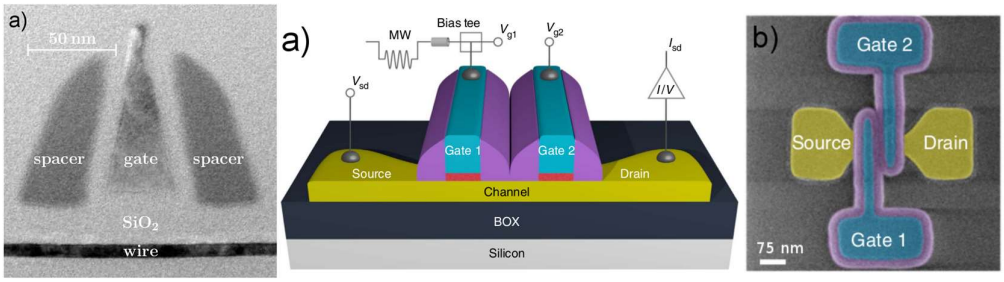


Figure 4-2: Left: MOSFET transistor with long spacers isolating a quantum dot under the gate, capacitively coupled to the source and drain contacts. Middle: 3D schematic of a double quantum dot. Right : top view of the device.

A key ingredient for the control of single charges is the design of the spacers between the source/drain contacts and the gate. These spacers control the tunnel barriers between the channel and source/drain electron reservoirs. If they are long enough, they lead at low temperature to the formation a small area under the gate in the regime of Coulomb blockade, meaning that electrons inside the area create a repulsion force preventing other electrons to flow. This makes a quantum dot. Acting on the gate potential then controls the number of electrons in the quantum dot.

A second step is to introduce another gate in this system, in order to create double quantum dot. Then, one quantum dot is formed under each gate, and they are tunnel coupled. In this configuration, the quantum mechanics property of Pauli Spin blockade occurs. Because of the tunnel coupling, and in virtue of the Pauli Exclusion Principle, some exclusive conditions must be respected for an electron to flow through the tunnel barrier. Hence, the spin of the electron in each quantum dot is interdependent.

A third step is to apply a static magnetic field in the region of the quantum dots, to obtain a split of the spin-up and spin-down energy level, leading to an oscillation regime of the spin. This energy splitting is called the Zeeman Effect. The oscillation regime can be projected into the Bloch sphere through a rotation around the Z axis, with a specific pulsation that is related to the static magnetic field. This pulsation is called the Larmor frequency.

Finally, an oscillation on the X or Y axis is made possible, either with an additional dynamic magnetic field (the full XYZ process is then called Electron Spin Resonance – ESR), or with a dynamic electric field (the full XYZ process is then called Electric Dipole Spin Resonance – EDSR).

In the case where the oscillation frequency applied around X or Y axis is the same as the Z axis oscillation frequency that is due to the static magnetic field (so, the Larmor frequency), a specific regime for the spin state appears, which can be observed. This experiment is known as the measurement of the Rabi frequency.

Now, let's have a look at a spin manipulation process, with the example of Rabi oscillations measurement. In figure below: starting from the spin blockade regime, the qubit from dot 2 is pulsed into the Coulomb blockade regime, thanks to a specific bias on Gate 2. Then a burst of resonant microwaves of duration τ is applied on Gate 2 and rotates the spin. Finally, the system is pulsed back to the spin blockade for spin projection.

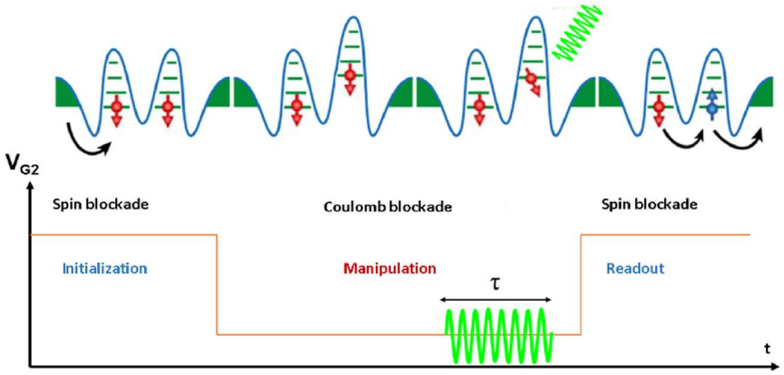


Figure 4-3: Protocol for manipulation and detection of the spin in a double quantum dot.

This sequence is repeated continuously, resulting in a measurable current I_D that oscillates as a function of τ, signature of the Rabi oscillations like shown in figure below. The number of oscillations observed during such a process is a direct technique for measurement of the coherence time.

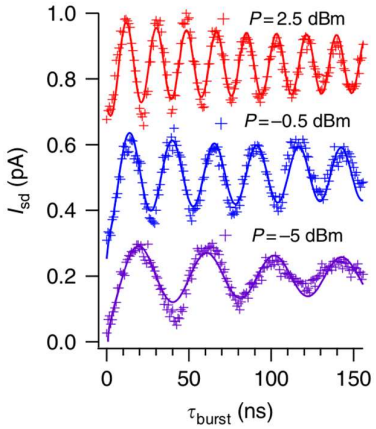


Figure 4-4: Rabi oscillations for different microwave powers.

In practice, current-based detection of Pauli spin blockade is not suited for quantum computation, which requires single-shot readout techniques. Therefore, a method for single-shot detection of charge transfer, called gate-reflectometry, is currently implemented.

4.3 Superconducting Qubits (IBM)

Superconducting qubits such as Transmons, are based on the energy levels of quantum anharmonic oscillators. These are similar to quantum LC circuits, but replace the inductive element with a Josephson junction – a MIMCAP comprising a thin oxide sandwhiched by two superconducting leads. Al/Al_2O_3 is a common material choice for these junctions. Compared to the quantum LC oscillator, the Josephson circuit has a sinusoidal energy landscape, making the energy states anharmonic. This allows precise control of the excitation states using the unique resonance frequency of each state.

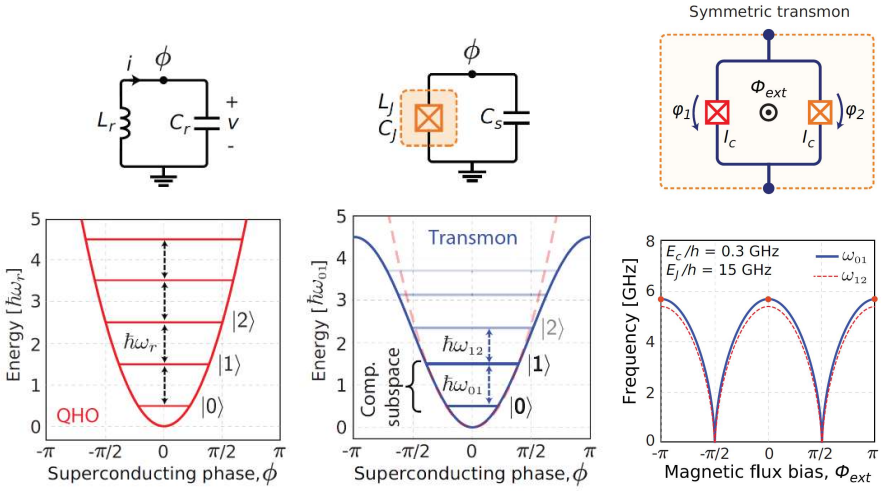


Figure 4-5: Schematic figures showing the characteristics of transmon qubits [4.3].

Typically, the bottom two energy states comprise the <0l and <1l states of the qubit. The resonance frequencies of energy levels can be tuned by applying a magnetic field. By tuning the magnetic field, the anharmonicity, i.e. the difference in resonance frequencies, can be made larger. This is beneficial to avoid accidental excitation to unwanted states. More advanced qubit architectures using a multitude of Josephson junction exists, that can further increase the anharmonicity. Operation of transmon quantum computers requires the generation of excitation pulses targeting the resonance frequencies of the qubit. These frequencies are typically around 4-8 GHz. Moreover, to implement arbitrary Bloch sphere rotations, precis phase control is required of the control signal. To create such control pulses, AWGs together with IQ modulation techniques are typically used. Readout is handled by detecting the phase or frequency shift of a readout resonator in response to an applied readout pulse. Transmon qubits have been widely adopted by major companies seeking to rapidly develop and commercialize quantum computers. While the decoherence times associated with transmons are relatively low, in the order of 10-100 µs, there are several reasons for this choice of qubit. Of primary importance is the availability and maturity of the hardware technology. Two-qubit gates are available for transmons. In addition, gate operation speed is relatively fast for transmons.

References

[4.1] Joseph C. Bardin *et al*, "Quantum Computing: An Introduction For Microwave Engineers", IEEE Microwave Magazine, 08/2020

[4.2] Leo Bourdet, "Modeling of electrical manipulation in silicon spin qubits", PhD report, 2019 https://tel.archives-ouvertes.fr/tel-02057677

[4.3] Krantz, Philip, et al. "A quantum engineer's guide to superconducting qubits." *Applied Physics Reviews* 6.2 (2019): 021318.

5 Evolution of Silicon technologies for cryogenic applications (LETI)

Even if we find a few papers on CMOS behavior at 77K or 4.2K since middle of 80's [5.1][5.2][5.3][5.4][5.5], research at low temperature is growing recently, due to the rise of quantum chip and quantum computer research. In a general picture, silicon-based technologies exhibit better performances at low temperature, with a sharper on/off state, and a better transconductance[5.5][5.6]. However, these benefits are mitigated by an increased variability, and a threshold voltage shift-up.

Projected technology benefits and trade-off related to the operation of CMOS down to cryogenic T° are identified and listed below [5.5]:

+ Higher I_{ON} values due to mobility and saturation velocity increase
+ Steeper subthreshold slope
+ High reduction of leakage current (I_{off})
+ Reduced thermal noise,
+ Better parasitic resistance/capacitance trade-off since BEOL metal resistance is reduced
+/- Reliability improvement due to bias temperature instability (BTI) reduction, but hot carrier degradation should become more severe [5.28]
- Variability increase
- $|V_T|$ increase for both PMOS and NMOS
- Increased thermal conductivity of Si, as well as of many materials

Subthreshold swing reduction and threshold shift with decreasing temperature are illustrated in the figure below (from Beckers *et al* JEDS 2018 [5.14]), for a 28nm bulk NMOS in linear regime.

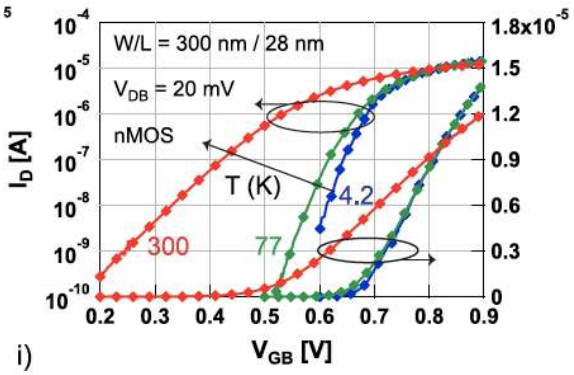


Figure 5-1: 28nm BULK CMOS transistor in linear regime along temperature (log scale on the left side, linear scale on the right side).

Until now, publications mainly focus on transistor DC characterization and modeling, although some RF-mmW analysis appear sporadically. In this section, we report works that brush a state of research on low temperature CMOS technology. We will conclude on the FD-SOI advantage, and future progress in characterization and modeling.

5.1 Bulk Silicon DC characterization (LETI)

In JSSC-2019, Google published a paper including characterization of 28nm BULK CMOS transistors with various flavors at 7K [5.12]. The next table illustrates some parameter changes at this low temperature which are consistent with previously observed values for a variety of recent advanced CMOS bulk technologies[5.13][5.14][5.15][5.16]. We notice the **enhancement of the peak of *gm*** when temperature

decreases, correlated with mobility improvement, and leading to a corresponding increase of the calculated f_t. The table also evidences the **increase of the threshold voltage $|V_T|$** (inherent to the Fermi level shift with temperature), then reducing the potential signal voltage dynamic, and pushing analog polarization point to higher DC power consumption.

On average, the threshold voltages increased by a factor of 40% (≈0.6 to 1mV/K depending on the technology and on the gate length, *i.e.* ≈180mV to 300mV at 4.2K), whereas the sub-threshold swing, extrinsic transconductance, and unity current gain cutoff frequency improved by factors of 4.6, 2, and 1.8, respectively. **The strong subthreshold swing variation comes along with a strong reduction of the off-current**, like it will be presented on FDSOI in the next subsection. This SS reduction with T-decrease however tends to saturates at roughly 20K-40K whatever the technology[5.19][5.20][5.21].

TABLE III
EXTRACTED TRANSISTOR PARAMETERS

	Temp.	uLVTn	LVTn	RVTn	HVTn	uLVTp	LVTp	RVTp	HVTp
V_T	300 K	0.18 V	0.22 V	0.26 V	0.34 V	0.21 V	0.24 V	0.29 V	0.37 V
	7 K	0.26 V	0.30 V	0.35 V	0.45 V	0.34 V	0.37 V	0.40 V	0.45 V
Sub-threshold	300 K	100 mV/dec	92 mV/dec	84 mV/dec	80 mV/dec	103 mV/dec	96 mV/dec	87 mV/dec	83 mV/dec
Swing	7 K	16 mV/dec	15 mV/dec	15 mV/dec	13 mV/dec	33 mV/dec	31 mV/dec	30 mV/dec	22 mV/dec
$G_{m,pk}$	300 K	1.5 S/mm	1.4 S/mm	1.3 S/mm	1.2 S/mm	1.3 S/mm	1.3 S/mm	1.2 S/mm	1.1 S/mm
	7 K	2.0 S/mm	2.0 S/mm	1.8 S/mm	1.7 S/mm	1.8 S/mm	1.7 S/mm	1.7 S/mm	1.5 S/mm
G_m	300 K	14 mS/mm	15 mS/mm	16 mS/mm	16 mS/mm	13 mS/mm	14 mS/mm	15 mS/mm	17 mS/mm
$J_D = 10$ mA/mm	7 K	33 mS/mm	32 mS/mm	32 mS/mm	31 mS/mm	23 mS/mm	27 mS/mm	30 mS/mm	28 mS/mm
$f_{t,pk}$	300 K	285 GHz	283 GHz	284 GHz	237 GHz	244 GHz	234 GHz	230 GHz	215 GHz
	7 K	360 GHz	365 GHz	347 GHz	313 GHz	349 GHz	335 GHz	327 GHz	291 GHz
f_t	300 K	52 GHz	53 GHz	46 GHz	53 GHz	49 GHz	49 GHz	49 GHz	49 GHz
$J_D = 10$ mA/mm	7 K	92 GHz	93 GHz	89 GHz	86 GHz	93 GHz	94 GHz	89 GHz	85 GHz

Figure 5-2: Extracted 28nm BULK CMOS transistors parameters at 7K.

Anomalous kink/hysteresis behavior have been reported in the past in the I-V characteristics of bulk CMOS transistors, with micron length scales and high operating voltages, operating at temperatures below carrier freeze-out [5.27] (see next figure). This phenomenon is also measured in contemporary commercial bulk CMOS technologies (*e.g.* 180nm CMOS, and 28nm CMOS)[5.32]. The appearance of the kink in 28nm CMOS technology at deep-cryogenic temperatures has been shown to occur at an operating voltage above 0.9V. It seems that such behaviour is not observed for this 28nm bulk CMOS technology below the standard supply voltage of 0.9V, and the output conductance, G_{ds}, remains practically constant with respect to temperature [5.14]. **Since the performance is unstable around the kink, the cryo-CMOS designers should be aware of this phenomenon**.

In FDSOI however with sufficiently thin Si channel (typically <10 nm) and/or very short transistors (L < 50 nm), the kink is totally absent as a consequence of super-coupling effect[5.28][5.33].

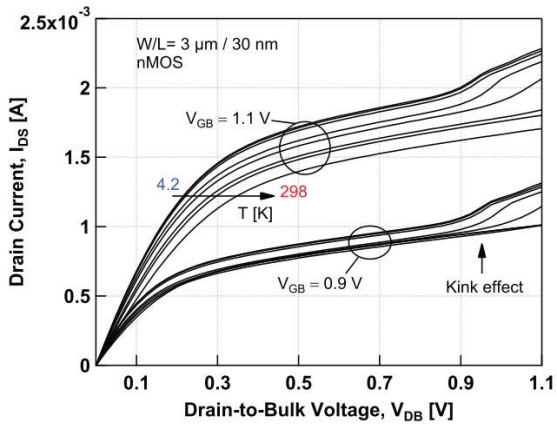


Figure 5-3: The kink effect sets in below T =110 K and above V_{DB} =0.9V in the output characteristics of a commercial 28nm bulk CMOS technology [5.32]

In VLSI-2020, TSMC published an evaluation of benefits of FinFET technology operating at low temperature down to 77K [5.29], in terms of power / performance balance. Hence, **significant enhancements on energy efficiency are expected using low temperature silicon technology**. By analyzing the trade-off between, in one hand, the additional consumed power of cooling, and in other hand, the gain on system power or performance, a range of optimal temperature of operation might be found for a given system. Nevertheless, to benefit from these advantages with FinFET at cryogenic temperatures, challenging engineering of the transistor gate stack is required in order to lower the V_T.

Mismatch was evaluated at low temperature down to 4.2K on various bulk technologies using a set of addressable transistors[5.7][5.8][5.9][5.10]. In JEDS 2020 [5.9][5.10], P. A. T Hart *et al* published an extended review of 40nm bulk CMOS transistors mismatch from 300K to 4K. The current factor, threshold voltage and subthreshold mismatch of CMOS devices at cryogenic temperatures was characterized. Mismatch increases at cryogenic temperatures (especially below 100K). In [5.10] the authors reported a large +75% temperature variations for A_β +22% for A_{VT} (+22%) for this 40nm bulk technology. **More generally, when temperature is decreased from RT down to 4.2K, mismatch increases in all operating regimes from moderate to strong inversion.**

T [K]	A_{VT} [mV · µm]						A_β [% µm]					
	NMOS			PMOS			NMOS			PMOS		
	40n	120n	400n	40n	120n	400n	40n	120n	400n	40n	120n	400n
300	3.2 ±0.2	4.5 ±0.2	5.7 ±0.3	3.2 ±0.1	3.4 ±0.2	3.9 ±0.3	5.6 ±0.2	7.4 ±0.4	9.5 ±0.5	5.3 ±0.3	6.3 ±0.3	6.0 ±0.4
4.2	3.4 ±0.2	5.0 ±0.3	7.0 ±0.4	3.0 ±0.1	4.0 ±0.2	5.9 ±0.4	7.4 ±0.3	11.6 ±0.7	18 ±1	7.6 ±0.4	11.3 ±0.6	15 ±1

Figure 5-4: Variability of $_{VT}$ and β measured on 40nm bulk CMOS technology as a function of temperature and device length.

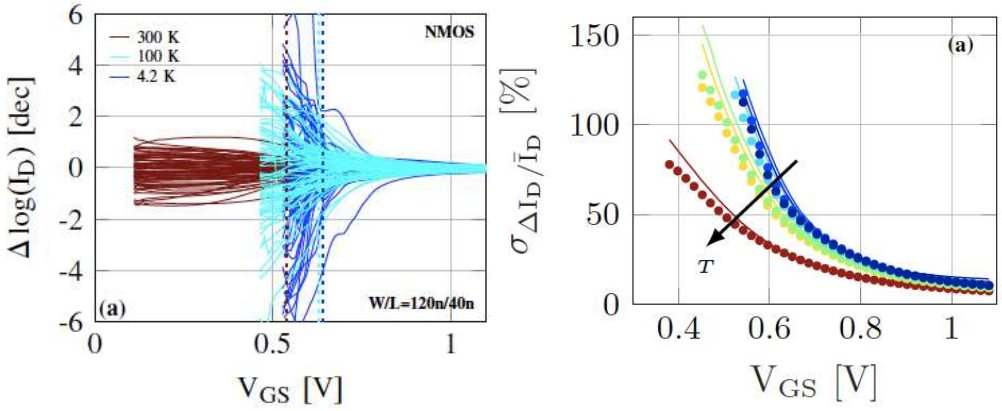


Figure 5-5: Drain-current mismatch for Bulk 40nmNMOS devices (W/L=120n/40n) as a function of temperature (VDS = 50mV). T = 4.2, 40, 100, 150, 200 and 300K.[5.9][5.10]

Authors also said that further measurements on linear arrays of transistors uncovered a systematic mismatch related to device placement at the array edges, which has been attributed to mechanical stress. It was shown as a countermeasure that, when random mismatch is not overshadowing such an edge effect, **placing dummy devices** at the array edge can alleviate such effect.

5.2 FD-SOI DC characterization (LETI)

FD-SOI devices offers back-gate control nodes to both PMOS and NMOS transistors enabling the threshold voltage control in a very efficient way down to mK[5.11][5.24][5.25][5.26][5.30]. In addition thin film FDSOI devices eliminate unwanted kink effects [5.27][5.28], and maintain low variability [5.11][5.37]. For these reasons, FDSOI-based technology appears to be a privileged candidate for low temperature applications

The figures below show that STM 28nm FDSOI technology exhibits gain on mobility, which is estimated as higher than bulk CMOS or FinFET. The maximum normalized *gm* for short channel exhibits 30% increase for V_{DS} between 0,3V and 1V, and more than 50% at 50mV [5.37].

The slope under threshold increases drastically, reducing the off current, and as in the case of bulk MOSFETs, comes along with a V_T increase as well. **Nevertheless, with FDSOI back-gate voltage, the threshold voltage can be tuned in order to recover room temperature V_T.**

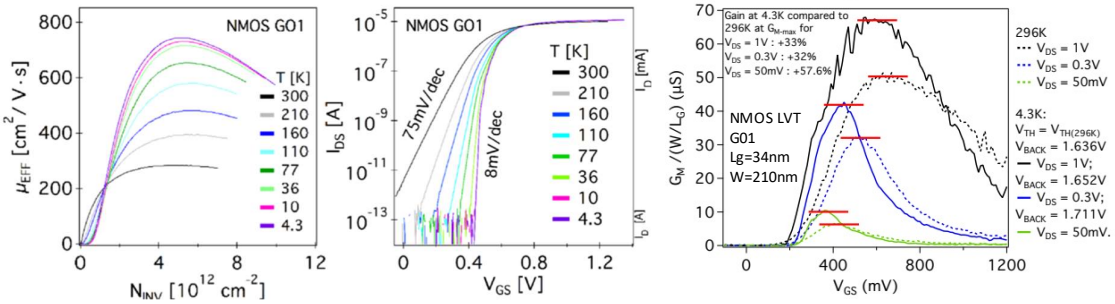


Figure 5-6: Mobility, subthreshold swing and gm evolution for 34nm FDSOI NMOS under T° variation.

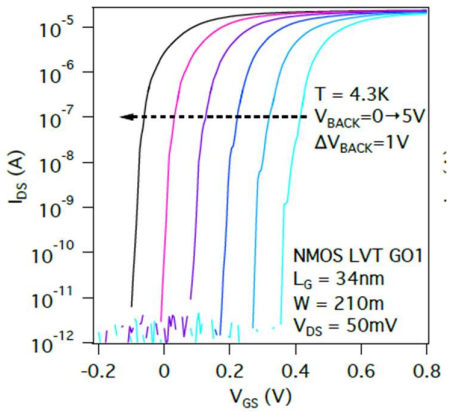


Figure 5-7: Impact of the back-gate voltage at on a 34nm FDSOI NMOS at 4.3K.

In Cardoso Paz *et al* [5.38], electrostatic and transport for 28nm FSDSOI NMOS and PMOS is evaluated and compared while lowering temperature from 300 K down to 4.2K. FD-SOI versatility is shown over the temperature range of operation, as the back gate (V_B) tuning efficiency is preserved at low temperatures, down to ultra-low temperatures as low as 100mK [5.11]. Very large threshold voltage tuning range is offered by means of forward and backward biasing ranging on +/- 5V depending on transistor flavor. The figure below shows that V_B=2V is enough to compensate the LVT NMOS threshold voltage from 300K to 4K. The drain current of short channel transistors can be further improved by the back-bias at 4K. This is due to the addition of both Front-Channel and Back-Channel contributions, with, in particular, a higher back channel mobility. Moreover, from room temperature down to 4.2 K, **access resistance decreases due to mobility improvement in source and drain regions**, which also contributes to drain current enhancement.

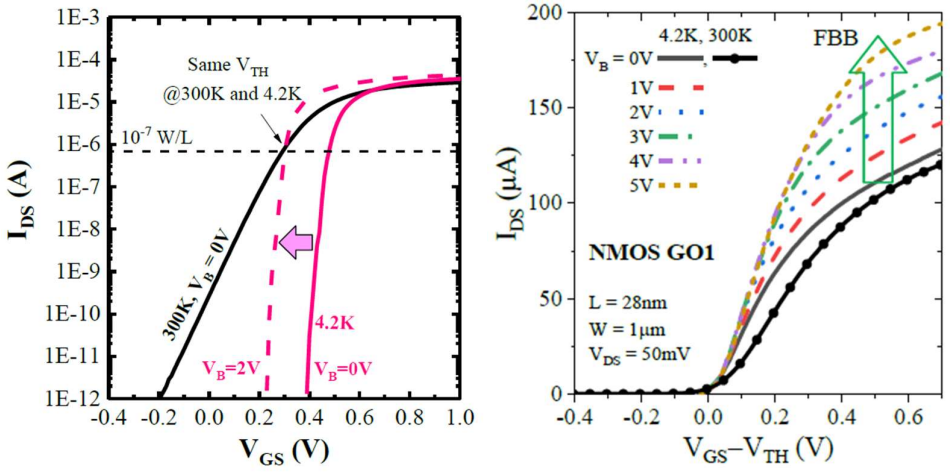


Figure 5-8: Impact of the back-gate voltage (left), and mobility gain in the weak inversion region (right) on 28nm LVT NMOS FDSOI.

28nm FDSOI variability is evaluated and published in Cardoso Paz *et al* ICMTS-2020 [5.42] and in Cardoso Paz *et al* VLSI-2020 [5.11] at 4.2K and below, and compared to 40nm BULK CMOS in term of mismatch performance. **The used configuration with on-chip addressing of a matrix of transistors forms a convenient approach for mismatch characterization at cryogenic temperatures, when cryogenic semi-automatic probing station is not available.** In [5.11], the authors demonstrate that **28nm FD-SOI outperforms other**

CMOS technologies at 4.3K in performance and variability. High performance is achieved even for short channel transistors, with I_{ON}>1mA/µm and I_{OFF} below the equipment accuracy <1fA, in particular by keeping advantage of back biasing. Comparison is summarized in the table below:

		28nm FDSOI L = 28nm		40nm Bulk L = 40nm	
		300K	4.2K	300K	4.2K
V_{TH} (V)	V_B= 0 V	0.29	0.49		
	V_B= 2 V		0.29		
I_{ON}/W (µA/µm)	V_B= 0 V	689	790	442	500
	V_B= 2 V		1088		
I_{OFF}/W (pA/µm)	V_B= 0 V	900	<0.01*		
	V_B= 2 V		<0.01*		
SS (mV/dec) @V_{DD}		74	4.8	88	28
Weak inversion: g_m/I_{DS} (V^{-1})		31.2	300	27	92
A_V (V/V)		27	36	20.8	23
A_{VT} (mV.µm)		0.98	1.24	3.0	3.5
$A_{\Delta\beta/\beta}$ (%.µm)		0.51	0.74	0.46	0.91

* Limited by the accuracy of the equipment.

Figure 5-9: Performance comparison of 28nm FDSOI at low temperature [5.39].

In this previous paper, **28nm FDSOI study is also pushed at Ultra Low Temperature (ULT) down to 100mK**, in order to evaluate FDSOI MOFET operation in the closest vicinity of the qubits. No loss of back biasing efficiency in DC operation in FD-SOI has been observed down to 100mK. **An increase in variability in subthreshold** regime related to subthreshold current oscillations have been reported due to disorders in the channel (see Figure 5-8). This phenomenon **can be avoided for L>160nm**, gets more pronounced as L decreases, is almost independent on W, and is stronger in PMOS. However this effect has no significant impact in inversion regime and only a small decrease of correlation between local fluctuations of V_{TH} in linear and saturation regime at LT (w.r.t. RT) was observed for short NMOS.

The physical origins of MOSFET mismatch at low temperature are studied, highlighting the impact of the band tail states on threshold voltage and current gain variabilities. However, these variabilities remain reasonably low in comparison to Room Temperature values and to other CMOS technologies, that's the reason why authors estimate that it should not be detrimental to circuit operation in this range of temperatures.

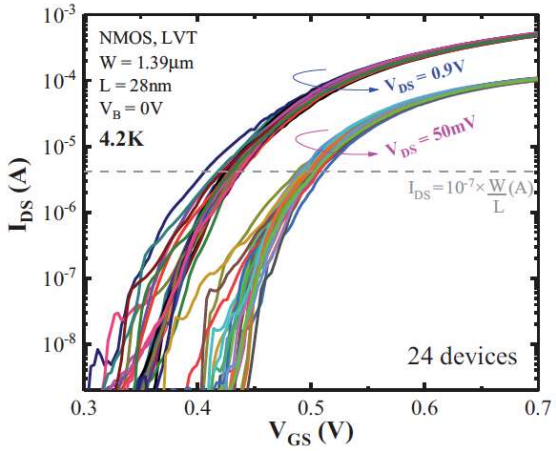


Figure 5-10: I_{DS} versus V_{GS} for 24 short channel (L = 28 nm) N-type LVT MOSFETs at 4.2 K, at V_{DS} = 50 mV and 0.9 V. With the logarithmic scale, variability is highlighted in the subthreshold regime. V_{TH} is extracted on the dashed line.

The results highlight the need of taking into account the slight degradation of variability at liquid helium temperatures in compact transistor models. **Extensive modelling effort (which is presented in a next subsection) is also convoked** in the observations reported by Cassé *et al* in the next published TED ESSDERC 2020 [5.41] and in [5.40]. Along with temperature decrease in long channel FDSOI devices, humps appear in the current, leading to oscillations of the transconductance with gate voltage. The inherent mobility discontinuity is demonstrated to be due to intersubband scattering, in relation with the 2D subband structure in thin film SOI. These features have to be taken into account for future cryo-CMOS compact models.

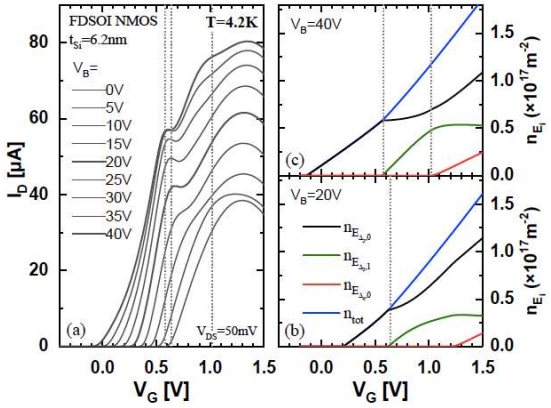


Figure 5-11: drain current I_D measured at T=4.2K on long channel NMOS transistors as a function of the gate voltage V_G, for different back gate biases V_B from 0 to 40V and a fixed drain voltage V_{DS}=50mV

In order to conclude this subsection, let's mention that GlobalFundries 22FDX get in the game.

In Bonen *et al* EDL 2019 [5.43], single finger 20nm NMOS and PMOS are characterized at 2K. The authors also observe I_{DS} oscillations in the subthreshold regime with large peak-to-trough ratios, appearing at V_{DS} < 50 mV in both n- and p-MOSFETs. They conclude on the signature of electron/hole resonant tunneling through the discrete energy levels of the single finger working as a quantum dot.

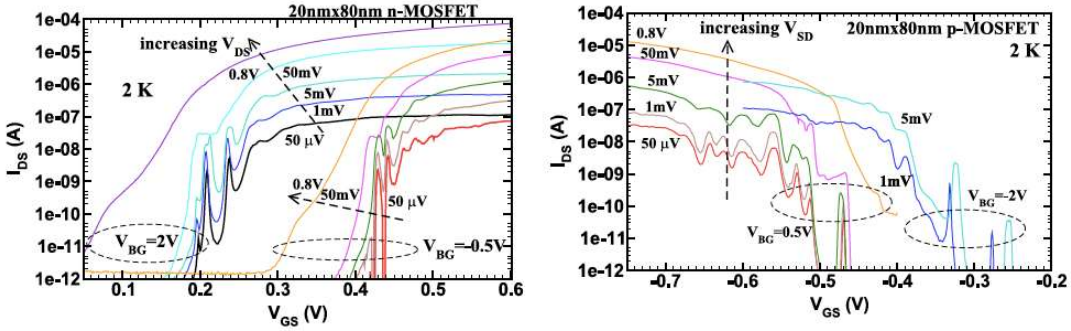


Figure 5-12: Measured MOSFET transfer characteristics vs. back-gate voltage and VDS at 2 K for (a) n-MOSFET and (b) p-MOSFET.

Cardoso Paz *et al* will publish in late 2020 [5.44] a characterization result of 22FDX technology. Figure below shows (at left) V_{TH} as a function of T for a nMOS at V_{DS} = 50mV and 0.9V, and at back gate bias (V_B) of 0V and 1.4V. The applied 1.4V at the back bias gives the same V_{TH} value as for the device operating at room temperature. On the right side, figure indicates the V_B required to shift V_{TH} back to its value at 300K for each temperature operation. FBB is an efficient way to correct the significant V_{TH} increase with temperature lowering and, therefore, enhance power efficiency.

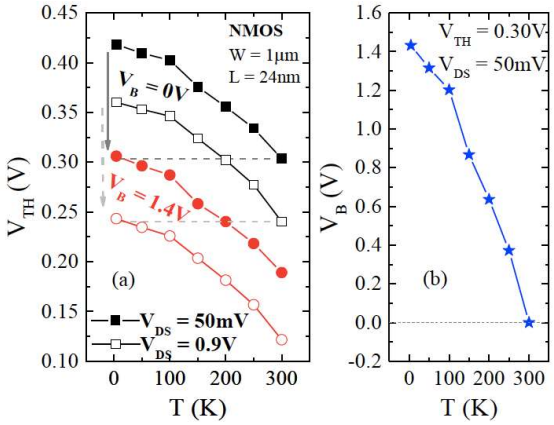


Figure 5-13: Impact of the backgate on 22FDX nMOS along temperature.

5.3 RF characterization (LETI)

Until now, very few publications on the RF properties of MOSFETs at low temperature can be found in the literature. A highly referenced paper was published by Siligaris *et al* in TED 2006 [5.48], reporting investigations on Small Signal Equivalent Model, f_t, f_{max}, Γ_{opt}, R_n and NF_{min} for Bulk 55nm nMOSFET at 78K in a measurement range from 5GHz to 45GHz. Siligaris *et al* have demonstrated that a 34% and a 78% improvement is obtained, respectively, on f_t and f_{max} at 78 K when compared to ambient temperature performances. In weak inversion, NF_{min} is reduced from ~ 1.4 to 0.5 dB at 30 GHz. Same trend is observed on the noise resistance R_n, which decreases from ~ 25 to 7 Ω at 30 GHz.

SSEC ELEMENTS (I_{ds} = 341 mA/mm, V_{ds} = 0.7 V)

T	g_m mS	g_d mS	C_{gin} fF	C_{miller} fF	$\frac{C_{miller}}{C_{gin}}$	R_g Ω	R_d Ω	R_s Ω	f_c GHz
296 K	136	20	49.3	31.4	0.64	10.6	1.3	1	439
78 K	183	37	58.2	31.4	0.54	1.65	0.6	0.2	501

T	R_i Ω	R_{gd} Ω	L_g, L_d pH	L_s pH	C_{pg} fF	C_{pd} fF	C_{ds} fF
296 K	0.2	3	20	6	6	39	2.4
78 K	0.3	1.9	28	3	6	40	3.1

Figure 5-14: This table summarizes the values of the electrical model for a 55nm MOSFET at 296K/78K.

Similar results down to 4.2K were obtained in 2008 by Hong *et al* on a 90nm Bulk CMOS[5.49]. 60% and 80% improvement is obtained on respectively f_t and f_{max}, although a maximum is reached for f_{max} around 25K.

According to the authors, "**R_g gradually decreases as the temperature decreases down to 50 K from the increase of mobility, and inversely, R_g increases with the temperature below 50 K because of the carrier freeze-out of the gate polysilicon**. The electrical conductivity of a semiconductor material is dependent on

the product of free carrier density and mobility. The gate resistance slightly increases because the free carrier density decreases, and the mobility is nearly saturated when the temperature drops below 50 K".

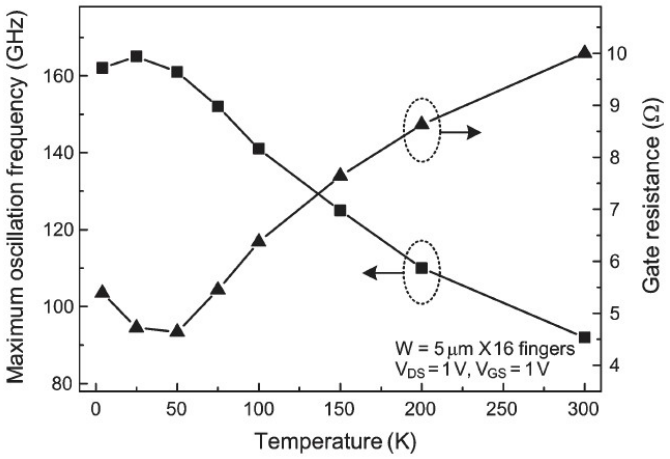


Figure 5-15: Dependence of the maximum oscillation frequency (f_{max}) and gate resistance (R_g) on temperature. The peak f_{max} is 165 GHz at 25 K [5.49].

In IMS 2014, Coskun and Bardin reported a cryogenic noise model for a 100 fingers 32nm SOI CMOS in the range of 6K-300K [5.16]. They proposed a noise measurement methodology that is based on room temperature noise measurement mixed with a temperature dependent small signal model. S parameters are performed until 67GHz. When temperature decreases from 300K to 6K, 18% and 78% improvement is obtained on respectively on f_t and f_{max}. The temperature dependent equivalent model is given in the table below (unfortunately, reported values for *gm* and f_t are inconsistent with curves and analysis in [5.16]). From their study, the team concludes on the ability for this technology to reach a minimum noise temperature of 3K, meaning a NF_{min} around 0,05dB.

NORMALIZED SMALL-SIGNAL NOISE MODEL PARAMETERS AND BIAS POINT. $V_{DS} = 0.9$ V.

T_a	R_g	R_s	R_d	R_{gs}	C_{gs}	C_{gd}	C_{ds}	g_m	g_{ds}	τ	T_d	f_t	f_{max}	I_D
K	$\Omega \cdot \mu$m				fF/μm			mS/μm		fs	K	GHz		mA/μm
6	43	27	47	380	0.55	0.31	0.27	1.3	0.11	128	2200	243	511	0.10
77	69	39	98	220	0.50	0.31	0.23	1.3	0.11	104	2200	240	437	0 .10
200	140	46	46	1050	0.54	0.31	0.18	1.3	0.12	195	2250	234	318	0.125
293	130	38	47	1150	0.57	0.31	0.23	1.3	0.15	25	2350	237	289	0.175

Figure 5-16: This table summarizes the electrical and noise model for a 100 fingers 32nm SOI CMOS from 293K to 6K [5.16].

More recently in 2019, a team from Université Catholique de Louvain presented (in S3S conference [5.50]) and published (in Journal of the Electron Devices Society [5.51]) a detailed analysis of 28nm FDSOI MOSFETs RF characteristics at 77K and extended recently down to 4,2K [5.52].

Gate lengths from 25 to 150 nm are investigated. The Si film, BOX and the equivalent gate oxide thicknesses are 7, 25 and 1.3 nm, respectively. Studied nMOSFETs feature 60 fingers of 2 µm width. S-parameters are measured in a frequency range from 10 MHz up to 40 GHz under V_{DS} = 0V, 0,6V and 1 V for different applied

gate voltages. **The back gate was kept grounded**. f_t and f_{max} as well as elements of small-signal equivalent circuit are extracted from the measured S-parameters.

In these studies, that's interesting to analyze the evolution of characterization methodology and extracted results that were obtained by this team between 2019 at 77K, and 2020 at 4,2K.

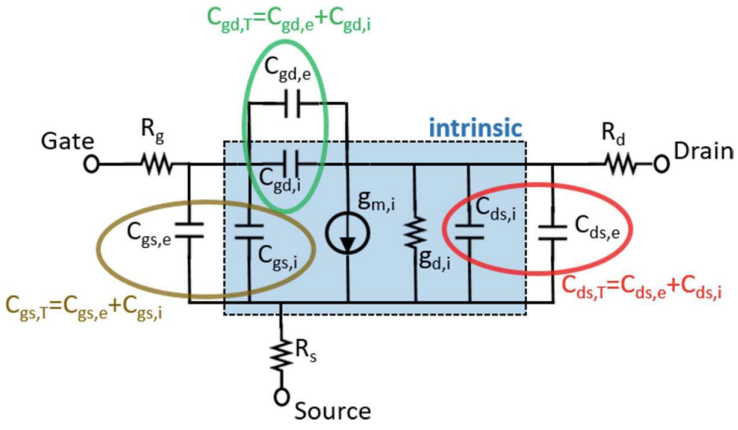


Figure 5-17: RF model of FDSOI UTBB MOSFET small-signal equivalent circuit.

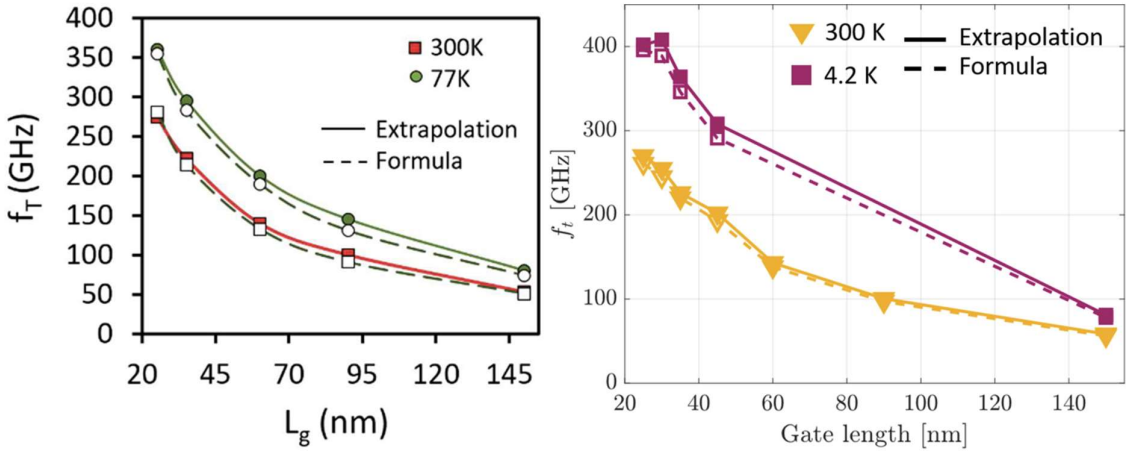


Figure 5-18: f_t vs. Lg for 4K, 77 K and 300 K at Vds = 1V and at Vgs that corresponds to maximum gm. Left : 2019; right : 2020

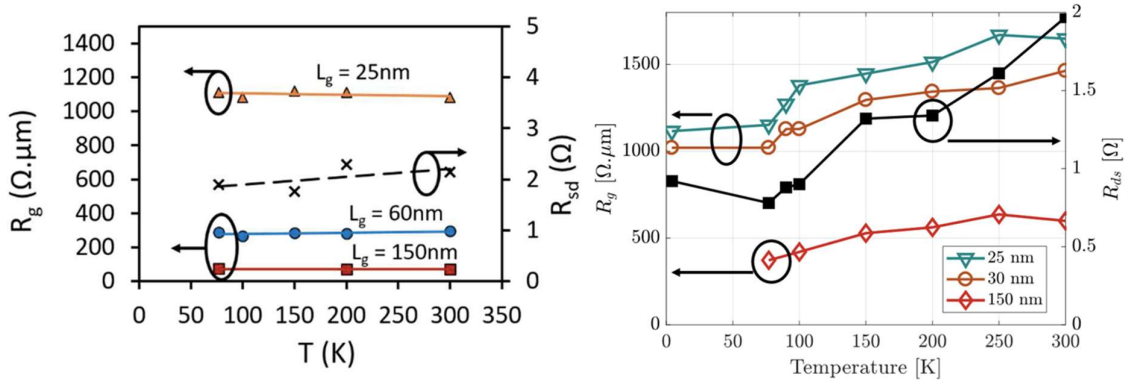


Figure 5-19: For various gate lengths Rg, Rsd versus T° extracted with Bracal method. Left : 2019; right : 2020

In 2019, f_t and f_{max} exhibited an improvement of 30% and 20% respectively for short channel, with a maximum around 100K. By the extraction of intrinsic and extrinsic parameters for the devices with different lengths and at various temperatures, it was concluded that Cgs, Rg, gds and Cgd/Cgs do not change much with temperature lowering and therefore, f_t and f_{max} temperature dependencies would be mostly related to gm temperature dependence.

In 2020, f_t and f_{max} are finally found to increase of 50%. Room temperature extraction exhibit higher values than in 2019 as well. And the team concludes that this improvement is **due to gm enhancement of 40%, Rg enhancement of 30%, and in a second order for short channel, to the Cgs decrease.** The maximum at 100K is observed again by this new characterization methodology. The authors observed a monotonic resistance decrease of ~30% with temperature lowering down to 70 K, then the value stabilizes. They think such a trend is in agreement with the TiN resistivity behavior at cryogenic temperatures.

In [5.52], this characterization and extraction methodology is described with details. At each temperature, an off-wafer Short-Open-Load-Thru (SOLT) calibration is first performed with an impedance standard substrate (ISS) calkit that is placed inside the manual probe station, so, at the same temperature as the samples. Then, the reference plane after de-embedding is supposed to be at the first metal layer of the FETs. The "load standard resistance" variation under temperature was found in the same range as the calkit load itself, therefore this variation was not taken into account. The related error in the small-signal model extraction was estimated to 5%, but the authors mention that **Rg and f_{max}, which are extracted above 10 GHz, are more sensitive to calibration accuracy. And they consider that maximum operant frequency of this calibration is limited to 20GHz. Another difficulty to be overcome is the quality of probing contacts at low temperature, leading to strong variation on Rg, Rds, *gm*, and f_t/f_{max} extractions.**

In 2019, Cassé *et al* in [5.35] also observed a strong variation of Rg along from 300K to 4K (see figure below).

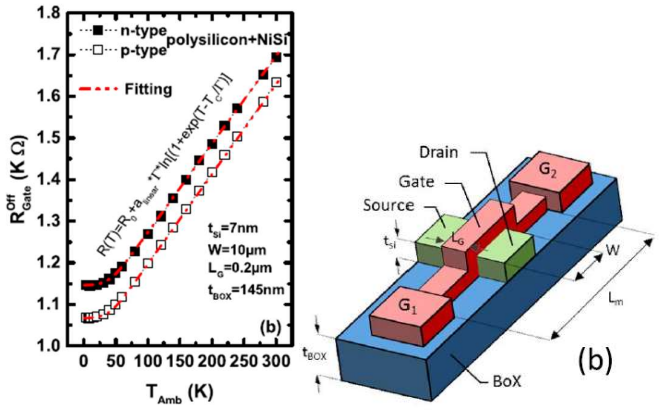


Figure 5-20: *(left) Gate resistance calibration curve showing R_{Gate} versus ambient temperature with no current flowing in transistor channel (off state). (right) Schematics of the structure used to measure R_{Gate}.*

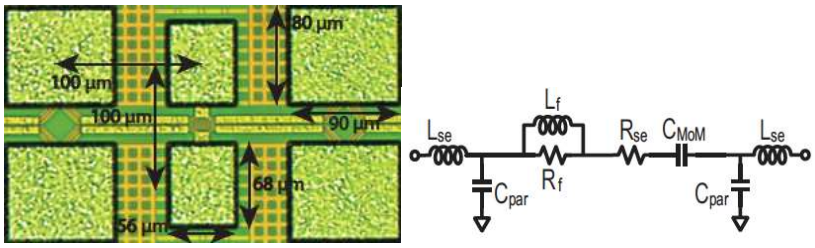
No measurement was performed under back-gate variation. **The lack of extensive data suggests that further investigations with RF measurements are necessary to understand extrinsic and intrinsic parameters variations under temperature down to 4K**.

Before closing this subsection on RF characterization, let's mention the only one paper that was found on back-end passive characterization. QuTech-Delft in JEDS 2020 [5.36] reported a study of on-chip MoM capacitor and transformer at 4K.

With a Lakeshore CPX cryogenic manual probe station, the two components from TSMC 40nm BULK CMOS were characterized, modelized and retro-simulated by EM simulation at room temperature and 4K.

A high-density rotative MoM capacitor with a poly shield was taped-out using stacked inter-digitated metal fingers in layers 1 to 5 with a finger width of 100nm and a spacing of 90nm. To increase the capacitance to a measurable value and in order not to be dominated by the parasitics of pads, 10 such capacitors were connected in parallel, with 6 horizontal and 38 vertical fingers. This provides a capacitance of 202 fF for an area of 150 µm2 (7.97 µm x 1.89 µm x 10).

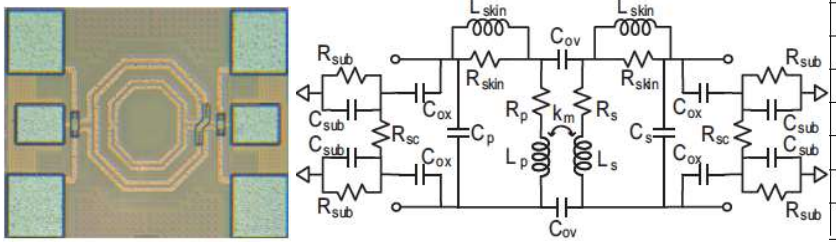
No strong changes were observed at low temperature.



Parameter	Units	RT	4 K
C_{MoM}	fF	191	197
R_{se}	Ω	0.5	0.1
C_{par}	fF	28	26.5
L_{se}	pH	20	20
R_f	Ω	3	3
L_f	pH	17	24

Figure 5-21:Micrograph. Equivalent model and T° fitting of the MoM capacitance.

The transformer is a two-turn primary winding with 190 µm diameter and 8 µm trace width and a two-turn secondary coil with 130 µm diameter and 7 µm trace width, using the ultra-thick metal layer. Shielding of the transformer was prevented due to a highly resistive substrate at 4K.



Parameter	Units	RT	4 K
L_p	pH	691	650
R_p	Ω	0.95	0.22
R_{skin}	Ω	1.2	0.53
L_{skin}	pH	48	48
C_p	fF	18	19
C_{ox}	fF	50	52
C_{sub}	fF	19	22
R_{sub}	kΩ	1.24	1000
R_{sc}	kΩ	2	2000
C_{ov}	fF	4.5	4
k_m	-	0.367	0.39

Figure 5-22:Micrograph. Equivalent model, and T° fitting of the transformer.

The series resistance of copper traces is strongly reduced by a factor 5, at low frequency. When skin effect takes advantages, this gain is reduced. But because of electric field confinement, the current flowing in the conductor interior reduces, decreasing Lint, and thus, the total inductance. the 5x increase in conductivity led to a 5% reduction in inductance.

For low resistive substrates, the capacitance from the windings to the ground plane is dominated by Cox, while for highly resistive substrates, the effective capacitance is lowered by Csub in series with Cox, resulting in a slight increase in the frequency where peak quality factor occurs. The self-resonance frequency of the transformer increases by 5%, due to the decrease in both inductance and effective parasitic capacitance to ground.

5.4 Thermal conductivity and Self Heating (LETI)

Self-heating is generally evaluated under DC condition[5.23][5.53], or due to extremely short thermal time constant in modern devices, using RF-based techniques [5.46][5.54].

The thermal conductivity of Si strongly decreases with Si thickness due to phonon-boundary scattering, reducing to 13 W/m.K in the 10-nm range at room temperature, compared to 140W/m.K for bulk Si. The Silicon-based material thermal conductivity exhibits a temperature dependence that is related to the material thickness and crystallographic arrangement. Scientific community understands this property thanks to the various phonon dispersion strategies [5.34][5.55].

The figure below shows that the thermal conductivity of bulk silicon exhibits a maximum value around 30-40K, and this maximum shifts towards higher temperatures for thin layers [5.35]:

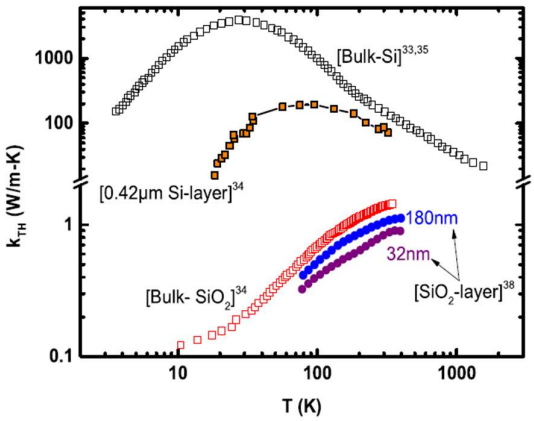


Figure 5-23: Thermal conductivity data versus temperature for bulk and Si-layer compared to that for bulk and SiO_2-layer..

Self-heating at cryogenics temperature in advanced bulk technologies has not been studied so far (see ref.[5.31] and references therein). One could expect improvement considering the increase of thermal conductivity of Si with decreasing temperature. **However the benefits are probably very limited in 3D FiNFET for which heat evacuation towards substrate is limited** [5.21][5.22][5.23].

In FDSOI devices, the BOX tends to confine the heat in the channel, and therefore the total thermal resistance depends on both the thermal conductivity of Si and SiO_2 which have different T-dependence and magnitude (see above figure). In ultra-thin film devices, like for 28nm FDSOI technology, the thermal resistance follows the temperature-dependence of thermal conductivity in silicon dioxide at room temperature [5.56][5.35]. At low temperature, typically below 10 K, both values of thermal conductivity of Si and SiO_2 have decreased as compared to the room temperature case, with k^{10K}_{SiO2}= 0.1 W/mK and k^{10K}_{Si} <1 W/mK for thin silicon layer or silicon nanowire. Experiments reported by Cassé *et al* in [5.35], based on gate DC resistance sensing, indicate that even at 4K, thermal resistance in the FDSOI transistor does not depend significantly on the SOI thickness (ranging from 7 up to 24 nm), whereas the BOX thinning (145 and 25 nm) strongly reduces the magnitude of the thermal resistance. As a result, the thermal resistance is strongly temperature dependent, especially at very low temperature, as illustrated below:

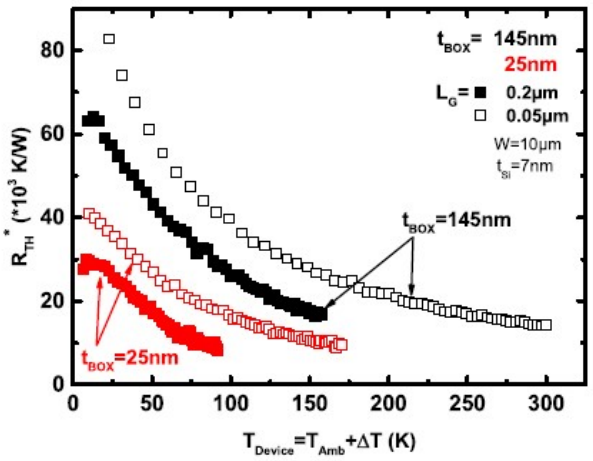


Figure 5-24: Thermal resistance R_{TH} versus device temperature, for wide and ultrathin FDSOI MOSFETs (from [5.35]).

This leads in particular to a nonlinear temperature increase of the device with the dissipated power, and potentially to a significant temperature increase of the device compared to the ambient cryogenic temperature.

Before macroscopic thermal diffusion in the complete chip and beyond, "thermal dissipation at the active region of electronic devices is a fundamental process of considerable importance. Inadequate heat dissipation can lead to prohibitively large temperature rises that degrade performance, and intensive efforts are under way to mitigate this self-heating" [5.45]. **In the context of low temperature (from few K to few mK), the power budget in coolers falls down drastically (typically, 1W at 1K, 1mW at 100mK), and then becomes an additional constraint leading to look for enhancement of cryoelectronic performances**.

Nyssens *et al* published in 2019 and 2020 [5.47] an extraction of the channel thermal resistance in 28nm FDSOI at 77K, based on the RF technique (S-parameter measurements). They confirm the increase in R_{th} when temperature is reduced down to 77 K, which is more important for the longest devices than for the shortest ones, suggesting different dominant heat evacuation paths according to the gate length. Contrary to their expectations, they conclude that the effect of self-heating on analog FoMs (gm, gd and intrinsic voltage gain) is slightly weaker at cryogenic temperatures with respect to room-temperature case. Thus they found that the voltage gain reduction due to SH is attenuated by 2 to 8%, depending on the gate length, when the ambient temperature goes to 77 K. They also measured the SH characteristic frequency $f_c=1/(2\pi R_{th}C_{th})$ at 77K and found that f_c increases about four-fold. They conclude that it is due to a **reduction of thermal capacitance in the device when temperature decreases. Nevertheless, the RF technique to evaluate self-heating requires to know accurately the channel temperature, whereas the authors were lead to assume a temperature range during their experiment.**

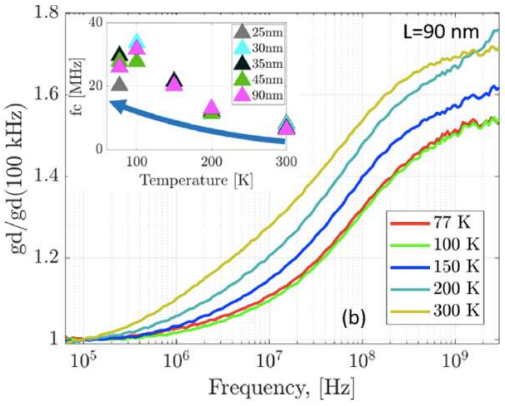


Figure 5-25: Normalized gd with frequency for different T° of a 90 nm-long MOSFET, at Vd=Vg=1V. The Inset shows the characteristic frequency of SH effect versus temperature for different device lengths.

A more complex analysis to evaluate self-heating might be done by mixing the studies from Cassé *et al* in [5.35], based on the gate DC resistance sensing, and studies from Nyssens *et al* in [5.47], based on the RF technique.

5.5 Cryo-CMOS Compact Modeling (EPFL)

This section highlights some of the challenges faced for the modeling of MOSFET devices for operation at cryogenic temperature (CT). A special focus is given on the modeling of the *threshold voltage* V_T and the *subthreshold swing* SS. The significant increase of V_T at CT reduces the available overdrive voltage and therefore needs to be modeled properly. The SS saturates to a constant value below a critical temperature T_c of typically 40 K. This mitigates the current saving that could be expected from reducing the temperature since the transconductance for a given current does not scale with $1/T$ below T_c. A correct modeling of these two phenomena is therefore key for developing an improved compact model (CM) that scales with T from RT down to CT.

The recent progress in the development of silicon qubits for quantum computing has led to the exploration of more scalable and integrable solution for the implementation of the control and read-out electronics of quantum computers [5.57]. Integrating all or part of the electronics, moving it closer to the quantum processor and operating it at CT can potentially solve this problem [5.58]. The scalability of CMOS technology, its high-speed and low-power operation makes it an excellent candidate for addressing this issue. This is the main driver behind the recent effort in better characterizing the operation of MOS transistors at CTs [5.59][5.60][5.61][5.62].

Moving the front-end electronics closer to the qubits and operating it at CT sets a strong constraint on the power dissipation. Reducing the temperature and taking advantage of the high transit frequency and low-noise features of advanced CMOS technologies can certainly help to meet this power constraint [5.63]. However, their reduced supply voltage combined with the increase of V_T at CT, is limiting the bias range to weak and moderate inversion. Although this can be a limitation at room temperature (RT), it becomes an asset at CT. Indeed, biasing the transistors of analog circuits in weak inversion (WI) can take advantage of the decrease of the subthreshold swing SS to reduce the current required to achieve a given transconductance and better accommodates the increase of V_T. The decrease of SS also significantly reduces the leakage current and enables the use of reduced logic swing resulting in lower power consumption of digital circuits. However, the reduction of power consumption resulting from the decrease of SS is unfortunately mitigated by the its saturation at CT.

5.5.1. Current Status of Standard Compact Models (EPFL)

Several key building blocks of the control and read-out electronics have already been validated at CT [5.63][5.64][5.65]. However, most of the circuits published so far have been designed without verification by simulation as it is done for circuits running at RT. This is because the CMs available today in most PDKs of advanced CMOS technologies do not scale correctly down to CT. This is illustrated in Figure 26 which shows the I_D versus V_G transfer characteristic in saturation at several temperatures simulated from the PDK of a commercially available 28 nm CMOS bulk process. CMs scale reasonably well with temperature down to 77 K, but at 4 K and below, some discontinuities start appearing in the moderate inversion region. This is partly due to the imperfect modeling of the device which was not initially planned to be used at CT, but also numerical issues appearing due to the combination of exponential functions with very large or small exponents [5.66]. As shown in Figure 27, the Fermi-Dirac distribution becomes almost a step function and the intrinsic carrier concentration n_i takes on extremely small values below 10 K, causing arithmetic underflow in implemented analytical expressions or convergence problems in computer-aided-design simulations [5.66]. As shown in Figure 27, this yields n_i values lying outside the range of the IEEE double precision (10^{-308}-10^{308}), e.g., at 4.2 K, $\mathrm{n_i} \cong 10^{-678}\ \mathrm{cm}^{-3}$. Therefore, an extension of the arithmetic precision needs to be used when computing the CM [5.66].

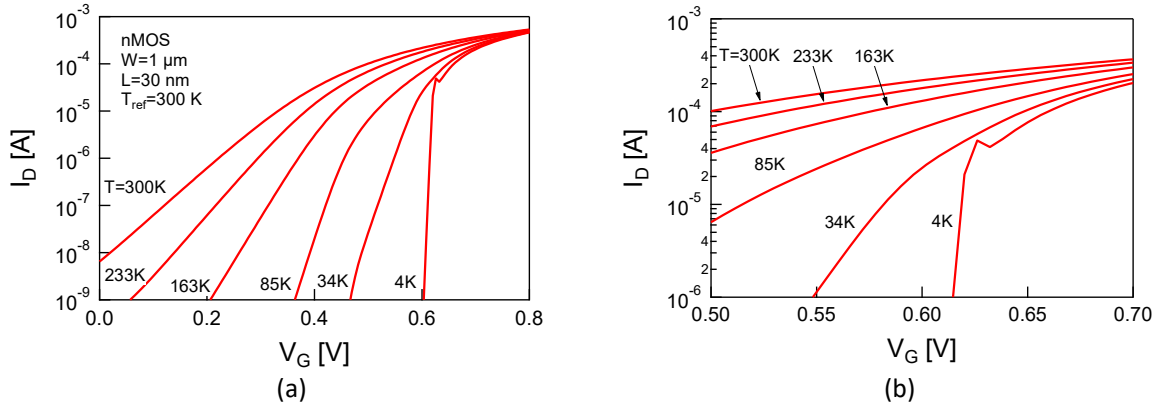


Figure 26: Simulated I_D versus V_G transfer characteristics in saturation simulated for different temperatures with the PDK of a 28-nm commercial bulk CMOS technology.

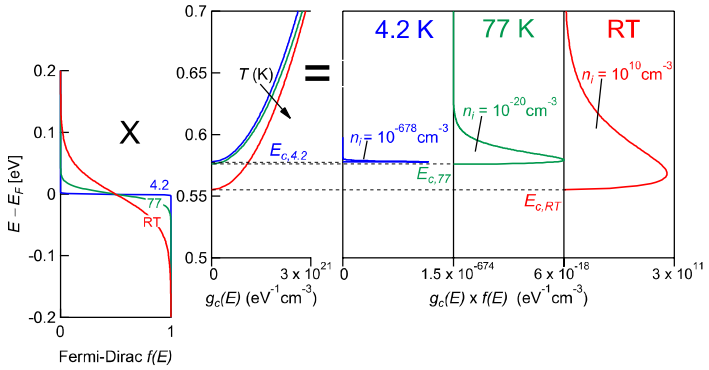


Figure 27: Intrinsic carrier concentration reaches extremely small values at 4.2 K. Typically $n_i \cong 10^{-678}\ cm^{-3}$ at $T = 4.2K$.

5.5.2. CT MOSFET Device Physics (EPFL)

Reducing T down to CT comes with several benefits, namely a substantial increase in mobility, saturated velocity and thermal conductivity, a significant reduction of the SS, leakage currents and reduced thermal noise. However, it also comes with some drawbacks including a significant V_T increase, possible oscillations

and kinks in the transfer and output characteristic, respectively. The behavior of mobility versus T has been extensively studied for various devices and is well understood and modeled [5.67][5.68]. Impurity freeze-out is becoming important for T below 100 K and can cause important kinks in the output characteristic of SOI devices at high V_{DS} voltage [5.68]. Because of the lower supply voltage, these kinks have not been observed for the 28 nm FDSOI devices presented in [5.61] at least within the nominal operating voltage range. Freeze-out has finally little impact on the dc characteristics of MOSFETs in advanced technologies.

5.5.3. Threshold Voltage Modeling (EPFL)

Even though the increase of V_T at CT can be circumvented by a proper back gate bias in FDSOI, V_T remains a key parameter in CMs and its increase with T must be understood and correctly modeled. It is sometimes claimed that dopant freezeout is responsible for the increase of V_T. However, dopant ionization due to the gate voltage is visible in the C-V plot as a bump near flatband [5.69][5.70]. This indicates that, in late depletion and inversion, dopants near the surface are ionized and that, despite the frozen substrate, the MOSFET can still operate normally. Figure 28 shows that the CT behavior of V_T is not so much impacted by dopant freezeout. V_T is a few mVs lower at all temperatures because E_F in the frozen substrate is positioned a few mVs further from E_c (in nMOSFETs). Therefore, dopant freezeout can be neglected in CMs for enhancement-type MOSFETs. It is however key to include a correct model of the Fermi potential $\Phi_F(T)$ as illustrated in Figure 29 to obtain the desired $V_T(T)$ behavior. Furthermore, in some pMOS devices, a kink is observed in $V_T(T)$, as shown in Figure 30 [5.71][5.73]. This kink cannot be explained by the saturation of $\Phi_F(T)$ nor by dopant freezout. Non-uniform interface traps close to E_c can add some additional V_T shift that starts at CT, but this is not sufficient to model the kink. Including an effective temperature dependence in C_{ox} works, which points to the pMOS gate stack [5.73].

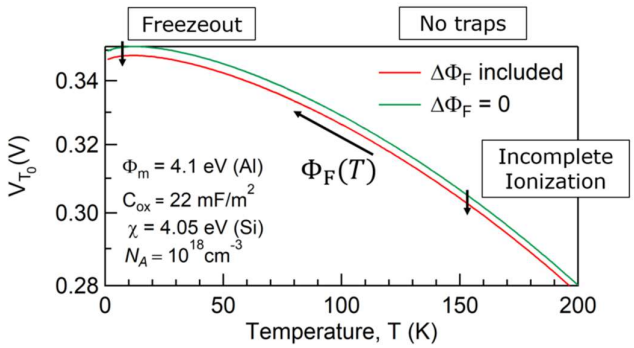


Figure 28: Equilibrium threshold voltage (V_{T0}) model versus temperature. The change of V_{T0} is mostly due to the change in the Fermi voltage as illustrated in Figure 29. The slight impact of incomplete ionization is indicated by the difference between the green curve, which does not account for incomplete ionization, and the red curve including the change of the Fermi potential $\Delta\Phi_F$ due to incomplete ionization. $\Delta\Phi_F$ is also represented in Figure 29.

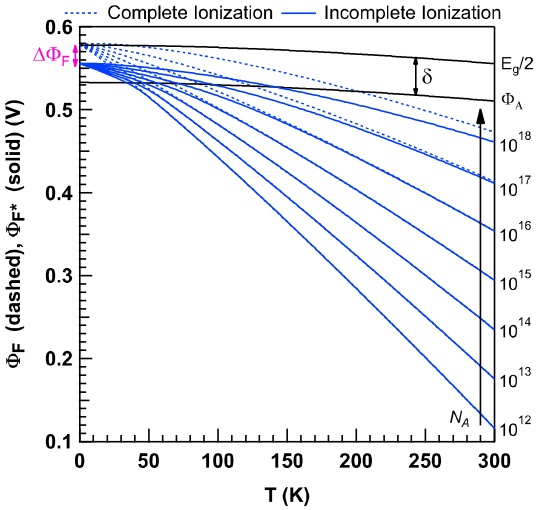


Figure 29: Bulk Fermi potential versus temperature including dopant freezeout (solid lines) and complete ionization (dotted lines). The figure illustrates the small impact of incomplete ionization compared to the strong variation of the Fermi voltage over temperature.

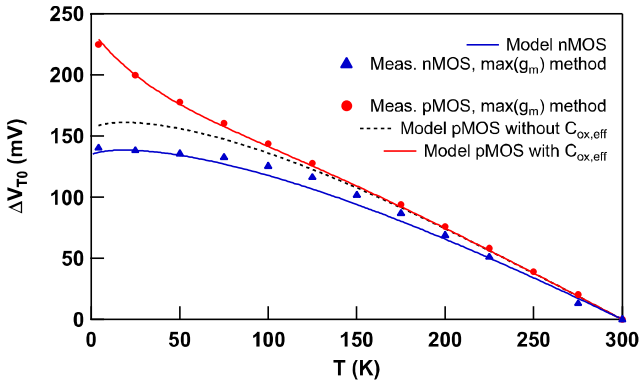


Figure 30: Change of the threshold voltage with temperature with respect to its value at RT $\Delta V_{T0}(T) = V_{T0}(T) - V_{T0}(300K)$. The deviation observed in ΔV_{T0} for pMOS devices below 50 K can be modelled with an effective T_{dep} in C_{ox} [5.70].

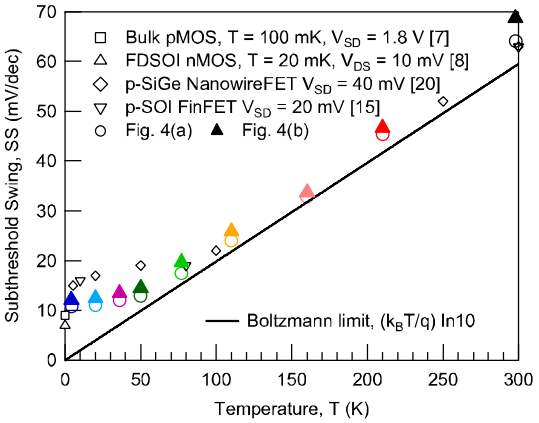


Figure 31: Saturation of SS for different advanced technologies [5.73].

5.5.4. Subthreshold Swing Modeling (EPFL)

The subthreshold swing SS starts to saturate at low temperature as illustrated for various devices in Figure 31. This is due to the disorder-induced band tails [5.72][5.73][5.74] which makes the SS saturate from $nU_T \ln 10$ to $nW_t/q \ln 10$ below T_c, where $U_T \triangleq kT/q$ and $W_t = kT_c$ is the band-tail width, typically around 4 meV (corresponding to $T_c \cong 46\,K$), and n is the standard *slope factor* $n = 1 + (C_{dep} + C_{it})/C_{ox}$. This behavior has drawn a lot of attention recently and several models have been proposed [5.72][5.73][5.74]. A physical model is proposed in [5.73] which leads to a closed form expression using Gauss hyper-geometric functions. Unfortunately, the latter are not available in Verilog-A and are therefore not appropriate for CM. An approximation using only simple functions available in Verilog-A was recently proposed in [5.74]

$$SS(T) = \ln 10\, n \frac{kT_c}{q}\left[1 + \alpha \ln\left(1 + exp\left(\frac{T - T_c}{\alpha\, T_c}\right)\right)\right] \quad (1)$$

where α is a fitting parameter typically equal to 0.1 [5.74]. The above equation (1) is plotted in Figure 33 a) and compared to the full hypergeometric theoretical model from [5.73] and measurements made at a constant current from the I_D-V_G characteristics shown in Figure 41. A good match is obtained between measurements and (1) for $\alpha = 0.4$.

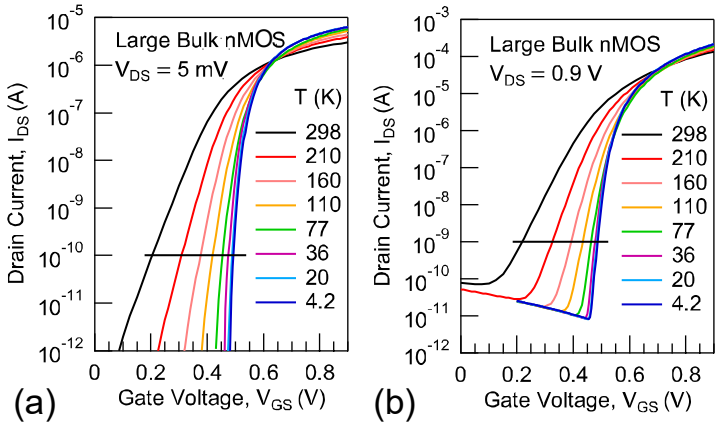


Figure 41: Low-temperature measurements in a large bulk MOSFET from a 28-nm bulk CMOS process \cite{bib:beckers:edl:41:2:feb:2020}.

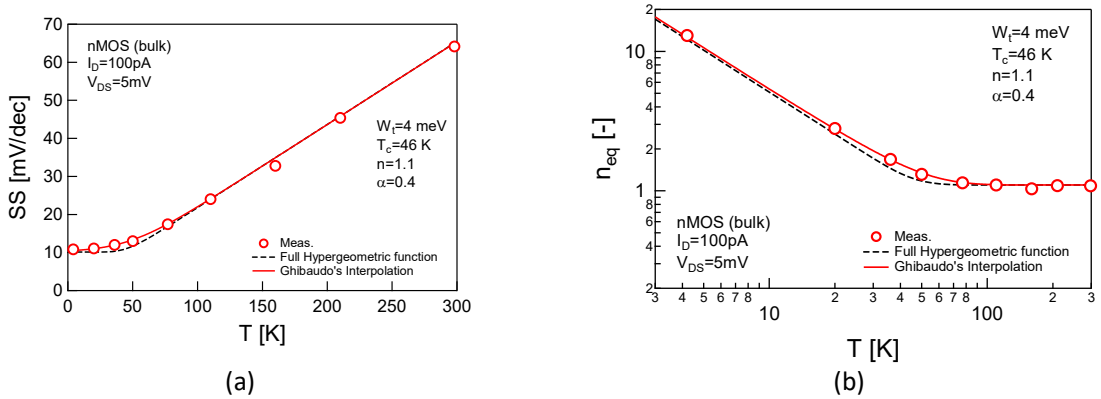


Figure 33: a) Saturation of the SS versus T at CT. The dashed black line corresponds to the physical model using the hypergeometric Gauss functions presented in [5.73], whereas the red line corresponds to the approximation proposed by Ghibaudo given by (1) [5.74]. The symbols represent the SS extracted at a constant current $I_D = 100\,pA$ from I_D-V_G characteristics shown in Figure 41 a). A good fit between the approximation and the measurements is obtained for $\alpha = 0.4$.

b) Equivalent slope factor $n_{eq}(T)$ as defined by (2) versus T. The strong increase of n_{eq} below 50 K compensates the decrease of U_T so that the product $n_{eq}(T)U_T$ saturate to $n\,kT_c/q$.

Above T_c, the drain current in WI is typically proportional to $exp[(V_G - V_{T0})/(nU_T)]$ where $U_T \triangleq kT/q$. For CM it is actually convenient to keep this expression and replace the slope factor n by a temperature-dependent slope factor $n_{eq}(T)$ defined as

$$n_{eq}(T) = \frac{SS(T)}{\ln 10\ U_T} \tag{2}$$

Eq. (2) is plotted versus T in Figure 33 b) which shows that the slope factor artificially increases for $T < T_c$ to make the product $n_{eq}(T)U_T$ saturate to $n\,kT_c/q$ for $T \ll T_c$. Note that the proposed $n_{eq}(T)$ is similar to what Tewksbury presented already in 1985 [5.75]. However, he attributed this increase to interface traps and not to band-tails.

5.5.5. Small-signal, Noise and RF Model (EPFL)

At RT, the transconductance in WI and saturation is given by $G_m = I_D/(nU_T)$ and therefore scales inversely with T. This means that the current required to achieve the same G_m at CT than at RT would ideally be divided by $T_{RT}/T_{LH} = 300/4 = 75$, which is a significant current saving! Unfortunately, for $T < T_c$, the transconductance becomes $G_m = I_D/(n_{eq}U_T)$ does not scale with $1/T$ anymore. The current saving then reduces to $T_{RT}/T_c = 300/40 = 7.5$ which is ten times less! This is why it is important to correctly model the SS and the related slope factor in the CM.

WI remains interesting for low-power at CT because it also offers the highest transconductance for a given current. This property is captured by the G_m/I_D figure-or-merit (FoM) which is plotted in normalized form in Figure 43 versus the inversion coefficient IC [5.76] for minimum length 28 nm bulk and FDSOI devices at RT and CT [5.59][5.61]. It is remarkable to observe that, after a correct normalization, the characteristic is identical for both devices and remains almost invariant wrt T. The only change occurs in strong inversion ($IC > 50$), where the G_m/I_D is increased at CT due to the increase of the saturated velocity. The measurements (symbols) are compared to the analytical expression of the G_m/I_D characteristic which only requires a single parameter, namely the velocity saturation parameter L_{sat} which corresponds to the part of the channel under full velocity saturation [5.76].

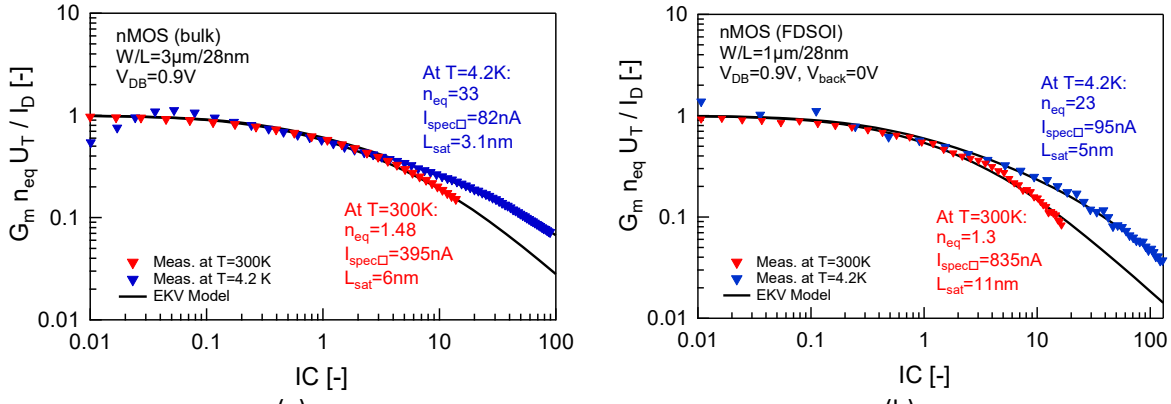


Figure 43: a) G_m/I_D measured on a 28 nm bulk nMOSFET at RT and CT [5.59]. b) G_m/I_D measured on a 28 nm FDSOI nMOSFET at RT and CT [5.61]. The transconductance is normalized to its value in WI accounting for the temperature dependence of $n_{eq}(T)$.

An accurate noise model is also key for the design of analog circuits operating at CT. As shown in [5.77], the input-referred noise can be accurately modeled at a given temperature by the combination of a white noise component (thermal in SI and shot noise in WI), a 1/f noise component and eventually some RTN with Lorentzian spectra [5.77]. However, making the model scalable with temperature and bias remains a challenge.

From an RF perspective, it has been shown in [5.78] that the temperature-dependence of the MOSFET capacitances and gate resistance is weak. The f_T and f_{max} temperature-dependence is therefore mostly related to the G_m temperature-dependence, emphasizing the importance of having an accurate dc model.

CMOS can potentially solve the interconnect bottleneck resulting from the increase of the number of qubits by moving the electronics closer to the qubits. The design of these circuits requires a CM which is accurate at CT. Unfortunately, the CMs available today do not scale correctly with T down to CT. They need to be improved with a particular attention to the increasing V_T and the decreasing SS which saturates below a critical temperature and hence mitigates the current savings. Although some important progress has been made in recent years, having a full CM able to correctly predict the behavior of a circuit operating at CT still requires a substantial effort.

5.6 Superconducting metals (IBM)

In this section, superconducting metals are explored as performance boosters for a cryogenic CMOS technology. These materials will target low-resistance interconnects at both middle and back end of line. The available materials will first be explored, with different cryostat temperature stages in mind. Key material parameters and limitations will be considered. Finally, we will consider the potential performance gain of this type of technology booster.

First, we consider the common interconnect metals. While Cu does not have a superconducting state, Al does, with a T_c of around 1.2 K. Al interconnects could be suitable for sub-Kelvin cryostat stages, but the relatively low T_c entails a low critical magnetic field (0.01 T), meaning that scaling of Al interconnects could be challenging. Indeed, prior experimental results have focused on Al as superconducting TSVs, which could be promising for 3D integration of CMOS chips with qubit chips [5.78]. As a performance booster in advanced CMOS circuits, however, Al is not an attractive choice, also considering that such advanced circuits likely must be placed at the 3 K stage due to their high thermal dissipation.

Other material choices that we will highlight are Nb and TiN, with T_c of 9 and 5.5 K, respectively. We note that In, while having a T_c of 3.4 K, could be a poor choice with the expectation of chip self-heating during operation. We do not consider here more exotic materials such as cuprates.

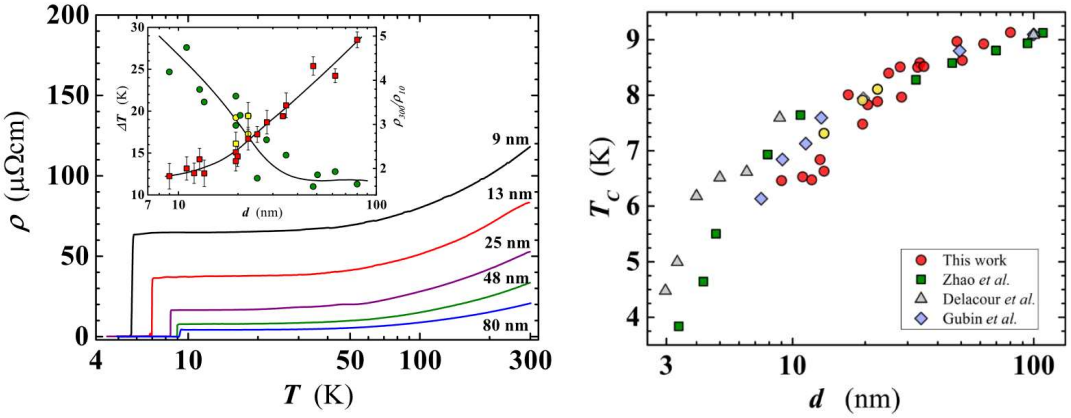


Figure 44: Critical temperatures for Nb films at various thicknesses [5.80].

The figure above shows resistivity of Nb films with various thicknesses [5.78]. Note that these films have a width of around 5 to 10 µm. The observation of superconduction above 4 K for films as thin as 9 nm is promising for the prospect of superconducting interconnects with line widths below 20 nm.

As for TiN, experimental reports show that superconductivity could be challenging to maintain in scaled dimensions. In F Pfuner et al. superconductivity is observed for one of two 18 nm TiN films with a T_c of about 3.5 K [5.78]. This T_c may not be sufficient considering self-heating effects. For lower temperature stages <1 K, the available materials is broader, such as demonstrated in [5.78] showing a stack of In/TiN/Al for BEOL interconnects.

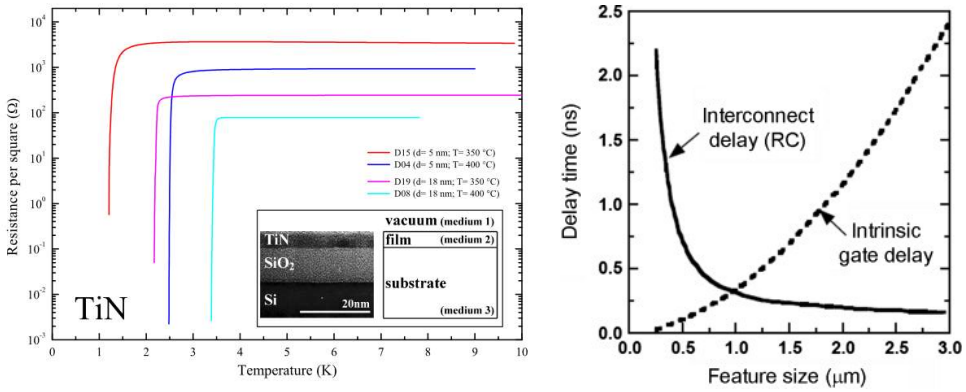


Figure 45: (a) Superconducting characteristics of TiN films with various thicknesses [5.81]. (b) RC delay versus feature size [5.83].

We note the challenge that standard deposition techniques for interconnects may not be available or suitable for high-quality superconductor interconnects such as a Damascene process for Nb. For TiN, deposition will likely be done through ALD, which may not be compatible with standard process schemes.

Predicting the overall performance gain of superconducting interconnects is highly challenging. While predicting the reduction of the RC delay is fairly straightforward, according to internal studies at IBM, a major gain of the technology is that it allows near dissipation-free communication with chip subsystems and modules that are widely physically separated [5.78]. This has huge implications for the overall chip architecture but cannot readily be quantified in terms of a performance metric without significant simulation work.

5.7 Perspectives (LETI)

With higher I_{ON} values due to mobility and saturation velocity increase, steeper subthreshold slope, high reduction of leakage current (I_{off}), reduced thermal noise, better parasitic resistance/capacitance trade-off, higher f_t and f_{max} and lower minimum Noise Figure, the CMOS technology exhibits very promising performances at low temperature for Digital, Analog and RF-mmW functions.

However, these performances are mitigated by kink effects at high voltage for short channels, degradation of hot carrier, variability increase, $|V_T|$ increase for both PMOS and NMOS and increased thermal conductivity of Si, as well as of many materials in the stack.

In comparison to bulk CMOS, FDSOI escapes to most of the previous drawbacks: the kink is totally absent with sufficiently thin Si, it outperforms other CMOS at 4.3K in variability, and the threshold voltage can be tuned with back-gate in order to recover room temperature $|V_T|$, even down to ultra-low temperatures as low as 100mK. Moreover, access resistance decreases more significantly, and maximum gm for short channel exhibits more than 50% increase at V_{DS} = 50mV, leading to higher performances especially at Low Power regime. Nontheless, along with temperature decrease, in long channel FDSOI devices, humps appear in the current, when playing on back-gate, which is a very specific behavior.

Further developments are foreseen on self-heating and thermal diffusion, variability, and reliability which is suspected to decrease due to hot carrier degradation at low temperature.

On back-end, the series resistance of copper traces is strongly reduced by a factor 5 at frequency below skin effect, leading to 5% reduction in inductance as well. The self-resonance frequency of reactive components increases due to the decrease of effective parasitic capacitance to ground. In the future (~7 years in standard process), we expect to benefit of superconducting materials into the back-end, even if proximal heating requires either high Tc superconducting metals or complex thermal diffusion strategies. In shorter term (~3 years), superconducting materials are expected to be a good lever of performance with TSV's, thus heterogeneous integration.

We anticipate a big effort on new physical phenomena understanding and their implementation into compact models, in order to offer a robust and parametric cryogenic design environment (~5 years for the today's observed phenomena in FDSOI). This task also depends on enhancements in low temperature test facilities.

References

[5.1] D.-H. Ju *et al*, "Transient substrate current generation and device degradation in CMOS circuits at 77K", IEDM 1985
[5.2] J.Y.-C. Sun *et al*, "A High Performance Liquid-Nitrogen CMOS SRAM Technology", ESSDERC 1988
[5.3] H. Hanamura *et al*, "Operation of bulk CMOS devices at very low temperatures", IEEE JSSC 1986
[5.4] F. Balestra and G. Ghibaudo, "Brief review of the MOS device physics for low temperature electronics," *Solid. State. Electron.*, vol. 37, no. 12, pp. 1967–1975, 1994.
[5.5] F. Balestra and G. Ghibaudo, Device and Circuit Cryogenic Operation for Low Temperature Electronics. 2001.
[5.6] F. Balestra and G. Ghibaudo, "Physics and performance of nanoscale semiconductor devices at cryogenic temperatures," *Semicond. Sci. Technol.*, vol. 32, no. 2, 2017.
[5.7] K. Das and T. Lehmann, "Effect of deep cryogenic temperature on silicon-on-insulator CMOS mismatch : A circuit designer's perspective," *Cryogenics*, vol. 62, pp. 84–93, 2014.
[5.8] Appaswamy, P. Chakraborty, and J. D. Cressler, "Cryogenic Matching Performance of 90 nm MOSFETs," in *International Semiconductor Device Research Symposium*, 2009.
[5.9] P. A. 'T Hart, J. P. G. van Dijk, M. Babaie, E. Charbon, A. Vladimirescu, and F. Sebastiano, "Characterization and Model Validation of Mismatch in Nanometer CMOS at Cryogenic Temperatures," in *Proceedings of European Solid-State Device Research Conference*, 2018, pp. 246–249.

[5.10] P. A. 'T Hart, M. Babaie, E. Charbon, A. Vladimirescu, and F. Sebastiano, "Characterization and Modeling of Mismatch in Cryo-CMOS," *IEEE J. Electron Devices Soc.*, vol. 8, no. March, pp. 263–273, 2020.
[5.11] Cardoso Paz *et al.*, "Variability Evaluation of 28nm FD-SOI Technology at Cryogenic Temperatures down to 100mK for Quantum Computing," in *Symposium on VLSI Technology, Digest of Technical Papers*, 2020, p. TN2.1
[5.12] Bardin *et al*, "Design and Characterization of a 28-nm Bulk-CMOS Cryogenic Quantum Controller Dissipating Less Than 2 mW at 3 K", JSSC 11/2019
[5.13] R. M. Incandela et al, « Nanometer CMOS characterization and compact modeling at deep-cryogenic temperatures », ESSDERC 2017
[5.14] Beckers, F. Jazaeri, and C. Enz, "Characterization and modeling of 28-nm bulk CMOS technology down to 4.2 K," IEEE JEDS 2018
[5.15] Rosario M.Incandela *et al*, "Characterization and Compact Modeling of Nanometer CMOS Transistors at Deep-Cryogenic Temperatures", IEEE JEDS 2018
[5.16] H. Coskun and J. C. Bardin, "Cryogenic small-signal and noise performance of 32nm SOI CMOS," IEEE MTT-S, 06/2014
[5.17] Beckers, F. Jazaeri, H. Bohuslavskyi, L. Hutin, S. De Franceschi, and C. Enz, "Design-oriented modeling of 28 nm FDSOI CMOS technology down to 4.2 K for quantum computing," in *Joint International EUROSOI Workshop and International Conference on Ultimate Integration on Silicon (EUROSOI-ULIS)*, 2018, pp. 1–4.
[5.18] H. Bohuslavskyi *et al.*, "Cryogenic Subthreshold Swing Saturation in FD-SOI MOSFETs Described With Band Broadening," *IEEE Electron Device Lett.*, vol. 40, no. 5, pp. 784–787, 2019.
[5.19] G. Ghibaudo, M. Aouad, M. Cassé, S. Martinie, T. Poiroux, and F. Balestra, "On the modelling of temperature dependence of subthreshold swing in MOSFETs down to cryogenic temperature," *Solid. State. Electron.*, vol. 170, no. April, p. 107820, 2020.
[5.20] Beckers, F. Jazaeri, and C. Enz, "Revised theoretical limit of the subthreshold swing in field-effect transistors," 2019.
[5.21] W. Ahn, S. H. Shin, C. Jiang, H. Jiang, M. A. Wahab, and M. A. Alam, "Integrated modeling of Self-heating of confined geometry (FinFET, NWFET, and NSHFET) transistors and its implications for the reliability of sub-20 nm modern integrated circuits," *Microelectron. Reliab.*, vol. 81, no. December 2017, pp. 262–273, 2018.
[5.22] Jang *et al.*, "Self-heating on bulk FinFET from 14nm down to 7nm node," in *IEEE International Electron Devices Meeting (IEDM)*, 2015, pp. 289–292.
[5.23] Prasad, "A Review of Self-Heating Effects in Advanced CMOS Technologies," *IEEE Trans. Electron Devices*, vol. 66, no. 11, pp. 4546–4555, 2019.
[5.24] J.-P. Noel et al., "Multi-VT UTBB FDSOI Device Architectures for Low-Power CMOS Circuit," IEEE Trans. Electron Devices, vol. 58, no. 8, pp. 2473–2482, 2011.
[5.25] L. Le Guevel *et al.*, "A 110mK 295 µW 28nm FDSOI CMOS Quantum Integrated Circuit with a 2.8GHz Excitation and nA Current Sensing of an On-Chip Double Quantum Dot," in *IEEE International Solid-State Circuits Conference (ISSCC)*, 2020, pp. 306–308.
[5.26] S. J. Pauka et al., "A Cryogenic Interface for Controlling Many Qubits." arXiv:1912.01299v1, 2019.
[5.27] Simoen, B. Dierickx, L. Warmerdam, J. Vermeiren, and C. Claeys, "Freeze-Out Effects on NMOS Transistor Characteristics at 4.2 K," IEEE Trans. Electron Devices, vol. 36, no. 6, pp. 1155–1161, 1989.
[5.28] H. J. Park, M. Bawedin, K. Sasaki, J. A. Martino, and S. Cristoloveanu, "Is there a kink effect in FDSOI MOSFETs?," in Joint EUROSOI-ULIS conference, 2017, pp. 212–215
[5.29] H. L. Chiang *et al*, "Cold CMOS as a Power-Performance-Reliability Booster for Advanced FinFETs", TSMC, VLSI2020
[5.30] Grill *et al.*, "Reliability and Variability of Advanced CMOS Devices at Cryogenic Temperatures," in *IEEE International Reliability Physics Symposium Proceedings*, 2020, vol. 2020-April.
[5.31] Cardoso Paz *et al*, "Electrostatics and channel coupling on 28nm FDSOI for cryogenic applications", EUROSOI 2020

[5.32] Jazaeri, A. Beckers, A. Tajalli, and J. Sallese, “A Review on Quantum Computing : From Qubits to Front-end Electronics and Cryogenic MOSFET Physics,” in *International Conference Mixed Design of Integrated Circuits and Systems (MIXDES)*, 2019, pp. 15–25.
[5.33] Beckers, F. Jazaeri, H. Bohuslavskyi, L. Hutin, S. De Franceschi, and C. Enz, “Characterization and modeling of 28-nm FDSOI CMOS technology down to cryogenic temperatures,” *Solid State Electron.*, vol. 159, no. 688539, pp. 106–115, 2019
[5.34] Asheghi *et al*, “Temperature-Dependent Thermal Conductivity of Single-Crystal Silicon Layers in SOI Substrates”, Journal of Heat Transfer, 02/1998
[5.35] Triantopoulos *et al*, “Self-Heating Effect in FDSOI Transistors Down to Cryogenic Operation at 4.2 K”, Electron Devices, 08/2019
[5.36] Patra, M. Mehrpoo, A. Ruffino, F. Sebastiano, E. Charbon, and M. Babaie, “Characterization and Analysis of On-Chip Microwave Passive Components at Cryogenic Temperatures,” IEEE J. Electron Devices Soc., vol. 8, pp. 448–456, Apr. 2020
[5.37] Heorhii BOHUSLAVSKYI, “Cryogenic electronics and quantum dots on silicon-on-insulator for quantum computing”, PhD report, 12/2018 : https://hal.archives-ouvertes.fr/tel-02183484/
[5.38] Cardoso Paz *et al*, "Electrostatics and channel coupling on 28nm FDSOI for cryogenic applications", EUROSOI2020 03/2020
[5.39] Cardoso Paz *et al*, "Variability Evaluation of 28nm FD-SOI Technology at Cryogenic Temperatures down to 100mK for Quantum Computing", VLSI 2020
[5.40] M. Cassé *et al.*, “Evidence of 2D intersubband scattering in thin film fully depleted silicon-on-insulator transistors operating at 4.2 K,” *Appl. Phys. Lett.*, vol. 116, p. 243502, 2020.
[5.41] M. Cassé *et al,* "Cryogenic Operation of Thin Film FDSOI NMOS Transistors: Effect of Back Bias on Drain Current and Transconductance", TED ESSDERC 2020
[5.42] Cardoso Paz *et al*, "Integrated Variability Measurements of 28 nm FDSOI MOSFETs down to 4.2 K for Cryogenic CMOS Applications", ICMTS 2020
[5.43] Bonen *et al*, “Cryogenic Characterization of 22-nm FDSOI CMOS Technology for Quantum Computing ICs”, IEEE EDL 2019
[5.44] Cardoso Paz *et al*, “Performance and Low Frequency Noise of 22nm FDSOI down to 4.2K for Cryogenic Applications”, 2020
[5.45] Schleeh *et al*, “Phonon black-body radiation limit for heat dissipation in electronics”, Nature Material 2015
[5.46] Makovejev *et al*, “Time and frequency domain characterization of transistor self-heating”, IEEE TED 2013
[5.47] L. Nyssens *et al.*, “Self-Heating in FDSOI UTBB MOSFETs at Cryogenic Temperatures and its Effect on Analog Figures of Merit,” *IEEE J. Electron Devices Soc.*, 2020
[5.48] Siligaris *et al*, “High-Frequency and Noise Performances of 65-nm MOSFET at Liquid Nitrogen Temperature”, IEEE TED 2006
[5.49] Hong *et al*, “Low-Temperature Performance of Nanoscale MOSFET for Deep-Space RF Applications”, IEEE EDL 2008
[5.50] Esfeh *et al*, “28 FDSOI RF Figures of Merits and Parasitic Elements at Cryogenic Temperature”, IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S) 2018
[5.51] Esfeh *et al*, “28-nm FDSOI nMOSFET RF Figures of Merits and Parasitic Elements Extraction at Cryogenic Temperature Down to 77 K”, IEEE JEDS 2019
[5.52] L. Nyssens *et al.*, “28-nm FD-SOI CMOS RF Figures of Merit down to 4.2 K,” *IEEE J. Electron Devices Soc.*, vol. 8, no. June, pp. 646–654, 2020.
[5.53] P. G. Mautry and J. Trager, “Investigation of Self-Heating in VLSI and ULSI MOSFETs,” in *IEEE International Conference on Microelectronic Test Structures (ICMTS)*, 1990, p. 221.
[5.54] Stellari, K. A. Jenkins, A. J. Weger, B. P. Linder, and P. Song, “Self-Heating Measurement of 14-nm FinFET SOI Transistors Using 2-D Time-Resolved Emission,” *IEEE Trans. Electron Devices*, vol. 63, no. 5, pp. 2016–2022, 2016.
[5.55] Cahill *et al.*, “Nanoscale thermal transport. II. 2003-2012,” *Appl. Phys. Rev.*, vol. 1, no. 1, 2014.

[5.56] T. Takahashi, T. Matsuki, T. Shinada, Y. Inoue, and K. Uchida, “Direct evaluation of self-heating effects in bulk and ultra-thin BOX SOI MOSFETs using four-terminal gate resistance technique,” *IEEE J. Electron Devices Soc.*, vol. 4, no. 5, pp. 365–373, 2016

[5.57] E. Charbon, "Cryo-CMOS Electronics for Quantum Computing Applications," ESSCIRC 2019 - IEEE 45th European Solid State Circuits Conference (ESSCIRC), pp. 1-6, Sept. 2019.

[5.58] E. Charbon, F. Sebastiano, A. Vladimirescu, H. Homulle, S. Visser, L. Song, and R. M. Incandela, "Cryo-CMOS for quantum computing," 2016 IEEE International Electron Devices Meeting (IEDM), pp. 13.5.1-13.5.4, Dec. 2016.

[5.59] A. Beckers, F. Jazaeri, and C. Enz, "Characterization and Modeling of 28-nm Bulk CMOS Technology Down to 4.2 K," IEEE Journal of the Electron Devices Society, vol. 6, pp. 1007-1018, March 2018.

[5.60] R. M. Incandela, L. Song, H. Homulle, E. Charbon, A. Vladimirescu, and F. Sebastiano, "Characterization and Compact Modeling of Nanometer CMOS Transistors at Deep-Cryogenic Temperatures," IEEE Journal of the Electron Devices Society, vol. 6, pp. 996-1006, April 2018.

[5.61] A. Beckers, F. Jazaeri, H. Bohuslavskyi, L. Hutin, S. D. Franceschi, and C. Enz, "Characterization and Modeling of 28-nm FDSOI CMOS Technology down to Cryogenic Temperatures," Solid-State Electronics, vol. 159, pp. 106-115, Sept. 2019.

[5.62] S. Bonen, U. Alakusu, Y. Duan, M. J. Gong, M. S. Dadash, L. Lucci, D. R. Daughton, G. C. Adam, S. Iordănescu, M. Păşteanu, I. Giangu, H. Jia, L. E. Gutierrez, W. T. Chen, N. Messaoudi, D. Harame, A. Müller, R. R. Mansour, P. Asbeck, and S. P. Voinigescu, "Cryogenic Characterization of 22-nm FDSOI CMOS Technology for Quantum Computing ICs," IEEE Electron Device Letters, vol. 40, no. 1, pp. 127-130, Jan. 2019.

[5.63] J. C. Bardin, E. Jeffrey, E. Lucero, T. Huang, S. Das, D. T. Sank, O. Naaman, A. E. Megrant, R. Barends, T. White, M. Giustina, K. J. Satzinger, K. Arya, P. Roushan, B. Chiaro, J. Kelly, Z. Chen, B. Burkett, Y. Chen, A. Dunsworth, A. Fowler, B. Foxen, C. Gidney, R. Graff, P. Klimov, J. Mutus, M. J. McEwen, M. Neeley, C. J. Neill, C. Quintana, A. Vainsencher, H. Neven, and J. Martinis, "Design and Characterization of a 28-nm Bulk-CMOS Cryogenic Quantum Controller Dissipating Less Than 2 mW at 3 K," IEEE Journal of Solid-State Circuits, vol. 54, no. 11, pp. 3043-3060, Nov. 2019.

[5.64] B. Patra, R. M. Incandela, J. P. G. v. Dijk, H. A. R. Homulle, L. Song, M. Shahmohammadi, R. B. Staszewski, A. Vladimirescu, M. Babaie, F. Sebastiano, and E. Charbon, "Cryo-CMOS Circuits and Systems for Quantum Computing Applications," IEEE Journal of Solid-State Circuits, vol. 53, no. 1, pp. 309-321, Jan. 2018.

[5.65] A. Ruffino, Y. Peng, F. Sebastiano, M. Babaie, and E. Charbon, "A Wideband Low-Power Cryogenic CMOS Circulator for Quantum Applications," IEEE Journal of Solid-State Circuits, vol. 55, no. 5, pp. 1224-1238, May 2020.

[5.66] A. Beckers, F. Jazaeri, and C. Enz, "Cryogenic MOS Transistor Model," IEEE Transactions on Electron Devices, vol. 65, no. 9, pp. 3617-3625, Sept. 2018.

[5.67] F. Balestra and G. Ghibaudo, "Physics and Performance of Nanoscale Semiconductor Devices at Cryogenic Temperatures," Semicond. Sci. Technol., vol. 32, no. 2, pp. 023002, Jan. 2017.

[5.68] G. Ghibaudo and F. Balestra, "Device physics and electrical performance of bulk Silicon MOSFETs," in Device and Circuit Cryogenic Operation for Low Temperature Electronics, F. Balestra and G. Ghibaudo, Eds.: Springer, 2001, pp. 3-35.

[5.69] A. Pirovano, A. L. Lacaita, A. Pacelli, A. Benvenuti, "Novel low-temperature {C}-{V} technique for {MOS} doping profile determination near the Si/SiO2 interface," IEEE Transactions on Electron Devices, vol. 48, no. 4, pp. 750-757, Apr. 2001.

[5.70] A. Beckers, F. Jazaeri, A. Grill, S. Narasimhamoorthy, B. Parvais, and C. Enz, "Physical Model of Low-Temperature to Cryogenic Threshold Voltage in MOSFETs," IEEE Journal of the Electron Devices Society, pp. 1-1, April 2020.

[5.71] N. C. Dao, A. El Kass, M. R. Azghadi, C. T. Jin,J. Scott, P. H. W. Leong, "An enhanced MOSFET threshold voltage model for the 6-300K temperature range," Microelectronics Reliability, vol. 69, pp. 36—39, Feb. 2017.

[5.72] H. Bohuslavskyi, A. G. M. Jansen, S. Barraud, V. Barral, M. Cassé, L. L. Guevel, X. Jehl, L. Hutin, B. Bertrand, G. Billiot, G. Pillonnet, F. Arnaud, P. Galy, S. D. Franceschi, M. Vinet, and M. Sanquer,

"Cryogenic Subthreshold Swing Saturation in FD-SOI MOSFETs Described With Band Broadening," IEEE Electron Device Letters, vol. 40, no. 5, pp. 784-787, May 2019.

[5.73] A. Beckers, F. Jazaeri, and C. Enz, "Theoretical Limit of Low Temperature Subthreshold Swing in Field-Effect Transistors," IEEE Electron Device Letters, vol. 41, no. 2, pp. 276-279, Feb. 2020.

[5.74] G. Ghibaudo, M. Aouad, M. Casse, S. Martinie, T. Poiroux, and F. Balestra, "On the Modelling of Temperature Dependence of Subthreshold Swing in MOSFETs down to Cryogenic Temperature," Solid-State Electronics, vol. 170, pp. 107820, April 2020.

[5.75] S. K. Tewksbury, "Attojoule MOSFET logic devices using low voltage swings and low temperature," Solid-State Electronics, vol. 28, no. 3, pp. 255-276, March 1985.

[5.76] C. Enz, F. Chicco, and A. Pezzotta, "Nanoscale MOSFET Modeling: Part 1: The Simplified EKV Model for the Design of Low-Power Analog Circuits," IEEE Solid-State Circuits Magazine, vol. 9, no. 3, pp. 26-35, Summer 2017.

[5.77] H. Achour, R. Talmat, B. Cretu, J. M. Routoure, A. Benfdila, R. Carin, N. Collaert, E. Simoen, A. Mercha, and C. Claey, "DC and low frequency noise performances of SOI p-FinFETs at very low temperature," Solid-State Electronics, vol. 90, pp. 160-165, Jan. 2013.

[5.78] B. K. Esfeh, V. Kilchytska, N. Planes, M. Haond, D. Flandre, and J. Raskin, "28-nm FDSOI nMOSFET RF Figures of Merits and Parasitic Elements Extraction at Cryogenic Temperature Down to 77 K," IEEE Journal of the Electron Devices Society, vol. 7, pp. 810-816, Aug. 2019.

[5.79] J. A. Alfaro-Barrantes *et al.*, "Superconducting High-Aspect Ratio Through-Silicon Vias With DC-Sputtered Al for Quantum 3D Integration," in *IEEE Electron Device Letters*, vol. 41, no. 7, pp. 1114-1117, July 2020, doi: 10.1109/LED.2020.2994862.

[5.80] Pinto, N., Rezvani, S.J., Perali, A. *et al.* Dimensional crossover and incipient quantum size effects in superconducting niobium nanofilms. *Sci Rep* **8,** 4710 (2018). https://doi.org/10.1038/s41598-018-22983-6.

[5.81] F Pfuner *et al* 2009 *New J. Phys.* **11** 113017, doi.org/10.1088/1367-2630/11/11/113017

[5.82] B Foxen *et al* 2017 *Quantum Sci. Technol.* **3** 014005, doi.org/10.1088/2058-9565/aa94fc

[5.83] Jaydeep Sarkar, in *Sputtering Materials for VLSI and Thin Film Devices*, Elsevier 2014.

6 Silicon-based cryogenic building blocks (LETI)

Previous section made emphases on the challenges that are still on the table in order to accurately characterize and model the silicon-based technology at low temperature, and to implement new technology add-ons like superconducting materials. This step is mandatory in the perspective of design of complex cryogenic building blocks that might reach the expectations in term of performance and power consumption, to support scaling-up of quantum computing.

Here, we introduce the ongoing works that are done in the domain of circuit design. Challenging issues are encountered in developing electronic functions which should require advancements on latter aspects as well as on test facilities. A global system architecture for control and readout of many qubits that is distributed along temperature stages, and involves hardware/software implementation, is also a complex issue to overcome. Nevertheless, we will see that ideas for circuit design are emerging and brilliant results are already achieved, sustaining promising developments.

6.1 Toward a large scale cryo-CMOS approach (LETI)

In Nature 2018 [6.23], Watson *et al* published an experience around a programmable two-qubits quantum processor. The setup of the experience was schematized like in figure below. While the quantum processor is located at the 20mK stage of the cryostat, the whole control/readout system relies on an extensive utilization of Room Temperature instruments.

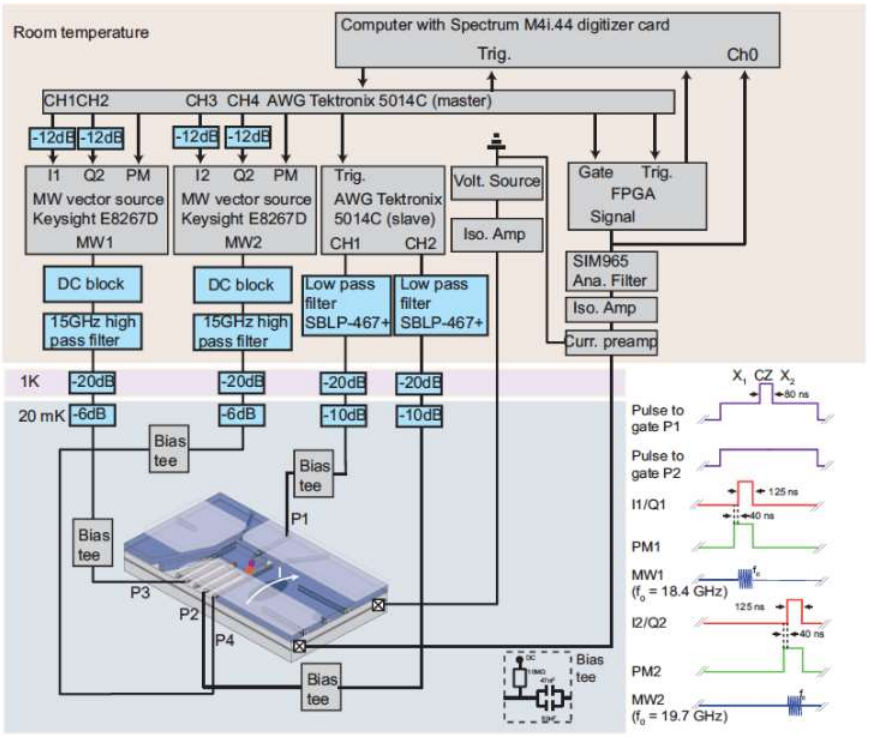


Figure 6-1: Schematic of measurement setup for a programmable two-qubit quantum processor[6.23].

In Microprocessors and Microsystems 2019, Van Dijk *et al,* from QuTech-Delft and Intel, disserted on the limitations of such implementation when scaling up, **even if Room Temperature high-speed FPGAs allows a more cost-effective mid-term implementation where a handful of qubits are driven** [6.24].

According to the consortium's knowledge, the first publication reporting a potential scalable approach for quantum computing with cryo-CMOS is the one from Fabio Sebastiano, Edoardo Charbon *et al*, from QuTech-Delft and Intel, in IEDM 2016 [6.25] and ISSCC 2017 [6.26].

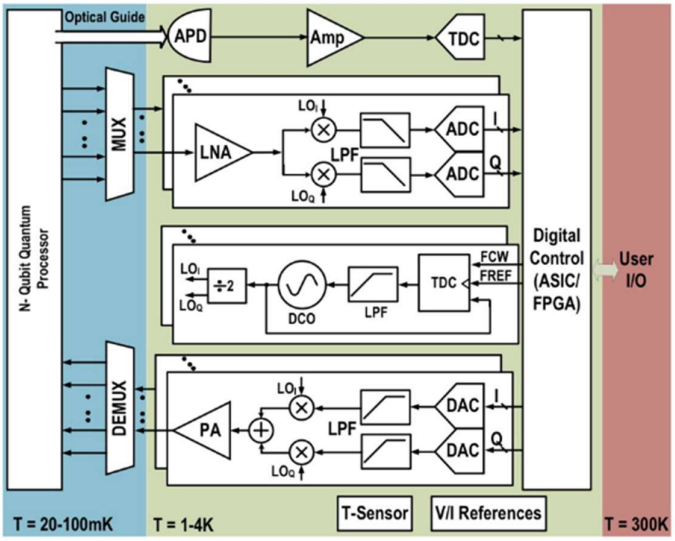


Figure 6-2: A scalable cryo-CMOS architecture for quantum computing.

This representation shows a generic system for a scalable qubit array where **the majority of the components are operating in the range of 1-4K**. The digital controller is thought as a local error corrector and a readout analyzer of the state of the qubit, but also to drive algorithm execution by digital signal synthesis. (De)multiplexers are used to reduce the number of interconnections to and from the cold areas of the circuit, so as to ensure large number of channels while minimizing thermal flux; these components might operate in TDMA, FDMA and SDMA mode. This system is supposed to be implemented in CMOS and on CMOS-compatible substrates. [6.25]

Nonetheless, as reported by Boter *et al* from QuTech-Delft with Intel in IEDM 2019 [6.27], "**one of the main issues in common with all nanoelectronic qubits is that current implementations require at least one external control line for every qubit. The small pitch of quantum dots permits extremely dense qubit arrays but aggravates the interconnect challenges.**"

They are driven in their thinking by Franke *et al* from the same organizations, saying in Microprocessors and Microsystems 2019 [6.28]: "In classical semiconductor technology, scaling was made possible by the invention of the integrated circuit, which allowed to interconnect large numbers of components without having to solder to each and every one of them. **Similarly, we expect that the scaling of interconnections and control lines with the number of qubits will be a central bottleneck in creating large-scale quantum technology**".

The today's tentative architectures (like the one from Charbon *et al*) are trying to manage a few qubits, with the perspective of many qubits at the corner of the eye. But they know that if we want to make scalability a reality, a breakthrough is necessary to circumvent the bottleneck of interconnects.

The option that is presented by Boter *et al* relies on a *sparse* qubit array, integrating a part of the electronic in something like a pixel of 12μm wide. **Qubits remain at the vertices of the pixel while idle and are supposed to shuttle to the operation regions between the vertices in order to perform single- and two-qubit**

operations as well as readout and initialization. Micromagnets are also supposed to be integrated in the qubit operations regions.

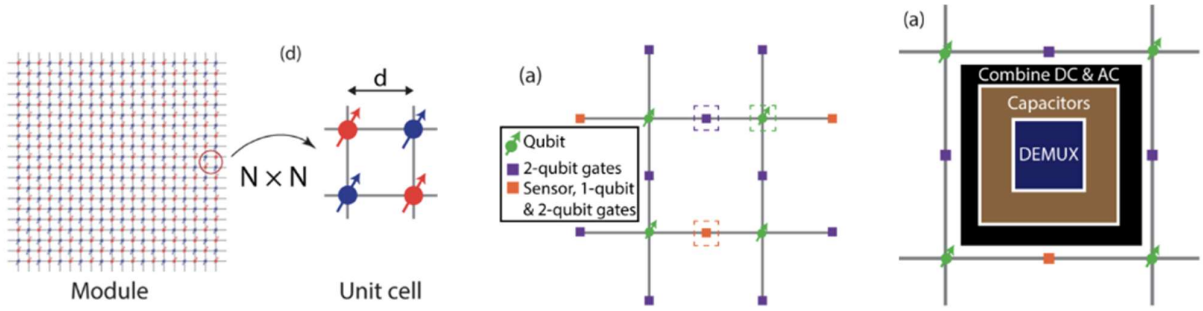


Figure 6-3: Pieces of the sparse qubit array representation. Please refer to [6.27] for more details

By the way, let's mention the recent PhD report from Baptiste Jadot in 2020, on "Coherent long-range transport of entangled electron spins" [3.7]. Integrated micromagnet is an active field of research as well [6.29].

On another aspect of the quantum computing scaling implementation, D. J. Reilly from Microsoft Quantum Sidney said in 2019 [6.30]: "Putting aside the obvious convenience and advantages of a tightly-integrated system over a machine with components that span meters, from the perspective of signal propagation, a distributed system presents at least two additional challenges. Firstly, at the radio or microwave frequencies used for quantum control signals (MHz - GHz), **the system size exceeds or is comparable to the wavelength. In this distributed regime, since currents and voltages are functions of space as well as time, it is necessary to work with controlled impedances**, bringing additional constraints in power dissipation or increased footprint of the interconnects and matching circuits relative to systems in which the impedance is a free-parameter. **The use of 50Ω transmission lines, for instance, impacts the power dissipated by circuits that must drive a 50Ω impedance.**"

Reilly talks also about **phase shifts in signals between different lines and interference between branches, the notion of synchronicity between signals on 1000s of different transmission lines, or the time-of-flight latency, whereas qubit manipulation underlies accurate phase and amplitude of microwave signals**. Many of these challenges to scaling appear to be **largely addressed by shrinking the distributed control system** and implementing it as a series of tightly packaged integrated circuits or components in close proximity to the qubits at cryogenic temperatures.

Therefore, the main issue becomes: **how to develop such control and readout circuits in the constraint of ultra-low power consumption and ultra-high sensitivity?** Pauka *et al* from Microsoft Quantum Sidney published a benchmark of system implementation, in 2019 [6.31].

The chip, enabling charge-manipulation process (not spin-qubit control) handles at 100mK the IO demultiplexing of DC and commuted signals in MHz range, was designed on 28nm FDSOI and gets advantage of back-gate bias to offset threshold voltage. Both high (1.8V) and low voltage (1V) devices, **individual back-gate control of n-type and p-type transistors or of entire circuit blocks**, are also properties that were appreciated by the designers. Authors mention the ability to control 32 quantum dots, even if the experiment was done with only 2 quantum dots (whose one for verifications).

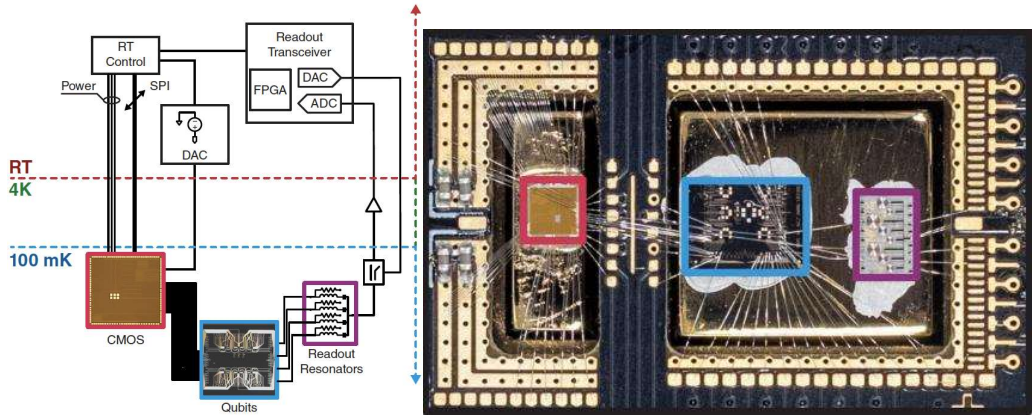


Figure 6-4: ***At left*** *: High-level schematic of the quantum-classical interface that leverages ultra-low power cryo-CMOS circuits for control signal generation. The control MUX block handles IO management, reducing the need for many high-bandwidth cables to higher temperature stages, and enables dense IO via chip-stacking techniques. For readout, signals are generated using direct digital synthesis (DDS), amplified at low temperatures and acquired by a high-speed ADC, operating at room temperature. A micro-controller, operating at 4K (not shown here but mentioned in [6.30]) coordinates triggering, clocking, and command buffering.* ***At right****: Photograph showing the cryo-CMOS chip (red box), GaAs qubit test chip (blue box), and resonator chip (purple box). Each chip is anchored to a gold-plated copper thermalization pillar, with a separate pillar used for the CMOS chip. This arrangement is intended to reduce the direct heat flow from the CMOS circuits to the qubit chip. [6.31]*

An extrapolation of power budget is shown, and suggests this circuit would be able to control 1000 qubits at 100mK. But, for the case of transmon or spin-qubits, the GHz range signal generation is not included, neither read-out power budget. Nonetheless, it is a very good demonstration for high volume polarization management.

According to Edoardo Charbon, given that cryo-CMOS circuits and systems will operate at temperatures close, ideally equal, to those of the qubits, **the main limitation is the power dissipation of classical circuits**, which have to be budgeted within the limits of thermal absorption by the refrigeration system. Up until a few years ago, this challenge was considered unsurmountable with CMOS devices [6.1]. In the next subsections, we present various building blocks that represent milestones for the achievement of cryo-CMOS integration of quantum computers.

6.2 Building blocks for control (LETI)

6.2.1. On control specifications (LETI)

Like seen in the short introduction on qubits technologies, physical qubit control is obtained by rotations along axes in the Bloch sphere representation. This is achieved by means of around **-100dBm microwave signals with frequencies ranging from 1GHz to 25-40GHz** (the Larmor frequency is used, and depends on qubits implementation – see section 4), and **amplitude modulated as a rectangular or raised cosine pulse of hundreds of nanoseconds**, so as to achieve the necessary spectral purity that ensures qubit fidelity preservation and operation accuracy [6.1].

Some parts of the control circuit operate at 4K, a stage of the dilution fridge where several Watts of thermal absorption is possible. Whereas some other parts of this control are located close to the qubit, in the 100mK (for today's Silicon spin-qubits) or 20mK (for today's transmons) stage. Here, the power dissipation must be

kept under, respectively, a few mW, or a few µW. This means that only a few transistors with very low voltage and current can be implemented in these stages. The electronic functions are then, mainly, passive functions or switches.

The most developed publication on control specifications was reported by van Dijk *et al* in 2019 [6.3], and partially revealed in [6.2]. Authors proposed a method for analyzing the effects of signal non idealities. Specifications for the microwave carrier, the microwave envelope, the idle mode, and some requirements for qubit frequency multiplexing, are declined, and the authors explore the case study for a single-qubit and a two-qubit operation.

In the table below, a single spin-qubit is controlled with a Larmor frequency (the microwave frequency of operation) fixed at 10GHz. A rectangular pulse is used, in the presence of noise and signals used to control other qubits in the context of a frequency division multiple-access (FDMA) scheme. The frequency spacing is set to 1GHz. The values provided for the microwave amplitude assume a qubit plane based on EDSR [see section 4] where an amplitude of 2mV at the gate is required for a Rabi frequency of 1MHz. **All specifications are valid at the gate so that wiring attenuation and altering might need to be factored in to refer the specifications back to the electronics**.

	Value	Infidelity contribution to an operation	Infidelity contribution to idling	Comment
Frequency				
nominal	10 GHz	0.64×10^{-9}		RWA when driving a qubit
spacing	1 GHz		1×10^{-6}	FDMA leakage with rectangular envelopes
inaccuracy	11 kHz	125×10^{-6}	308×10^{-6}	
oscillator noise	11 kHz_{rms}	125×10^{-6}	308×10^{-6}	ENBW = 2.5 MHz, $\mathcal{L}$(1 MHz) = -106 dBc/Hz
nuclear spin noise	1.9 kHz_{rms}	3.6×10^{-6}	8.9×10^{-6}	From [33], $T_2^* = 120\,\mu s$
wideband noise	12 μV_{rms}	125×10^{-6}		ENBW = 2.9 MHz, $S_{add} = 7.1\,nV/\sqrt{Hz}$
Phase				
inaccuracy	0.64°	125×10^{-6}	31×10^{-6}	FDMA Z-corrections limit the no operation
Amplitude				
nominal	2 mV			Full-scale: 4 mV, RMS: 1.4 mV_{rms}
inaccuracy	14 µV	125×10^{-6}		
noise	14 μV_{rms}	125×10^{-6}		ENBW = 1.0 MHz, PSD = $14\,nV/\sqrt{Hz}$, SNR = -40 dB
off-spur	19 µV		217×10^{-6}	-41 dBc
off-noise	10 μV_{rms}		125×10^{-6}	ENBW = 2.0 MHz, PSD = $7.1\,nV/\sqrt{Hz}$
Duration				
nominal	500 ns			
inaccuracy	3.6 ns	125×10^{-6}		
noise	3.6 ns_{rms}	125×10^{-6}		
		$F_{X,Y} = 99.9\,\%$	$F_I = 99.9\,\%$	

Figure 6-5: Example specifications for the control electronics of a single qubit in the context of FDMA scheme. The PSD values provided as a comment assume a white spectrum for the amplitude and frequency noise (i.e. -20 dB/dec for the phase noise).

Following these specifications, the authors said the corresponding microwave signal (envelope shape, amplitude and duration) could be generated by an AWG with a sample rate of at least 150 MS/s, such that the sample time is less than 6.7 ns, resulting in a maximum inaccuracy of 3.3 ns. Furthermore, the AWG should have a resolution of 8 bits, such that at a full-scale swing of 4 mV, the quantization step is sufficiently low. To meet the noise requirement and the specifications on the residual driving when not operating the qubit ('off-spur' in the table below), an effective number of bits (ENOB) of 6.5 bits is required.

The LO used for the up-conversion requires a frequency resolution of around 20 kHz (for the inaccuracy). Assuming a -20 dB/dec slope of the phase noise, the single-side band phase noise at 1MHz from the carrier needs to be below -106 dBc/Hz. Furthermore, the LO's phase inaccuracy needs to be below 0.64°.

So, a controller that would be developed with cryo-CMOS in the context of such a theoretical processing, should meet the same range of specifications. This study demonstrates the importance of qubit-level simulations to extrapolate some minimal requirements expected from the electronics. In order to conduct these simulations, J. van Dijk, *et al.* have developed their own toolset for the co-design and co-optimization of electronic/quantum systems. It comprises a SPICE simulator enhanced with a Verilog-A model based on a Hamiltonian solver emulating the quantum behavior of single-electron spin qubit [6.32].

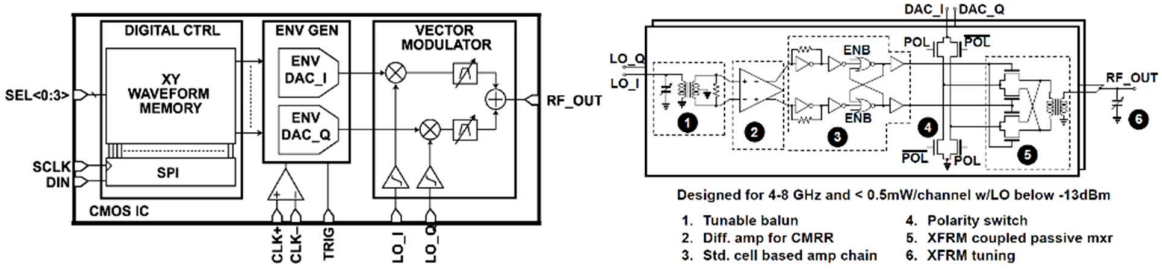
6.2.2. Building blocks (LETI)

Building blocks designed on **CMOS and dedicated for control** started to be reported with the ISSCC 2017 conference by Charbon *et al*. [6.4]. Up to now, about 20 papers were referenced on this topic specifically. Most of them are from a joint research with **QuTech-Delft, EPFL and Intel**, and supervised by Fabio Sebastiano, Edoardo Charbon, and Stefano Pellerano [6.1] to [6.14]. Another group is from **Google, with University of Massachusetts and University of California**. Supervision is mainly ensured by Joseph C. Bardin [6.15][6.16]. And a third one is from **Microsoft and the University of Sydney**, supervised by D.J. Reilly and S.J. Pauka [6.17] to [6.19]. Few other publications come from University of Science and Technology of China [6.20] or CEA [6.21]. Here, we focus on the main contributions.

In the expectation of addressing 1000's of Qubits with a power budget intermediate milestone around 1mW/Qubit , Bardin *et al.* reported in ISSCC 2019 [6.15] and JSSC 2019 [6.16] a **28nm BULK CMOS** mixed Digital/RF pulse generator allowing a single Qubit XY control (Z rotation is obtained by combining XY), working at 3K with 2mW.

The range of operation (4-8GHz microwave pulse) corresponds to the typical transition frequency of transmon. The pulse shaping is digitally generated and stored in local memory to drive a vector modulator. On-chip memory greatly reduces the number of I/Os, thereby resulting in significant savings in power and system complexity. The clock port of the DACs and the LO ports of the vector modulator are buffered (−20dBm at 2 GHz and −10dBm at 5.6 GHz) to minimize the external RF power required to drive the chip, **with the small penalty of increased power budget**.

This architecture has been chosen as a tradeoff between power consumption, performance, and robustness to the inherent uncertainty associated with designing for operation at 3 K, where foundry design models are currently unavailable, said Bardin *et al*.



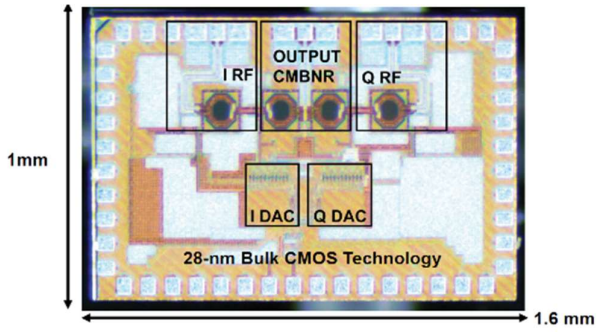


Figure 6-6: Block diagram of the pulse shape generator, and architecture of the vector modulator. Die micrograph.

Performance comparison with Room temperature instrumentation is summarized in the table below.

	Standard	Cryo CMOS
Form Factor	Rack Mount	IC
Physical Temp	300 K	**3 K**
Update Rate	1 Gsps	1 Gsps
Instruction Set	N/A	**4-bit**
Dig. Data Rate	28 Gbps	**< 0.5 Gbps**
Measured T_1	18.3 μs	17.8 μs
$\|2\rangle$ population	Negligible	Negligible (T_G>15ns)
π-Rabi P{|1⟩}	~95%	~95%
3 Gate RMS Err	2.5%	11.7% (uncalibrated)
Total AC+DC PWR	> 1W	**< 2mW**

Figure 6-7: Comparison of CMOS IC @ f_{CLK} = 1 GHz To a state-of-the-art quantum control system.

The authors had concluded: "**While the demonstration of a cryogenic quantum control interface dissipating <2mW is an important step towards the development of a scalable quantum control and measurement system, much research is still required to implement such a system. Future work could focus on calibration algorithms or the development of ICs for control of multiple qubits, qubit readout, or other related applications.**" In 2019, Bardin *et al.* expect that a reasonable long-term limit for the dissipation of an XY control electronics should be 250 µW/qubit.

"Horse-Ridge" is probably the most mature circuit that was published about qubits control at few K. QuTech introduced their research during ISSCC 2020 [6.13]. The team has developed a low temperature multiplexed signal generator. The **22nmFinFET** circuit is able to drive 4x32 TX signals at 3K, and covering a set of critical specifications for spin-qubit and transmon control. Even if, according to the authors, this technology is not optimized and requires adjustments for cryogenic applications, and even if no mature models are available, they demonstrate the ability for such a circuit to operate low temperature. However, **the analog part dissipates 1.7 mW/qubit when low-frequency band only is used, and the digital part dissipates 330 mW (for 32 channels). So this circuit is located at the 3K stage where a power budget of ~1W is available, about 1meter from the qubits 20mK stage**.

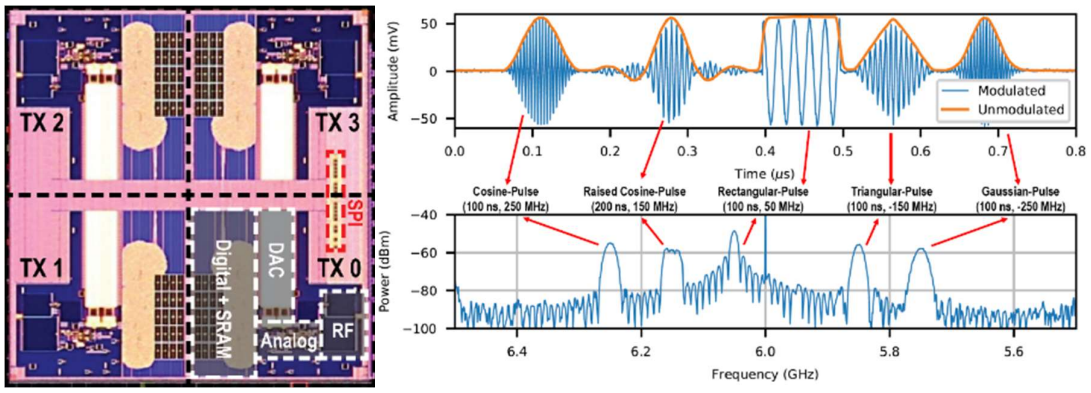


Figure 6-8: 4x4mm 22nmFinFET = 4x32TX, and time/frequency domain representation of pulse shaping that can be produced by this integrated digital AWG.

The mixed digital/RF circuit generates highly accurate signal, required for Qubit control:

- $f_{carrier}$~2-20GHz, t_{pulse}~20-200ns
- SFDR > 44dB on frequency range of operation
- SNR > 44dB in ~25MHz BW
- Programmable pulse shaping
- Phase tracking and correction

The architecture of this circuit is schematically represented in the figure below (1x32TX).

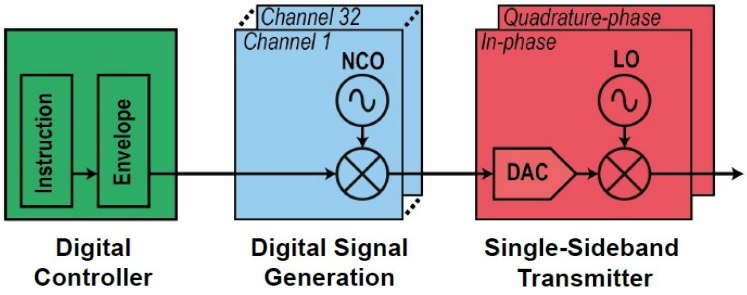


Figure 6-9: bloc-schematic of Horse-Ridge circuit .

In each of the 4 identical parts, a digital envelope generator, producing an on-chip stored data of 18Gbps, feeds a 32 channels digital IQ modulator. Thanks to on-chip memory, no high-speed connection from room temperature is required during quantum algorithm execution. For coherent control of a 32 qubits tile, 32 numerically controlled oscillators (NCO) with phase tracking and frequency accuracy ~0.025 ppm are used. Phase and amplitude is tuned by the digital data to reach a theoretical qubit fidelity of 99,99%: phase imbalance < 0.7° and gain imbalance < 0.1 dB leads to Spurious-Free Dynamic Range > 44 dB. The signal is then converted in the analog world, filtered, equalized and up-converted. To cover a 2-to-20GHz output band, the RF path is split into 2 parts: 2-to-15GHz and 15-to-20GHz.

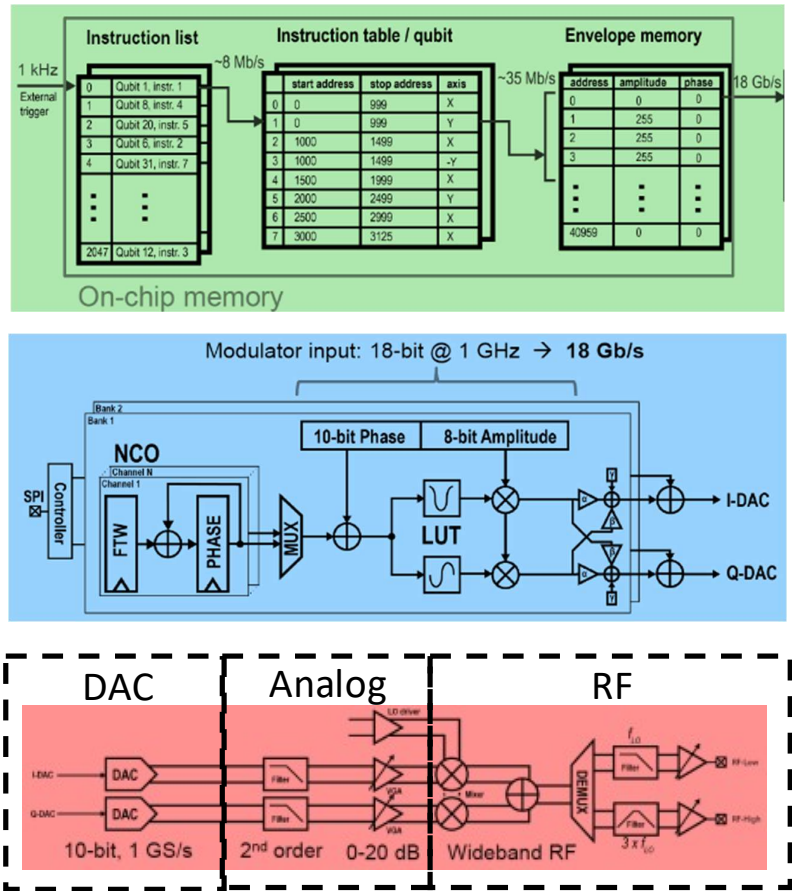


Figure 6-10: bloc-schematic of Horse-Ridge circuit (more details)

The authors report an experiment of typical XZ rotation and Rabi measurement on one qubit using the low-RF band, and got a comparable accuracy as with room temperature instrumentation. This component is clearly a big milestone for the quantum computer community. **However, since 4x32 channels are available, one would like to read a report on multi-qubit analysis**. Therefore, the frequency multiplexed and synchronized control is not demonstrated. **Should we speculate on the challenge that multiplexed control and readout represents?** Moreover, it's regrettable that Local Oscillator for RF up-conversion remains off-chip, whereas this signal generator is critical for spectral purity, like we will see in the next paragraph.

Another team from QuTech presented what could be the missing VCO, as a standalone component, in ISSCC 2020 [6.12]. Starting from van Dijk [6.3] for signal frequency noise specification < $2kHz_{rms}$, they obtained that **a classical architecture of VCO is not enough, even with improvement of the LC tank at low temperature**. The authors developed a **specific calibration loop** in order to reach the required Phase Noise.

The circuit was implemented on **40nm Bulk CMOS**, and exhibit the following performances:

- Frequency range: 4-5GHz
- Power consumption: 0,4mW for calibration,2mW for the oscillator
- Phase Noise: -100dBc/Hz @ 100KHz
- AM/FM noise: $1,6KHz_{rms}$

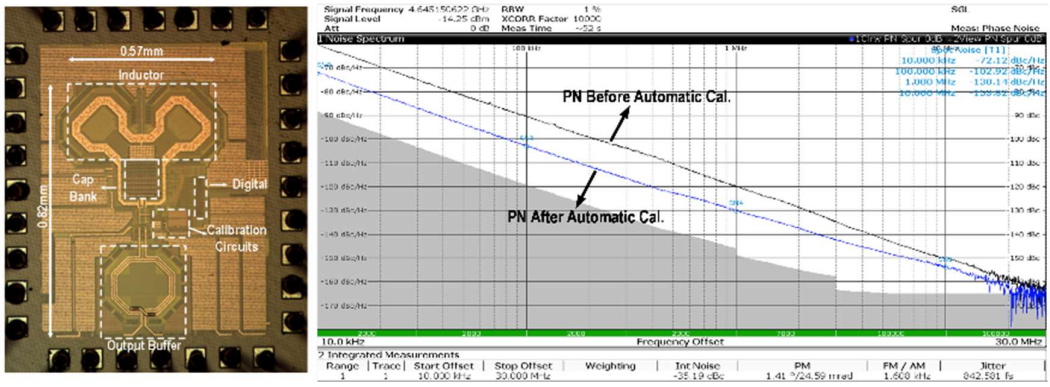


Figure 6-11: A 40nm Bulk CMOS VCO with calibration loop. Phase Noise at 4K (f_0=4.6GHz).

During ISSC 2020, Le Guevel *et al.* from CEA presented a Quantum Integrated Circuit (QIC) : **a single chip in industrial CMOS technology (28nm FDSOI) integrating a double quantum dot with mixed digital-analog circuits** including biasing, signal generation and manipulation at 2.8GHz, and read-out [6.21]. To the authors' knowledge, this is **the very first CMOS integration of quantum-dots together with digital and analog circuitry, thus working at 100mK**.

As **no cryogenic design-kits are available at 100mK**, the FDSOI technology has been selected using the already published transistor-level characterizations at low-temperatures and the **back-biasing feature** which creates more design flexibility, said the authors.

The chip contains a voltage-controlled ring oscillator, a digital multiplexer, a level shifter, a digitally-controlled capacitive divider, a dynamic flip-flop, a digital output buffer, a trans-impedance amplifier, and an analog multiplexer. And the double quantum dot.

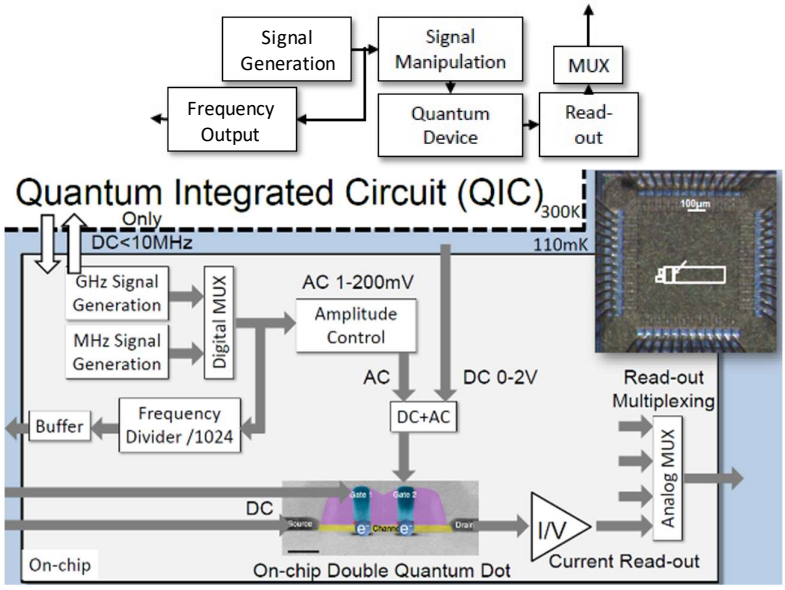


Figure 6-12: Functional block diagram of the mixed digital-analog signal circuitry interfacing a quantum device at 110mK.

The figure below represents the tradeoff between GHz signal generation and the related power consumption, when back-biasing is tuned in the frequency generator (a ring oscillator). **This experience demonstrates the great interest of FDSOI capability in the power management at low temperature**.

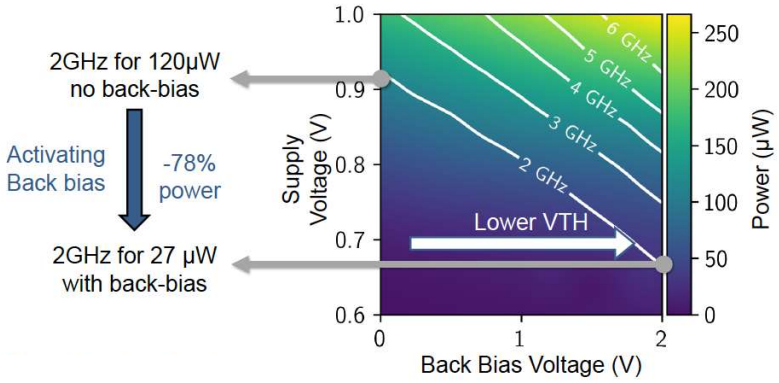


Figure 6-13: Forward back-gating decreases power for same frequency.

The overall IC power consumption is 295µW when all functions are enabled and frequency is set to 2.8GHz, leading to a temperature of 110mK in a He3/He4 dilution-type cryostat. **Thus, the whole power budget available at this temperature is used for manipulation and read-out of one quantum dot.**

Moreover, the circuit is not a qubit controller, but a charge controller. Accurate spin manipulation is out of the circuit's specifications. However, this circuit acts as a demonstrative performance. It is an important milestone since it demonstrates the viability of heterogeneous functions (analog, digital, and quantum) down to 110mK on a single industrial chip while sustaining quantum effects. And **it makes possible the embedding of electronics dedicated to a quantum dot characterization**.

We will come back to this realization in the next subsection, to talk about the read-out part.

6.3 Building blocks for read-out (LETI)

Spin-qubits and transmon require different readout techniques. Especially, Transmon commonly use resonators embedded with the transmon chip, and III/V Josephson Parametric Amplifiers. For a general information, please refer to “IEEE microwave magazine” of August 2020, which develops two papers on the subject [6.33][6.34]. In the next subsection, we intend to make the focus on cryo-CMOS blocks that we find in the processing chain of the both. And we will talk about spin-qubit sensing.

6.3.1. On read-out techniques (LETI)

If the electronics for qubit control seems close to a transmission chain (TX) in a RF-mmW communication system, qubit read-out techniques differ substantially from a RX part, by the fact that **a quantum state is sensed**. The qubit state is not a quantity that is easily captivated and made available to an amplification chain, like with an antenna, or more generally, a channel. **It requires an interface to sample without demolition a physical information in relation to the qubit state**, and convert this information into a signal, which can then be amplified and processed.

Qubit read-out techniques can then be separated into two significant domains of research: sensing & processing.

In IEDM 2019, Meunier *et al* presented a review on challenges and perspectives on spin-qubit readout, regarding the known techniques for qubit sensing [6.35]. In the semiconductor, spin read-out is performed by

converting the spin into a charge information and probe the charge properties of the electron on timescales approaching the μs. **Two main physical concepts are used to perform single shot charge read-out**: direct charge read out thanks to an **electrometer, or capacitive** measurement. The both require a couple of quantum dots linked by tunnel effect. Based on the Pauli exclusion principle, tunneling from one dot to the other is made dependent on the spin state, and then, the charge detection is spin-dependent.

An alternative concept to electrometer is commonly used for quantum dot characterization: the acquisition, by **trans-impedance amplification**, of the current emerging from one of the two dots when this one acts as a Single Electron Transistor. Le Guevel et al, who co-integrated a double quantum dot with its control and read-out circuit, demonstrated such a technique with a quantum charge pumping circuit in ISSCC 2020 [6.21]. However, the acquisition time (state-of-the-art in the range of ms) needs to be enhanced to meet quantum computing requirements.

Readout of spin-states via quantum point contact (QPC) or single electron transistor (SET) charge sensors, proximal to each quantum dot, however, pose a significant challenge to scale-up, in that they require separate surface gates or large contact leads, crowding the device and tightly constraining the on-chip architecture[6.50].

J. I. Colless *et al* [6.36] also report that charge sensing using QPCs or SETs requires that the sensor be kept at a value of conductance where sensitivity is maximized. This is typically achieved by applying additional compensating voltages to gates when acquiring a charge-stability diagram. Whereas gate sensors do not require such offset charge compensation or gate voltage control. On further practical use, they find gate reflectometry is a robust detector at elevated temperatures, in contrast to QPC charge sensors which suffer from a thermally broadened conductance profile and suppressed sensitivity with increasing temperature. (Comparison @ T ~20 mK and T ~1100 mK). **The perspective of a future Silicon spin-qubit at 1K then pleads in favor of gate-reflectometry technique**.

In capacitive sensing, or gate-coupled RF reflectometry, sensing is accomplished by measuring the dispersive response of an electromagnetic RF resonator connected to one of the qubit gates (already in place to define the quantum dots) and excited at its resonance frequency. The dispersive response to the incident RF wave comes from shifts in the quantum capacitance when electrons undergo tunneling (tunneling modifies the gate capacitance beyond the geometric contribution). Changes in the quantum capacitance alter the resonator frequency, which in turn leads to a shift in the phase and magnitude of the reflected carrier. This response of the resonator is detected by fast sampling of the in-phase and quadrature components of the reflected signal to produce a baseband signal, proportional to the dispersive shift. J. I. Colless *et al* demonstrated such a technique on double quantum dot early in 2010's [6.36]. It was surprisingly found as a fast and sensitive readout technique in the single-electron regime.

About read-out specifications comprising the qubit properties and the attached interface, we report, again, van Dijk *et al* in 2019 [6.3], where the exercise was done with the assumption of QPC charge sensor device.

The readout fidelity is evaluated in regards with three factors:

- The probability $\mathbf{P_{charge}}$ that the spin state is correctly projected to the charge state. This is limited by the **quantum-dot control** electronics. A large singlet-triplet splitting is desired to limit the influence of the control electronics on the readout.
- The probability $\mathbf{P_{sense}}$ that the charge sensor correctly detects the charge state. This depends highly on the **type of sensor** employed.
- The probability $\mathbf{P_{detect}}$ that the readout circuit correctly discriminates the signal of the charge sensor. This depends on the **readout processing**. For a direct current readout flowing from a Quantum Point Contact (QPC), like assumed by van Dijk *et al*., the contribution is considered to be only related to the added noise and then to the signal to noise ratio.

Van Dijk *et al* investigate specifications for the electronics that control the spin to charge conversion and specifications for the electronics processing the read-out signal, and they decline a case study of the specifications for a spin-qubit read-out, **when the three above probabilities equally contribute**.

The publication from van Dijk *et al* [6.3] **underlines the approach of building specifications by including the quantum object properties with the electronics. Coming from two worlds, with two formalisms, working according to such an approach is challenging, however probably necessary.**

	Value	Infidelity contribution to the readout	
Detuning energy			
Nominal	83.2 mV (4.2 meV, 1.0 THz)		
Error	0.24 mV (12 μeV, 2.8 GHz) $\sigma = 0.24\ \mathrm{mV_{rms}}$, PSD $= 0.24\ \mu\mathrm{V}/\sqrt{\mathrm{Hz}}$	167×10^{-6}	
Tunnel coupling			
Nominal	39 MHz (0.16 μeV)	167×10^{-6}	$P_{\mathrm{charge}} = 99.967\%$
Charge sensor		333×10^{-6}	$P_{\mathrm{sense}} = 99.967\%$
QPC			
Signal	400 pA		
Noise	53 $\mathrm{pA_{rms}}$, PSD $= 57\ \mathrm{fA}/\sqrt{\mathrm{Hz}}$	222×10^{-6}	
Readout circuit			
Input-referred noise	26 $\mathrm{pA_{rms}}$, PSD $= 28\ \mathrm{fA}/\sqrt{\mathrm{Hz}}$	111×10^{-6}	$P_{\mathrm{detect}} = 99.967\%$ $F = 99.9\%$

Figure 6-14: Example of specification for a Pauli spin blockade readout scenario with direct charge measurement (QPC is for Quantum Point Contact).

6.3.2. Building blocks for read-out processing (LETI)

Most of publications talking about CMOS read-out building blocks come from the research group with QuTech [6.37] to [6.49], since 2018. Based on a supposed architecture where the RF reflectometry is implemented, the authors evaluate a Low Noise Amplifier, a Voltage Control Oscillator, an integrated circulator in place of directional coupler.

We also mention Le Guevel *et al.* who presented a Trans-Impedance Amplifier on a Quantum Integrated Circuit during ISSCC 2020 [6.21].

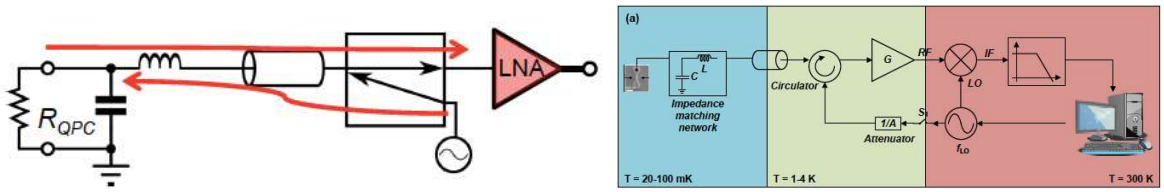


Figure 6-15: Readout architecture by reflectometry (left : directional coupler, right : circulator) Charbon et al.

In JSSC 2018 [6.37], the authors presented a CMOS LNA incorporating a noise cancellation technique. The architecture was proposed by Bruccoleri *et al* in 2004 [6.38] and prototyped by Charbon *et al* a first time in 2017 [6.26]. The circuit is done on 160nm BULK. Within this reflectometry scenario, a weak reflected power of −135 dBm in 1-MHz bandwidth per qubit is considered, to preserve qubits from strong electric field variations. It means that the readout circuitry should be operating at noise level below 40 pV/√Hz.

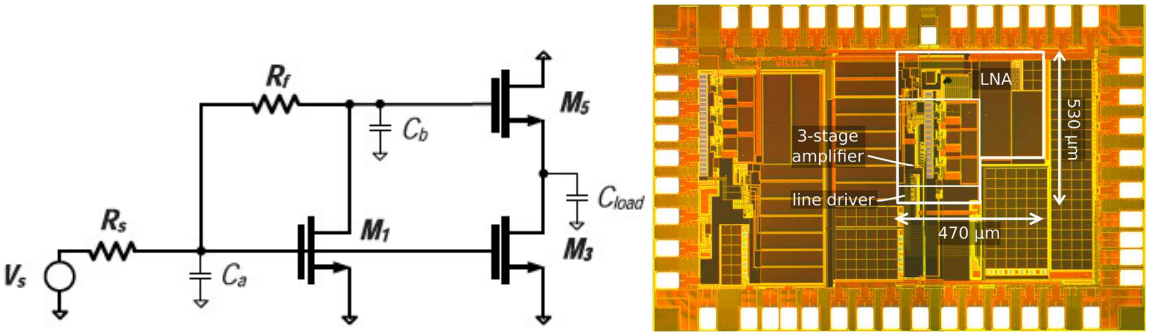


Figure 6-16: Simplified architecture of the LNA stage. Micrograph with three-stage diode loaded amplifier and 50Ohms driver.

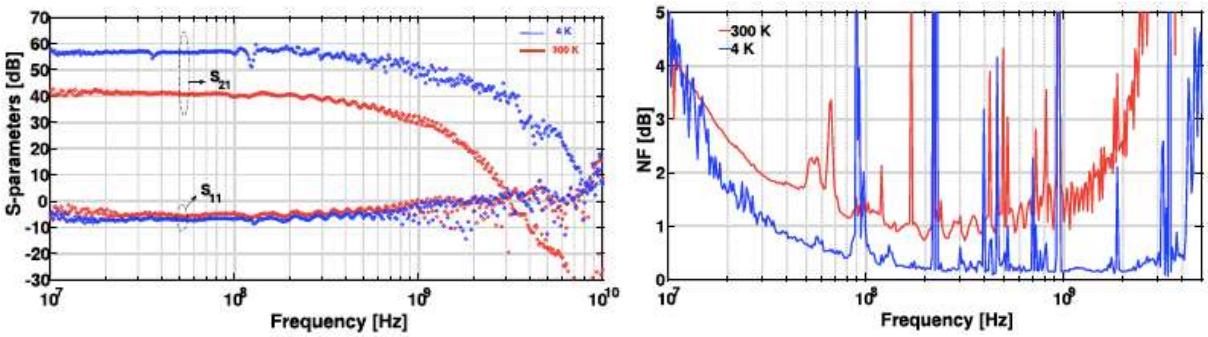


Figure 6-17: S-parameters and NF of the reported LNA.

S-parameters are representing gain (S21) and input matching (S11) both at room temperature and 4 K. The LNA stage exhibit a gain of 21V/V, and the whole function including a three stage amplifier and a 50 Ω driver has a gain of 40 dB and a −3-dB bandwidth of 400 MHz at room temperature, which improves to 57 dB and 500 MHz at 4 K, respectively. Input matching improves by 3 dB by going to lower temperatures, reaching a value of −8 dB at 4 K.

Noise measurements reveal an in-band NF of 0.1 dB±0.05 dB at 4 K and 0.8 dB±0.3 dB at 300 K, as shown in the above figure, which is equivalent to a noise temperature of 7 K and 60 K, respectively. The measured NF then translates to an improvement of a ~10× noise factor at 4 K with respect to 300 K, that can be attributed to the large decrease in thermal noise.

Although the linearity performance (not shown here) is not comparable with the state-of-the-art LNAs, high linearity is not required in the target application because of extremely low level of the input power, which is fixed below −110 dBm.

The chip dissipates 91 mW (80 mW) at 4 K (300 K), of which 54.9 mW (45.9 mW) is for the LNA, and the remaining part is dissipated by bias and gain stages.

The authors mentioned **this LNA and the following VCO benefited from the development of a modified MOS11 model for transistors operating at 4 K. This is noticeable information since for now, there is no compact model for any technology at low temperature.**

In [6.37] again, the authors present a VCO that is commonly used for control and read-out. In order to save the qubit dephasing time T2, the frequency noise (FN) specification for the PLL is set to be lower than 1.9 kHz_{rms} for a 6-GHz carrier. The lower integration bound is set by the quantum operation cycle (worst case: 1/T2) and the higher limit is determined by the qubit operation speed.

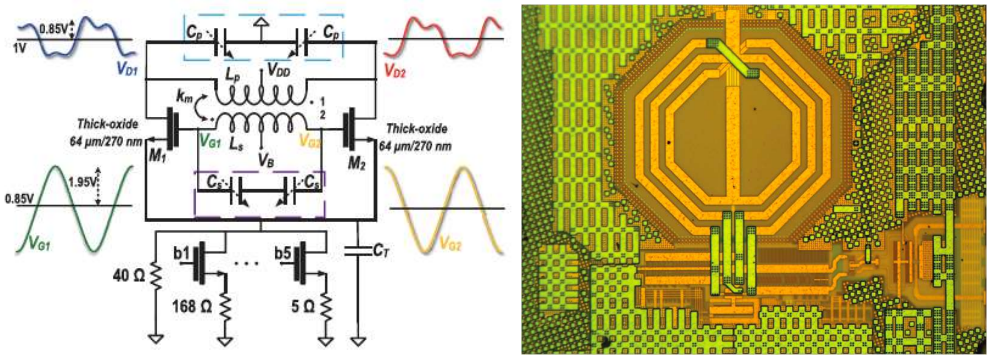


Figure 6-18: Class-F2,3 oscillator with simulated waveforms. Chip micrograph of the 40-nm CMOS.

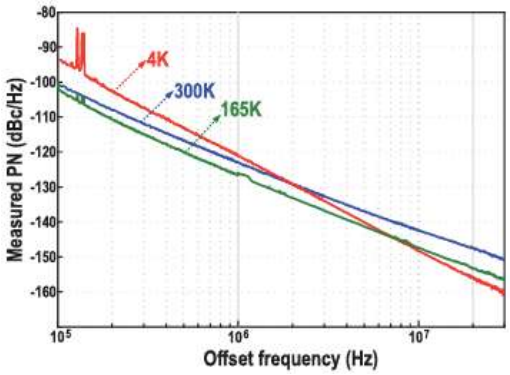


Figure 6-19: Measured PN at 6.3 GHz at various temperatures.

About the Phase Noise, the authors said: Even though the flicker noise upconversion is significantly suppressed by virtue of the chosen topology, the oscillator's **$1/f^3$ PN corner increases dramatically at lower temperatures**, thus hinting at a **larger 1/f noise corner for MOS transistors at cryogenic temperatures. Another reason could be the worsening of mismatch between the two core transistors at cryogenic temperatures**, resulting in an asymmetric rise/fall times of oscillation waveforms and, consequently, resulting in a larger dc value of the impulse sensitivity function and higher flicker noise upconversion.

This Phase Noise response is to be compared to the VCO that was shown by another team of the same group of research during ISSCC2020 [6.12], who developed a calibration technique in order to reduce the phenomena mentioned here. Please refer to section 6.2.

In 2019, Charbon *et al* presented an integrated circulator on 40nmCMOS[6.44] centered at 6.5GHz.

CMOS implementations of circulators have been demonstrated in the K-band (18-27GHz), based on switched transmission lines. Switched transmission lines require lambda/4 transmission lines at the switching frequency (i.e., 1/3 of the 6.5-GHz operating frequency), thus resulting in a significant penalty on the circulator's die area and insertion loss. N-path filters require non-overlapping clock phases at the 6.5-GHz operating frequency, thus resulting in excessive power dissipated by the clock drivers.

By exploiting a novel architecture based on all-pass filters, the authors avoid both large-area transmission lines and power-hungry high-frequency clock drivers. This results in a 0.45mm^2 passive circulator operating down to 4.2K with large bandwidth, while dissipating only 2.1 mW with 1.3-dB minimum insertion loss and 17-dB maximum isolation over the 1-dB bandwidth from 5.8 GHz to 7.6 GHz.

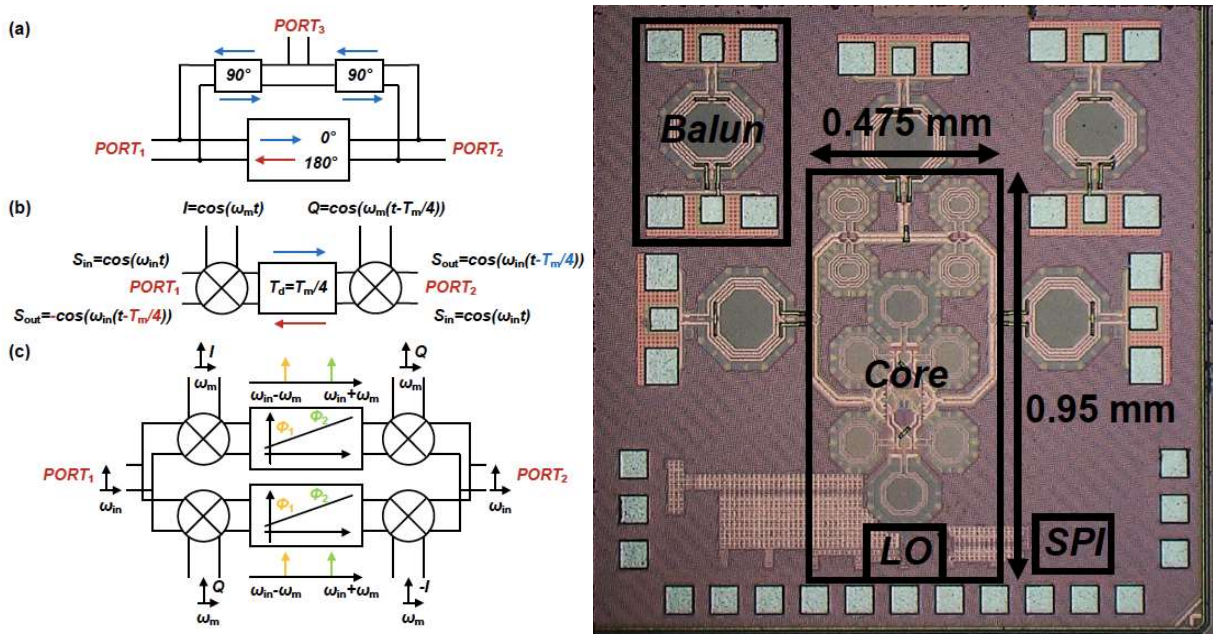


Figure 6-20: Block diagram of (a) the circulator, including (b) the time domain analysis and (c) the frequency domain analysis of the non-reciprocal branch. Micrograph of the die.

The measured S-parameters at 4.2K, after calibration and balun de-embedding, are reported in figure below. Compared to Room Temperature, improvements are observed thanks to increase in quality factor of passives, in particular inductors, due to the reduction of substrate losses thanks to carrier freeze-out, and thanks to a reduction of R_{ON} resistance of the mixer switches in series with the signal path, due to larger carrier mobility.

However, **the authors don't comment the higher insertion loss on the 1-2 path and the lower isolation on 1-3 and 2-3. The 1-2 path is probably the more critical since it reduces the signal-to-noise ratio before the LNA**.

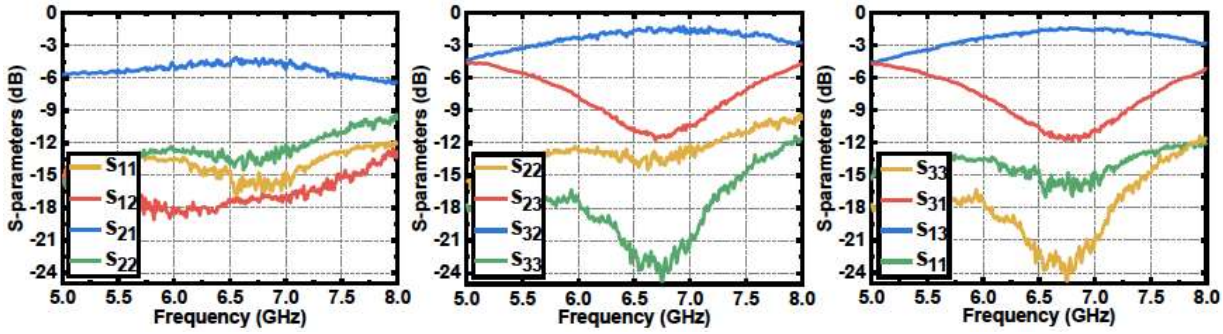


Figure 6-21: Circulator S-parameter measurements at 4.2 K.

in ISSCC 2020 Le Guevel *et al,* a double gate transistor (L=40nm, W=80nm, S=96nm) acts as a double quantum dot at 100mK, and the current induced by quantum charge pumping is measured by a Trans-Impedance Amplifier, followed by a pass-gate analog multiplexer for read-out [6.21]. The TIA schematic and current response is given figure below.

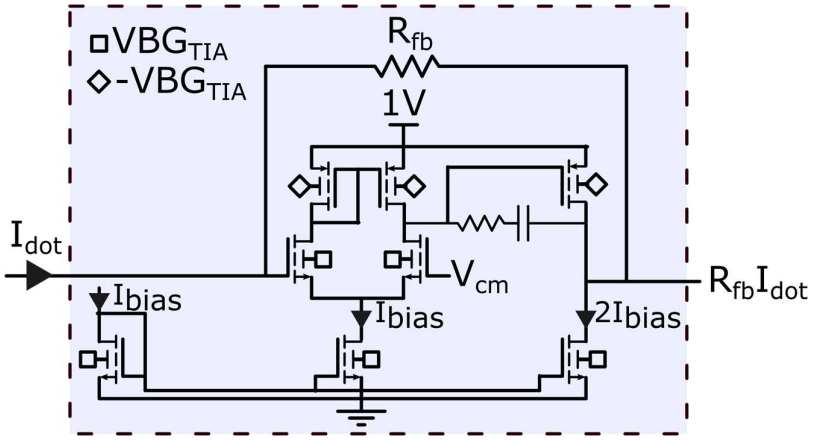


Figure 6-22: low power TIA for quantum charge pumping.

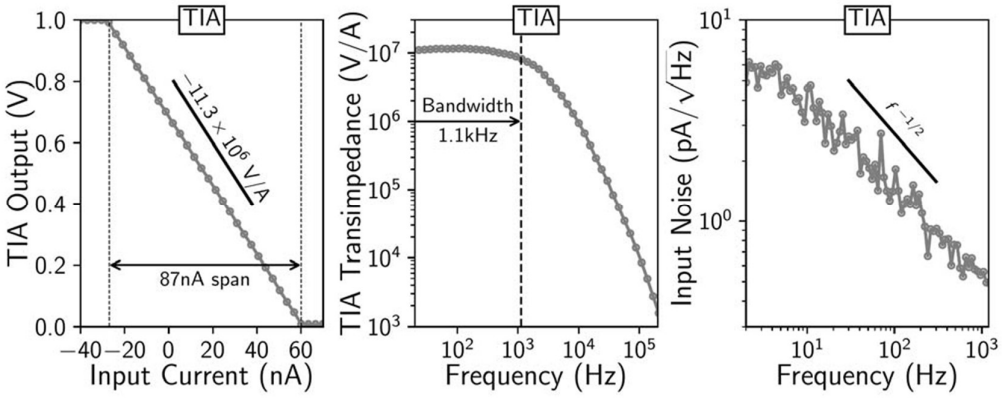


Figure 6-23: TIA current read-out.

With a linear response on 87nA, a trans-impedance of 141dBΩ on a 1.1kHz bandwidth, and 50 pA/√Hz maximum input noise, the TIA consumes 1µW on 10µm^2, allowing expectation of massive parallel sensing within the limited cryostat power budget.

6.3.3. About read-out RF reflectometry multiplexing (LETI)

Working on strategies for read-out multiplexing since 2010's, Reilly *et al* presented in 2014 a prototype for, let's say a tile, driving three to ten qubits [6.50]. And they presented in 2019 a more complex experimentation with a 28nmFDSOI I/O control chip at the qubit stage, where this kind of read-out multiplexer was also implemented to drive two qubits [6.51][6.52], in order to validate the scalability of their approach.

Whereas most of cryo-CMOS circuits are sized (in power consumption) to live in the 4K stage, **the multiplexed RF reflectometry presumes that a set of matching resonators are close to the qubits, hence, at or below 100mK. Commonly, this kind of function is a passive function with limited power dissipation. However, this network, whose complexity increases with the number of channels, is susceptible to alter the signal at the early level of sensing, leading to a significant signal-to-noise ratio degradation.**

Moreover, like Meunier *et al* said in IEDM 2019 [6.35], "the typical size of the tank circuit (inductance, capacitance) used in the circuit read-out is at the lowest equal to 100µmx100µm. Therefore, with the current technology, it is difficult to maintain a small foot print for the spin qubit taking into account the whole system requested to perform read out. One approach to fabricated small footprint inductors consists in exploiting

the very large inductance of superconductor thin film which are compatible with CMOS processing. For instance, TiN, which is widely used in gate stacks, allows reaching few tens of nH in a few micron squares."

Waiting for superconducting coils with nanometer resolution on width and space, the experience that is reported here with GaAs qubits enlighten some of the challenges in the realization of such a function based on superconducting materials.

The figure below exhibits a three-path circuit. The feed-lines, bias tees, and resonators are fabricated on a sapphire chip using superconducting Niobium. The Niobium remains superconducting at the moderate magnetic fields needed to operate spin qubits. Each inductor L_i in resonance with the parasitic capacitance C_p defines a unique frequency channel f_i for addressing each readout detector.

The ultimate goal of the approach is to define gates of the quantum dots as readout sensors. Here, the experience is made with one gate, and two QPC. Each resonant circuit contains an integrated bias tee for independent dc voltage biasing. The on-chip bias tees are constructed using interdigitated capacitors with interspace high permittivity material. The self-resonance frequency of all the inductors is increased by over-etching the sapphire dielectric between adjacent turns, decreasing the effective dielectric constant and reducing the capacitance.

Since QPC acts as a capacitance in parallel with a variable resistance, it is possible to modulate the resonance matching with the gate bias level. Whereas on the gate, biasing strongly changes the resonance frequency. This is a property of GaAs that we should not see with Silicon-Spin qubits, where the quantum capacitance variation due to tunnelling is weaker. And given the large separation in resonance frequencies, crosstalk is negligible in this 3-channel implementation.

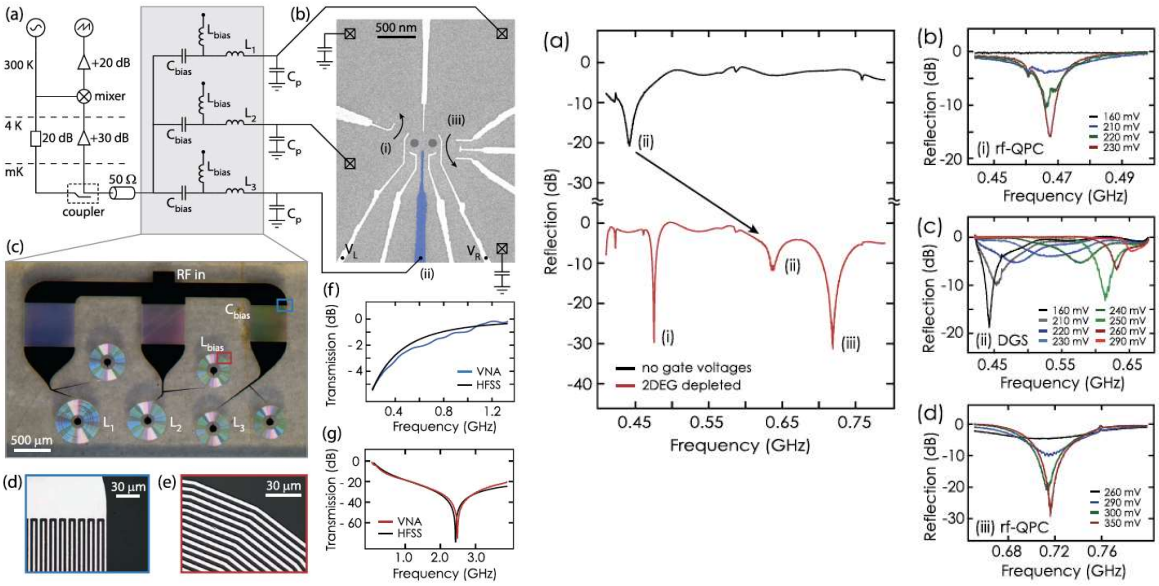


Figure 6-24: (left) Three channel frequency multiplexing scheme for spin qubit readout, and micrograph of the multiplexing chip which is patterned using niobium on a sapphire substrate, comprising interdigitated capacitors and spiral inductors. (right) Reflection analysis for variable gate voltage on each multiplexed channel [6.50].

The same experience was done with a 10-channels chip, shown in the figure below. The 10 channels are defined using inductors Li with values between 60 nH and 250 nH that form a resonant circuit with parasitic

capacitance Cp as described above. Each channel again integrates a bias tee, needed for independent biasing of the gate sensors.

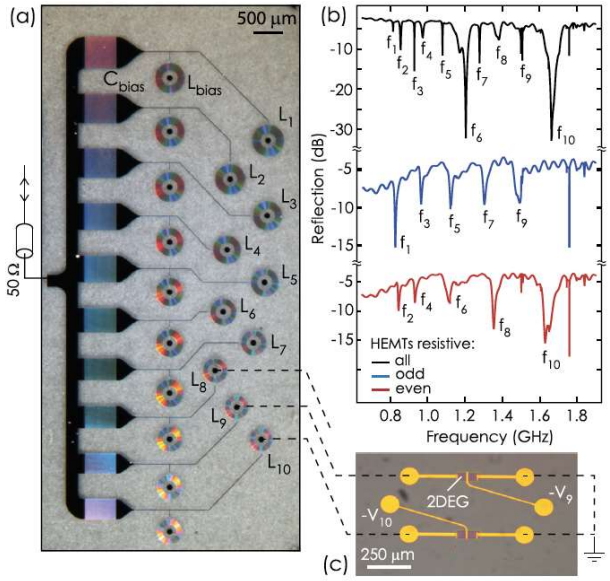


Figure 6-25: (a) Optical micrograph of a 10:1 MUX chip with integrated bias tees. (b) Frequency response of the MUX chip with inductors Li each connected to GaAs HEMT with a gate-controllable conductance to mimic the response of 10 QPCs. Data show the response with all HEMTs in the resistive state (black), odd HEMTs resistive (blue), and even HEMTs resistive (red). (c) shows an optical micrograph of a section of the HEMT device with dashed lines indicative of bondwires. [6.50].

Reducing the size of this type of component is possible, by playing on bias strategies, and reducing critical dimensions of interdigitated capacitor and inductors. **A more serious challenge is frequency crowding arising from the limited bandwidth available using planar lumped element inductors**. For a maximum resonance frequency of about 5 GHz and given the need to separate channels by several linewidths to suppress crosstalk, the total number of independent gate sensors that can be read out simultaneously is around 100, said the authors.

If the constraint of simultaneous readout is relaxed, time domain multiplexing via cryogenic switching elements would allow readout of banks of frequency multiplexed channels to be interleaved in time, they conclude.

6.4 Perspectives (LETI)

A wide set of functions are already explored, especially for control, which benefit from the large experience on integrated RF & Digital Communications. However, the absence of robust parametric models does not allow to optimise simulations. Therefore, we might expect enhanced performances simply by optimisation in the next 2-3 years, thanks to back-simulations and specific models.

- Review on control: voltage reference, SPI, digital signal generator, active filters, VGA, mixers, VCO, filters, vector modulator
- Review on read-out: TIA, LNA, analog multiplexer, high Q resonators, mixers, VCO, circulator

Nonetheless, referring to the most advanced 'Horse Ridge' dissipating 12 mW/qubit (digital and analog parts) at 4K where power budget is around 1W, it means that we would be able to control "only" 80 qubits… Power

budget is clearly the main parameter that drives circuit design. Perhaps, some further enhancements might occur by transferring know-how from Ultra-Low Power IoT architectures, including complex spectral utilization (~6 years).

On read-out part, it seems that according to Reilly in [6.19], the act of measuring the state of qubits does not alarm the community, assuming that measuring a system with a sensor is generally easier (requiring less energy) than controlling it with an actuator. Nevertheless, considering the need of efficient multiplexing at 20-100mK for thousands of qubits, followed by a wideband amplification of very weak signals susceptible to be disturbed by the crowded spectrum, requires much right now in order to follow the next increasing number of qubits.

Notice that the most advanced 20mK control chip, which handles IO demultiplexing of signals for 32 qubits, was designed on 28nm FDSOI. That is an encouraging milestone for the fully depleted SOI, which surpasses other CMOS technologies on most properties, especially at Ultra Low Power operation.

References

[6.1] E. Charbon, “Cryo-CMOS Electronics for Quantum Computing Applications,” IEEE ESSDERC, Sep. 2019.
[6.2] J. P. G. van Dijk, E. Charbon, and F. Sebastiano, “The electronic interface for quantum processors,” Microprocessors and Microsystems, vol. 66, pp. 90–101, Apr. 2019.
[6.3] J. P. G. van Dijk *et al*., “The impact of classical control electronics on qubit fidelity” Phys. Rev. Applied, vol. 12, no. 4, Art. no. 4, Oct. 2019
[6.4] E. Charbon *et al*., “15.5 Cryo-CMOS circuits and systems for scalable quantum computing,” in 2017 IEEE International Solid-State Circuits Conference (ISSCC), Feb. 2017
[6.5] B. Patra et al., “Cryo-CMOS Circuits and Systems for Quantum Computing Applications,” IEEE Journal of Solid-State Circuits, vol. 53, no. 1, Art. no. 1, Jan. 2018
[6.6] H. Homulle, F. Sebastiano, and E. Charbon, “Deep-Cryogenic Voltage References in 40-nm CMOS,” IEEE Solid-State Circuits Letters, vol. 1, no. 5, Art. no. 5, May 2018
[6.7] F. Sebastiano, “WSJ-6 - Cryogenic CMOS interfaces for Quantum computers,” RFIC, Jun. 2019.
[6.8] B. P. Wuetz *et al*., “Multiplexed quantum transport using commercial off-the-shelf CMOS at sub-kelvin temperatures,” p. 8, Jul. 2019.
[6.9] J. van Staveren *et al*., “Voltage References for the Ultra-Wide Temperature Range from 4.2K to 300K in 40-nm CMOS,” in ESSCIRC 2019 - IEEE 45th European Solid State Circuits Conference (ESSCIRC), Sep. 2019.
[6.10] J. P. G. van Dijk *et al*., “The Impact of Classical Control Electronics on Qubit Fidelity”, Phys. Rev. Applied, vol. 12, no. 4, Art. no. 4, Oct. 2019.
[6.11] J. M. Boter *et al*., “A sparse spin qubit array with integrated control electronics,” in 2019 IEEE International Electron Devices Meeting (IEDM), Dec. 2019.
[6.12] J. Gong, Y. Chen, F. Sebastiano, E. Charbon, and M. Babaie, “A 200dB FOM 4-5GHz Cryogenic Oscillator with an Automatic Common-Mode Resonance Calibration for Quantum Computing Applications,” Feb. 2020.
[6.13] B. Patra *et al*., “A Scalable Cryo-CMOS 2-to-20GHz Digitally Intensive Controller for 4×32 Frequency Multiplexed Spin Qubits/Transmons in 22nm FinFET Technology for Quantum Computers,” p. 3, Feb. 2020.
[6.14] B. Patra, M. Mehrpoo, A. Ruffino, F. Sebastiano, E. Charbon, and M. Babaie, “Characterization and Analysis of On-Chip Microwave Passive Components at Cryogenic Temperatures,” IEEE J. Electron Devices Soc., vol. 8, pp. 448–456, Apr. 2020
[6.15] J. C. Bardin *et al*., “29.1 A 28nm Bulk-CMOS 4-to-8GHz ¡2mW Cryogenic Pulse Modulator for Scalable Quantum Computing,” in 2019 IEEE International Solid- State Circuits Conference - (ISSCC), San Francisco, CA, USA, Feb. 2019

[6.16] J. C. Bardin *et al*., "Design and Characterization of a 28-nm Bulk-CMOS Cryogenic Quantum Controller Dissipating Less Than 2 mW at 3 K," IEEE Journal of Solid-State Circuits, vol. 54, no. 11, Art. no. 11, Nov. 2019.
[6.17] J. M. Hornibrook et al., "Cryogenic Control Architecture for Large-Scale Quantum Computing," Phys. Rev. Applied, vol. 3, no. 2, Art. no. 2, Feb. 2015, doi: 10.1103/PhysRevApplied.3.024010.
[6.18] S. J. Pauka et al., "A Cryogenic Interface for Controlling Many Qubits," arXiv:1912.01299 [cond-mat, physics:physics, physics:quant-ph], Dec. 2019, Accessed: Feb. 12, 2020. [Online]. Available: http://arxiv.org/abs/1912.01299.
[6.19] D. J. Reilly, "Challenges in Scaling-up the Control Interface of a Quantum Computer," arXiv:1912.05114 [cond-mat, physics:physics, physics:quant-ph], Dec. 2019, Accessed: Jul. 15, 2020. [Online]. Available: http://arxiv.org/abs/1912.05114.
[6.20] J. Lin et al., "Scalable and customizable arbitrary waveform generator for superconducting quantum computing," AIP Advances, vol. 9, no. 11, Art. no. 11, Nov. 2019.
[6.21] L. Le Guevel, "A 110mK 295µW 28nm FDSOI CMOS Quantum Integrated Circuit with a 2.8GHz Excitation and nA Current Sensing of an On-Chip Double Quantum Dot," vol. ISSCC, Feb. 2020.
[6.22] J. P. G. van Dijk *et al*., "The impact of classical control electronics on qubit fidelity" Phys. Rev. Applied, vol. 12, no. 4, Art. no. 4, Oct. 2019
[6.23] T. F. Watson *et al*., « A programmable two-qubit quantum processor in silicon », Nature, vol. 555, 2018
[6.24] J. P. G. van Dijk, E. Charbon, et F. Sebastiano, « The electronic interface for quantum processors », Microprocessors and Microsystems, vol. 66, p. 90‑101, apr. 2019
[6.25] Charbon *et al*, « Cryo-CMOS for Quantum Computing ». IEEE IEDM 2016
[6.26] Charbon *et al*, « 15.5 Cryo-CMOS circuits and systems for scalable quantum computing ». IEEE ISSCC 2017
[6.27] Boter *et al*, « A sparse spin qubit array with integrated control electronics ». IEEE IEDM 2019
[6.28] Franke *et al*, « Rent's Rule and Extensibility in Quantum Computing ». *Microprocessors and Microsystems* 2019
[6.29] Struck *et al*, « Low-Frequency Spin Qubit Energy Splitting Noise in Highly Purified 28Si/SiGe ». Npj Quantum Information 2020
[6.30] D. J. Reilly, « Challenges in Scaling-up the Control Interface of a Quantum Computer », Microsoft Quantum Sidney, arXiv, 2019
[6.31] S. J. Pauka et al., « A Cryogenic Interface for Controlling Many Qubits », Microsoft Quantum Sidney, arXiv, 2019
[6.32] J. van Dijk, *et al*. "SPINE (SPIN Emulator) - A Quantum-Electronics Interface Simulator," in 2019 IEEE 8th International Workshop on Advances in Sensors and Interfaces (IWASI), Otranto, Italy, Jun. 2019.
[6.33] J. C. Bardin, D. Sank, O. Naaman, and E. Jeffrey, "Quantum Computing: An Introduction for Microwave Engineers," IEEE Microwave, vol. 21, no. 8, pp. 24–44, Aug. 2020.
[6.34] J. Aumentado, "Superconducting Parametric Amplifiers: The State of the Art in Josephson Parametric Amplifiers," IEEE Microwave, vol. 21, no. 8, pp. 45–59, Aug. 2020.
[6.35] T. Meunier *et al.*, "Qubit read-out in Semiconductor quantum processors: challenges and perspectives," in *2019 IEEE International Electron Devices Meeting (IEDM)*, Dec. 2019
[6.36] J. I. Colless *et al.*, "Dispersive Readout of a Few-Electron Double Quantum Dot with Fast rf Gate Sensors," *Phys. Rev. Lett.*, vol. 110, no. 4, p. 046805, Jan. 2013
[6.37] B. Patra et al., "Cryo-CMOS Circuits and Systems for Quantum Computing Applications," IEEE Journal of Solid-State Circuits, vol. 53, no. 1, Art. no. 1, Jan. 2018, doi: 10.1109/JSSC.2017.2737549.
[6.38] F. Bruccoleri, E. A. M. Klumperink, and B. Nauta, "Wide-Band CMOS Low-Noise Amplifier Exploiting Thermal Noise Canceling," IEEE J. Solid-State Circuits, vol. 39, no. 2, pp. 275–282, Feb. 2004, doi: 10.1109/JSSC.2003.821786.
[6.39] H. Homulle, F. Sebastiano, and E. Charbon, "Deep-Cryogenic Voltage References in 40-nm CMOS," IEEE Solid-State Circuits Letters, vol. 1, no. 5, Art. no. 5, May 2018, doi: 10.1109/LSSC.2018.2875821.

[6.40] A. Ruffino, Y. Peng, and E. Charbon, “Interfacing Qubits via Cryo-CMOS Front Ends,” in 2018 IEEE International Conference on Integrated Circuits, Technologies and Applications (ICTA), Nov. 2018, pp. 42–44, doi: 10.1109/CICTA.2018.8705712.
[6.41] E. Charbon, “The Role of Cryo-CMOS in Quantum Computers,” LBL– Berkeley, Feb. 2019.
[6.42] J. P. G. van Dijk, E. Charbon, and F. Sebastiano, “The electronic interface for quantum processors,” Microprocessors and Microsystems, vol. 66, pp. 90–101, Apr. 2019, doi: 10.1016/j.micpro.2019.02.004.
[6.43] M. Mehrpoo et al., “Benefits and Challenges of Designing Cryogenic CMOS RF Circuits for Quantum Computers,” in 2019 IEEE International Symposium on Circuits and Systems (ISCAS), May 2019, pp. 1–5, doi: 10.1109/ISCAS.2019.8702452.
[6.44] A. Ruffino, Y. Peng, F. Sebastiano, M. Babaie, and E. Charbon, “A 6.5-GHz Cryogenic All-Pass Filter Circulator in 40-nm CMOS for Quantum Computing Applications,” in 2019 IEEE Radio Frequency Integrated Circuits Symposium (RFIC), Jun. 2019, pp. 107–110, doi: 10.1109/RFIC.2019.8701836.
[6.45] F. Sebastiano, “WSJ-6 - Cryogenic CMOS interfaces for Quantum computers,” RFIC, Jun. 2019.
[6.46] J. van Staveren et al., “Voltage References for the Ultra-Wide Temperature Range from 4.2K to 300K in 40-nm CMOS,” in ESSCIRC 2019 - IEEE 45th European Solid State Circuits Conference (ESSCIRC), Sep. 2019, pp. 37–40, doi: 10.1109/ESSCIRC.2019.8902861.
[6.47] J. Gong, Y. Chen, F. Sebastiano, E. Charbon, and M. Babaie, “A 200dB FoM 4-to-5GHz Cryogenic Oscillator with an Automatic Common-Mode Resonance Calibration for Quantum Computing Applications,” p. 3, Feb. 2020.
[6.48] B. Patra, M. Mehrpoo, A. Ruffino, F. Sebastiano, E. Charbon, and M. Babaie, “Characterization and Analysis of On-Chip Microwave Passive Components at Cryogenic Temperatures,” IEEE J. Electron Devices Soc., vol. 8, pp. 448–456, Apr. 2020, doi: 10.1109/JEDS.2020.2986722.
[6.49] A. Ruffino, Y. Peng, F. Sebastiano, M. Babaie, and E. Charbon, “A Wideband Low-Power Cryogenic CMOS Circulator for Quantum Applications,” IEEE J. Solid-State Circuits, vol. 55, no. 5, pp. 1224–1238, May 2020, doi: 10.1109/JSSC.2020.2978020.
[6.50] J. M. Hornibrook et al., “Frequency multiplexing for readout of spin qubits,” Applied Physics Letters, vol. 104, no. 10, Art. no. 10, Mar. 2014, doi: 10.1063/1.4868107.
[6.51] S. J. Pauka et al., “A Cryogenic Interface for Controlling Many Qubits,” arXiv:1912.01299 [cond-mat, physics:physics, physics:quant-ph], Dec. 2019, Accessed: Feb. 12, 2020. [Online]. Available: http://arxiv.org/abs/1912.01299.
[6.52] D. J. Reilly, “Challenges in Scaling-up the Control Interface of a Quantum Computer,” arXiv:1912.05114 [cond-mat, physics:physics, physics:quant-ph], Dec. 2019, Accessed: Jul. 15, 2020. [Online]. Available: http://arxiv.org/abs/1912.05114.

7 Novel technologies and associated building blocks

7.1 Superconducting logic (IBM)

Superconducting logic is based on superconducting metals and Josephson junctions. In this type of logic, information is stored as quantized magnetic flux LI inside a superconducting loop, where L is the inductance and I the current running through it. As the current through a Josephson junction surpasses the critical value, the superconducting phase across the junction increases by 2π and a single flux quantum SFQ (a voltage pulse) is procuded at the output. Due to the low losses, the associated switching energies are extremely small and can be in the order of aJ. This represents the main advantage of the technology compared to a cryogenic CMOS solution.

While there are many attractive applications for superconducting logic, thugh of course limited to cryogenic operation, quantum computing is one of the more promising. Superconducting logic, due to the extremely low switching energies, could enable logic circuits at millikelvin temperature stages, where the power dissipation of CMOS would be limiting.

Table: Available superconducting logic families

Name	Power	Static Power	Dynamic power (per JJ)	Static Gates	JJ count
RSFQ: rapid single flux quantum	DC	High	$\alpha I_c \Phi_0 f$	No	10^5
LR-RSFQ: inductor-resistor RSFQ	DC	Low	$\alpha I_c \Phi_0 f$	No	
LV-RSFQ: low-voltage RSFQ	DC	Low	$\alpha I_c \Phi_0 f$	No	
ERSFQ: energy-efficient RSFQ	DC	0 *	$I_b \Phi_0 f$	No	10^4
eSFQ: energy-efficient SFQ	DC	0 *	$I_b \Phi_0 f$	No	10^3
RQL: reciprocal quantum logic	AC	~0	$\alpha I_c \Phi_0 f\, 2/3$	Some	10^5
PML: phase mode logic	AC	~0	$\alpha I_c \Phi_0 f/3$	Some	
AQFP: adiabatic quantum flux parametron	AC	~0	$\alpha I_c \Phi_0 2\tau_{sw}/\tau_x$	No	10^3

The table above shows the many different superconducting logic families. Single flux quantum SFQ digital logic was described above. Rapid SFQ circuits use resistors to distribute DC supply currents to the superconducting loops. The use of resistors means that static power dissipation is relatively high, 10-100 times the dynamic power. As the required bias current per switching Josephson junction is ~0.7 times the critical current, typically around 100 μA, this limits the number of logic elements to the thousands. Nevertheless, quite advanced circuits such as 8-bit ALU were fabricated using this technology. Energy efficient RSFQ (ERSFQ) is a related technology where the logic gates are biased using a Josephson junction in series with a large bias inductor. While the static power becomes approximately 0, the biasing network requires a significant overhead to the chip area.

More mature superconducting logic processes demonstrated today almost exclusively make use of Nb as the superconductor and almunina as the Josephson junction tunneling barrier. The minimum feature sizes are typically in the range of 200 – 1000 nm. The acceptable process varations for this type of logic is typically smaller than for CMOS, which represents the main process challenge, especially for smaller junctions – which are key to improved energy efficiency. However, switching speed does not depend on the device size, opposite to CMOS, so the minimum feature sizes can be larger than state-of-the-art CMOS.

The figure below shows delay versus switching energy for several different logic technologies. Solid circles with error bars show the energy range including refrigeration power for different refrigeration technologies. We note however that this metric is not immediately relevant to quantum computer applications where scalability and power dissipation inside the cryostat are the main concerns. Dashed lines show constant energy-delay products. Generally, the superconducting logic families perform significantly better than CMOS at lower frequencies. The limitation with regards to number of logic devices to the thousands is however reiterated. The benchmark values for the CMOS technologies (red squares) is done at room-temperature, while the SFQ technologies are of course strictly cryogenic.

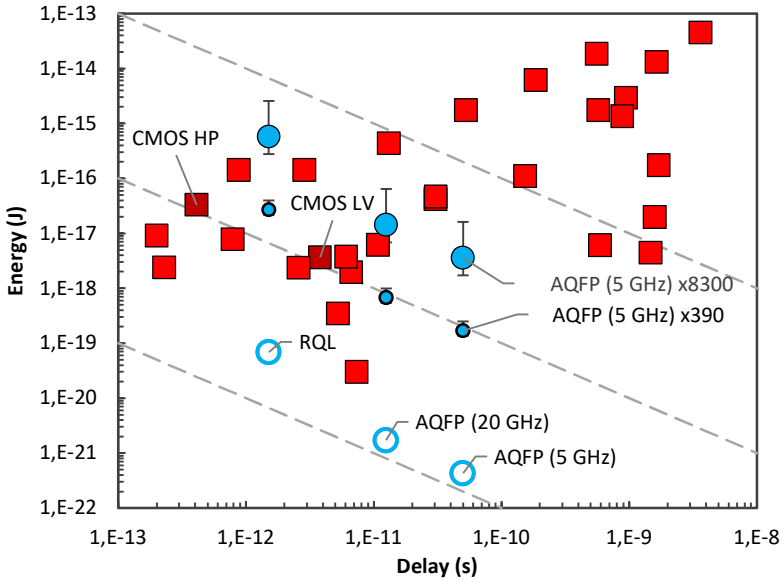


Figure 7-1: Energy versus delay for intrinsic elements.

Several limitations and challenges for the wide implementation of SCE have been predicted (IRDS).

- Customized EDA tools are required, as standard CMOS tools are not adequate. Circuit simulations and timing analysis must be performed for pulse-based signal operation.
- Complete PDKs must be developed for the SCE fabrication processes and related components.
- High yield, reliability and variability Josephson junctions must be achieved to reduce operating margins and improve performance. This becomes especially critical for junction numbers approaching 1 million.
- CMOS integration: The temperature sensistivity of the Al/AlO_x Josephson junctions is the main limiting factor. Thermal budget is here 200C, while it is 400C for CMOS.
- Input/output: For communication with classical systems, high data-rate I/Os are required, which must also be low dissipation. Electrical to optical conversation will be important here.
- Magnetic materials: Magnetic materials are required for SCE fabrication. However, integration of such materials in CMOS is challenging.

7.2 Novel cryogenic materials (IBM)

Cryogenic operation temperature can enable new properties of materials in the class of topological conductors, such as Weyl semimetals (WSM). In these materials, toplogical protection leads to the formation

of linear band dispersion crossing near the Fermi level, giving rise to a range of unusual properties. One such property is extremely large magnetoresistance (MR) at cryogenic temperatures. MR in the order of 1e6% at a few Tesla has been demonstrated experimentally. Leveraging this type of MR, a novel type of cryogenic amplifier was proposed by IBM [7.1].

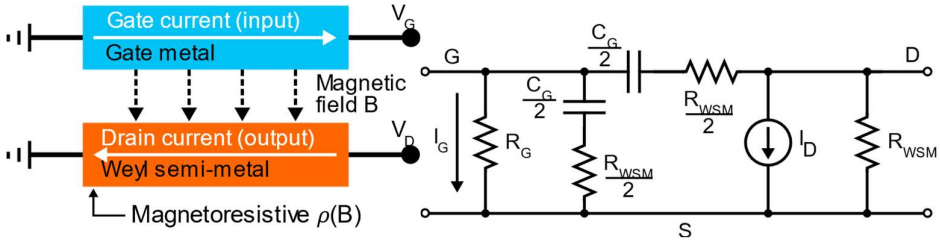


Figure 7-2 Schematic and small signal model of Weyl semimetal-based amplifier [7.1].

The amplifier operates by the generation of a local magnetic field through a local conductor. The magnetic field modulates the conductivity of the WSM through the MR effect. By implementing a superconducting gate electrode, i.e. the magnetic field-generating electrode, the DC power of the device can be extremely low compared to standard HEMTs. The WSM amplifier is similar to an HBT, i.e. has a current-controlled transconductance, but with the channel being essentially a metal. This enables high gain at very low supplied bias. Performance simulations estimate that the WSM amplifier could give about 20 dB gain at less than 50 µW DC power which compared to HEMTs at iso-current is an improvement of around a factor 50.

7.3 Cryogenic HEMTs/LNAs (IAF)

InGaAs channel HEMTs are key enablers for existing quantum computers, providing the lowest RF noise temperatures of all transistor technologies. Cryogenic InGaAs-based HEMT LNAs are used by radio astronomers since many years. In existing quantum computer systems, they are applied as second stage amplifiers at the ~4 K temperature level. As preamplifiers, Josephson parametric amplifiers (JPAs) with a noise temperature of about 1.5 times the quantum noise limit are used. State-of-the-art packaged 4-8 GHz amplifiers distributed by the company Low Noise Factory achieve a noise temperature of 1.5 K, which is about four times the quantum noise limit. These single channel hybrid amplifiers have a power consumption of about 5 mW. The limitation of utilized HEMT technologies and their usage for quantum computer systems can be classified in three categories that are also linked with each other. First, the achievable noise performance of InGaAs-HEMT-based LNAs in relation to the quantum noise limit, second, the power consumption of the each LNA, and third, the scalability of the number of LNA channels.

The limitation of the noise performance of InGaAs-based HEMT LNAs can be divided in effects of the active device and the realization of an LNA. The active device faces two major challenges. The dissipated power of the HEMTs is not only from a system perspective an issue. Also the intrinsic performance of HEMTs strongly depends on the dissipated power and the question of the required bias condition and the corresponding achievable noise performance. As it is exemplarily described in [7.1] (Schleeh et al., “Phonon black-body radiation limit for heat dissipation in electronics”; Nature materials Letters, vol. 14, pp.187, 2014) that the LNA noise temperature decreases linearly with the ambient temperature down to about 40 K, whereas the noise temperature flattens for lower ambient temperatures. This can be explained by the actual temperature of the InGaAs channel. Due to a limited transport of phonons, self-heating of the channel is understood as one of the major limitations of the achievable noise performance. The deviation of the channel temperature in a HEMT compared to the ambient temperature depends even more on the dissipated power density when reducing the ambient temperature. Thus, the lower the ambient temperature is the more important is a reduction of the dissipated power in order to obtain the advantage of a reduced ambient temperature. This fact interacts as well with system aspects. On the one hand, the available cooling power is limited even more

when going down in temperature. On the other hand, it is expected that the number of LNAs increases while scaling up the number of Qubits in a quantum computer, which emphasizes the importance of a low-power operation even further. The ohmic contact of InGaAs-based HEMT technologies is a determining parameter of the corresponding source resistance and should be also considered. While resistors, such as the gate resistance, decrease with the ambient temperature, the resistance of the ohmic contact tends to remain constant or increases. Thus, a low ohmic contact is even more important as compared to room temperature applications.

In addition to the active device, the implementation of an LNA faces a limitation of the noise performance. When comparing the demonstrated noise performance over operating frequency, it can be observed that the noise temperature flattens for low frequencies. This has mainly two reasons – first, for design reasons, the total gate width of HEMT-based LNAs is roughly inverse proportional to the operating frequency. Thus, for a constant bias condition, the power consumption scales with the total gate width. This is, as described before, problematic. A second point is that the loss mainly in the first input matching network are not necessarily scaling over frequency. Especially for low frequencies, matching networks are getting bulky for which reason the losses are considerably high so that RF losses are, beside the active device, a dominating part of the overall noise temperature of an LNA. A hybrid implementation of LNAs, where individual transistor chips are assembled with separately realized low-loss matching networks is a common possibility to counteract the losses of the matching networks at lower frequencies and is currently the state of the art for LNAs used in quantum computers. However, especially under the requirement of a scalability of quantum computers in general and the RF part in special, hybrid LNAs are disadvantages and an MMIC solution is highly preferable. Furthermore, when thinking about heterointegration technologies for quantum computers, the implementation of LNAs as MMICs seems to be inevitable. However, a major challenge in doing so are the losses of matching networks in MMICs.

Consequently, the necessary steps to improve the performance of III-V HEMT-based LNAs for quantum computing systems are a reduction of the noise temperature down to values of JPAs in the order of 1.5 to 2 times the quantum noise limit, a reduction of the power consumption ideally below 1 mW, and scalable MMIC-based implementation. Therefore, the following action items are important:

- Reduction of the LNA power consumption by improved devices and circuit designs
- Improvement of the heat dissipation from the transistor channel
- Reduction of RF losses in the matching networks, e.g., by implementing superconductors as interconnection metallization layers
- Bringing the noise temperature down to the JPA level
- Multi-channel InGaAs-based HEMT LNA MMICs
- Heterointegration of the HEMT technology with Qubits for short RF interconnects and replacing of JPAs

References

[7.1] A Toniato, B Gotsmann, E Lind, CB Zota, Weyl Semi-Metal-Based High-Frequency Amplifiers, 2019 IEEE International Electron Devices Meeting (IEDM), 9.4. 1-9.4. 4.

[7.2] Schleeh et al., “Phonon black-body radiation limit for heat dissipation in electronics”; Nature materials Letters, vol. 14, pp.187, 2014.

8 Thermal environment (LETI)

Regularly in this document, one can be sensitized to the power budget issue, that is linked to the low temperature environment required for quantum computing. That's the reason why we think that an information on the thermal environment is welcome in this document, even it is a side issue in cryo-CMOS challenges.

The efficiency of a refrigeration system is fundamentally limited by the second law of thermodynamics, here in the term of Carnot Coefficient of Performance (COP):

$$\eta \leq \frac{T_C}{T_H - T_C}$$

Where T_C and T_H are lower (cold) temperature and higher (hot) temperature, respectively. From this, an isothermal/isentropic (reversible adiabatic) refrigeration cycle from 300K to 4.2K would lead to η=0,0142. It means we would need 70W of work to maintain 1W of thermal dissipation at 4K. In reality, irreversible losses occur during the process, and the real Coefficient of Performance is far below the Carnot cycle. Hence, 10 to 50kW are spent to get a rough capacity of 1W at 4K, of 1mW at 100mK, of 20µW at 20mK. Nonetheless, dilution refrigeration systems which are used to cooldown to mK range rely on specific properties of He^3 and He^4 : Carnot cycle has no signification anymore in this regime.

8.1 Cryostats and dilution units for complete experimentation (LETI)

Wet baths (vase in liquid phase) down to 4.2K are widely available, and power availability at this temperature is today considered as "open" (there is no upper limitation). For example, *Linde Kryotechnik 'forced-flow helium refrigeration'* systems offer cooling capacities of up to 900 W at 4.4 K, and over 18 kW at 4.5 K for the very large CERN installations [8.1], with a power cost from 250KW to several MW.

At lower temperature, dilution systems are wet or dry chambers, and the power availability is drastically limited. The figure below shows the performance of the *Oxford Instruments TritonXL* dilution unit [8.2]. **0.9mW is expected at 100mK, and 25µW at 20mK**. *CryoConcept* offers dilution units exhibiting 400µW at 100mK, and 10µW at 20mK [8.3]. *Leiden Cryogenics* recently claim the most powerful dilution refrigerator in the world : 1.6mW at 98mK, 2.5mW at 120mK and 20mW at 300 mK [8.11].

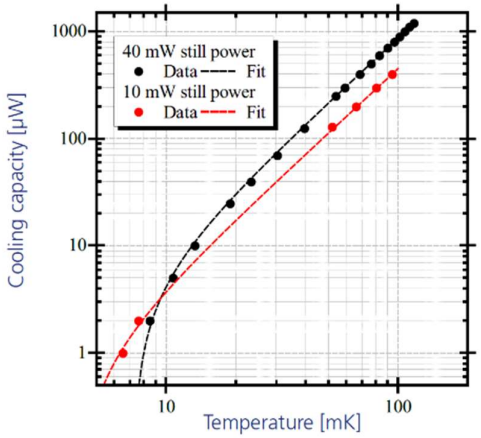


Figure 8-1: Cooling capacity at low temperature (the still is a ^{3}He distillation chamber for flow rate control)

The dilution units take place into a cryostat, managing the cooling from room temperature to a base temperature from 1K to 10K in general. The "4K system" from *Bluefors*, which is able to handle their upper mentioned dilution unit, offers a base temperature at 3.2K, with **0.4W available power at the 4.2K plate** [8.4]. *Cryomech* system exhibits a cooling capacity of 0.5W at 4.2K [8.10].

Additional Pulse Tube Refrigerators (PTR or cryocoolers) may be used in order to boost available power at **4.2K in the range of several Watts**. For example, the *Crymech PT420* reaches **2W at 4.2K** [8.12].

Typical power cost of such complete dilution fridge is in the range of 10kW to 50kW.

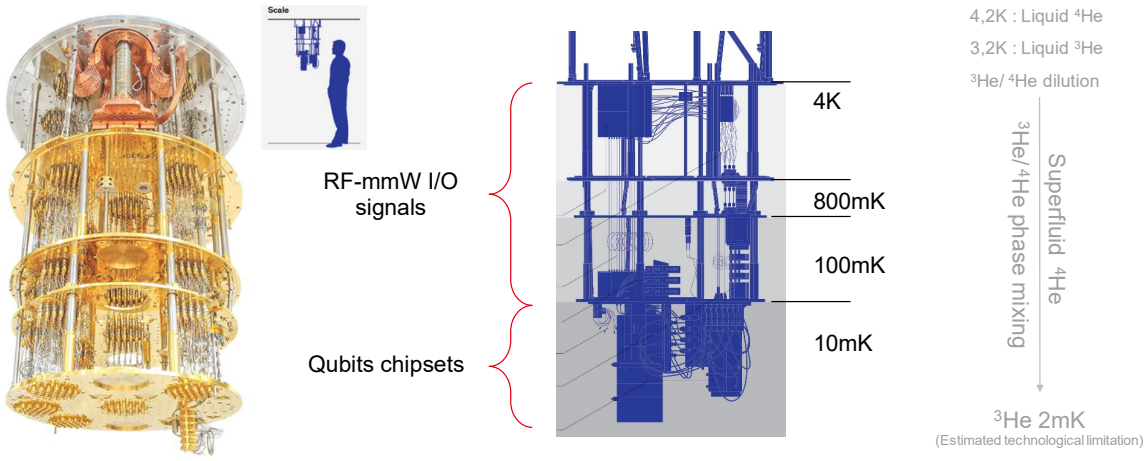


Figure 8-2:Dilution unit from Bluefors (schematic is arranged from [8.5])

It is interesting to notice that necessary time to cool down to qubit temperature is in the range of a day, like shown in the figure below.

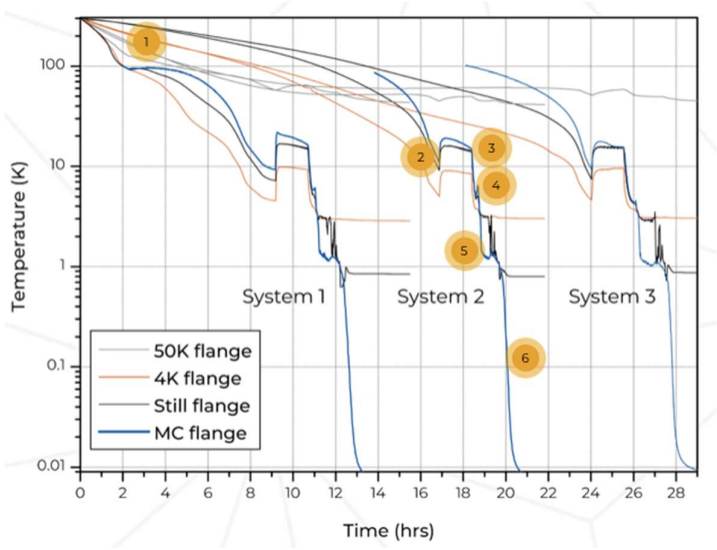


Figure 8-3: cool-down curves for three different configurations of BLUEFORS systems measured with calibrated resistance thermometers.

An installation that would mix a wide Helium bath for 'unlimited' power at 4K, and a dilution unit for lower temperatures, is not available today. That is probably a field of future research, in order to relax the 4K stage in quantum computers. However, like Reilly said in [6.19], even when large quantities of liquid helium are used, the limited thermal conductivity of materials still presents a challenge for heat removal in a complex and heterogeneous system.

8.2 Cryo-CMOS characterization facilities (LETI)

For the characterization of chips at 4K, a simple installation requires to implement the chip on a board with dedicated packaging and interconnects compatible with this temperature. Then, the board is connected at the extremity of a dipstick [8.6], which is plunged into a human-size vase.

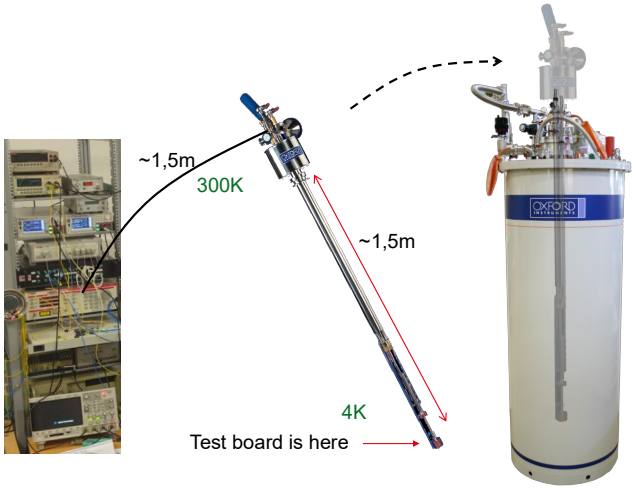


Figure 8-4: Typical installation for chip-on-board characterization.

In such an experience, however, one can assume that **long cables and temperature gradient along the dipstick needs a deembedding procedure**, which is quite challenging when we want to characterize GHz frequency range signals.

For 4K on-probe manual characterization, *Lakeshore* provides solutions [8.7]. Nevertheless, with 25 to 50 mm diameter of probable regions, and limited probe positioning degrees of freedom, this kind of prober is not really suitable for high volume DC characterization. And when RF characterization is performed, in order to prevent from chamber opening during the campaign, most of the probable region is occupied by the calibration kit.

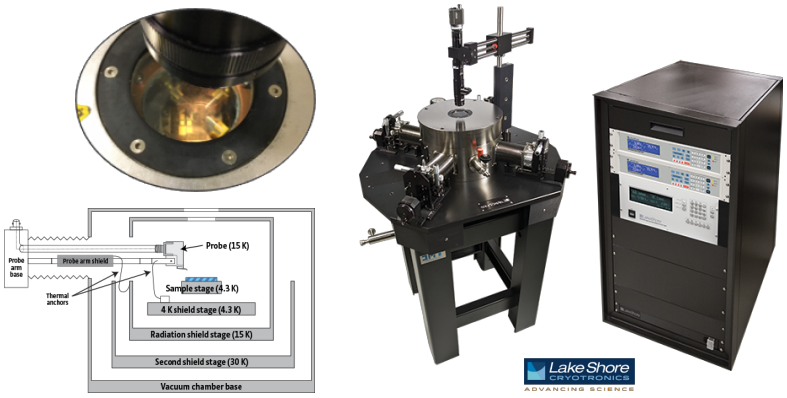


Figure 8-5: Typical installation for manual probing.

Recently, *Bluefors* has developed a fully automatic 300mm probe station, with chuck temperature below 4K, including cassette-to-cassette loading [8.8]. It **efficiently addresses fully automatic DC characterization of low temperature devices**.

Figure 8-6: Bluefors cryogenic automatic 300mm probe station.

However, **RF characterization at low temperature remains challenging up to now**. This kind of measurement requires a suitable calibration phase with available calibration kit in the chamber, single & differential RF-mmW probes with bias tee maintaining signal integrity at low temperatures, and customized DC probe cards for complex biasing and control. According to our knowledge, only a few unitary DC probes and GSG probes are available [8.9].

References

[8.1] https://www.linde-kryotechnik.ch/wp-content/uploads/2016/10/helium_solutions_eng_en.pdf
[8.2] https://nanoscience.oxinst.com/assets/uploads/products/nanoscience/documents/TritonXL_Product_Brochure.pdf
[8.3] http://cryoconcept.com/products/hexa-dry-family/References
[8.4] https://bluefors.com/products/4k-system-cryostat/
[8.5] https://spectrum.ieee.org/computing/hardware/quantum-computers-strive-to-break-out-of-the-lab
[8.6] https://nanoscience.oxinst.com/assets/uploads/products/nanoscience/documents/Brochure_4868_NS_Heliox_Feb2019.pdf
[8.7] https://www.lakeshore.com/products/categories/overview/material-characterization-products/cryogenic-probe-stations/model-crx-4k-cryogenic-probe-station
[8.8] https://bluefors.com/products/the-cryogenic-wafer-prober/
[8.9] https://www.formfactor.com/product/probes/z-probe/z-probe-coaxial/
[8.10] https://www.cryomech.com/products/4k/
[8.11] https://leiden-cryogenics.com/news/113-most-powerful-cf-dilution-refrigerator
[8.12] https://www.cryomech.com/products/pt420/

9 Conclusions (LETI)

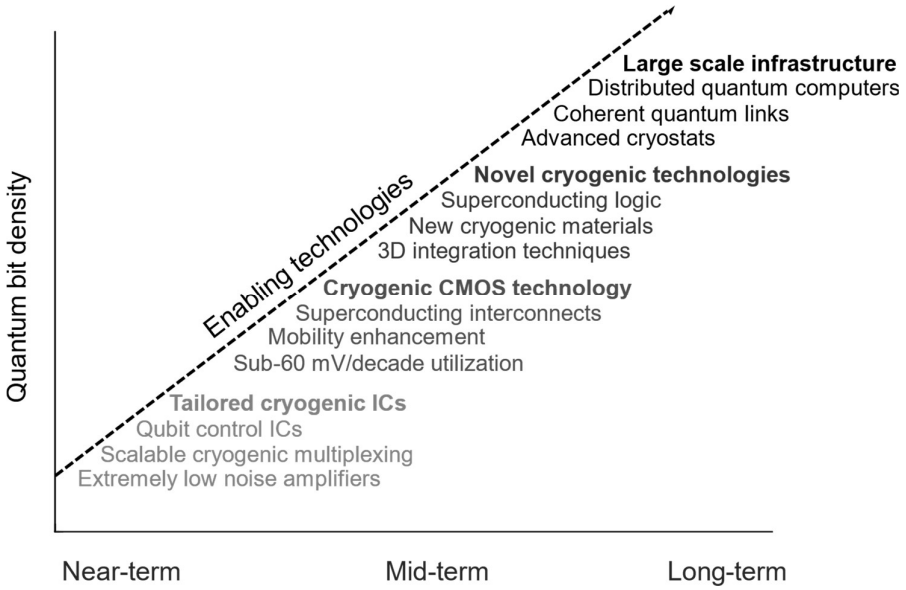


Figure 9-1: Overview for quantun infrastructure electronics supporting the scaling of qubit numbers.

While a handful of qubits are driven, Room temperature high-speed FPGAs and parallel reflectometry authorizes a relatively cost-effective implementation compared to expensive instrumentation. However, one of the main issues in common with all nanoelectronic qubits is that current implementations require at least one external control line for every qubit. In addition, the small pitch of quantum dots permits extremely dense qubit arrays but aggravates the interconnect challenges. We expect that the scaling of interconnections and control lines with the number of qubits will be a central bottleneck in creating large-scale quantum technology.

For few years, several teams around the world are being working on cryo-CMOS designs, and each achievement is an important milestone on the road to a scalable architecture that is able to manage many qubits within a realistic implementation, and power budget.

Until now, one assumes that most of the electronics stands at the 4K stage because of ~1W available power. The use of 50Ω transmission lines for power handling, phase shifts in signals between different lines and interference between branches, synchronicity between signals on thousands of different transmission lines, or time-of-flight latency… all these challenges to overcome for qubit manipulation underlying accurate phase and amplitude of microwave signals, would then be circumvented for the long path between room temperature and 4K.

Nonetheless, referring to the most advanced 'Horse Ridge' dissipating 12 mW/qubit (digital and analog parts) at 4K where power budget is around 1W, it means that we would be able to control "only" 80 qubits…

A part of the community thinks that the link between Room temperature and 4K is not so critical, as soon as most of the signal processing is kept at Room temperature. Furthermore, a road to increase substantially power budget at 4K seems open. In comparison, much more challenging is the multiplexing and demultiplexing that are required to reduce the number of interconnections to and from 20-100mK stage, so as to ensure a large number of channels while minimizing thermal flux.

How to develop such control and readout circuits in the constraint of ultra-low power consumption and ultra-high sensitivity? Probably thanks to smart designs, architectures, and signal processing strategies; but more surely with help of several technology developments like superconducting back-ends, III-V enhancements, III-V on Si implementations, or heterogeneous 3D-TSV packages. And by working on increasing the cooling availability and efficiency so as to relax the so crucial power budget.

Following the study performed in this work, we have assembled an overview for cryogenic quantum infrastructure technologies, shown in Figure 9-1. To enable significant qubit number scaling, we believe some amount of integrated electronics will be required. Considering the availability of around 4 W at the 4K in current cryostats, the electronics must target around 4 mW/qubit dissipation. Adding to this some amount of multiplexing and results already published on cryo control circuits, this seems like a feasible target. Beyond this point, however, with the electronics pushing a need for below 1 mW/qubit, we believe cryogenic technology performance boosters will be required. As we approach qubit numbers above the millions the power budget will be extremely tight and we foresee the need for novel approaches such as new materials and device architectures, in addition to innovations in cooling. With the combination of cryostats increased cooling power at 4 K, multiplexing, efficient readout schemes and tailored cryogenic CMOS, a path towards such qubit numbers can, however, be discerned.