Cryogenic CMOS for Qubit Control and Readout

Pellerano, Stefano; Subramanian, Sushil ; Park, Jong-Seok ; Patra, Bishnu ; Xue, Xiao ; Vandersypen, Lieven M.K.; Babaie, Masoud; Charbon, Edoardo; Sebastiano, Fabio

DOI
10.1109/CICC53496.2022.9772841

Publication date
2022

Document Version
Final published version

Published in
2022 IEEE Custom Integrated Circuits Conference, CICC 2022 - Proceedings

Cryogenic CMOS for Qubit Control and Readout

Stefano Pellerano[1], Sushil Subramanian[1], Jong-Seok Park[1], Bishnu Patra[1], Todor Mladenov[1], Xiao Xue[2], Lieven M. K. Vandersypen[2], Masoud Babaie[2], Edoardo Charbon[3], Fabio Sebastiano[2]

[1]Intel Corporation, OR, USA, [2]Delft University of Technology, The Netherlands, [3]EPFL, Lausanne, Switzerland

Introduction

Quantum computers have been heralded as a novel paradigm for the solution of today's intractable problems, whereas the core principles of quantum computation are superposition, entanglement and interference, three fundamental properties of quantum mechanics [1]. A quantum computer generally comprises a quantum processor, made of an array of quantum bits or qubits, and a classical controller, which is used to control and read out the qubits. Quantum algorithms are generally mapped onto a circuit of quantum gates that operate on multiple qubits. Unlike conventional digital bits, qubits can take a coherent state ranging from $|0\rangle$ to $|1\rangle$ on a continuous sphere, known as the Bloch Sphere and they are implemented based on several mechanisms. While many solid-state implementations of qubits exist, an exhaustive description of available technologies is beyond the scope of this paper [2][3].

To execute a quantum algorithm, analog/RF signals are required to manipulate the state of individual qubits and entangle multiple qubits. Additional control signals are required to perform qubit state readout. These signals are generated by the classical controller, which is generally operated at room temperature (RT), while qubits must be cooled to cryogenic temperature (CT), usually from few tens of milli-Kelvin to a few Kelvin, to preserve their coherence. The resulting large temperature gradient must be handled to enable a direct electrical interconnection, while preventing energy transfer from RT to CT. To keep phonon flow to a minimum, while ensuring fast and compact electrical connections, multiple thermalization stages are used, where each stage attenuates control or amplifies readout signals. The goal is to match the impedance of each signal generator and only generate the noise equivalent to that temperature. An impedance of 50Ω is typically used to easily adapt existing RT instrumentation. This approach has worked well with <100 qubits. However, with an increasing size of quantum processors, there has been a need of moving classical control to lower temperatures, close, or ideally identical, to those of the qubits, to enable scalable, compact, and reliable machines in the future. An additional concern is the capability of such a system to be quickly debugged and upgraded, something hard to implement in a RT approach.

To address these issues, in 2016 we have proposed the use of cryogenic electronics, and in particular, cryogenic CMOS, or cryo-CMOS, as a good candidate to accomplish the tasks of a generic classical controller [4]. With the lack of thermalization requirements, cryo-CMOS electronics can generate control signals at an amplitude far lower than if they were generated at RT. In addition, readout circuits can also benefit from low thermal noise, while the proximity in space and temperature to qubits can reduce the complexity of interconnect, where superconductors can be used to maximize thermal isolation with virtually zero electrical losses. With high-temperature qubits [5]-[7], this configuration could be simplified further, eliminating the requirement for advanced refrigeration and milli-Kelvin temperatures.

The main limitation of the cryo-CMOS controller is the power of classical circuits, which must be within the limits of thermal absorption by the refrigeration unit. Non-CMOS devices, such as high-electron-mobility transistors (HEMTs), SiGe heterojunction bipolar transistors (HBTs), GaAs logic, and rapid single-flux quantum (RSFQ) circuits have been proposed to achieve high-performance circuit functionality of the type needed in the classical controller [8]-[10]. However, none of these technologies, except perhaps HBTs, can take advantage of 60 years of industrial development aimed at scalability, like CMOS [5][11][12]. An advanced CMOS process, Intel 22nm FinFET technology, was selected in this work to implement complex functionality at low power.

The classical control systems proposed in this paper are part of a larger context and require therefore specific interfaces to the quantum stack that supports quantum algorithms. The quantum stack in Fig. 1 is the ensemble of architectural components designed to translate such algorithms to quantum circuits and finally to quantum control via assembly languages, like QASM. Such architectures are complex and significant development is ongoing [13].

Fig. 1. Quantum computing stack. Credit: Harald Homulle, 2016.

In this paper, we first give a brief overview of the current state of the art in cryogenic electronics for qubit control and readout. We then describe the architecture, circuit design and measurement results of two cryogenic SoCs, code-named "*Horse Ridge*", designed to operate at 4K and implemented in Intel 22nm FinFET technology. Horse Ridge 1 is designed for state manipulation of spin qubits and transmons. Coherent qubit control, randomized benchmarking and two-qubit algorithm are demonstrated with a 2-qubit quantum processor. Horse Ridge 2 is targeted at spin qubits and can implement all the required control functions of drive, readout and gate bias pulsing. Its performance is fully characterized at 4K. We then conclude the paper with some considerations on future directions.

State of the Art Review

To demonstrate the reliability and performance of CMOS technology for realizing complex cryogenic integrated circuits, several building blocks have been published over the past few years [14]-[32]. However, this section only focuses on the prior art that demonstrated promising performance at several hundreds of millikelvin or the ability to directly control and readout qubits.

The first integrated implementation of a cryogenic pulse modulator was presented in [15], where a pair of current-mode digital-to-analog converters (DAC) controlled by a waveform memory and a single-sideband upconverter are employed to manipulate the state of superconducting qubits. Although the pulse modulator consumes <2mW/qubit, its scalability is limited by the need for one high-frequency local oscillator (LO) per qubit, and the use of fixed-length symmetric pulses limits its compatibility with other qubit technologies. To alleviate these issues, [29] proposed a rotational phase accumulator in the digital domain to generate a microwave pulse with any arbitrary duration, phase, and frequency. This work also proposed a sinusoid shaping nonlinear DAC followed by a linear interpolating DAC to improve the controller power efficiency. However, the scalability of this structure is limited since it can only track the phase of one qubit at a time.

The work in [25] reported a switched-capacitor-based Cryo-CMOS controller in a 28nm FDSOI CMOS technology that operates at 100mK and can generate static and dynamic voltages for the parallel control of qubits. To generate the required static voltages of qubits electrodes, the output of a room-temperature DAC is multiplexed by using capacitive sample-and-hold circuits periodically refreshed to counteract the charge leakage of the hold capacitors. The required dynamic voltage pulses are then implemented by toggling the potential of the bottom plate of the hold capacitors with two fixed external voltage sources. The reported average power consumption for generating voltage pulses of 100mV is about 18nW per cell. Consequently, this structure can potentially be scaled up to generate voltage for a few tens of thousands of electrodes in a commercially available dilution refrigerator.

To demonstrate the viability of various functions down to 110mK, [20] presented a Quantum Integrated Circuit (QIC) in a 28nm FDSOI

2022 IEEE Custom Integrated Circuits Conference (CICC) | 978-1-6654-0756-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/CICC53496.2022.9772841

978-1-6654-0756-4/22/$31.00 ©2022 IEEE

CMOS technology. The QIC chip comprises a double quantum dot (QDUT) with digital and analog blocks for QDUT biasing, manipulation, and read-out. This work successfully demonstrated a quantum effect called “charge pumping” by applying a 2.8GHz signal from an on-chip modulator to the QDUT and measuring its current through an on-chip trans-impedance amplifier (TIA) while sweeping the voltages of the plunger gates.

In [26], a fully-integrated cryo-CMOS I/Q receiver is proposed integrating frequency synthesis, low-noise amplifiers, Gilbert cell active mixers, and polyphase filters to perform phase-sensitive readout of qubit signals. The receiver, operating at frequencies of 4.5 to 8.5GHz, is based on a non-zero IF down-conversion with >20dB image-rejection ratio (IRR). At a temperature of 3.5K, it achieves a conversion gain up to 70dB with an effective noise equivalent temperature of 39.2K, IIP3>-72dBm, and P1dBin>-85dBm. The proposed receiver supports up to 70 qubits in frequency-division multiple-access (FDMA), dissipating 1.5mW/qubit.

The work in [27] proposed a cryogenic receiver for gate-based RF readout of spin qubits, achieving a 58dB gain and a 45K effective noise temperature at 4.2K while consuming 0.17mW/qubit. Compared to the prior-art cryogenic DC readout IC in [20], this work supports FDMA for scalability, 1000× faster readout, with 5× lower energy consumption. However, in order not to degrade the inherent signal-to-noise ratio (SNR) of the qubit signals, a superconducting parametric amplifier with a near quantum-limit noise temperature (T_n~0.4K) or a HEMT LNA (T_n~2.2K) must be placed prior to the receiver to amplify the signal by >20dB.

The work in [31] proposed a single-pole-4-throw (SP4T) reflective Cryo-CMOS multiplexer using standard 28nm CMOS technology to reduce the number of required output RF lines from the base temperature stage of a dilution refrigerator to the control/readout electronics placed at 4K stage. The multiplexer is controlled with a parallel or serial interface and exhibits DC-to-10GHz bandwidth, 1.6dB insertion loss and 34dB isolation at 6GHz, <2ns switching time and 36.2μW heat dissipation while operating at 32mK. This work can improve cryogenic characterization throughput and alleviate the wiring problem in future large-scale quantum computing systems.

In [33], dispersive readout of the charge state is demonstrated on an array of 3×3 quantum dots, accessed by a combination of FDMA with resonators in the 6-8GHz range and time-division multiple-access (TDMA) with MOS switches. The chip is fully integrated in a 40nm standard CMOS technology and operates at 50mK.

Horse Ridge 1: Qubit Drive

Horse Ridge 1 (HR1) is our first-generation control SoC, which implements the function of qubit drive: the generation of RF amplitude- and phase-modulated pulses to manipulate the state of qubits [34]. The controller integrates four identical transmitters, each capable of controlling 32 frequency-multiplexed single-electron spin qubits or transmons in the frequency range from 2 to 20GHz, for a total of 128 qubits over 4 RF cables. Each transmitter has been designed following the methodology presented in [35] to achieve a 99.99% fidelity, with both an SNR (25MHz bandwidth) and a spurious-free dynamic range (SFDR) better than 44dB. Figure 2 shows the simplified block diagram of one of the transmitters, integrating an instruction controller, a multi-channel direct digital synthesizer (DDS) and an analog/RF front-end.

The digital is clocked at 1GHz and implements two polar modulators (bank 1 & 2). The 32 numerically controlled oscillators (NCOs) are used for coherent control of 32 qubits with a single LO. The 22-bit frequency control word enables 0.2kHz resolution and a 12-bit phase word is used to adjust the NCO phase and implement Z-axis qubit rotations [35]. Phase and amplitude are modulated by a 10-bit and 8-bit sequence, respectively. These sequences are stored in the on-chip memory and calibrated for each qubit to implement specific instruction sets. *Sin(x)* and *Cos(x)* functions are generated through look-up tables (LUT) to reduce power consumption. Signal bit widths have been defined by system simulations to target 99.999% fidelity, 10dB better than the overall controller target [35]. Digital IQ and offset compensation networks (α, β and γ in Fig. 2) are implemented for image and LO leakage suppression. When used at the same time, the two identical banks allow for simultaneous control of two qubits.

Two high-speed 10-bit (5-bit thermometer + 5-bit binary) current-steering DACs followed by a 2nd order GmC reconstruction filter convert the IQ digital modulated pulses into analog signals. Current bleeding in the DAC unit cell [36] and current-mode filtering enable high-linearity operation. The VGA is implemented as a 4-bit programmable current mirror, with a single-stage amplifier to boost the current mirror loop gain for increased linearity. A double-balanced Gilbert cell IQ mixer with current bleeding is used to upconvert the baseband pulse to the qubit Larmor frequency. Two

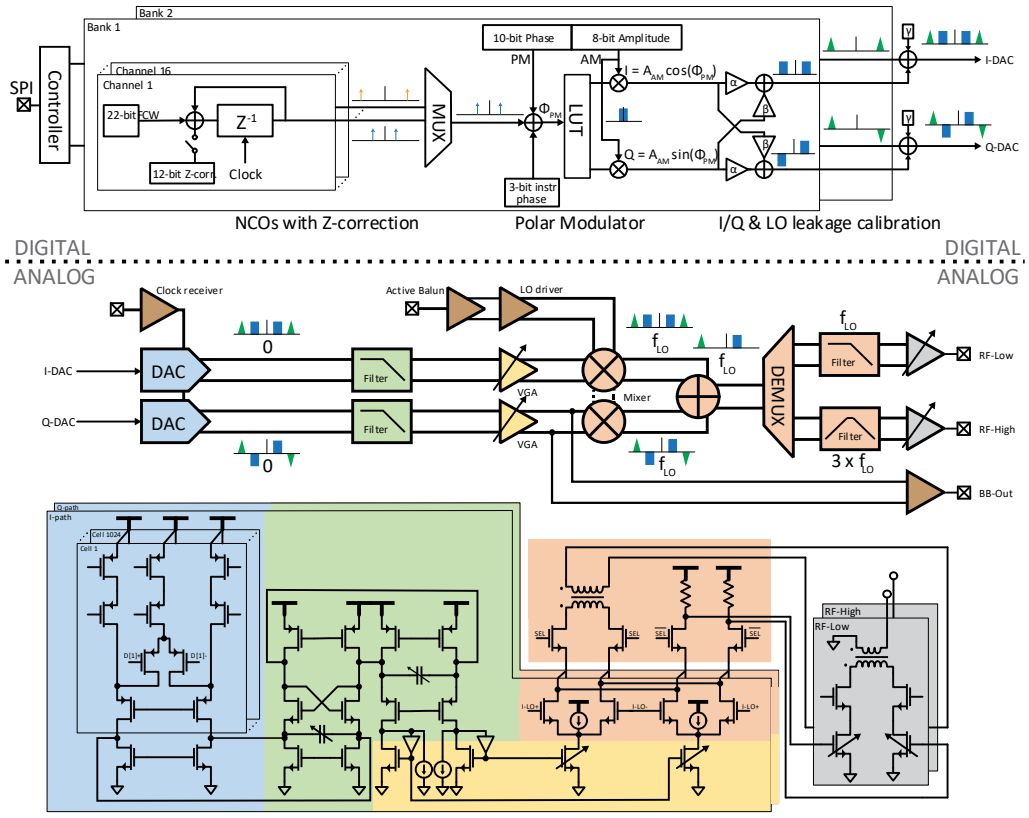
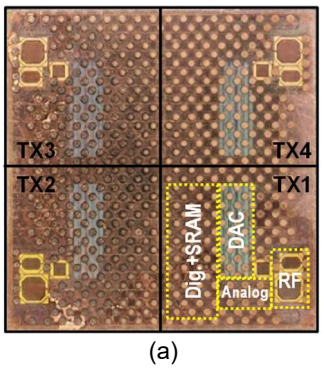
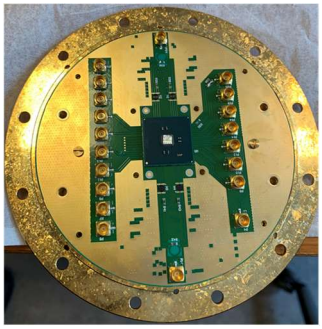


Fig. 2. Simplified block diagram of one of the four identical transmitters integrated in HR1 SoC.



(a)



(b)

Fig. 3. (a) HR1 die and (b) FCBGA package assembled on cryogenic PCB.

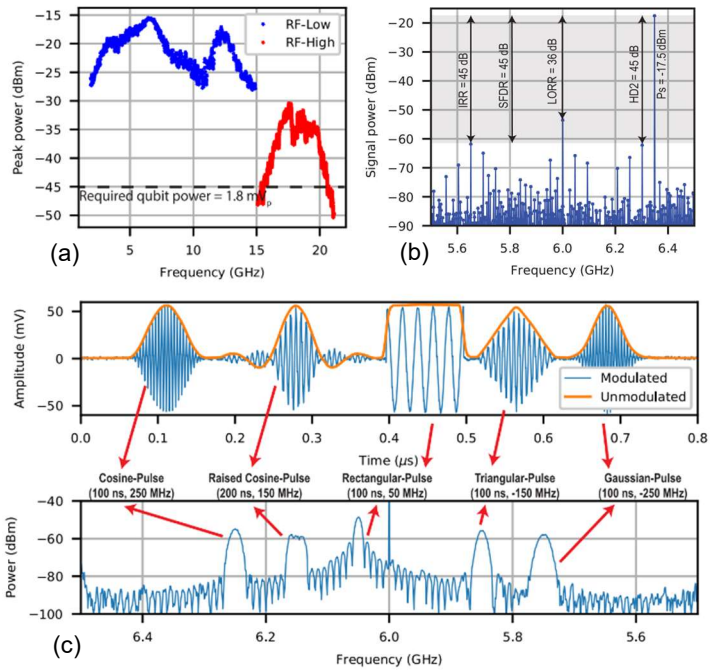


Fig. 4. (a) RF bandwidth. (b) CW SFDR at 6.35GHz. (c) Time-domain signal at baseband (top) and upconverted output spectrum (bottom).

operating modes are used to cover the 2-20GHz RF output bandwidth, selected by a mux embedded in the mixer output with cascode devices. In the low-band mode (2-15GHz), the fundamental harmonic of the LO is used with the mux steering the mixer output current into a resistive load. For the high-band (15-20GHz), the 3rd harmonic of the LO is used and the mixer current is steered into a resonant load. Two separate class-A amplifiers with wideband transformer-based matching networks are used to drive the 50Ω cables to the qubits. With 15 switchable unit cells, ~24dB of power control is achieved.

The 2-15GHz LO signal is provided externally. Wilkinson power dividers and discrete wire-bonded quadrature hybrids are used on the cryogenic PCB to generate multiple single-ended IQ signals. Single-ended to differential LO drivers with 20dB voltage gain and 15GHz bandwidth are integrated on-chip to deliver the required voltage swing to the mixers. The target phase noise of -116dBc/Hz at 1MHz offset [35] is achieved over the entire frequency range with ~7mW power for both IQ paths. The 1GHz clock for the DACs and digital baseband is also provided externally. This is shared across the 4 transmitters, each using a 200Ω on-chip termination. The analog/RF front-end receives the external clock and distributes it to the DACs and digital baseband after proper synchronization. All bias currents are generated by a constant-gm block, tunable from 0.5× to 2× the nominal value to account for the shift in biasing points between 300K and 4K. An external bias current can also be applied directly.

HR1 die is fabricated in Intel 22nm FinFET process and occupies an area of 16mm^2 for 4 transmitters [Fig. 3(a)]. A custom-made annealed and gold-plated copper enclosure shown in Fig. 3(b) is used for efficient thermalization when the PCB is installed in the dilution refrigerator.

The measured peak output power at 4K is shown in Fig. 4(a). Flatness is limited by additional ground inductance erroneously introduced in the layout. The SFDR at peak output power is >45dB, limited by HD2 and IRR [Fig. 4(b)]. The LO rejection ratio (LORR) after calibration is 36dB. This is not a limitation for qubit control since the LO leakage can be placed far away from any qubit by proper choice of LO and NCO frequencies. Measured IM3 and SNR (25MHz bandwidth) are better than 50dB and 48dB respectively [34].

The amplitude and phase modulation information for pulse shaping is stored in on-chip SRAM, allowing envelopes up to 41μs, which are referenced by a look-up table (LUT) that can define 8 instructions per qubit, each with a phase offset to efficiently reuse envelopes for different rotation axes. Figure 4(c) demonstrates the versatility of this design by showing the measurement of an instruction sequence containing 5 unique waveforms targeting 5 different qubit frequencies. At 1GHz clock frequency, the power consumption is 330mW and 54mW for the digital and analog/RF, respectively. The digital power consumption can be substantially reduced by design optimization and clock-gating. Nevertheless, thanks to the frequency multiplexing architecture, the total power consumption of 12mW/qubit would still allow the control of >320 qubits in a state-of-the-art dilution refrigerator over just 10 RF cables.

Extensive testing of HR1 with spin qubits is conducted [37]. We used the 2-qubit quantum processor shown in Fig. 5(a). First, to demonstrate coherent qubit control, Rabi oscillations are measured. As shown in Fig. 5(b), the qubit is initialized to ground state and then an RF pulse of duration *t* is generated by HR1 at the qubit Larmor frequency (13.4GHz for this experiment) and applied to the gate MW in Fig. 5(a). The qubit state is then measured. Figure 5(c) shows that as the pulse duration *t* is increased the measured state oscillates between |0⟩ and |1⟩, as the qubit state is rotated along the Bloch sphere, demonstrating coherent qubit control. The limited amplitude of the measured Rabi oscillations shown in Fig. 5(c) is due to the non-zero error probability in the readout process. To quantify the fidelity of a single-qubit gate operation and compare with RT control electronics, a randomized benchmarking experiment is performed [37]. A random set of qubit rotations ["*Clifford Operations*" in Fig. 5(d)] is executed and the error in the final qubit state is measured as a function of the number of consecutive operations. Results show that within the measurement error margin, HR1 achieves the same fidelity

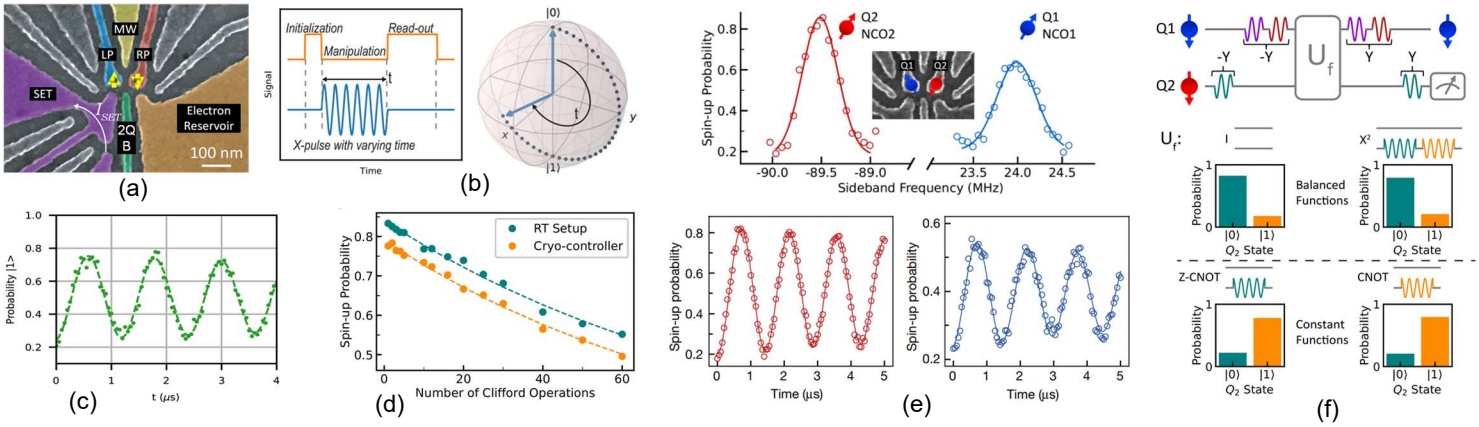


Fig. 5. (a) Two-qubit processor die micrograph. Two qubits are formed under the LP and RP gates. RF pulses are applied to the MW gate and 2Q-B gate is used to control the qubit-to-qubit interaction. A charge sensor based on a single electron transistor (SET) is used for readout. (b) Rabi oscillation experiment and (c) measured result. (d) Randomized benchmarking experiment, showing same fidelity between HR1 and conventional room temperature (RT) electronics. The two curves are offset by 0.05 along the Y axis for better visibility. (e) Simultaneous coherent control of 2 qubits with two-tone pulse. (f) Pulse sequence and measured results for the Deutsch-Josza 2-qubit algorithm.

as the RT control electronics at 99.7%, limited by the qubit sample. The two qubits can be simultaneously controlled by HR1 through frequency multiplexing. As shown in Fig. 5(e), simultaneous coherent control is demonstrated by using two tones spaced by about 100MHz around a 13.5GHz LO frequency, corresponding to the two qubit resonance frequencies. Two-qubit operations can also be implemented. Once the potential barrier between the two qubits is lowered [2Q-B gate in Fig. 5(a)], due to reciprocal interaction each qubit resonance frequency splits in two (by ~10MHz for this experiment), depending on the state of the other qubit. The controlled-NOT (CNOT) gate is the quantum equivalent of an XOR gate: the target qubit state is flipped only if the control qubit state is $|1\rangle$. If f_0 and f_1 correspond to the resonance frequency of the target qubit when the control qubit state is $|0\rangle$ or $|1\rangle$ respectively, then by pulsing at f_1 a CNOT is implemented (with some residual phase rotation on the control qubit), since the pulse would be at the target qubit resonance and hence able to rotate it only if the control qubit state is $|1\rangle$. To perform a single-qubit operation on the target qubit, two consecutive pulses at f_0 and f_1 are generated, so that the target qubit is rotated regardless the state of the control qubit. Leveraging these single- and two-qubit operations, HR1 is programmed to execute the 2-qubit Deutsch-Josza algorithm, which determines if a function gives a constant or balanced outcome [38]. Figure 5(f) shows the pulse sequence (Q1 is the input qubit and Q2 is the output qubit) and the measured results for the 4 different implementations of the function U_f. Q2 final state correctly predicts if U_f is constant or balanced with about 80% probability. This experiment highlights the ability of HR1 to perform arbitrary sequences of operations leveraging frequency multiplexing and DDS-based frequency generation.

Horse Ridge 2: Qubit Drive, Readout and Gate Pulsing

Horse Ridge 2 (HR2) is the second-generation cryogenic controller SoC. It extends the capabilities of HR1 by adding qubit state readout and gate voltage pulsing functionality [39]. Figure 6 shows HR2 system architecture. The qubit drive is similar to HR1: 16 NCOs shared across two polar modulators allow for a frequency multiplexed control of up to 16 qubits, with simultaneous pulsing of any pair. To reduce memory usage, this implementation shares a single envelope template that is scaled by 2^{12} 16-bit coefficients for each qubit to implement different rotations, each with 2^{12} additional 22-bit phase settings for each NCO. This corresponds to 2^{19} programmable codewords per qubit. As for HR1, NCO phase update is also used for rotation along the Z-axis for correction of unwanted Z-rotations due to AC-Stark shift induced on adjacent qubits by frequency multiplexing [35]. To further improve fidelity, a digital FIR is integrated to create two notches in each pulse power spectral density at programmable frequencies for each qubit to suppress energy leakage to adjacent qubits. The SoC also integrates a microcontroller for increased flexibility in implementing the instruction set. The digital and DACs speed has been increased to 2.5GHz with respect to HR1 to extend the complex signal bandwidth to 2GHz. For the 4K characterization, a clock of 1.6GHz has been used to reduce power consumption.

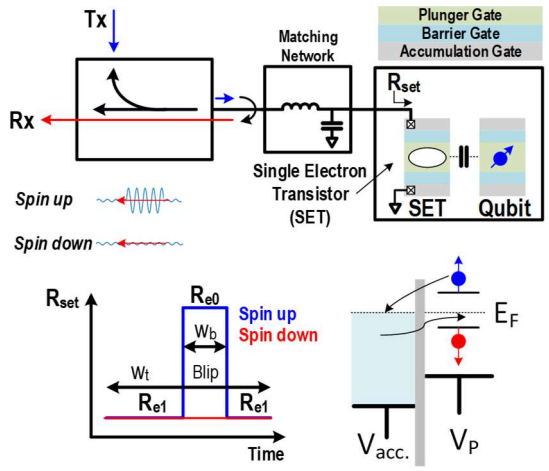


Fig. 7. Basic principles of qubit state readout based on RF reflectometry and spin-to-charge conversion. The SET coupled to the qubit dot is used as charge sensor.

To read the qubit state, spin-to-charge conversion is adopted [40]. As shown in Fig. 7, the qubit dot plunger gate potential V_P is adjusted such that the Fermi energy level (E_f) of the nearby electron reservoir (metal contact) is between the qubit spin-up and spin-down energy levels. Only if the qubit is in the spin-up state, the electron energy is sufficient to tunnel through the barrier into the reservoir and a spin-down electron successively tunnels back to the dot. This charge

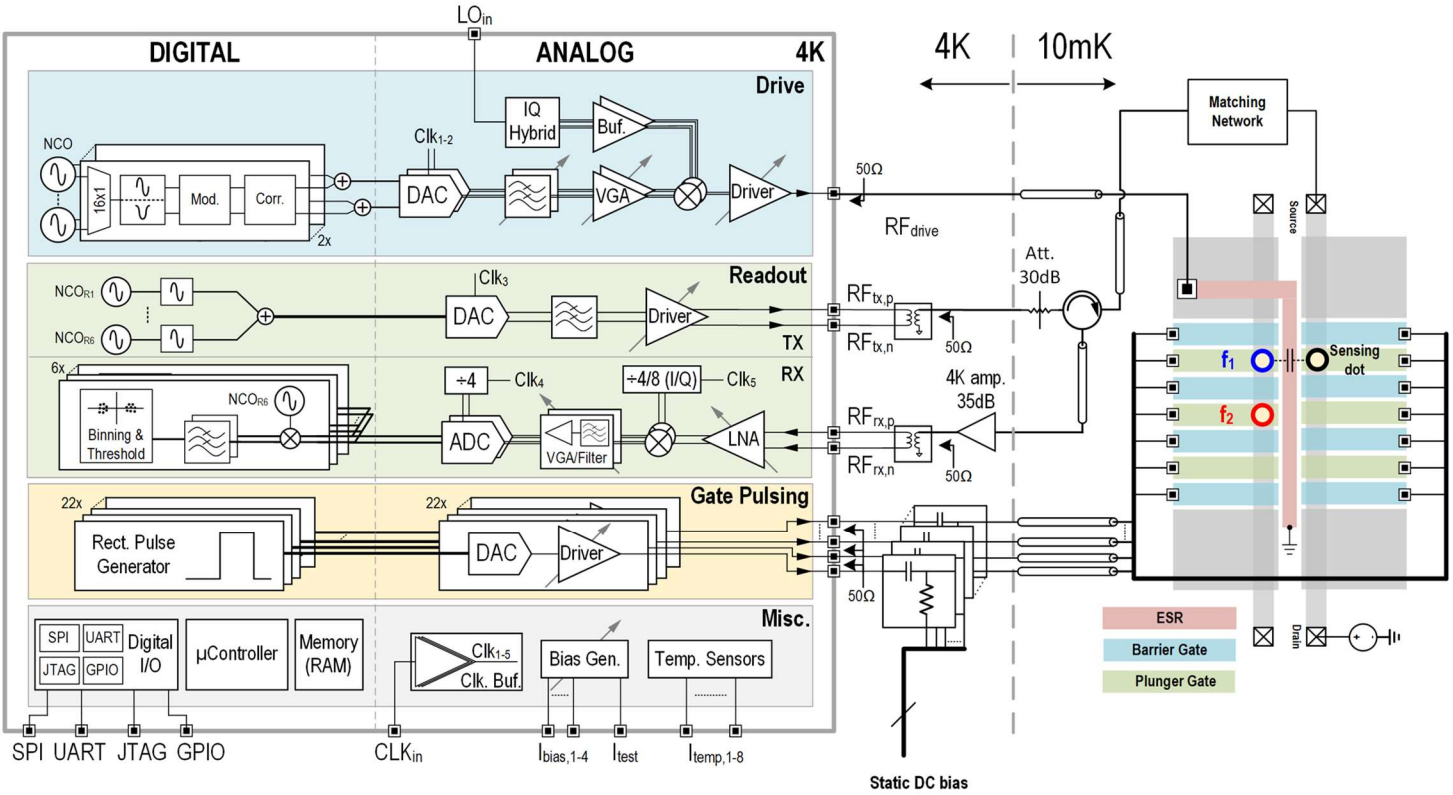


Fig. 6. Simplified block diagram of HR2 SoC with intended connections to the spin qubit chip at 10mK

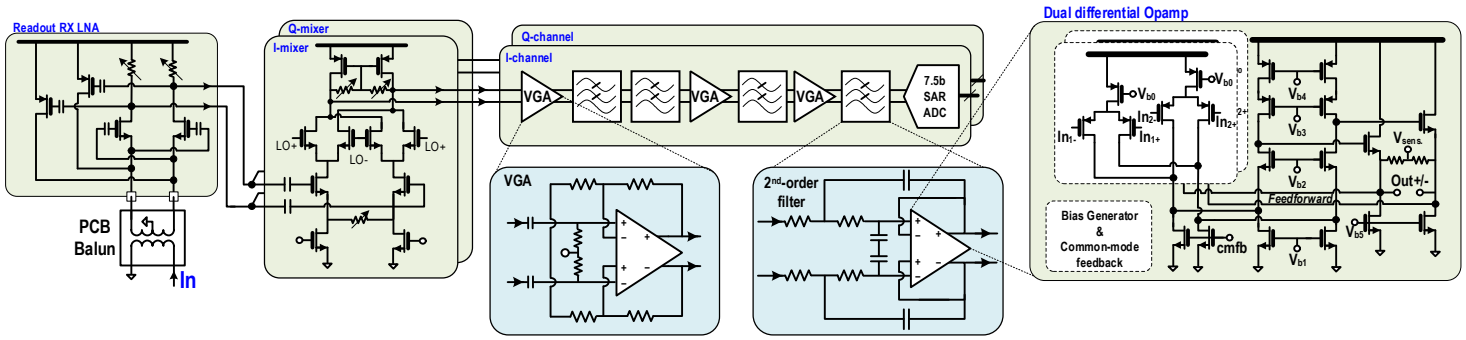


Fig. 8. Simplified schematic of the receiver for qubit state readout.

displacement can be detected by a nearby charge sensor implemented with a single electron transistor (SET in Fig. 7). The SET gate is capacitively coupled to the qubit dot such that a charge displacement in the qubit dot modulates the SET gate bias. As shown in Fig. 7, this results in a modulation of the SET channel resistance from R_{e1} to R_{e0} and back to R_{e1} as the qubit electron tunnels to the reservoir and then back to the dot. This impedance modulation can be sensed by RF reflectometry: by detecting the amplitude and/or phase modulation of the reflected RF pulse, the spin-up qubit state can be detected.

To implement RF reflectometry readout, a multi-tone signal generator (TX) and an IQ receiver (RX) are integrated (Fig. 6). The TX uses a digital-to-RF architecture to avoid LO and other spurs introduced by an upconversion mixer which would be reflected by the SET matching network and saturate the receiver. Six NCOs are used to generate a multi-tone signal so that up to 6 qubits can be read simultaneously by frequency multiplexing (Fig. 6). The same 10-bit DAC is reused from the drive, together with the GmC filter. A high-linearity variable attenuator delivers from -70 to -40dBm power into a 50Ω load, so that a signal with <-70dBm power and ~100mK noise temperature can be delivered to the SET through an additional 30dB external attenuator placed at the 10mK stage (Fig. 6). The total TX power consumption is 7.6mW. To detect the reflected signal (~ -90dBm at the receiver input), the readout receiver is designed for >90dB of gain. An IQ downconversion mixer is used to limit the ADC bandwidth and avoid instability by partitioning the receiver gain between RF and baseband stages. The LO signal is generated by integer division of the system clock so that all clock harmonics fall either at DC or out of band, where they are suppressed by AC-coupling and low-pass filtering. Due to the intrinsic full-duplex operation of the readout, TX/RX isolation is essential, with more than 60dB required for this implementation. By leveraging differential signaling with discrete baluns integrated on the PCB and TX/RX floorplan separation with reduced conductivity of the substrate at cryogenic temperature, more than 80dB isolation has been measured at 4K. The readout receiver is shown in Fig. 8. The LNA uses dual-feedback [41] to achieve 0.1-0.6GHz wideband operation with a simulated NF and gain of <3dB and 27dB respectively at 300K. To suppress spurs, an 8th-order adjustable low-pass filter with VGA is implemented after the double-balanced IQ mixer. Fully-differential op-amps with dual-differential inputs and feedforward compensation are used in both filters and VGAs (Fig. 8). The baseband IQ signal is then sampled by two 7.5-bit 400MS/s SAR ADCs. The receiver bandwidth can be adjusted from 60 to 200MHz with 40 to 90dB total gain. The total power consumption of the receiver is 38mW.

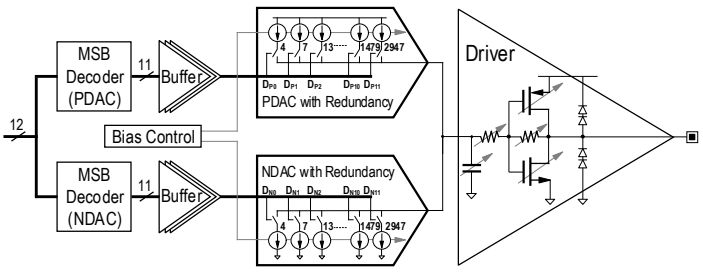


Fig. 9. Simplified schematic of the high-speed qubit gate pulsing circuit.

For a linear array of 7 qubits and respective SETs, 22 DACs are integrated to control all the gate potentials. Each DAC is composed of two 11-bit pMOS and nMOS current source banks to support both positive and negative pulses while saving power when idling (Fig. 9). To reduce area, a <2 radix is used in each DAC with a look-up table (LUT) to map the 11-bit code to the redundant 12-bit input. A 50Ω output driver for each DAC is used to absorb pulse reflections from the open termination at the qubit gates. Since we use square pulses, each DAC code is only updated twice for each pulse. This also saves power in the LUTs. All 22 DACs can be pulsed simultaneously with different amplitudes, which is critical to cancel crosstalk due to capacitive coupling among the gates on the qubit chip. Each DAC static power consumption is 2.3mW, limited by the 50Ω output impedance of the driver.

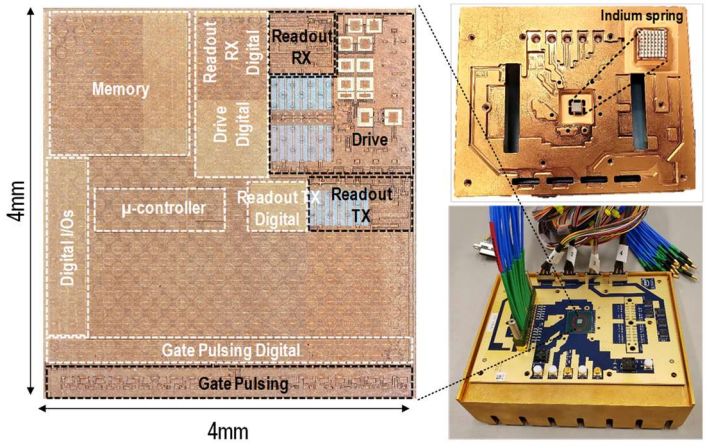


Fig. 10. HR2 SoC and cryogenic PCB with copper enclosure for thermalization. Indium springs for back-side silicon thermal contact are visible on the top part of the enclosure.

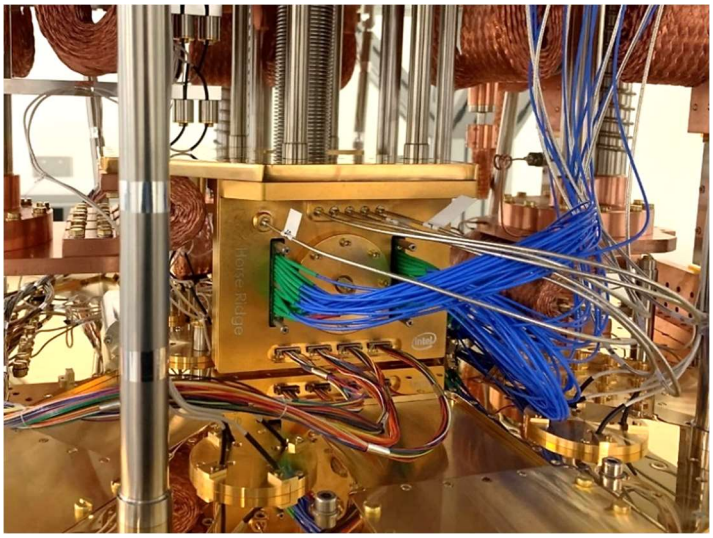


Fig. 11. Horse Ridge 2 cryogenic controller assembled in the copper enclosures and installed in the 4K plate of the dilution refrigerator.

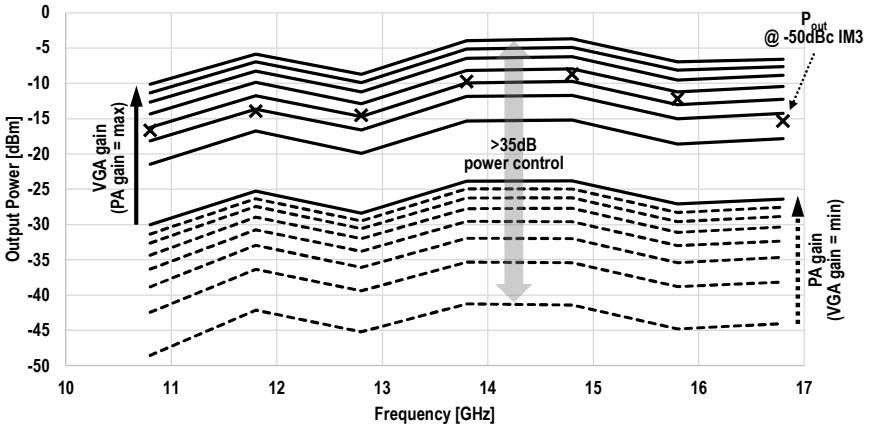


Fig. 12. RF output power of qubit drive pulse generator in CW mode, showing >35dB power control and better than -15dBm power for IM3<-50dBc

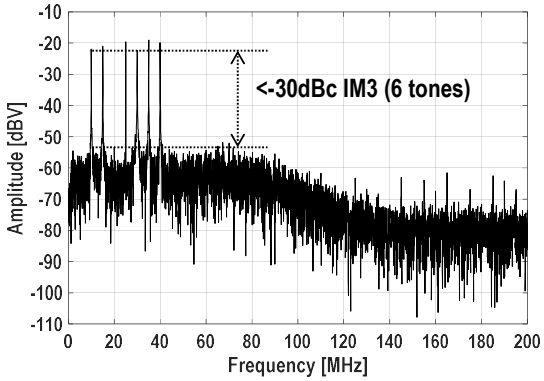


Fig. 13. Readout receiver linearity test with 6 tones generated by the readout transmitter, showing the FFT of the ADC sampled signal.

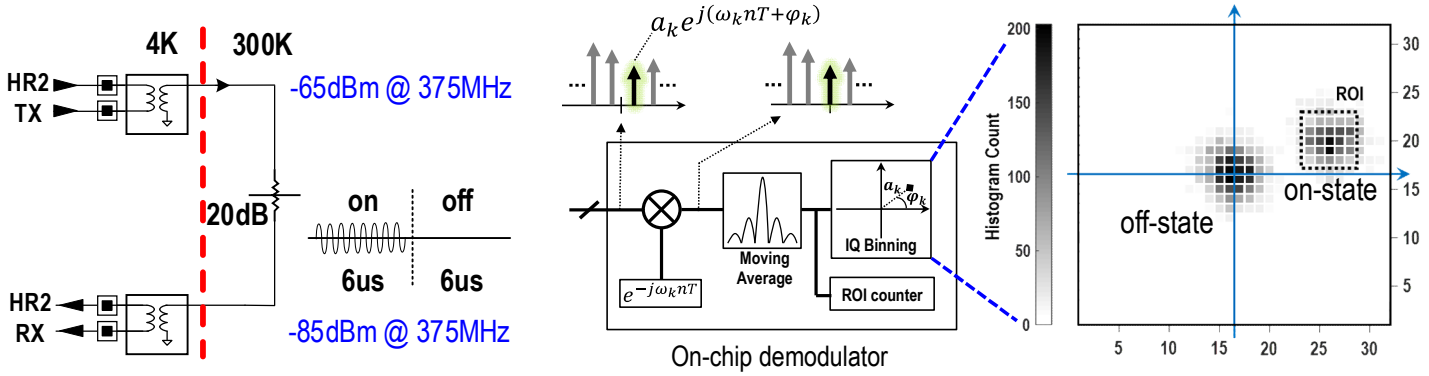


Fig. 14. Qubit readout loop-back test setup and measured results. Details of digital demodulator are also shown.

HR2 is also implemented in Intel 22nm FinFET technology. Figure 10 shows the 16mm^2 die micrograph integrating more than 100M transistors. The PCB hosting the FCBGA packaged SoC is enclosed in a two-piece copper enclosure for thermalization at 4K. Thermal resistance is minimized by a large number of ground bumps and BGA balls and an array of indium springs for direct back-side silicon contact. Figure 11 shows HR2 assembled in the copper enclosure and installed on the 4K plate of the dilution refrigerator. The SoC has been fully characterized at 4K inside the fridge, clocked at 1.6GHz [39]. All the following measurements are reported at 4K, unless noted otherwise. As shown in Fig. 12, RF drive achieves >-10dBm peak power and >-15dBm power for IM3<-50dBc over 11-17GHz. Measured SNR over 25MHz bandwidth is >44dB, meeting the 99.99% single-qubit gate fidelity target [35]. Receiver linearity has been tested with a 6-tone signal (Fig. 13), showing intermodulation products <-30dBc, sufficient for a 99.99% readout fidelity [39]. The measured receiver noise temperature is 44K, which we believe might be limited by device self-heating [42]. Analysis of readout fidelity based on receiver noise figure and other parameters can be found in [39]. A loop-back configuration is used to test qubit readout (Fig. 14). The TX generates a pulse at 375MHz for 6µs while the RX receives for 12µs. The TX signal is first routed outside the fridge, attenuated and fed back to the RX with -85dBm input power to match the level expected for RF reflectometry. After amplification, IQ downconversion with a 400MHz LO and further amplification and filtering at baseband, the signal is sampled by the ADCs and processed by the digital demodulator (Fig. 14). One of the 6 channels is used to downconvert the complex tone to DC by digital mixing. Each IQ sample is then filtered with a moving average and binned into a 32×32 Re/Im plane. As can be seen in Fig. 14, the cluster of points corresponding to the -85dBm signal is clearly distinguishable. By setting a total count threshold for a specific region of interest (*ROI* in Fig. 14), the qubit state corresponding to signal reflection can be detected.

The table in Fig. 15 summarizes the performance and compares features of HR1 and HR2 SoCs. Substantial improvement can be seen in HR2 digital power consumption, thanks to an optimized architecture and implementation of clock gating.

Future Directions and Conclusion

As state-of-the-art quantum processors comprise only a few qubits (<100), each qubit can still be individually wired from the cryogenic chamber to the room-temperature electronics. Current setups even allow for a few cables per qubit to provide different signals and functionalities, such as biasing, microwave drive for single-qubit operations, and fast pulsing for 2-qubit gates [43]. Consequently, the use of cryogenic electronics is minimal in most quantum-computer prototypes and typically limited to a few low-noise cryogenic amplifiers that exploit the low-temperature to ensure the signal-to-noise ratio in the read-out [43][44]. Not requiring the development of integrated circuits, this approach offers fast deployment and flexibility. However, the complexity of room-temperature electronics has already forced the shift from general-purpose off-the-shelf equipment to specialized discrete electronics implementations. Furthermore, ensuring the reliable operation of hundreds of cables through the multiple stages of a dilution refrigerator from room-temperature down to 100mK is an engineering feat in itself, and it is arguable that we are already very close to the limits imposed by refrigerator size, wiring heat load, and reliability.

A first approach to overcome the wiring bottleneck is re-using each cable for multiple functions, e.g., exploiting frequency multiplexing on a single line to drive single-qubit operations on multiple qubits, as proposed in the Horse Ridge SoCs. Although this can alleviate the issue in the short term, it is unlikely to allow scaling from today's ~100 qubits to the required millions of qubits, due to inherent multiplexing limits, e.g., crosstalk and limited bandwidth for frequency multiplexing. A more promising alternative is arranging N^2 qubits in a 2D array and addressing them using a cross-bar scheme requiring ~N bit and word lines, similar to standard semiconductor memories [45]. While cross-bar addressing can bring more aggressive scalability of the control signals, it imposes stringent requirements on the uniformity of the quantum devices, which is extremely challenging in itself.

As a consequence, an increasing number of researchers are developing cryogenic interfaces, as previously discussed, with more and more functionalities of the electronic control interface being

		HR1	HR2
Qubit platform		Transmons + spin qubits	Spin qubits
Controller capability		Drive	Drive, Readout with Digital Detector, Gate Pulsing, µ-Controller
Power	Analog	1.7mW/qubit	Drive: 5.2mW/qubit[a] Readout TX/RX: 1.3/9.3mW/qubit Gate pulsing: 2.9mW/channel
	Digital	Digital: 330mW @ 1GHz clock	10 – 140mW[b] @ 1.6GHz clock
Chip area		4mm^2	16mm^2
Technology		22nm FinFET CMOS	22nm FinFET CMOS
Drive			
Freq. range		2 – 20GHz @ >-45dBm	11 – 17GHz @ >-10dBm
Freq. multiplexing		Yes, 32 qubits	Yes, 16 qubits
Sampling rate		1GS/s	up to 2.5GS/s
IM3		<-50dBc @ -18dBm (6.25GHz)	<-50dBc @ >-17dBm (11 to 17GHz)
Data bandwidth		1GHz	up to 2GHz
Digital FIR		No	Yes, 2 notches per qubit
Envelope size		Up to 40,960 points AWG[d]	16,384 points AWG[c]
Instruction set		2^3 codewords per qubit	2^{19} codewords per qubit
Output impedance		N/A	50Ω
LO IQ generation		External off-chip hybrid (PCB)	Integrated on-chip hybrid

HR2 – Readout	
Freq. multiplexing	Yes, 6 qubits
TX output freq.	DC – 0.6GHz
TX output power	-70 to -40dBm
RX RF freq.	200 – 600MHz
RX gain	40 – 90dB
RX noise temperature[e]	44K
RX baseband BW	60 – 200MHz
Filter order	8th
ADC	SAR, 7.5 bits, 400MS/s
Qubit state detector	Integrated
HR2 – Gate Pulsing	
# Channels	22 (simultaneous)
Amplitude range	±0.4V
Amplitude res.	11 bits
Pulse width	10ns – 2.6ms
Pulse width resolution	2.5ns
Rise/fall time	~50ns[f]
Output rms noise	200µV$_{rms}$
Output impedance	50Ω

[a]Max P_{out} [b]Depending on activity [c]Per codeword, per qubit
[d]Shared for 8 codewords and 16 qubits [e]400MHz LO, 100MHz baseband [f]14pF load

Fig. 15. HR1 and HR2 performance summary and comparison.

integrated for operation at cryogenic temperature. Those ICs are typically operated at around 4K, a temperature at which existing refrigerators offer a cooling capability of a few Watts. Although serving thousands of qubits or more will surpass such capability, the cooling power of the cooling infrastructure may be enormously improved at 4K by adopting custom-made refrigerators [46]. However, operating the electronics above 1K due to the cooling requirements also forces the electronics to be far from the qubits, which typically operate at 20-100mK on a different temperature stage, thus still resulting in a wiring bottleneck. The straightforward solution would be to design "hot" qubits operating at the same temperature of the electronics [47], and encouraging results in this direction have been shown, demonstrating silicon spin qubits operating above 1K [5][6]. As an alternative, an electronic front-end with simple functionality, and hence with power requirements compatible with the mK cryogenic stage (typically < 1mW at 100mK), can be operated close to the sub-K qubits to just multiplex the electronic signals, thus alleviating the wiring bottleneck. An example is shown in [25].

Even with the electronic interface and the quantum processor operating at the same temperature, the challenge remains on how to connect the classical and the quantum section. 3D packaging approaches could be employed, similar to room-temperature System-in-Package (SiP), by fabricating the qubits and the electronics in separate dies to exploit processes optimized for the specific task [48][49]. A bolder solution is monolithic co-integration, which is very suitable for spin qubits based on semiconductor quantum dots that are in principle compatible with advanced CMOS manufacturing [50]. As quantum-dot-based qubits require a very small pitch in the order of 100nm, handling the fan-out from a large array of 100-nm-pitch metal lines to transistor-based electronics is still demanding, even in a fully integrated solution. The cross-bar addressing scheme mentioned above can ease this issue by lowering the number of required control lines. Another option would be to space the qubits using long-range quantum interconnects so that the resulting inter-qubit silicon area is available for electronic integration [51]. In both cases, electromagnetic and thermal crosstalk from the electronics must be tightly controlled not to spoil qubit performance. For instance, cryogenic self-heating from CMOS transistors can significantly affect the temperature of on-chip devices even at tens of µm distance [42], thus potentially limiting the integration density of such a quantum/classical system.

As sketched above, the path towards the integration of the cryogenic electronic interface to support large-scale quantum processors will still require significant innovations. Although the main research challenge in today's quantum-computing research is the demonstration of a scalable quantum processor architecture, the bottlenecks in wiring and in the electronic interfacing will have to be tackled next. Cryo-CMOS electronics can offer a large scale of integration and high-speed low-power performance coupled with a mature EDA infrastructure and the potential for qubit co-integration, thus representing a suitable candidate to support future large-scale quantum computers.

Acknowledgment

The authors thank M. Fredette, C. Le, C. Paulino, C. Vasel, S. Kale, B. Carlton, P. Sagazio, M. Chakravorti, A. B. K. T., M. Rahman, D. Vemparala, T. Huynh, S. Peinkofer, Z. Tuli, V. Hoang, R. Lee, R. Mckee, E. Hernandez, B. Esparza, L. Chai, P. Bach, J. Feng, A. R. Flores, J. A. Batres, R. Yarlagadda, Y. Liao, L. Abramowitz, A. Srinivasan, V. Parvatha, R. Sao, S. B. K, M. Bhartiya, I. Klotchkov, D. Kurian, J. Timbadiya, S. Soni, A. Kshirsagar, T. Nguyen, T. Hewage, S. Vangal, U. Jalan, R. Stingel, T. Prabhakaran, J. Lim, S. Suzuki, D. Souza, H. Carlo, L. Chang, D. Lake, N. Patel, M. Woodruff, T. Huusari, H.J. Lee, S. Mudanai, S. Kim, S. Ravikumar, C. Nieva, F. Sheikh, H. Rampurawala, L. Lampert, A. Hubert, J. Roberts, S. Premaratne, T. Watson, C. Hull, J. Clarke, J. Held, A. Matsuura, C. Jeon , J. van Dijk, N. Samkharadze, A. Corna, A. Sammak, G. Scappucci, M. Veldhorst, Z. Y. Chang, O. Benningshof, M. Sarsby, N. Alberts, T. Alkemade, B. Bakker, H. Homulle, R. N. Schouten and B. P. Wuetz.

References

[1] R.P. Feynman, "Simulating physics with computers", *International journal of theoretical physics*, 21(6), 467–488 (1982).

[2] L. Vandersypen, "Quantum computing – the next challenge in circuit and system design", *IEEE International Solid-State Circuits Conference* (ISSCC), San Francisco, CA, 2017.

[3] H. Bluhm and L.R. Schreiber, "Semiconductor spin qubits – A scalable platform for quantum computing?", *IEEE International Symposium on Circuits and Systems*, Sapporo, Japan (2019).

[4] E. Charbon, et al., "Cryo-CMOS for Quantum Computing", *International Electron Device Meeting* (IEDM), San Francisco, CA, 2016.

[5] C.H. Yang, et al., "Operation of a silicon quantum processor unit cell above one Kelvin", *Nature* 580, 350–354 (2020).

[6] L. Petit, et al., "Universal quantum logic in hot silicon qubits", *Nature* 580, 355–359 (2020).

[7] L.C. Camenzind, et al., "A spin qubit in a fin field-effect transistor", *arXiv:2103.07369v1* (2021).

[8] J.D. Cressler and H.A. Mantooth, Eds., *Extreme environment electronics*. Boca Raton, Florida: CRC Press., 2013.

[9] J.D. Cressler, et al., "On the profile design and optimization of epitaxial Si- and SiGe-base bipolar technology for 77 K applications. I. Transistor DC design considerations," *IEEE Tran. Electron Devices*, 40(3), 525-541 (1993).

[10] L. Najafizadeh, et al., "Sub-1-K Operation of SiGe Transistors and Circuits," *IEEE Electron Device Lett.*, 30(5), 508-510 (2009).

[11] S.R. Ekanayake, et al., "Characterization of SOS-CMOS FETs at Low Temperatures for the Design of Integrated Circuits for Quantum Bit Control and Readout," *IEEE Tran. Electron Devices*, 57(2), 539-547 (2010).

[12] T. Lehmann, "Cryogenic Support Circuits and Systems for Silicon Quantum Computers", *IEEE International Symposium on Circuits and Systems*, Sapporo, Japan (2019).

[13] L. Riesebos, et al., "Quantum Accelerated Computer Architectures", *IEEE International Symposium on Circuits and Systems*, Sapporo, Japan (2019).

[14] B. Patra et al., "Cryo-CMOS circuits and systems for quantum computing applications," *IEEE J. Solid-State Circuits*, vol. 53, no. 1, pp. 309–321, Jan. 2018.

[15] 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 J. Solid-State Circuits*, vol. 54, no. 11, pp. 3043–3060, Nov. 2019.

[16] J. van Staveren et al., "Voltage references for the ultra-wide temperature range from 4.2K to 300K in 40-nm CMOS," *in Proc. IEEE 45th Eur. Solid State Circuits Conf. (ESSCIRC)*, Sep. 2019, pp. 37–40.

[17] M. J. Gong *et al.,* "Design Considerations for Spin Readout Amplifiers in Monolithically Integrated Semiconductor Quantum Processors," *IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, 2019, pp. 111-114.

[18] I. Bashir et al., "A Mixed-Signal Control Core for a Fully Integrated Semiconductor Quantum Computer System-on-Chip," *in Proc. IEEE 45th Eur. Solid State Circuits Conf. (ESSCIRC)*, Sep. 2019, pp. 125-128.

[19] M. Mehrpoo, et al., "A cryogenic CMOS parametric amplifier," *IEEE Solid-State Circuits Lett.*, vol. 3, pp. 5–8, 2020.

[20] L. L. Guevel et al., "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," *2020 IEEE International Solid- State Circuits Conference - (ISSCC)*, 2020, pp. 306-308.

[21] J. Gong, et al., "A 200dB FoM 4-to-5GHz cryogenic oscillator with an automatic common-mode resonance calibration for quantum computing applications," *in IEEE ISSCC Dig. Tech. Papers*, Feb. 2020, pp. 308–310.

[22] P. Vliex et al., "Bias Voltage DAC Operating at Cryogenic Temperatures for Solid-State Qubit Applications," *IEEE Solid-State Circuits Lett.*, vol.3, pp. 218-221, 2020.

[23] A. Esmailiyan et al., "A Fully Integrated DAC for CMOS Position-Based Charge Qubits with Single-Electron Detector Loopback Testing," in *IEEE Solid-State Circuits Letters*, vol. 3, pp. 354-357, 2020.

[24] A. Ruffino, et al., "A wideband low-power cryogenic CMOS circulator for quantum applications," *IEEE J. Solid-State Circuits*, vol. 55, no. 5, pp. 1224–1238, May 2020.

[25] S. J. Pauka, et al., "A cryogenic CMOS chip for generating control signals for multiple qubits," *Nature Electronics*, vol. 4, pp.64–70, Jan. 2021.

[26] A. Ruffino, et al., "A fully-integrated 40-nm 5-6.5 GHz cryo-CMOS system-on-chip with I/Q receiver and frequency synthesizer for scalable multiplexed readout of quantum dots," *in IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2021, pp. 210–211.

[27] B. Prabowo et al., "A 6-to-8 GHz 0.17 mW/qubit cryo-CMOS receiver for multiple spin qubit readout in 40 nm CMOS technology," *in IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2021, pp. 212–213.

[28] G. Kiene et al., "A 1 GS/s 6-to-8b 0.5 mW/qubit cryo-CMOS SAR ADC for quantum computing in 40 nm CMOS," *in IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2021, pp. 214–215.

[29] K. Kang et al., "A 5.5mW/Channel 2-to-7 GHz Frequency Synthesizable Qubit-Controlling Cryogenic Pulse Modulator for Scalable Quantum Computers," *2021 Symposium on VLSI Circuits*, 2021, pp. 1-2.

[30] Y. Peng, et al., "A cryogenic broadband sub-1-dB NF CMOS low noise amplifier for quantum applications," *IEEE J. Solid-State Circuits*, vol. 56, no. 7, pp. 2040–2053, Jul. 2021.

[31] A. Potocnik, et al. "Millikelvin temperature cryo-CMOS multiplexer for scalable quantum device characterization." *Quantum Science and Technology*, Oct. 2021.

[32] J. Gong, et al., "A 2.7mW 45fsrms-Jitter Cryogenic Dynamic-Amplifier-Based PLL for Quantum Computing Applications," *2021 IEEE Custom Integrated Circuits Conference (CICC),* 2021, pp. 1-2.

[33] A. Ruffino et al., "Integrated multiplexed microwave readout of silicon quantum dots in a cryogenic CMOS chip," *ArXiv:2101.08295*, 2021.

[34] J. P. G. Van Dijk et al., "A scalable cryo-CMOS controller for the wideband frequency-multiplexed control of spin qubits and transmons," *IEEE J. Solid-State Circuits*, vol. 55, no. 11, pp. 2930–2946, Nov. 2020.

[35] J. P. G. van Dijk et al., "Designing a DDS-based SoC for high-fidelity multiqubit control," *IEEE Trans. Circuits Syst. I, Reg. Papers,* vol. 67, no. 12, pp. 5380-5393, Dec. 2020.

[36] C.-H. Lin et al., "A 12 bit 2.9 GS/s DAC with IM3 -60 dBc beyond 1 GHz in 65 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 44, no. 12, pp. 3285–3293, Dec. 2009.

[37] X. Xue et al., "CMOS-based cryogenic control of silicon quantum circuits," *Nature*, vol. 593, no. 7858, pp. 205–210, May 2021.

[38] Deutsch, D. & Jozsa, R. "Rapid solution of problems by quantum computation" *Proc. R. Soc. Lond. A* 439, 553–558 (1992).

[39] J. Park et al., "A Fully Integrated Cryo-CMOS SoC for State Manipulation, Readout, and High-Speed Gate Pulsing of Spin Qubits," *IEEE J. Solid-State Circuits*, vol. 56, no. 11, pp. 3289-3306, Nov. 2021.

[40] J. Elzerman, et al., "Single-shot read-out of an individual electron spin in a quantum dot," *Nature 430*, 431-435 (2004).

[41] S. Woo, et al., "A 3.6 mW differential commongate CMOS LNA with positive-negative feedback," in *Proc. IEEE Int. Solid-State Circuit Conf.,* Feb. 2009, pp. 218–219.

[42] P. A. T Hart, et al., "Characterization and Modeling of Self-Heating in Nanometer Bulk-CMOS at Cryogenic Temperatures," in *IEEE Journal of the Electron Devices Society,* vol. 9, pp. 891-901, 2021.

[43] F. Arute, et al., "Quantum supremacy using a programmable superconducting processor," *Nature*, vol. 574, 505–510 (2019)

[44] G. Zheng, "Rapid gate-based spin read-out in silicon using an on-chip resonator.," *Nat. Nanotechnol*, vol.14, pp. 742–746, 2019.

[45] R. Li, et al., "A crossbar network for silicon quantum dot qubits" *Sci. Adv.* 4, eaar3960 (2018).

[46] L. Tavian, *Latest Developments in Cryogenics at CERN*, Tech. rep. CERN-AT2005-011-ACR., July 2005, http://cds.cern.ch/record/851586.

[47] L. M. K. Vandersypen et al., "Interfacing Spin Qubits in Quantum Dots and Donors— Hot Dense and Coherent", *npj Quantum Information*, vol. 3, no. 34, 2017.

[48] M. Vinet et al., "Towards scalable silicon quantum computing," *2018 IEEE International Electron Devices Meeting (IEDM)*, 2018, pp. 6.5.1-6.5.4.

[49] D. Rosenberg et al., "Solid-State Qubits: 3D Integration and Packaging," in *IEEE Microwave Magazine*, vol. 21, no. 8, pp. 72-85, Aug. 2020.

[50] A.M.J. Zwerver, et al., "Qubits made by advanced semiconductor manufacturing", *arXiv:2101.12650*, 2021.

[51] Jelmer M. Boter, et. al, "The spider-web array - a sparse spin qubit array," *arXiv:2110.00189*, 2021.