Optimizing the Nonlinearity and Dissipation of a SNAIL Parametric Amplifier for Dynamic Range

N. E. Frattini, V. V. Sivak, A. Lingenfelter, S. Shankar, and M. H. Devoret

Phys. Rev. Applied **10**, 054020 — Published 8 November 2018

DOI: 10.1103/PhysRevApplied.10.054020

Optimizing the nonlinearity and dissipation of a SNAIL Parametric Amplifier for dynamic range

N. E. Frattini,* V. V. Sivak,† A. Lingenfelter, S. Shankar, and M. H. Devoret‡
Department of Applied Physics, Yale University, New Haven, CT 06520, USA
(Dated: September 20, 2018)

We present a quantum-limited Josephson-junction-based 3-wave-mixing parametric amplifier, the SNAIL Parametric Amplifier (SPA), which uses an array of SNAILs (Superconducting Nonlinear Asymmetric Inductive eLements) as the source of tunable nonlinearity. We show how to engineer the nonlinearity over multiple orders of magnitude by varying the physical design of the device. As a function of design parameters, we systematically explore two important amplifier nonidealities that limit dynamic range: the phenomena of gain compression and intermodulation distortion, whose minimization are crucial for high-fidelity multi-qubit readout. Through a comparison with first-principles theory across multiple devices, we demonstrate how to optimize both the nonlinearity and the input-output port coupling of these SNAIL-based parametric amplifiers to achieve higher saturation power, without sacrificing any other desirable characteristics. The method elaborated in our work can be extended to improve all forms of parametrically induced mixing that can be employed for quantum information applications.

I. INTRODUCTION

Quantum-limited Josephson parametric amplifiers [1, 2] are a key component in many precision microwave measurement setups such as for the readout of superconducting qubits [3–7], the high sensitivity detection of electron spin resonance [8, 9], and the search for axions [10]. As the first component of a microwave amplification chain, the main desired specifications for a Josephson amplifier are: i) low added noise: the noise added by the amplifier should be no larger than the minimum imposed by quantum mechanics, ii) high gain: the amplifier power gain should be large enough to overwhelm the noise temperature of the following amplification chain (in practice, at least 20 dB), iii) large bandwidth: the amplifier gain should be constant over a bandwidth that is large enough for the desired application, iv) large dynamic range: the output signal power should be linearly proportional to the input signal power over a wide enough power range, v) unidirectionality: the amplifier should, ideally, amplify only signals incident from the system being probed and isolate the signal source from spurious noise that propagates back from subsequent devices in the amplification chain, vi) ease of operation: the energy necessary for amplification should be delivered to the amplifier in a simple and robust manner without requiring precise tuning, vii) robustness of construction: the amplifier circuit should not require too delicate tolerances.

Among these characteristics, dynamic range is a particularly important requirement for scaling up superconducting qubit setups to larger size systems [11]. The dynamic range characterizes the input power range over which the amplifier behaves as a linear device for a single- or multi-tone input. For quantum-limited amplifiers, the lower limit on the dynamic range is set fundamentally by quantum mechanics, so improving dynamic range corresponds to increasing the upper limit. The upper limit is controlled by two distinct but closely related nonidealities in the large-signal amplifier response. The first nonideality is the phenomenon of amplifier saturation, also called gain compression. This limits the maximum output power that can be produced by the device for an arbitrary input signal. The second nonideality, previously unexplored for quantum-limited amplifiers, is the phenomenon of intermodulation distortion for multi-tone inputs, where the amplifier produces spurious tones on its output in addition to the desired amplified copies of the input tones. Together, these two nonidealities limit the signal powers that can be processed by the amplifier and thus are a problem for faster or higher-power qubit readout as well as for the readout of multiple qubits [6].

Is it possible to improve the amplifier dynamic range without sacrificing other desirable characteristics? Here, we answer affirmatively by demonstrating systematic improvement of the dynamic range of a 3-wave-mixing degenerate parametric amplifier, named the SNAIL Parametric Amplifier (SPA). The SPA is based on an array of Superconducting Nonlinear Asymmetric Inductive eLements (SNAILs) [12, 13], which provides the flexibility needed to optimize the 3-wave-mixing amplification process, while simultaneously minimizing the 4-wave-mixing Kerr nonlinearity suspected to cause amplifier saturation [14, 15]. With this flexibility, we have engineered an SPA that achieves a 1 dB compression power ($P_{-1\mathrm{dB}} \in [-102, -112]$ dBm for 20 dB gain) on par with the best quantum-limited resonant parametric amplifiers [16–18], but over the entire tunable bandwidth of 1 GHz without sacrificing any other desirable characteristics including quantum-limited noise performance.

Our demonstration of dynamic range improvement is crucially accompanied by first-principles theory that elucidates the link between the physical realization of the

* Equal contribution; nicholas.frattini@yale.edu
† Equal contribution; vladimir.sivak@yale.edu
‡ michel.devoret@yale.edu

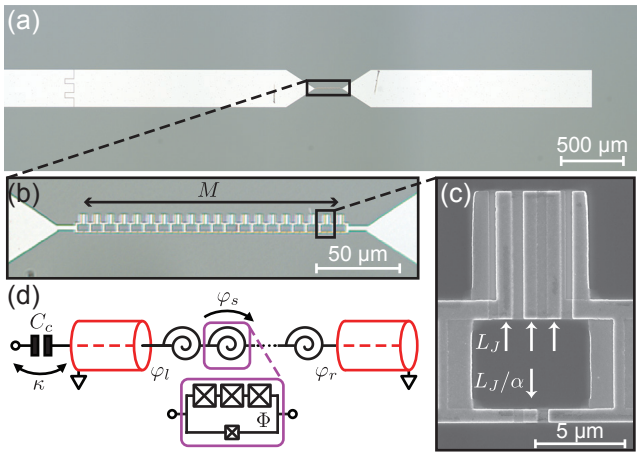

FIG. 1. (a) Optical microscope image and (d) corresponding circuit model of a SNAIL Parametric Amplifier (SPA). An array of M SNAILs is inserted at the center of a $\lambda/2$ section of microstrip transmission line, colored red in (d). (b) Image of an array of $M = 20$ SNAILs. (c) Electron micrograph of a single SNAIL with 3 large Josephson junctions (inductance L_J) in a loop with one smaller junction (inductance L_J/α). Arrows indicate the junctions and the inset of (d) gives the SNAIL circuit schematic. In (d), φ_s denotes the phase drop across each SNAIL. The node phase φ_l (φ_r) denotes the location where the left (right) side of the array of SNAILs connects to the linear embedding structure. The dissipation rate κ is set by capacitive coupling with capacitance C_c to the transmission line.

amplifier and the nonidealities of its response to large input signals. This link is accomplished in two steps. First, we show how to map the physical layout of the SPA to the phenomenological parameters that enter the input-output description of the device. These parameters consist of the 3-wave- and 4-wave-mixing nonlinear components of the SPA Hamiltonian, as well as the damping induced through coupling to a transmission line via an input-output port. Second, we describe and validate experimentally how these phenomenological parameters directly determine the nonidealities in the amplifier's response to large input signals. Such first-principles theoretical description opens the door to further improvements in the amplifier dynamic range as well as the optimization of any other form of parametrically induced mixing for quantum information processing.

The article is organized as follows. In Sec. II, we introduce the physical realization of the SPA. Sec. III briefly describes the relevant parameters of the SPA model and their influence on the small-signal gain of the amplifier. Sec. IV validates our mapping between the physical layout and the SPA parameters, with the theoretical details given in Appendix A. In Sec. V, we explore the mechanisms responsible for amplifier saturation, and characterize the intermodulation distortion of the SPA in Sec. VI, with theoretical details given in Appendix B.

Device	L_J (pH)	M	α	C_c (pF)	$\omega_0/2\pi$ (GHz)
A	60	1	0.29	0.048	8.4
B	67	10	0.29	0.039	11.4
C	47	20	0.09	0.068	17.9
D	44	20	0.09	0.075	23.5
E	34	20	0.09	0.088	23.4

TABLE I. Constitutive parameters of 5 devices measured in the experiment: Josephson inductance of largest junction (L_J), number of SNAILs (M), junction inductance ratio (α), coupling capacitance to the 50 Ω transmission line (C_c), and frequency of the $\lambda/2$ microstrip embedding structure when the array of SNAILs is replaced by a short (ω_0).

II. SPA PHYSICAL REALIZATION

Akin to the Josephson Parametric Amplifier (JPA) [19, 20], the SPA is realized by placing M SNAILs at the center of a $\lambda/2$ section of microstrip transmission line (Fig. 1a). Fig. 1b depicts an array of $M = 20$ SNAILs, where each SNAIL consists of an array of 3 large Josephson junctions (Josephson inductance L_J) in a loop with one smaller junction (inductance L_J/α). In practice, we chose the smallest L_J that was still larger than the parasitic geometric inductance of the 24 μm perimeter SNAIL loop. As shown in the electron micrograph (Fig. 1c), the Josephson junctions are fabricated using a Dolan bridge process for aluminum (Al) on silicon (Si).

The microstrip transmission line sections are formed by a 2 μm thick silver (Ag) layer deposited on the back of the 300 μm thick Si wafer to act as a ground plane, and by center traces of Al whose length l_{MS} and width w_{MS} adjust the frequency ω_0 and the characteristic impedance Z_c. For all devices in this work, we held the microstrip width constant at $w_{\mathrm{MS}} = 300\,\mu$m to set $Z_c = 45\,\Omega$, and adjusted l_{MS} (in conjunction with M, α and L_J) to set the operating frequencies of the devices (see Section IV A). The coupling to the 50 Ω transmission line κ is set by a gap capacitor (capacitance C_c) at one end of the SPA resonator (Fig. 1a). Later devices (E in Table I) also have a second weakly capacitively coupled port on the opposite end of the resonator for the delivery of the pump (not shown in Fig. 1a). By design, κ is much larger than the internal dissipation rate and the coupling to the pump port.

The experimental characterization was performed in a helium dilution refrigerator (temperature ≈ 20 mK) with a standard microwave reflection measurement setup. While cold, a magnet coil mounted beneath the sample applies a magnetic flux Φ to each SNAIL, which we assume to be uniform across the array. All measurements were performed with a PNA-X network analyzer [21], which contains two microwave sources and the capability to quickly perform intermodulation distortion measurements (see Section VI). The strong pump tone needed for amplification was either combined with the signal tone at room temperature or applied on a separate pump line.

III. SPA MODEL

The device is modeled with the circuit schematic of Fig 1d. Following Ref. 12, we treat the SNAIL as a nonlinear inductor that provides an asymmetric potential energy $U_{\mathrm{SNAIL}}(\varphi_s)$ and corresponding current-phase relation $I_s(\varphi_s) = \frac{2\pi}{\Phi_0}\frac{dU_{\mathrm{SNAIL}}}{d\varphi_s}$ (where $\Phi_0 = h/2e$ is the superconducting magnetic flux quantum and φ_s is the phase drop across the small junction of the SNAIL). These functions are engineered via the junction inductance ratio α and the externally applied magnetic flux Φ. To include the linear embedding circuit, we enforce the constraint of current conservation at the left and right boundary nodes of the SNAIL array (phases denoted φ_l and φ_r in Fig. 1d), which are connected to the ends of the respective transmission lines. As shown in Appendix A, properly handling this nonlinear constraint equation is crucial for the prediction of higher order Hamiltonian terms, such as Kerr.

We next quantize the system and express the Hamiltonian of the lowest frequency mode of the SPA up to fourth order as

$$\boldsymbol{H}_{\mathrm{SPA}}/\hbar = \omega_a \boldsymbol{a}^\dagger \boldsymbol{a} + g_3 \left(\boldsymbol{a} + \boldsymbol{a}^\dagger\right)^3 + g_4 \left(\boldsymbol{a} + \boldsymbol{a}^\dagger\right)^4, \quad (1)$$

where $\boldsymbol{a}^\dagger$ ($\boldsymbol{a}$) is the harmonic oscillator creation (annihilation) operator, ω_a is the resonant frequency, and the third-order and fourth-order nonlinearities are denoted g_3 and g_4 respectively. These three Hamiltonian terms are all tuned in situ via the applied magnetic flux Φ through each SNAIL loop. Along with the coupling rate to the transmission line κ, the parameters of $\boldsymbol{H}_{\mathrm{SPA}}$ determine the behavior of the SPA as a degenerate parametric amplifier as we show next.

To operate the SPA as a 3-wave-mixing amplifier, we apply a strong microwave pump tone at $\omega_p \approx 2\omega_a$, with mean intra-cavity amplitude α_p. As shown in Appendix B, input-output theory [22] gives the phase-preserving power gain G for a signal at frequency ω_s scattering in reflection off of an SPA as

$$G = 1 + \frac{4\kappa^2|g|^2}{(\Delta_p^2 - \omega^2 + \frac{\kappa^2}{4} - 4|g|^2)^2 + (\kappa\omega)^2}, \quad (2)$$

where $\omega = \omega_s - \omega_p/2$ is the detuning of the input signal from $\omega_p/2$, $g = 2g_3\alpha_p$, and $\Delta_p = \Delta + \frac{32}{3}g_4|\alpha_p|^2$ with $\Delta = \omega_a - \omega_p/2$.

For this work, we always set the pump frequency so that $\Delta = 0$. Note that the maximum of gain G always occurs at $\omega_s = \omega_p/2$, similar to the flux-pumped JPA [23], making this amplifier particularly easy to tune up and operate. This property is in contrast with the tune-up procedure for Josephson Parametric Converters (JPCs) and even 4-wave-mixing JPAs, as outlined in Ref. 15 and Ref. 24 respectively.

As shown by Eq. 2, designing an amplifier operating at ω_s reduces to engineering ω_a, g_3, g_4, and κ. This task is accomplished by the appropriate choice of the physical knobs described in Section II. To illustrate control over these Hamiltonian parameters and provide intuition on this mapping, we compare the set of devices listed in Table I.

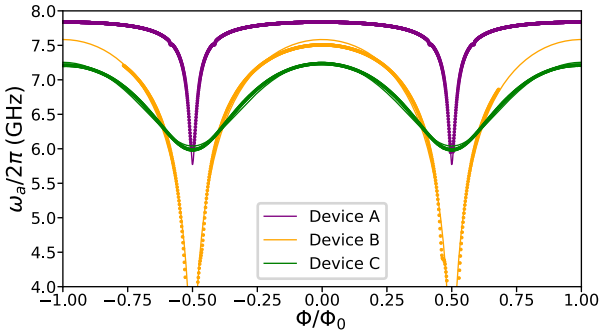

FIG. 2. Resonant frequency ω_a as a function of applied magnetic flux Φ for three devices. Thin solid lines are fits to a model based on the schematic in Fig. 1d.

IV. SPA HAMILTONIAN CHARACTERISTICS

A. Resonant Frequency Tunability

We first compare the linear response characteristics of these devices, specifically the resonant frequency ω_a as a function of applied magnetic flux Φ (Fig. 2). The tunability range of ω_a depends on two factors: (1) the flux-tunable SNAIL inductance $L_{\mathrm{SNAIL}}(\Phi)$, and (2) the participation of the SNAIL array in its embedding structure. Multiple physical knobs affect both of these factors; here for simplicity we focus on the influence of α and M.

The first factor, the flux dependence of $L_{\mathrm{SNAIL}}(\Phi)$, is strongly analogous to that of a DC-SQUID (Superconducting QUantum Interference Device) or an RF-SQUID [25]: the inductance is tunable between a minimum at $\Phi/\Phi_0 = 0$ and maximum at $\Phi/\Phi_0 = 0.5$. The range of this tunability is given by the asymmetry between the inductances on either arm of the superconducting loop which, in the SNAIL, is controlled by the junction inductance ratio α. $\alpha = 1/3$ (where 3 is the number of large junctions in a SNAIL) corresponds to perfect inductive symmetry, resulting in $L_{\mathrm{SNAIL}}(0.5\,\Phi_0) \to \infty$. $\alpha > 1/3$ causes the SNAIL potential to have multiple inequivalent minima and results in hysteretic behavior, which we wish to avoid in an amplifier. $\alpha < 1/3$ gives some asymmetry, where smaller α corresponds to a smaller inductive tunability range. However, as we will show in Section IV B, smaller α is advantageous for achieving the optimal flux profile of g_3 and g_4.

The second factor influencing the tunable range of ω_a is the fraction of the mode inductance coming from the SNAILs. For a given SNAIL design with an $L_{\mathrm{SNAIL}}(\Phi)$, this is controlled by the number of SNAILs M in series as well as the length l_{MS} and width w_{MS} of the surrounding microstrip embedding structure. In practice, M provides

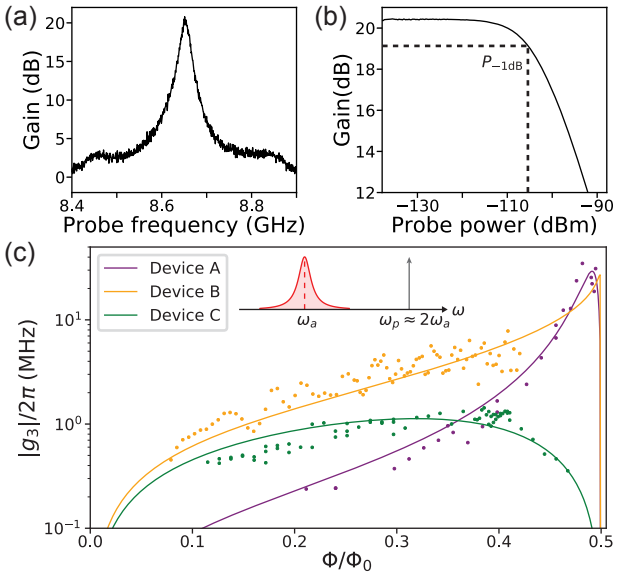

FIG. 3. (a) Reflection gain versus input signal frequency when the SPA is biased with a strong tone at $\omega_p \approx 2\omega_a$, frequency landscape inset of (c). (b) Gain as a function of input signal power shows amplifier saturation. Input power at which the gain reduces by 1 dB is denoted $P_{-1\mathrm{dB}}$. (c) Third-order nonlinearity g_3 versus applied magnetic flux Φ. Solid curves are first-principles theory for g_3.

more control over ω_a due to the practical difficulty in realizing microstrip embedding structures with impedances significantly different from 50 Ω. Thus, M and α are chosen first and then we adjusted l_{MS} (while keeping w_{MS} fixed) to hit our desired operating frequency range.

Focusing on the Φ dependence of ω_a in devices B and C (Fig. 2), we see the ability of α and M to engineer the frequency tunability. For $\alpha = 0.29$ as in Device B, the total inductances on either arm of the SNAIL are nearly equal so the SNAIL inductance changes drastically from $\Phi/\Phi_0 = 0$ to $\Phi/\Phi_0 = 0.5$. Conversely, the inductance of each $\alpha = 0.09$ SNAIL in device C changes only a little and the aggregation of these small changes for all 20 SNAILs gives the device its ≈ 1 GHz of tunability.

B. Nonlinear Characteristics

Having described the linear response of the SPA resonator, we next demonstrate its operation as a 3-wave-mixing degenerate parametric amplifier. We applied a strong microwave pump tone at $\omega_p = 2\omega_a$ and adjusted the pump power to achieve 20 dB of small-signal reflection gain (example shown Fig. 3a). One standard phenomenon that limits amplifier quality is the saturation of the gain with increasing input signal power. Shown in Fig. 3b, we measured the input-referred 1 dB compression point $P_{-1\mathrm{dB}}$ as the input signal power where the gain drops by 1 dB. To understand this phenomenon and improve the $P_{-1\mathrm{dB}}$, we perform a systematic study across multiple devices. Toward this goal, we begin by first measuring the nonlinearity of the SPA Hamiltonian (Eq. 1) as a function of applied magnetic flux Φ for all devices in Table I.

The dependence of third-order nonlinearity g_3 on Φ is shown in Fig. 3c for three representative devices. g_3 is extracted from Eq. 2 by tuning up a 20 dB gain point and using the measured values for ω_a, κ and a calibration on the applied pump power. Also shown is our first-principles theory calculation, which uses only the linear characteristics fit from Fig. 2 and room temperature measurements of the resistance of the SNAIL array. A global scale factor of ≈ 2 has been applied to the extracted g_3, which could arise from pump power miscalibration or the enhanced coupling of the pump to the SNAILs through higher frequency modes not considered in our simple model. Comparing devices A and C, we note the relatively constant g_3 for device C ($\alpha = 0.09$) except near $\Phi/\Phi_0 = 0$ and $\Phi/\Phi_0 = 0.5$ where symmetry forbids 3-wave-mixing terms. In contrast, device A ($\alpha = 0.29$) shows a two order of magnitude variation in g_3 over the same flux range. This comparison highlights the drastic difference in the flux profile of g_3, here mainly arising from the difference in the junction inductance ratio α.

The fourth-order nonlinearity g_4 is extracted from a Stark shift measurement. In this experiment, we applied a strong ≈ 500 MHz detuned drive that populates the resonator with $\bar{n}$ average steady-state photons and shifts its resonant frequency. Here, $\bar{n}$ is calibrated using fits of ω_a, κ and room temperature line attenuation. In Fig. 4a we plot the measured frequency shift $\Delta\omega_a$ of a typical SPA resonator as a function of $\bar{n}$ and applied magnetic flux Φ (color). The frequency shift changes from negative to positive over half of a flux quantum. The solid lines are fits to $\Delta\omega_a = K\bar{n} + K'\bar{n}^2$. From this fit, we extract the Stark shift per photon K, which is related to the Hamiltonian parameters g_3 and g_4 up to second order in perturbation theory by $K = 12(g_4 - 5g_3^2/\omega_a)$ (see Appendix A).

The dependence of K and thus g_4 on Φ is shown in Fig. 4b for three representative devices together with our first-principles theory calculation. The contrast between device A and device C again highlights the effect of α on the flux profile. Specifically, device A shows a three order of magnitude change in g_4, while device C's g_4 is relatively constant over most of the flux range. Additionally, both devices nominally support a region of suppressed Kerr. However, device A attains this region over a very narrow flux range, making the suppression practically useless, while device C shows a robust suppression regime by more than an order of magnitude from its $\Phi/\Phi_0 = 0$ value. This suppression could be useful in applications where the circuit designer wants some nonlinearity for mixing purposes, but would prefer to suppress spurious Kerr interactions.

While our previous comparison of devices A-C focused on the flux profile of g_3 and g_4, their overall magnitudes must also be engineered for optimizing amplifier nonide-

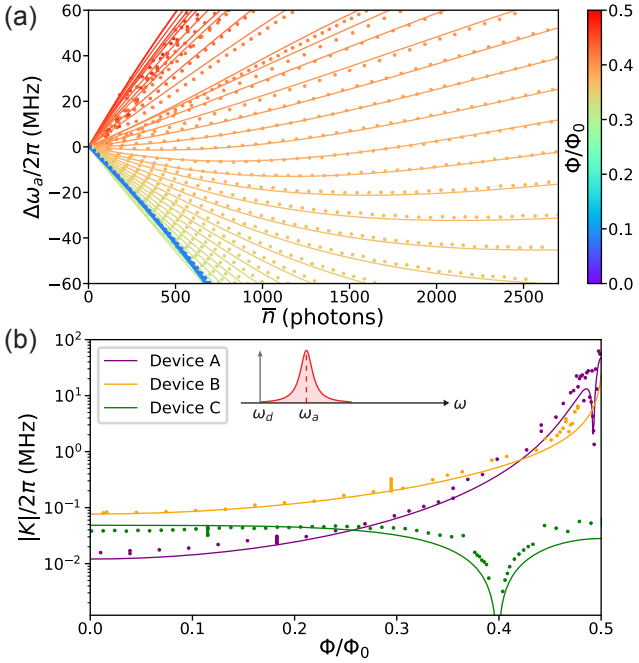

FIG. 4. (a) Frequency shift $\Delta\omega_a$ versus the number of steady state photons populating the resonator $\bar{n}$ induced by a drive at ω_d (inset of (b)). The measured shift is plotted for a few different applied magnetic fluxes Φ (denoted with color). Solid lines are fits to $\Delta\omega_a = K\bar{n} + K'\bar{n}^2$. (b) Magnitude of Stark shift per photon $|K|$ as a function of applied magnetic flux Φ. Solid lines are first-principles theory for $|K|$.

Device	$\omega_a/2\pi$ (GHz)	$\kappa/2\pi$	$\lvert g_3\rvert/2\pi$	$\lvert g_4\rvert/2\pi$
A	<6 - 7.84	35 - 55	0.3 - 30	0.001 - 4.9
B	<4 - 7.51	30 - 35	0.5 - 60	0.006 - 0.5
C	5.99 - 7.24	90 - 120	0.4 - 1.5	0.004
D	7.09 - 8.37	180 - 250	0.5 - 1.8	0.003
E	7.76 - 9.24	270 - 440	0.7 - 2.0	0.004

TABLE II. In situ tunable range of phenomenological parameters of 5 devices measured in the experiment: resonant frequency (ω_a), coupling to the 50 Ω transmission line (κ), third-order nonlinearity (g_3), and fourth-order nonlinearity (g_4) where we quote the average for devices C, D and E disregarding the $0.1\,\Phi_0$ region around the Kerr-free point. All parameters given in MHz, except for $\omega_a/2\pi$ in GHz.

alities, such as saturation power. Besides α and L_J, these magnitudes are also influenced by the number of SNAILs M (see Appendix A). At small values, changing M strongly affects $|g_3|$ and $|g_4|$, but we note its influence substantially weakens for $M \gtrsim 20$. Subsequent devices (D and E) have similar magnitudes of the nonlinearities and flux profiles to device C, but instead vary the coupling to the transmission line κ. A summary of these phenomenological parameters for all devices is given in Table II. As we will show next, these factors affect the gain compression.

V. GAIN COMPRESSION

Having established the connection between the physical parameters of our device and the properties of $\boldsymbol{H}_{\rm SPA}$ (Eq. 1), we now optimize the nonlinearities (g_3 and g_4) and the coupling to the transmission line (κ) to achieve higher dynamic range. But first, let us review the causes of amplifier saturation.

The previous formula for gain (Eq. 2) shows no dependence on input signal power, and therefore does not capture the phenomenon of amplifier saturation. To include this dependence, we need to account for the population of the resonator by signal photons at frequency ω_s. We therefore introduce the mean intracavity amplitude α_s. Furthermore, 3-wave mixing creates an image tone (often called the idler) at frequency $\omega_i = \omega_p - \omega_s$, with intracavity amplitude α_i which is comparable to α_s. As shown by a semiclassical harmonic balance analysis (see Appendix B) that includes both of these amplitudes in the input-output theory on equal footing with α_p, the gain G can be recast into a formula similar to Eq. 2. We find

$$G = 1 + \frac{4\kappa^2|g_{\rm eff}|^2}{(\Delta_{\rm eff}^2 - \omega^2 + \frac{\kappa^2}{4} - 4|g_{\rm eff}|^2)^2 + (\kappa\omega)^2}, \qquad (3)$$

where $\omega = \omega_s - \omega_p/2$, $g_{\rm eff} = 2g_3\alpha_p$, and $\Delta_{\rm eff} = \Delta + 12g_4\left[\frac{8}{9}|\alpha_p|^2 + |\alpha_s|^2 + |\alpha_i|^2\right]$ with $\Delta = \omega_a - \omega_p/2$. Considering the on resonance response ($\omega \to 0$), we see that we can tune the pump strength and thus $g_{\rm eff}$ such that the denominator of Eq. 3 goes to 0 and the gain G diverges. Resonant parametric amplifiers operate very close to this parametric instability point, with $g_{\rm eff}$ chosen such that $G = 20$ dB. As a result, slight changes in this denominator are enough to significantly affect the gain G.

Two causes of gain compression can be associated with changes in the denominator of Eq. 3. The first, Kerr-induced Stark shifts, comes from shifts in $\Delta_{\rm eff}$ with increasing signal power. More signal power increases α_s and shifts the resonant frequency due to g_4. Under the approximation that α_p is independent of α_s (often termed the stiff-pump approximation [26, 27]), we can calculate the compression power due to Stark shifts as

$$P_{\rm -1dB}^{\rm Stark} \sim \frac{\kappa}{|g_4|}\frac{1}{G_0^{5/4}}\hbar\omega_a\kappa, \qquad (4)$$

where G_0 is the small-signal gain (see derivation in Appendix B).

The second cause of gain compression visible from Eq. 3 is that of pump depletion, which arises from the breakdown of the stiff-pump approximation [26, 27]. Pump depletion results from the intrinsic nonlinear coupling between the intracavity pump amplitude α_p and the signal amplitude α_s. Thus, increasing α_s changes α_p and consequently the denominator of Eq. 3. Assuming $g_4 = 0$, we can estimate the compression power due to

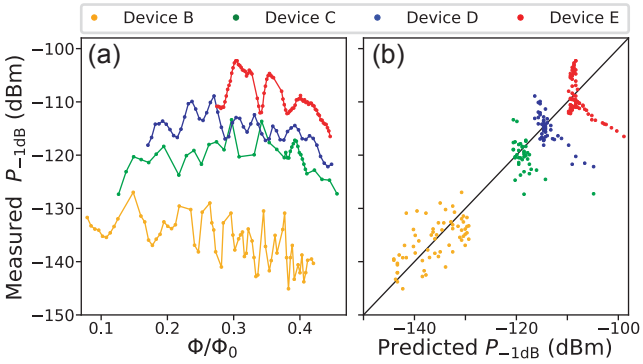

FIG. 5. (a) Measured 1 dB compression power ($P_{-1\mathrm{dB}}$) as a function of applied magnetic flux Φ for four devices biased at 20 dB gain. (b) Comparison between measured value and first-principles theory, which semiclassically treats pump depletion and Stark shifts to second order in harmonic balance.

pump depletion as

$$P_{-1\mathrm{dB}}^{\text{pump dep}} \sim \frac{\kappa}{g_3^2/\omega_a} \frac{1}{G_0^{3/2}} \hbar\omega_a\kappa, \tag{5}$$

where G_0 is the small-signal gain. We note that this compression mechanism arises directly from the third-order nonlinearity that we need for amplification, and is thus unavoidable.

Given these limits on dynamic range, we examine Eq. 4 and Eq. 5 to formulate a recipe for higher compression powers: decrease nonlinearities g_3 and g_4, and increase the dissipation κ. Intuitively, this recipe pushes the optimization closer to a system which obeys the assumptions underlying Eq. 2, namely a more linear oscillator.

Following this recipe requires more applied pump power to reach a desired gain. However, we must be mindful that the current through the SNAIL does not approach the critical current of its Josephson junctions. In practice, applying pump currents that approach the critical current does not directly cause gain compression, but instead determines whether the amplifier achieves the desired small-signal gain in the first place. This limitation translates to $pQ \gtrsim 1$, where p is the inductive participation ratio of all nonlinear elements and Q is the total quality factor of the SPA mode [17]. This is also rigorously equivalent to ensuring the validity of the Taylor expansion of the SNAIL potential in deriving $\boldsymbol{H}_{\mathrm{SPA}}$ (Eq. 1). All amplifiers we consider here satisfy $pQ > 15$ to ensure that the amplifier produced 20 dB of small-signal gain.

We followed the recipe of reducing nonlinearities and increasing dissipation in designing our devices (see Table II), and compare their 1 dB compression powers in Fig. 5a as a function of applied flux Φ. For each point, we measured the resonant frequency ω_a, applied a pump at $\omega_p = 2\omega_a$, and adjusted its power to get $G = 20$ dB. We then measured the $P_{-1\mathrm{dB}}$ compression point (as shown in Fig. 3b). Fig. 5b shows the correlation of our first-principles theory predictions of saturation power with the measured $P_{-1\mathrm{dB}}$, where the black line indicates agreement between theory and experiment. This theory numerically solves the semiclassical Langevin equations of motion to second order in harmonic balance to obtain α_s, α_i, and α_p for given input pump and signal powers (see Appendix B). The gain is then calculated using Eq. 3. We find that, for our devices, the Stark shift mechanism of gain compression closely approximates the full numerical solutions.

To confirm the dependence of $P_{-1\mathrm{dB}}$ on g_4, we first focus on device B, which has the largest $|g_4|/2\pi \in [6, 530]$ kHz for different flux bias points. This change in g_4 results in a systematic 15 dB change in $P_{-1\mathrm{dB}}$ and the theory predicts the trend. We note that the scatter in the data of Fig. 5 results from ripples in the impedance of the transmission line seen by the SPA, which affects its linewidth κ. The compression power is highly sensitive to this parameter, which can be seen in Eq. 4 and Eq. 5.

For device C, we engineered $\kappa/2\pi \in [90, 120]$ MHz and $|g_4|/2\pi \approx 4$ kHz except near its Kerr-free region (see Fig. 4b). These changes in κ and g_4 directly result in device C's increased performance compared to device B. Devices D and E are similar to device C but with increasing $\kappa/2\pi$ to $[180, 250]$ MHz and $[270, 400]$ MHz respectively, and again show improved performance. Specifically, the best device, device E, achieves $P_{-1\mathrm{dB}} \in [-102, -112]$ dBm, which is on par with the best known quantum-limited resonant parametric amplifiers [16–18]. We stress that this performance, achieved with a dynamic bandwidth $\approx 30-40$ MHz, is consistent over the entire tunable bandwidth of 1 GHz.

Despite this increase in dynamic range, Fig. 5b shows that theory predicted that we should have achieved higher saturation powers at certain applied magnetic fluxes. The flux bias points where theory overpredicts $P_{-1\mathrm{dB}}$ are those where g_4 is suppressed (see Eq. 4). Specifically, device C in Fig. 4b shows a tenfold reduction in measured g_4 at around $\Phi/\Phi_0 = 0.4$. However, the measured $P_{-1\mathrm{dB}}$ did not increase near this flux bias point (Fig. 5a). Devices D and E show similar Kerr-free regions as measured by Stark shift, but also do not show increased $P_{-1\mathrm{dB}}$. This puzzle suggests that either our theory has misidentified the cause of amplifier compression, despite its success at all other flux points, or that Kerr is not in fact suppressed in these regions when the pump is on and the amplifier is operational.

VI. INTERMODULATION DISTORTION

To investigate this discrepancy, we measured the Kerr nonlinearity in the presence of the strong pump tone using a third-order intermodulation distortion (IMD) measurement [28]. This standard nomenclature of third-order IMD originates from the fact that fourth-order Kerr terms in the Hamiltonian generate third-order terms in the equations of motion. As we will show, this measurement provides clues about the causes of amplifier satu-

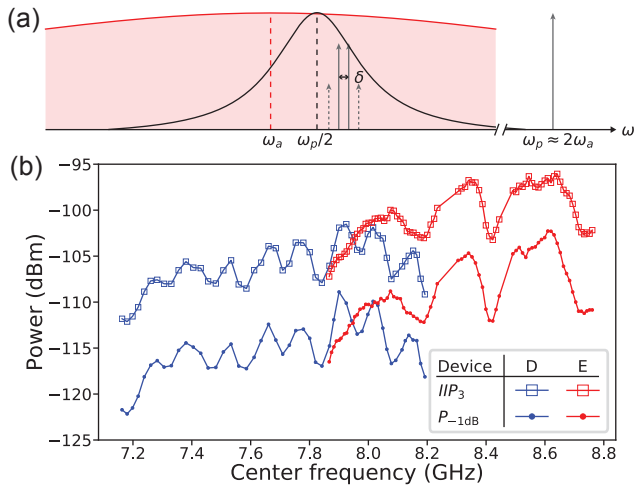

FIG. 6. (a) Caricatured (not-to-scale) frequency spectrum for the measurement of third-order intermodulation distortion products of an SPA. Red shaded region is the Lorenzian lineshape of the linear mode of width κ. Black is the reflection gain of the amplifier when pumped with a strong microwave tone at $\omega_p \approx 2\omega_a$. Solid arrows show two tones applied above $\omega_p/2$ with a relative detuning δ. Dashed arrows denote sidebands generated by the SPA detuned from the main tones by δ. (b) IIP_3 and $P_{-1\mathrm{dB}}$ as a function of the center frequency ($\omega_p/2$) of the 20 dB gain curve for two devices. Neighboring experimental data points have been joined to emphasize correlations between the two experiments.

ration and also probes the response of the amplifier to multi-tone or broadband input signals. Understanding the response to such input signals is particularly crucial for employing quantum-limited amplifiers in any multiplexed readout scheme of superconducting qubits.

A third-order IMD experiment is performed according to the frequency landscape in Fig. 6a. With the pump on and the amplifier biased to $G = 20$ dB, we applied two main signal tones (solid gray arrows) centered at $\omega_p/2 + 2\pi\times 500\,\mathrm{kHz}$ with a relative detuning $\delta/2\pi = 100\,\mathrm{kHz}$ and measured the power in the resulting sidebands (dashed gray arrows). Intuitively, two signal photons from one input tone and one from the other combine in a 4-wave-mixing process to generate the resulting sideband. Thus, the measured relative power between the main tones and the sidebands indicates the amount of spurious 4-wave mixing occurring in the device.

Sweeping the applied power on the two main signal tones, we extracted the input-referred third-order intercept point (IIP_3) at which the measured sideband power would equal the main input signal power (details in Appendix B 2 and Fig. 9a). We conform to the usage in standard microwave amplifier data sheets to take IIP_3 as the metric for third-order IMD. To characterize our amplifiers, in Fig. 6b, we compare the IIP_3 and $P_{-1\mathrm{dB}}$ as a function of the center frequency of the Lorenzian gain curve for two different devices. Each point corresponds to a point tuned up in Fig. 5. Strikingly, the features in $P_{-1\mathrm{dB}}$, which are caused by ripples in the line impedance, are exactly reproduced in IIP_3. Such a comparison indicates that the cause of IIP_3, which is spurious 4-wave mixing, is most likely responsible for the saturation of the amplifier. This confirms Ref. 15's assertion that Kerr is responsible for the saturation of state-of-the-art parametric amplifiers.

Quantitatively, lowest-order harmonic balance theory predicts (see Appendix B) that the measured IIP_3 is related to the Kerr nonlinearity g_4 by the equation

$$IIP_3 = \frac{\kappa}{12|g_4|}\frac{1}{G_0^{3/2}}\hbar\omega_a\kappa, \tag{6}$$

where G_0 is the small-signal gain. However, upon a closer examination of Fig. 6b, we do not observe a distinct peak in IIP_3 for devices D and E, which both support regions where g_4 is suppressed. Thus, we see again as with $P_{-1\mathrm{dB}}$ that the nonlinear properties of the amplifier in the Kerr-free region (as measured by Stark shifts) did not show the expected improvement.

IIP_3 is a measure of nonlinear 4-wave-mixing scattering in the presence of the amplification pump, and thus gives us a clue as to the origin of this discrepancy near the Kerr-free region. Such nonlinear scattering can arise from the multiple terms in the Hamiltonian (Eq. 1): for example, a g_4 process as well as, for instance, two cascaded g_3 processes. While for most of the flux range the g_4 process dominates, near the Kerr-free region g_3 is maximal and the cascaded processes become important. Taking these processes into account is equivalent to going to higher order in harmonic balance, which is shown to improve the agreement between theory and experiment for $P_{-1\mathrm{dB}}$ (see Appendix B).

Not only do IMD measurements help us understand the causes of amplifier saturation, they are also interesting in their own right since reduction of these spurious mixing processes is important for many applications. For instance, any scheme for multiplexed readout of superconducting qubits requires the independence of the readout channels. These spurious intermodulation products will directly limit the isolation between channels either by directly mixing them or by distorting pulses. Furthermore, such intermodulation products put an upper bound on the quantum efficiency of any practical amplifier since, without careful calibration, distortion of the incident quantum signal is unlikely to be accounted for in the experimentalist's demodulation scheme.

VII. CONCLUSION

In conclusion, we have introduced the SPA, a 3-wave-mixing degenerate parametric amplifier, which is simple to design, fabricate, and operate. Through a systematic study across multiple devices, we have confirmed that the fourth-order Kerr term in the amplifier Hamiltonian is the primary cause of gain compression and intermodulation distortion. With this insight, we have optimized

the SPA to achieve 1 dB compression powers on par with the best reported values for resonant quantum-limited parametric amplifiers, but over the entire tunable bandwidth of 1 GHz of the device, and without sacrificing any other desirable characteristics.

Importantly, the most precious of these characteristics, quantum-limited noise performance, was confirmed through comparing noise visibility ratio (NVR) measurements. A proxy for noise temperature, NVR is the ratio between the noise power spectral density with the pump on and the pump off. All amplifiers in this work were measured to have comparable NVRs and thus comparable noise performances to other quantum-limited amplifiers measured in the same system [12, 29]. Moreover, an SPA was shown to improve the readout of a superconducting qubit in Ref. 30, where the quantum efficiency of a phase-sensitive measurement chain involving an SPA was measured to be $\eta = 0.6$ in a self-calibrated manner.

Our work on the improvement of amplifier performance can be carried out further. One puzzling observation was the absence of a peak in the saturation power at the Kerr-free point, despite the confirmation that Kerr-induced Stark shifts are the primary cause of gain compression. The IMD measurements suggested that Stark shifts caused by higher harmonics limit the saturation power at the Kerr-free point. A natural next step would be to understand how to reduce these spurious harmonics in the presence of a strong pump drive. An alternative strategy would be to further reduce amplifier nonlinearity, which should increase the saturation power at the cost of requiring more pump power to achieve the same gain. As such, effectively engineering the pump power delivery network to achieve the desired pump strength, without introducing excess noise or heating up the base plate of the dilution refrigerator, will become increasingly more crucial for higher dynamic range amplifiers. More broadly, the optimizations performed in this work for higher dynamic range do not conflict with recent approaches for enhanced dynamic bandwidth via impedance engineering [18, 31], nor with approaches for directional amplification [32–37]. The reduction of Kerr and the use of arrays of nonlinear elements should also increase the dynamic range of traveling wave amplifiers [13, 38–42].

Furthermore, our results indicate that 3-wave mixing with an array of SNAILs is a particularly robust building block for information processing with superconducting quantum circuits. With this versatile tool, both the third-order and the Kerr nonlinear parameters can be controlled over many orders of magnitude. Moreover, the sign of Kerr can be changed and its magnitude suppressed in situ by tuning an applied magnetic flux. Such control can be convenient in parameter regimes which are rather different from quantum-limited amplifiers, such as for instance, in superconducting qubits [43, 44]. The method of arraying multiple SNAILs is more generally applicable for optimizing parametrically induced mixing, such as in quantum-limited switches [45, 46], frequency converters [47–53] and other quantum devices [54, 55].

ACKNOWLEDGEMENTS

We acknowledge fruitful discussions with Shantanu Mundhada and Luigi Frunzio. We are grateful to Vladislav Kurilovich and Pavel Kurilovich for performing Born-Oppenheimer elimination of the fast SNAIL mode (see Appendix A) to verify our semiclassical model. We also acknowledge the Yale Quantum Institute. Facilities use was supported by YINQE, and the Yale SEAS cleanroom. This research was supported by ARO under Grant No. W911NF-14-1-0011 and W911NF-18-1-0212, and S.S. acknowledges support from ARO W911NF-14-1-0563.

Appendix A: SPA Hamiltonian

In this Appendix, we discuss the map between the physical layout of the SPA and the Hamiltonian parameters. Section A 1 describes a single-mode lumped element model of the SPA and shows the importance of the nonlinear current conservation for predicting the Kerr nonlinearity. Section A 2 provides further insight about the renormalization of Kerr by high energy modes. In Section A 3 we treat the SPA resonator as a distributed circuit element and show the expressions referred to as "first-principles theory" in the main text.

1. Lumped-element model of SPA

The SNAIL [12] is a one-loop dipole element composed of three identical large Josephson junctions (inductance L_J) in one arm and a smaller junction (inductance L_J/α) in the other arm (see Fig. 1c). The loop is threaded with an applied magnetic flux Φ. The internal capacitances of the junctions are small and the self-resonance frequencies of the element are expected to be above 30 GHz. Therefore, at typical frequencies of circuit QED experiments, it behaves as a nonlinear inductor and its lowest energy configuration corresponds to having equal phase drops across the three large junctions. Thus, the SNAIL provides a 6π-periodic potential energy

$$U_{\mathrm{SNAIL}}(\varphi_s) = -\alpha E_J \cos\varphi_s - 3E_J \cos\frac{\varphi_{\mathrm{ext}} - \varphi_s}{3}, \quad \text{(A1)}$$

where E_J is the Josephson energy, and $\varphi_{\mathrm{ext}} = 2\pi\Phi/\Phi_0$. We operate in a regime where phase fluctuations are suppressed, and the potential can be Taylor expanded near one of its equivalent minima. We denote the expansion coefficients by $c_n = \frac{1}{E_J}\frac{d^n U_{\mathrm{SNAIL}}}{d\varphi_s^n}(\varphi_{\mathrm{min}})$ and the minimum location φ_{min} is determined from the condition

$$c_1 \equiv \alpha \sin\varphi_{\mathrm{min}} + \sin\frac{\varphi_{\mathrm{min}} - \varphi_{\mathrm{ext}}}{3} = 0. \quad \text{(A2)}$$

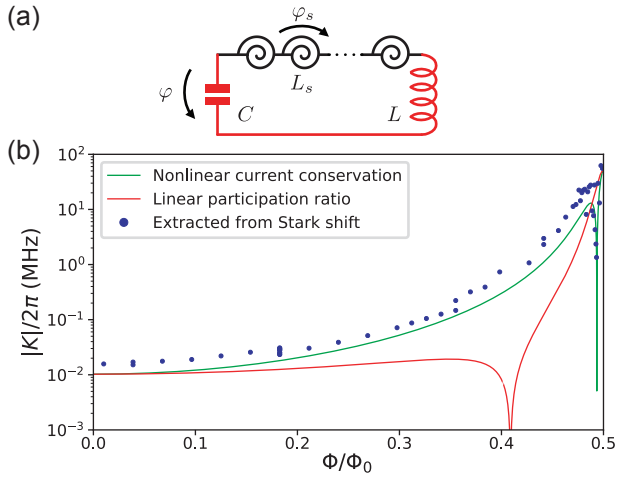

FIG. 7. (a) Single-mode lumped element circuit model for an SPA with M SNAILs, where the microstrip resonator is approximated by a series LC circuit (red). Phase drop across each SNAIL is denoted as φ_s, while φ is the canonical phase coordinate for the mode. (b) Magnitude of Stark shift per photon $|K|$ of device A from Fig. 4, together with theoretical predictions: green includes nonlinear current conservation, red is linear participation ratio based approach in which $K \propto c_4$.

These flux-dependent coefficients $c_n = c_n(\varphi_{\rm ext})$, together with L_J, completely characterize the behavior of the SNAIL when embedded in a larger electrical circuit. For instance, the linear inductance of the SNAIL is $L_s(\varphi_{\rm ext}) = L_J/c_2(\varphi_{\rm ext})$.

To realize an SPA with a resonance in the 4-10 GHz range, an array of M identical SNAILs is embedded into a transmission line resonator. The simplified circuit model, shown in Fig. 7a, consists of a series combination of capacitance C, inductance L, and an array of M SNAILs. We will further assume that the SNAIL array can be considered as a lumped sub-circuit, in which the phase splits equally among the individual SNAILs. This assumption is justified for small capacitance to ground of the inter-SNAIL islands [56] and for identical SNAILs.

In this case the total inductance of the emergent electromagnetic mode becomes flux-tunable, as it consists of flux-independent inductance L and flux-dependent inductance $L_s^{\rm array}(\varphi_{\rm ext}) = ML_s(\varphi_{\rm ext})$ coming from the SNAIL array. The resonance frequency of this SPA mode is

$$\omega_a(\varphi_{\rm ext}) = \frac{1}{\sqrt{C(L + L_s^{\rm array}(\varphi_{\rm ext}))}} = \frac{\omega_0}{\sqrt{1 + M\xi_J/c_2(\varphi_{\rm ext})}}, \tag{A3}$$

where we have defined the dimensionless coefficient $\xi_J = L_J/L$ and the resonance frequency in the absence of the array $\omega_0 = 1/\sqrt{LC}$.

The Lagrangian of this system is

$$\mathcal{L} = \frac{C\varphi_0^2}{2}\dot{\varphi}^2 - U(\varphi_s, \varphi), \tag{A4a}$$

$$U(\varphi_s, \varphi) = MU_{\rm SNAIL}(\varphi_s) + \frac{1}{2}E_L(\varphi - M\varphi_s)^2, \tag{A4b}$$

where φ is the mode canonical phase coordinate, φ_0 is the reduced flux quantum, and $E_L = \varphi_0^2/L$.

The coordinate φ_s is not an independent variable, as it does not have its own kinetic energy. Therefore, prior to quantization, we need to eliminate it by minimizing the nonlinear potential energy $U(\varphi_s, \varphi)$ as a function of φ_s. The resulting trivial Lagrange equation of motion is equivalent to imposing a full nonlinear current conservation condition at the node between the SNAIL array and the linear inductor. Using (A1) we can write it as

$$\alpha \sin\varphi_s + \sin\frac{\varphi_s - \varphi_{\rm ext}}{3} + \xi_J(M\varphi_s - \varphi) = 0. \tag{A5}$$

This equation determines the SNAIL phase $\varphi_s[\varphi]$ as a function of mode canonical coordinate φ, which has to be further quantized. The potential energy can now be written in terms of a single degree of freedom φ

$$U(\varphi) = MU_{\rm SNAIL}(\varphi_s[\varphi]) + \frac{1}{2}E_L(\varphi - M\varphi_s[\varphi])^2. \tag{A6}$$

Given that we operate in the regime of small phase fluctuations, we can again Taylor expand the renormalized potential $U(\varphi)$, resulting in the coefficients $\tilde{c}_n = \frac{1}{E_J}\frac{d^nU}{d\varphi^n}(\overline{\varphi}_{\rm min})$, where $\overline{\varphi}_{\rm min}$ is determined from the condition $\tilde{c}_1 = 0$. Using the current conservation equation (A5), we can write the first four Taylor coefficients as

$$\tilde{c}_1 = \xi_J(\overline{\varphi}_{\rm min} - M\varphi_s[\overline{\varphi}_{\rm min}]), \tag{A7}$$

$$\tilde{c}_2 = \xi_J(1 - M\frac{\partial\varphi_s}{\partial\varphi}[\overline{\varphi}_{\rm min}]), \tag{A8}$$

$$\tilde{c}_3 = -M\xi_J\frac{\partial^2\varphi_s}{\partial^2\varphi}[\overline{\varphi}_{\rm min}], \tag{A9}$$

$$\tilde{c}_4 = -M\xi_J\frac{\partial^3\varphi_s}{\partial^3\varphi}[\overline{\varphi}_{\rm min}]. \tag{A10}$$

The derivatives of the implicit function $\varphi_s[\varphi]$ can be found by differentiating (A5). For example, by differentiating it once we get

$$\frac{\partial\varphi_s}{\partial\varphi} = \frac{\xi_J}{\alpha\cos\varphi_s + \frac{1}{3}\cos\frac{\varphi_s - \varphi_{\rm ext}}{3} + M\xi_J}, \tag{A11}$$

In addition, we can show from (A5) that the presence of the series linear inductor does not change the location of the SNAIL potential minimum $\varphi_s[\overline{\varphi}_{\rm min}] = \varphi_{\rm min}$. Therefore, we can express the Taylor coefficients $\tilde{c}_n$ for the renormalized potential $U(\varphi)$ in terms of the bare ones c_n

introduced earlier for a single SNAIL

$$\tilde{c}_2 = \frac{p}{M} c_2, \tag{A12}$$

$$\tilde{c}_3 = \frac{p^3}{M^2} c_3, \tag{A13}$$

$$\tilde{c}_4 = \frac{p^4}{M^3}\left(c_4 - \frac{3c_3^2}{c_2}(1-p)\right), \tag{A14}$$

where we have defined the array linear participation ratio

$$p \equiv \frac{ML_s}{L + ML_s} = \frac{M\xi_J}{c_2 + M\xi_J}. \tag{A15}$$

After performing the Legendre transformation and canonical quantization, the Hamiltonian of the mode can be written as

$$\boldsymbol{H} = 4E_C \boldsymbol{N}^2 + E_J\left(\frac{\tilde{c}_2}{2!}\boldsymbol{\varphi}^2 + \frac{\tilde{c}_3}{3!}\boldsymbol{\varphi}^3 + \frac{\tilde{c}_4}{4!}\boldsymbol{\varphi}^4 + ...\right), \tag{A16}$$

where $E_C = e^2/2C$ and $[\boldsymbol{\varphi}, \boldsymbol{N}] = i$. For more convenience, we can introduce the bosonic raising and lowering operators $\boldsymbol{a}^\dagger$ and $\boldsymbol{a}$ that diagonalize the quadratic part of the Hamiltonian in the excitation number basis. The SPA Hamiltonian after this second quantization can be written, truncated to fourth order, as

$$\boldsymbol{H}_{\rm SPA}/\hbar = \omega_a \boldsymbol{a}^\dagger \boldsymbol{a} + g_3(\boldsymbol{a} + \boldsymbol{a}^\dagger)^3 + g_4(\boldsymbol{a} + \boldsymbol{a}^\dagger)^4, \tag{A17}$$

where

$$\hbar g_3 = \frac{1}{6}\frac{p^2}{M}\frac{c_3}{c_2}\sqrt{E_C \hbar \omega_a}, \tag{A18}$$

$$\hbar g_4 = \frac{1}{12}\frac{p^3}{M^2}\left(c_4 - \frac{3c_3^2}{c_2}(1-p)\right)\frac{1}{c_2}E_C. \tag{A19}$$

In general this Hamiltonian has both odd and even nonlinearities for $\Phi/\Phi_0 \neq n/2$ (where n is any integer), unlike the symmetric transmon Hamiltonian [57]. These nonlinearities inherit their flux dependence from that of the SNAIL potential, and thus are tunable in situ.

We can now relate these nonlinearities to the Stark shift per photon K that is measured in the experiment. In a nearly harmonic oscillator, K can be calculated as the dispersion of transition frequencies between the neighboring energy levels,

$$\hbar K = \frac{d^2 E(n)}{dn^2}, \tag{A20}$$

where $E(n)$ is the n'th energy level of the Hamiltonian (A17). In the case $g_3 = 0$, which is the well-known Duffing oscillator model, K is simply related to the Hamiltonian parameter g_4 via $K = 12g_4$. However, in the asymmetric SPA potential this relation is modified to $K = 12\left(g_4 - 5g_3^2/\omega_a\right)$, where the last term comes from the second-order perturbation theory correction to the energy levels. Using expressions (A18-A19) we thus obtain:

$$\hbar K = \frac{p^3}{M^2}\left(c_4 - \frac{3c_3^2}{c_2}(1-p) - \frac{5}{3}\frac{c_3^2}{c_2}p\right)\frac{1}{c_2}E_C. \tag{A21}$$

We would like to stress that the calculation of the Kerr effect that we have outlined is significantly different from previous calculations in 3-wave-mixing amplifiers, such as for the JPC, in Ref. 15, 58, and 59. Previous calculations considered first the linearized circuit, in which the total phase drop φ splits between the nonlinear circuit elements and the linear inductor in proportion to their respective participation ratios p and $1-p$. Then they assumed that the nonlinearity is diluted by the corresponding power of the participation ratio, in which case we would have $g_n \propto p^{n-1}c_n$. As we can see from (A18) and (A19), this approach yields the correct values for the lowest order cubic nonlinearity, but fails to predict higher order nonlinearities, such as g_4 and therefore K, correctly.

This discrepancy arises because the linear participation ratio based approach does not properly account for the nonlinear current conservation (A5) in the SPA between the SNAIL array and the inductor. This effect leads to a renormalization of g_4 due to c_3 evident in (A19). One can see that in the limit of small participation ratio $p \to 0$ the additional contribution is equal to $-3c_3^2/c_2$ which does not contain any small parameters relative to c_4. This significantly shifts the Kerr-free point in flux and modifies the whole Kerr nonlinearity profile. The comparison of data taken on device A with predictions of both approaches is shown in Fig. 7b. Note that at certain fluxes the predictions differ by several orders of magnitude.

Moreover, apart from this renormalization effect, there is a trivial second order perturbation theory correction to energy levels due to the g_3 term in the Hamiltonian, which affects the Stark shift per photon K (last term in Eq. A21). Note that both g_4 and this correction scale identically with E_C. This perturbative correction is insignificant in the limit of small participation ratio p. On the other hand, in the limit $p \to 1$, the renormalization effect due to the linear inductance becomes irrelevant and the perturbative contribution becomes important instead.

2. Renormalization of Kerr

We can also explain the physics of the nontrivial contribution to g_4 in (A19) from a different perspective. This correction is nothing but a renormalization of the potential due to high frequency modes that are inevitably present in the system (for example, the plasma modes of the SNAIL junctions). Such modes cannot be directly probed by low energy experiments, but their effect is evident in measurable quantities such as the Stark shift per photon K.

To demonstrate this point, consider the circuit in Fig. 8a, which includes the capacitance C_s shunting the SNAIL (we consider $M = 1$ for simplicity). Such a circuit has two eigenmodes: the SPA low frequency mode ($\omega_a/2\pi \sim 7\,\mathrm{GHz}$) and the high frequency SNAIL mode ($\Omega_s/2\pi > 30\,\mathrm{GHz}$). In the low participation ratio limit,

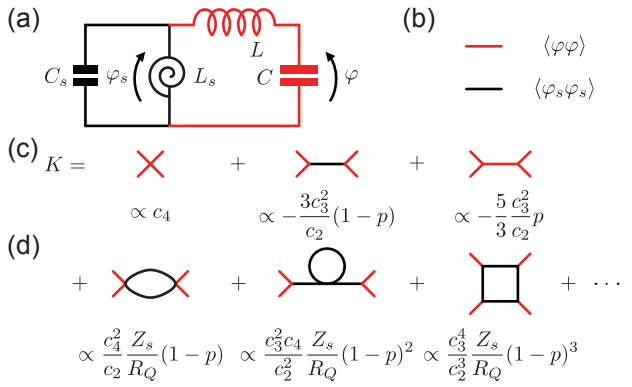

FIG. 8. (a) Two-mode lumped element circuit model for an SPA with $M = 1$ SNAIL. One mode is the SPA mode, frequency $\omega_a \approx 1/\sqrt{C(L+L_s)}$, and the other is a high frequency SNAIL mode, frequency $\Omega_s \approx 1/\sqrt{C_s L_s} \gg \omega_a$. (b) Definitions for elements of Feynman diagrams used to eliminate the high frequency SNAIL mode. Red (black) lines represent propagator for low (high) frequency mode with canonical coordinate φ (φ_s). (c) Diagrammatic series to calculate Stark shift per photon K of SPA (red) mode. Each diagram corresponds to a term in Eq. A21, where third (fourth) order interaction vertices are weighted by c_3 (c_4) from the SNAIL potential. (d) Example higher order 1-loop diagrams, each suppressed by Z_s/R_Q where $Z_s = \sqrt{L_s/C_s}$ and $R_Q = \hbar/4e^2$.

the SPA mode is mostly localized in the series LC circuit (red in Fig. 8a), with the high frequency mode in the SNAIL and its shunting capacitor (black).

In this system, in contrast with (A4a) (see also Fig. 7a), the variable φ_s becomes a real quantum mechanical coordinate with its own conjugate momentum. Therefore, it should be quantized on equal footing with φ. However, since $\Omega_s \gg \omega_a$, this fast degree of freedom can be integrated out using the Born-Oppenheimer method or more sophisticated QFT techniques [60].

In the Feynman diagram language, the Kerr nonlinearity of the SPA mode can be represented as a fourth-order self-interaction vertex. The example diagrams that contribute to the renormalization of this vertex are depicted in Fig. 8c,d.

The first diagram in Fig. 8c comes directly from the quartic term in the potential energy of the SNAIL element (A1) and is therefore proportional to the corresponding Taylor coefficient c_4. Including only this diagram results in $K \propto c_4$, which is equivalent to the linear participation ratio based approach discussed previously.

The interesting and nontrivial correction to K comes from the high energy SNAIL mode, and is represented by the second diagram in Fig. 8c. This contribution is equivalent to the second term in (A21), and it does not depend on Ω_s as long as the requirement $\Omega_s \gg \omega_a$ is satisfied. In the previous single-mode model this contribution arose from imposing the full nonlinear current conservation (A5).

The last diagram in Fig. 8c depicts two cascaded c_3-self-interactions of the SPA mode. This is equivalent to the trivial second-order perturbation theory correction, as seen in the last term of (A21). Together, the three diagrams in Fig. 8c give Kerr to the leading order and fully replicate the result (A21).

In principle, there are other contributions to the renormalization of the fourth-order self-interaction vertex. Some example one-loop processes are shown in Fig. 8d. However, such diagrammatic corrections are suppressed by a factor $(Z_s/R_Q)^l$, where Z_s is the impedance of the fast mode and l is the number of loops. In practice, $Z_s/R_Q \ll 1$ in our fabrication process. In fact, it is fundamentally difficult to achieve $Z_s \gtrsim R_Q$ [56].

Comparing the two electrical circuits in Fig. 7a and Fig. 8a, we see that the circuit in 7a corresponds to taking the limit $C_s \to 0$ for the one in 8a. Rigorously taking this limit would lead to large quantum fluctuations of phase, and the diagrammatic series would diverge. To ensure the convergence, C_s still has to be large enough to satisfy $Z_s \ll R_Q$. In this case, classical elimination of the fast degree of freedom φ_s via nonlinear current conservation (A5) is justified.

Understanding these effects is important because the general approach for designing circuit QED systems relies on pushing the spurious modes up in frequency and then neglecting their influence by arguing that the detuning to these modes is large. We have shown that the presence of these modes can influence low energy observables, and they have to be accounted for either by means of full nonlinear current conservation, or equivalently by integrating out the high energy modes.

3. Distributed-element model of SPA

While the lumped element model above elucidates the important details in the theoretical treatment of the SPA and provides physical intuition, it cannot strictly be applied to our devices. In the SPA, the SNAIL array is embedded in a transmission line resonator, which is a distributed circuit element, see Fig. 1d.

The Lagrangian of such a system can be written as

$$\mathcal{L} = \left(\int_{-l_{\rm MS}/2}^{-0} + \int_{+0}^{l_{\rm MS}/2}\right)\left[\frac{c}{2}(\partial_t\phi) - \frac{1}{2\ell}(\partial_x\phi)^2\right]dx - MU_{\rm SNAIL}\left(\frac{\varphi_r - \varphi_l}{M}\right), \tag{A22}$$

where c is the capacitance per unit length and ℓ is the inductance per unit length. The generalized flux $\phi(x,t)$ on the transmission line is a one-dimensional massless Klein-Gordon field which has a discontinuity at $x = 0$, where the transmission line is interrupted by the lumped element SNAIL array. For convenience, we have introduced $\varphi_l = \phi(-0,t)/\varphi_0$ and $\varphi_r = \phi(0+,t)/\varphi_0$ to denote the superconducting phase on both sides of the array.

Using zero current boundary conditions at $x = \pm l_{\rm MS}/2$ and linearizing the Lagrange equation of motion for this system, we can perform an eigenmode decomposition and

find the resonant frequency ω_a of the structure as the smallest nontrivial solution of the equation

$$\omega_a \tan\left(\frac{\pi}{2}\frac{\omega_a}{\omega_0}\right) = \frac{2Z_c}{ML_s(\varphi_{\rm ext})}, \tag{A23}$$

where $Z_c = \sqrt{\ell/c}$ is the characteristic impedance of the transmission line, and $\omega_0 = \pi/l_{\rm MS}\sqrt{\ell c}$ is the resonant frequency when the array of SNAILs is replaced with a short. We use (A23) for the fits of ω_a in Fig. 2.

Following Refs. 61 and 62, we calculate the nonlinearities of the SPA as

$$g_3 = \frac{4Z_c c_3}{3M^2 L_J}\sqrt{\frac{Z_c}{R_Q}}\left(\frac{\cos^2\frac{\pi}{2}\frac{\omega_a}{\omega_0}}{\frac{\pi}{2}\frac{\omega_a}{\omega_0}+\sin\pi\frac{\omega_a}{\omega_0}}\right)^{3/2}, \tag{A24}$$

$$K = \frac{2\omega_a\cos^2\frac{\pi}{2}\frac{\omega_a}{\omega_0}}{c_2 M^3(\frac{\pi}{2}\frac{\omega_a}{\omega_0}+\sin\pi\frac{\omega_a}{\omega_0})^2}\frac{Z_c}{R_Q} \times\left(c_4 - \frac{c_3^2}{c_2}\frac{3+5\left(\frac{\omega_a M L_s}{2Z_c}\right)^2}{1+3\left(\frac{\omega_a M L_s}{2Z_c}\right)^2}\right), \tag{A25}$$

where $R_Q = \hbar/(2e)^2$ is the resistance quantum. The limits of small and unity participation ratio correspond in this model to $ML_s \ll Z_c/\omega_a$ and $ML_s \gg Z_c/\omega_a$ respectively. In these limits, the corrections to Kerr due to c_3 coincide in both the distributed- (A25) and lumped-element (A21) models.

We use the expressions (A24) and (A25) for g_3 and K to plot the first-principles theory curves in Fig. 3c and Fig. 4b respectively.

4. Scaling with number of SNAILs

The above expressions for g_3 and K give these nonlinearities as functions of designable parameters, such as the number of SNAILs M. However, the scaling with M can be rather unintuitive when applied to practical device design changes. For example, we consider the goal of decreasing K by increasing M. Using the lumped-element expression (A21) for simplicity, the dependence on M is $K \propto 1/M^2$ when p, α, and E_C are held constant. To then hold the operating frequency ω_a constant, increasing M requires a corresponding increase in E_J (smaller L_J), as is suggested in Ref. 14 for higher dynamic range in JPAs. Unfortunately, parasitic geometric inductance in the leads used to make the junctions prevents the reduction of L_J much below its current $\approx$ 40 pH value without significantly affecting p.

In practice, when increasing M, we hold L_J constant and increase E_C to realize the desired frequency ω_a. The predicted scaling in this instance is weaker ($K \propto p^2/M$). Moreover, this tactic also tends to increase p, which further weakens this scaling. As such, we find that increasing M much past 20 for our design does not strongly reduce K until $M \gtrsim 200$, at which point the SNAIL array can no longer be considered a lumped element. Future work must model these larger arrays as SNAIL transmission lines, where increasing M reduces K if the characteristic impedance is independent of M.

Appendix B: Dynamic Range of an Amplifier

In this Appendix we discuss the map between the Hamiltonian and input-output port coupling parameters and the amplifier characteristics, specifically the saturation power and intermodulation distortion. In Section B 1, we theoretically treat amplifier saturation by performing harmonic balance analysis of arbitrary-signal-power scattering. Section B 2 contains theoretical treatment of IMD. Section B 3 compares in detail the measured saturation power and theory.

1. Semiclassical solution: harmonic balance

We analyze the response of the system using standard input-output theory together with the Quantum Langevin Equation (QLE) for mode $\boldsymbol{a}$

$$\dot{\boldsymbol{a}} = \frac{i}{\hbar}[\boldsymbol{H}_{\rm SPA} + \boldsymbol{H}_{\mathbf{drive}}, \boldsymbol{a}] - \frac{\kappa}{2}(\boldsymbol{a} - \boldsymbol{a}^\dagger), \tag{B1}$$

$$\boldsymbol{H}_{\mathbf{drive}} = \hbar(\boldsymbol{u}_{\rm in} + \boldsymbol{u}_{\rm in}^\dagger)(\boldsymbol{a} + \boldsymbol{a}^\dagger), \tag{B2}$$

$$\boldsymbol{u}_{\rm out} = i\kappa(\boldsymbol{a} - \boldsymbol{a}^\dagger) - \boldsymbol{u}_{\rm in}, \tag{B3}$$

written here without the rotating wave approximation (RWA) since we are applying an off-resonant pump. The classical drive amplitude $u_{\rm in(out)} = \sum_\omega u^\omega_{\rm in(out)} e^{-i\omega t}$ is related to the input (output) power at the corresponding frequency ω via $P_{\rm in(out)} = \frac{\hbar\omega_a}{\kappa}|u^\omega_{\rm in(out)}|^2 \cdot \frac{\omega_a^2}{\omega^2}$ (for the capacitive coupling).

A linear harmonic oscillator which is pumped at ω_p and probed at ω_s responds independently to each of these frequency components of the incoming field. In contrast the SPA, which consists of a weakly nonlinear resonator, will produce a response at all harmonics $\omega_{mn} = m\omega_p + n\omega_s$. Conventional degenerate parametric amplifier theory [2, 22] takes into account only one additional harmonic (the idler) at $\omega_i = \omega_p - \omega_s$, although in practice all higher intermodulation products (IMDs) will be created. Their magnitudes are often small and therefore neglecting these harmonics is a reasonable starting point.

Restricting ourselves to signal ω_s, idler ω_i, and pump ω_p frequencies, we can solve the QLE (B1-B3) using the semiclassical harmonic balance method [15, 26, 63]. We will denote the input drive strengths at ω_s, ω_i, and ω_p as u_s, u_i, and u_p respectively, and treat them as classical drives (i.e. c-numbers). After applying harmonic balance conditions, we obtain a self-consistent system of coupled equations that relates all harmonics to each other:

$$\Big(\omega_p - \omega_a + i\frac{2}{3}\kappa - g_4\big[\frac{32}{9}|\alpha_p|^2 + 16|\alpha_s|^2 + 16|\alpha_i|^2\big]\Big)\alpha_p = u_p + 6g_3\alpha_i\alpha_s, \tag{B4}$$

$$\Big(\omega_s - \omega_a + i\frac{\kappa}{2} - g_4\big[\frac{32}{3}|\alpha_p|^2 + 12|\alpha_s|^2 + 12|\alpha_i|^2\big]\Big)\alpha_s = u_s + 4g_3\alpha_p\alpha_i^*, \tag{B5}$$

$$\Big(\omega_i - \omega_a + i\frac{\kappa}{2} - g_4\big[\frac{32}{3}|\alpha_p|^2 + 12|\alpha_i|^2 + 12|\alpha_s|^2\big]\Big)\alpha_i = u_i + 4g_3\alpha_p\alpha_s^*, \tag{B6}$$

where α_s, α_i and α_p denote the mean intracavity amplitudes at the corresponding frequencies.

To cast these equations into a more familiar form, we introduce the signal detuning $\omega = \omega_s - \frac{\omega_p}{2}$, the pump detuning $\Delta = \omega_a - \frac{\omega_p}{2}$, and the effective pump detuning that includes Stark shift effects

$$\Delta_{\rm eff} = \Delta + g_4\big[\frac{32}{3}|\alpha_p|^2 + 12|\alpha_s|^2 + 12|\alpha_i|^2\big]. \tag{B7}$$

With the help of these notations, equations (B5-B6) can be written via a susceptibility matrix as

$$\begin{pmatrix} \alpha_s \\ \alpha_i^* \end{pmatrix} = \frac{1}{\mathcal{D}}\begin{pmatrix} -\omega - \Delta_{\rm eff} - i\frac{\kappa}{2} & 2g_{\rm eff} \\ 2g_{\rm eff}^* & \omega - \Delta_{\rm eff} + i\frac{\kappa}{2} \end{pmatrix}\begin{pmatrix} u_s \\ u_i^* \end{pmatrix},$$

$$\mathcal{D} = \Delta_{\rm eff}^2 - \omega^2 + \frac{\kappa^2}{4} - 4|g_{\rm eff}|^2 - i\kappa\omega, \tag{B8}$$

where $g_{\rm eff} = 2g_3\alpha_p$. Using (B8) and the input-output relation (B3), we calculate the phase-preserving power gain $G = |u_{\rm out}^{\omega_s}|^2/|u_s|^2$. After some algebraic transformations, it can be cast into the form

$$G = 1 + \frac{4\kappa^2|g_{\rm eff}|^2}{(\Delta_{\rm eff}^2 - \omega^2 + \frac{\kappa^2}{4} - 4|g_{\rm eff}|^2)^2 + (\kappa\omega)^2}. \tag{B9}$$

We note the close similarity between this expression for gain and (2). However, here all parameters are effective ones that depend on the input signal power. For example, the parameter g in (2) depends on α_p which is treated as a constant. In (B9) however, α_p is determined from the self-consistent system of equations (B4-B6), and therefore depends on the signal power.

To achieve large gain, the denominator in (B9) should be tuned close to zero, often called the parametric instability point. Altering this denominator by even a small amount in response to increased input signal power will reduce the gain and cause amplifier saturation.

The small change in gain can therefore be written as

$$\delta G = G_0^{3/2}\frac{8g_3}{\kappa\sqrt{n_p}}\delta n_p - G_0^{3/2}\frac{1}{2\kappa g_3\sqrt{n_p}}\delta\Delta_{\rm eff}^2, \tag{B10}$$

where G_0 denotes the small-signal gain and $n_p = |\alpha_p|^2$. We have also restricted to the case $\omega = 0$ and neglected the weak influence of numerator in (B9).

The first term in (B10) shows the effect of changes in pump population n_p. Pump population can be depleted in response to increasing signal power due to the intrinsic g_3-induced coupling of α_p and α_s in (B4). This back-action on the pump due to the signal [26], often called pump depletion, will cause amplifier saturation even when $g_4 = 0$. Moreover, pump depletion can also arise due to the amplification of quantum fluctuations within the amplifier band [27].

The second term in (B10) shows the effect of changes in $\Delta_{\rm eff}$. This change originates from the Stark shift contributions of the input signal in the detuning (B7). In the first-order harmonic balance theory discussed so far, this effect vanishes when $g_4 = 0$.

From (B10) it is straightforward to estimate how the saturation power scales with the parameters of the SPA for these two mechanisms

$$P_{-1\rm dB}^{\rm Stark} \sim \frac{\kappa}{|g_4|}\frac{1}{G_0^{5/4}}\hbar\omega_a\kappa, \tag{B11}$$

$$P_{-1\rm dB}^{\rm pump\,dep} \sim \frac{\kappa}{g_3^2/\omega_a}\frac{1}{G_0^{3/2}}\hbar\omega_a\kappa, \tag{B12}$$

where for brevity we have assumed $\Delta = 0$.

Before comparing this semiclassical harmonic balance theory to the experimental data, let us cover one more topic which is relevant for amplifiers: spurious intermodulation distortion.

2. Intermodulation distortion (IMD)

Let us now extend the analysis to consider the response of an amplifier to two monochromatic input tones. Standard parametric amplifier theory predicts that the output will consist of two amplified tones at the same frequencies and two idler tones. However, because of the resonator nonlinearity, higher-order-mixing products will inevitably be created. The third-order IMD products are most relevant to the amplifier quality, as they are created within the amplifier bandwidth [28].

Third-order IMDs measure the nonlinear 4-wave-mixing scattering processes of a device, which can be understood from the following frequency conditions, see Fig. 6a. If we send two signals at ω_{s1} and ω_{s2}, the nearest third-order-mixing products will be created at sideband frequencies $2\omega_{s1} - \omega_{s2}$ and $2\omega_{s2} - \omega_{s1}$. Focusing on the IMD product at $\omega_{\rm imd} = 2\omega_{s1} - \omega_{s2}$, this 4-wave-mixing process corresponds to the annihilation of 2 photons at

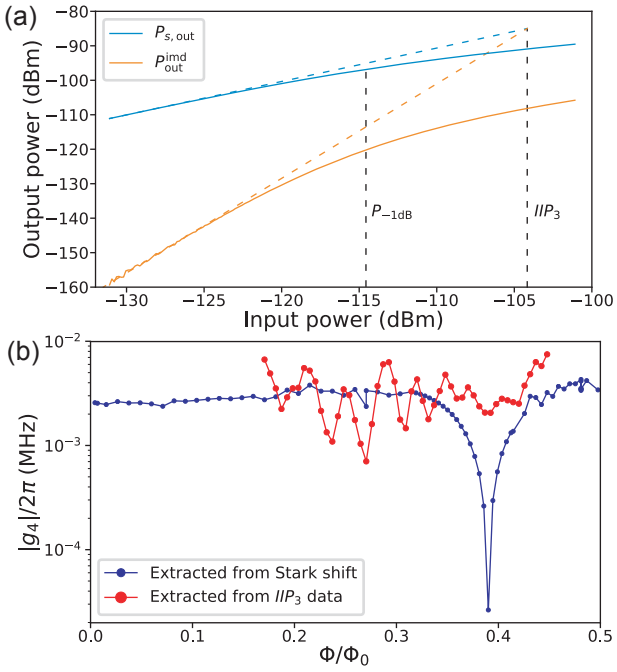

FIG. 9. (a) Example IIP_3 experiment showing measured signal power $P_{s,\mathrm{out}}$ (solid blue) and third-order IMD sideband power $P_{\mathrm{out}}^{\mathrm{imd}}$ (solid orange) as a function of applied input power $P_{s,\mathrm{in}}$. Intersection of low-power asymptotes (dashed) gives IIP_3. The 1 dB compression power $P_{-1\mathrm{dB}}$ is also indicated for reference. (b) 4-wave-mixing nonlinearity g_4 as a function of applied magnetic flux Φ for device D, extracted from the Stark shift (blue) or IIP_3 (red) experiments.

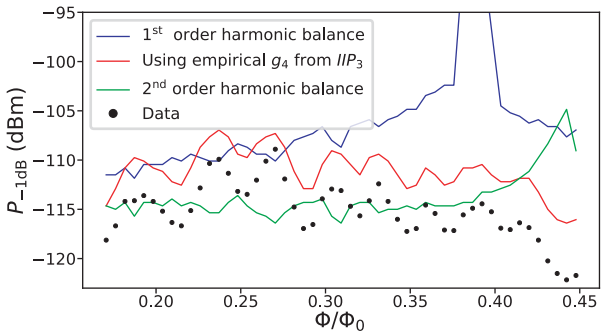

FIG. 10. Measured $P_{-1\mathrm{dB}}$ (black) as a function of Φ for device D biased at 20 dB gain. Solid lines are predictions of first-order harmonic balance theory (B4-B6) using g_4 calculated from first-principles (blue) or empirically extracted from IIP_3 (red). Green depicts the second-order harmonic balance theory (B18-B20) using the same parameters as blue.

ω_{s1} to create a photon each at ω_{s2} and ω_{imd}. Thus, measuring the power in the sideband at ω_{imd} informs us about the strength of spurious 4-wave-mixing processes.

We can quantitatively relate this sideband power to the g_4 of the device. Including ω_{imd} in the harmonic balance analysis of the QLE (B1), we find that the Kerr term in the Hamiltonian acts as an effective drive at the respective IMD frequencies. For example, for the IMD product at $\omega_{\mathrm{imd}} = 2\omega_{s1} - \omega_{s2}$, the effective drive strength is equal to

$$u_{\mathrm{in(eff)}}^{\mathrm{imd}} = 12 g_4 \alpha_{s1}^2 \alpha_{s2}^*. \tag{B13}$$

Therefore, we can pretend that there is an input at the IMD frequencies, and the amplifier treats it as any other input signal, namely creates an amplified output $P_{\mathrm{out}}^{\mathrm{imd}} = GP_{\mathrm{in(eff)}}^{\mathrm{imd}}$ and the corresponding idler. For near-resonant tones,

$$P_{\mathrm{out}}^{\mathrm{imd}} = G\frac{\hbar\omega_a}{\kappa}|u_{\mathrm{in(eff)}}^{\mathrm{imd}}|^2 = G\frac{\hbar\omega_a}{\kappa}(12g_4)^2 n_{s1}^2 n_{s2}, \tag{B14}$$

where $n_{s1(2)} = |\alpha_{s1(2)}|^2$ is the average intra-resonator population, which can be related to the input and output powers by $n_{s1(2)}\hbar\omega_a\kappa = GP_{s1(2),\mathrm{in}} = P_{s1(2),\mathrm{out}}$ (for $G \gg 1$). Thus, we can relate the output power at ω_{imd} to the output powers at ω_{s1} and ω_{s2} by

$$P_{\mathrm{out}}^{\mathrm{imd}} = \frac{|12g_4|^2 G}{\kappa^4(\hbar\omega_a)^2} P_{s1,\mathrm{out}}^2 P_{s2,\mathrm{out}}. \tag{B15}$$

Assuming $P_{s2,\mathrm{in}} = P_{s1,\mathrm{in}} = P_{s,\mathrm{in}}$, we note that $P_{\mathrm{out}}^{\mathrm{imd}}$ scales cubically with applied input power, whereas $P_{s,\mathrm{out}} = GP_{s,\mathrm{in}}$ scales linearly (see Fig. 9a). Thus, as in Ref. 28, we can define the input-referred third-order intercept point IIP_3 as the input power where the low power asymptotes of $P_{s,\mathrm{out}}$ (dashed blue) and $P_{\mathrm{out}}^{\mathrm{imd}}$ (dashed orange) intersect. Using (B15) we can then obtain

$$IIP_3 = \frac{\kappa}{12|g_4|}\frac{1}{G^{3/2}}\hbar\omega_a\kappa. \tag{B16}$$

While the above derivation for brevity assumes $G \gg 1$, it can be shown that for arbitrary gain G the IIP_3 is given by

$$IIP_3 = \frac{\kappa}{12|g_4|}\left(\frac{\sqrt{G}-1}{G-1}\right)^3 \hbar\omega_a\kappa. \tag{B17}$$

This result can be used to extract the fourth-order nonlinearity g_4 in the presence of the amplification pump and compare it to the Stark shift experiment described in the main text (Fig. 4).

As shown in Fig. 9b, there is reasonably good agreement between g_4 measured by these independent experiments at fluxes away from the Kerr-free point. Ripples in the g_4 extracted from IIP_3 arise from variation of the impedance seen by the SPA (see discussion of Fig. 5 in main text). Strikingly, near the Kerr-free point the g_4 extracted from IIP_3 does not exhibit the reduction of the magnitude observed in the Stark shift measurement. We conjecture that this effect is related to the discrepancy between theory and experiment for $P_{-1\mathrm{dB}}$ seen in Fig. 5b. We explore this question further in the next section.

3. Comparison to experimental data

In this Section, we compare the predictions of the saturation power by the harmonic balance theory with the experimental data for device D, see Fig. 10.

Let us focus first on the blue first-order harmonic balance theory curve. To obtain it, we map the signal and pump powers onto the corresponding drive strengths u_s and u_p, solve the system of equations (B4-B6) numerically, and use the resulting amplitudes to calculate the gain (B9). In this numerical procedure, we use ω_a and κ extracted from the resonance fits at each flux point, and g_3 and g_4 calculated from first principles. As one can see from Fig 10, the agreement with data is not satisfactory. In particular, the theory predicts a sharp peak near the Kerr-free point, which is not observed in the data.

Similarly, from (B16) a peak near the Kerr-free point is expected in IIP_3, which is not observed in the data (Fig. 6b). This suggests trying to instead use g_4 extracted from IIP_3 (see Fig. 9b) in our numerical calculation to predict the compression power. The result of this is the red curve on Fig. 10, which shows a much better agreement with the data at all fluxes.

We attempt to explain the failure of the first-order harmonic balance theory near the Kerr-free point by going to higher order. We had previously included only the main frequency components ω_p, ω_s and $\omega_i = \omega_p - \omega_s$ in harmonic balance. The relevant frequencies to next order are $2\omega_p$, $2\omega_s$, $2\omega_i$, $\omega_p + \omega_s$, $\omega_p + \omega_i$, $\omega_s - \omega_i$ and 0. The corresponding intra-resonator amplitudes of these harmonics are suppressed by a factor of $\sim \kappa/\omega_a$ since they fall outside the linewidth of the SPA resonator. However, increasing the population at these frequencies can cause additional Stark shifts and contribute to the reduction of the saturation power. Using the smallness of the amplitudes of these harmonics, the new extended system of harmonic balance equations can be partially solved and reduced to three equations similar to (B4-B6) as shown below:

$$\Big(\omega_p - \omega_a + i\frac{2}{3}\kappa - g_4\big[\frac{32}{9}|\alpha_p|^2 + 16|\alpha_s|^2 + 16|\alpha_i|^2\big] + \frac{g_3^2}{\omega_c}\big[\frac{928}{45}|\alpha_p|^2 + 42|\alpha_s|^2 + 42|\alpha_i|^2\big]\Big)\alpha_p = u_p + 6g_3\alpha_i\alpha_s, \quad \text{(B18)}$$

$$\Big(\omega_s - \omega_a + i\frac{\kappa}{2} - g_4\big[\frac{32}{3}|\alpha_p|^2 + 12|\alpha_s|^2 + 12|\alpha_i|^2\big] + 4\frac{g_3^2}{\omega_c}\big[7|\alpha_p|^2 + 15|\alpha_s|^2 + 36|\alpha_i|^2\big]\Big)\alpha_s = u_s + 4g_3\alpha_p\alpha_i^*, \quad \text{(B19)}$$

$$\Big(\omega_i - \omega_a + i\frac{\kappa}{2} - g_4\big[\frac{32}{3}|\alpha_p|^2 + 12|\alpha_i|^2 + 12|\alpha_s|^2\big] + 4\frac{g_3^2}{\omega_c}\big[7|\alpha_p|^2 + 36|\alpha_s|^2 + 15|\alpha_i|^2\big]\Big)\alpha_i = u_i + 4g_3\alpha_p\alpha_s^*. \quad \text{(B20)}$$

Using this system of equations with the g_3 and g_4 calculated from first principles, we are able to reconstruct the green line on Fig. 10. We can see that the agreement is significantly improved compared to the first-order harmonic balance theory (blue), except in the small region close to half flux. This second-order harmonic balance theory is compared to data across multiple devices in Fig. 5b. The remaining discrepancy (tails in Fig. 5b) can possibly be explained by going to even higher order in harmonic balance, or by including higher order terms in the expansion of the Hamiltonian (A17).

Finally, we find that using the stiff-pump approximation for Stark-shift-limited saturation yields results similar to the full numerical solution. This further confirms that the effect of pump depletion is negligible and Kerr-induced Stark shifts are the primary mechanism responsible for amplifier saturation.

[1] R. Vijay, M. H. Devoret, and I. Siddiqi, *Invited Review Article: The Josephson bifurcation amplifier*, Review of Scientific Instruments **80**, 111101 (2009).

[2] A. Roy and M. Devoret, *Introduction to parametric amplification of quantum signals with Josephson circuits*, Comptes Rendus Physique **17**, 740 (2016).

[3] J. E. Johnson, C. Macklin, D. H. Slichter, R. Vijay, E. B. Weingarten, J. Clarke, and I. Siddiqi, *Heralded State Preparation in a Superconducting Qubit*, Phys. Rev. Lett. **109**, 050506 (2012).

[4] D. Risté, J. G. van Leeuwen, H.-S. Ku, K. W. Lehnert, and L. DiCarlo, *Initialization by Measurement of a Superconducting Quantum Bit Circuit*, Phys. Rev. Lett. **109**, 050507 (2012).

[5] M. Hatridge, S. Shankar, M. Mirrahimi, F. Schackert, K. Geerlings, T. Brecht, K. M. Sliwa, B. Abdo, L. Frunzio, S. M. Girvin, et al., *Quantum Back-Action of an Individual Variable-Strength Measurement*, Science **339**, 178 (2013).

[6] E. Jeffrey, D. Sank, J. Y. Mutus, T. C. White, J. Kelly, R. Barends, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, et al., *Fast Accurate State Measurement with Superconducting Qubits*, Phys. Rev. Lett. **112**, 190504 (2014).

[7] T. Walter, P. Kurpiers, S. Gasparinetti, P. Magnard, A. Potočnik, Y. Salathé, M. Pechal, M. Mondal, M. Oppliger, C. Eichler, et al., *Rapid High-*

Fidelity Single-Shot Dispersive Readout of Superconducting Qubits, Phys. Rev. Applied **7**, 054020 (2017).

[8] A. Bienfait, J. J. Pla, Y. Kubo, M. Stern, X. Zhou, C. C. Lo, C. D. Weis, T. Schenkel, M. L. W. Thewalt, D. Vion, et al., *Reaching the quantum limit of sensitivity in electron spin resonance*, Nature Nanotechnology **11**, 253 (2016).

[9] A. Bienfait, P. Campagne-Ibarcq, A. H. Kiilerich, X. Zhou, S. Probst, J. J. Pla, T. Schenkel, D. Vion, D. Esteve, J. J. L. Morton, et al., *Magnetic Resonance with Squeezed Microwaves*, Phys. Rev. X **7**, 041011 (2017).

[10] B. M. Brubaker, L. Zhong, Y. V. Gurevich, S. B. Cahn, S. K. Lamoreaux, M. Simanovskaia, J. R. Root, S. M. Lewis, S. Al Kenany, K. M. Backes, et al., *First Results from a Microwave Cavity Axion Search at* 24 μeV, Phys. Rev. Lett. **118**, 061302 (2017).

[11] S. B. Bravyi and A. Y. Kitaev, *Quantum codes on a lattice with boundary*, arXiv:quant-ph/9811052 (1998), arXiv: quant-ph/9811052.

[12] N. E. Frattini, U. Vool, S. Shankar, A. Narla, K. M. Sliwa, and M. H. Devoret, *3-wave mixing Josephson dipole element*, Appl. Phys. Lett. **110**, 222603 (2017).

[13] A. B. Zorin, *Josephson Traveling-Wave Parametric Amplifier with Three-Wave Mixing*, Phys. Rev. Applied **6**, 034006 (2016).

[14] C. Eichler and A. Wallraff, *Controlling the dynamic range of a Josephson parametric amplifier*, EPJ Quantum Technology **1**, 2 (2014).

[15] G. Liu, T.-C. Chien, X. Cao, O. Lanes, E. Alpern, D. Pekker, and M. Hatridge, *Josephson parametric converter saturation and higher order effects*, Appl. Phys. Lett. **111**, 202603 (2017).

[16] J. Y. Mutus, T. C. White, R. Barends, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, E. Jeffrey, J. Kelly, A. Megrant, et al., *Strong environmental coupling in a Josephson parametric amplifier*, Appl. Phys. Lett. **104**, 263513 (2014).

[17] C. Eichler, Y. Salathe, J. Mlynek, S. Schmidt, and A. Wallraff, *Quantum-Limited Amplification and Entanglement in Coupled Nonlinear Resonators*, Phys. Rev. Lett. **113**, 110502 (2014).

[18] T. Roy, S. Kundu, M. Chand, A. M. Vadiraj, A. Ranadive, N. Nehra, M. P. Patankar, J. Aumentado, A. A. Clerk, and R. Vijay, *Broadband parametric amplification with impedance engineering: Beyond the gain-bandwidth product*, Appl. Phys. Lett. **107**, 262601 (2015).

[19] M. A. Castellanos-Beltran, K. D. Irwin, G. C. Hilton, L. R. Vale, and K. W. Lehnert, *Amplification and squeezing of quantum noise with a tunable Josephson metamaterial*, Nat Phys **4**, 929 (2008).

[20] C. Eichler, D. Bozyigit, C. Lang, M. Baur, L. Steffen, J. M. Fink, S. Filipp, and A. Wallraff, *Observation of Two-Mode Squeezing in the Microwave Frequency Domain*, Phys. Rev. Lett. **107**, 113601 (2011).

[21] Note1, pNA-X network analyzer model Keysight N5242A.

[22] A. A. Clerk, M. H. Devoret, S. M. Girvin, F. Marquardt, and R. J. Schoelkopf, *Introduction to Quantum Noise, Measurement and Amplification*, Reviews of Modern Physics **82**, 1155 (2010), arXiv: 0810.4729.

[23] T. Yamamoto, K. Inomata, M. Watanabe, K. Matsuba, T. Miyazaki, W. D. Oliver, Y. Nakamura, and J. S. Tsai, *Flux-driven Josephson parametric amplifier*, Appl. Phys. Lett. **93**, 042510 (2008).

[24] M. Hatridge, R. Vijay, D. H. Slichter, J. Clarke, and I. Siddiqi, *Dispersive magnetometry with a quantum limited SQUID parametric amplifier*, Phys. Rev. B **83**, 134501 (2011).

[25] J. Clarke and A. I. Braginski, eds., *The SQUID Handbook*, vol. 1 (Wiley, Weinheim, 2004).

[26] A. Kamal, A. Marblestone, and M. Devoret, *Signal-to-pump back action and self-oscillation in double-pump Josephson parametric amplifier*, Phys. Rev. B **79**, 184301 (2009).

[27] A. Roy and M. Devoret, *Quantum-limited parametric amplification with Josephson circuits in the regime of pump depletion*, Phys. Rev. B **98**, 045405 (2018).

[28] D. M. Pozar, *Microwave Engineering* (John Wiley & Sons, Inc., 2012), 4th ed.

[29] N. Bergeal, F. Schackert, M. Metcalfe, R. Vijay, V. E. Manucharyan, L. Frunzio, D. E. Prober, R. J. Schoelkopf, S. M. Girvin, and M. H. Devoret, *Phase-preserving amplification near the quantum limit with a Josephson ring modulator*, Nature **465**, 64 (2010).

[30] S. Touzard, A. Kou, N. E. Frattini, V. V. Sivak, S. Puri, A. Grimm, L. Frunzio, S. Shankar, and M. H. Devoret, *Gated conditional displacement readout of superconducting qubits*, arXiv:1809.06964 [quant-ph] (2018), arXiv: 1809.06964.

[31] O. Naaman, D. G. Ferguson, and R. J. Epstein, *High Saturation Power Josephson Parametric Amplifier with GHz Bandwidth*, arXiv:1711.07549 [cond-mat, physics:physics, physics:quant-ph] (2017), arXiv: 1711.07549.

[32] B. Abdo, K. Sliwa, L. Frunzio, and M. Devoret, *Directional Amplification with a Josephson Circuit*, Phys. Rev. X **3**, 031001 (2013).

[33] K. M. Sliwa, M. Hatridge, A. Narla, S. Shankar, L. Frunzio, R. J. Schoelkopf, and M. H. Devoret, *Reconfigurable Josephson Circulator/Directional Amplifier*, Phys. Rev. X **5**, 041020 (2015).

[34] F. Lecocq, L. Ranzani, G. A. Peterson, K. Cicak, R. W. Simmonds, J. D. Teufel, and J. Aumentado, *Nonreciprocal Microwave Signal Processing with a Field-Programmable Josephson Amplifier*, Phys. Rev. Applied **7**, 024028 (2017).

[35] A. Metelmann and A. A. Clerk, *Quantum-Limited Amplification via Reservoir Engineering*, Phys. Rev. Lett. **112**, 133904 (2014).

[36] A. Metelmann and A. A. Clerk, *Nonreciprocal Photon Transmission and Amplification via Reservoir Engineering*, Phys. Rev. X **5**, 021025 (2015).

[37] L. Ranzani and J. Aumentado, *Graph-based analysis of nonreciprocity in coupled-mode systems*, New J. Phys. **17**, 023024 (2015).

[38] C. Macklin, K. O'Brien, D. Hover, M. E. Schwartz, V. Bolkhovsky, X. Zhang, W. D. Oliver, and I. Siddiqi, *A near–quantum-limited Josephson traveling-wave parametric amplifier*, Science **350**, 307 (2015).

[39] T. C. White, J. Y. Mutus, I.-C. Hoi, R. Barends, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, E. Jeffrey, et al., *Traveling wave parametric amplifier with Josephson junctions using minimal resonator phase matching*, Appl. Phys. Lett. **106**, 242601 (2015).

[40] M. R. Vissers, R. P. Erickson, H.-S. Ku, L. Vale, X. Wu, G. C. Hilton, and D. P. Pappas, *Low-noise kinetic inductance traveling-wave amplifier using three-wave mixing*, Appl. Phys. Lett. **108**, 012601 (2016).

[41] M. T. Bell and A. Samolov, *Traveling-Wave Parametric Amplifier Based on a Chain of Coupled Asymmetric SQUIDs*, Phys. Rev. Applied **4**, 024014 (2015).

[42] W. Zhang, W. Huang, M. E. Gershenson, and M. T. Bell, *Josephson Metamaterial with a Widely Tunable Positive or Negative Kerr Constant*, Phys. Rev. Applied **8**, 051001 (2017).

[43] S. O. Mundhada, A. Grimm, S. Touzard, U. Vool, S. Shankar, M. H. Devoret, and M. Mirrahimi, *Generating higher-order quantum dissipation from lower-order parametric processes*, Quantum Sci. Technol. **2**, 024005 (2017).

[44] S. Puri, S. Boutin, and A. Blais, *Engineering the quantum states of light in a Kerr-nonlinear resonator by two-photon driving*, npj Quantum Information **3**, 18 (2017).

[45] O. Naaman, M. O. Abutaleb, C. Kirby, and M. Rennie, *On-chip Josephson junction microwave switch*, Appl. Phys. Lett. **108**, 112601 (2016).

[46] B. J. Chapman, E. I. Rosenthal, J. Kerckhoff, B. A. Moores, L. R. Vale, J. A. B. Mates, G. C. Hilton, K. Lalumière, A. Blais, and K. W. Lehnert, *Widely Tunable On-Chip Microwave Circulator for Superconducting Quantum Circuits*, Phys. Rev. X **7**, 041043 (2017).

[47] B. Abdo, K. Sliwa, F. Schackert, N. Bergeal, M. Hatridge, L. Frunzio, A. D. Stone, and M. Devoret, *Full Coherent Frequency Conversion between Two Propagating Microwave Modes*, Phys. Rev. Lett. **110**, 173902 (2013).

[48] M. S. Allman, J. D. Whittaker, M. Castellanos-Beltran, K. Cicak, F. da Silva, M. P. DeFeo, F. Lecocq, A. Sirois, J. D. Teufel, J. Aumentado, et al., *Tunable Resonant and Nonresonant Interactions between a Phase Qubit and LC Resonator*, Phys. Rev. Lett. **112**, 123601 (2014).

[49] E. Flurin, N. Roch, J. D. Pillet, F. Mallet, and B. Huard, *Superconducting Quantum Node for Entanglement and Storage of Microwave Radiation*, Phys. Rev. Lett. **114**, 090503 (2015).

[50] W. Pfaff, C. J. Axline, L. D. Burkhart, U. Vool, P. Reinhold, L. Frunzio, L. Jiang, M. H. Devoret, and R. J. Schoelkopf, *Controlled release of multiphoton quantum states from a microwave cavity memory*, Nature Physics **13**, 882 (2017).

[51] C. J. Axline, L. D. Burkhart, W. Pfaff, M. Zhang, K. Chou, P. Campagne-Ibarcq, P. Reinhold, L. Frunzio, S. M. Girvin, L. Jiang, et al., *On-demand quantum state transfer and entanglement between remote microwave cavity memories*, Nature Physics p. 1 (2018).

[52] P. Kurpiers, P. Magnard, T. Walter, B. Royer, M. Pechal, J. Heinsoo, Y. Salathé, A. Akin, S. Storz, J.-C. Besse, et al., *Deterministic quantum state transfer and remote entanglement using microwave photons*, Nature **558**, 264 (2018).

[53] P. Campagne-Ibarcq, E. Zalys-Geller, A. Narla, S. Shankar, P. Reinhold, L. Burkhart, C. Axline, W. Pfaff, L. Frunzio, R. J. Schoelkopf, et al., *Deterministic Remote Entanglement of Superconducting Circuits through Microwave Two-Photon Transitions*, Phys. Rev. Lett. **120**, 200501 (2018).

[54] Z. Leghtas, S. Touzard, I. M. Pop, A. Kou, B. Vlastakis, A. Petrenko, K. M. Sliwa, A. Narla, S. Shankar, M. J. Hatridge, et al., *Confining the state of light to a quantum manifold by engineered two-photon loss*, Science **347**, 853 (2015).

[55] U. Vool, A. Kou, W. C. Smith, N. E. Frattini, K. Serniak, P. Reinhold, I. M. Pop, S. Shankar, L. Frunzio, S. M. Girvin, et al., *Driving Forbidden Transitions in the Fluxonium Artificial Atom*, Phys. Rev. Applied **9**, 054046 (2018).

[56] N. A. Masluk, I. M. Pop, A. Kamal, Z. K. Minev, and M. H. Devoret, *Microwave Characterization of Josephson Junction Arrays: Implementing a Low Loss Superinductance*, Phys. Rev. Lett. **109**, 137002 (2012).

[57] J. Koch, T. M. Yu, J. Gambetta, A. A. Houck, D. I. Schuster, J. Majer, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, *Charge-insensitive qubit design derived from the Cooper pair box*, Phys. Rev. A **76**, 042319 (2007).

[58] F. Schackert, Ph.D., Yale University (2013).

[59] E. Flurin, Ph.D. thesis, École Normale Supérieure (2015).

[60] A. Altland and B. D. Simons, *Condensed Matter Field Theory* (Cambridge University Press, Cambridge, 2010).

[61] M. Wallquist, V. S. Shumeiko, and G. Wendin, *Selective coupling of superconducting charge qubits mediated by a tunable stripline cavity*, Phys. Rev. B **74**, 224506 (2006).

[62] S. E. Nigg, H. Paik, B. Vlastakis, G. Kirchmair, S. Shankar, L. Frunzio, M. H. Devoret, R. J. Schoelkopf, and S. M. Girvin, *Black-Box Superconducting Circuit Quantization*, Phys. Rev. Lett. **108**, 240502 (2012).

[63] K. M. Sundqvist and P. Delsing, *Negative-resistance models for parametrically flux-pumped superconducting quantum interference devices*, EPJ Quantum Technology **1**, 6 (2014).