


Phonon-Trapping-Enhanced Energy Resolution in Superconducting Single-Photon Detectors

Pieter J. de Visser^{1,*}, Steven A.H. de Rooij^{1,2}, Vignesh Murugesan,¹ David J. Thoen^{2,3} and Jochem J.A. Baselmans^{1,2}

¹*SRON Netherlands Institute for Space Research, Niels Bohrweg 4, Leiden 2333 CA, Netherlands*

²*Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Mekelweg 4, Delft 2628 CD, Netherlands*

³*Faculty of Applied Sciences, Delft University of Technology, Kavli Institute of NanoScience, Lorentzweg 1, Delft 2628CJ, Netherlands*

 (Received 12 March 2021; revised 30 March 2021; accepted 17 August 2021; published 30 September 2021)

A noiseless, photon-counting detector, which resolves the energy of each photon, could radically change astronomy, biophysics, and quantum optics. Superconducting detectors promise an intrinsic resolving power at visible wavelengths of $R = E/\delta E \approx 100$ due to their low excitation energy. We study superconducting energy-resolving microwave kinetic inductance detectors (MKIDs), which hold particular promise for larger cameras. A visible and near-infrared photon absorbed in the superconductor creates a few thousand quasiparticles through several stages of electron-phonon interaction. Here we demonstrate experimentally that the resolving power of MKIDs at visible to near-infrared wavelengths is limited by the loss of hot phonons during this process. We measure the resolving power of our aluminum-based detector as a function of photon energy using four lasers with wavelengths between 1545–402 nm. For detectors on thick SiN/Si and sapphire substrates the resolving power is limited to 10–21 for the respective wavelengths, consistent with the loss of hot phonons. When we suspend the sensitive part of the detector on a 110-nm-thick SiN membrane, the measured resolving power improves to 19–52, respectively. The improvement is equivalent to a factor 8 ± 2 stronger phonon trapping on the membrane, which is consistent with a geometrical phonon propagation model for these hot phonons. We discuss a route towards the Fano limit by phonon engineering.

DOI: [10.1103/PhysRevApplied.16.034051](https://doi.org/10.1103/PhysRevApplied.16.034051)

I. INTRODUCTION

A camera in which each pixel can count individual photons with high efficiency and which simultaneously resolves each photon's color is a dream for many (bio)physicists, astronomers, and anybody working in low light applications. In semiconductor-based detectors, each visible or near-infrared photon creates only one or a few excitations due to the large band gap. Therefore, energy information about absorbed photons is lost. Dark excitations are indistinguishable from photon excitations, which causes the well-known dark current and read noise that limit the detector sensitivity. In a superconductor, the energy needed to create excitations can be orders of magnitude less than in a semiconductor. A single visible photon can break a few thousand Cooper pairs. A measurement of the number of broken pairs is therefore a direct measure of the photon energy. Microwave kinetic inductance

detectors (MKIDs) [1,2] measure the number of quasiparticles excited in the superconductor through a change in the complex conductivity. The change in kinetic inductance and resistance of the superconductor are measured with a microwave resonator. Microwave resonators can be naturally multiplexed into large arrays. Several visible and near-infrared MKID instruments are deployed at telescopes, with up to 20,000 pixels [3–5]. Other energy-resolving detectors for these wavelengths are transition edge sensors (TES) and superconducting tunnel junction detectors (STJ), which are harder to multiplex into arrays, although small arrays have been demonstrated in astronomical [6,7] and biological [8] applications.

One of the main characteristics of photon-counting MKIDs, the energy-resolving power, is poorly understood. The resolving power can be limited by different mechanisms, such as the signal to noise, quasiparticle diffusion, nonuniformity of the microwave current, and phonon losses. Ultimately, the resolving power is limited by the statistical variation in the number of quasiparticles generated per photon. This limit arises because the initial photon

*p.j.de.visser@sron.nl

energy (a few eV) is downconverted into quasiparticles with a typical energy of 0.1–0.2 meV through a number of electron-electron and electron-phonon interaction steps. When all photon energy is kept inside the detector during this predetection stage, the resolving power is described by Fano statistics [9,10], and given by

$$R_{\text{Fano}} = \frac{1}{2\sqrt{2 \ln 2}} \sqrt{\frac{\eta_{pb}^{\text{max}} E}{\Delta F}}, \quad (1)$$

with E the photon energy, $\Delta = 1.76k_B T_c$ the gap energy of the superconductor, and T_c the critical temperature. η_{pb} expresses that not all energy can be converted to new quasiparticles. Even when all photon energy stays inside the superconductor during down-conversion, $\eta_{pb} = \eta_{pb}^{\text{max}} = 0.59$, which is equivalent to an average energy of 1.7Δ per generated quasiparticle [11]. $F \approx 0.2$ is the Fano factor for superconductors [9,10]. For aluminum with $T_c = 1.25$ K, $R_{\text{Fano}} = 47\text{--}92$ for wavelengths of 1550–400 nm, respectively. Throughout this paper we use the term resolving power for $R = E/\delta E$, with δE the FWHM energy resolution, expressed in energy units.

Energy-resolution measurements on MKIDs have been reported for different materials. For TiN (with $T_c = 0.8$ K), $R \approx 10$ at 400 nm [3,12]. For PtSi, with $T_c = 0.9$ K, $R = 5.8 - 8.1$ at wavelengths of 1310–808 nm (a best $\delta E = 0.16$ eV) [13]. For Hf with $T_c = 0.4$ K, $R = 6.8 - 9.4$ (best $\delta E = 0.14$ eV) for the same wavelengths [14]. In devices optimized for photon-number resolution using TiN/Ti/TiN trilayers $R = 3.6$ at 1550 nm [15]. All of these results are partially due to noise limitations and partially due to unknown reasons. Care was taken to not be limited by current nonuniformity [3,16]. The best resolving powers of $R = 9.6 - 17$ for wavelengths of 1310–406 nm (best $\delta E = 0.10$ eV) were recently reported using a parametric amplifier to reduce the amplifier noise [16,17]. In Ref. [16], the signal-to-noise limit was $R_{\text{SN}} = 24$ at 1120 nm, which clearly shows that another mechanism dominates R , possibly phonon losses. These numbers show that the Fano limit is still far out of reach for MKIDs. On top of that, to study the intrinsic resolving-power limits, a much higher signal to noise is desirable.

TES detectors have shown resolving powers of 11–23 at 780–350 nm [18] (a constant $\delta E = 0.15$ eV) and 7 at 1570 nm [19]. TES have proven very powerful in counting the number of photons in pulses [19–21], for which a moderate resolving power suffices. The zero-noise property has enabled several notable experimental milestones in quantum physics [22–24]. For STJs, which are also pair-breaking detectors, the resolving power has been studied in much more depth, resulting in a measured resolving power of 12–25 at wavelengths of 1500–400 nm (best $\delta E = 0.07$ eV) [25]. This corresponds to $R \approx R_{\text{Fano}}/2$ in Ta, which has a relatively high $T_c \approx 4.4$ K. This resolving power was shown to be partially due to tunneling noise, which is specific to STJs, and partially due to the loss of hot phonons. These hot phonons are lost during the down-conversion process of the initial photon energy to the detected quasiparticles excitations, which involves several stages of electron-electron and electron-phonon interaction. The down-conversion process was modeled in detail [26–28] and verified experimentally in Ref. [25]. To what extent the phonon flow can be impeded during the energy down-conversion and whether that indeed improves the resolving power has not been experimentally addressed.

Here we demonstrate that the resolving power of MKIDs at visible to near-infrared wavelengths is indeed limited by hot phonon losses due to the down-conversion process after photon absorption. We use a hybrid (Nb, Ti)N-Al MKID with aluminum as the sensitive material, which is optimized for this resolving-power study. We use a geometry for which we have studied the quasiparticle physics in detail and for which the sensitivity is high enough not to be limited by signal to noise [29–31]. We measure the resolving power as a function of photon energy by exposing the detector to continuous illumination of four lasers with wavelengths between 1545–402 nm. For devices on a SiN/Si substrate and devices on a sapphire substrate, the resolving power is limited to 10–21 (best $\delta E = 0.08$ eV) for the respective wavelengths, which is consistent with the loss of hot phonons. These numbers are in the same ballpark as the best reported MKID, TES, and STJ devices on substrates. When we suspend the sensitive part of the detector on a 110-nm-thick SiN membrane, the measured

TABLE I. Overview of the measured devices. Devices *A* and *B* have eight resonators on a chip. Devices *C* and *D* are fabricated on the same chip. We explicitly verify that the Al T_c and ρ are the same on the SiN membrane and substrate. The Al linewidths are measured from SEM images of each chip. For device *A*, the cofabricated dc structure had a design error. The data shown here is measured on a device with the same geometry, Al thickness, substrate and fabrication steps.

Device	Substrate	Al thickness	Al width	Resonator type	T_c (K)	ρ ($\mu\Omega$ cm)
<i>A</i>	Sapphire	50 nm	0.9 μm	CPW	1.25*	0.9*
<i>B</i>	Sapphire	150 nm	0.7 μm	CPW	1.12	0.4
<i>C</i>	SiN on Si	50 nm	1.7 μm	IDC	1.25	1
<i>D</i>	110 nm SiN membrane	50 nm	1.7 μm	IDC	1.25	1

resolving power improves to 19–52 for 1545–402 nm (best $\delta E = 0.04$ eV), due to the strong phonon trapping. We discuss a route towards reaching the Fano limit by further reducing the phonon flow.

The main results of this work are summarized in Fig. 3(c) and Fig. 4. In Fig. 4(a) we directly compare the measured resolving power for devices on a substrate and on a membrane.

II. DETECTOR DESIGNS

We use hybrid (Nb, Ti)N-Al resonators as single-photon detectors. The devices used in this work are hybrid (Nb, Ti)N-Al resonators [30] and their properties are listed in Table I. We use either 350- μm -thick *C*-plane sapphire substrates (for devices *A* and *B*) or 150-nm LPCVD grown SiN on 350 μm Si (for devices *C* and *D*). Sapphire and SiN are isolators with a large enough band gap to avoid the excitation of charge carriers with 402-nm photons. We use two different designs, one based on the conventional coplanar waveguide (CPW) geometry and a design where a large part of the CPW is replaced by an interdigitated capacitor (IDC) [32]. Resonators are coupled capacitively to the microwave readout line. Micrographs of the devices are shown in Fig. 1(a), 1(b), 1(d), and 1(e), respectively. In both cases a large part of the resonator is made out of wide (Nb, Ti)N structures, which strongly reduces two-level system (TLS) noise due to the favorable properties of (Nb, Ti)N [33] and the width of the structures [34,35]. (Nb, Ti)N is deposited using reactive magnetron sputtering in an argon-nitrogen plasma [36,37], with a thickness of 200 nm for devices on SiN and 300 nm for devices on sapphire. Near the shorted end of the quarter-wave resonator we put a short section of sputter deposited Al, which is the sensitive part. The ground planes of the readout line are balanced using aluminium bridges on top of a polyimide support. The T_c of the (Nb, Ti)N layer varies somewhat between wafers, and is typically 15 K. The SiN membranes are defined with a potassium hydroxide (KOH) etch from the backside of the wafer, after finishing the layers on the front side. After the KOH etch the wafer is cleaned with RCA2. The membrane of device *D* ends up as 110 nm thick, due to overetching with SF6 when patterning the (Nb, Ti)N layer. For devices *C* and *D* the measured Al line width is 1.7 μm . For phonon trapping in device *D* we rely on the geometrical effect of phonon retrapping due to the 1.7/0.110 aspect ratio of the line width and the membrane thickness. Phonons that leave the Al film have a high chance to reflect on the bottom of the thin membrane, and scatter back into the Al film (see also Appendix A). We analyze the resulting improvement in resolving power in Sec. VII.

For the membrane-based device we find that the number of measured photon hits is strongly dominated by hits in the (Nb, Ti)N ground-plane strip, when that (Nb, Ti)N strip

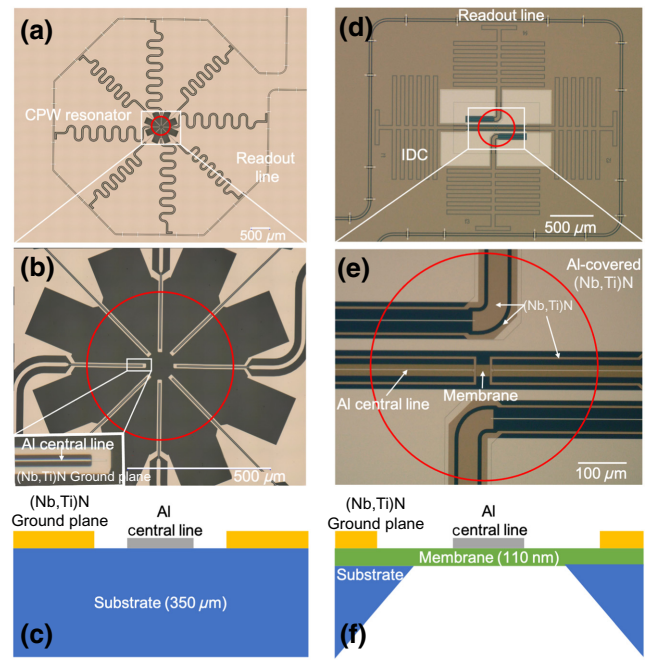


FIG. 1. (a),(b) Micrographs of device *A*, with eight CPW resonators, which are capacitively coupled to the readout line. The red circle indicates the size and position of the 500- μm -diameter aperture, which is mounted in front of the sample and therefore limits the area that is illuminated by laser light. Only near the shorted end of the resonator there is a small section of central line made out of Al (260 or 530 μm long), which is the sensitive part of the resonator and points towards the center of the chip. In the center of the chip most of the (Nb, Ti)N ground plane is cut away to avoid absorption of light in there. (c) Schematic cross section of the sensitive shorted end of the detector, a CPW section with (Nb, Ti)N ground planes and Al central line on a thick sapphire substrate (not to scale). (d),(e) Micrographs of devices *C* on the substrate (top and bottom) and *D* on the membrane (middle two), which have a large interdigitated capacitor (IDC) to enhance the responsivity. The CPW end section is particularly wide to fit the whole membrane width inside. (f) Schematic cross section of the sensitive shorted end of the detector (not to scale). The 1.7- μm -wide Al strip is fabricated on a 110-nm-thick membrane to allow phonons to be retrapped into the film before escaping to the bath. The (Nb, Ti)N ground planes are on the substrate outside the membrane area, to avoid photons absorbed in the (Nb, Ti)N ground plane to generate a measurable pulse in the Al strip through phonon transport.

is also placed on the membrane. Therefore, we make the end section of these resonators much wider, to only have the central Al strip on the membrane [Fig. 1(f)]. The reference detector *C* on the substrate has the same geometry to allow a direct quantitative comparison. For these wide sections, we find that the signal to noise is degraded about $3\times$ compared to a narrow CPW [Fig. 1(b)], which is compensated by the use of the IDC in these designs. As shown in Sec. VI, the designs indeed have sufficient signal to noise to study the intrinsic limits to the resolving power.

We choose these designs to study the resolving power and the effect of phonon losses in a device that we understand well. The downside of this geometry, together with the choice for Al as the active material, is that the absorption efficiency of photons (or quantum efficiency) of the devices is low. The efficiency is limited by the Al fill fraction in the aperture and the Al surface impedance at the measured wavelengths.

The design excludes two possible limitations to the energy resolution. Quasiparticle diffusion out of the sensitive part of the detector could cause pulse height variations. $\Delta_{(\text{Nb,Ti})\text{N}}/\Delta_{\text{Al}} \approx 12$, which prevents the quasiparticles generated in the Al to diffuse out of the Al volume during the single-photon pulse. Secondly microwave current density nonuniformity in the detector could lead to a different pulse height for the same photon energy when absorbed in different locations [3,16]. The current density in the Al part of our resonators is uniform to 99%. Together with quasiparticle diffusion *inside* the Al volume, which leads to a uniform quasiparticle density, we expect current nonuniformity not to play a role for $R < 100$.

To limit absorption of radiation in unwanted areas on the chip we place a 500- μm -diameter aperture on top of the chips, as indicated with a red circle in Fig. 1. Since (Nb, Ti)N absorbs a lot more visible and near-infrared light than Al, which is unwanted, we cut away most of the (Nb, Ti)N ground plane inside the aperture for devices *A* and *B*. For devices *C* and *D* we cover most of the ground plane with Al, since absorption in the Si substrate (with a 1.2-eV band gap) is also undesirable.

III. EXPERIMENTAL SETUP

The samples are cooled in a pulse-tubed precooled adiabatic demagnetization refrigerator. Measurements are carried out at a temperature of 120 mK, unless otherwise indicated. The sample stage is carefully shielded from stray light from the 3-K stage of the cooler, using a box-in-box concept with coax cable filters in the outer box [29,38]. The system is schematically depicted in Fig. 2(a). Laser light is coupled into the cryostat with an optical fiber (Thorlabs SMF-28-J9, with a 8.2- μm -diameter core and 125- μm diameter cladding, which is single moded at 1550 nm). The fiber is thermalized at 3 K with eight windings on a copper spool with a 32-mm diameter. The fiber ends at the 100-mK box as a bare cleaved end, which is mounted with a custom copper clamp, to avoid standard commercial parts inside the magnetic shields, which are typically made from stainless steel and therefore magnetic. The only possible entry for (stray) light into the 100-mK box is therefore through the fiber. Light enters the light tight box through a fused silica engineered diffuser (RPC Photonics), which spreads out the narrow beam from the fiber over an angle of 32° , and also compensates for multimode patterns,

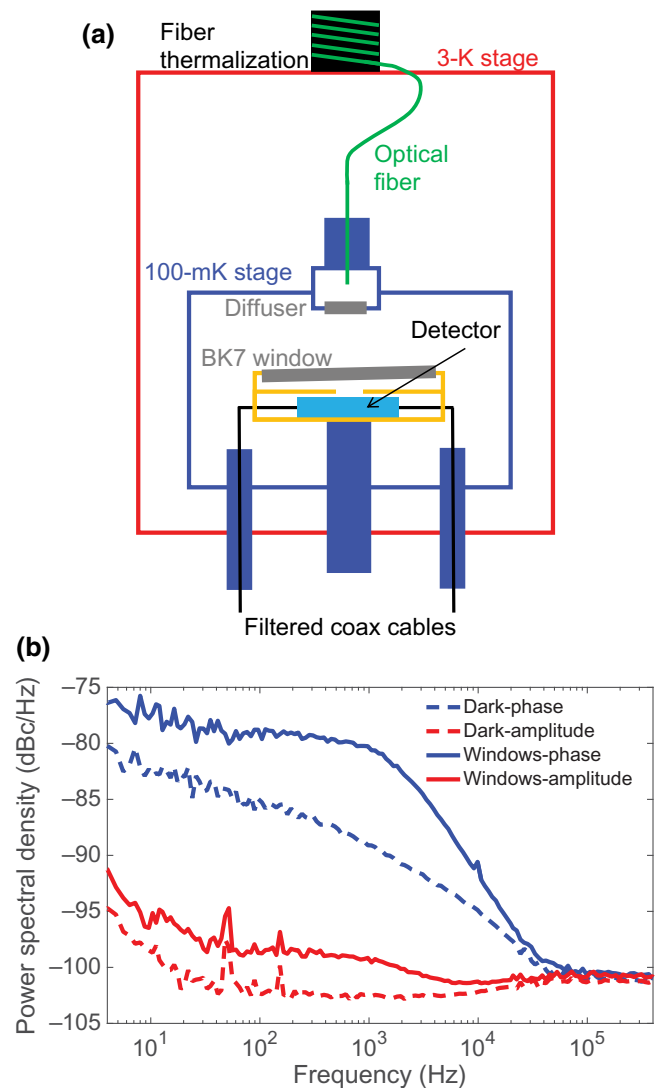


FIG. 2. (a) Schematic of the inner parts of the cryostat, showing the box-in-box concept where the only hole at 100 mK is made for the optical fiber. A detailed description can be found in the text. (b) Noise spectra of the phase and amplitude response of a CPW MKID in a dark environment and in an environment with open (BK7 and fused silica) windows to the 3-K stage of the setup. Both measurements are performed without laser illumination.

which can occur for the shorter wavelengths. At the sample box we mount a 5-mm-thick BK7 glass window. Right above the sample we place a 500- μm -diameter aperture as described above. Outside the aperture light is absorbed using black, carbon-loaded epoxy. The samples are front illuminated.

Microwave measurements are performed with a standard homodyne detection scheme. The signal from the microwave generator is first attenuated at every temperature stage, sent through the sample and amplified with a low noise HEMT amplifier at 3 K and with a room-temperature amplifier. The noise floor we observe in the

measurement corresponds to an effective system noise temperature of about 5 K. The output is mixed with the original signal using an IQ mixer. A detailed overview of all components can be found in Ref. [39].

In an early experiment of this kind, the optical fiber was not mounted at the lowest temperature stage, but ended at the 3-K stage of the cooler. The detector, illuminated through the same fused silica diffuser and BK7 glass window (both at 100 mK), was therefore also exposed to the 3-K environment. In this configuration we find a strong increase in the noise, compared to a dark measurement with the sample box closed with a copper lid, as shown in Fig. 2(b). This demonstrates that the thermal radiation of the 3-K environment is not filtered by the fused silica and BK7 glass windows and is visible as photon noise in the detector response. Although BK7 is one of the best possible glass filters for wavelengths from near-infrared to submillimeter, it is known to be transparent for frequencies lower than approximately 500 GHz [40]. The Al detector is sensitive to any light with frequencies higher than its gap frequency of 92 GHz. For the experiments reported here we solve this by mounting the optical fiber directly to the 100-mK stage as described above. For imaging experiments with these detectors through a window, a low-frequency filter will be indispensable to preserve the resolving power.

IV. MEASUREMENT PROCEDURE

For every new measurement condition, we first perform an S_{21} frequency sweep to find the resonant frequency and to measure the resonance circle as a reference for the MKID response. At the resonant frequency we first perform a noise measurement with the lasers off at 1 Msample/s. For the resolving-power measurements, we use four room temperature, fiber-coupled diode lasers (Thorlabs) with wavelengths 1545, 986, 673, and 402 nm, with output powers between 1–10 mW in continuous-wave mode and a linewidth corresponding to $R > 300$. To enable single-photon absorption at a rate of approximately 100 Hz, around 80 dB of attenuation is used before entering the cryostat. Note that also laser power is lost due to the inefficient absorption as explained above and because the illumination spot after the diffuser is larger than the aperture. We take 40 s of data for devices A and B and 160 s for devices C and D , sampled at 1 Msample/s, which aims at a few thousand single-photon events per wavelength.

Single-photon pulses are processed with a standard optimal filter, applied to the MKID phase response, θ , which gives the highest signal to noise for our devices. The optimal filter is applied in the frequency domain, to extract the optimal pulse amplitude A from the measurement [41]. We consider $\theta(f) = AM(f) + N(f)$, with $\theta(f)$ the Fourier transform of the measured pulse, $M(f)$ the pulse model, normalized to 1, and $N(f)$ the noise. We can estimate A

using [41]

$$A = \int_{-\infty}^{\infty} \frac{\theta(f)M^*(f)}{|N(f)|^2} df \bigg/ \int_{-\infty}^{\infty} \frac{|M(f)|^2}{|N(f)|^2} df, \quad (2)$$

where the bandwidth is limited at low frequencies by the duration of the pulses, which we take 3 ms, including 0.5 ms of prepulse data, and at high frequencies by the sampling rate. Time traces that have multiple pulses (on top of each other) within this time window are left out of the analysis. The pulse component at $\theta(f = 0)$ is left out of the analysis since it does not improve, or even deteriorates the resolving power. The pulse template $M(f)$ is calculated for each measurement condition as the average of the detected pulses, where we iterate once to use only pulses that have a pulse amplitude within 3σ around the maximum in the histogram. Typical average pulses are shown in Fig. 3(a) for all wavelengths for devices C and D . The noise power spectral density $|N(f)|^2$ is calculated from the time stream with lasers off. We average the spectrum of thousand samples with the same length as the pulses. Care is taken to filter out large energy events that also occur with lasers off, due to, e.g., cosmic rays [42,43], before calculating $|N(f)|^2$. The average pulse and noise spectra are shown for device C in Fig. 3(b). In the plot we distinguish between the spectrum of the average pulse (averaging the time-domain pulses before taking the spectrum) and the average spectrum of pulses (averaging the spectra of each pulse) to check that the noise floor in the pulse measurements is the same as in the noise measurement. The “average residual” represents a spectrum of the time traces where pulses are detected, after subtracting the optimally filtered pulse, $\theta(f) - AM(f)$, to check the consistency of the analysis.

In the analysis we distinguish several aspects of the resolving power, expressed by different symbols. R is the measured resolving power of the detector. A typical histogram for 673-nm illumination of devices C and D is shown in Fig. 3(c), which already illustrates the large difference in R for substrate- and membrane-based devices, to which we come back to in Sec. VI. R is calculated as the FWHM of a kernel density estimate, shown as the lines in Fig. 3(c), where we choose the kernel bandwidth to be the same as the bin size of the histogram. In Appendix B, we show histograms for all wavelengths and for the whole energy range. In particular, for the membrane device we also observe low-energy hits, which occur because the uniform illumination also causes photons to be absorbed in the (Nb, Ti)N surrounding the Al strip. On the membrane devices, part of that energy can reach the Al strip, causing a measurable pulse, whereas for the substrate devices that energy is quickly lost in the much thicker substrate.

R_{SN} is the estimate of the signal-to-noise contribution to the resolving power from the noise spectrum and the pulse

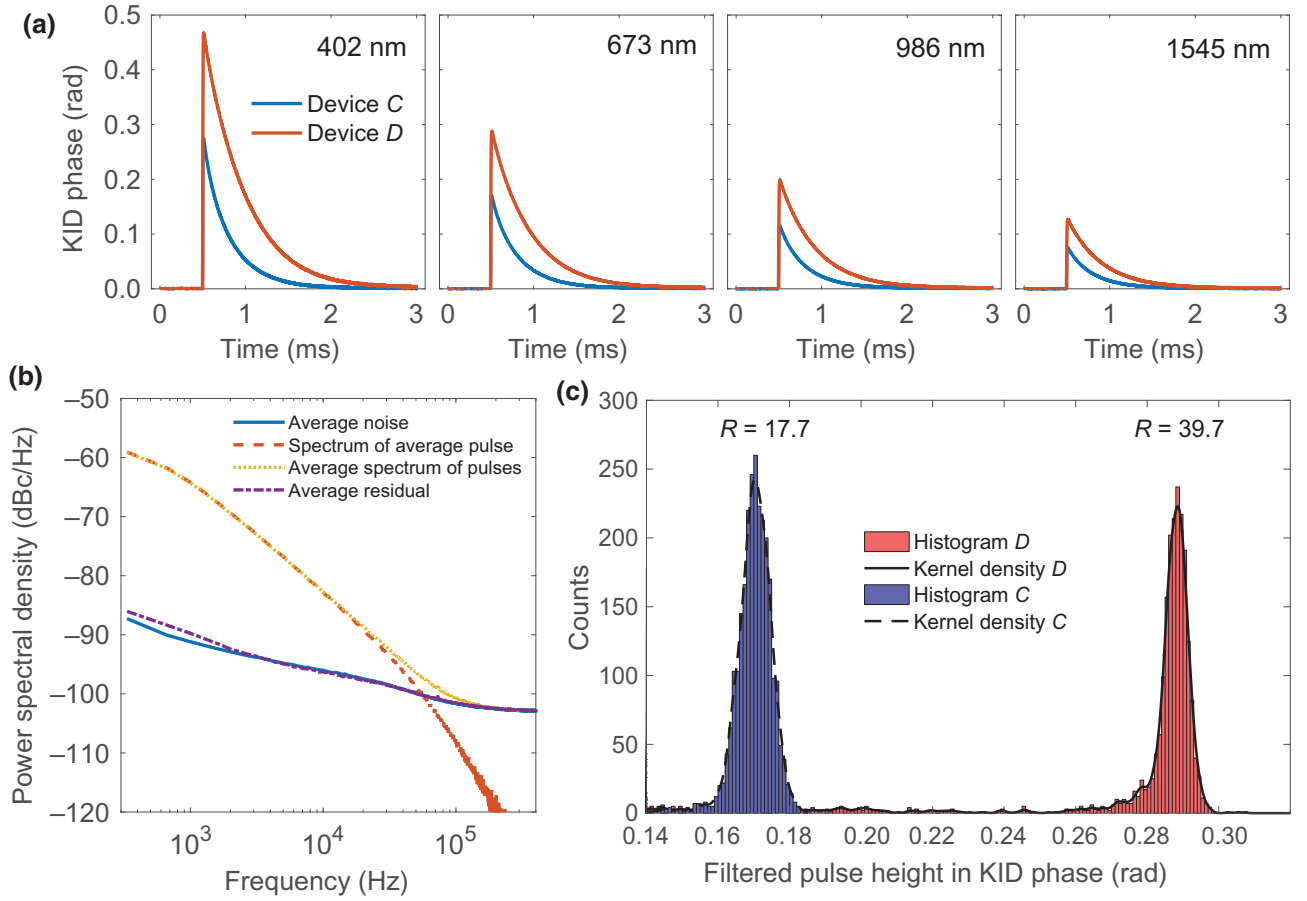


FIG. 3. (a) Average pulse shape, obtained by averaging typically thousand single-photon pulses measured at the indicated wavelengths for devices *C* (substrate) and *D* (membrane). The pulse height is partially due to the different Q_c (37.6×10^3 for *C* and 46.0×10^3 for *D*), but mainly due to the stronger phonon trapping in device *D*. (b) Power spectral densities of the average noise (solid) and average pulse (dashed), which together constitute the optimal filter for device *C* at 120 mK, with $P_{\text{read}} = -83$ dBm illuminated with a 673-nm laser. The area in between the pulse and noise curves determines the effective R_{SN} and bandwidth. As a consistency check we also plot the average of each pulse's spectrum (dotted) and average residual, $\theta(f) - AM(f)$ (dash dotted). (c) Histogram of pulses of device *C* and *D* at 120 mK, illuminated with the 673-nm laser and with $P_{\text{read}} = -83$ dBm and $P_{\text{read}} = -85$ dBm, respectively. The comparison of the FWHM of these histograms demonstrates the strong improvement in R due to phonon trapping from $R = 17.7$ to $R = 39.7$. The change in the average pulse height in the histogram corresponds to that in (a).

template [41,44]:

$$R_{\text{SN}} = \frac{\bar{A}}{2\sqrt{2 \ln 2}} \sqrt{\int_{-\infty}^{\infty} \frac{|M(f)|^2}{|N(f)|^2} df}, \quad (3)$$

with \bar{A} the average pulse height. R_{SN} is also called R_{NEP} in the literature. R_0 is the measured resolving power using the optimal filter [Eq. (2)] on pieces of noise. The resulting histogram is centered around $E = 0$ and R_0 represents the FWHM of that histogram with respect to the photon energy, $R_0 = E/\delta E_{E=0}$. It is a measure of the potential influence of, e.g., baseline fluctuations that are not taken into R_{SN} . Finally we analyze the contributions to the measured R using $1/R_{\text{intrinsic}}^2 = 1/R^2 - 1/R_{\text{SN}}^2$. $R_{\text{intrinsic}}$ is a

measure of the resolving-power contributions that are not due to noise.

V. PULSE HEIGHT ANALYSIS

Figure 3(a) shows the pulse response to single photons at the four measured wavelengths, each averaged over a few thousand single-photon hits, for device *C* on the SiN substrate and device *D* on the membrane.

We can estimate the pulse height using the number of quasiparticles generated by the photon $N_{qp} = \eta_{pb}E/\Delta$. For a linear detector, the resulting detector phase response is then given by $\theta = N_{qp}(d\theta/dN_{qp})$. We calculate the responsivity factor using $d\theta/dN_{qp} = -4Q/f_{\text{res},0} \times df_{\text{res}}(T)/dN_{qp}(T)$, where the resonant frequency $f_{\text{res}}(T)$ is obtained from a measurement of S_{21} as a function of bath

temperature [45]. Q is taken from the S_{21} measurement that we perform before each pulse measurement. This leaves η_{pb} as the only unknown, which we therefore use as a fit parameter. We calculate η_{pb} for each wavelength and for three different P_{read} , which results in an average $\eta_{pb} = 0.30 \pm 0.02$ for device C and $\eta_{pb} = 0.40 \pm 0.04$ for device D . The uncertainty represents the standard deviation of the 12 measurements for each device. For both devices, lower P_{read} leads to higher pulses and therefore higher estimated η_{pb} . The difference in η_{pb} between devices C and D means that in device D approximately 33% more energy is kept in the film during the down-conversion process. The numbers are close to the expected η_{pb} [11,46], although it has not been calculated in aluminum at high energies. For perfect phonon trapping, $\eta_{pb} = 0.59$ is expected. For the trapping factor on the membrane of approximately 5 we have here, this maximum is not yet reached, as shown in Ref. [46] for low energies. The present experiment should not be taken as a measurement of η_{pb} (as in Ref. [47]), because of the large range of energies involved, but mainly as a verification that the single-photon pulse heights in Al MKIDs are consistent with our current understanding of the response.

Capturing the behavior of the resonator in a single responsiveness term $d\theta/dN_{qp}$, measured for a quasiparticle population at elevated bath temperature is a strong

assumption. A more detailed pulse model, including the pulse rise and decay, should start from the rate equations for quasiparticle recombination [48,49], with an excitation due to the absorbed photon energy. The resulting $N_{qp}(t)$ can be converted to complex conductivity [50], $\sigma_1(t)$ and $\sigma_2(t)$, and surface impedance. Together with an estimate of the kinetic inductance fraction α_k [51], and the measured quality factors, one can calculate $S_{21}(t)$. The only leftover fit parameter would also be η_{pb} . We find however that the results of such a model strongly depend on whether one interprets the observed saturation in the pulse decay times at low temperatures as an effective quasiparticle temperature, or as a saturation of the lifetime only. This lifetime saturation is commonly observed [2,42,52]. We have recently performed new experiments, part of which on the here presented devices, which support a lifetime saturation without quasiparticle number saturation [31,53].

VI. ENERGY RESOLUTION ON THE SUBSTRATE AND ON THE MEMBRANE

The measured resolving power for the different devices as a function of photon energy is shown in Fig. 4(a), which is the main result of this work, together with the characteristic histogram in Fig. 3(c). We observe that for

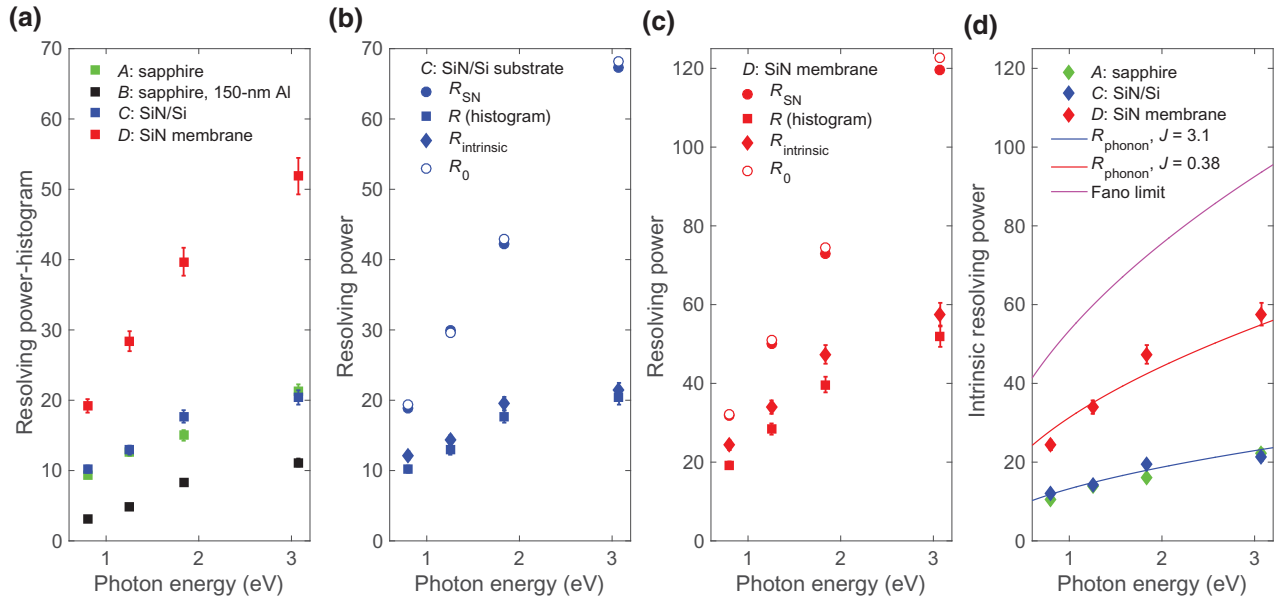


FIG. 4. Measured resolving power. (a) The resolving power R as obtained from the histogram of pulses for each of the four measured wavelengths for devices A – D . The much higher R due to phonon trapping for the device on a membrane is the main result of this work. The error bars are described in the text. (b),(c) The different contributions to the resolving power for devices C and D , respectively. The R_{SN} is the signal-to-noise contribution obtained from the average pulse and noise (filled circles). R_0 represents the FWHM of the histogram when the optimal filter is applied on traces of noise (open circles). R is the measured histogram resolving power as shown also in (a) (filled squares). $R_{\text{intrinsic}}$ quantifies all contributions to R that are not due to signal to noise, in this case dominated by hot phonon losses. (d) $R_{\text{intrinsic}}$ for devices A , C , and D . The R_{phonon} lines are fits of Eq. (4) to the data and represent the expected R for different levels of hot phonon loss. The magenta line is the Fano limit for the Al film properties of device D in Table I, using Eq. (1). Note that the y scale of the panels is different.

devices *A* and *C* on sapphire and SiN substrates the resolving power is similar and limited to 10–21 for wavelengths 1545–402 nm, respectively. We also observe that roughly $R \propto \sqrt{E}$, which already hints upon an intrinsic effect, since $R_{\text{SN}} \propto E$. For device *D* on the membrane, the resolving power increases to 19–52, which is a large improvement compared to devices *A* and *C*, and demonstrates that R on the substrate is indeed limited by the loss of hot phonons. The measured R at 1545 nm corresponds to a best measured energy resolution of 0.04 eV. The error bars are obtained from varying the kernel width of the kernel density estimate from a too coarse to a too fine representation of the histogram, which can lead to a $\pm 5\%$ variation in the estimated R .

On the way to a higher R , we have also measured a 150-nm Al film on sapphire, device *B*. The phonon escape time increases linearly with film thickness, which should enhance a phonon-loss limited $R_{\text{intrinsic}}$ significantly. However, as we observe in Fig. 4(a), the measured R is even lower than for devices *A* and *C*. R_{SN} strongly limits R for this thick film due to the reduced responsivity, which is caused by the increase in Al volume and decrease in α_k .

We separate the contributions of R_{SN} and $R_{\text{intrinsic}}$ to R as described in Sec. IV, which is shown in Fig. 4(b) for device *C* and in Fig. 4(c) for device *D* [the y scale of (a),(b) is different from (c),(d)]. We observe that going from the substrate to membrane increases R_{SN} by a factor 1.7, which is due to the increased pulse height on the membrane, where the noise stays the same (see Fig. 5 and Sec. VIII). We also observe that R_{SN} is close to linear in E , as expected in the linear regime of the MKID response, and as expected from the pulse heights in Fig. 3(a).

In both devices *C* and *D*, R_{SN} is a small contribution to R and so $R_{\text{intrinsic}}$ dominates. We exclude known limits to $R_{\text{intrinsic}}$ due to quasiparticle out diffusion and nonuniform current density by design (Sec. II). It is also unlikely that baseline fluctuations that are not captured into $N(f)$ play a dominant role since R_0 is very close to R_{SN} for both devices. It is therefore plausible that $R_{\text{intrinsic}}$ for the membrane device is also dominated by hot phonon losses. We therefore interpret the difference between the substrate and membrane as entirely due to the enhanced phonon trapping.

VII. ANALYSIS OF PHONON LOSS AND PHONON TRAPPING

We first derive the experimental phonon-loss factors from the resolving-power measurements in Sec. VII A and analyze the relative phonon trapping between devices on the substrate and on the membrane in Sec. VII B. In Sec. VII C we analyze the absolute value of the phonon-loss factor for the substrate devices.

A. Measured phonon-loss factor

We quantify the effective level of phonon loss and phonon trapping on the membrane by comparing the $R_{\text{intrinsic}}$ for the devices on the substrate (*C*) and on the membrane (*D*), as shown in Fig. 4(d). The influence of hot phonon loss on the resolving power can be expressed as [27,28]

$$R_{\text{phonon}} = \frac{1}{2\sqrt{2 \ln 2}} \sqrt{\frac{\eta_{pb}^{\text{max}} E}{\Delta(F+J)}}, \quad (4)$$

with J accounting for the hot phonon loss. Per definition the Fano limit is reached when $J = 0$ [Eq. (1)]. Note that the minimum energy to generate a quasiparticle, $\Delta/\eta_{pb}^{\text{max}}$, should be used here also for significant phonon losses. The generally used η_{pb} is an expression of average energy lost. R_{phonon} expresses the variance of that process, for which the effect of phonon loss is taken into account by $J > 0$.

Using the known values for F , Δ , and η_{pb}^{max} , and assuming that $R_{\text{intrinsic}}$ is indeed dominated by phonon losses, we can fit Eq. (4) to $R_{\text{intrinsic}}$ for devices *C* and *D*, which results in the curves plotted in Fig. 4(d), with $J_C = 3.1 \pm 0.4$ and $J_D = 0.38 \pm 0.08$. We conclude that in device *D*, hot phonon loss still dominates over the Fano factor ($F = 0.2$).

B. Enhanced phonon trapping on the membrane

The phonon loss factor J is expected to scale linearly with the phonon trapping factor $\tau_{\text{esc}}/\tau_{pb}$, with τ_{esc}

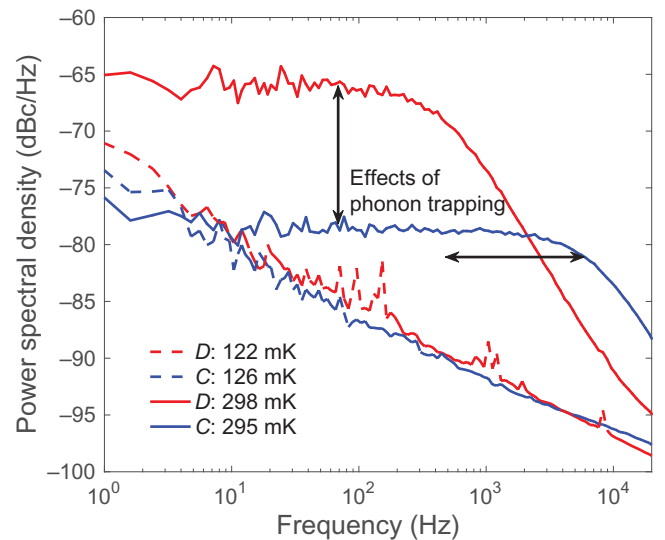


FIG. 5. Power spectral density of the resonator phase for devices *C* and *D* measured at temperatures close to 120 mK (dashed lines), where we measure the resolving power, and at high temperatures (solid lines). The arrows indicate the enhanced generation-recombination (GR) noise level and lifetime due to phonon trapping, which show up at high temperature. The low-temperature spectra have no signature of GR noise.

the average time a phonon spends in the superconductor and τ_{pb} the average time it takes to break a Cooper pair [27,28]. $\tau_{pb} = 0.26$ ns for Al [54]. The ratio $J_C/J_D = 8 \pm 2$ derived from the experiments therefore expresses the *relative* phonon-trapping enhancement between the device on the membrane compared to the substrate.

To model the enhanced phonon trapping we consider that the phonon energies, which dominate the down-conversion process range from the Debye energy to about 10Δ [11]. These energies correspond to phonon wavelengths of 0.5–10 nm, which are all smaller than the dimensions of our device. Therefore we use a geometrical ray-tracing model, in which phonons are generated in the superconducting film with random angles. Upon each interaction with the interface we calculate the transmission probability from the acoustic matching [55]. A phonon that ends up in the substrate is lost. Such a phonon can interact with the metal layers all over the substrate, therewith losing its energy, and finally end up in the bath (the sample holder). For the membrane a phonon has a (high) chance to be reflected back into the film, dependent on the membrane thickness and the width of the film. After the phonon leaves the membrane area underneath the superconductor, we introduce a 50% chance of return to account for diffusive reflection at the membrane surfaces. Details of the model are discussed in Appendix A. The resulting average phonon dwell time in the 1.7- μm -wide Al strip is 0.16 ns for the substrate and 1.3 ns for a 110-nm membrane. The resulting phonon-trapping enhancement is a factor 8.1, which agrees well with the value derived above from the resolving-power measurement, despite the simplicity of the model.

C. Absolute phonon-loss calculation on the substrate

So far we only compared the relative $R_{\text{intrinsic}}$ between the substrate and membrane. The absolute value of J_C , on the substrate, can be compared to the detailed model of the different down-conversion stages in Refs. [11,27,28].

For the high-energy part of the down-conversion process (from E_{photon} to around E_{Debye}), we evaluate Eq. (25) from Ref. [27], for photons incident on the top surface. We first consider that for the highest phonon energies, around the Debye energy, only longitudinal phonons are available for Al [56,57], and therefore the input parameters are taken only for longitudinal phonons. The result is $J_{\text{high}} = 2.8$. The input parameters are taken for Al from Table I and the timescales and characteristic energies from Refs. [11,55] and the acoustical properties of SiN from Ref. [58].

The low energy part of the down-conversion involves energies ranging from E_{Debye} to the characteristic energy for Al, which is 10.6Δ [11]. We use Eq. (6) from Ref. [28], with a correction [57] as given in Appendix C. In this phase, both longitudinal and transversal phonons play a role. The result from this calculation is $J_{\text{low}} = 0.24$. It is

clear that the dominant effect on the energy variance, and therefore on the resolving power, originates from the first steps of the down-conversion process as expressed in J_{high} .

The phonon-loss factors from both stages can be added together, $J = J_{\text{high}} + J_{\text{low}} = 3.0$, which agrees well with the measured J_C .

When we fit in the same way the measured $R_{\text{intrinsic}}$ for device A on sapphire, we obtain $J_A = 3.7 \pm 0.3$. We then use the same equations with the acoustical properties of sapphire to obtain theoretically $J = 2.7$, which is lower than the measured J_A .

We realise that not all the input parameters to calculate J_{high} and J_{low} are accurately measured. For example, we assume a diffusion constant for thin Al of $20 \text{ cm}^2/\text{s}$, where measured values range from 18–60 cm^2/s [59–61]. Next to that, the Debye model is poorly justified for the high energies involved. Additionally, the acoustic model assumes a perfect interface for transmission and reflection, which may be rough in reality. Therefore, a realistic error margin on the calculation for J is at least ± 0.5 . In the light of these considerations and the large range of energies involved, a correct order-of-magnitude estimate is already satisfactory.

VIII. DISCUSSION

A. Phonon trapping factor in quasiparticle fluctuations

We have also measured quasiparticle fluctuations (generation recombination noise) in the same devices, which we present in detail in Ref. [31]. The quasiparticle recombination times on the substrate and membrane, fitted from the spectra of these fluctuations, show a factor 66 ± 17 difference in phonon escape time. The effect is illustrated for the high temperature ($T \sim 300$ mK) noise spectra of devices C and D in Fig. 5. The phonon-trapping factor that applies to quasiparticle fluctuations and lifetime in the spectra is $(1 + \tau_{\text{esc}}/\tau_{pb})$. The quasiparticle lifetime is a factor 16 longer and the noise 12 dB higher on the membrane. The discrepancy between the numbers for τ_{esc} from the measured R and from the quasiparticle fluctuations needs further study. The quasiparticle fluctuations are maintained by phonons with energies close to 2Δ , much lower than those that determine R . At the corresponding phonon wavelength, around 50 nm, the phonon dimensionality and the distribution over different phonon modes starts to play a role, for which the ray-tracing model is inappropriate.

As shown in Fig. 5, the effect of phonon trapping is clearly visible in the GR noise spectra at higher temperatures (> 250 mK). However, we observe roughly the same lifetime for the main part of the pulse decay in both cases at 120 mK (319 μs for C and 431 μs for D). At the same time, we also observe a strong reduction in generation-recombination noise at low temperatures, as shown for

temperatures close to 120 mK in Fig. 5, which was unexpected [42,49]. This is nevertheless the main reason why we get the required R_{SN} for a membrane-based device at 120 mK. Both these observations and their possible relation with quasiparticle trapping are extensively discussed in a separate publication [31]. For comparison we measure the resolving power for device D at 250 mK at 1545 nm. At this temperature $R = 11$, limited by $R_{\text{SN}} = 13.8$, which is much lower than at 120 mK, despite a longer pulse decay time of 1.0 ms ($R_{\text{SN}} = 31.8$ at 120 mK). Understanding this reduction in noise is critical to understand the high R_{SN} measured here, but we emphasize it does not influence the limit to R due to hot phonon loss.

B. Towards a Fano-limited resolving power

The next question that is raised by this work is how to experimentally reach a Fano-limited energy resolution. In the case of Al on a thin membrane as in device D , we have $J_D = 1.9F$, whereas $J \ll F$ is needed for the Fano factor to dominate the resolving power. If we set a goal of $R = 0.9R_{\text{Fano}}$, we need $J < F/5$, and therefore another factor of 10 better phonon trapping. This is unrealistic with the geometrical phonon-trapping approach we take now, since either the membrane becomes impractically thin, or the Al strip becomes too wide to conserve R_{SN} (only with a more resistive metal, a wider strip remains an option). We therefore envision the fabrication of a phononic crystal (PnC) to better trap the phonons that flow to the bath. It is helpful for MKIDs to have a PnC that blocks high-frequency phonons, but is transparent for phonons at the microwave readout frequency (2 – 6 GHz) to reduce microwave nonequilibrium effects [62,63]. It is possible to achieve a factor 10 phonon trapping at particular energy bands (compared to an unpatterned membrane), while keeping high phonon transparency at low energies [64,65].

The analysis above shows that it is realistic to get much closer to the Fano limit. However, reaching $R > 0.9R_{\text{Fano}}$ is unrealistic from a phonon-trapping point of view. Additionally, the requirement $R_{\text{SN}} \gg R_{\text{Fano}}$ is also difficult. For our Al MKIDs, apart from the additional phonon trapping, we aim to improve R_{SN} further by reducing the film thickness to lower the volume and improve the kinetic inductance fraction. Additionally, we can use a pulse filter combining the amplitude and phase response [16,66]. For the present data, we find the R_{SN} of phase only to be sufficient.

It is striking that most substrate-based MKIDs with a high enough R_{SN} are all limited to $R \approx 10 - 20$ [3,12–14,16,17]. Here we conclusively show that in Al MKIDs on substrate R is limited by hot phonon loss. For other materials this is likely the case, although the acoustic properties of these materials need to be better characterized. We emphasize that the loss of hot phonons arises at the initial stages of down-conversion at energies much higher than

the quasiparticle excitations that are eventually detected. Therefore, hot phonon loss arises in pair-breaking detectors (MKID, STJ, QCD), in thermal (e.g., TES) detectors, and also in SNSPD detectors [67].

IX. CONCLUSION

We measure the resolving power of superconducting, energy-resolving, single-photon counting MKIDs for wavelengths of 1545–402 nm. By comparing devices where the superconductor is deposited on the substrate with devices on a 110-nm membrane, we show that the resolving power is limited by the loss of hot phonons. These phonons are involved in the down-conversion process, in which the absorbed single-photon energy is converted into measurable quasiparticle excitations through multiple stages of electron-electron and electron-phonon interaction. For devices on substrate the resolving power is limited to 10–21 for 1545–402 nm, respectively. On the membrane this improves to 19–52, which is consistent with a factor 8 ± 2 stronger phonon trapping compared to the substrate. Even stronger phonon trapping, possibly using phononic crystals, is needed to get closer to a Fano-limited resolving power, together with a further improvement of the signal to noise (thinner film, smaller volume).

All presented data, raw data and analysis software will be made available in a reproduction package uploaded to Zenodo [69].

ACKNOWLEDGMENTS

We acknowledge help with the experimental setup and initial data analysis from Max van Strien, Robert Huiting, Klaas Keizer, and Juan Bueno. We acknowledge help from Alex Kozorezov to interpret his equations for the down-conversion process, and insightful discussions with Karwan Rostem, Tuomas Puurtinen, Ilari Maasilta, Henk Hoevers, Sae Woo Nam, Jonas Zmuidzinis, Harvey Moseley, and Ben Mazin. P.J.dV. is financially supported by the Netherlands Organisation for Scientific Research NWO (Veni Grant No. 639.041.750 and Projectruimte 680-91-127). J.J.A.B. is supported by the European Research Council ERC (Consolidator Grant No. 648135 MOSAIC).

APPENDIX A: PHONON-TRAPPING MODEL

The phonon trapping we rely on in the membrane experiment is purely geometrical: the Al film is much wider (1.7 μm) than the membrane thickness (110 nm), which leads to reflection of the phonons back into the Al film. To estimate the effective phonon trapping due to the thin membrane underneath the superconducting strip, we perform a simulation with a geometrical ray-tracing model. In the down-conversion process from photon energy to quasiparticles, phonons play a prominent role. For this problem we are mainly interested in phonons with relatively high

energies and therefore short phonon wavelengths. Phonons with energies ranging from the Debye energy (37 meV in Al) to around $10\Delta = 1.9$ meV dominate this process [11], which corresponds to typical phonon wavelengths of 0.5–10 nm. These wavelengths are small enough to expect that phonon dimensionality does not play a role yet and a geometrical model is a reasonable assumption.

In the model, we initialize a phonon at the top surface of the metal film at a random position and with a random angle of propagation and trace the propagation of the phonon. Reflection at any interface is assumed to randomize the angle of reflection due to roughness, in comparison to the short wavelengths. When hitting the metal-substrate interface we calculate the probability of the transmission from the acoustic mismatch model with the equations from Ref. [55]. Upon transmission the phonon is refracted according to Snell's law.

When the phonon enters the area underneath the film we distinguish between the substrate and membrane. For a film on the substrate, we consider a phonon that enters the substrate to be lost, since the substrate is $350\ \mu\text{m}$ thick, compared to the metal strip of $1.7\ \mu\text{m}$ wide.

For the membrane, a phonon that hits the bottom of the membrane is reflected back at random angle, and therefore has a probability to re-enter the film, again with a probability from the acoustic matching. When the phonon propagates outside the area underneath the metal, we estimate in this case a 50% probability of returning underneath the film at a random angle, to take into account that any (diffusive) interaction with the membrane interfaces outside of this area can lead to the phonon returning. Diffusive scattering at the membrane interfaces is plausible considering that the membrane is exposed to the reactive ion etch for the (Nb, Ti)N on the front side. Phonons that do not return we consider lost.

The effect of phonon trapping is quantified using the total dwell time of the phonon in the superconducting film, usually called the phonon escape time τ_{esc} . We average the dwell time of 100 000 phonons for each set of parameters. The results are shown in Table II. For an Al film on a $350\text{-}\mu\text{m}$ -thick sapphire substrate using this model we find reasonable agreement with the analytical expression $\tau_{\text{esc}} = 4d/\eta u$, with d the film thickness, u the

average phonon velocity, and η the interface transparency (when integrated over all angles), using numbers from Ref. [55].

Comparing the dwell times on the SiN substrate and the SiN membrane, we find a factor of 8.1 enhancement due to the thin membrane. This corresponds very well to the trapping factor measured from resolving power 8 ± 2 , especially given the simplicity of the model.

The phonon-trapping enhancement obtained from the measurements of quasiparticle fluctuations at temperatures from 250–350 mK is 66 ± 17 [31]. The equilibrium generation-recombination process involves only phonons with energies close to 2Δ . An energy of 2Δ corresponds to a phonon wavelength of approximately 50 nm, which is comparable to the thicknesses of film and membrane. Therefore, one would not *a priori* expect a close agreement between the effective trapping at high energies and these low energies. A complete model of phonon propagation in a thin membrane at the different energies involved in the down-conversion process (from the Debye energy down to 2Δ) would be needed to microscopically model the phonon propagation. This model should include the preferential phonon modes emitted from the aluminium film, the density of phonon states of film and membrane and the phonon scattering at the (etched) surfaces of the membrane.

APPENDIX B: LOW-ENERGY EVENTS FOR MEMBRANE DEVICES

The full histograms of devices *C* and *D*, corresponding to the resolving-power measurements in Fig. 4(a), are shown in Figs. 6 and 7, respectively. Comparing the substrate- and membrane-based MKIDs, we find that low-energy hits are more frequent in the membrane-based device *D*. In the membrane device there is typically a strong peak in the histogram at the expected pulse height for direct Al hits (at the highest pulse height). Then there is a broad shoulder at low energies, most likely related to (Nb, Ti)N hits, since this shoulder is much stronger for 402-nm light, where (Nb, Ti)N is more absorptive. The third class of hits is represented by a peak at low energy, with typically 30% the pulse height of the primary hits. For the substrate MKIDs, the signal to noise is also high

TABLE II. Overview of the phonon properties used in the ray-tracing model and the resulting phonon escape times from an Al film on top of that substrate. In the model we take the 110-nm SiN on the Si substrate as $350\ \mu\text{m}$ SiN. Properties of Al and sapphire are taken from Ref. [55] and SiN from Refs. [58,68].

Material	Thickness	c_t (km/s)	c_l (km/s)	density (g/cm ³)	τ_{esc} (ns) 50% return	τ_{esc} (ns) no return
Al	50 nm	3.26	6.65	2.73		
Sapphire	$350\ \mu\text{m}$	6.45	10.9	3.99		0.18
SiN	$350\ \mu\text{m}$	6.2	10.3	3.1		0.16
SiN	110 nm	6.2	10.3	3.1	1.3	0.74

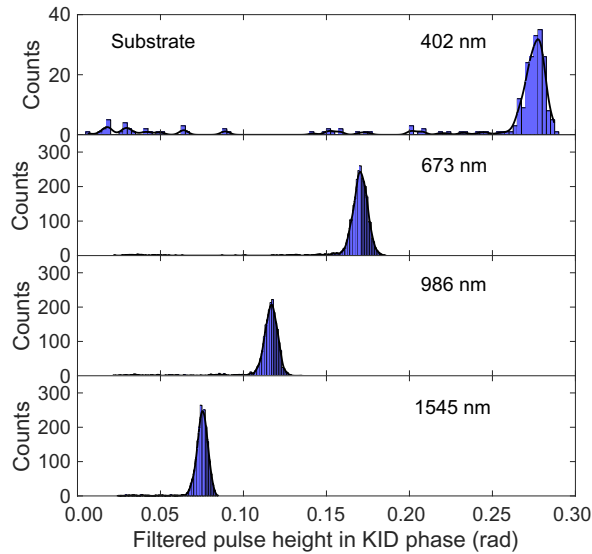


FIG. 6. Full histograms for the laser wavelengths as indicated for device *C*, at -83 dBm and 120 mK. The resolving powers following from the kernel density estimate (lines) are presented in Fig. 4(a).

enough to see both classes of low-energy hits if they would occur, but this is not the case (Fig. 6). On top of that, the number of direct Al hits per unit time is similar for devices *C* and *D* for the same laser power. We therefore attribute both classes of low-energy hits to photons being absorbed in other parts of the detector and leading to a response through phonon transport in the membrane. The broad shoulder in the histogram can therefore be attributed to the large (Nb, Ti)N areas on the left and right of the membrane [see Fig. 1(e)]. These hits lead to different pulse

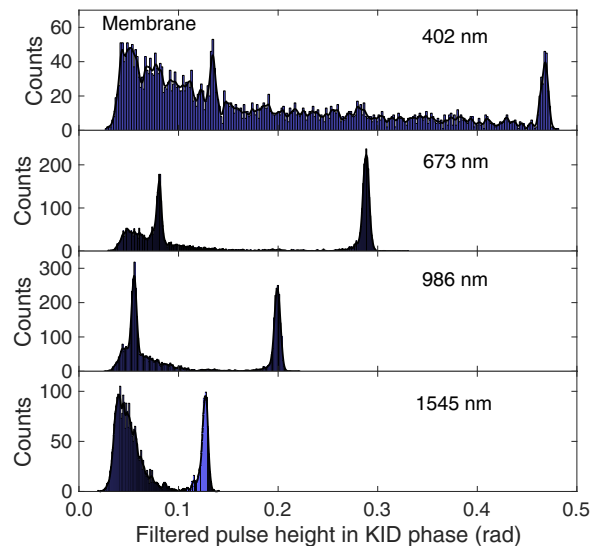


FIG. 7. Full histograms for the laser wavelengths as indicated for device *D*, at -85 dBm and 120 mK. The resolving powers following from the kernel density estimate (lines) are presented in Fig. 4(a).

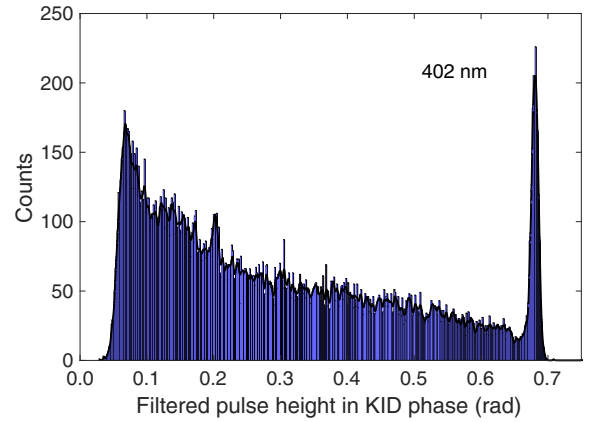


FIG. 8. Full histogram for the 402-nm laser for device *D*, at -87 dBm and 120 mK, based on 26 306 counts (3987 in Fig. 7). The resulting resolving power is the same as in Fig. 7.

heights because energy can leak out due to diffusion and phonon transport into the substrate. The low-energy peak is most likely related to the (Nb, Ti)N connections in the center of the membrane where it is more likely that a well-defined amount of energy reaches the Al. The low-energy hits show up in the histogram far enough from the primary peak to not negatively affect the measured R , which we take as the FWHM of the primary peak. For the substrate device, the energy of such a hit is quickly lost in the thick substrate, before reaching the Al strip, and does not lead to a measurable pulse.

The low-energy events could be avoided by properly focusing the radiation onto the sensitive part of the MKID with a microlens. As explained in Sec. II we choose a well-studied geometry with sufficient signal to noise for energy-resolution studies, with the downside of absorption of photons also outside of the most sensitive, Al strip.

For device *D* at 402 nm, as shown in the top panel of Fig. 7, we also measure a longer time trace (1024 s) in a separate cooldown to check the resolving power with more counts in the histogram, which is shown in Fig. 8, and gives a consistent R with Fig. 7. The average pulse height is higher than in Fig. 7, because the coupling quality factor changes from cooldown to cooldown, here from 48.5×10^3 to 77.7×10^3 , most likely due to the very open IDC structure.

Because of the high signal to noise, subsequent single-photon hits that are too close in time are generally easy to distinguish and filter out before making the histogram. However, in the presence of low-energy hits on device *D* this is not always effective. We therefore need to remove double hits that involve a primary and a low-energy hit in a postprocessing step, for which we reject these double hits based on the residual after applying the optimal filter, $\theta(f) - AM(f)$, integrated over frequency [see Fig. 3(b)]. We verify that this step removes only double hits and does not artificially reduce the number of low-energy hits, which is evident from Figs. 7 and 8.

APPENDIX C: EQUATIONS FOR THE LOW-ENERGY PART OF THE DOWN-CONVERSION

The equations for the low-energy part of the down-conversion, J_{low} , are taken from Ref. [28] [Eq. (6) therein]. We reproduce them here in the original notation with corrections [57]

$$J_{\text{low}} = \eta (1 - \xi_c^2) \frac{\Omega_D l_{pe,D}}{\epsilon_0 d} \frac{3Z_1(0)}{11Z_1(0) + 3} g_1 \left(\frac{\Omega_D}{\Omega_1} \right). \quad (\text{C1})$$

Here we take $Z_1(0) = 1$. $g_1(x)$ is given by

$$g_1(x) = \int_{1/x}^1 dz f(z) \left(Ei[1, z(x-1)] - Ei(1, 1-z) + \ln \left[\frac{(x-1)z}{1-z} \right] - \frac{1}{4} \eta (1 - \xi_c) \left\{ Ei[1, 2z(x-1)] - Ei[1, 2(1-z)] + \ln \left[\frac{(x-1)z}{1-z} \right] \right\} \right), \quad (\text{C2})$$

where the $\ln(\dots)$ terms are corrected with respect to Ref. [28]. $f(x)$ is given by

$$f(x) = 1 - x - \frac{1}{12} x \left\{ x^2 \left[\cos(\sqrt{2} \ln x) - 7\sqrt{2} \sin(\sqrt{2} \ln x) \right] - 1 \right\}.$$

Here Ω_D is the Debye energy, Ω_1 is the characteristic energy at which this stage ends, η is the acoustical transmission coefficient, $l_{pe,D}$ is the mean free path, d the film thickness, $\epsilon_0 = 1.7\Delta$ the average quasiparticle energy and $\xi_c = \cos(\theta_c)$, with θ_c the critical angle for phonon transmission at the interface.

-
- [1] P. K. Day, H. G. LeDuc, B. A. Mazin, A. Vayonakis, and J. Zmuidzinas, A broadband superconducting detector suitable for use in large arrays, *Nature* **425**, 817 (2003).
 - [2] Jonas Zmuidzinas, Superconducting microresonators: Physics and applications, *Annu. Rev. Condens. Matter Phys.* **3**, 169 (2012).
 - [3] Benjamin A. Mazin, Bruce Bumble, Seth R. Meeker, Kieran O'Brien, Sean McHugh, and Eric Langman, A superconducting focal plane array for ultraviolet, optical, and near-infrared astrophysics, *Opt. Express* **20**, 1503 (2012).
 - [4] Seth R. Meeker *et al.*, DARKNESS : A microwave kinetic inductance detector integral field spectrograph for high-contrast astronomy, *Publ. Astron. Soc. Pac.* **130**, 65001 (2018).
 - [5] Alex Walter, Benjamin B. Mazin, Clint Bockstiegel, Neelay Fruitwala, Paul Szypryt, Isabel Lipartito, Seth Meeker, Nicholas Zobrist, Giulia Collura, Grégoire Coiffard, Paschal Strader, Olivier Guyon, Julien Lozi, and Nemanja Jovanovic, in *Ground-based and Airborne Instrumentation for Astronomy VII*, edited by Christopher J. Evans, Luc Simard, and Hideki Takami, International Society for Optics and Photonics (SPIE, Austin, United States, 2018), Vol. 10702.

-
- [6] Roger W. Romani, A. J. Miller, B. Cabrera, S. W. Nam, and John M. Martinis, Phase-resolved crab studies with a cryogenic transition-edge sensor spectrophotometer, *Astrophys. J.* **563**, 221 (2001).
 - [7] P. Verhoeve, D. D. E. Martin, R. A. Hijmering, J. Vermeer, A. van Dordrecht, G. Sirbi, T. Oosterbroek, and A. Peacock, S-cam 3: Optical astronomy with a stj-based imaging spectrophotometer, *Nucl. Instrum. Methods Phys. Res. A: Accel. Spectrom. Detect. Assoc. Equip.* **559**, 598 (2006).
 - [8] Kazuki Niwa, Takayuki Numata, Kaori Hattori, and Daiji Fukuda, Few-photon color imaging using energy-dispersive superconducting transition-edge sensor spectrometry, *Sci. Rep.* **7**, 45660 (2017).
 - [9] M. Kurakado, Possibility of high resolution detectors using superconducting tunnel junctions, *Nucl. Instrum. Methods* **196**, 275 (1982).
 - [10] N. Rando, A. Peacock, A. van Dordrecht, C. Foden, R. Engelhardt, B. G. Taylor, P. Gare, J. Lumley, and C. Pereira, The properties of niobium superconducting tunneling junctions as x-ray detectors, *Nucl. Instrum. Methods Phys. Res. A: Accel. Spectrom. Detect. Assoc. Equip.* **313**, 173 (1992).
 - [11] A. G. Kozorezov, A. F. Volkov, J. K. Wigmore, A. Peacock, A. Poelaert, and R. den Hartog, Quasiparticle-phonon downconversion in nonequilibrium superconductors, *Phys. Rev. B* **61**, 11807 (2000).
 - [12] B. A. Mazin, S. R. Meeker, M. J. Strader, P. Szypryt, D. Marsden, J. C. van Eyken, G. E. Duggan, A. B. Walter, G. Ulbricht, M. Johnson, B. Bumble, K. O'Brien, and C. Stoughton, ARCONS: A 2024 pixel optical through near-IR cryogenic imaging spectrophotometer, *Publ. Astron. Soc. Pac.* **125**, 1348 (2013).
 - [13] P. Szypryt, S. R. Meeker, G. Coiffard, N. Fruitwala, B. Bumble, G. Ulbricht, A. B. Walter, M. Daal, C. Bockstiegel, G. Collura, N. Zobrist, I. Lipartito, and B. A. Mazin, Large-format platinum silicide microwave kinetic

- inductance detectors for optical to near-IR astronomy, *Opt. Express* **25**, 25894 (2017).
- [14] Nicholas Zobrist, Grégoire Coiffard, Bruce Bumble, Noah Swimmer, Sarah Steiger, Miguel Daal, Giulia Collura, Alex B. Walter, Clint Bockstiegel, Neelay Fruitwala, Isabel Lipartito, and Benjamin A. Mazin, Design and performance of hafnium optical and near-IR kinetic inductance detectors, *Appl. Phys. Lett.* **115**, 213503 (2019).
- [15] W. Guo, X. Liu, Y. Wang, Q. Wei, L. F. Wei, J. Hubmayr, J. Fowler, J. Ullom, L. Vale, M. R. Vissers, and J. Gao, Counting near infrared photons with microwave kinetic inductance detectors, *Appl. Phys. Lett.* **110**, 212601 (2017).
- [16] Nicholas Zobrist, Byeong Ho Eom, Peter Day, Benjamin A. Mazin, Seth R. Meeker, Bruce Bumble, Henry G. LeDuc, Grégoire Coiffard, Paul Szypryt, Neelay Fruitwala, Isabel Lipartito, and Clint Bockstiegel, Wide-band parametric amplifier readout and resolution of optical microwave kinetic inductance detectors, *Appl. Phys. Lett.* **115**, 042601 (2019).
- [17] Nicholas Zobrist, Nikita Klimovich, ByeongHo Eom, Grégoire Coiffard, Miguel Daal, Noah J. Swimmer, Sarah Steiger, Bruce Bumble, Henry G. LeDuc, Peter K. Day, and Benjamin A. Mazin, Improving the dynamic range of single photon counting kinetic inductance detectors, *J. Astron. Telescopes, Instrum., Syst.* **7**, 010501 (2021).
- [18] B. Cabrera, R. M. Clarke, P. Colling, A. J. Miller, S. Nam, and R. W. Romani, Detection of single infrared, optical, and ultraviolet photons using superconducting transition edge sensors, *Appl. Phys. Lett.* **73**, 735 (1998).
- [19] L. Lolli, E. Taralli, C. Portesi, E. Monticone, and M. Rajteri, High intrinsic energy resolution photon number resolving detectors, *Appl. Phys. Lett.* **103**, 041107 (2013).
- [20] Danna Rosenberg, Adriana E. Lita, Aaron J. Miller, and Sae Woo Nam, Noise-free high-efficiency photon-number-resolving detectors, *Phys. Rev. A* **71**, 061803 (2005).
- [21] Adriana E. Lita, Aaron J. Miller, and Sae Woo Nam, Counting near-infrared single-photons with 95% efficiency, *Opt. Express* **16**, 3032 (2008).
- [22] Marissa Giustina, Alexandra Mech, Sven Ramelow, Bernhard Wittmann, Johannes Kofler, Jörn Beyer, Adriana Lita, Brice Calkins, Thomas Gerrits, Sae Woo Nam, Rupert Ursin, and Anton Zeilinger, Bell violation using entangled photons without the fair-sampling assumption, *Nature* **497**, 227 (2013).
- [23] Marissa Giustina *et al.*, Significant-Loophole-Free Test of Bell's Theorem with Entangled Photons, *Phys. Rev. Lett.* **115**, 250401 (2015).
- [24] Lynden K. Shalm *et al.*, Strong Loophole-Free Test of Local Realism, *Phys. Rev. Lett.* **115**, 250402 (2015).
- [25] D. D. E. Martin, P. Verhoeve, A. Peacock, A. G. Kozorezov, J. K. Wigmore, H. Rogalla, and R. Venn, Resolution limitation due to phonon losses in superconducting tunnel junctions, *Appl. Phys. Lett.* **88**, 123510 (2006).
- [26] M. Kurakado and H. Mazaki, Possibility of a high-resolution nuclear radiation detector using a superconductor: I. mean energy loss per excess quasiparticle, *Nucl. Instrum. Methods Phys. Res.* **185**, 141 (1981).
- [27] A. G. Kozorezov, J. K. Wigmore, D. Martin, P. Verhoeve, and A. Peacock, Electron energy down-conversion in thin superconducting films, *Phys. Rev. B* **75**, 094513 (2007).
- [28] A. G. Kozorezov, J. K. Wigmore, D. Martin, P. Verhoeve, and A. Peacock, Phonon noise in thin metal films in an advanced energy down-conversion stage, *J. Low Temp. Phys.* **151**, 51 (2008).
- [29] P. J. de Visser, J. J. A. Baselmans, J. Bueno, N. Llombart, and T. M. Klapwijk, Fluctuations in the electron system of a superconductor exposed to a photon flux, *Nat. Commun.* **5**, 3130 (2014).
- [30] J. J. A. Baselmans, J. Bueno, S. J. C. Yates, O. Yurduseven, N. Llombart, K. Karatsu, A. M. Baryshev, L. Ferrari, A. Endo, D. J. Thoen, P. J. de Visser, R. M. J. Janssen, V. Murugesan, E. F. C. Driessen, G. Coiffard, J. Martin-Pintado, P. Hargrave, and M. Griffin, A kilo-pixel imaging system for future space based far-infrared observatories using microwave kinetic inductance detectors, *A&A* **601**, A89 (2017).
- [31] Steven A. H. de Rooij, Jochem J. A. Baselmans, Vignesh Murugesan, David J. Thoen, and Pieter J. de Visser, Strong reduction of quasiparticle fluctuations in a superconductor due to decoupling of the quasiparticle number and lifetime, [arXiv:2103.04777](https://arxiv.org/abs/2103.04777) (2021).
- [32] J. J. A. Baselmans *et al.*, (to be published).
- [33] R. Barends, H. L. Hortensius, T. Zijlstra, J. J. A. Baselmans, S. J. C. Yates, J. R. Gao, and T. M. Klapwijk, Noise in NbTiN, Al, and ta superconducting resonators on silicon and sapphire substrates, *IEEE Trans. Appl. Supercond.* **19**, 936 (2009).
- [34] Jiansong Gao, Jonas Zmuidzinas, Benjamin A. Mazin, Henry G. LeDuc, and Peter K. Day, Noise properties of superconducting coplanar waveguide microwave resonators, *Appl. Phys. Lett.* **90**, 102507 (2007).
- [35] Jiansong Gao, Miguel Daal, John M. Martinis, Anastasios Vayonakis, Jonas Zmuidzinas, Bernard Sadoulet, Benjamin A. Mazin, Peter K. Day, and Henry G. LeDuc, A semiempirical model for two-level system noise in superconducting microresonators, *Appl. Phys. Lett.* **92**, 212504 (2008).
- [36] Boy Gustaaf Cornelis Bos, David Johannes Thoen, E. A. F. Haalebos, P. M. L. Gimbel, T. M. Klapwijk, J. J. A. Baselmans, and Akira Endo, Reactive magnetron sputter deposition of superconducting niobium titanium nitride thin films with different target sizes, *IEEE Trans. Appl. Supercond.* **27**, 1 (2017).
- [37] David Johannes Thoen, Boy Gustaaf Cornelis Bos, E. A. F. Haalebos, T. M. Klapwijk, J. J. A. Baselmans, and Akira Endo, Superconducting NbTiN thin films with highly uniform properties over a 100 mm diameter wafer, *IEEE Trans. Appl. Supercond.* **27**, 1 (2017).
- [38] J. Baselmans, S. Yates, P. Diener, and P. de Visser, Ultra low background cryogenic test facility for far-infrared radiation detectors, *J. Low Temp. Phys.* **167**, 360 (2012).
- [39] Pieter de Visser, Ph.D. thesis, Delft University of Technology, 2014, <https://doi.org/10.4233/uuid:ee4c9fc-f90d-4c12-a878-8428ee4adb4c>.
- [40] M. Naftaly and R. E. Miles, Terahertz time-domain spectroscopy of silicate glasses and the relationship to material properties, *J. Appl. Phys.* **102**, 043517 (2007).
- [41] Kent D. Irwin, Ph.D. thesis, Stanford University, 1995, <https://www.slac.stanford.edu/exp/cdms/Science/Results/Theses/irwin.pdf>.
- [42] P. J. de Visser, J. J. A. Baselmans, P. Diener, S. J. C. Yates, A. Endo, and T. M. Klapwijk, Number Fluctuations

- of Sparse Quasiparticles in a Superconductor, *Phys. Rev. Lett.* **106**, 167004 (2011).
- [43] K. Karatsu, A. Endo, J. Bueno, P. J. de Visser, R. Barends, D. J. Thoen, V. Murugesan, N. Tomita, and J. J. A. Baselmans, Mitigation of cosmic ray effect on microwave kinetic inductance detector arrays, *Appl. Phys. Lett.* **114**, 032601 (2019).
- [44] Megan E. Eckart, Ph.D. thesis, California Institute of Technology, 2007, <https://thesis.library.caltech.edu/2508>.
- [45] J. Baselmans, S. J. C. Yates, R. Barends, Y. J. Y. Lankwarden, J. R. Gao, H. Hoevers, and T. M. Klapwijk, Noise and sensitivity of aluminum kinetic inductance detectors for sub-mm astronomy, *J. Low Temp. Phys.* **151**, 524 (2008).
- [46] T. Guruswamy, D. J. Goldie, and S. Withington, Quasiparticle generation efficiency in superconducting thin films, *Supercond. Sci. Technol.* **27**, 055012 (2014).
- [47] P. J. de Visser, S. J. C. Yates, T. Guruswamy, D. J. Goldie, S. Withington, A. Neto, N. Llombart, A. M. Baryshev, T. M. Klapwijk, and J. J. A. Baselmans, The non-equilibrium response of a superconductor to pair-breaking radiation measured over a broad frequency band, *Appl. Phys. Lett.* **106**, 252602 (2015).
- [48] Allen Rothwarf and B. N. Taylor, Measurement of Recombination Lifetimes in Superconductors, *Phys. Rev. Lett.* **19**, 27 (1967).
- [49] C. M. Wilson and D. E. Prober, Quasiparticle number fluctuations in superconductors, *Phys. Rev. B* **69**, 094524 (2004).
- [50] D. C. Mattis and J. Bardeen, Theory of the anomalous skin effect in normal and superconducting metals, *Phys. Rev.* **111**, 412 (1958).
- [51] J. Gao, J. Zmuidzinas, A. Vayonakis, P. Day, B. Mazin, and H. Leduc, Equivalence of the effects on the complex conductivity of superconductor due to temperature change and external pair breaking, *J. Low Temp. Phys.* **151**, 557 (2008).
- [52] R. Barends, S. van Vliet, J. J. A. Baselmans, S. J. C. Yates, J. R. Gao, and T. M. Klapwijk, Enhancement of quasiparticle recombination in ta and al superconductors by implantation of magnetic and nonmagnetic atoms, *Phys. Rev. B* **79**, 020509 (2009).
- [53] Steven de Rooij, Master's thesis, Delft University of Technology, 2020, <http://resolver.tudelft.nl/uuid:1d935dc8-58d1-438f-995b-872e9d5f077e>.
- [54] S. B. Kaplan, C. C. Chi, D. N. Langenberg, J. J. Chang, S. Jafarey, and D. J. Scalapino, Quasiparticle and phonon lifetimes in superconductors, *Phys. Rev. B* **14**, 4854 (1976).
- [55] Steven B. Kaplan, Acoustic matching of superconducting films to substrates, *J. Low Temp. Phys.* **37**, 343 (1979).
- [56] S. Y. Savrasov and D. Y. Savrasov, Electron-phonon interactions and related physical properties of metals from linear-response theory, *Phys. Rev. B* **54**, 16487 (1996).
- [57] Private communication with A. G. Kozorezov.
- [58] T. Kühn, D. V. Anghel, J. P. Pekola, M. Manninen, and Y. M. Galperin, Heat transport in ultrathin dielectric membranes and bridges, *Phys. Rev. B* **70**, 125425 (2004).
- [59] C. Wang, Y. Y. Gao, I. M. Pop, U. Vool, C. Axline, T. Brecht, R. W. Heeres, L. Frunzio, M. H. Devoret, G. Catelani, L. I. Glazman, and R. J. Schoelkopf, Measurement and control of quasiparticle dynamics in a superconducting qubit, *Nat. Commun.* **5**, 5836 (2014).
- [60] S. Y. Hsieh and James L. Levine, Diffusion of Quasiparticles in Superconducting Aluminum Films, *Phys. Rev. Lett.* **20**, 1502 (1968).
- [61] S. Friedrich, K. Segall, M. C. Gaidis, C. M. Wilson, D. E. Prober, A. E. Szymkowiak, and S. H. Moseley, Experimental quasiparticle dynamics in a superconducting, imaging x-ray spectrometer, *Appl. Phys. Lett.* **71**, 3901 (1997).
- [62] P. J. de Visser, D. J. Goldie, P. Diener, S. Withington, J. J. A. Baselmans, and T. M. Klapwijk, Evidence of a Nonequilibrium Distribution of Quasiparticles in the Microwave Response of a Superconducting Aluminum Resonator, *Phys. Rev. Lett.* **112**, 047004 (2014).
- [63] L. Cardani, N. Casali, A. Cruciani, H. le Sueur, M. Martinez, F. Bellini, M. Calvo, M. G. Castellano, I. Colantoni, C. Cosmelli, A. D'Addabbo, S. Di Domizio, J. Goupy, L. Minutolo, A. Monfardini, and M. Vignati, Al/Ti/Al phonon-mediated KIDs for UV-vis light detection over large areas, *Supercond. Sci. Technol.* **31**, 075002 (2018).
- [64] K. Rostem, P. J. de Visser, and E. J. Wollack, Enhanced quasiparticle lifetime in a superconductor by selective blocking of recombination phonons with a phononic crystal, *Phys. Rev. B* **98**, 014522 (2018).
- [65] T. A. Puurtinen, K. Rostem, P. J. de Visser, and I. J. Maasilta, A composite phononic crystal design for quasiparticle lifetime enhancement in kinetic inductance detectors, *J. Low Temp. Phys.* **199**, 577 (2020).
- [66] L. Cardani, N. Casali, I. Colantoni, A. Cruciani, F. Bellini, M. G. Castellano, C. Cosmelli, A. D'Addabbo, S. Di Domizio, M. Martinez, C. Tomei, and M. Vignati, High sensitivity phonon-mediated kinetic inductance detector with combined amplitude and phase read-out, *Appl. Phys. Lett.* **110**, 033504 (2017).
- [67] A. G. Kozorezov, C. Lambert, F. Marsili, M. J. Stevens, V. B. Verma, J. P. Allmaras, M. D. Shaw, R. P. Mirin, and Sae Woo Nam, Fano fluctuations in superconducting-nanowire single-photon detectors, *Phys. Rev. B* **96**, 054507 (2017).
- [68] K. E. Petersen, Silicon as a mechanical material, *Proc. IEEE* **70**, 420 (1982).
- [69] Pieter J. de Visser, Steven A. H. de Rooij, Vignesh Murugesan, David J. Thoen, and Jochem J. A. Baselmans, Data for: "Phonon-trapping-enhanced energy resolution in superconducting single-photon detectors," <https://doi.org/10.5281/zenodo.5499985> (2021).