A Compact Low-Power Cryo-CMOS Readout With Active Quenching for Superconducting Nanowire Single-Photon Detectors

Giovanni Carboni, *Graduate Student Member, IEEE*, Jad Benserhir, *Member, IEEE*, Maoran Li, Lin Jin, Marinus C. van der Maas, Luc Enthoven, Jan Riegelmeyer, Ryoichi Ishihara, *Senior Member, IEEE*, Carlos Errando-Herranz, Masoud Babaie, *Senior Member, IEEE*, and Fabio Sebastiano, *Senior Member, IEEE*

***Abstract*—Superconducting nanowire single-photon detectors (SNSPDs) have emerged as leading cryogenic photon detectors, thanks to their high detection efficiency and low jitter. However, their large-scale integration remains limited by the wiring bottleneck between the cryogenic detectors and their room-temperature readout electronics. In applications such as color-center-based quantum computers (QCs), thousands of detectors may need to operate in parallel within a limited cryogenic cooling budget, thus asking for a scalable, low-power cryogenic electronic readout. To address these needs, this work introduces a cryogenic readout circuit directly wire-bonded to the SNSPD and using a high-impedance input to maximize the quality of the detector signal, thus relaxing the requirement of the cascaded amplifier and reducing its power consumption. An active quenching circuit is then adopted to ensure a reliable reset after the latching of the detector induced by such high input impedance. Implemented in 40-nm CMOS with an active area of <0.14 mm^2, the system achieves competitive performance at 0.1 K, delivering low timing jitter (<40 ps), high speed (dead time of ≈5 ns), and dark count rates (DCRs) below 1 Hz, while achieving a 5× reduction in power consumption (down to 20 μW) with respect to the cryogenic-readout state-of-the-art. Its ultralow-power operation and compact footprint make the proposed solution well-suited for integration within large-scale quantum-computing architectures.**

***Index Terms*—Active quenching, color-center, cryo-CMOS, quantum computing, readout, superconducting nanowire single-photon detector (SNSPD).**

Received 9 December 2025; revised 17 February 2026; accepted 18 March 2026. Date of publication 10 April 2026; date of current version 1 July 2026. This article was approved by Associate Editor Piero Malcovati. This work was supported by the "Modular Quantum Computers" by Fujitsu Ltd. and Delft University of Technology, co-funded by The Netherlands Enterprise Agency under Project PPS2007. The work of Carlos Errando-Herranz was supported by the Dutch Research Council Dutch Research Council (NWO) through the Project "QuTech Phase II" under Project 601.QT.001. *(Giovanni Carboni and Jad Benserhir contributed equally to this work.) (Corresponding author: Fabio Sebastiano.)*

Giovanni Carboni, Jad Benserhir, Maoran Li, Lin Jin, Marinus C. van der Maas, Luc Enthoven, Ryoichi Ishihara, and Fabio Sebastiano are with the Department of Quantum and Computer Engineering, Delft University of Technology, 2628 CD Delft, The Netherlands, and also with QuTech, Delft University of Technology, 2628 CJ Delft, The Netherlands (e-mail: F.Sebastiano@tudelft.nl).

Jan Riegelmeyer and Carlos Errando-Herranz are with the Department of Quantum and Computer Engineering, Delft University of Technology, 2628 CD Delft, The Netherlands, also with QuTech, Delft University of Technology, 2628 CJ Delft, The Netherlands, and also with the Kavli Institute of Nanoscience, Delft University of Technology, 2628 CJ Delft, The Netherlands.

Masoud Babaie is with the Department of Microelectronics, Delft University of Technology, 2628 CD Delft, The Netherlands, and also with QuTech, Delft University of Technology, 2628 CJ Delft, The Netherlands.

Digital Object Identifier 10.1109/JSSC.2026.3678808

I. Introduction

QUANTUM computers (QCs) hold the promise of delivering computational advantages far beyond what is achievable with classical machines. However, current QC prototypes remain far from offering the performance required for real-world use. One of the key obstacles to scaling these systems is the realization of an electronic interface capable of reliably controlling and reading out qubits at scale [1], [2], [3]. Among available qubit platforms, color-center qubits, such as nitrogen-vacancy (NV) and tin-vacancy (SnV) centers in diamond, stand out as promising candidates for large-scale architectures [4], [5], since their robust, high-fidelity quantum operations, relatively high operating temperatures (∼1 K), and compatibility with optical interconnects provide a solid groundwork for large-scale integration. Their operation at relatively higher cryogenic temperatures (CTs) than other qubit platforms increases the available cooling power, while the ability to place qubits several millimeters apart, thanks to optical interconnects [6], [7], enables practical 3-D integration of the qubits with their cryogenic electro-optical control and readout circuitry [Fig. 1(a) and (b)] [5], [8], [9].

Since NV-center readout relies on detecting the few photons emitted during resonant excitation [10], a detector with high efficiency is required to reliably capture these photons, despite optical losses and the limited cyclicity of the optical transition [11]. At the same time, a low dark-count rate is essential to avoid spurious detection events that could mimic real fluorescence and degrade spin-state discrimination. Thus, single-photon detectors with high detection efficiency and minimal dark count are critical for achieving high-fidelity NV-center readout [5]. Considering the operating temperature of color-center qubits, along with the stringent requirements on detection efficiency and ultralow dark count rate (DCR), superconducting nanowire single-photon detectors (SNSPDs) stand out as the only viable technology capable of meeting these requirements, thanks to both high-efficiency photon detection and precise photon timing at CT [12], [13]. Unlike several qubit platforms based on superconducting circuits or semiconductor spins that only use purely electrical readout [14], color-center qubits fundamentally require optical fluorescence for state discrimination, making single-photon detection an intrinsic component of system architecture. This work,

therefore, focuses on the electronic challenges introduced by scalable cryogenic photon readout, with color-center quantum processors serving as a representative and technologically demanding application. As shown in Fig. 1(b), several components in the target system, such as the optical switch and qubit drivers, already require a cryogenic-CMOS (cryo-CMOS) interface, thus easing the addition of a co-integrated cryo-CMOS readout for the SNSPD. Thanks to the proximity to the detector, such a readout can be readily connected to the SNSPD, thus circumventing the wiring bottleneck between the cryogenic detector and the typically adopted room-temperature readout. However, the area and the power consumption of the detector and its cryo-CMOS interface must be minimized to comply with the integration requirements. To address those needs, we demonstrate an integrated SNSPD readout architecture designed to meet the requirements for robust color-center detection and entanglement operations, and operating even below 1 K. By eliminating the need for long transmission lines between the SNSPD and its readout, the conventional 50-Ω matching constraint, imposed by offchip cabling, is lifted, making it possible to employ a higher input impedance that eases the amplifier requirements and allows for a low-power design. Despite the potential of this approach, prior work still required a power consumption exceeding 100 μW for specifications comparable to this work's target application, even when adopting a low-noise SiGe BiCMOS technology [15]. To bridge the gap between existing readout solutions and the requirements of color-center QCs, this work presents a fully integrated 40-nm cryo-CMOS readout circuit that achieves a 5× reduction in static power consumption compared to the state-of-the-art, enabled by a minimalistic ultralow-power architecture that leverages a high input impedance and active quenching to prevent the detector latching otherwise induced by such a large input impedance, with the achieved performance primarily enabled by the use of a high input impedance, a single-stage LNA architecture, and direct bonding to minimize input parasitics, rather than by CMOS technology scaling alone. This article extends on [16] by providing a review of state-of-the-art SNSPD readout in Section II, deriving an analytical framework to model the readout circuitry in Section III, and presenting a comprehensive circuit description and experimental validation in Sections IV and V, respectively. Concluding remarks are drawn in Section VI.

II. SNSPD Read-Out Circuits: Functionality and State-of-the-Art

A. Superconducting Nanowire Single-Photon Detectors

The detection mechanism of SNSPDs relies on breaking Cooper pairs in a current-biased superconducting nanowire. Upon absorption of a photon with sufficient energy, the superconducting state is locally suppressed, forming a resistive domain that generates a measurable voltage pulse across the nanowire [17]. Following absorption, Joule heating raises the local temperature above the critical threshold T_c, interrupting the superconducting current. The nanowire kinetic inductance L_k(typically in the range from 100 nH to 10 μH, depending

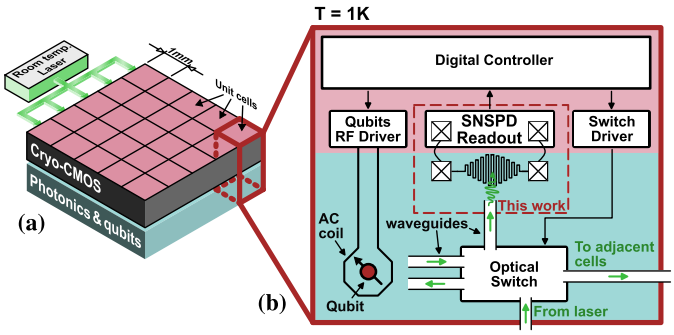


Fig. 1. Architecture of a QC based on color centers. (a) 3D-integrated cryo-CMOS electrical interface, photonic interface, and qubits. (b) Unit cell with one color center, qubit driver, switch driver, optical switch, and single-photon detector (one per unit cell, with one readout circuit per detector).

on the geometry) prevents an instantaneous redistribution of the bias current, thereby temporarily sustaining current flowthrough the emerging resistive domain and shaping the initial transient voltage response. The resulting current dynamics can be approximated using the effective inductive time constant of the nanowire. As shown in Fig. 2(a), once the resistive domain forms, the bias current decays with a characteristic time constant

$$\tau_d \approx \frac{L_k}{R_L + R_n(t)} \tag{1}$$

where $R_n(t)$denotes the time-dependent resistance of the non-superconducting domain. As the hotspot develops and $R_n$increases, the effective resistance seen by $L_k$rises, accelerating the current decay. During this decay, $R_n$typically reaches values in the order of a few kΩ, while it can be up to tens of kΩif the full nanowire exits the superconducting regime. During the reset phase, as $R_n(t)$collapses and the nanowire returns to the superconducting state, the current recovers with a time constant approximately given by

$$\tau_L \approx \frac{L_k}{R_L}. \tag{2}$$

These inductive dynamics govern the pulse shape illustrated in Fig. 2(a). Through heat diffusion and thermal conduction to the substrate, the nanowire cools and returns to the superconducting state during the reset phase, setting the minimum recovery time and maximum count rate. Thanks to this principle, SNSPDs combine high sensitivity and low timing jitter. For wavelengths around 1550 nm, state-of-the-art detectors achieve system jitters of 6–8 ps [18], significantly outperforming semiconductor photodetectors like InGaAs/InP SPADs, which are typically limited to timing jitters in the 50–200 ps range at telecom wavelengths [19].

A key limitation in SNSPD operation is the occurrence of latching. When a photon creates a resistive hotspot, the bias current must rapidly divert from the resistive area to allow the nanowire to return to the superconducting state. If the bias current has no low-impedance path available, such as when the typically-adopted load resistance $R_L$in parallel to the SNSPD is too large, a significant portion of the current continues to flow through the resistive nanowire. This sustained current generates additional Joule heating, reinforcing the hotspot instead of allowing it to collapse. As a result,

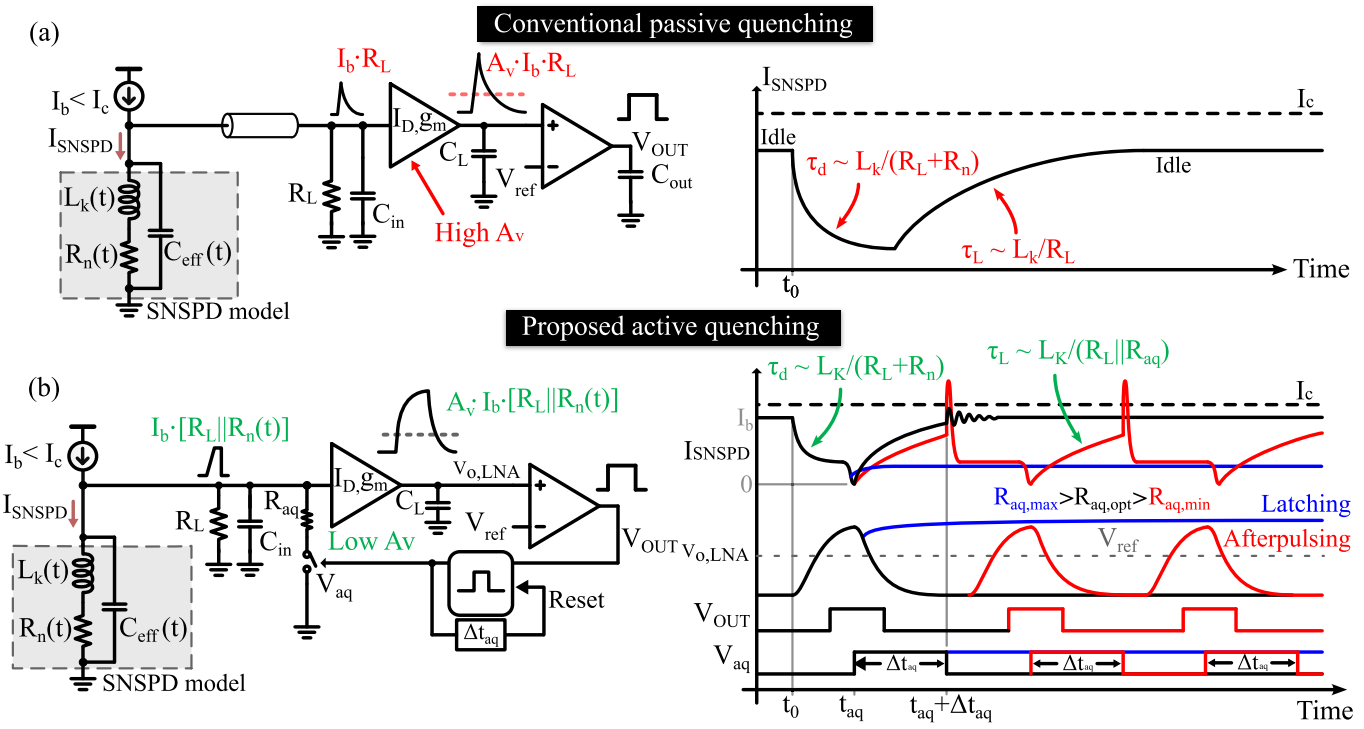


Fig. 2. Readout architectures and corresponding SNSPD current dynamics during a detection event with photon arrival at t_0. (a) Conventional configuration with a 50-Ωtransmission line connecting the SNSPD to the readout circuit (typically operating at higher CT or even at room temperature), which enforces the use of a 50-ΩLNA and passive quenching. (b) Proposed scheme with the detector and the readout co-located at CT, thereby eliminating the transmission line and the 50-Ωmatching requirement but asking for active quenching. The qualitative waveforms shown here are discussed in detail in Section III-C, where the underlying current and voltage evolution is analyzed.

the nanowire remains in a resistive state and is effectively latched, disabling subsequent photon detection until the bias is reduced or interrupted [20], [21]. Consequently, the choice of $R_L$is critical: it must be low enough to provide an efficient path for the bias current to escape during hotspot formation and avoid latching, yet high enough to ensure sufficient output swing and timing performance for the readout electronics [15]. According to the conventional approach, the design of the readout and amplification circuitry must then balance two main requirements: maintaining a bias current slightly lower than the critical current $I_c$for maximum detection efficiency while ensuring low load impedance and adequate thermal dissipation to prevent latching.

B. SNSPD Readout Circuits

The readout of SNSPDs has conventionally relied on commercial cryogenic or room-temperature LNAs with a standard 50-Ω input impedance cascaded with a comparator to generate a digital pulse retaining the photon-timing information [see Fig. 2(a)]. Although this value of input impedance primarily ensures proper matching to coaxial transmission lines, it also provides sufficient electrothermal damping to suppress latching, as demonstrated in both experimental and modeling studies [20], [21], [22]. However, the low input impedance significantly affects the characteristics of the SNSPD voltage pulse, requiring the LNA to provide substantial gain and/or lower noise, which increases power consumption and complicates thermal management in large cryogenic systems. These amplifiers, including cryogenic InP HEMT and SiGe wideband LNAs, can achieve noise temperatures in the range from 2 to 5 K, bandwidths up to 12 GHz, and power consumptions in the range from 5 to 50 mW, depending on device technology and noise requirements. Such performance is feasible because power dissipation is less restrictive at CTs above 1K than in sub-Kelvin detector environments [23], [24]. When a single detector is read out, this approach becomes advantageous: using a dedicated cryogenic LNA reduces the readout noise temperature well below the intrinsic SNSPD noise, typically in the 10–20-ps_{rms} range, with state-of-the-art measurements demonstrating rms jitters of ≈20 ps with optimized cryogenic amplifiers [25], and below 10 ps for advanced detector geometries using ultralow-noise readout electronics [26].

As cryogenic systems transition from single-channel to multi-pixel SNSPD arrays, conventional methods become less feasible. Implementing a high-power cryogenic LNA for each detector conflicts with the limited cooling capabilities of cryostats and would increase the complexity of the wiring between detectors and their readout. To address this issue, on-chip multiplexing methods, such as time-division, frequency-division, and amplitude multiplexing, have been developed [27], [28], [29]. While these methods simplify wiring, they require additional superconducting circuits on the detector side, leading to significant tradeoffs. Depending on the multiplexing strategy, designers may face lower count rates, increased timing jitter, or changes to SNSPD geometry, possibly reducing the system detection efficiency. Unlike detector multiplexing strategies, our work targets scaling down the readout electronics power and footprint without modifying

the detectors. Such a compact, ultra-low-power cryo-CMOS readout would then be compatible with large-scale quantum processors, eliminating power-intensive wideband cryogenic LNAs and on-chip multiplexing elements. Thanks to the low power, the readout electronics for multiple channels can be operated very close to the detector chip, thus simplifying the interconnects and paving the way to a possible (3-D) hybrid integration scheme for the SNSPD and the electronics [30].

To achieve the target specifications, the proposed readout leverages the absence of long 50-Ω-matched transmission lines to adopt a high-input readout impedance that enhances the SNSPD voltage pulse but introduces the risk of latching. To prevent latching in high-impedance SNSPD readout architectures, active quenching can be employed. In particular, the work by Ravindran et al. [15]demonstrated that replacing the conventional low-impedance (50 Ω) input with a high input impedance enables the readout to leverage the inherent gain of the SNSPD during detection. By allowing the hotspot to drive a substantial voltage across a large load resistance, the SNSPD generates a larger electrical pulse (i.e., higher slew rate), significantly relaxing front-end noise requirements. A controlled reset loop with programmable timing then allows the nanowire to return to the superconducting state without entering a latched regime. While this prior work established the feasibility and qualitative benefits of active quenching, it did not address how the readout circuit parameters, specifically transconductance g_m, input capacitance, bandwidth, and load impedance, quantitatively affect timing performance. The dependence of timing jitter on the interplay between readout amplifier output-referred noise, input-network dynamics, and amplifier slew rate remains largely unexplored. To bridge this gap, Section III-A presents an analytical framework linking the parameters of the readout circuit to the system-level performance.

III. System Design and Optimization

A. Target Specifications for NV and SnV Center Readout

For entanglement generation between color centers, the SNSPD readout must resolve arrival times on a timescale shorter than the emitter's optical decay. For instance, the radiative lifetime of SnV centers outside a cavity is ~6 ns, implying that detector jitter should be limited to a small fraction of this, typically 200–300 ps, to avoid temporal smearing of the photon wave packet [31]. Embedding the center in a nanophotonic cavity further shortens the lifetime and tightens these constraints. High timing resolution is also required to preserve photon indistinguishability in the presence of frequency detuning $\Delta\omega$ [32]: the fidelity of the two-photon interference drops as the detector response time approaches the beat period $2\pi/\Delta\omega$. Sub-nanosecond timing jitter ensures that the detector response time remains much shorter than the characteristic beat period associated with frequency detuning between photons, preventing detector-induced temporal averaging from introducing additional temporal-mode distinguishability. This preserves the phase coherence necessary for high-fidelity two-photon interference and remote entanglement between non-identical, spatially separated color centers [33].

Beyond timing constraints, the high detection efficiency of SNSPDs (>98% at visible and telecom wavelengths) is a major advantage for this application [13]. Entanglement schemes based on single-photon heralding suffer from low success probabilities, since both emitters must generate and transmit a photon in each trial. Any detector inefficiency quadratically reduces the success probability.

In addition to jitter and efficiency, the SNSPD readout must support short post-detection time-filtering windows after signal digitization. Experiments with NV centers with a 12-ns radiative lifetime [34]demonstrate that the SNR is maximized for gate lengths of approximately 6 ns, with typical operating points around 10 ns to balance fidelity and success rate. For SnV centers, with lifetimes approximately half as long, these values scale proportionally. Taken together, these considerations impose strict requirements on the readout chain: low-noise operation (<300-ps timing jitter), high count rates of at least 20 MHz (requiring a recovery time below 50 ns), high detection efficiency, and minimal static power (<100 μW to satisfy the cooling budget of the system in Fig. 1, including the SNSPD bias circuitry), while enabling nanosecond-scale windowing to preserve fidelity and the rate of remote entanglement operations.

B. Timing Jitter Optimization

As depicted in Fig. 2, the simplified front-end features an SNSPD modeled with an inductance L_k, a state-dependent resistance $R_n(t)$, and the current-dependent effective capacitance of the nanowire $C_{\text{eff}}(I_b)$, with $I_b$being the SNSPD bias current [20], [21]. To accurately reproduce the SNSPD dynamics, including rise time, fall time, and reset behavior, we use the established SPICE electrothermal model introduced in [21], with device parameters detailed in Section IV-A. The SNSPD is directly connected to the load resistance $R_L$and to the LNA. We model the input impedance of the readout circuit with a parasitic input capacitance C_{in}in parallel to R_L.

The timing jitter at the LNA output is determined by the slew rate and the integrated noise at its output. Following the small-signal approximation, the rms jitter $\sigma_t$can be approximated as

$$\sigma_t \approx \frac{\sigma_{v,\text{out}}}{\text{SR}_{\text{out}}} \tag{3}$$

where $\sigma_{v,\text{out}}$is the rms noise voltage at the threshold crossing and $\text{SR}_{\text{out}}$is the effective slew rate of the LNA output. After photon detection, the LNA input voltage increases at a finite slew rate SR_{in}. By assuming that the passive-quenching readout in Fig. 2(a)adopts a low $R_L(R_L \ll R_n)$ to prevent latching and the active-quenching readout adopts a high $R_L(R_L \gg 200\,\Omega)$ [35], $\text{SR}_{\text{in}}$for a given bias current $I_b$in the active-quenching regime and in the passive-quenching is restricted by the total input capacitance or by L_k, and the load resistance R_L, respectively, and can be expressed as [15]

$$\text{SR}_{\text{in}} = \begin{cases} \dfrac{I_b R_L}{\sqrt{\frac{8 L_k C_{\text{eq}}(I_b)}{3}}}, & \text{for } \dfrac{L_K}{R_L + R_n} \gg (R_L \parallel R_n)\, C_{\text{eq}} \\ \dfrac{I_b}{C_{\text{eq}}(I_b)}, & \text{for } \dfrac{L_K}{R_L + R_n} \ll (R_L \parallel R_n)\, C_{\text{eq}} \end{cases} \tag{4}$$

where $C_{eq}(I_b) = C_{in} + C_{eff}(I_b)$is the equivalent capacitance at the LNA input. Two main mechanisms can limit the slew rate at the LNA output: the finite amplifier bandwidth and the input slew rate, with one of the two mechanisms dominating depending on the detector size and the amplifier speed [25]. In the operating regime of interest, we assume that the detector-induced transient does not limit the output slew rate, as we do not require extreme timing jitter specifications, and we target a very-low-power readout. In this regime, the detector sets the input slew rate, while the amplifier limits the output slew rate relevant for timing jitter. This results in the amplifier limiting the slew rate and the noise bandwidth at its output. We model the amplifier as a linear element with an equivalent transconductance g_m, output impedance R_{out}, DC gain by $A_v = g_m R_{out}$, and output load C_L. The peak voltage at the LNA input is $I_b \cdot (R_L \| R_n)$, with $R_n$being the detector resistance at the threshold crossing. With those assumptions, the slew-rate at the amplifier output can be approximated as

$$SR_{out} \approx \frac{g_m I_b (R_L \| R_n)}{C_L}. \tag{5}$$

For noise estimation, the contribution of flicker $(1/f)$ noise is neglected, as its effect is confined to low frequencies and becomes negligible when the operating bandwidth significantly exceeds the device's flicker-noise corner frequency. The noise contributed by the SNSPD bias current source is neglected, since it is typically much smaller than the noise of the input circuit ($R_n$and R_L). Under those assumptions, the LNA output-referred integrated noise can be modeled by combining the amplifier's thermal and shot noise (assuming it is dominated by an input MOS transistor with transconductance $g_m$and drain current I_d) with the thermal noise from $R_n$and R_L

$$\sigma_{v,out} = \sqrt{\frac{A_v}{C_L}} \sqrt{kT\left[\gamma + g_m (R_L \| R_n)\right] + \frac{qF}{2 (g_m/I_d)}} \tag{6}$$

where $k$denotes Boltzmann's constant, $q$is the elementary charge, $T$is the absolute temperature, $\gamma$is the excess noise factor of the input transistor (taken as 2/3), and $F$is the Fano factor describing the shot-noise component of I_d, assumed constant and equal to 0.1 at both 300 and 1 K [36], [37].

Substituting (5)and (6)into (3)yields a compact expression for the LNA output-referred timing jitter, valid under the amplifier-limited regime, that is, $(R_L \| R_n) C_{in} \ll R_{out} C_L$

$$\sigma_t \approx \frac{\sqrt{A_v C_L}}{g_m I_b (R_L \| R_n)} \cdot \sqrt{kT (\gamma + g_m (R_L \| R_n)) + \frac{qF}{2 (g_m/I_d)}}. \tag{7}$$

The timing jitter in (7)indicates a tradeoff among transconductance, load capacitance, and bias currents. Increasing either $g_m$or the SNSPD bias current $I_b$enhances the bandwidth-limited slew rate, reducing jitter. However, a higher $g_m$amplifies the relative contribution of the input-network thermal noise $kT(R_L \| R_n)$. In contrast, increasing the amplifier bandwidth, that is, reducing the load capacitance C_L, is beneficial, as it improves the slew rate and lowers the integrated output noise, at least in the amplifier-limited regime.

To validate the extracted model, a circuit-level transient simulation was performed using the LNA architecture shown in Section IV-B, excluding the feedback amplifier. The LNA is implemented as an ideal 40-nm CMOS single-transistor stage with $g_m/I_d = 16\,V^{-1}$and a DC gain of approximately 25 dB. The total load capacitance was assumed to be $C_L = 40\,fF$. The SNSPD SPICE model was directly connected to the LNA input to reproduce the system's time-domain behavior. The input capacitance was set to $C_{in} = 350\,fF$, and the SNSPD bias current to $I_b = 20\,\mu A$. A few simplifying assumptions were introduced: only the noise contribution of the main amplifying transistor (M_1) was considered, and the feedback device (M_4) was biased with an ideal voltage source instead of being regulated by the feedback loop. The timing jitter has been estimated by assuming the circuit is tuned to trigger the comparator at the instants of maximum output slew rate, which is non-trivial to estimate analytically due to the non-linearity of the SNSPD lumped elements [$R_n(t)$, $L_k(t)$, and $C_{eff}(t)$] and the presence of multiple poles in the system. Thus, $SR_{out}$is extracted from the transient response as the maximum slope of the LNA output, while the output-referred noise $\sigma_{v,out}$is obtained by integrating the total white noise over the LNA's bandwidth. The timing jitter as a function of $g_{m1}$is then computed using (3). All simulations were conducted at 300 K, as the available transistor models are unreliable below −55 °C [38].

As discussed in Section II-A, upon photon absorption, the time-varying SNSPD resistance $R_n(t)$at the comparison instant becomes significantly more influential when using a large load resistance compared with a small load. The term $R_n(t) \| R_L$affects both the slew rate (5)and the output-referred noise (6), and therefore influences the estimated timing jitter (7). The SNSPD resistance $R_n(t)$can be written as [15]

$$R_n(t) = R_{n,pk}\left(1 - e^{-2.2t/t_{rise}}\right) \tag{8}$$

where t_{rise}denotes the hotspot growth time constant and $R_{n,pk}$represents the peak normal-domain resistance determined by the nanowire geometry and its kinetic inductance. For the SNSPD parameters discussed in Section IV-A, we estimate $t_{rise} \approx 27\,ns$and $R_{n,pk} \approx 34\,k\Omega$. A direct estimation of $R_n$at the maximum slew rate instant yields values ranging from roughly 600 Ωat $g_m = 1\,mS$to about 30 kΩat $g_m = 5\,\mu S$. Those extracted values are used when computing the jitter with (7).

Fig. 3shows the circuit simulation results at 300 K compared to the analytical model expressed by (3)as a function of g_m, along with the model extrapolated to 1 K. The comparison is conducted for two load configurations ($R_L = 10\,k\Omega$and $R_L = 50\,\Omega$), thus highlighting the benefits of using a large load resistance over a smaller one. In the amplifier-limited regime, the slew-rate term scales with the LNA input voltage $I_b \cdot (R_L \| R_n)$, while the contribution of the input network to $\sigma_{v,out}$increases with $(g_m (R_L \| R_n))^{1/2}$. As a result, a larger $R_L$leads directly to a smaller contribution of the input network to the jitter. Furthermore, the contribution to the jitter due to the amplifier directly improves, thanks to the higher slew rate. This results in more than an order-of-magnitude reduction in jitter by moving from tens of ohms [Fig. 2(a)] to several kilo-ohms [Fig. 2(b)], even with identical transistor parame-

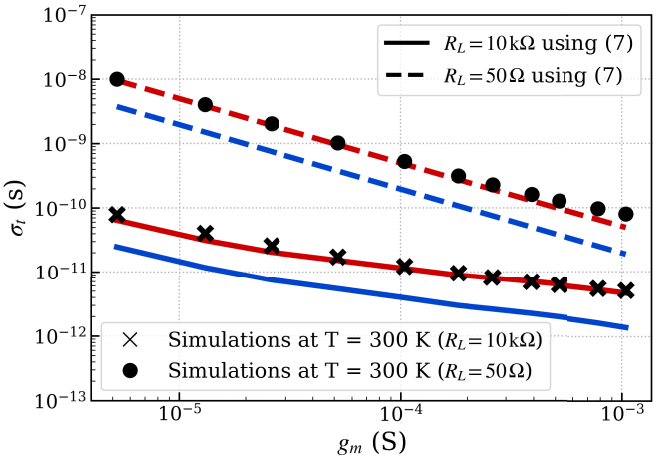


Fig. 3. Timing jitter at the LNA output as a function of g_m. Symbols correspond to circuit-level transient simulations performed only at $T = 300\,\mathrm{K}$using the Cadence Spectre simulator and including the full SNSPD interfacing network and the SNSPD electrical model. Solid and dashed lines show the analytical timing-jitter predictions given by (7)for $R_L = 10\,\mathrm{k}\Omega$and $R_L = 50\,\Omega$, respectively. The analytical model is evaluated at $T = 300\,\mathrm{K}$(red) and $T = 1\,\mathrm{K}$(blue).

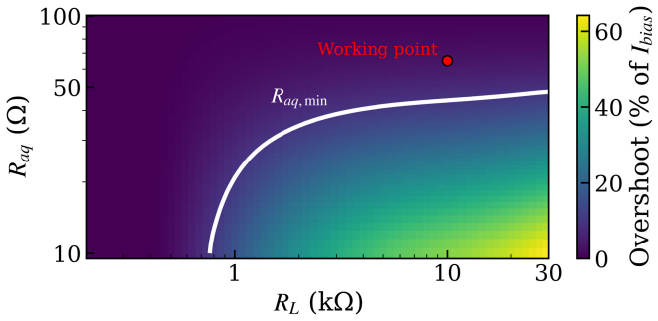


Fig. 4. SNSPD overshoot current after opening of R_{aq}. In red, the selected working configuration of $R_L$and R_{aq}in the measurements. In the plot, L_k = 200 nH, C_{in} = 350 fF, Δt_{aq} = 10 ns, and I_b = 20 μA.

ters. This benefit decreases as $g_m$increases, as shown in (7)and highlighted in Fig. 3.

As temperature decreases, the thermal-noise terms decrease, while the slew rate remains constant, if not better, as the amplifier could benefit from a wider bandwidth for the same transconductance at CT [modeled in (7)by a smaller C_L] [40].

To meet the target rms system jitter of 40 ps, it is necessary to account not only for the jitter contribution of transistor M1, as estimated in Fig. 3, but also for additional noise sources in the readout chain, as well as the intrinsic jitter of both the SNSPD and the laser. Considering the design margin required to achieve the jitter target without incurring excessive power consumption, a transconductance value of g_{m1} = 390 μS was selected. As indicated in Fig. 3, this operating point yields a jitter contribution of only a few picoseconds at 1 K. For such sizing, the noise contributions appearing in (6)distribute differently depending on the load resistance and temperature. At $T = 300\,\mathrm{K}$with $R_L = 10\,\mathrm{k}\Omega$, approximately 17% of the total output noise originates from the shot noise of transistor M1, 47% from its thermal noise, and the remaining 36% from the input network. When using a smaller load ($R_L = 50\,\Omega$), the proportions would shift to 26% from M1 shot noise, 72% from M1 thermal noise, and only 2% from the input network. At $T = 1\,\mathrm{K}$, however, the shot noise term dominates, accounting for approximately 97% of the total output-referred noise. It should also be noted that including additional noise sources, such as those contributed by the amplifier active load and the comparator's input-referred noise, will further degrade the overall timing jitter beyond the estimates shown in Fig. 3. It should be noted that these simulations are performed using the proposed LNA architecture and are intended to compare input load configurations within this specific design; they do not represent the ultimate performance achievable with dedicated high-power 50-Ωcryogenic amplifier implementations.

C. Active Quenching

Since the proposed architecture leverages the latching behavior of the SNSPD, the reset mechanism significantly affects both the count rate and the reliability of the system. As highlighted in the waveform in Fig. 2, when a photon drives the SNSPD out of its superconducting state, the bias current redistributes between the load resistor $R_L$and the SNSPD, now mainly characterized by a resistive impedance $R_n$and a negligible inductance. When the resulting voltage pulse is detected by the cascade of the LNA and the comparator, the active quenching switch is closed (at t_{aq}) to short the bias circuit away from the SNSPD, consequently allowing the detector to cool down and recover superconductivity. The time constant associated with the thermal cool-down of the SNSPD is typically much faster than the electrical circuit dynamics, so that it is approximated as instantaneous in the following. However, if the switch impedance $R_{aq}$is too high, that is, $R_{aq} > R_{aq,max}$, not enough current is diverted, and the SNSPD remains in the latched state (blue curve). If the switch resistance is low enough (black and red curve), enough current is shunted, so that the SNSPD recovers its superconductivity and starts absorbing more current thanks to its low (inductive) impedance. The current in the SNSPD inductance undergoes a transient from the initial condition at t_{aq}(determined by I_B, $R_n(t_{aq})$, R_L, and R_{aq}) toward its steady state value I_b, following an approximately exponential settling typically dominated by the inductor time constant $\tau_L = L_k/(R_L \| R_{aq})$, since typically $(R_L \| R_{aq})C_{in} \ll \tau_L$. However, at $t_{aq} + \Delta t_{aq}$, the active quenching switch is opened, creating a resonant circuit composed of L_k, C_{in}, and $R_L$with a relatively high quality factor. Depending on the value of the inductor current at $t_{aq} + \Delta t_{aq}$, which is a strong function of R_{aq}, the SNSPD current can overshoot beyond its critical current I_c, leading to the SNSPD triggering again and causing a sequence of afterpulsing. This happens for a too small switch resistance, that is, $R_{aq} < R_{aq,min}$, which leads to a too slow current transient between $t_{aq}$and $t_{aq} + \Delta t_{aq}$. A relatively small overshoot can already be detrimental, as the bias current $I_b$is typically chosen close enough to the critical current to reach the onset of internal efficiency saturation, that is, $I_b = p \cdot I_C$, with $p = 80\%$–95%. To analyze the interaction between the active-quenching circuit and the SNSPD, a simplified SPICE model for $t > t_{aq}$, based on

the schematic in Fig. 2(b), has been developed. The SNSPD is modeled solely through its kinetic inductance L_k, as the nanowire gains back its superconducting state on a very short timescale. The behavior of this model has been validated through cross simulations with the same SNSPD model based on [21] and used for the simulation of the jitter performance in Section III-B. The maximum overshoot in SNSPD current extracted from the model is plotted in Fig. 4, showing that for a higher $R_L$the effect of the higher quality factor of the resonant tank must be compensated by a larger R_{aq}that speeds up the current recovery during the active quenching. The minimum resistance $R_{aq,min}$can then be determined by setting the maximum allowed overshoot, for instance, the 10% line adopted in this work. The upper bound $R_{aq,max}$for the switch resistance is the same upper bound for the load resistance for the passive quenching (see Section II-A), that is, a few hundred ohms, thus not representing a limitation in practice.

D. Design Tradeoff Summary

Designing a low-power readout architecture based on active quenching involves balancing several interdependent tradeoffs between timing performance, stability, and energy efficiency. Increasing the input-stage transconductance improves the slew rate and reduces the LNA-limited jitter, but also raises power consumption and enhances the shot-noise contribution associated with higher bias currents. Similarly, a larger load resistance boosts the output swing and improves jitter performance, but drives the input section toward the underdamped regime, increasing current overshoot and the likelihood of afterpulsing. As a result, $R_L$can be optimized for jitter performance, as described in Section III-B, and R_{aq}chosen accordingly. To ensure a safety margin, the design presented here adopts $R_L = 10\,\text{k}\Omega$and $R_{aq} = 70\,\Omega$(see Fig. 4).

IV. Proposed Architecture

A. SNSPD Fabrication

For the realization of photonic circuits with broadband transparency and minimal absorption loss, we use stoichiometric silicon nitride (Si_3N_4) as waveguide material. A 250-nm layer of $Si_3N_4$is deposited via low-pressure chemical vapor deposition (LPCVD) onto a 2-μm thermally oxidized SiO_2layer on a silicon substrate. Before circuit fabrication, a nominal 10-nm-thick NbTiN film is deposited on top of the nitride layer by magnetron sputtering. Afterward, the electrical contact pads and alignment markers are patterned by photolithography. The structures are then realized by evaporation of 100-nm Au, followed by a lift-off process. 10-nm Cr is deposited before the Au layer, for the benefit of improving adhesion. Following this step, the nanowire geometries are defined using electron beam lithography (EBL) on positive-toned resist AR-P 6200.04. After removing unexposed resist, a short oxygen plasma step is performed to eliminate any remaining resist residuals. Subsequently, nanowire patterns are transferred from the resist into the NbTiN film using reactive ion etching (RIE) with $SF_6$and O_2plasma. The Si_3N_4waveguides are patterned with EBL using AR-P 6200.06 e-beam resist and then etched with RIE using $CHF_3$and Ar plasma. To mitigate the measurement uncertainty associated with low coupling efficiency, flood illumination has been employed to obtain a more direct estimate of the intrinsic detection efficiency. A scanning electron microscope image of the resultant nanowire atop Si_3N_4waveguide is shown in Section V-A. The resulting device consists of a single-bend nanowire with an active detection region approximately 100-μm long and 10-μm wide. Preliminary measurements indicate a room-temperature normal-state resistance of the nanowire $R_n \approx 375\,\text{k}\Omega$. To simulate the behavior of the system, an SNSPD model based on [21]was employed with the following parameters: $L_k = 200\,\text{nH}$, normal-state resistance $R_n = 375\,\text{k}\Omega$, critical current $I_c = 18.5\,\mu\text{A}$, and hotspot-related capacitance $C_{eff} = 473\,\text{fF}$.

B. Front-End Circuit

Fig. 5illustrates the schematic of the cryo-CMOS readout design. The CMOS chip includes two on-chip pads used to directly bond the SNSPD. Since the operating temperature remains highly stable, the on-chip bias current is generated using an internal unsilicided P^+poly resistor R_{bias}in Fig. 5(c), which provides sufficient accuracy for this design iteration. A more precise current-reference architecture could be incorporated in future versions if required. A PMOS current source M_{bias}provides the bias current for the SNSPD. Programmable diodes $M_{b1-2}$are used to program the current in the LNA, as explained later in this section. To experimentally explore the tradeoffs discussed in Section III, the readout input resistance R_L can be selected from a set of 5 silicided P^+poly resistor (nominal value at RT of 50 Ω, 500 Ω, 2 kΩ, 10 kΩ, 100 kΩ) and their parallel combinations, which are expected to decrease by ∼20% when cooling to CT [41]. Each resistor is enabled through an individually addressable NMOS switch (D_{0-4}). To enable the analysis of the impact of the quenching resistance on the recovery dynamics, the active-quenching switch $M_{aq}$is implemented using 7 programmable parallel NMOS transistors with binary-weighted widths, schematically represented by the resistor R_{aq}in Fig. 2. The active-quenching resistance is selected by routing the V_{AQ}signal through a programmable multiplexer. This multiplexer is realized with 2.5-V-thick-oxide NMOS transistors, which offer improved robustness to cryogenic threshold-voltage shifts, therefore guaranteeing sharp rail-to-rail transitions. Moreover, because the same shift-register infrastructure is reused across all programmable blocks, operating the multiplexer at 2.5 V eliminates the need for a dedicated level-shifting stage, reducing circuit complexity and overhead. With this topology, the effective active-quenching resistance R_{aq} from approximately 15 Ωto 2 kΩat RT. The activation and quenching duration is determined by a delay chain, as described later in Section IV-D. The NMOS devices used for active quenching employ cryogenic-aware forward body biasing [41], achieved by connecting the body contact to an external 1 V bias to counteract the cryogenic increase in V_{th}. This technique ensures sufficient overdrive for the switch during quenching by inducing a threshold voltage reduction from 550 to 300 mV, effectively restoring the room-

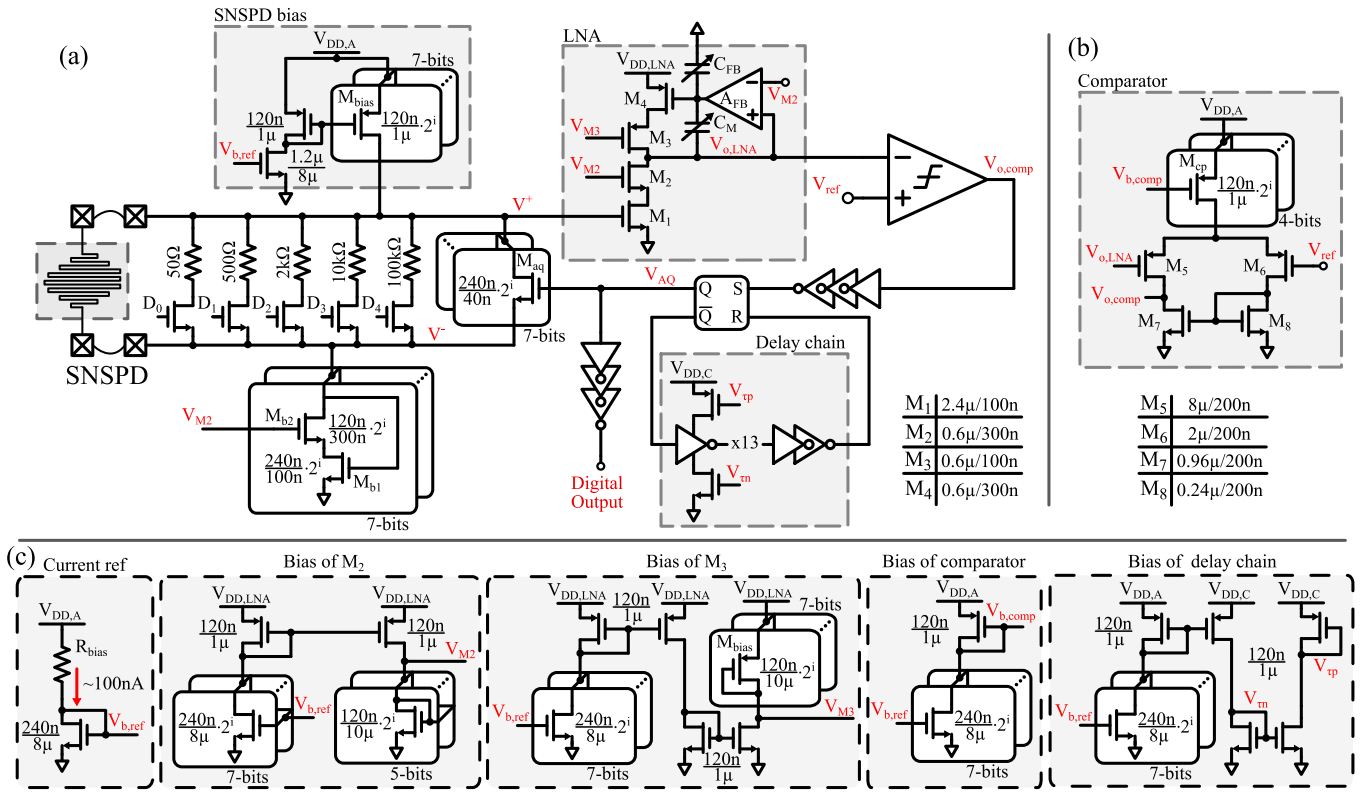


Fig. 5. Overview of the cryo-CMOS readout circuit: (a) transistor-level schematic of the cryo-CMOS readout circuit, (b) transistor-level schematic of the comparator, and (c) transistor-level schematics of all the biasing circuitry with their programmable elements.

temperature V_{th}. As a result, a smaller switch can be used, reducing parasitic loading at the readout input node (C_{in}).

C. Low Noise Amplifier

The LNA adopts a single-stage cascoded common-source amplifier, with M_1as the input device, M_2as its cascode, and an active load formed by M_{3-4}, biased at approximately 7 μA to meet the jitter specifications. This topology was selected for its simplicity, as it provides the required gain with the minimum number of branches, which is advantageous for reducing overall power consumption. The gain transistor $M_1$of the LNA is designed to operate in weak inversion to maximize transconductance efficiency, which is known to benefit jitter performance, as discussed in Section III-B. The MOSFETs typically exhibit a substantial increase of the subthreshold slope at CT, leading to a 3×increase in the transconductance g_{m_1}(from $g_{m_1} \simeq 130$ μS estimated from RT simulations to 390 μS) for the same bias current [38]. The bias voltages ($V_{M2}$and V_{M3}) of the LNA cascode transistors $M_{2,3}$are generated by 5-bit programmable NMOS and PMOS diodes fed by 7-bit programmable current mirrors. This approach ensures that the cascode transistors remain in saturation across temperature, compensating for the threshold-voltage increase observed at CTs. Connected to the negative terminal of the SNSPD, a 7-bit programmable input branch ($M_{1b}$and M_{2b}) of a wide-swing cascode current mirror exploits the superconducting state of the SNSPD to achieve significant current reuse. The output branch of this current mirror coincides with $M_{1,2}$in the LNA. In this configuration, the circuit can tune the bias current delivered to the SNSPD while maintaining a constant quiescent current in the LNA, thereby ensuring a consistent transconductance for M_1across the entire SNSPD bias-current range. A slow feedback amplifier (A_{FB}), consisting of a PMOS differential input pair loaded by an NMOS current mirror, ensures proper bias of the amplifier active load M_{3-4}, and sets the amplifier DC output $V_{o_{LNA}} \approx 400$ mVusing an externally provided voltage reference V_{FB}. A_{FB}must detect the DC drift at the amplifier output while maintaining a narrow bandwidth (<10 kHz) to filter out high-frequency components. This is crucial for preventing LNA bias drift at high event rates. A programmable Miller compensation capacitance $C_M$ensures the loop stability, and a programmable load capacitance $C_{FB}$is used to tune the bandwidth of the feedback amplifier. Although this feedback loop slightly reduces the amplifier gain, it does not affect the detection efficiency or the system jitter.

D. Back-End Circuit

The comparator, shown in Fig. 5(b), consists of a PMOS differential input pair (M_{5-6}) with an NMOS current-mirror load (M_{7-8}). Its role is to detect the LNA output crossing an externally provided threshold voltage, generating a sharper digital-like transition that helps contain the overall timing jitter. For a given total current, the comparator speed is optimized by sizing the current in $M_{6,8}$ 4×shorter than the current in $M_{5,7}$, which directly affects the signal path. Although this current distribution worsens the intrinsic noise performance of the comparator, a quantitative noise analysis shows that its contribution remains negligible compared to the LNA. In

particular, the dominant comparator noise originates from the PMOS input differential pair and is found to be more than two to three orders of magnitude lower than the LNA output-referred noise over the relevant range of transconductance, load resistance, and operating temperature. The high gain of the LNA ensures that the signal is significantly amplified before reaching the comparator stage, further reducing the relative impact of comparator noise on the overall timing jitter. The total bias current of the comparator is programmable through a two-stage scheme: a 7-bit NMOS-based current mirror defines the coarse current level, which in turn sets the resolution of the 4-bit PMOS-based current mirror M_{cp}.

The choice of the comparator threshold is critical for minimizing timing jitter. Ideally, the threshold must be placed near the point of maximum slope in the LNA output waveform. However, setting the threshold too close to the LNA DC output voltage increases the probability of noise-induced false triggers. Conversely, placing the threshold too far from the DC output level is also detrimental [see Fig. 2(b)]. As the LNA output approaches its peak value, the waveform flattens, reducing the local slope and thereby degrading the jitter.

The comparator buffered output is sampled by a standard-cell-based SR latch. The latch Qoutput (V_{AQ}) serves both as the readout output and the driving signal for the active quenching switch, whereas the $\overline{Q}$output drives a current-starved inverter-based delay chain that resets the latch. This delay chain controls the output pulsewidth and defines the quenching-switch on-time during the SNSPD post-detection cooldown. Since a too short delay can lead to improper quenching while a longer delay limits the count rate, the delay time can be programmed from 5 to 75 ns through a 7-bit NMOS-based current mirror that sets the inverter bias current.

V. Measurement Results

A. Measurement Setup

Fabricated in a 40-nm CMOS process, the readout circuit occupies an active area of approximately 0.14 mm^{2}and draws a total current of 29 μA from a 1.1-V supply. The overall power is distributed among the key functional blocks as follows: 38% for the SNSPD biasing branch, 42% for the analog amplifier, and the remaining 20% for the comparator and digital control logic. A preliminary characterization performed in liquid helium using an emulated SNSPD pulse generated by a Tektronix[1] AWG5014C showed a system jitter of approximately 30 ps_{rms} at 4 K. The measurement data and analysis files are found in [39].

Thanks to its low power consumption, the circuit can be operated at sub-Kelvin temperatures without exceeding the cryostat's thermal budget. For subsequent experiments, the chip was installed in the mixing chamber stage of a BlueFors[1] SD dilution refrigerator, operating at a base temperature of about 70 mK (see Fig. 6.). The CMOS die and the SNSPDs are glued on the same substrate, allowing direct electrical connection through short (approximately 5 mm) aluminum wirebonds. The resulting bonding-wire inductance ($\approx$ 5 nH) is negligible compared to the intrinsic kinetic inductance of the SNSPD ($L_k \approx 200$ nH), while parasitic capacitance at the input node remains the dominant factor influencing timing performance, as described by (4).

Multiple SNSPDs were integrated on the same chip: one has been connected to the cryo-CMOS readout under test, and another has been routed to a conventional RT readout chain for reference measurements. The RT path employed two ZFL-1000LN+ amplifiers in cascade, providing a benchmark for timing and pulse shape comparison. Both SNSPDs were illuminated simultaneously through a shared optical fiber delivering a Gaussian spot large enough to cover both detectors, ensuring comparable photon fluxes in each channel.

The CMOS chip provides the bias current directly to its associated SNSPD, and its operating conditions were configured via on-chip shift registers accessible through an SPI interface. These registers set the bias levels and overall readout configuration, allowing flexible tuning of circuit parameters from room temperature. For proof-of-concept validation, the system was illuminated with a Cobolt[1] 06-MLD continuous-wave (CW) 520-nm laser. For jitter characterization, a picosecond pulsed NKT (SuperK FIANIUM) laser was employed, with its electrical trigger used as a reference to compute the overall system jitter. The optical power was attenuated to the single-photon regime using calibrated optical attenuators and a power meter positioned outside the cryostat. Although fiber coupling losses were not directly quantified, the power incident at the cryostat input was estimated to be on the order of a few microwatts. The cryo-CMOS output waveforms were recorded using a Teledyne LeCroy WavePro[1] 7 Zi-A real-time oscilloscope (1.5–6-GHz bandwidth), and the timing jitter was extracted using its built-in time-interval histogram function by measuring the delay between the laser trigger edge and the corresponding chip-output threshold crossing; the oscilloscope intrinsic timing jitter (sub-ps rms) is well below the jitter levels reported in this work.

B. Result Discussion

The impact of R_{aq}on afterpulsing behavior is depicted in Fig. 7. Three distinct operational regimes emerge: as explained in Section III-C, when $R_{aq}$is too low ($R_{aq} \lesssim 50\ \Omega$), a sharp reinjection of the entire bias current at the end of the recovery phase leads to pronounced afterpulses, as clearly visible in the figure. Conversely, as explained in Section III-C, when $R_{aq}$is too high (not shown), the circuit behaves similar to a high-impedance readout, causing latching and preventing proper operation. An optimal intermediate resistance value ($R_{aq} \simeq 65\ \Omega$) results in photon-induced pulses without afterpulses, thus demonstrating stable and efficient reset behavior.

Fig. 8 shows the count rate and the DCR for the full range of SNSPD bias currents for both the cryo-CMOS and RT readout. The similar response trend for the two readout systems demonstrates that the cryo-CMOS is not limiting the readout, thereby validating its suitability for low-power cryogenic quantum applications without compromising performance. The observed variations between RT and CT are attributed to the fabrication-related spread between the two individual detectors used for the RT readout and the cryo-CMOS readout.

[1]Registered trademark.

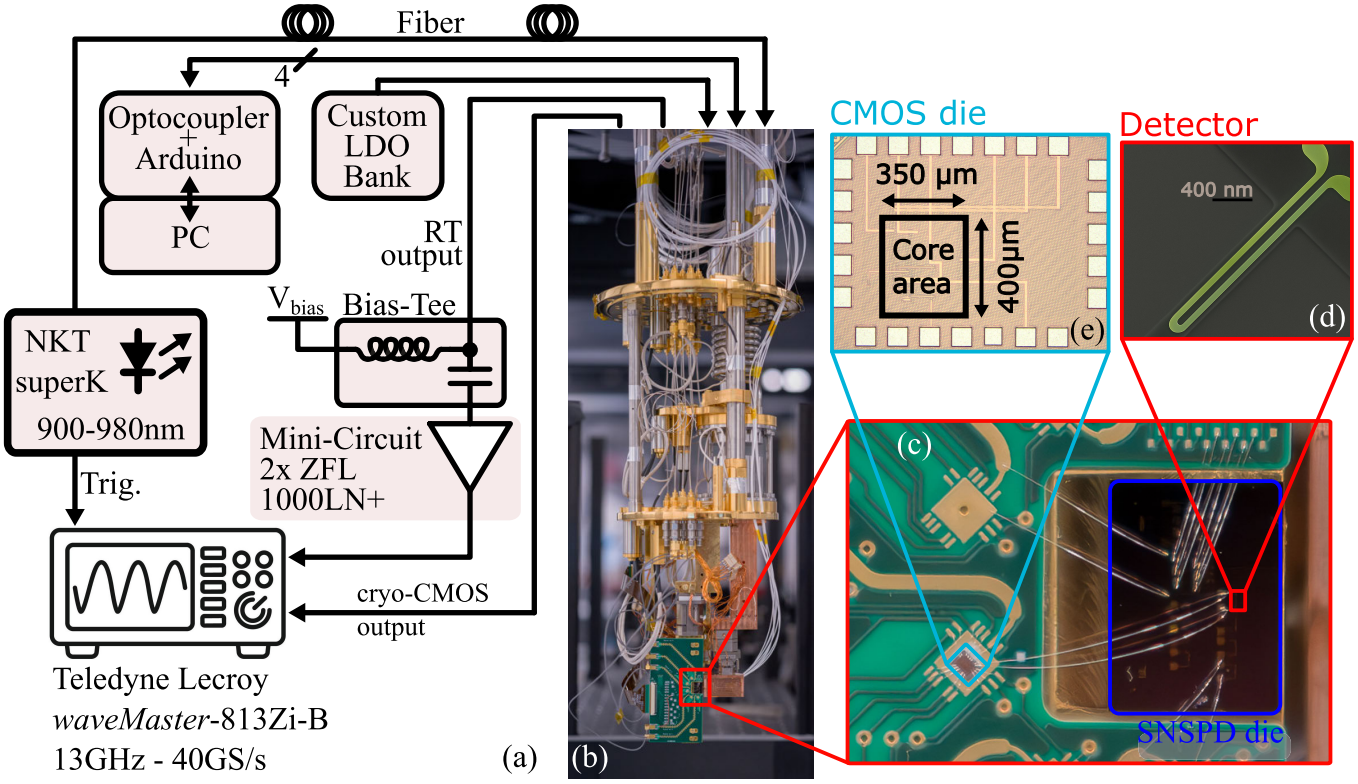


Fig. 6. Measurement setup. (a) Schematic representation of the room-temperature setup used for the characterization of both the cryo-CMOS readout and the bare SNSPDs connected to room-temperature LNAs. The LDO bank is implemented with Analog Devices LT3042 linear regulators. (b) Photograph of the BlueFors SD dilution refrigerator hosting the test PCB. (c) Close-up view of the test PCB mounted on the cold finger, showing the cryo-CMOS die and the SNSPD die wire-bonded to the PCB and to each other. (d) Micrograph of the SNSPD consisting of a single-bend nanowire with an active detection area of 100×10 μm (0.04-mm^{2}additional area required for the bonding pads). Next to the two SNSPDs used for the RT and cryo-CMOS readouts, the remaining area on the SNSPD die contains additional unused test structures. (e) Micrograph of the CMOS readout chip.

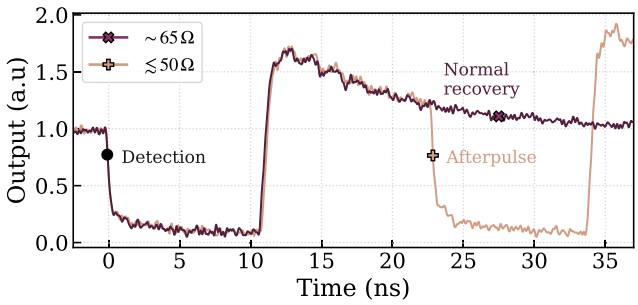


Fig. 7. Measured normalized cryo-CMOS digital output. The digital output of the chip exhibits reversed polarity, meaning that the falling edge of the signal corresponds to the detection event. The plot shows the impact of active quenching resistance on afterpulsing. In the plot, $R_L = 10\,\text{k}\Omega$, and $I_b = 12\,\mu\text{A}$. Overshoot and ringing on the rising edge are due to the combined effect of a non-matched driver driving a ~4 m-long 50-Ω line with a 50-Ω load and an excessive external resistance on the output driver supply.

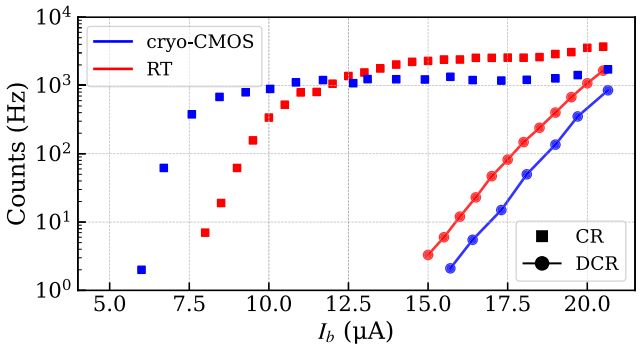


Fig. 8. Measured count rate and DCR as a function of the SNSPD bias for cryo-CMOS (blue) and RT readout (red) for $R_L = 10\,\text{k}\Omega$. The RT readout was used as a reference to establish the expected SNSPD characteristics under identical bias and temperature conditions, providing a baseline for comparison with the cryo-CMOS readout path.

The relevance of the comparator's decision time is highlighted in Fig. 9, which shows both the measured DCR and the system timing jitter as functions of the comparator reference voltage, expressed as a percentage of the LNA output swing (i.e., lower percentages correspond to a reference voltage closer to the LNA's DC output level). Importantly, the dark counts reported in this measurement primarily originate from the readout circuitry itself, rather than from the intrinsic dark-count processes of the SNSPD. The shape of the jitter curve qualitatively reflects the LNA voltage response, since a steeper LNA output slope directly translates into lower jitter. This suggests that V_{ref}should ideally be set close to the LNA's DC output voltage $V_{DC,LNA}$, where the signal slope is maximal. However, when $V_{ref}$is set too close to 0% of the LNA output

TABLE I
PERFORMANCE COMPARISON OF CRYOGENIC SNSPD READOUT

Parameter	Peng et al. [25]	Niu et al. [41]	Li et al. [23]	Ravindran et al. [15]	This work
Technology	130-nm SiGe BiCMOS	130-nm SiGe BiCMOS	130-nm SiGe BiCMOS	130-nm SiGe BiCMOS	40-nm CMOS
Temperature (K)	3.6	4.2	4.2	2.8	0.1 [a]
Active Quenching	No	No	No	Yes	Yes
Internal SNSPD Bias	No	No	No	No	Yes
Dark Count rate (Hz)	–	–	–	–	<1 [b]
Rms jitter (ps)	~17	–	–	~25	~40
Area (mm^2)	0.038	0.5	0.073 [c]	0.46	0.14
Power (μW)	8200 [d]	1800 [d]	3100 [d]	100 [d]	20 (+13 μW SNSPD bias)

[a] Measured at the cold-finger. [b] At the optimal detection efficiency bias point. [c] Core area only. [d] SNSPD bias not included.

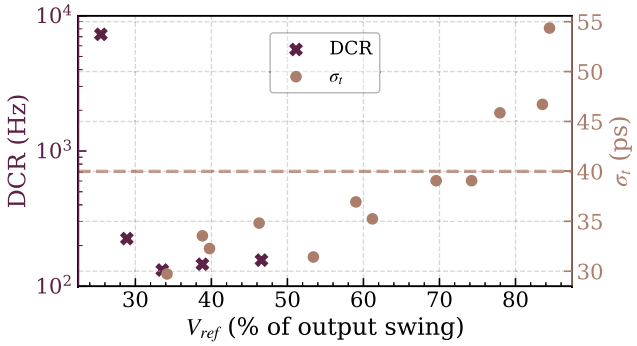


Fig. 9. Measured system timing jitter and DCR as a function of the voltage reference of the comparator, expressed in percentage of the LNA's output swing for R_L = 10 kΩand I_b = 12 μA.

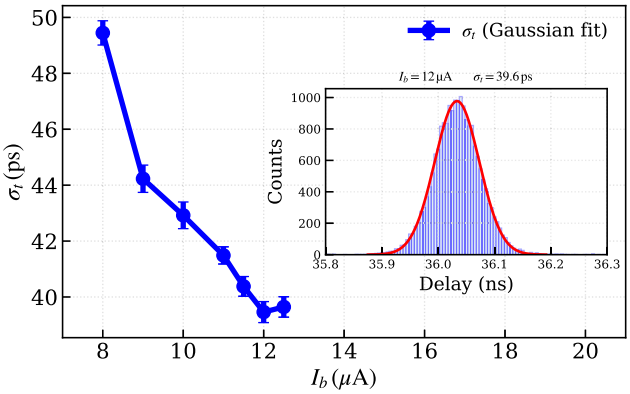


Fig. 10. Measured system timing jitter as a function of the SNSPD bias current I_b. The inset illustrates the extraction procedure based on a Gaussian fit for a representative case at I_b = 12 μA and R_L = 10 kΩ.

swing, noise-induced triggering appears, as highlighted by the increase in DCR in Fig. 9. As a tradeoff between jitter and DCR, the reference voltage is fixed to 60% in the following characterization, ensuring reliable system operation.

Fig. 10 presents the measured system rms jitter as a function of the SNSPD bias current I_b. As $I_b$increases, the detector output amplitude rises accordingly, leading to a higher slew rate at the LNA input; this improvement in signal edge sharpness is directly reflected in the reduction of the measured system jitter. The inset illustrates the extraction procedure, where the delay between the fast optical trigger and the electrical response is histogrammed and fit with a Gaussian distribution, providing a consistent estimate of the system timing jitter across all bias points.

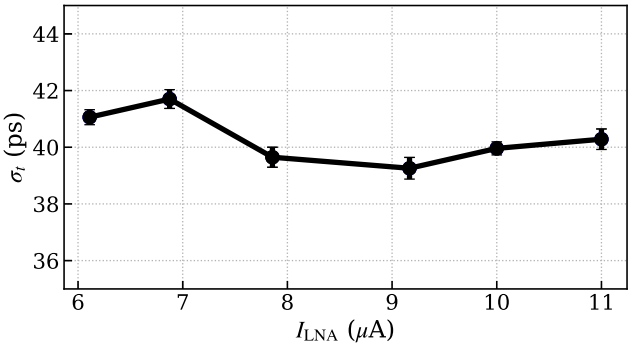


Fig. 11. Measured timing jitter as a function of the LNA bias current for R_L = 10 kΩand I_b = 12 μA.

Fig. 11 presents the measured rms jitter as a function of the LNA bias current. The curve is obtained by sweeping the width of the transistors $M_{b1}$and M_{b2}, while keeping the supply at the nominal value $V_{DD,LNA}$ = 1.1V. The resulting jitter stays essentially flat over the full bias-current range, suggesting that the intrinsic LNA contribution is negligible compared to other dominant noise sources, most likely arising from the SNSPD itself or from the measurement setup. This is consistent with the analysis in Section III-Bthat predicts the jitter contribution by the cryo-CMOS to be below 10 ps.

Table I compares the proposed SNSPD readout circuit with previously reported readouts. It is worth noting that the prior works summarized in Table I employ SiGe BiCMOS technologies with HBT-based amplification stages, which inherently provide superior noise performance compared to the nanoscale CMOS devices used in this work. However, relative to other cryogenic implementations, the presented architecture achieves the lowest power consumption while maintaining comparable system jitter performance. Furthermore, this work demonstrates the lowest operating temperature for an SNSPD/readout combination and is the first architecture that integrates on-chip bias-current generation for the SNSPD. The reported power consumption refers to operation at 100 mK, which represents a particularly stringent condition for the readout circuit. On the one hand, at this temperature, the available cooling power at this temperature is extremely limited, requiring the electronic interface to dissipate only a few tens of microwatts. On the other hand, SNSPDs exhibit their best performance at deep CTs, with reduced timing jitter and DCR [43], [44], so that the overall system performance is primarily limited by the readout electronics rather than the detector itself. Scalable system operation is expected at

higher CTs (e.g., ~1 K [5], [45]), the performance of cryo-CMOS devices is expected to remain unchanged [46], while substantially larger cooling power is available. In this context, the low per-channel power consumption enables the readout of large arrays, although any potential crosstalk, for example, via shared biasing, should be addressed in future large-scale implementations.

VI. Conclusion

This work demonstrates a low-power cryo-CMOS readout circuit that combines a high-impedance input stage with active quenching to interface with SNSPDs. Measurements confirm that the fabricated detectors can be properly biased and read out, enabling high-fidelity single-photon detection. Owing to its compact footprint and low power consumption, the proposed solution is well-suited for integration in scalable color-center QCs, preserving a substantial portion of the power budget for more demanding system components.

Acknowledgment

The authors would like to thank Erwin Van Zwet for the useful discussions. They also thank Roy Schoonenboom for lab planning for installing Bluefors, Olaf Benningshof and Tom Orton for the installation and maintenance of Bluefors, Tim Hiep and Nico Alberts for manufacturing the cryogenic compatible sample and nanopositioner holders.

References

[1] 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.

[2] J.Anders et al., "CMOS integrated circuits for the quantum information sciences," *IEEE Trans. Quantum Eng.*, vol. 4, pp. 1–30, 2023.

[3] S.Pellerano et al., "Cryogenic CMOS for qubit control and readout," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, Apr. 2022, pp. 01–08.

[4] H. P.Bartling et al., "Universal high-fidelity quantum gates for spin qubits in diamond," *Phys. Rev. Appl.*, vol. 23, no. 3, Mar.2025, Art. no. 034052.

[5] L.Enthoven, M.Babaie, and F.Sebastiano, "Optimizing the electrical interface for large-scale color-center quantum processors," *IEEE Trans. Quantum Eng.*, vol. 5, pp. 1–17, 2024.

[6] D.Riedel et al., "Efficient photonic integration of diamond color centers and thin-film lithium niobate," *ACS Photon.*, vol. 10, no. 12, pp. 4236–4243, Dec.2023.

[7] K. C.Chen et al., "A scalable cavity-based spin-photon interface in a photonic integrated circuit," *Optica Quantum*, vol. 2, no. 2, pp. 124–132, 2024.

[8] L.Li et al., "Heterogeneous integration of spin–photon interfaces with a CMOS platform," *Nature*, vol. 630, no. 8015, pp. 70–76, 2024.

[9] R.Ishihara et al., "3D integration technology for quantum computer based on diamond spin qubits," in *IEDM Tech. Dig.*, Dec.2021, pp. 1–4.

[10] S.Pezzagnaand J.Meijer, "Quantum computer based on color centers in diamond," *Appl. Phys. Rev.*, vol. 8, no. 1, Mar.2021, Art. no. 011308.

[11] D.Hopper, H.Shulevitz, and L.Bassett, "Spin readout techniques of the nitrogen-vacancy center in diamond," *Micromachines*, vol. 9, no. 9, p. 437, Aug.2018.

[12] I.Holzmanand Y.Ivry, "Superconducting nanowires for single-photon detection: Progress, challenges, and opportunities," in *Proc. Adv. Quantum Technol.*, vol. 2, nos. 3–4, Apr. 2019, Art. no. 1800058.

[13] D. V.Reddy, R. R.Nerem, S. W.Nam, R. P.Mirin, and V. B.Verma, "Superconducting nanowire single-photon detectors with 98% system detection efficiency at 1550 nm," *Optica*, vol. 7, no. 12, pp. 1649–1653, 2020.

[14] F.Sebastiano, "Fundamentals of cryo-CMOS circuits and systems for quantum computing: Challenges and prospects in interfacing future quantum computers," *IEEE Solid State Circuits Mag.*, vol. 17, no. 3, pp. 61–71, Mar.2025.

[15] P.Ravindran, R.Cheng, H.Tang, and J. C.Bardin, "Active quenching of superconducting nanowire single photon detectors," *Opt. Exp.*, vol. 28, no. 3, pp. 4099–4114, Feb.2020.

[16] G.Carboni et al., "Cryo-CMOS readout of single-photon detectors for color-center quantum computers," in *Proc. IEEE Eur. Solid-State Electron. Res. Conf. (ESSERC)*, Sep. 2025, pp. 253–256.

[17] C. M.Natarajan, M. G.Tanner, and R. H.Hadfield, "Superconducting nanowire single-photon detectors: Physics and applications," *Superconductor Sci. Technol.*, vol. 25, no. 6, Jun.2012, Art. no. 063001.

[18] M.Caloz et al., "Intrinsically-limited timing jitter in molybdenum silicide superconducting nanowire single-photon detectors," *J. Appl. Phys.*, vol. 126, no. 16, 2019, Art. no. 164501.

[19] E.Amri, G.Boso, B.Korzh, and H.Zbinden, "Temporal jitter in free-running InGaAs/InP single-photon avalanche detectors," *Opt. Lett.*, vol. 41, no. 24, pp. 5728–5731, 2016.

[20] A. J.Kerman, J. K. W.Yang, R. J.Molnar, E. A.Dauler, and K. K.Berggren, "Electrothermal feedback in superconducting nanowire single-photon detectors," *Phys. Rev. B, Condens. Matter*, vol. 79, no. 10, Mar.2009, Art. no. 100509.

[21] K. K.Berggren et al., "A superconducting nanowire can be modeled by using SPICE," *Superconductor Sci. Technol.*, vol. 31, no. 5, May2018, Art. no. 055010.

[22] A.Engel, J. J.Renema, K.Il'in, and A.Semenov, "Detection mechanism of superconducting nanowire single-photon detectors," *Superconductor Sci. Technol.*, vol. 28, no. 11, Nov.2015, Art. no. 114003.

[23] L.Li et al., "Wideband cryogenic amplifier for a superconducting nanowire single-photon detector," *Frontiers Inf. Technol. & Electron. Eng.*, vol. 22, no. 12, pp. 1666–1676, Dec.2021.

[24] J.Benserhir, Y.Zou, H.-C.Han, Y.Peng, and E.Charbon, "Broadband noise characterization of SiGe HBTs down to 4K," *IEEE J. Electron Devices Soc.*, vol. 13, pp. 983–996, 2025.

[25] Y.Peng, J.Benserhir, M.Castaneda, A.Fognini, C.Bruschini, and E.Charbon, "A 0.32 × 0.12 mm^2cryogenic BiCMOS 0.1–8.8 GHz low noise amplifier achieving 4 K noise temperature for SNWD readout," *IEEE T-MTT*, vol. 72, no. 4, pp. 2179–2192, Apr.2024.

[26] B.Korzh et al., "Demonstration of sub-3 ps temporal resolution with a superconducting nanowire single-photon detector," *Nature Photon.*, vol. 14, no. 4, pp. 250–255, 2020.

[27] M.Hofherr et al., "Time-tagged multiplexing of serially biased superconducting nanowire single-photon detectors," *IEEE Trans. Appl. Supercond.*, vol. 23, no. 3, p. 2501205, Jun.2013.

[28] S.Doerner et al., "Frequency-multiplexed bias and readout of a 16-pixel superconducting nanowire single-photon detector array," *Appl. Phys. Lett.*, vol. 111, no. 3, Jul.2017, Art. no. 032603.

[29] A.Gaggero et al., "Amplitude-multiplexed readout of single photon detectors based on superconducting nanowires," *Optica*, vol. 6, no. 6, pp. 823–828, 2019.

[30] I. EsmaeilZadeh et al., "Superconducting nanowire single-photon detectors: A perspective on evolution, state-of-the-art, future developments, and applications," *Appl. Phys. Lett.*, vol. 118, no. 19, May2021, Art. no. 190502.

[31] S. D.Tchernij et al., "Single-photon-emitting optical centers in diamond fabricated upon Sn implantation," *ACS Photon.*, vol. 4, no. 10, pp. 2580–2586, Oct.2017.

[32] J.Metzand S. D.Barrett, "Effect of frequency-mismatched photons in quantum-information processing," *Phys. Rev. A-At., Mol., Opt. Phys.*, vol. 77, no. 4, 2008, Art. no. 042323.

[33] C. M.Knaut et al., "Entanglement of nanophotonic quantum memory nodes in a telecom network," *Nature*, vol. 629, no. 8012, pp. 573–578, May2024.

[34] L.Childressand R.Hanson, "Diamond NV centers for quantum computing and quantum networks," *MRS Bull.*, vol. 38, no. 2, pp. 134–138, Feb.2013.

[35] A. J.Annunziata et al., "Reset dynamics and latching in niobium superconducting nanowire single-photon detectors," *J. Appl. Phys.*, vol. 108, no. 8, Oct.2010, Art. no. 084507.

[36] G.Kiene et al., "A 1-GS/s 6–8-b cryo-CMOS SAR ADC for quantum computing," *IEEE J. Solid-State Circuits*, vol. 58, no. 7, pp. 2016–2027, Jul.2023.

[37] S.Das, S.Raman, and J. C.Bardin, "Experimental characterization of the MOSFET Fano factor at cryogenic temperatures for accurate cryo-CMOS RF modeling," *IEEE Trans. Microw. Theory Techn.*, vol. 73, no. 10, pp. 7164–7176, Oct.2025.

[38] 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 J. Electron Devices Soc.*, vol. 6, pp. 996–1006, 2018.

[39] G. Carboni and J. Benserhir, "Data underlying the publication: A compact low-power Cryo-CMOS readout with active quenching for superconducting nanowire single-photon detectors," Delft Univ. Technol., Delft, The Netherlands, 2026. [Online]. Available: https://doi.org/10.4121/26073c87-7fdf-429a-a381-e6648f5a0e52
[40] A.Akturk, N.Goldsman, Z.Dilli, and M.Peckerar, "Effects of cryogenic temperatures on small-signal MOSFET capacitances," in *Proc. Int. Semicond. Device Res. Symp.*, Dec. 2007, pp. 1–2.
[41] R. W. J.Overwater, M.Babaie, and F.Sebastiano, "Cryogenic-aware forward body biasing in bulk CMOS," *IEEE Electron Device Lett.*, vol. 45, no. 2, pp. 152–155, Feb.2024.
[42] X.Niu, L.Li, X.Wu, and D.Wang, "A wideband cryogenic readout amplifier with temperature-insensitive gain for SNSPD," *Sensors*, vol. 22, no. 3, p. 1225, Feb.2022.
[43] T.Yamashita, S.Miki, W.Qiu, M.Fujiwara, M.Sasaki, and Z.Wang, "Temperature dependent performances of superconducting nanowire single-photon detectors in an ultralow-temperature region," *Appl. Phys. Exp.*, vol. 3, no. 10, Oct.2010, Art. no. 102502.
[44] R.Gourgues et al., "Superconducting nanowire single photon detectors operating at temperature from 4 to 7 K," *Opt. Exp.*, vol. 27, no. 17, pp. 24601–24609, 2019.
[45] N.Fakkel et al., "A cryo-CMOS controller with class-DE driver and DC magnetic-field tuning for quantum computers based on color centers in diamond," *IEEE J. Solid-State Circuits*, vol. 59, no. 11, pp. 3627–3643, Nov.2024.
[46] C.Enz, H.-C.Han, and E.Charbon, "Modeling of the MOSFET for the design of cryo-CMOS circuits," in *Proc. IEEE Eur. Solid-State Electron. Res. Conf. (ESSERC)*, Sep. 2024, pp. 5–8.

Giovanni Carboni (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees (cum laude) in electrical engineering from the University of Modena and Reggio Emilia, Modena, Italy, in 2020 and 2022, respectively. He is currently pursuing the Ph.D. degree with Delft University of Technology (TU Delft), Delft, The Netherlands.

In 2023, he was a Research Assistant with the Microwave Engineering Group, University of Modena and Reggio Emilia. His research interests include cryogenic CMOS circuits for superconducting nanowire single-photon detector (SNSPD) front-end design for quantum computing applications.

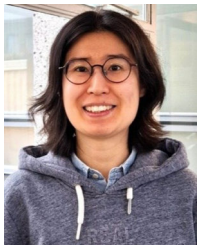
Jad Benserhir (Member, IEEE) received the B.Sc. degree in engineering physics from the Institut National Polytechnique de Grenoble (INPG), Grenoble, France, in 2017, the triple joint M.Sc. degree in micro- and nanotechnologies for integrated systems from Politecnico di Torino, Turin, Italy, the INPG, and the École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, in 2019, and the Ph.D. degree from the Advanced Quantum Architecture Laboratory (AQUA), EPFL, in 2024, where his research focused on CMOS/BiCMOS circuits for superconducting nanowire single-photon detectors (SNSPDs) front-end design and quantum applications.

He was a Research Assistant with the City University of Hong Kong, Hong Kong, in 2018. He is currently a Post-Doctoral Researcher with QuTech, Delft University of Technology (TU Delft), Delft, The Netherlands. His research interests include cryogenic electronics for quantum computing, RF and millimeter-wave integrated circuits, and advanced readout architectures for SNSPDs.

Dr. Benserhir serves on the Technical Program Committee of the IEEE MTT-S International Microwave Symposium (IMS), RFIC Technology and Techniques (RFTT) track, for the 2026 edition. He has also contributed to several international conferences and journals in the field of solid-state circuits and quantum technologies.

Maoran Li was born in Xi'an, China, in 1998. He received the B.E. degree in integrated circuit design and integrated systems from the University of Electronic Science and Technology of China, Chengdu, China, in 2020, and the M.Sc. degree in microelectronics from Delft University of Technology (TU Delft), Delft, The Netherlands, in 2023.

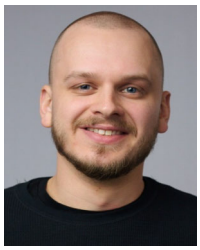
He is currently an Analog IC Design Engineer with Silicon Integrated, Eindhoven, The Netherlands.



Lin Jin received the Ph.D. degree in physics from the University of Münster, Münster, Germany, in 2024, under the supervision of Prof. Wolfram Pernice, where her research focused on the realization of monolithic integrated circuits on diamond and SiC.

She was a Post-Doctoral Researcher at the University of Münster, where she worked on the 3-D integration of advanced photonics, cryo-CMOS electronics, and SnV qubits in diamond. She is currently an Ada Lovelace Post-Doctoral Fellow at QuTech, Delft University of Technology (TU Delft), Delft, The Netherlands, jointly affiliated with the research groups of Prof. Tim Taminiau and Prof. Carlos Errando-Herranz. Her research interests include the spin dynamics of V2 centers in SiC and the use of integrated photonics to facilitate the development of interconnected spin chains.

Dr. Jin is a member of the American Physical Society (APS). She was awarded the Marie Skłodowska-Curie Fellowship in 2026.



Marinus C. van der Maas was born in Goes, The Netherlands, in 1996. He received the B.Sc. degree in electrical engineering and the M.Sc. degree in microelectronics from Delft University of Technology (TU Delft), Delft, The Netherlands, in 2020 and 2022, respectively, where he is currently pursuing the Ph.D. degree.

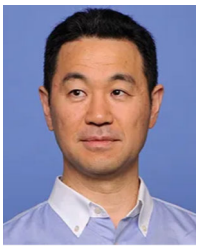
His research interests include integrated optics with optical MEMS switches and single-photon detectors.

Luc Enthoven received the B.Sc. and M.Sc. degrees (cum laude) in electrical engineering from Delft University of Technology (TU Delft), Delft, The Netherlands, where he is currently pursuing the Ph.D. degree in cryo-CMOS circuits for scalable quantum computing at the Group of Fabio Sebastiano.

Since 2025, he has been the CTO and one of the founders of FrostByte, where he works on cryo-CMOS for scaling quantum devices. In 2019, he worked as an Intern at Analog Devices, BMS Group, Munich, Germany. His research interests are in analog and mixed-signal circuit design as well as low-power cryo-CMOS circuits and systems.

Jan Riegelmeyer was born in Münster, Germany, in 1997. He received the B.Sc. and M.Sc. degrees in physics from the University of Münster, Münster, in 2019 and 2023, respectively. He is currently pursuing the Ph.D. degree with QuTech, Delft University of Technology (TU Delft), Delft, The Netherlands.

After his studies, he joined the Quantum Start-Up Pixel Photonics, Münster, in 2023, contributing to the development of superconducting nanowire single-photon detectors (SNSPDs). His current research focuses on the integration of color centers into nanophotonic circuitry.



Ryoichi Ishihara (Senior Member, IEEE) received the Ph.D. degree in physical electronics from Tokyo Institute of Technology, Tokyo, Japan, in 1996.

In 1996, he joined Delft University of Technology (TU Delft), Delft, The Netherlands, where he established research on large-area electronics using thin-film transistors (TFTs). He became a Principal Investigator at QuTech and a member of the Kavli Institute of Nanoscience, Delft, in 2014. Since 2013, he has been a Visiting Professor at Japan Advanced Institute of Science and Technology, Nomi, Japan, and a Specially Appointed Associate Professor at Tokyo Institute of Technology. He is currently an Associate Professor and the Head of the Quantum Integration Technology Group, Department of Quantum and Computer Engineering, TU Delft. He is a co-author of more than 90 journal articles and 120 international conference papers, and a co-inventor of 13 granted patent families. Highlights of his career include the first demonstration of TFTs inside a large Si grain in 1996, location and orientation control of Si crystals by laser crystallization in 2000 and 2006, monolithic 3D-IC using single-grain TFTs in 2008, ultrahigh-mobility Ge TFTs in 2010, carbon nanotube vertical interconnect vias in 2013, and the first solution-processed flexible Si TFTs on plastic in 2013 and a paper in 2014. His research interests include monolithic integration and advanced packaging technologies for scalable silicon spin qubit processors, integration engineering of diamond color centers for quantum computing and sensing, and printed silicon on paper substrates for flexible, biodegradable, and recyclable electronics.

Carlos Errando-Herranz was born in Valencia, Spain, in 1989. He received the Industrial Engineering degree from the Universitat Politècnica de València (UPV), Valencia, in 2013, with specialization in electronics and automatics, and the Ph.D. degree in micro- and nanosystems from the KTH Royal Institute of Technology, Stockholm, Sweden, in 2018.

He has held post-doctoral positions at KTH and at the Massachusetts Institute of Technology, Cambridge, MA, USA, and in 2023 joined Delft University of Technology (TU Delft), Delft, The Netherlands, as an Assistant Professor of quantum and computer engineering at the Electrical Engineering, Mathematics, and Computer Science School, and as the Group Leader at QuTech, Delft.

Dr. Errando-Herranz was a recipient of an ERC Starting Grant in 2025, a Vetenskapprådet Postdoctoral Fellowship from Sweden, and a Marie-Curie Global Postdoctoral Fellowship.

Masoud Babaie (Senior Member, IEEE) received the B.Sc. degree (Hons.) in electrical engineering from the Amirkabir University of Technology, Tehran, Iran, in 2004, the M.Sc. degree in electrical engineering from the Sharif University of Technology, Tehran, in 2006, and the Ph.D. degree (cum laude) in electrical engineering from Delft University of Technology (TU Delft), Delft, The Netherlands, in 2016.

He is currently an Associate Professor with the Department of Microelectronics, TU Delft. His research interests include RF and millimeter-wave integrated circuits and systems for wireless communications, as well as cryogenic electronics for quantum computing.

Dr. Babaie served on the Technical Program Committee for the IEEE International Solid-State Circuits Conference (ISSCC) from 2021 to 2026, including as the RF Subcommittee Vice Chair in 2025 and 2026. Since 2022, he has been the Chair of the Emerging Computing Devices and Circuits Subcommittee of the IEEE European Solid-State Electronics Research Conference (ESSERC). He also serves as an Associate Editor for IEEE SOLID-STATE CIRCUITS LETTERS(SSCL) and has served as a Guest Editor for IEEE JOURNAL OF SOLID-STATE CIRCUITS(JSSC). He has received several awards, including the 2015–2016 IEEE Solid-State Circuits Society Pre-Doctoral Achievement Award, the 2019 IEEE ISSCC Demonstration Session Certificate of Recognition, the 2020 IEEE ISSCC Jan Van Vessem Award for Outstanding European Paper, the 2022 IEEE CICC Best Paper Award, the 2023 IEEE IMS Best Student Paper Award (Second Place), the 2024 IEEE Solid-State Circuits Society Outstanding Reviewer Award, and the 2025 IEEE Solid-State Circuits Society New Frontier Award. In 2019, he was awarded the Veni Grant from the Netherlands Organization for Scientific Research (NWO).

Fabio Sebastiano (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees (cum laude) in electrical engineering from the University of Pisa, Pisa, Italy, in 2003 and 2005, respectively, the M.Sc. degree (cum laude) from the Sant'Anna School of Advanced Studies, Pisa, in 2006, and the Ph.D. degree from Delft University of Technology (TU Delft), Delft, The Netherlands, in 2011.

From 2006 to 2013, he was with NXP Semiconductors Research, Eindhoven, The Netherlands, where he researched fully integrated CMOS frequency references, nanometer temperature sensors, and area-efficient interfaces for magnetic sensors. In 2013, he joined TU Delft, where he is currently an Associate Professor. He has authored or co-authored one book, 12 patents, and over 120 technical publications. His main research interests are cryogenic electronics, quantum computing, sensor read-outs, and fully integrated frequency references.

Dr. Sebastiano has been on the technical program committee of the ISSCC, IMS, and the IEEE RFIC Symposium. He is currently serving as an Associate Editor of the IEEE JOURNAL OF SOLID-STATE CIRCUITSand has also served as a Guest Editor for IEEE JOURNAL OF SOLID-STATE CIRCUITSand as an Associate Editor of IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION SYSTEMS. He was a co-recipient of several awards, including the 2008 ISCAS Best Student Paper Award, the 2017 DATE Best IP Award, the ISSCC 2020 Jan van Vessem Award for Outstanding European Paper, the 2022 IEEE CICC Best Paper Award, and the 2023 Best Student Paper Award of the Joint ESSDERC/ESSCIRC tracks. He has served as a Distinguished Lecturer of the IEEE Solid-State Circuit Society.