

Ultra-sensitive THz microwave kinetic inductance detectors for future space telescopes

Baselmans, J. J.A.; Facchin, F.; Laguna, A. Pascual; Bueno, J.; Thoen, D. J.; Murugesan, V.; Llombart, N.; De Visser, P. J.

DOI

[10.1051/0004-6361/202243840](https://doi.org/10.1051/0004-6361/202243840)

Publication date

2022

Document Version

Final published version

Published in

Astronomy and Astrophysics

Citation (APA)

Baselmans, J. J. A., Facchin, F., Laguna, A. P., Bueno, J., Thoen, D. J., Murugesan, V., Llombart, N., & De Visser, P. J. (2022). Ultra-sensitive THz microwave kinetic inductance detectors for future space telescopes. *Astronomy and Astrophysics*, 665, Article A17. <https://doi.org/10.1051/0004-6361/202243840>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.


Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Ultra-sensitive THz microwave kinetic inductance detectors for future space telescopes

J. J. A. Baselmans^{1,2} , F. Facchin², A. Pascual Laguna^{1,2}, J. Bueno¹, D. J. Thoen², V. Murugesan¹, N. Llombart², and P. J. de Visser^{1,2}

¹ SRON – Netherlands Institute for Space Research, Niels Bohrweg 4, 2333 CA - Leiden, The Netherlands
e-mail: J.Baselmans@sron.nl

² THz Sensing group, Microelectronics Department, Delft University of Technology, Mekelweg 1, 2628 CD - Delft, The Netherlands

Received 22 April 2022 / Accepted 14 June 2022

ABSTRACT

Aims. Future actively cooled space-borne observatories for the far-infrared, loosely defined as a 1–10 THz band, can potentially reach a sensitivity limited only by background radiation from the Universe. This will result in an increase in observing speed of many orders of magnitude. A spectroscopic instrument on such an observatory requires large arrays of detectors with a sensitivity expressed as a noise equivalent power $NEP = 3 \times 10^{-20} \text{ W}/\sqrt{\text{Hz}}$.

Methods. We present the design, fabrication, and characterisation of microwave kinetic inductance detectors (MKIDs) for this frequency range reaching the required sensitivity. The devices are based on thin-film NbTiN resonators which use lens-antenna coupling to a submicron-width aluminium transmission line at the shorted end of the resonator where the radiation is absorbed. We optimised the MKID geometry for a low NEP by using a small aluminium volume of $\approx 1 \mu\text{m}^3$ and fabricating the aluminium section on a very thin (100 nm) SiN membrane. Both methods of optimisation also reduce the effect of excess noise by increasing the responsivity of the device, which is further increased by reducing the parasitic geometrical inductance of the resonator.

Results. We measure the sensitivity of eight MKIDs with respect to the power absorbed in the detector using a thermal calibration source filtered in a narrow band around 1.5 THz. We obtain a $NEP_{\text{exp}}(P_{\text{abs}}) = 3.1 \pm 0.9 \times 10^{-20} \text{ W}/\sqrt{\text{Hz}}$ at a modulation frequency of 200 Hz averaged over all measured MKIDs. The NEP is limited by quasiparticle trapping.

Conclusions. The measured sensitivity is sufficient for spectroscopic observations from future, actively cooled space-based observatories. Moreover, the presented device design and assembly can be adapted for frequencies up to ≈ 10 THz and can be readily implemented in kilopixel arrays.

Key words. instrumentation: detectors – methods: miscellaneous

1. Introduction

Radiation in the far-infrared (FIR) part of the electromagnetic spectrum, loosely defined as the 0.03–1 mm wavelength range, represents about half the energy generated in the Universe since the Big Bang and includes information from processes mostly invisible at other wavelengths (Dole et al. 2006). Unfortunately, observations are notoriously difficult: The Earth's atmosphere is at best partially transparent. Even at the altitudes of stratospheric balloons, line emission from residual water vapour can contaminate astrophysical emission lines. Additionally, self-emission from the telescope creates significant radiation loading, limiting instrument sensitivity. Only an actively cooled space telescope with a temperature of about 4 K in combination with background-limited detectors allows an instrument that is limited only by the Universe background. A spectroscopic instrument on such an observatory requires large-format detector arrays with a pixel count of $\approx 10^5$ with a sensitivity expressed in a noise equivalent power (NEP) of $\sim 3 \times 10^{-20} \text{ W}/\sqrt{\text{Hz}}$ (Farrah et al. 2019; Hailey-Dunsheath et al. 2021). The combination of sensitivity, high radiation frequency, and pixel count presents a major challenge for future detector systems. Several detector systems exist that are progressing towards this goal.

Transition edge sensors (TESs; Irwin & Hilton 2005) are used in many ground-based and balloon-borne observatories. Recent work (Audley et al. 2016; Khosropanah et al. 2016; Williams et al. 2020) demonstrates the progress in horn-coupled devices operating in a 60–110 μm wavelength band reaching a $NEP \sim 3 \times 10^{-19} \text{ W}/\sqrt{\text{Hz}}$ using a thermal calibration source (Williams et al. 2020). Nagler et al. (2020) suggest a possible route towards even better sensitivities, reaching an electrical $NEP < 10^{-20} \text{ W}/\sqrt{\text{Hz}}$ by means of a proximity-effect TES.

Quantum capacitance detectors (QCDs; Echternach et al. 2013) are a less mature technology, but are the most sensitive: they have shown single-photon counting at 1.5 THz and background limited performance at power levels as low as 10^{-20} W , corresponding to a $NEP < 10^{-20} \text{ W}/\sqrt{\text{Hz}}$ using a 451 pixel array of antenna-coupled detectors operating at 1.5 THz (Echternach et al. 2021).

Microwave kinetic inductance detectors (MKIDs) are microwave resonators whose resonance frequency depends on the amount of radiation absorbed (Day et al. 2003). Antenna-coupled MKIDs have reached sensitivities of $NEP = 3 \times 10^{-19} \text{ W}/\sqrt{\text{Hz}}$ for a 961 pixel array operating around a radiation frequency of 850 GHz, read out by a single readout system

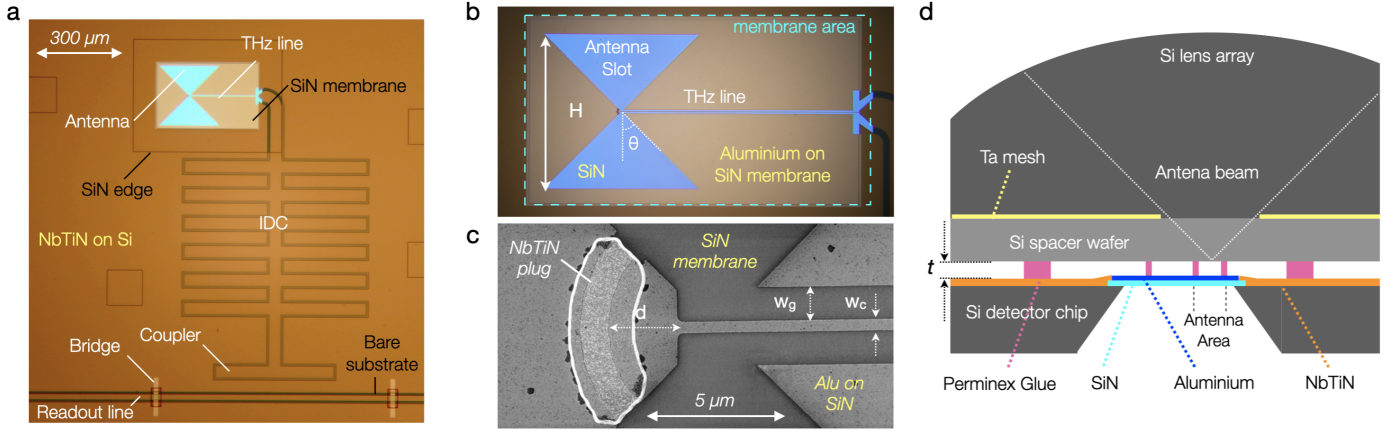


Fig. 1. Detector geometry. (a) Micrograph of a single detector, consisting of a NbTiN CPW line loaded with an IDC coupled to the readout line via a coupling structure. Its shorted end consists of an aluminium section fabricated on a thin SiN membrane, which is highlighted by the backlighting in the micrograph. (b) Zoom onto the aluminium section on the SiN membrane with the leaky-slot antenna coupled to a co-planar waveguide line (referred to as THz line), where radiation absorption takes place. (c) Scanning electron microscope image of the antenna centre showing the THz line in detail and the NbTiN quasiparticle plug. The THz line dimensions are: $w_g = 1.2 \mu\text{m}$ and $w_c = 0.4 \mu\text{m}$. (d) Cross-sectional diagram of the detector assembly (not to scale). The detector chip, as depicted in panels a–c, is coupled to a Si lens array using a spacer wafer with a Ta absorbing mesh, with an opening aligned to the antenna to enable radiation coupling. The vacuum gap t is created using spin-on PermiNex[®] glue pillars as indicated.

(Baselmans et al. 2017). Operation at radiation frequencies above 1 THz with a similar sensitivity and array size was demonstrated by Bueno et al. (2017, 2018). Lumped-element KIDs have almost reached similar sensitivities for devices coupled to an on-chip filterbank (McGeehan et al. 2018). MKIDs are especially attractive because of their ease of multiplexing, allowing kilopixel arrays with only a single readout line. They are also the only detectors whose sensitivity to ionising radiation has been tested extensively for large arrays with relevant sensitivities (Karatsu et al. 2019), with an extrapolated dead time in L2 of less than 4% depending on array geometry. Additionally, Karatsu et al. (2016) found negligible changes in device performance after irradiating the devices with a proton beam simulating the worst-case scenario for radiation absorption of a five-year observation in L2.

In this paper, we present a MKID design that is scalable for a 0.1–10 THz frequency band and for which we measure a $\text{NEP}_{\text{exp}}(P_{\text{abs}}) = 3.1 \pm 0.9 \times 10^{-20} \text{ W}/\sqrt{\text{Hz}}$ at a modulation frequency of 200 Hz averaged over all devices, using a thermal calibration source band-pass filtered around 1.5 THz. Here, P_{abs} is the THz power absorbed in the detector. The design is optimised for low NEP and high responsivity, and will be easily scalable to kilopixel arrays because the readout requirements and fabrication are similar to earlier large system demonstrators (Baselmans et al. 2017).

2. Device design

We explain the device design referring to Fig. 1: the detector is an antenna-coupled hybrid MKID made from a wide niobium-titanium-nitride (NbTiN) coplanar waveguide (CPW) section loaded with an interdigitated capacitor (IDC) fabricated on a silicon wafer. At the shorted end of the CPW, the MKID consists of an aluminium section on a silicon-nitride (SiN) membrane of 100 nm in thickness. The latter has a leaky-slot antenna and aluminium CPW line (hereafter referred to as the THz line) to couple to and absorb the THz radiation. The entire structure is a microwave resonator with a resonance frequency $F_0 \approx 3 \text{ GHz}$. Radiation absorption in the aluminium THz line modifies the Cooper pair- and quasiparticle density which results

in a frequency shift of the resonator read-out using homodyne detection at F_0 (Day et al. 2003). Aluminium is chosen as the active material for its superior performance in terms of NEP with respect to other materials.

The first step in optimising the MKID design is to minimise NEP_{GR} , which is the fundamental limit in MKID sensitivity, and is given by (Day et al. 2003; de Visser et al. 2014):

$$\text{NEP}_{\text{GR}} = \frac{2\Delta}{\eta_{\text{pb}}} \sqrt{\frac{n_{\text{qp}} V}{\tau_{\text{R}}^*}}, \quad (1)$$

where Δ is the energy gap of the superconductor, n_{qp} the density of thermally excited quasiparticles, τ_{R}^* the experimentally observed quasiparticle recombination time (Flanigan et al. 2016), and V the volume of the central strip of the THz line. To minimise NEP_{GR} , we aim to reduce the volume (Hailey-Dunsheath et al. 2021), which implies that the THz line must be short while still absorbing all the power coupled in from the antenna. This requires a thin and narrow central line. We use a central conductor width $w_c = 0.4 \mu\text{m}$ (Fig. 1c) fabricated from a sputter-deposited 16 nm thick aluminium film with a sheet resistance of $R_s = 4.1 \Omega/\square$, corresponding to a resistivity $\rho = 6.5 \times 10^{-8} \Omega\text{m}$, and critical temperature $T_c = 1.54 \text{ K}$. To match the line to the input impedance of the antenna, $Z_{\text{antenna}} = 141 + 36i \Omega$, obtained from simulations in CST[®], we use a gap width $w_g = 1.2 \mu\text{m}$, resulting in a characteristic impedance of the THz line of $Z_{\text{THz line}} = 194 - 103i \Omega$. The high imaginary component is caused by the high aluminium sheet resistance. Despite this, we find a good match between the antenna and the THz line, with the reflection coefficient $|\mathcal{S}_{11}| = -14.5 \text{ dB}$. Subsequently we use simulations in SONNET[®] based upon the method described by Endo et al. (2020) to calculate the fraction of the power absorbed in the THz central line with respect to the total power absorbed. We find a high value of 91.5% caused by the high line impedance. Most of the remaining power, 8.2%, is absorbed in the ground plane, and the remaining losses are due to SiN absorption and re-radiation. We note that quasiparticles created in the ground plane do not contribute to the detector response because they diffuse away from the THz

line. A THz line length of 50 μm is sufficient to absorb 90% of the radiation, the minimum THz line length in our devices is 113 μm .

We further reduce the aluminium volume by using a single THz line coupled to the antenna, in contrast to the dual THz line design from [Bueno et al. \(2017\)](#). To prevent quasiparticle diffusion from the central line to the ground plane, we use a small NbTiN plug creating an Andreev mirror close to the antenna feed point ([Andreev 1964](#)), as shown in Fig. 1c. The distance $d = 2.5 \mu\text{m}$ in Fig. 1c is motivated by two factors. First, the THz current density, and with that the ohmic losses in the NbTiN plug, strongly depend on the distance d from the start of the THz line, which we simulate in CST[®]. For $d > 2.5 \mu\text{m}$, we find that the losses become independent of d and are limited only by losses in the aluminium film. Second, quasiparticle creation is most intense at the start of the narrow section of the THz line where the current density peaks. Initial quasiparticle excitations will relax to the gap edge of aluminium in about $t_{II} = 1.3 \text{ ns}$ ([Kozorezov et al. 2000](#)), which corresponds to a diffusion distance $d = \sqrt{\frac{t_{II}}{2\rho N_0 e^2}} \approx 1.9 \mu\text{m}$, using N_0 the single particle density of states in aluminium $N_0 = 1.08 \times 10^{47} \text{ m}^{-3} \text{ J}^{-1}$ and e the electron charge. Hence, d must be larger than this value to prevent quasiparticles created at the start of the narrow section, which have not yet relaxed to the gap energy of the aluminium to escape over the NbTiN Andreev barrier.

A low NEP_{GR} alone is not sufficient for the detector to actually reach this sensitivity experimentally. It is also crucial that the measured noise, and thereby the NEP, is dominated by photon fluctuations and not by excess noise sources. We therefore maximise the photon noise power spectral density as measured with an MKID, which is given by ([de Visser et al. 2014](#))

$$S_{\theta}^P = 2P_{\text{abs}} hF \frac{(d\theta/dP_{\text{abs}})^2}{1 + (2\pi f \tau_{\text{R}}^*)^2} \quad (2)$$

when using the MKID phase θ as observable, as is the case in this work. Here, we omit the photon bunching term which is negligible for the power and frequency used in this work. The first term represents the photon number fluctuations, with F the photon frequency, h Planck's constant, f the modulation frequency, and $d\theta/dP_{\text{abs}}$ the MKID phase responsivity to the absorbed power. It is useful to reformulate this equation in terms of quasiparticle excitations, as an MKID measures photon fluctuations indirectly by measuring the quasiparticles created. Using $dN_{\text{qp}}/dP_{\text{abs}} = \tau_{\text{R}}^* \eta_{\text{pb}}/\Delta$, $P_{\text{abs}} = N_{\text{qp}} \Delta / (2\tau_{\text{R}}^* \eta_{\text{pb}})$ ([Flanigan et al. 2016](#)) and the Mattis-Bardeen result for $d\theta/dn_{\text{qp}}$ ([Mattis & Bardeen 1958](#)), we get

$$S_{\theta}^P = \frac{n_{\text{qp}} \tau_{\text{R}}^* \eta_{\text{pb}}}{1 + (2\pi f \tau_{\text{R}}^*)^2} \frac{hFV}{\Delta} \cdot \left(-\frac{\alpha\beta Q}{|\sigma|V} \frac{d\sigma_2}{dn_{\text{qp}}} \right)^2, \quad (3)$$

where the first two terms express the fluctuations in quasiparticle number due to photon fluctuations and the last term represents the MKID responsivity to quasiparticle number changes, $d\theta/dN_{\text{qp}}$ where $N_{\text{qp}} = n_{\text{qp}} V$ is the total number of quasiparticles. Furthermore, $\beta = 1 + \frac{2t/\lambda}{\sin h2t/\lambda} \sim 2$ for thin films, with λ the bulk magnetic penetration depth and t the film thickness, $|\sigma|$ the absolute value of the aluminium surface conductance $\sigma = \sigma_1 - i\sigma_2$, which depends on the energy gap, quasiparticle density, and microwave frequency ([Mattis & Bardeen 1958](#)). Q is the resonator quality factor, $\alpha = L_{k,c}/L_{\text{tot}}$, which is the ratio of the kinetic inductance of the THz central line with respect to the total inductance. We maximise S_{θ} in three ways: (i) We

increase both τ_{R}^* and η_{pb} by placing the THz line on a very thin membrane ([Fyhrie et al. 2016](#)), in contrast with the 1 μm thick membrane used by [Bueno et al. \(2017\)](#). This enhances the re-trapping of 2Δ phonons emitted by quasiparticle recombination. The result is an increase in the phonon escape time τ_{esc} which increases the experimental quasiparticle recombination time $\tau_{\text{R}}^* = 0.5 \cdot \tau_{\text{R}} \cdot (1 + \tau_{\text{esc}}/\tau_{\text{B}})$. Here τ_{R} is the single particle recombination time as defined by [Kaplan et al. \(1976\)](#) and $\tau_{\text{B}} = 0.26 \text{ ns}$ is the pair-breaking time in aluminium ([Guruswamy et al. 2014](#)). The membrane thickness is approximately 100 nm, which is limited by its mechanical strength. We find $\tau_{\text{esc}}/\tau_{\text{B}} = 0.65$ using a simple phonon-trapping model ([de Visser et al. 2021](#)), which is a factor four increase with respect to a solid substrate. A larger ratio of $\tau_{\text{esc}}/\tau_{\text{B}}$ also increases η_{pb} ; we estimate $\eta_{\text{pb}} = 0.37$ based on the work of [Guruswamy et al. \(2014\)](#), which will be smaller for the same THz line on a solid substrate. (ii) We minimise the THz central line volume V as discussed before, and (iii) we maximise the kinetic inductance fraction α . For the kinetic inductance of the THz line itself, this is automatically achieved by using a narrow and thin THz line. However, α in Eq. (2) is defined with respect to the entire resonator and a $\lambda/4$ distributed resonator has a significant geometric inductance. To reduce the geometric inductance, we include a large IDC to ground in the MKID as shown in Fig. 1a, following [Noroozian et al. \(2009\)](#). The IDC-section has a much larger capacitance-to-inductance ratio than a CPW line, and therefore the same resonance frequency is obtained with less parasitic inductance in the NbTiN section. The effect is greatest for the shortest THz lines: $\alpha \approx 0.6$ for a 113 μm -long THz line, while a device without the IDC would have $\alpha \approx 0.42$. This is a relatively small difference because the THz line itself has a very high (kinetic) inductance due to its small dimensions and high sheet resistance.

The last step in our effort to maximise the photon noise (and GR noise) visibility is to reduce the TLS noise, which we achieve by widening the capacitive section of the resonator, including the IDC ([Gao et al. 2008](#)). We use a central line width $w = 40 \mu\text{m}$ and a gap width of $g = 8 \mu\text{m}$. Wider structures become impractically large because of the onset of self-resonances within the IDC itself. Using a narrower gap increases the IDC capacitance, but will increase the TLS noise ([Wenner et al. 2011](#)). We therefore limit the ratio $w/(w + 2g)$ to about 0.70. We note that an increase in Q does not provide a useful optimisation, because both the photon noise level and the TLS noise level scale with Q^2 .

For the radiation coupling, we use a leaky wave antenna ([Neto 2010](#)) because this antenna allows for relatively large feature sizes, enabling high radiation frequencies ([Yurduseven et al. 2016](#)), and because it must be fabricated on a membrane, compatible with our design optimisation discussed above. The antenna feed is placed in the second focus of a synthesised elliptical Si lens with a diameter of 1.550 mm. The antenna uses a small vacuum gap $t \ll \lambda$ to launch a leaky wave into the Si ([Neto 2010](#)). The antenna design is optimised for 1.5 THz resulting in a slot opening angle of $\theta = 45^\circ$, a slot length of $H = 200 \mu\text{m}$, and a vacuum gap of $t = 4 \mu\text{m}$, as shown in Figs. 1b,d. The entire design of the KID can be optimised for frequencies up to $\sim 10 \text{ THz}$ by changing the leaky slot dimensions and vacuum gap height.

3. Fabrication and assembly

The device fabrication is largely identical to that of [Bueno et al. \(2017\)](#) and we use UV lithography and dry etching to

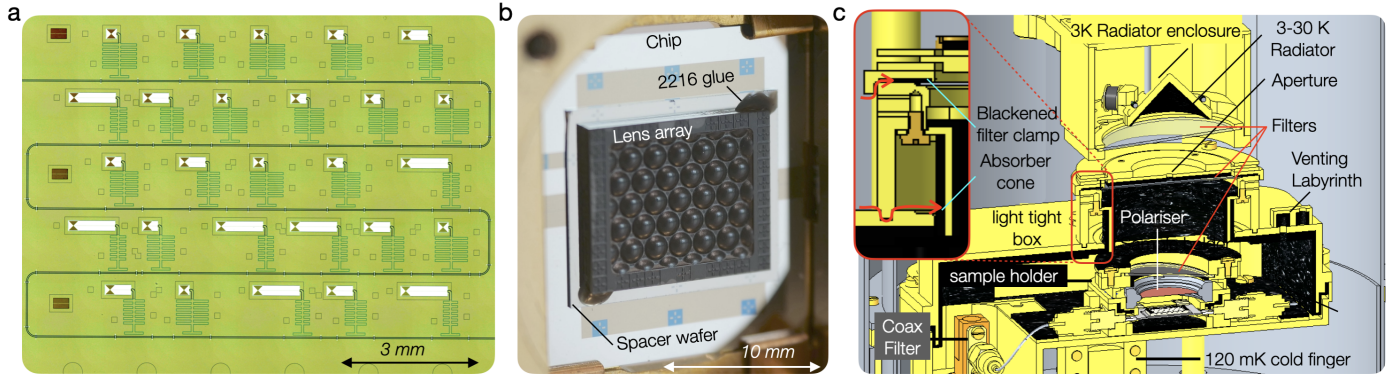


Fig. 2. Detector chip and experimental setup. (a) A microscope image of the detector chip prior to mounting of the lens array, showing the 27 different detectors and the four different geometries used, which vary in THz line length. We note that the centre pixel is one of the designs with the shortest THz line length. (b) Photograph of the detector assembly after lens array and spacer wafer mounting. (c) Cross-sectional CAD render of the experimental setup. The upgrades with respect to the setup of [de Visser et al. \(2014\)](#) are highlighted in the insert.

define most of the structures. We start with a $\rho = 10 \text{ k}\Omega\text{cm}$ Si wafer on which we deposit 150 nm SiN on both sides using low-pressure chemical vapour deposition (LPCVD). The SiN is patterned to define the membrane areas on the front side of the wafer and etch masks on the back side, which are used in the last fabrication step to remove the Si, creating a free-standing membrane. Subsequently, we sputter deposit 207 nm NbTiN ($T_c = 15 \text{ K}$, $\rho = 228 \times 10^{-8} \Omega\text{m}$) to define the resonators. The NbTiN etch requires a short over-etch which removes approximately 50 nm from the SiN. To define the THz line and antenna we sputter deposit a 16 nm aluminium film and pattern it in two steps: First we perform UV contact lithography and wet etching to define the readout line bridges and a large patch of aluminium covering the membrane area. We then use electron beam lithography and wet etching to define the final patterns, similarly to [Mirzaei et al. \(2020\)](#), resulting in a high-resolution definition of the structures as shown in Fig. 1c. The small black structures on the aluminium are small voids created during the 15-min 175°C bake of the PMMA e-beam resist, likely caused by thin-film dewetting ([Thompson 2012](#)). In the last step, we open the membranes by potassium-hydroxide (KOH) etching (8 h in 75°C KOH:H₂O 1:4), followed directly with a RCA2 clean (10 min in 70°C HCl:H₂O₂:H₂O 1:1:5) to remove potassium ions on the back side of the membrane.

The assembled detector array consists of the device chip, a 250 μm Si spacer wafer, and a laser-machined Si lens array commercially obtained from Veldlaser. A cross-sectional sketch of the detector assembly is given in Fig. 1d, and a photograph is shown in Fig. 2b. The spacer wafer top has a 40 nm β -Ta absorbing mesh ($R_s = 51 \Omega/\square$) with a design similar to that used by [Baselmans et al. \(2017\)](#) that acts as a low-pass filter at the MKID readout frequencies and as THz radiation absorber to eliminate stray radiation. Above each antenna, a 600 μm diameter hole allows radiation coupling to the antenna. On the other side, the spacer wafer has 4 μm thick lithographically defined pillars made from PermiNex[®], a spin-on photo-sensitive glue used to define the vacuum gap t as indicated in Fig. 1d and to provide a permanent bond between the chip and the spacer wafer. To assemble the final detector, we align the spacer wafer with the lens array and glue them together with two spots of 2216 epoxy (see Fig. 2b), which is cured for 30 minutes at 90°C . At this temperature the PermiNex[®] is unaffected. In the next step, we align the lens-spacer assembly to the chip and press the assembly together using a spring-loaded tool. The PermiNex[®] is cured for 15 minutes at 180°C to create a permanent bond between spacer

and chip, which is found to be resilient against more than 20 thermal cycles down to 120 mK. A photograph of the entire detector chip without the lens array is given in Fig. 2a. The chip has 27 detectors, with four different device designs that differ in THz line length, which can be 913, 663, 238, or 113 μm . This corresponds to $V = 6.0, 4.4, 1.6,$ and $0.85 \mu\text{m}^3$ respectively, including the small volume near the NbTiN plug. All devices are designed with a coupling Q factor $Q_c = 120 \times 10^3$.

4. Experiment

In the experiment, we measure the sensitivity, $\text{NEP}_{\text{exp}}(P_s, f)$, at several powers from a thermal radiator source P_s using the setup depicted in Fig. 2. We denote with f the post-detection modulation frequency of the spectral shape of the NEP. We mount the detector assembly in a holder which itself is mounted inside a light-tight box mechanically anchored to the cold stage of an adiabatic demagnetisation refrigerator (ADR). All interfaces and venting holes of the holder and light-tight box are equipped with labyrinths coated with a radiation-absorbing layer from carbon-loaded epoxy and SiC grains. A thermal radiation source weakly coupled to the 3 K stage of the cooler enclosed in a 3 K box is placed directly above, and can be heated to 30 K. In the first experiment, we used a 24 mm diameter radiator with a radiation absorbing layer from carbon-loaded epoxy and 0.2 mm SiC grains, as shown in Fig. 2c. Radiation coupling between the source and the detector is achieved through a 1.85 mm aperture at 42 mm from the lens surface. The use of the aperture reduces the throughput to the detectors, and therefore higher temperatures are needed to reach the same amount of source power at the detectors. This reduces the sensitivity of the radiation power to errors in the measured temperature. Furthermore, it allows a check as to whether the coupled power matches the beam pattern of the detector. We use eight filters in total – mounted on the radiator box, the light-tight box, and the sample holder – to define a pass-band of around 1.5 THz. These are the exact same filters as discussed by [de Visser et al. \(2014\)](#). The second experiment was identical, except that we used a 40 mm diameter radiator with 1 mm SiC grains and no aperture, to cross-check our results: In this case, the antenna pattern does not play a role because the opening angle to the radiator is much larger than the beam opening angle. Also, the total power coupled to the detector chip largely exceeds the power coupled to the single-mode antenna beams. Stray coupling to this excess radiation will result in

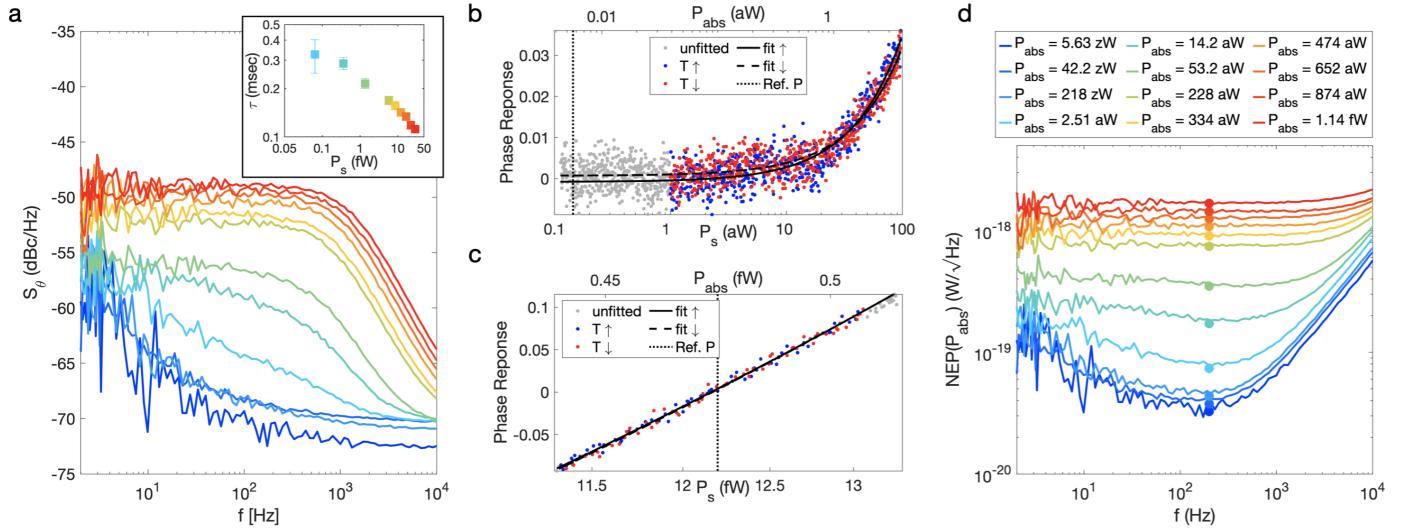


Fig. 3. Noise, response and NEP for a single detector. (a) Measured phase noise power spectral density for several source powers, increasing from bottom to top. There is a clear transition from detector-limited noise (lower curves) to photon-noise-limited noise, characterised by a white spectrum rolled-off with τ_R^* (highest curves). The insert shows the τ_R^* for the same source powers. (b) KID phase response around the lowest source power used to obtain the NEP, $P_s = 0.145$ aW ($P_{abs} = 5.63$ zW), which corresponds to a radiator temperature of 4 K. The data represent a measurement obtained by sweeping the radiator temperature from 3.95 to 6.21 K and back to 3.95 K. The data is plotted on a semi-logarithmic scale to show the absence of any response at low powers. The fits and fitted range are indicated as well. (c) MKID response and fits for increasing and decreasing radiator temperature around a higher source power, $P_s = 12.2$ fW. We indicate the source power as well as the absorbed power in *panels b and c* for convenience. (d) Measured NEP for the same set of powers, now expressed in power absorbed in the detector. The legend is also valid for *panel a*.

overestimation of the coupling efficiency. In all cases, a polariser is used on the holder to allow only radiation in co-polarisation with respect to the antenna. We note that, because of the increased sensitivity of the detectors, we needed to upgrade the light-tight box with respect to the setup in [de Visser et al. \(2014\)](#) – as indicated in the insert of [Fig. 2c](#) – with better stray-light control around the filter mount. We use a CRYOPHY[®] outer shield and niobium inner shield at 3 K surrounding the light-tight box and radiator to limit time-dependent magnetic fields. The theoretical coupling between the absorbing aluminium volume in the MKID aligned with the aperture is calculated using a CST simulation of the antenna beam and the tool provided by [Zhang et al. \(2019\)](#) to be $\eta_{opt,calc.} = 4.3\%$. The reason for the low coupling is the (intended) large spill-over loss between the small pin-hole aperture and the lens–antenna beam pattern. This includes the losses in the ground plane of the THz line and the antenna to THz line mismatch, which are discussed above. For the 40 mm diameter radiator without aperture, we find $\eta_{opt,calc.} = 76\%$.

We began experiments with a frequency sweep with the chip at 120 mK and the radiator at 3 K, which represents fully dark conditions. We find 23 out of 27 MKIDs with an average $Q_i = 0.23 \times 10^6 \pm 0.13 \times 10^6$ and an average $Q_c = 6 \times 10^4 \pm 2 \times 10^4$ at a readout power of -118 dBm. We selected a subset of the detectors that couple well to the aperture. For each detector we obtained $NEP_{exp}(P_s, f)$ using

$$NEP_{exp}(P, f) = \sqrt{S_\theta(f)} \left(\frac{d\theta}{dP} \right)^{-1} \sqrt{1 + (2\pi f \tau_R^*)^2}, \quad (4)$$

where P is the radiation power, given now by P_s , which is the power from the source that can couple to a single mode. P_s is calculated from the radiator temperature and the measured filter transmission using the formalism explained in detail in [Appendix B of Ferrari et al. \(2018\)](#). [Figure 3a](#) shows the phase noise power spectral density $S_\theta(f)$ for one representative device, which has a resonance frequency of 2.74 GHz and an aluminium

length of 913 μm . The spectrum is obtained from 64 s of time domain data sampled at 50 kilosample/sec with the readout tuned to the MKID resonance frequency while maintaining a constant black body temperature. At low powers, shown by the lower (blue) lines, we observe a noise spectrum with a $1/f$ slope. At increasing power, the noise level at frequencies $f < 1$ kHz increases sharply until the noise becomes white for $P_s > 1$ fW. The insert in the figure shows τ_R^* for the same powers obtained from a fit to the cross power spectral density of the resonator amplitude and phase noise ([de Visser et al. 2012](#)), which is given in [Appendix A](#). We observe a saturation of the recombination time at around 0.3 ms and a reduction for increasing powers as expected. Subsequently, we obtain the responsivity $d\theta/dP_s$ by a linear fit to the MKID phase response upon a small change of the black-body temperature. Care was taken to sweep the radiator temperature slowly to prevent hysteresis caused by a temperature difference between the thermometer and the radiating surface. In [Fig. 3b](#), we show the response measurement and fit around the lowest P_s of 0.145 aW (corresponding to a 4 K radiator temperature). We fit the response curve over a range of $1 \text{ aW} < P_s < 0.1 \text{ fW}$ (which corresponds to a range in absorbed power $0.039 \text{ aW} < P_{abs} < 39 \text{ aW}$, as we show later): the high power limit is used so as to reach a statistically significant response which is still linear; the low end was chosen so as to reduce the sensitivity to drifts. It is clear from the figure that there is negligible hysteresis and that the fits are excellent over the whole power range spanning a factor 1000, with no response below $P_s = 1$ aW. This proves that the setup is light tight; radiation leaks would allow unfiltered radiation to couple to the detector. The total integrated power per mode for a 4 K radiator is 10^6 times larger than the power in the narrow band around 1.5 THz, and it is strongly temperature dependent. Even a weak coupling to broad-band stray radiation would therefore result in a significant response of the detector at 4 K ($P_s = 0.145$ aW) in [Fig. 3b](#). For higher powers, we fit a symmetric power range around the power at which we measure the noise, as shown in [Fig. 3c](#).

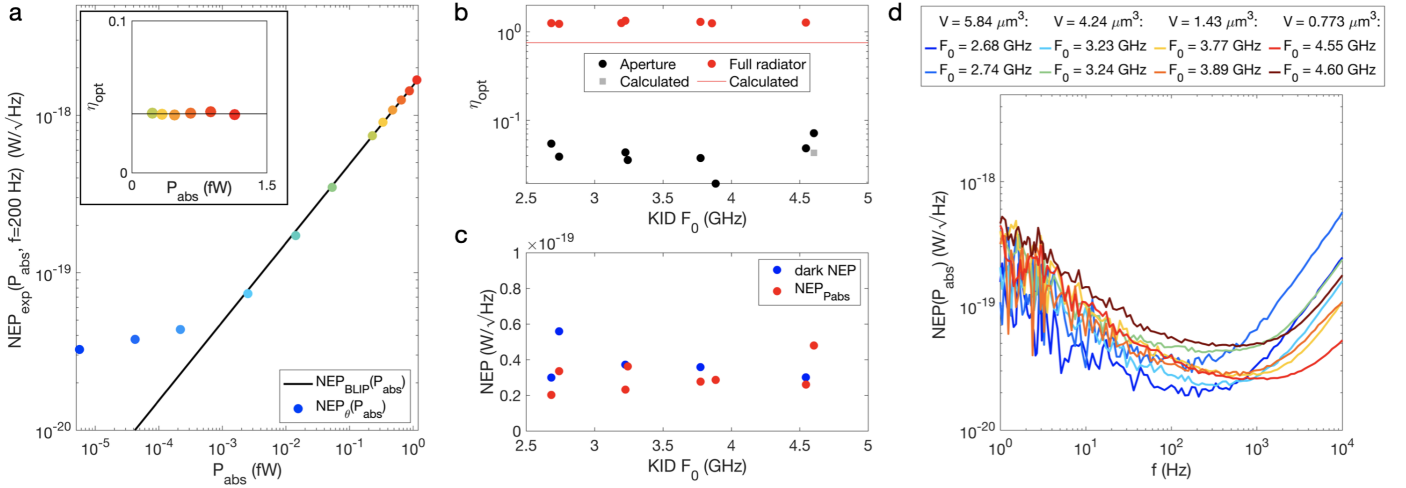


Fig. 4. NEP and coupling efficiency. (a) The NEP at $f = 200$ Hz (see panel d), as function of absorbed power using the pinhole setup for the same device as in Fig. 3. The black line shows the theoretical curve. The absorbed power is evaluated from the source power using the measured efficiency as indicated in the insert, and the mean value of the points, $\eta_{\text{opt}} = 0.0389 \pm 0.004$, is given by the line. The statistical error bars are smaller than the dots. (b) The averaged coupling efficiencies for the two experiments. The red dots represent the results for the large radiator, where all detectors couple almost identically to the source as expected; the calculated coupling is given by the line. The black dots give the coupling for the pinhole setup, where only the best aligned detector is expected to reach the theoretical coupling value given by the grey square. We note that, in both cases, the measured coupling exceeds the theoretical value by the same factor of ~ 1.65 . (c) The measured NEP_{exp}(P_{abs} , $f = 200$ Hz) for the pinhole experiment, as displayed in Fig. 2c, together with the dark NEP obtained at the same modulation frequency using a change in bath temperature to obtain the response. (d) The measured NEP_{exp}(P_{abs} , f) for the pinhole experiment at the lowest radiator temperature for all measured MKIDs. We observe no dependence of the NEP on aluminium volume. We note that panel a and Fig. 3 are given for the detector with $F_0 = 2.74$ GHz.

A possible source of systematic error is the radiator temperature readout: the detector response starts at $P_s = 1$ aW, corresponding to a radiator temperature of 6.3 K, where a 50 mK temperature change would result in a 10% change in P_s . For the full aperture setup, the response starts at a temperature of 4.5 K, which is due to the higher coupling. At this temperature, a 50 mK temperature change will result in a 20% change in P_s . We therefore performed a cross-calibration between the Cernox[®] radiator thermometer and another similar thermometer. The temperature difference between the two is less than 15 mK in the temperature range of the experiment, contributing less than 3% uncertainty in radiator power.

We can now obtain NEP_{exp} using Eq. (4), and use it to determine the detector coupling efficiency η_{opt} using the photon noise from the radiator as an absolute calibration source (Ferrari et al. 2018). This procedure is only valid if the MKID is photon-noise limited, which is the case for source powers exceeding 5 fW; here the spectrum is white with a single Lorentzian roll-off (Janssen et al. 2013). The optical efficiency is given by (Ferrari et al. 2018)

$$\eta_{\text{opt}} = \frac{\int 2P_{s,\nu} h\nu d\nu + \int 4\Delta P_{s,\nu} / \eta_{\text{pb}} d\nu}{NEP_{\text{exp}}^2 - \int 2P_{s,\nu} h\nu F_\nu O_\nu d\nu}, \quad (5)$$

where F_ν is the filter transmission, O_ν is the photon occupation number, and $P_{s,\nu}$ is P_s per unit frequency ν . The result is given in the insert of Fig. 4a. We get $\eta_{\text{opt}} = 0.039 \pm 0.004$ with very little scatter between the individual datapoints. We can now obtain the experimental NEP as a function of the power absorbed, NEP_{exp}(P_{abs} , f), using Eq. (4) with $P = P_{\text{abs}} = \eta_{\text{opt}} P_s$. The result is shown in Fig. 3d. At the lowest powers, the NEP at $f = 200$ Hz reaches a value of NEP_{exp}(P_{abs} , $f = 200$ Hz) = $3.3 \pm 0.3 \times 10^{-20}$ W/√Hz. At lower frequencies, the NEP is higher because of the $1/f$ noise. For higher powers, the NEP spectral shape becomes flatter, and for $P_{\text{abs}} > 53$ aW the NEP

is frequency independent down to 1 Hz. In Fig. 4a, we give NEP_{exp}(P_{abs} , $f = 200$ Hz) together with the theoretical value of the photon noise background from the radiation source and obtain a result very similar to those of de Visser et al. (2014), Baselmans et al. (2017) and Ferrari et al. (2018), but with ten times greater sensitivity. For a comparison of the NEP as a function of the absorbed power as shown in Fig. 4a for all measured MKIDs, we refer to Appendix C. For all MKIDs, the microwave readout power absorbed by the quasiparticle system, $P_{\text{abs},\text{qp}}$, is significant, especially at low absorbed powers: For example, at the lowest value of $P_{\text{abs}} = 5.63$ zW, we estimate that $P_{\text{abs},\text{qp}} \approx 20$ aW. At values of $P_{\text{abs}} \geq 1$ fW, $P_{\text{abs},\text{qp}}$ and P_{abs} become similar.

In Fig. 4b, we give the coupling efficiency for all MKIDs for the aperture-coupled radiator by the black dots, and the theoretical coupling for a well-aligned detector by the grey square. We observe that the MKID with the highest resonance frequency has the maximum coupling, as expected as this is the detector in the chip centre as can be seen in Fig. 2a. The measured $\eta_{\text{opt}} = 7\%$, exceeding the expected value of 4.3% by a factor 1.63. We observe six MKIDs with similar coupling but some scatter, which can be explained by a small misalignment of the aperture, causing unequal coupling to the six detectors at equal distance from the centre device. The red dots represent the results with the large radiator; we now see that all detectors couple identically with a mean value of $\eta_{\text{opt}} = 1.27 \pm 0.04$, exceeding the expected value by a factor 1.68. The fact that both experiments yield an identical coupling with respect to the theoretical value implies that both setups are equivalent and systematic errors such as an error in the beam pattern calculation, in-band stray radiation coupling with the large radiator, or a difference in temperature between the radiator surface and thermometer can be excluded. The higher measured transmission is therefore most likely caused by a higher transmission of our eight quasi-optical filters compared to the data provided by their supplier QMCI[®], which gives a peak transmission of 13.5%. A

potential reason is that QMCI[®] performs the filter measurements at room temperature.

In Fig. 4c we give $\text{NEP}_{\text{exp}}(P_{\text{abs}}, f = 200 \text{ Hz})$ for eight KIDs in the pinhole experiment and find a mean value of $\text{NEP}_{\text{exp}}(P_{\text{abs}}, f = 200 \text{ Hz}) = 3.1 \pm 0.9 \times 10^{-20} \text{ W}/\sqrt{\text{Hz}}$. We also give the dark NEP (Baselmans et al. 2008; Janssen et al. 2014). To obtain the dark NEP, we measure the response of the MKIDs to a change in chip temperature while keeping the radiator in dark conditions, that is, $T_{\text{BB}} = 3 \text{ K}$. Under these conditions, the number of quasiparticles in the aluminium N_{qp} can be calculated from the chip temperature, the volume of the aluminium section of the resonator, and the energy gap (Janssen et al. 2014). Rewriting N_{qp} in terms of FIR power allows the dark NEP to be calculated using Eq. (4) by replacing the responsivity term $d\theta/\delta P$ by the dark responsivity given by Baselmans et al. (2008):

$$\frac{d\theta}{dP_{\text{dark}}} = \frac{\eta_{\text{pb}} \tau_{\text{R}}^*}{\Delta} \frac{d\theta}{dN_{\text{qp}}(T)}, \quad (6)$$

where we obtain τ_{R}^* using a fit to the cross power spectral density of the phase- and amplitude noise as discussed above. For the operating temperature of 120 mK, a good fit cannot always be obtained. We find a mean value of $\text{NEP}_{\text{dark}}, (f = 200 \text{ Hz}) = 3.8 \pm 1.3 \times 10^{-20} \text{ W}/\sqrt{\text{Hz}}$, in good agreement with $\text{NEP}_{\text{exp}}(P_{\text{abs}}, f = 200 \text{ Hz})$. A similar agreement between dark and optical NEP was found by Janssen et al. (2014). Finally, panel d shows the $\text{NEP}_{\text{exp}}(P_{\text{abs}}, f)$ at the lowest radiator power for the same set of MKIDs. All have a similar sensitivity.

5. Discussion

The observed reduction in photon noise level for $P_{\text{abs}} < 228 \text{ aW}$ (see Fig. 3a) is inconsistent with a NEP limited by a thermal reservoir of residual quasiparticles, which was the conclusion in previous work (de Visser et al. 2014). However, it is consistent with quasiparticle trapping as discussed by de Rooij et al. (2021). The optimisation of the MKID responsivity, as discussed in the context of Eq. (3), is now especially relevant. A high responsivity allows for the device to remain background limited even when the photon-noise level drops. The use of a thin membrane, increasing both τ_{R}^* and η_{pb} , is therefore an important optimisation. The downside is a strong reduction in instantaneous dynamic range when compared to the results of Baselmans et al. (2017). We find an instantaneous dynamic range of a few hundred aW, with readout re-tuning the devices operate up to 30 fW absorbed power.

Interestingly, we observe in Fig. 4 that the $\text{NEP}_{\text{exp}}(P_{\text{abs}}, f = 200 \text{ Hz})$ is independent of the aluminium volume. More precisely, we observe that both the noise S_{θ} and responsivity $d\theta/dP$ are independent of volume. For the noise, this is not surprising, because we are limited in this regime by amplifier noise and device noise, but the responsivity is expected to scale with the inverse of the volume, as can be seen from Eqs. (2) and (3). We find from the dark NEP measurement that $d\theta/dN_{\text{qp}}$ does scale inversely with volume as expected. We also find that the low temperature quasiparticle lifetime τ_{R}^* scales inversely with volume. A smaller effect is that α is slightly smaller for the smallest volume devices due to parasitic inductance in the IDC, and the net result is the observed volume independence of the responsivity to power. Interestingly, the observed volume (or line length length) dependence in τ_{R}^* is not seen in the Appendix of de Rooij et al. (2021).

A better understanding of the trapping mechanism and the transition between trapping-dominated hybrid MKIDs and devices limited by quasiparticle creation due to the readout power will be important for further optimisation, which is especially relevant for the $1/f$ noise. If trapping is caused by surface states, as argued by de Rooij et al. (2021), device optimisation is possible using thicker aluminium films as this will reduce trapping and increase τ_{R}^* . Interestingly, thicker films will also increase τ_{R}^* as well as η_{pb} due to enhanced 2Δ phonon recycling caused by the longer dwell time of the phonons in the film. This might partially compensate the responsivity reduction caused by an increase in the volume and reduction in α .

It is interesting to consider the intrinsic energy resolution for the possibility of single photon counting in the FIR. This resolution is given by Moseley et al. (1984): $dE \approx \text{NEP}_{\text{exp}}(P_{\text{abs}}, f = 200 \text{ Hz}) \cdot \sqrt{\tau_{\text{R}}^*}$. We find an energy resolution averaged over all MKIDs of $dE/h = 0.5 \pm 0.2 \text{ THz}$, with a best value of $dE/h = 0.24 \text{ THz}$ for one of the devices with the smallest volume. Reliable single photon detection should therefore be possible for frequencies above a few THz. For more details, we refer to Appendix B.

6. Conclusions

In conclusion, we present a new MKID design, where radiation coupling and absorption take place in a small aluminium volume fabricated on a 100 nm SiN membrane embedded in a NbTiN resonator. The device is optimised for high quasiparticle responsivity and low NEP. We obtain a mean $\text{NEP}_{\text{exp}}(P_{\text{abs}}, f = 200 \text{ Hz}) = 3.1 \pm 0.9 \times 10^{-20} \text{ W}/\sqrt{\text{Hz}}$ with a time constant varying from 0.33 msec to 0.03 msec. The NEP is limited by $1/f$ noise and quasiparticle trapping, limiting the performance in dark conditions at low modulation frequencies. At absorbed powers of 53 aW and higher the device reaches background-limited operation with a flat noise level down to 1 Hz. The antenna design, device geometry, and assembly using PermiNex[®] spin-on glue allows scaling to frequencies up to $\approx 10 \text{ THz}$. Importantly, creating kilopixel arrays from the presented device design is straightforward; the fabrication and readout are largely identical to those in the work presented by Bueno et al. (2018). All presented data, measurement data, and analysis scripts will be made available in Baselmans et al. (2022).

Acknowledgements. P.J.d.V. is financially supported by the Netherlands Organisation for Scientific Research NWO (Veni Grant 639.041.750 and Projectruimte 680-91-127).

References

- Andreev, A. 1964, *Zh. Eksperim. i Teor. Fiz.*, **46**, 1823
- Audley, M. D., de Lange, G., Gao, J.-R., et al. 2016, *Rev. Sci. Instrum.*, **87**, 043103
- Baselmans, J., Yates, S. J. C., Barends, R., et al. 2008, *J. Low Temp. Phys.*, **151**, 524
- Baselmans, J. J. A., Bueno, J., Yates, S. J. C., et al. 2017, *A&A*, **601**, A89
- Baselmans, J., Facchin, F., Pascual Laguna, A., et al. 2022, <https://doi.org/10.5281/zenodo.6607050>
- Bueno, J., Yurduseven, O., Yates, S. J. C., et al. 2017, *Appl. Phys. Lett.*, **110**, 233503
- Bueno, J., Murugesan, V., Karatsu, K., Thoen, D. J., & Baselmans, J. J. A. 2018, *J. Low Temp. Phys.*, **193**, 96
- Day, P. K., LeDuc, H. G., Mazin, B. A., Vayonakis, A., & Zmuidzinas, J. 2003, *Nature*, **425**, 817
- de Rooij, S. A. H., Baselmans, J. J. A., Murugesan, V., Thoen, D. J., & de Visser, P. J. 2021, *Phys. Rev. B*, **104**, L180506
- de Visser, P. J., Baselmans, J. J. A., Yates, S. J. C., et al. 2012, *Appl. Phys. Lett.*, **100**, 162601

- de Visser, P. J., Baselmans, J. J. A., Bueno, J., Llombart, N., & Klapwijk, T. M. 2014, *Nat. Commun.*, **5**, 3130
- de Visser, P. J., de Rooij, S. A. H., Murugesan, V., Thoen, D. J., & Baselmans, J. J. A. 2021, *Phys. Rev. Appl.*, **16**, 034051
- Dole, H., Lagache, G., Puget, J.-L., et al. 2006, *A&A*, **451**, 417
- Echternach, P. M., Stone, K. J., Bradford, C. M., et al. 2013, *Appl. Phys. Lett.*, **103**, 053510
- Echternach, P. M., Beyer, A. D., & Bradford, C. M. 2021, *J. Astron. Teles. Instrum. Syst.*, **7**, 1
- Endo, A., Laguna, A. P., Hähnle, S., et al. 2020, *Proc. SPIE*, **11453**, 114532T
- Farrah, D., Smith, K. E., Ardila, D., et al. 2019, *J. Astron. Teles. Instrum. Syst.*, **5**, 020901
- Ferrari, L., Yates, S. J. C., Eggens, M., Baryshev, A. M., & Baselmans, J. J. A. 2018, *IEEE Trans. Terahertz Sci. Technol.*, **8**, 572
- Flanigan, D., McCarrick, H., Jones, G., et al. 2016, *Appl. Phys. Lett.*, **108**, 083504
- Fyhrie, A., McKenney, C., Glenn, J., et al. 2016, *SPIE Conf. Ser.*, **9914**, 99142B
- Gao, J., Daal, M., Martinis, J. M., et al. 2008, *Appl. Phys. Lett.*, **92**, 212504
- Guruswamy, T., Goldie, D. J., & Withington, S. 2014, *Supercond. Sci. Technol.*, **27**, 055012
- Hailey-Dunsheath, S., Janssen, R. M. J., Glenn, J., et al. 2021, *J. Astron. Teles. Instrum. Syst.*, **7**, 1
- Irwin, K., & Hilton, G. 2005, *Transition-Edge Sensors*, ed. C. Enss (Berlin, Heidelberg: Springer), 63
- Janssen, R. M. J., Baselmans, J. J. A., Endo, A., et al. 2013, *Appl. Phys. Lett.*, **103**, 203503
- Janssen, R. M. J., Endo, A., de Visser, P. J., Klapwijk, T. M., & Baselmans, J. J. A. 2014, *Appl. Phys. Lett.*, **105**, 193504
- Kaplan, S. B., Chi, C. C., Langenberg, D. N., et al. 1976, *Phys. Rev. B*, **14**, 4854
- Karatsu, K., Dominjon, A., Fujino, T., et al. 2016, *J. Low Temp. Phys.*, **184**, 540
- Karatsu, K., Endo, A., Bueno, J., et al. 2019, *Appl. Phys. Lett.*, **114**, 032601
- Khosropanah, P., Suzuki, T., Ridder, M. L., et al. 2016, *SPIE Conf. Ser.*, **9914**, 99140B
- Kozorezov, A. G., Volkov, A. F., Wigmore, J. K., et al. 2000, *Phys. Rev. B*, **61**, 11807
- Mattis, D. C., & Bardeen, J. 1958, *Phys. Rev.*, **111**, 412
- McGeehan, R., Barry, P. S., Shirokoff, E., et al. 2018, *J. Low Temp. Phys.*, **193**, 1024
- Mirzaei, M., Barrentine, E. M., Bulcha, B. T., et al. 2020, *Proc. SPIE (Digital Library)*, **11453**, 114530M
- Moseley, S. H., Mather, J. C., & McCammon, D. 1984, *J. Appl. Phys.*, **56**, 1257
- Nagler, P. C., Sadleir, J. E., & Wollack, E. J. 2020, ArXiv e-prints [arXiv:2012.06543]
- Neto, A. 2010, *IEEE Trans. Antennas Propag.*, **58**, 2238
- Noroozian, O., Gao, J., Zmuidzinas, J., LeDuc, H. G., & Mazin, B. A. 2009, *AIP Conf. Proc.*, **1185**, 148
- Thompson, C. V. 2012, *Ann. Rev. Materials Res.*, **42**, 399
- Wenner, J., Barends, R., Bialczak, R. C., et al. 2011, *Appl. Phys. Lett.*, **99**, 113513
- Williams, E. A., Withington, S., Goldie, D. J., et al. 2020, *Rev. Sci. Instrum.*, **91**, 123104
- Yurduseven, O., Llombart Juan, N., & Neto, A. 2016, *IEEE Trans. Antennas Propag.*, **64**, 3330
- Zhang, H., Dabironezare, S. O., Carluccio, G., Neto, A., & Llombart, N. 2019, in 44th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz), 1

Appendix A: Cross power spectral density and lifetime fits

In Fig. 3a of the main text we show in the insert the apparent quasiparticle recombination time τ_R^* . These datapoints were obtained from the Lorentzian fits to the cross power spectral density of the noise measured at each radiator power. The data and fits that resulted in the values of the recombination time are given in Fig.A.1.

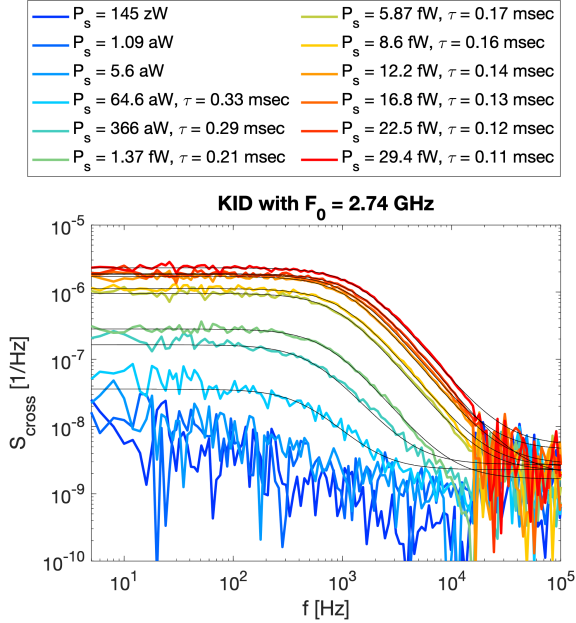


Fig. A.1. Cross power spectral density at each source power and the fits resulting in the apparent recombination time as indicated. Only at the lowest radiator powers does the photon-noise signal disappear and we cannot fit the data with a single Lorentzian roll-off.

Appendix B: Energy resolution

Using $NEP_{exp}(P_{abs}, f = 200 \text{ Hz})$ and the lifetimes obtained from the fits to the photon noise cross power spectral density as discussed above, we can obtain an estimate of the energy resolution, $dE = NEP_{exp}(P_{abs}, f = 200 \text{ Hz}) \cdot \sqrt{\tau_R^*}$. The result is given in Fig. B.1. We find an energy resolution averaged over all MKIDs of $dE/h = 0.5 \pm 0.2 \text{ THz}$ and a clear improvement towards the smaller device volumes, which results from the fact that the smaller devices have a shorter quasiparticle lifetime and a similar NEP.

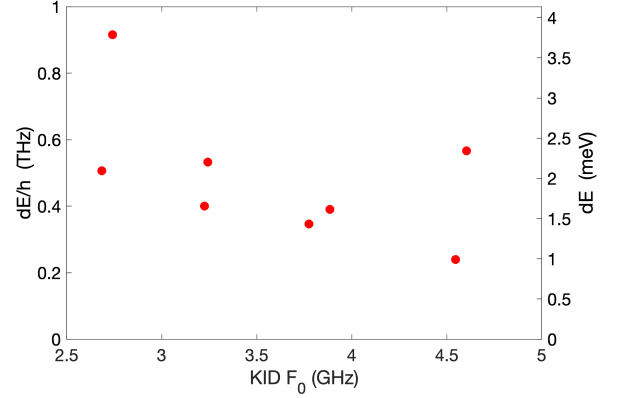


Fig. B.1. Estimated energy resolution for all MKIDs discussed in this work, given as a function of the MKID resonance frequency.

Appendix C: $NEP_{exp}(P_{abs}, f = 200 \text{ Hz})$ as function of absorbed power

In Fig. 4a of the main manuscript we give $NEP_{exp}(P_{abs}, f = 200 \text{ Hz})$ for the MKID with a resonant frequency of 2.74 GHz. The other measured MKIDs show a similar behaviour, as shown in the top left figure of Fig. C.1. However, in the power interval between 1 and 10 aW, we see for most detectors that the measured $NEP_{exp}(P_{abs}, f = 200 \text{ Hz})$ is slightly lower than the photon noise limit. We re-plot the same data normalised to the photon noise limit in the other two panels in the top row to make this more clear. To experimentally address this issue, we repeat the measurement using a large (40 mm diameter) radiator coupled to the detector chip without the small aperture used in the rest of the experiments. We observe an identical behaviour, which shows both the same limited NEP and an NEP lower than the photon noise limit in the same 1-10 aW absorbed power range. This correspondence in power and not in temperature excludes possible issues with the filter stack or thermometer: The coupling efficiency with the large body is 1.27, whereas it is 7% with the aperture, that is, about a factor 16 difference. This implies that the same absorbed power levels are obtained at different radiator temperatures. This is very clear from the rightmost figures: At 6K we observe in the top panel that $NEP_{exp}(P_{abs}, f = 200 \text{ Hz})$ is below the photon noise limit. However, in the bottom panel $NEP_{exp}(P_{abs}, f = 200 \text{ Hz})$ is identical to the background limit. To further exclude issues with the thermometer, we recalibrated our thermometer as discussed in the main text. The exact mechanism of the discussed effect remains uncertain, but it could be related to the fact that the photon arrival rate is 1 kHz for 1 aW, respectively. This is of the order of the bandwidth given by the recombination time, $\frac{1}{2\pi\tau_R^*}$, which is 500 Hz for $\tau_R^*=0.3 \text{ msec}$.

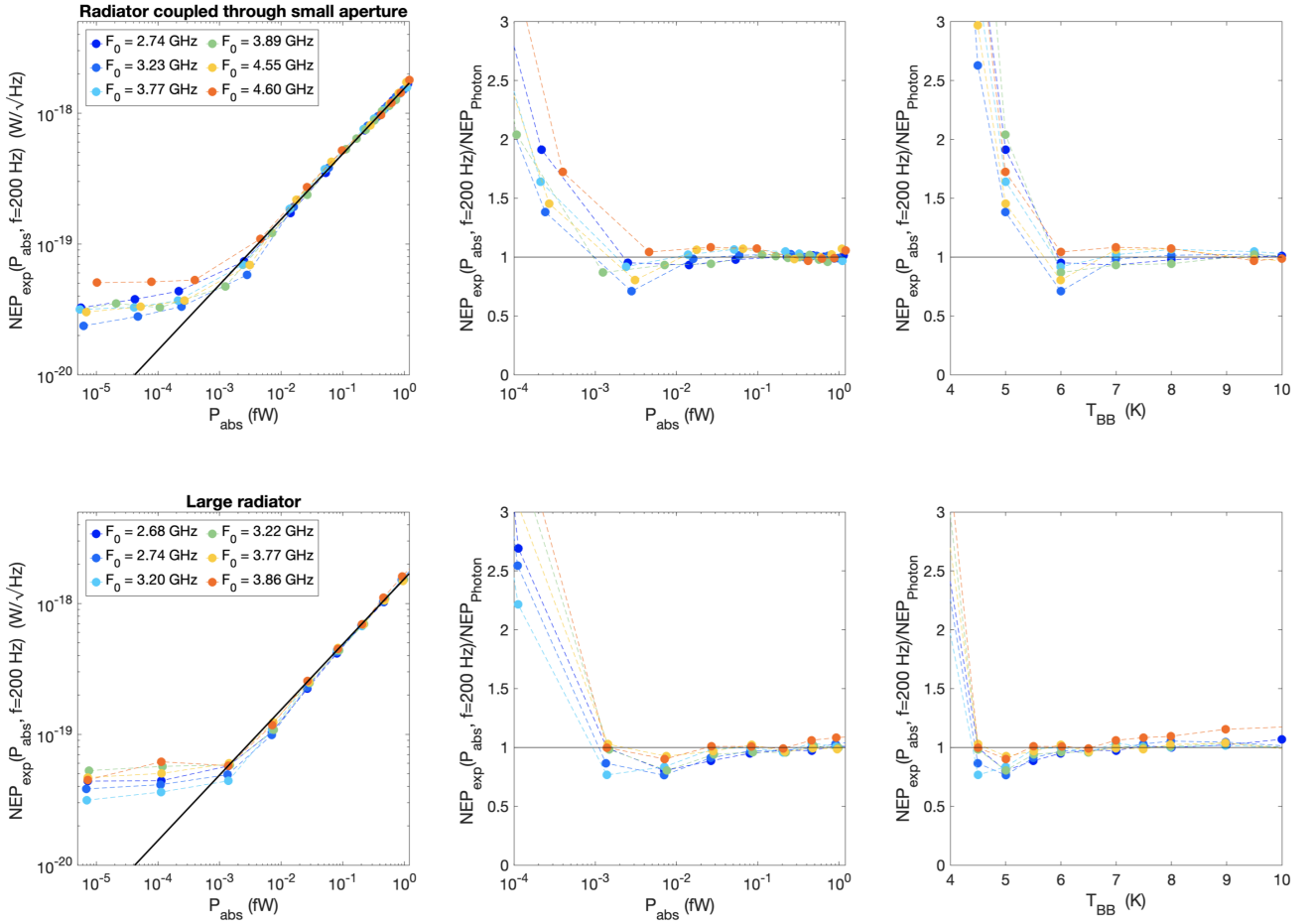


Fig. C.1. NEP as function of absorbed power for all measured devices. Top row: Data of all measured MKIDs using radiation coupling via a small pinhole to the small radiator source. The left figure gives the data for all KIDs identical to Fig3.e in the main text. The centre and right plots give the NEP divided by the photon noise limit, showing more clearly that the measured NEP can be below the photon noise limit for $10 \text{ aW} < P_{abs} < 100 \text{ aW}$. Bottom row: Similar data, but now using the large radiator coupled without pinhole to the detector chip. We observe that the region where the experimental NEP is below the photon noise limit scales with P_{abs} , and not with radiator temperature.