

A Cryogenic Broadband Sub-1-dB NF CMOS Low Noise Amplifier for Quantum Applications

Yatao Peng[✉], Member, IEEE, Andrea Ruffino[✉], Member, IEEE, and Edoardo Charbon[✉], Fellow, IEEE

Abstract—A cryogenic broadband low noise amplifier (LNA) for quantum applications based on a standard 40-nm CMOS technology is reported. The LNA specifications are derived from the readout of semiconductor quantum bits at 4.2 K, whose quantum information signals are characterized as phase-modulated signals. To achieve broadband input matching impedance and low noise figure, the gate-to-drain capacitance of the input transistor is exploited. The goal is to involve a resistive and capacitive load into the input impedance match of a common-source stage with source inductive degeneration. The capacitive load is created by an *LC* parallel tank whose resonant frequency is lower than the operating frequency. The achieved non-constant in-band equivalent capacitance is proven to be beneficial to input impedance matching. The resistive part of the load is provided by the transconductance of the cascode stage implicitly. An inductor is added to the gate of the cascode transistor to suppress its noise, and a transformer-based resonator with two resonant frequencies serves as the load of the first stage, thus extending the operating bandwidth. Design considerations for the cryogenic temperature operation of the LNA are proposed and analyzed. The LNA achieves a measured gain (S_{21}) of 35 ± 0.5 dB, return loss > 12 dB, and NF of 0.75–1.3 dB across the band (4.1–7.9 GHz), with 51.1-mW power consumption at room temperature, while it shows a measured gain of 42 ± 3.3 dB, and NF of 0.23–0.65 dB with 39-mW power consumption at 4.2 K between 4.6 and 8 GHz. To the best of our knowledge, this is the first report of a cryogenic LNA based on a bulk CMOS process working above 4 GHz showing sub-1-dB NF both at room and cryogenic temperatures.

Index Terms—Capacitive loading, cryogenic CMOS, low noise amplifier (LNA), noise reduction, quantum computing, qubit readout, transformer.

I. INTRODUCTION

QUANTUM computing platforms have already reached the level of tens of qubits and they are rapidly progressing, so a full-fledged quantum computer is expected to be built in the near future [1]. Along with it, developing high-accuracy and low noise quantum-state readout and control electronics has become an active study field [2]. The required radio frequency control and readout circuitries are currently implemented by discrete devices wired to the quantum devices on one end and to room-temperature instruments on the other [3], [4]. In order to reduce the system complexity for

Manuscript received July 27, 2020; revised December 29, 2020 and February 19, 2021; accepted April 8, 2021. Date of publication April 26, 2021; date of current version June 29, 2021. This article was approved by Associate Editor Alyosha Molnar. This work was supported by the European Union's Horizon 2020 Research And Innovation Program under Grant 688539. (Corresponding author: Yatao Peng.)

The authors are with the Advanced Quantum Architecture Laboratory, École Polytechnique Fédérale de Lausanne (EPFL), 2002 Neuchâtel, Switzerland (e-mail: yatao.peng@epfl.ch).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/JSSC.2021.3073068>.

Digital Object Identifier 10.1109/JSSC.2021.3073068

large-scale qubit platforms, the readout and control circuits are proposed to be integrated on a single chip working at cryogenic temperature. This also reduces the round-trip delay across cables and filters, thus improving the overall system's fidelity [5].

The cryogenic low noise amplifier (cryo-LNA) is a critical block in the qubit readout chain since it is the first block that magnifies the weak quantum information signals from the qubits. RF cryo-LNAs were mainly implemented using III–V devices [6], [7] or silicon-germanium (SiGe) HBTs [8] for a long time because of their superior cryogenic RF characteristics. Recently, it was verified in [2] that deep-sub-micrometer CMOS transistors work well at deep cryogenic temperature, thus providing an indication that cryo-circuit blocks can be designed to reach fully integrated readout and control for quantum processors in standard CMOS technology. This is very attractive due to its low cost, high flexibility, and extreme reliability.

Early attempts to reach this goal started with a 500-MHz cryo-CMOS LNA for quantum computing designed in a 0.16- μ m bulk CMOS process [9]. The LNA was based on a noise-canceling topology and achieved \sim 0.1-dB noise figure (NF) at 4.2 K. In [10], a narrowband (2.025–2.12 GHz) 77-K LNA fabricated in 0.18- μ m CMOS technology was presented for space applications and it showed 0.5-dB cryogenic NF. Moreover, a 7-GHz cryogenic CMOS parametric amplifier with > 1.25 -dB NF at 8.5 K was presented in [11]. To the best of our knowledge, no works on CMOS LNA with sub-1-dB NF operating at high frequency (\sim 6 GHz) have been reported to date. In particular, designs for qubit readout requiring high performance at cryogenic temperature are not yet reported. In this work, we present a broadband cryo-CMOS LNA with \sim 0.23-dB NF at 4.2 K and 0.75-dB NF at room temperature.

Common approaches for wideband CMOS LNAs include common-gate (CG) amplifiers [12], noise-canceling (NC) amplifiers [13], and common source (CS) amplifiers with broadband *LC*-ladder input impedance matching [14]. CG amplifiers are broadband since the input impedance matching can be easily achieved by proper sizing of the transistor transconductance. These structures, however, suffer from inferior NF (> 2 dB), which is heavily deviating from the intrinsic minimum NF NF_{min} of the device. The NC amplifier utilizes an auxiliary amplifier to generate an out-of-phase noise signal path with respect to the main amplifier and adds the noise signals at the output port. Ideally, the noise of the main amplifier can be totally nullified. However, the unequal parasitic capacitances of the auxiliary and main signal paths and especially the deterioration of mismatch for

CMOS devices at cryogenic temperature [15] result in noise phase misalignment at the output port, which weakens the effect of noise-cancellation at high frequencies. In addition, the auxiliary amplifier still contributes considerable noise. Thus, these structures are not attractive for high-frequency qubit readout. In this article, the CS cascode amplifier with source inductive degeneration (ID) and resistive and capacitive ($R-C$) loading impedance is adopted as the basic structure [16], [17]. Several modifications are applied to suit our application. First, the off-chip large inductor (> 100 nH), which increases the noise contribution of the cascode transistor, is discarded, so as to achieve a fully integrated LNA. Second, an inductor is added to the gate of the cascode transistor to suppress its noise contribution and further lower the NF. Third, a transformer-based tank is utilized as the load of the first stage to extend the bandwidth. In addition, some critical considerations for cryogenic temperature (4.2 K) operation are also proposed in the LNA design.

This article is organized as follows. Section II introduces the gate-dispersive spin qubit readout circuits and analyzes the specifications for the LNA. Section III details the LNA topology and analyzes its gain, input impedance, and NF characteristics. Section IV illustrates the design considerations for the cryogenic operation of the LNA. Based on our analysis, the implementation and circuit parameters of the wideband cryo-LNA designed in a standard 40-nm CMOS process are illustrated. The measurement results and cryogenic NF measurement method are shown in Section V. Finally, Section VI concludes this article.

II. SEMICONDUCTOR SPIN QUBIT READOUT FRONT END CIRCUIT OVERVIEW

Radio frequency reflectometry techniques were proposed for semiconductor spin qubit (or quantum dot) readout to enhance charge sensitivity by one order magnitude compared with conventional dc-based techniques [18]. The RF-reflectometry diagram for a single qubit readout is depicted in Fig. 1(a). It is an *in situ* gate dispersive readout technique, which suppresses external electrometers and, hence, it is a promising candidate for scaled-up qubit arrays [19], [20]. The single qubit device is a double-gate nanowire FET implemented in a fully-depleted silicon-on-insulator (SOI) process. L -type matching network (L_M and C_M) is used to convert the high gate impedance into the 50Ω cable impedance. In such a technique, the semiconductor quantum device is addressed at its gate with an RF carrier signal and the phase of the reflected signal is read out. According to the state of the qubit, the capacitance at the gate will be different, due to the presence in the $|1\rangle$ state of an additional quantum or tunneling capacitance ΔC with respect to the state $|0\rangle$. This contributes to a phase shift $\Delta\varphi$ in the response of the $L-C$ matching network, which carries the quantum information about the qubit state, as shown in Fig. 1(b). The phase shift $\Delta\varphi$ is expressed as

$$\Delta\varphi \cong \tan^{-1} \left[\left(\frac{1 - \Gamma_0}{1 + \Gamma_0} \right) Q_M \cdot \frac{2\Delta C}{C_p + C_M} \right]. \quad (1)$$

Here, C_p is the parasitic gate capacitance of the quantum device, Q_M is the Q -factor of the matching network, and Γ_0

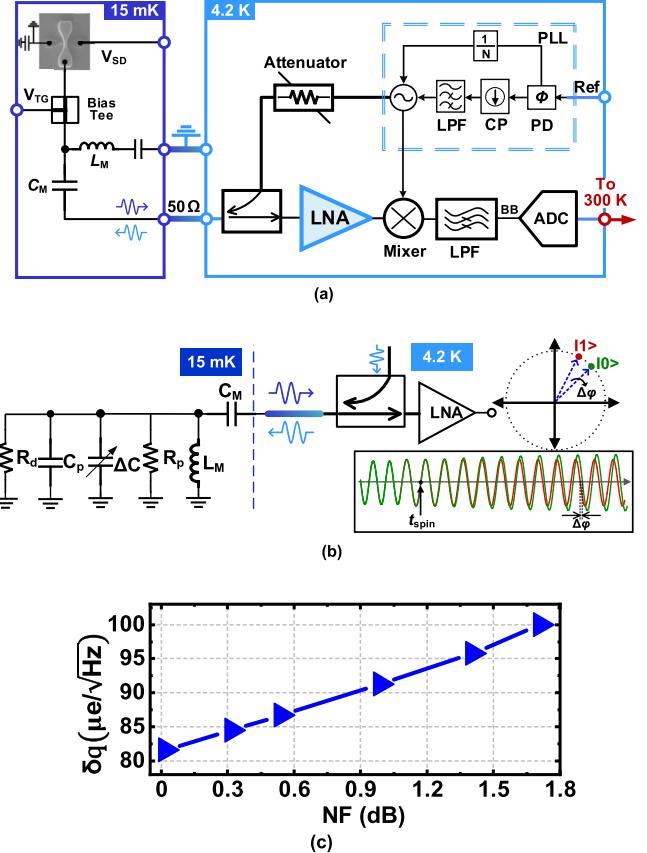


Fig. 1. (a) Semiconductor spin qubit readout front end diagram. (b) Simple model for the demodulation of quantum information signals [19]. R_d is the gate to ground parasitic resistance of the quantum device and L_M is the equivalent resistance of the inductor. At t_{spin} , the spin state of the qubit changes and the phase of the reflected signal changes by $\Delta\varphi$ after stabilization. (c) Readout charge sensitivity dependence on NF.

is the return loss of state $|0\rangle$. The tunneling capacitance ΔC is generally less than 1 fF [19]. To increase $\Delta\varphi$, the impedance matching network should be endowed with a high Q -factor and well matched to 50Ω . For instance, given $Q_M = 50$, $C_p + C_M = 200$ fF, $\Gamma_0 = 0.05$, and $\Delta C = 0.5$ fF, $\Delta\varphi$ is $\sim 12.5^\circ$. The reflected signal will then show a 12.5° phase shift compared with the original signal. This phase shift can be readily extracted by I/Q demodulation when the probe signal and the local oscillator (LO) signal are synchronized well. In general, we can model the reflected signals as phase-modulated M-PSK signals with $M > 360^\circ/\Delta\varphi$. Thus, we model the reflected signals as a pair of 32-PSK-modulated signals; the phase shift resolution of which is $11.25^\circ < 12.5^\circ$, assuming some margin. The minimum signal-to-noise ratio (SNR) for demodulating 32-PSK signals with a 0.01% bit error rate (BER) is 21.8 dB [21].

The noise at the output of the LNA at cryogenic temperature is mostly contributed by the LNA and the insertion loss of the directional coupler or the circulator (~ 0.2 dB). The charge sensitivity (δq) is inversely proportional to the readout bandwidth BW [22]. The NF requirements of the LNA for a required charge sensitivity can be then calculated based on the receiver sensitivity equation [23] once the parameters of

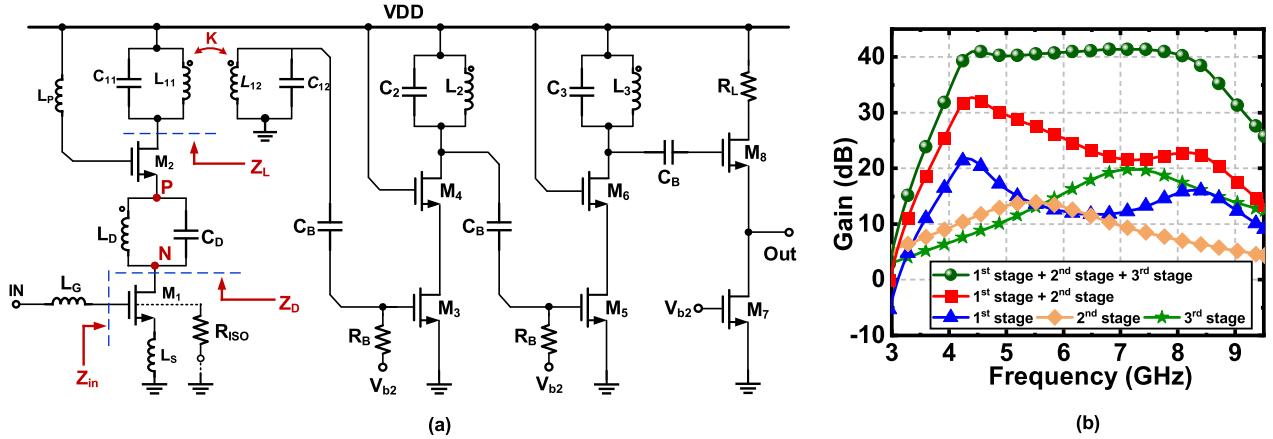


Fig. 2. (a) LNA structure. (b) Simulation gains of the different stages.

the quantum device and the matching network are known, as shown in Fig. 1(c). In the calculation, the probe signal power is set to -100 dBm, which is weak enough to avoid affecting the quantum state (kick-back) while achieving high charge sensitivity [19].

When scaling up the number of qubits, frequency multiplexing should be used. For this, N rows of $L-C$ matching networks centered at different but unique frequencies (f_1, f_2, \dots, f_N) should be designed to guide the signals of each quantum device (or each cluster of quantum devices) and act as a frequency multiplexer (F-MUX). In [24], a 10:1 $L-C$ F-MUX is implemented using a GaAs process for a 10-qubit readout. The multiplexer works around 1.2 GHz and occupies an area of 9.5×4.5 mm 2 , which is yet too large for a scenario where hundreds of qubits exist. Nevertheless, the trend in this direction is clear. Sub-1-GHz frequencies are extensively used in RF-reflectometry-based qubit detection, because low frequency is more suitable for discrete RF circuit design, as used so far. In our proposal, for a fully integrated approach, we choose higher readout frequencies up to 6 GHz since at higher frequencies, a more compact footprint of integrated $L-C$ matching networks can be achieved, along with higher Q -factor inductors in the readout circuit, wider absolute bandwidth to handle more qubits, and higher charge sensitivity [25].

Assuming the Q -factor of the $L-C$ network of the n th quantum device is Q_n (3-dB readout bandwidth f_n/Q_n), to reduce the crosstalk between the rows and ensure a realizable F-MUX design, frequency spacing $\Delta f_n = f_n - f_{n-1}$ between row $n-1$ and row n should be reasonably large. One can define $X = f_n/\Delta f_n$, where X is determined by Q_n and probe signal quality. The bandwidth for N -qubit readout can be then calculated as

$$BW = \sum_{n=1}^{N-1} \left(\frac{f_n}{Q_n} + \frac{f_n}{X} \right). \quad (2)$$

Assuming $Q_n = 50$ and $X = Q_n$, which means the frequency separation is the same as the readout bandwidth, the LNA bandwidth required for 20 qubits is calculated as ~ 3.6 GHz with adequate design margin to surmount the parameters'

TABLE I

LNA SPECIFICATIONS AT 4.2 K FOR GATE DISPERSIVE QUBIT READOUT

SPECS	Frequency	BW	Gain	NF	P _{1db}	Power
Values	6 GHz	3.6 GHz	>40 dB	<0.8 dB	>-80 dBm	<50 mW

deviation due to temperature and process variation. When time-multiplexing is also applied [20], [26], >100 qubit readout can be achieved. The power consumption is set to 1 mW/qubit for the entire readout, due to cooling power limitations [9], while the gain is set to 40 dB to reduce the gain requirement for broadband intermediate-frequency signals. Note that the NF of RF CMOS amplifiers scales down limitedly at cryogenic temperatures due to shot noise. Thus, we target an LNA with NF ~ 1 dB at room temperature (300 K). Based on the above analysis, the target specifications of the LNA are listed in Table I.

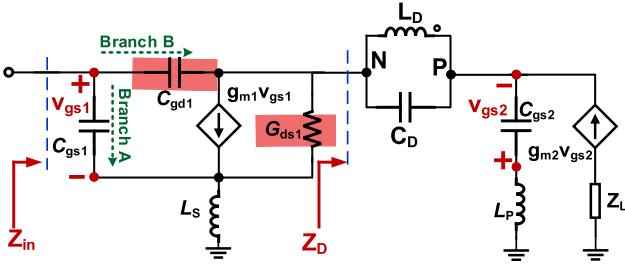
III. ANALYSIS OF THE CAPACITIVE LOADED LNA

The LNA topology is shown in Fig. 2(a): it features three cascode stages, i.e., one input low noise stage and two gain boost stages, with an output buffer for impedance matching. The first stage is a CS cascode amplifier with ID, loaded by an LC parallel tank (L_D and C_D). The gate-drain capacitance of the input transistor (M_1) C_{gd1} will involve the L_D-C_D tank and the transconductance of M_2 into the input impedance matching while virtually not affecting noise. The transformer-based tank constructed by $L_{11}, C_{11}, L_{12}, C_{12}$ is applied as the load of the cascode stage to extend the operating bandwidth.

A. Gain Analysis

The effective transconductance of the CS stage with source ID can be deduced as

$$G_{m1} = \frac{g_{m1}}{j\omega C_{gs1}(j\omega L_S + j\omega L_G + R_S) + j\omega g_{m1}L_S} \frac{g_{m2}}{j\omega C_{gs2} + g_{m2}} \quad (3)$$

Fig. 3. Small-signal equivalent circuit of the input stage for Z_{in} derivation.

where g_{m1} (g_{m2}) and C_{gs1} (C_{gs2}) are the transconductance and gate–source capacitance of M_1 (M_2) and R_S is the source impedance. When the input impedance is matched and $g_{m2} \gg 1/j\omega C_{gs2}$, then $G_{m1} \cong g_{m1}/[2(1+j\omega L_S g_{m1})]$. The first stage gain is calculated as $G_{m1} Z_L$, where Z_L is the input impedance of the transformer-based LC-tank. Its two resonant frequencies are [27]

$$f_{1,2}^2 = \frac{1 + \eta \mp \sqrt{1 + \eta^2 - 2\eta(1 - k^2)}}{8\pi^2 L_{12} C_{12}(1 - k^2)} \quad (4)$$

where $\eta = L_{12}C_{12}/L_{11}C_{11}$ and k is the coupling coefficient. We set $f_1 = 4.2$ GHz, $f_2 = 8.2$ GHz to cover the operating band with $k \cong 0.55$. As shown in Fig. 2(b), the simulated gain of the first stage is higher than 11 dB between 3.7 and 9 GHz, which is sufficient to suppress the noise contribution of subsequent stages by a factor ~ 12.6 . The magnitude ratio of the impedances at resonances is $R_{p2}/R_{p1} \approx (1-k^2)(1+\eta)/6$ [27]. R_{p2} tends to be smaller than R_{p1} when η is set to a reasonable value (<5), which results in gain inequality at resonances. Also, the gain decreases as the frequency deviates from f_1 and f_2 , thus a deep gain valley appears at the central frequency between the two peaks. The large gain ripples of the first stage are flattened by the second and third stages. The second and third cascode stages are loaded by LC parallel tanks resonating at 5.5 GHz with a 6.5 Q -factor and 7.2 GHz with an 8.2 Q -factor, respectively. The gain ripple of the three-stage amplifier is within ± 0.5 dB in the operating band [Fig. 2(b)]. The transformer-based tank load has the advantage of being compact and consuming low power for wideband operation. Moreover, suppression of harmonics is also achieved thanks to its bandpass characteristics.

B. Input Impedance Match Analysis

The input impedance of the first stage of the LNA (Z_{in}) is derived based on the small-signal equivalent circuit in Fig. 3

$$Z_{in} = \frac{(1 + j\omega C_{gd1} Z_D)(1 + T) + Z_D G_{ds1} (1 - \omega^2 L_S C_{gs1})}{j\omega \{ (Z_D + j\omega L_S) [G_{ds1} C_T + C_{gd1} (g_{m1} + j\omega C_{gs1})] + C_T \}} \quad (5)$$

$$Z_D = \frac{1 - \omega^2 L_P C_{gs2}}{j\omega C_{gs2} + g_{m2}} + \frac{j\omega L_D}{1 - \omega^2 L_D C_D} \quad (6)$$

where $C_T = C_{gs1} + G_{ds1}$, $T = j\omega L_S(g_{m1} + j\omega C_{gs1} + G_{ds1})$. Z_D is the impedance seen from the drain of M_1 and g_{DS1}

is the output conductance of M_1 . Z_{in} is composed of two parallel networks from Branch A by C_{gs1} and Branch B by C_{gd1} . If $C_{gd1} = 0$ and $g_{DS1} = 0$ are assumed, only Branch A in Fig. 3 conducts ac signals and performs the input impedance matching. The real part of Z_{in} is $\text{Re}\{Z_{in}\} = g_{m1} L_S / C_{gs1}$, which is independent of frequency, while the imaginary part of Z_{in} , $\text{Im}\{Z_{in}\} = \omega L_S - 1/\omega C_{gs1}$, is canceled by L_G , which results in the narrowband behavior for the classical CS-inductively source degenerated LNA [28]. For a wideband LNA design, the transistor size (W) of M_1 needs to be set large enough to achieve less-sensitive noise circles (dependent on frequency and source impedance) while obtaining a high g_{m1} (gain), then C_{gd1} is generally large enough to involve Z_D into input impedance matching. From a qualitative perspective, if frequency is high enough, the impedance seen from branch B is resistive and determined by G_{ds1} and $\text{Re}\{Z_D\}$, which is less sensitive to frequency. When taken in parallel with the impedance seen from Branch A, the real part of which ($\sim g_{m1} L_S / C_{gs1}$) is also insensitive to frequency, then the real part of Z_{in} can be designed to be $\sim 50\text{-}\Omega$ wideband. On the other hand, $\text{Im}\{Z_{in}\}$ is nulled by the resonance of an effective capacitance introduced by the capacitive load and the effective inductance due to L_S , without using matching components at the input port [16]. Due to the low Q -factor of the resonance, relatively broadband matching was achieved. In [16], Z_D is realized by a resistor in series with a capacitor explicitly, connecting in parallel to an RF choke for voltage supply. This, however, relies on the gallium arsenide transistor's excellent intrinsic gain (g_m/G_{ds}) performance to achieve high-voltage gain without a cascode transistor for the input transistor to suppress the noise contribution of the subsequent gain stages.

In general, a cascode transistor is imperative for the input CS-transistor to enhance output impedance and thus the intrinsic voltage gain in deep-submicrometer CMOS processes. For the cascode structure, an $L-C$ parallel tank was proposed in [17] to be inserted between the drain of the input transistor and the cascode transistor so as to create a capacitive load, while the resistive part is provided by the transconductance of the cascode CG stage implicitly. In [17], the inductor in the tank is an off-chip 100-nH inductor, to obtain a more pronounced and easily controlled capacitive load. However, the off-chip connection introduces extra parasitic capacitances at nodes P and N due to the bonding pads and PCB routes, which increases the noise contribution of M_2 . Hence, we integrate the LC -tank (L_D and C_D) on chip. Neglecting the parasitic capacitances at nodes P and N, $Z_D \approx 1/g_{m2} + j\omega L_D / (1 - \omega^2 L_D C_D)$ given that $g_{m2} \gg \omega C_{gs2}$ and $\omega^2 L_D C_{gs2} \ll 1$. If the resonant frequency $1/(L_D C_D)^{1/2}$ is well below the operating frequency, we still can have a proper capacitive load at the drain of M_1 . Now, the effective capacitance $C_D^e(\omega)$ increases with frequency in the operating band. This is different from the cases in [16] and [17], where the effective capacitance is almost constant in the operating band due to the large L_D or RF choke. The small $C_D^e(\omega)$ at low frequency may lead to a real part of impedance seen from branch B more sensitive to frequency. However, as the calculated results of (5) and a simulation using BSIM4 transistor models shown in Fig. 4(a), with the help of non-constant $C_D^e(\omega)$, $\text{Im}\{Z_{in}\}$

Type, should be $C_{gs1} + C_{gd1}$ unless unit mismatch

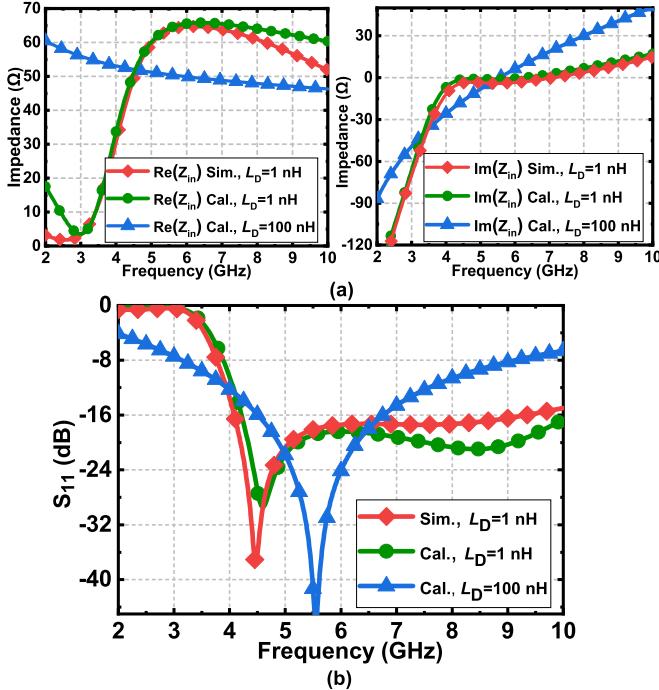


Fig. 4. Calculation and simulation results of (a) Z_{in} and (b) S_{11} for the first stage. The circuit parameters are: $L_G = 0.85 \text{ nH}$, $L_S = 0.25 \text{ nH}$, $C_{gs1} = 0.24 \text{ pF}$, $C_{gd1} = 90 \text{ fF}$, $C_D = 2.5 \text{ pF}$, $G_{ds1} = 33 \text{ mS}$, $g_{m1} = 0.22 \text{ S}$, $g_{m2} = 0.1 \text{ S}$ for $L_D = 1 \text{ nH}$, while $L_G = 0.9 \text{nH}$, $L_S = 0.19 \text{ nH}$, $C_D = 2.9 \text{ pF}$ are set to obtain optimum S_{11} for $L_D = 100 \text{ nH}$.

remains near zero in a broadband. This result is better than using a large LD where the imaginary part increases constantly. The gate inductor (L_G) used for noise matching can further enhance input impedance matching performance. In our design, we also adjust g_{m2} to optimize S_{11} ; hence, the dimensions of M1 and M2 are not necessarily identical. The comparison of S_{11} calculation ($L_D = 1$ and 100 nH) and simulation results ($L_D = 1 \text{ nH}$) is shown in Fig. 4(b), and both cases can achieve wideband matching.

At the resonant frequency of L_D and C_D , the high impedance of $|Z_D|$ may induce the LNA to oscillate. To avoid oscillation, an off-chip high-pass bias circuit was applied in [17] to guide the potential low-frequency oscillating signals to ground. Actually, as long as $\text{Re}\{Z_{in}\}$ is positive, the circuit is stable. This condition is not so difficult to satisfy because of the limited Q -factor of the tank and the high G_{ds1} of the short-channel transistor. Relatively large L_S and lower g_{m2} help to ensure $\text{Re}\{Z_{in}\} > 0$. Moreover, the multistage cascode structure of the LNA has good isolation, which ensures global stability. The stability simulation results of the LNA are shown in Fig. 5. There is no overlap between the source stability circles (SSCs) and the unit return loss circle at different frequencies, which illustrates the stability of the designed LNA.

C. Noise Analysis of M_1

The noise analysis of the LNA is based on the small-signal equivalent circuit including the noise sources of the first stage shown in Fig. 6. We neglected the noise contribution of L_S and the gate-induced noise. The noise of R_g and the channel thermal noise are considered in our design. According to linear

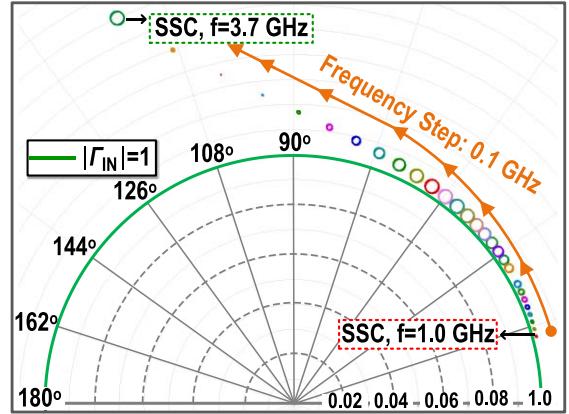


Fig. 5. SSCs versus frequency.

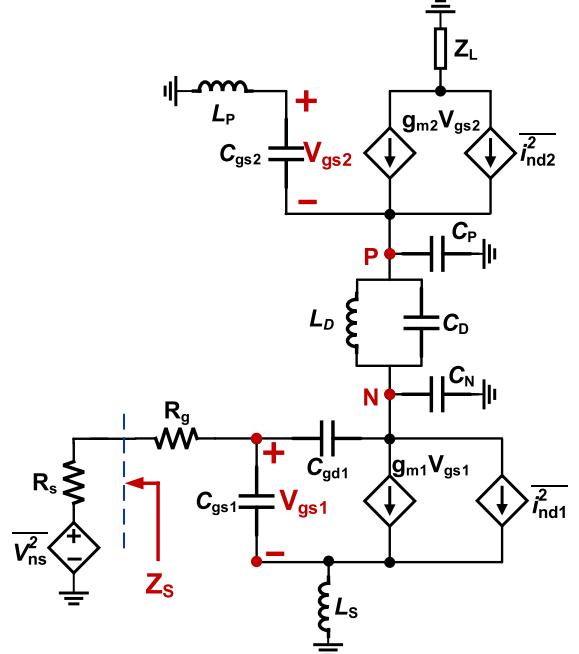


Fig. 6. Small-signal equivalent circuit for NF analysis.

noise network circuit theory [29], the noise parameters of M_1 with L_S (F_{\min} , $Z_{\text{opt}} = R_{\text{opt}} + jX_{\text{opt}}$) and its NF (F_{M1}) can be derived as

$$Z_{\text{opt}} = \sqrt{R_g^2 + \frac{\alpha g_{m1} R_g}{\gamma \omega^2 C_T^2}} + j \left(\frac{1}{\omega C_T} - \omega L_S \right) \quad (7)$$

$$F_{\min} = 1 + 2 \frac{\gamma \omega^2 C_T^2}{\alpha g_{m1}} \left(\sqrt{R_g^2 + \frac{\alpha g_{m1} R_g}{\gamma \omega^2 C_T^2}} + R_g \right) \quad (8)$$

$$F_{M1} = F_{\min} + \frac{\gamma \omega^2 C_T^2}{\alpha g_{m1} R_S} \left[(R_S - R_{\text{opt}})^2 + (X_S - X_{\text{opt}})^2 \right] \quad (9)$$

where R_S and X_S are the real and imaginary part of the impedance seen from the gate of M_1 , i.e., $Z_S = R_S + jX_S$, and γ is the noise excess factor. From (8), F_{\min} is determined by a parameter $P = R_g C_T^2 / \alpha g_{m1}$. For a short-channel MOSFET working in the saturation region, we have [28]

$$I_D = W C_{\text{ox}} v_{\text{sat}} \frac{V_{\text{od}}^2}{V_{\text{od}} + \text{LE}_{\text{sat}}} \quad (10)$$

$$g_{m1} = \mu_{\text{eff}} C_{\text{ox}} \frac{W}{L} V_{\text{od}} \alpha \quad (11)$$

$$\alpha = \frac{1 + 0.5 \cdot \rho}{(1 + \rho)^2} \quad (12)$$

where W/L is M_1 's dimension, $\rho = V_{\text{od}}/\text{LE}_{\text{sat}}$, V_{od} is the gate overdrive voltage, E_{sat} is the velocity saturation field strength, μ_{eff} is effective channel mobility, and v_{sat} is the saturation velocity. Based on (8) and (10)–(12), small W and high V_{od} are preferred to lower P , thus F_{min} . On the other hand, V_{od} needs to follow $V_{\text{od}} < V_{\text{ds}}$ well to keep M_1 in the saturation region, and W needs to adhere to metal current density rules for circuit reliability considerations. Moreover, a small transistor size may lead to Z_{opt} and Z_{in} impossible to match using on-chip components. Thus, F_{min} has a limited optimization margin (<0.1 dB) for a power-constrained design. Generally, noise impedance matching is more significant compared with lowering F_{min} when the transistor is working in moderate inversion [30]. *LC*-ladder matching networks can be exploited for a broadband match of 50Ω to Z_{opt} [14], but due to the loss of inductors, even at cryogenic temperature [31], more NF performance is easily lost rather than gained. In our design, a single inductor L_G is used at the gate. Thus, for noise matching, we have $\text{Re}\{Z_{\text{opt}}\} = 50 \Omega$ and $\text{Im}\{Z_{\text{opt}}\} = \omega L_G$, that is

$$\frac{\alpha g_{m1} R_g}{\gamma \omega^2 C_T^2} = 50 \quad (13)$$

$$\frac{1}{\omega C_T} - \omega L_S = \omega L_G. \quad (14)$$

Since F_{min} rises rapidly with frequency according to (8), the matching frequency is set slightly higher than the center frequency to balance the in-band NF variation. The simulated noise circles of M_1 without L_S are plotted in Fig. 7. It can be seen that both $\text{Re}Z_{\text{opt}}$ and $\text{Im}Z_{\text{opt}}$ decrease with frequency, which is consistent with (7), and the discrepancies between calculation and simulation of Z_{opt} are because of the neglected gate-induced noise and $G_{\text{ds}1}$'s noise in the analysis. The choice of Z_{opt} needs to consider the noise of the transistor and the extra noise brought by the necessary inductors for noise match. If $Z_{\text{opt}} \sim (50 + j80)\Omega$ at 7 GHz is set as the target matching impedance, $L_S + L_G$ is ~ 1.5 nH based on (14), which tunes 50Ω to $Z_{\text{LF}} \sim (50 + j45)\Omega$ at 4 GHz (low-frequency edge f_L), and Z_{LF} locates on a noise circle of 4 GHz with $\text{NF} = 0.49$ dB, deviating only 0.15 dB (ΔNF) from its NF_{MIN} . Another observation is that, if we match 50Ω to an impedance $Z_{\text{OPT}}^A \sim (49 + j50)\Omega$ at 7 GHz instead of Z_{OPT} , NF degrades only 0.1 dB at 7 GHz, while the necessary inductance $L_G + L_S$ scales down by 38%, so the noise contribution of the inductors' parasitic resistances will reduce accordingly. For an inductor with Q -factor around 15, the NF reduction due to the inductance decrease is ~ 0.2 dB from simulation. If the inductor area is also considered, Z_{OPT}^A is more attractive. As shown in Section IV, setting a target impedance that needs a smaller inductance for noise matching is more desirable also at low temperatures, since the NF is less sensitive to source impedance, while the NF of an inductor maybe comparable with NF_{min} at cryogenic temperature.

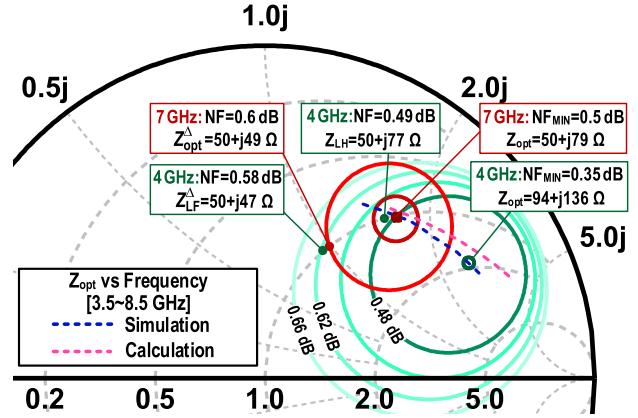


Fig. 7. Noise circles and NF_{min} versus frequency of the input transistor.

D. Noise of the Cascode Transistor

The channel noise of the cascode transistor M_2 ($i_{\text{nd}2}^2$) contributes considerable noise at high frequency if the parasitic capacitances at nodes P and N (C_P and C_N) are too large to be ignored [32]. C_N is composed by gate-to-drain capacitance of M_1 ($C_{\text{gd}1}$) and drain-to-bulk capacitance of M_1 ($C_{\text{db}1}$), while C_P is comprised of source-to-bulk capacitance of M_2 ($C_{\text{sb}2}$). In addition, both C_N and C_P are raised by the routing capacitance ($C_{\text{rt}N}$ and $C_{\text{rt}P}$). Thus, $C_N = C_{\text{gd}1} + C_{\text{db}1} + C_{\text{rt}N}$ and $C_P = C_{\text{sb}2} + C_{\text{rt}P}$. The admittance at node P can be derived as

$$Y_P \cong j\omega C_P + \frac{j\omega C_N (1 - \omega^2 L_D C_D)}{1 - \omega^2 L_D (C_D + C_N)}. \quad (15)$$

If $C_D \gg C_N$, $Y_P \cong j\omega(C_P + C_N)$. Actually, the equivalent parasitic capacitance at node P with the $L_D - C_D$ tank is smaller compared with the case without $L_D - C_D$ tank, which is beneficial to design a smaller L_P as described below. The noise factor of M_2 is calculated as

$$F_{M2} = \frac{(Y_P - j\omega C_{\text{gs}2} - \omega^2 L_P C_{\text{gs}2} Y_N)^2 i_{\text{nd}2}^2}{4g_{m2}^2 G_{m1}^2 k_B T R_S} \left(\frac{Z_{\text{in}} + R_S}{Z_{\text{in}}} \right)^2. \quad (16)$$

Equation (16) reveals that if $L_P = (Y_P - j\omega C_{\text{gs}2})/\omega^2 C_{\text{gs}2} Y_N$, the noise contribution of M_2 will be nulled. On the other hand, L_P cannot be set too large since it also decreases $\text{Re}\{Z_{\text{in}}\}$ according to (6), which further enlarges the discrepancy of Z_{in}^* and Z_{opt} of M_1 . The layout parasitic extraction shows $C_P + C_N \cong 195$ fF and $C_{\text{gs}2} \cong 95$ fF. Then, L_P is calculated as 0.95 nH. Fig. 8 shows the simulated NF of LNA with and without L_P , where a ~ 0.1 -dB NF reduction thanks to L_P can be seen.

IV. CRYOGENIC CMOS LNA DESIGN AND IMPLEMENTATION

A. Design Considerations for Cryogenic Temperature Operation

Standard foundry PDKs do not provide device models at ultralow temperatures. Therefore, we designed the LNA based on the above analysis, using compact room temperature models of the PDK. Then, we took cryogenic behavior into

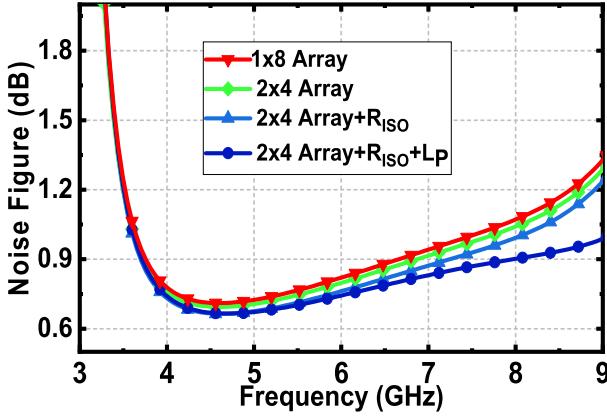
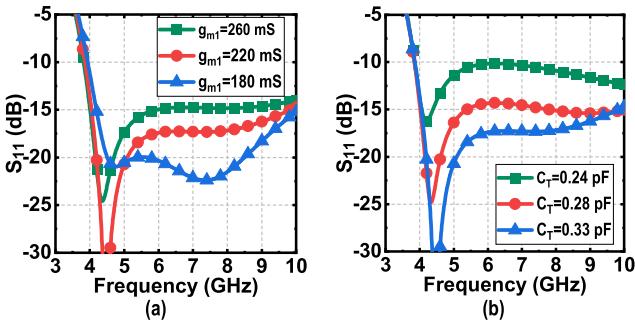


Fig. 8. NF comparison of the used noise optimization techniques.

Fig. 9. S_{11} dependence on (a) g_{m1} and (b) C_T .

consideration by electromagnetic (EM) simulations of passive inductive devices and by extracting the parameter variation of transistors and capacitors at cryogenic temperature from cryogenic measurement results in the literature. On top of that, some critical modifications have been applied to the LNA to adapt its operation to cryogenic temperature.

Due to the decrease of electron scattering in the channel, the electron mobility of the NMOS transistors increases $\sim 2 \times$, which will increase the transconductance [33], [34]. The S_{11} dependence on g_{m1} is shown in Fig. 9(a). It can be seen that return loss performance can be maintained well even assuming g_{m1} variation $>30\%$ at 4.2 K. On the other hand, the widening of semiconductor bandgap and the sharpness of the Fermi–Dirac statistics at low temperature raise the threshold voltage by 30%. Thus, all the cascode stages (M_2 , M_4 , and M_6) are designed with low threshold voltage transistors in the PDK, and the bias voltage of M_1 , M_3 , and M_5 can be set externally.

Since the gate-to-source/drain capacitances of the transistors have considerable influence on impedance matching and noise matching, their variation at cryogenic temperature should be taken seriously. For a transistor operating in the saturation region, the gate-to-source capacitance mainly consists of the oxide capacitance between the gate and the channel, $C_1 \propto WLC_{ox}$, where C_{ox} is the gate-oxide capacitance per unit area, and the capacitance due to the overlap of the gate polysilicon with the source and drain areas, $C_2 \propto C_{ov}W$, where C_{ov} is the overlap capacitance per unit area [35].

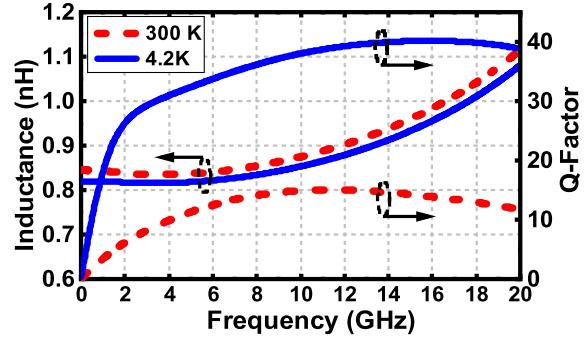


Fig. 10. EM simulation of an inductor at 300 and 4.2 K.

The second part also contributes to most of the gate-to-drain capacitance. At cryogenic temperature, on the one hand, the neutral regions around the interface of the polysilicon gate and n^+ source/drain regions will be frozen-out, resulting in the reduction of electric conduction [36], thus C_{ov} (which determines C_{gd1}) will decrease at 4.2 K. On the other hand, the depletion region at the polysilicon-SiO₂ interface enlarges at low temperature, which acts as an additional capacitor in series with the gate oxide capacitor [37]. This will reduce C_{ox} (which mainly determines C_{gs1}). Hence, we believe that the gate-to-source/drain capacitances of the MOSFET should reduce during cryogenic operation.

The measurement results in [37] demonstrated that gate-to-source/drain capacitances slightly decrease as the temperature is reduced from 300 to 100 K, but there is a steep decrease in capacitance below 100 K, and the total capacitance reduction is about 35% at 4.2 K compared with room temperature for a 90-nm channel length CMOS transistor. In [38], a negligible ($\sim 3.5\%$) change of capacitance is declared when temperature decreases from 293 to 6 K for a 32-nm SOI CMOS transistor. If also the 4.2-K measured results of a transistor with 0.7- μ m channel length in MOS-cap-type connection [36], where the capacitance decreases about 90%, are considered, then we can conclude that the reduction of transistor capacitances at cryogenic temperature will decrease with technology nodes. Using a linear fitting, we predicted that the capacitance of M_1 ($L = 60$ nm) decreases $\sim 18\%$ at 4.2 K. The S_{11} dependence on C_T is illustrated in Fig. 9(b), and it can be seen that S_{11} deteriorates when C_T decreases. The tolerance for C_T variation is $\sim 25\%$. For the above assumption of 18% capacitance reduction at cryogenic temperature for M_1 , an input matching $S_{11} < -10$ dB can still be maintained in theory.

Concerning the inductors (L_G , L_S , L_D , L_P , L_2 , L_3) and transformer (L_{11} and L_{12}), we accounted for their cryogenic characteristics by EM simulation. Fig. 10 shows a comparison of the EM simulation results of a ~ 0.85 -nH inductor at 300 and 4.2 K. In the simulation, the conductivity of the ultra-thick metal at 4.2 K is set to 5 \times the one at 300 K (5.8×10^7 S/m), and the substrate resistance is increased 1000 \times the value at 300 K ($10 \Omega \cdot \text{cm}^{-1}$) due to the freeze-out of the substrate. It can be seen that the inductance decreases slightly ($\sim 3\%$) at 4.2 K, while the Q -factor is 2.7 \times higher than at 300 K. Apart from the self-inductances and

Q-factor variation, EM simulation results of the transformer show that its coupling coefficient increases 5%, due to the change in the capacitive coupling, which draws the two resonances of the transformer-based tank apart by ~ 180 MHz. For the MoM capacitors, a negligible capacitance variation and sharp *Q*-factor increment at 4.2 K are shown from measurement [31]. The increase of the *Q*-factors of passive devices will boost the gain at the resonances of the loaded *LC* tanks. This will deteriorate the in-band gain flatness, which increases the design complexity of the analog-to-digital converters in the readout chain. Thus, the inductors are put in series with small polysilicon resistors (L_2 with $6\ \Omega$, L_3 with $4\ \Omega$), so that the *Q*-factor of the *LC* tanks will be partially determined by the polysilicon resistors and not only by the metal parasitic resistance. This reduces the in-band gain fluctuation by more than 4 dB at cryogenic temperature.

Now, we discuss the noise design considerations for the cryogenic temperature (4.2 K) operation of the LNA. With the shrinking of transistor dimensions to deep-submicrometer, quasi-ballistic transport of the carriers causes shot-like noise to appear in the channel. Therefore, the channel noise is not solely thermal noise. The channel noise including both noise types can be expressed as

$$\frac{\bar{I}_{nd}^2}{\Delta f} = 4k_B T \frac{\gamma}{\alpha} g_{m1} \cdot (1 - \tau)^2 + 2q\sigma I_D \cdot \tau^2 \quad (17)$$

$$\tau = \frac{R_b}{R_{ch} + R_b} \quad (18)$$

where q is the elementary charge, R_b is the equivalent barrier resistance (inducing shot noise), which is inversely proportional to the gate-source voltage V_{gs} , R_{ch} is the channel resistance including the source and drain contact resistance, τ is known as the shot noise suppression factor [39], and σ is the excess noise factor for shot noise. For long-channel devices, $R_b \ll R_{ch}$, and the thermal noise model is effective. For 40-nm channel length transistors, the shot-like noise contributes $\sim 25\%$ of the total noise power [40]. For a given bias current, in contrast to thermal noise, which scales with temperature, shot noise is independent of temperature [41]. In general, for LNAs designed for room temperature operation, the minimum channel length (40 nm) is set since the NF is proportional to the channel length, as explained in Section III. However, if the LNA needs to target cryogenic applications, a smaller shot-like noise proportion is desirable, i.e., the larger channel length is preferable. Thus, in our design, the channel length of M_1 and M_2 is set as 60 nm to decrease the shot-like noise proportion. With this choice, we believe that the proportion to total noise can be less than 10%. For M_3 – M_8 , we still use the minimum channel length to reduce their parasitic capacitance. As analyzed above, the transistor capacitance (C_T) decreases at cryogenic temperatures. Apart from C_T , the cryogenic gate parasitic resistance R_g will also scale down $\sim 3\times$ from its room temperature value [38]. Also, a much smaller V_{od} can obtain the same g_{m1} at 4.2 K due to the increase of carrier mobility (by $3\times$). Thus, an increase of $\sim 10\%$ for α is calculated based on (12). These variations will influence the noise parameters considerably. The calculated noise circles at a signal frequency point (7 GHz) and the Z_{opt}

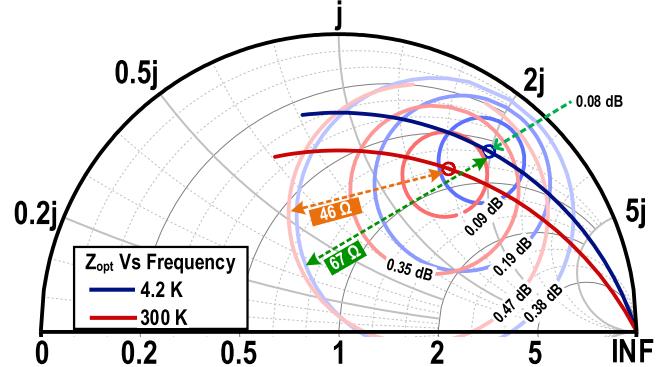


Fig. 11. Calculated noise circles and Z_{opt} versus frequency.

versus frequency (0–10 GHz) are shown in Fig. 11, where the noise excess factor γ is assumed to be the same for different temperatures [42]. As expected, $\text{Re}Z_{opt}$ decreases at 4.2 K due to the considerable drop of R_g , and $|\text{Im}\{Z_{opt}\}|$ increases due to the reduction of C_T . According to (9), $\partial F_{M1}/\partial R_S$ and $\partial F_{M1}/\partial X_S$ are proportional to the equivalent noise resistance ($R_n = \gamma \omega^2 C_T^2 / \alpha g_{m1}$); hence, the noise circles are less sensitive to the source impedance change compared with room temperature, due to the reduction of R_n at 4.2 K. For example, for an NF degradation (ΔNF) of 0.3 dB from NF_{\min} , the noise circle radiiuses are 67 and 46 Ω for 300 and 4.2 K, respectively. Therefore, the choice of the target noise matching impedance at cryogenic temperature is more flexible, without sacrificing NF too much. Moreover, a flatter in-band NF can be expected at 4.2 K. From the perspective of practical circuit design, we should comprehensively consider the above qualitative and idealized analysis and the PDK models. So, first, we obtained Z_{opt} and the noise circles at room temperature using the compact PDK models and we set the target noise matching impedance. Then, we scaled the impedance for cryogenic temperature operation by a proper proportion. Therefore, to keep the noise impedance match at cryogenic temperature, we needed to increase $L_G + L_S$ by $\sim 18\%$ from the room temperature design value (~ 0.9 nH) to compensate for the capacitance reduction at 4.2 K. The impact of the $L_G + L_S$'s scale-up on S_{21} , S_{11} and NF at room temperature (with 300-K EM model) is shown in Fig. 12, together with all the simulated results using the 4.2-K EM model. The drop of $\text{Re}\{Z_{opt}\}$ is partially compensated by the adjustment of g_{m1} through the bias voltage of M_1 .

B. LNA Implementation

Based on the above considerations, the LNA was designed according to the following steps: 1) for given power consumption (i.e., I_D and the power supply voltage), V_{od} and W/L for M_1 are determined to obtain a maximum g_{m1} while satisfying (13) at 7 GHz, where channel length is pre-set to 60 nm to reduce the shot-type noise; 2) the noise circles of M_1 are simulated and Z_{OPT}^Δ is set, so as to obtain the value of $L_G + L_S$ based on (14) at room temperature, which is then scaled up to compensate for the transistor capacitance reduction at cryogenic temperature; 3) the LD-CD tank and

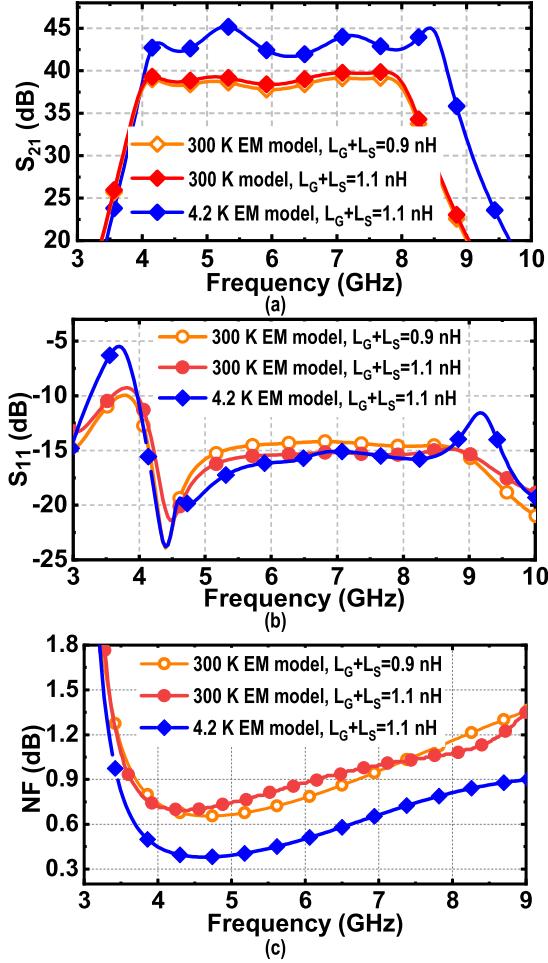


Fig. 12. Post-layout simulation results. (a) S_{21} , (b) S_{11} , and (c) NF for different design choices.

transistor M_2 are added, then S_{11} is optimized by L_D , C_D , g_{m2} , and proper distribution of LS and LG, while keeping LS + LG constant; 4) the parasitic capacitances at nodes P and N are extracted to define the inductance LP; and 5) the transformer-based tank, second, third stage, and the output buffer are designed to satisfy the gain specifications. After some optimizations, the main circuit parameters in Fig. 2(a) are illustrated in Table II. All the transistors are designed in deep N-wells to reduce substrate noise coupling. To further suppress the noise coupled through body terminals, large resistors RISO ($\sim 30 \text{ k}\Omega$) were used to bias the p-type body inside deep N-wells to the ground [Fig. 2(a)]. The noise reduction effect of RISO is demonstrated in Fig. 8.

The floorplan of the transformer of the first stage is illustrated in Fig. 13(a). It is a combined tapped and interleaved transformer to keep a compact area ($125 \times 115 \mu\text{m}^2$) and obtain a medium coupling coefficient ($k = 0.55$). The layout of the wide transistor M_1 is also critical for noise optimization since it determines R_g and the parasitic source resistance to some extent. To minimize the interconnect parasitic resistances, the large transistor is divided into eight unit cells. Each cell consists of 32 fingers with a width of $1.6 \mu\text{m}$. Each gate node (finger) was laid out from both sides by metal 1

TABLE II
LNA CIRCUIT PARAMETERS

Transistors	Dimensions	Components	Values
M_1	$384 \mu\text{m}/60 \text{ nm}$	L_G	0.86 nH
M_2	$160 \mu\text{m}/60 \text{ nm}$	L_D	0.96 nH
$M_3 \sim M_6$	$64 \mu\text{m}/40 \text{ nm}$	L_S	0.25 nH
M_7	$24 \mu\text{m}/40 \text{ nm}$	C_D	2.5 pF
M_8	$32 \mu\text{m}/240 \text{ nm}$	L_{11}	1.9 nH
		L_{12}	1.6 nH
		C_{11}	0.4 pF
		C_{12}	0.4 pF
		L_P	0.85 nH

and then connected to the ultra-thick top metal, which can theoretically reduce the effective gate resistance by a factor of 4 [43]. The eight cells were placed in a 2-row and 4-column array (2×4 array, fingers in vertical), as shown in Fig. 13(b). This floorplan shortens the effective interconnect route length of gate nodes of the units, and the post-layout simulation showed about 0.03-dB NF improvement when compared with the 1×8 array [17], as depicted in Fig. 8.

V. MEASUREMENT RESULTS AND DISCUSSION

The cryogenic LNA has been fabricated in a bulk 40-nm process with seven metal layers, including an ultra-thick metal. Fig. 14 shows the micrograph of the LNA die, and the total area, including pads, is $\sim 0.96 \times 0.64 \text{ mm}^2$. The die has been wire-bonded onto a PCB for dc biasing, while the RF signals have been tested by a pair of GSG probes. During the design stage, the influence of the dc bond wires has been considered. In the on-chip decoupling capacitors, the ground layers (metal layers 1, 3, 5, and 7) and power layers (metal layers 2, 4, and 6) are alternately stacked to cover free areas, and small capacitor ($\sim 200 \text{ fF}$) cells are evenly distributed on the ground and power layers, thereby ensuring that the decoupling structures are effective wideband. The LNA was measured in a Lake Shore CPX probe station at 300 K in vacuum (10^{-5} torr). The cryogenic temperature was achieved by flowing liquid helium (4.2 K) into the vacuum chamber. The chip was tested using a Keysight vector network analyzer (PNA-X N5245A), with embedded options for NF measurement and linearity measurement. The short-open-load-through calibrations were performed before the real circuit measurements, and the calibration power levels were set to -48 dBm at 300 K and -55 dBm at 4.2 K, respectively, which are within the linear region of the LNA.

At room temperature, the chip power supply was set to 1.4 V and the static current of the first stage was 24.5 mA, with an M_1 gate bias voltage $V_{b1} = 0.72 \text{ V}$. The other gain stages and the output source follower consumed about 12 mA in total, with bias voltage $V_{b2} = 0.75 \text{ V}$. The S-parameter measurement results are shown in Fig. 15. It can be seen that at 300 K, S_{21} is 35 dB with a good flatness ($< \pm 0.6 \text{ dB}$ variation) from 4.15 to 7.5 GHz, while the 3-dB bandwidth

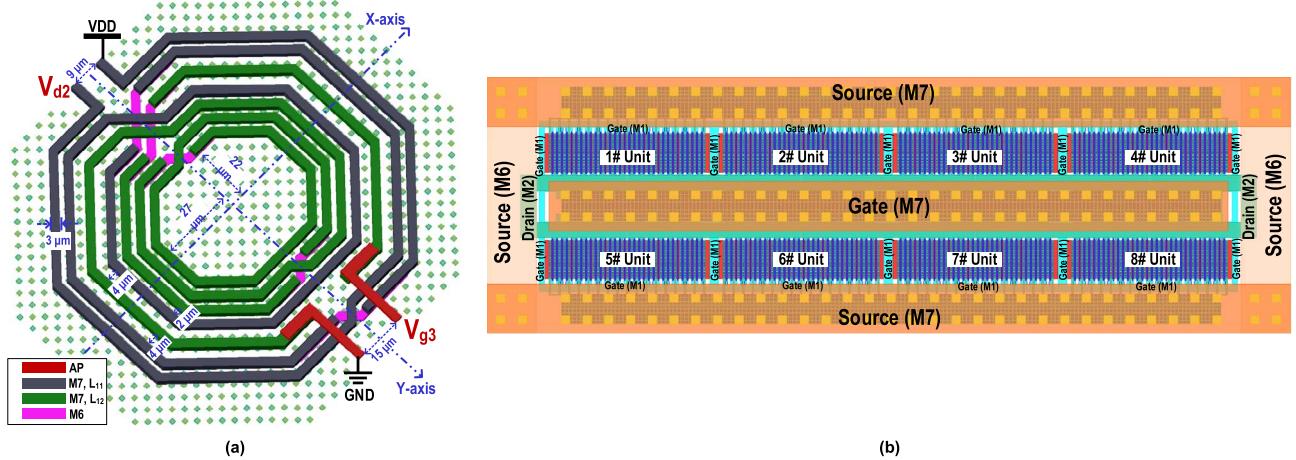


Fig. 13. (a) Layout of the transformer in the used 40-nm CMOS process with ultra-thick top metal. (b) M1 layout.

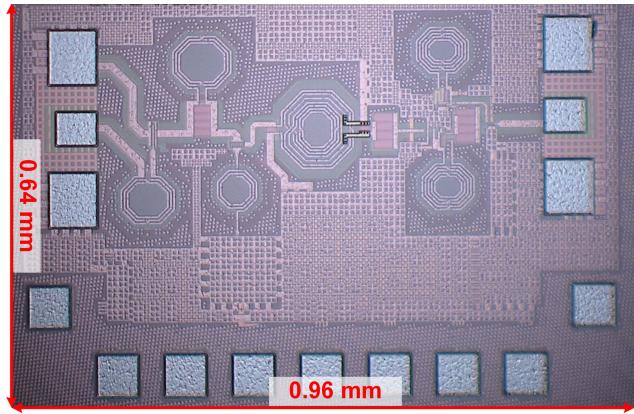


Fig. 14. Micrograph of the LNA.

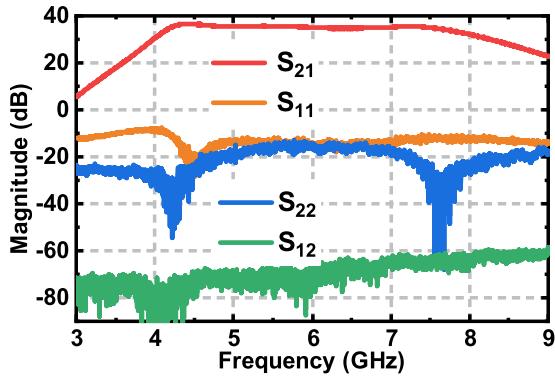


Fig. 15. S-parameter measurement results of the LNA at room temperature.

is 4.1–8.2 GHz ($\sim 65\%$ fractional bandwidth). Within the 3-dB frequency band, S_{11} is smaller than -10 dB, which indicates the broadband power matching works well at room temperature. At the same time, the output port shows good return loss ($S_{22} < -15$ dB) since the transconductance of the output buffer is designed to be ~ 20 mS.

At 4.2 K, the supply voltage was kept at 1.4 V. By tuning the input transistor bias, the bias current of the first stage was set to

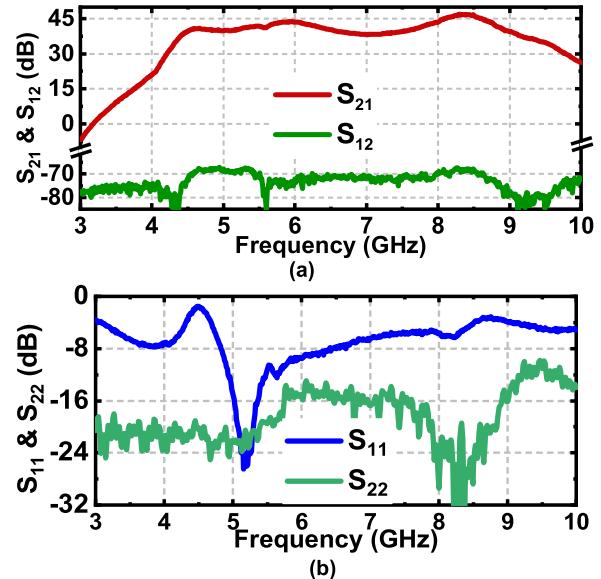


Fig. 16. S-parameter measurement results of the LNA at 4.2 K. (a) S_{21} and S_{12} . (b) S_{11} and S_{22} .

19.3 mA. The bias current change from room temperature was set to obtain low return loss at 4.2 K, and this is a reasonable adjustment that benefits noise optimization since the optimum noise impedance at cryogenic temperature tends to be closer to the conjugate of input impedance [38]. The second, third, and output stages consume a total of 8.6 mA, which is enough to keep the total gain higher than 40 dB at 4.2 K. The band center frequency deviates to a higher frequency, which can be explained by the decrease in the inductance of the LC tanks and transistor capacitance at 4.2 K. As shown in Fig. 16, the gain of the LNA at 4.2 K is 42 ± 3.3 dB in the frequency band from 4.6 to 8.2 GHz, while the $S_{11} < -10$ dB frequency band is $4.82 \sim 6.1$ GHz with optimum $S_{11} = -26$ dB at 5.2 GHz, which deteriorates to -5.8 dB at 7.9 GHz. The worse input match performance at 4.2 K is most likely due to the variation of C_{gs1} and C_{gd1} . Based on the analysis in Section IV, the impedance matching performance is sensitive

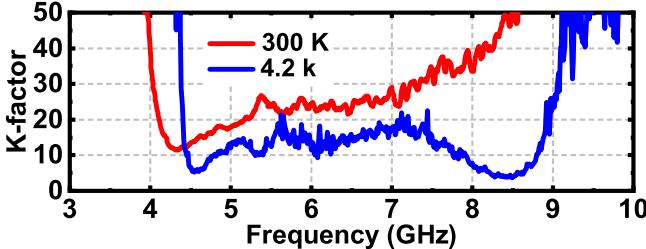
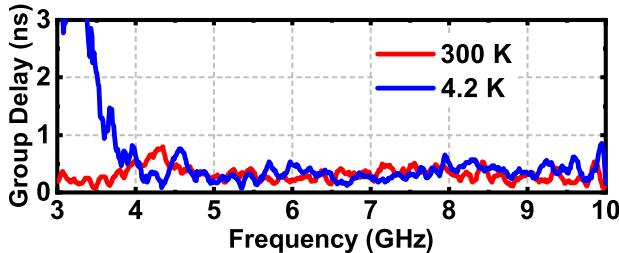
Fig. 17. Measurement results of K -factor.

Fig. 18. Measurement results of group delay.

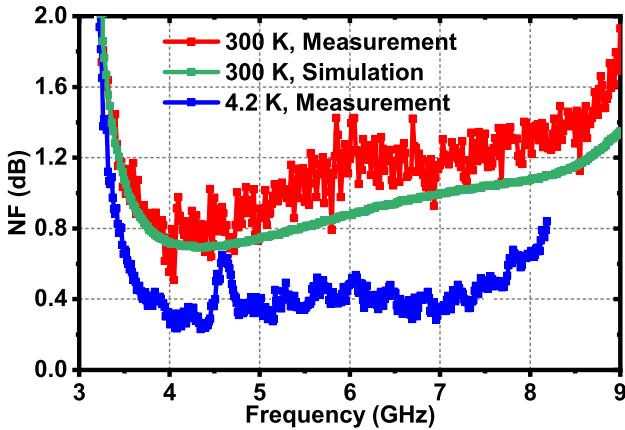
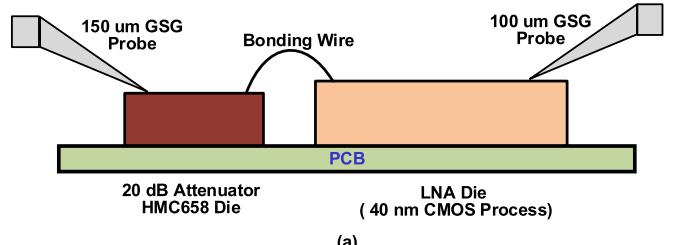


Fig. 19. NF simulation and measurement results.

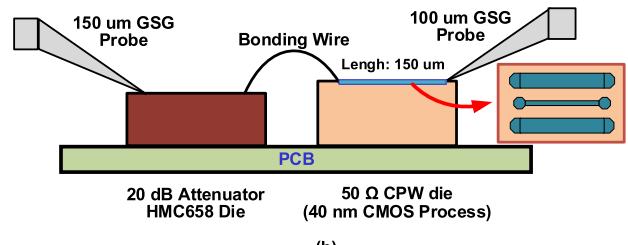
to C_{gd1} . The observation that the best impedance matching frequency point increases from 4.5 GHz at 300 K to 5.2 GHz at 4.2 K supports our speculation. The stability factor (K -factor) measured results show that $K > 4$ both at 300 and 4.2 K (Fig. 17), which indicates that the LNA is stable. Moreover, the measured group delays within the passband are found to be less than 0.6 and 0.81 ns at 300 and 4.2 K, respectively (Fig. 18).

The noise measurement results at room temperature are shown in Fig. 19. It can be seen that the NF is as low as 0.75 dB at 4.5 GHz, and it increases with frequency. The NF ranges from 0.75 to 1.3 dB over the entire operating frequency. The simulation and measurement results show good agreement.

The NF at 4.2 K was measured based on the setups illustrated in Fig. 20. A 20-dB attenuator was used for noise isolation between the cryogenic and room temperature environments. The attenuator bare die (HMC658) was mounted close to the LNA die on the same PCB, and wire-bonding was



(a)



(b)

Fig. 20. Cryogenic NF measurement setups for (a) NF measurements and (b) de-embedding.

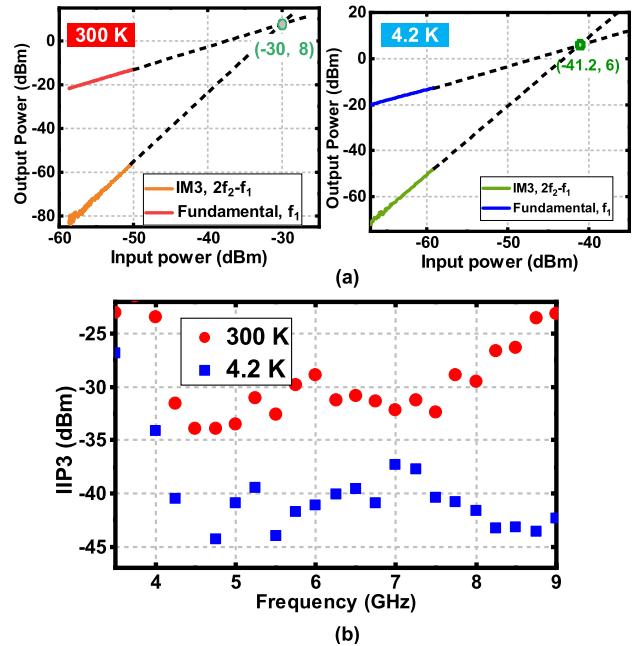


Fig. 21. IIP3 measurement results. (a) 6-GHz power sweep. (b) IIP3 versus frequency sweep.

employed for connection between the dies [Fig. 20(a)]. This setup can ensure that the attenuator's temperature is 4.2 K, the same as the LNA.

In order to de-embed the noise of attenuator and bonding wire, the same model attenuator was wire-bonded to a die with 50- Ω CPW line [Fig. 20(b)], implemented in the same process as the die. Once the noise equivalent temperature of the attenuator and LNA ($T_{eq,LNA+ATT}$), and the insertion loss of the de-embedding structure (L_{ATT}) are measured, the noise equivalent temperature of the LNA ($T_{eq,LNA}$) at 4.2 K can be calculated by [51]

$$T_{eq,LNA} = \frac{T_{eq,LNA+ATT} - 4.2 K \cdot (L_{ATT} - 1)}{L_{ATT}}. \quad (19)$$

TABLE III
PERFORMANCE SUMMARY AND COMPARISON WITH THE STATE-OF-THE-ART LNAs

	Topology	Frequency (GHz)	V _{DD} (V)	Temperature	Power (mW)	Gain (dB)	NF (dB)	S ₁₁ (dB)	Area (mm ²)	Off-chip components	Technology
JSSC'14 [44]	Complementary CS + capacitive loading	0.9~1.2	2.5	300 K	50	10.5	1.3~1.5	-14~5	0.68	Yes ¹	0.18 μ m SOI CMOS
TCAS-I'19 [45]	CG-CS + current bleeding + NC	0.05~1	2.2	300 K	19.8	27±3	2.3~3.3	-22~15	0.078 ²	No	65 nm CMOS
JSSC'18 [13]	Complementary CS + mutual NC with XFMR	2.3~2.5	0.7	300 K	0.48	16~17.4	2.8~3.5	-25~14	0.42 ²	No	60 nm CMOS
SSCL'20 [46]	CS with ID	4~8	1.6	300 K	24.8	13.5~15	2~2.5	-15~12	0.4	Yes ¹	0.13 μ m CMOS
JSSC'20 [47]	22~32	1.05	300 K	17.3	18.7~21.5	1.7~2.2	-14.5~10	0.168	No	22 nm FD-SOI	
TCAS-I'19 [48]	Common emitter with ID + output matching	8~20	1.6/3.1	300 K	48.5	10~17.5	1.25~3	-30~10	0.69	No	0.13 μ m SiGe HBT
JSSC'07 [17]	CS cascode with ID + capacitive load	0.7~1.4	1	300 K	43	17~22	0.29~0.55	-15~3	0.83	Yes	90 nm CMOS
IMS'16 [49]	Common emitter with ID + resistive feedback	18~25	1.2	300 K	0.9	17~23	3.2~4.2	---	0.25	Yes ¹	0.13 μ m SiGe HBT
TMTT'18 [7]	3 stage cascade	0.3~14	2	300 K	100	39.8~41.6	0.7~1.4	-22~1	1.5	Yes	100 nm InP HEMT
MWCL'09 [8]	Cascode with resistive feedback	0.1~5	---	300 K	70	25.2~31	0.9~1.29	-10~6	---	Yes ¹	0.13 μ m SiGe HBT
77 K	15 K	20	28.5~33.5	0.04~0.07	-22~10						
TMTT'19 [10]	Cascode CS	2.0~2.12	1.8	300 K		15	1.52~1.9	-18~12	---	Yes	0.18 μ m CMOS
TMTT'12 [50]	Resistive feedback	0.5~4	1.4	4.2 K	15	18	0.5	-40~15			
JSSC'17 [9]	NC	0.1~0.5	---	300 K	80	35~40	0.8~1.25	-9~5	0.249 ²	No	0.16 μ m CMOS
This Work	Cascode + R-C loading + XFMR load	4.1~7.9	1.4	300 K	51.1 ³	35.5~36.5	0.75~1.3	-22~12	0.72	No	40 nm CMOS
		4.6~8.2		4.2 K	39 ⁴	39.2~44.8	0.23~0.65	-26~5.8			

1. Bonding wire, 2. Circuit core without pads, 3. Includes output buffer's power consumption 2.8 mW, 4. Includes output buffer's power consumption ~1.92 mW.

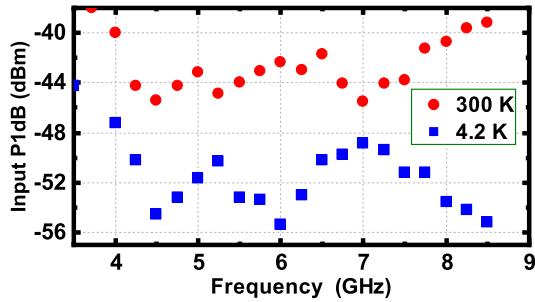


Fig. 22. Input P1dB measurement results against frequency.

The cryogenic measurement results for the NF are shown in Fig. 19. It can be seen that the NF is smaller than 0.5 dB from 4 to 7.8 GHz, and the best NF is 0.23 dB at ~5.2 GHz, i.e., 15-K equivalent noise temperature.

The weak signal linearity performance (IIP3) was tested using two-tone signals with 1-MHz frequency spacing. The IIP3 at the central frequency of 6 GHz is -30 dBm at room temperature and -41.2 dBm at 4.2 K [Fig. 21(a)]. The

cryo-IIP3s are greater than -45 dBm in the operating frequency band [Fig. 21(b)]. The lower IIP3 at cryogenic temperature is due to the higher gain and lower bias current of the last stage. The large signal non-linearity performance (input 1-dB compression point, i.e., P1dB) measurement results of the LNA at 300 and 4.2 K versus frequency are shown in Fig. 22. It can be seen that the P1dB is always higher than -56 dBm, which is much higher than the reflected signal power from the quantum devices.

Table III compares the performance of the presented cryogenic CMOS LNA with published LNAs. Our LNA exhibits excellent performance in terms of bandwidth and gain. Without using high-quality off-chip impedance matching components, the sub-1-dB NF performance is comparable with LNAs designed with InP HEMT and SiGe devices both at room and cryogenic temperatures.

VI. CONCLUSION

A CMOS cryogenic LNA was proposed for the gate-dispersive readout of semiconductor spin qubits. The detection of the quantum information signal was

modeled as demodulation of phase-modulated PSK signals, from which the LNA specifications were first derived. The proposed LNA is based on a CS cascode stage with inductive source degeneration, and a resistive and capacitive impedance loading technique is applied. The capacitive load is created by an on-chip L - C parallel tank and the resistive load is provided by the transconductance of the cascode transistor implicitly. An inductor is added to the gate of the cascode transistor to suppress its noise contribution and a transformer-based resonator is employed as the load of the first stage to extend the LNA bandwidth. Design considerations for the cryogenic temperature operation of the LNA are proposed and analyzed. The LNA achieves an NF between 0.75 and 1.3 dB with a bandwidth from 4.1 to 8.2 GHz at room temperature, and an NF between 0.23 and 0.65 dB with a bandwidth from 4.6 to 8.2 GHz at 4.2 K. The cryo-CMOS LNA meets the performance requirements for the gate dispersive readout of spin qubits, which marks a major step toward a fully integrated qubit readout platform.

ACKNOWLEDGMENT

The authors would like to thank Dr. Claudio Bruschini for the project management and Mr. Atef Akhnoukh for tape-out support.

REFERENCES

- [1] F. Arute *et al.*, “Quantum supremacy using a programmable superconducting processor,” *Nature*, vol. 574, no. 7779, pp. 505–510, 2019.
- [2] E. Charbon *et al.*, “Cryo-CMOS for quantum computing,” in *IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2016, pp. 13.5.1–13.5.4.
- [3] Y. Chen *et al.*, “Multiplexed dispersive readout of superconducting phase qubits,” *Appl. Phys. Lett.*, vol. 101, no. 18, pp. 26013–26014, 2012.
- [4] N. Ares *et al.*, “Sensitive radio-frequency measurements of a quantum dot by tuning to perfect impedance matching,” *Phys. Rev. A, Gen. Phys. Appl.*, vol. 5, no. 3, pp. 40111–40116, Mar. 2016.
- [5] 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.
- [6] C. C. Yang, B. Nelson, W. Jones, and B. Allen, “A cryogenically-cooled wide-band HEMT MMIC low-noise amplifier,” *IEEE Microw. Guided Wave Lett.*, vol. 2, no. 2, pp. 58–60, Feb. 1992.
- [7] E. Cha *et al.*, “0.3–14 and 16–28 GHz wide-bandwidth cryogenic MMIC low-noise amplifiers,” *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 11, pp. 4860–4869, Nov. 2018.
- [8] J. C. Bardin and S. Weinreb, “A 0.15 GHz cryogenic SiGe MMIC LNA,” *IEEE Microw. Wireless Compon. Lett.*, vol. 19, no. 6, pp. 407–409, Jun. 2009.
- [9] 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.
- [10] A. Caglar and M. B. Yelten, “Design of cryogenic LNAs for high linearity in space applications,” *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 66, no. 12, pp. 4619–4627, Dec. 2019.
- [11] M. Mehrpoo, F. Sebastian, E. Charbon, and M. Babaie, “A cryogenic CMOS parametric amplifier,” *IEEE Solid-State Circuits Lett.*, vol. 3, pp. 5–8, Dec. 2020.
- [12] A. Liscidini, M. Brandolini, D. Sanzogni, and R. Castello, “A $0.13\ \mu\text{m}$ CMOS front-end, for DCS1800/UMTS/ 802.11b-g with multiband positive feedback low-noise amplifier,” *IEEE J. Solid-State Circuits*, vol. 41, no. 4, pp. 981–988, Apr. 2006.
- [13] M. Rahman and R. Harjani, “A 2.4-GHz, sub-1-V, 2.8-dB NF, 475-W dual-path noise and nonlinearity cancelling LNA for ultra-low-power radios,” *IEEE J. Solid-State Circuits*, vol. 53, no. 5, pp. 1423–1430, May 2018.
- [14] A. Ismail and A. A. Abidi, “A 3–10-GHz low-noise amplifier with wideband LC-ladder matching network,” *IEEE J. Solid-State Circuits*, vol. 39, no. 12, pp. 2269–2277, Dec. 2004.
- [15] P. A. Hart, J. P. G. van Dijk, M. Babaie, E. Charbon, A. Vladimirescu, and F. Sebastian, “Characterization and model validation of mismatch in nanometer CMOS at cryogenic temperatures,” in *Proc. 48th Eur. Solid-State Device Res. Conf. (ESSDERC)*, Sep. 2018, pp. 246–249.
- [16] R. Hu, “Wide-band matched LNA design using transistor’s intrinsic gate-drain capacitor,” *IEEE Trans. Microw. Theory Techn.*, vol. 54, no. 3, pp. 1277–1286, Mar. 2006.
- [17] L. Belostotski and J. W. Haslett, “Sub-0.2 dB noise figure wideband room-temperature CMOS LNA with non- $50\ \Omega$ signal-source impedance,” *IEEE J. Solid-State Circuits*, vol. 42, no. 11, pp. 2492–2502, Nov. 2007.
- [18] D. J. Reilly, C. M. Marcus, M. P. Hanson, and A. C. Gossard, “Fast single-charge sensing with a RF quantum point contact,” *Appl. Phys. Lett.*, vol. 91, no. 16, pp. 21011–21013, Oct. 2007.
- [19] I. Ahmed *et al.*, “Radio-frequency capacitive gate-based sensing,” *Phys. Rev. A, Gen. Phys. Appl.*, vol. 10, no. 1, pp. 40181–40189, Jul. 2018.
- [20] S. Schaal *et al.*, “A CMOS dynamic random access architecture for radio-frequency readout of quantum devices,” *Nature Electron.*, vol. 2, no. 6, pp. 236–242, Jun. 2019.
- [21] Y. Jang, D. Yoon, and S.-K. Lee, “Generalized BER expression of MPSK in the presence of phase error,” *IEEE Commun. Lett.*, vol. 17, no. 12, pp. 2213–2216, Dec. 2013.
- [22] T. Müller, “Radio-frequency quantum point contact charge detection,” Ph.D. dissertation, Eidgenössische Technische Hochschule Zürich, Zürich, Switzerland, 2012.
- [23] B. Razavi and R. Behzad, *RF Microelectronics*, vol. 2. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.
- [24] J. M. Hornibrook *et al.*, “Frequency multiplexing for readout of spin qubits,” *Appl. Phys. Lett.*, vol. 104, no. 10, Mar. 2014, Art. no. 103108.
- [25] A. Crippa *et al.*, “Gate-reflectometry dispersive readout and coherent control of a spin qubit in silicon,” *Nature Commun.*, vol. 10, no. 1, pp. 1–6, Dec. 2019.
- [26] A. Ruffino, T.-Y. Yang, J. Michniewicz, Y. Peng, E. Charbon, and M. F. Gonzalez-Zalba, “Integrated multiplexed microwave readout of silicon quantum dots in a cryogenic CMOS chip,” 2021, *arXiv:2101.08295*. [Online]. Available: <http://arxiv.org/abs/2101.08295>
- [27] M. Babaie and R. B. Staszewski, “A class-F CMOS oscillator,” *IEEE J. Solid-State Circuits*, vol. 48, no. 12, pp. 3120–3133, Dec. 2013.
- [28] D. K. Shaeffer and T. H. Lee, “A 1.5-V, 1.5-GHz CMOS low noise amplifier,” *IEEE J. Solid-State Circuits*, vol. 32, no. 5, pp. 745–759, May 1997.
- [29] U. L. Rohde, A. K. Poddar, and G. Böck, *The Design of Modern Microwave Oscillators for Wireless Applications*. Hoboken, NJ, USA: Wiley, 2005.
- [30] C.-T. Fu, C.-N. Kuo, and S. S. Taylor, “Low-noise amplifier design with dual reactive feedback for broadband simultaneous noise and impedance matching,” *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 4, pp. 795–806, Apr. 2010.
- [31] B. Patra, M. Mehrpoo, A. Ruffino, F. Sebastian, E. Charbon, and M. Babaie, “Characterization and analysis of on-chip microwave passive components at cryogenic temperatures,” *IEEE J. Electron Devices Soc.*, vol. 8, pp. 448–456, 2020.
- [32] H. Samavati, H. R. Rategh, and T. H. Lee, “A 5-GHz CMOS wireless LAN receiver front end,” *IEEE J. Solid-State Circuits*, vol. 35, no. 5, pp. 765–772, May 2000.
- [33] R. M. Incandela, L. Song, H. Homulle, E. Charbon, A. Vladimirescu, and F. Sebastian, “Characterization and compact modeling of nanometer CMOS transistors at deep-cryogenic temperatures,” *IEEE J. Electron Devices Soc.*, vol. 6, pp. 996–1006, Mar. 2018.
- [34] A. Beckers, F. Jazaeri, A. Ruffino, C. Bruschini, A. Baschirotto, and C. Enz, “Cryogenic characterization of 28 nm bulk CMOS technology for quantum computing,” in *Proc. 47th Eur. Solid-State Device Res. Conf. (ESSDERC)*, Sep. 2017, pp. 62–65.
- [35] B. Razavi, *Design of Analog CMOS Integrated Circuits*, 2nd ed. New York, NY, USA: McGraw-Hill, 2005.
- [36] E. A. Gutierrez-D, O. Gonzalez-C, and R. S. Murphy-A, “Electron transport through accumulation layers and its effect on the series resistance of MOS transistors,” in *Proc. 2nd IEEE Int. Caracas Conf. Devices, Circuits Syst. ICCDCS. 70th Anniversary MOSFET 50th BJT*, Mar. 1998, pp. 51–54.
- [37] S.-H. Hong, G.-B. Choi, R.-H. Baek, H.-S. Kang, S.-W. Jung, and Y.-H. Jeong, “Low-temperature performance of nanoscale MOSFET for deep-space RF applications,” *IEEE Electron Device Lett.*, vol. 29, no. 7, pp. 775–777, Jul. 2008.
- [38] A. H. Coskun and J. C. Bardin, “Cryogenic small-signal and noise performance of 32 nm SOI CMOS,” in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2014, pp. 13–16.

- [39] C. Spathis, A. Birbas, and K. Georgakopoulou, "Semi-classical noise investigation for sub-40 nm metal-oxide-semiconductor field-effect transistors," *AIP Adv.*, vol. 5, no. 8, Aug. 2015, Art. no. 087114.
- [40] V. M. Mahajan *et al.*, "A physical understanding of RF noise in bulk nMOSFETs with channel lengths in the nanometer regime," *IEEE Trans. Electron Devices*, vol. 59, no. 1, pp. 197–205, Jan. 2012.
- [41] P. R. Gray, *Analysis and Design of Analog Integrated Circuits*. Hoboken, NJ, USA: Wiley, 2009.
- [42] G. D. J. Smit, A. J. Scholten, R. M. T. Pijper, L. F. Tiemeijer, R. van der Toorn, and D. B. M. Klaassen, "RF-noise modeling in advanced CMOS technologies," *IEEE Trans. Electron Devices*, vol. 61, no. 2, pp. 245–254, Feb. 2014.
- [43] D. Zhao and P. Reynaert, "A 60-GHz dual-mode class AB power amplifier in 40-nm CMOS," *IEEE J. Solid-State Circuits*, vol. 48, no. 10, pp. 2323–2337, Oct. 2013.
- [44] B.-K. Kim, D. Im, J. Choi, and K. Lee, "A highly linear 1 GHz 1.3 dB NF CMOS low-noise amplifier with complementary transconductance linearization," *IEEE J. Solid-State Circuits*, vol. 49, no. 6, pp. 1286–1302, Jun. 2014.
- [45] S. Kim and K. Kwon, "A 50-MHz–1-GHz 2.3-dB NF noise-cancelling balun-LNA employing a modified current-bleeding technique and balanced loads," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 66, no. 2, pp. 546–554, Feb. 2019.
- [46] E. Zailer, L. Belostotski, and R. Plume, "Wideband LNA noise matching," *IEEE Solid-State Circuits Lett.*, vol. 3, pp. 62–65, Apr. 2020.
- [47] B. Cui, J. R. Long, and D. L. Harame, "A 1.7-dB minimum NF, 22–32 GHz low-noise feedback amplifier with multistage noise matching in 22-nm SOI-CMOS," in *Proc. IEEE Radio Freq. Integr. Circuits Symp. (RFIC)*, Jun. 2019, pp. 211–214.
- [48] C. Çaiskan, I. Kalyoncu, M. Yazici, and Y. Gurbuz, "Sub-1-dB and wideband SiGe BiCMOS low-noise amplifiers for X-band applications," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 66, no. 4, pp. 1419–1430, Apr. 2019.
- [49] W.-T. Wong, P. Ravindran, S.-W. Chang, and J. C. Bardin, "A SiGe Ka-band cryogenic low-noise amplifier," in *IEEE MTT-S Int. Microw. Symp. Dig.*, May 2016, pp. 1–3.
- [50] D. Russell and S. Weinreb, "Low-power very low-noise cryogenic SiGe IF amplifiers for terahertz mixer receivers," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 6, pp. 1641–1648, Jun. 2012.
- [51] R. Hu and S. Weinreb, "A novel wide-band noise-parameter measurement method and its cryogenic application," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 5, pp. 1498–1507, May 2004.



Yatao Peng (Member, IEEE) received the B.Eng. degree in information engineering from Tianjin Normal University, Tianjin, China, in 2010, and the Ph.D. degree in electronics engineering from the University of Chinese Academy of Sciences, Beijing, China, in 2015, with a focus on multiband/wideband RF front-end circuits for wireless communication.

From 2016 to 2017, he was a Post-Doctoral Researcher with the University of Macau, Macau, China, where he was involved in CMOS millimeter-wave integrated circuit design. From 2017 to 2018, he was with the National University of Singapore, Singapore, where he served as a Research Fellow to develop CMOS phase-shifter modules for hybrid integrated flexible electronic systems. He is currently a Scientist with the Advanced Quantum Architecture Laboratory, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. His research interest includes CMOS/BiCMOS cryogenic circuits for quantum applications.



Andrea Ruffino (Member, IEEE) received the B.Sc. degree (*cum laude*) in engineering physics from the Politecnico di Torino, Turin, Italy, in 2013, the triple joint M.Sc. degree (*cum laude*) in micro and nanotechnologies for integrated systems from the Politecnico di Torino, the Institut National Polytechnique de Grenoble (INPG), Grenoble, France, and the École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, in 2015, and the Ph.D. degree in microsystems and microelectronics from EPFL in 2021.

From 2015 to 2016, he was with Hypres, Inc., Elmsford, NY, USA, working on the design of superconducting readout circuits in rapid single flux quantum (RSFQ) technology for superconducting nanowire single-photon detectors. In 2016, he joined EPFL as a Research Assistant and from 2017 to 2018, he was a Visiting Research Assistant at the Delft University of Technology, Delft, The Netherlands. Since 2016, he has been with EPFL, where he is working on cryogenic CMOS electronics for qubit readout and control, focusing on single-chip cryo-CMOS transceivers for scalable silicon quantum computers. His main research interests include analog and RF integrated circuit design, cryogenic CMOS electronics for quantum computing applications, and superconducting electronics and sensors.

Dr. Ruffino was also among the Best Student Paper Finalists at the IEEE Radio-Frequency Integrated Circuits (RFIC) Symposium 2019 and he was a recipient of the IEEE Solid-State Circuits Society (SSCS) Predoctoral Achievement Award for 2020–2021.



Edoardo Charbon (Fellow, IEEE) received the Diploma degree from ETH Zürich, Zürich, Switzerland, in 1988, the M.S. degree from the University of California at San Diego, La Jolla, CA, USA, in 1991, and the Ph.D. degree from the University of California at Berkeley, Berkeley, CA, USA, in 1995, all in electrical engineering and EECS.

He has consulted with numerous organizations, including Bosch, X-Fabs, Texas Instruments, Maxim, Sony, Agilent, and the Carlyle Group. He was with Cadence Design Systems, San Jose, CA, USA, from 1995 to 2000, where he was an Architect of the company's initiative on information hiding for intellectual property protection. In 2000, he joined Canesta Inc., as the Chief Architect, where he led the development of wireless 3-D CMOS image sensors. Since 2002, he has been a Faculty Member with the École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland, where he has been a Full Professor since 2015. From 2008 to 2016, he was with the Delft University of Technology, Delft, The Netherlands, as the Chair of VLSI design. He has been the driving force behind the creation of deep-submicrometer CMOS SPAD technology, which is mass-produced since 2015 and is present in telemeters, proximity sensors, and medical diagnostics tools. He has authored or coauthored over 250 articles and two books. He holds 20 patents. His interests span from 3-D vision, FLIM, FCS, and NIROT to super-resolution microscopy, time-resolved Raman spectroscopy, and cryo-CMOS circuits and systems for quantum computing.

Dr. Charbon is a Distinguished Visiting Scholar of the W. M. Keck Institute for Space at Caltech and a fellow of the Kavli Institute of Nanoscience Delft.