

MIT Open Access Articles

*Superconducting nanowire single-photon detector  
with integrated impedance-matching taper*

The MIT Faculty has made this article openly available. **Please share**  
how this access benefits you. Your story matters.

**Citation:** Zhu, Di et al. "Superconducting nanowire single-photon detector with integrated impedance-matching taper." Applied Physics Letters 114 (January 2019): 22901 © 2019 Author(s)

**As Published:** <http://dx.doi.org/10.1063/1.5080721>

**Publisher:** American Institute of Physics (AIP)

**Persistent URL:** <https://hdl.handle.net/1721.1/121343>

**Version:** Original manuscript: author's manuscript prior to formal peer review

**Terms of use:** Creative Commons Attribution-Noncommercial-Share Alike



# Superconducting nanowire single-photon detector with integrated impedance-matching taper

Di Zhu,<sup>1</sup> Marco Colangelo,<sup>1</sup> Boris A. Korzh,<sup>2</sup> Qing-Yuan Zhao,<sup>1</sup> Simone Frasca,<sup>2</sup> Andrew E. Dane,<sup>1</sup> Angel E. Velasco,<sup>2</sup> Andrew D. Beyer,<sup>2</sup> Jason P. Allmaras,<sup>2</sup> Edward Ramirez,<sup>2</sup> William J. Strickland,<sup>3</sup> Daniel F. Santavicca,<sup>3</sup> Matthew D. Shaw,<sup>2</sup> and Karl K. Berggren<sup>1, a)</sup>

<sup>1</sup>*Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, United States*

<sup>2</sup>*Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, United States*

<sup>3</sup>*University of North Florida, Jacksonville, FL, United States*

(Dated: 22 October 2018)

Conventional readout of a superconducting nanowire single-photon detector (SNSPD) sets an upper bound on the output voltage to be the product of the bias current and the load impedance,  $I_B \times Z_{\text{load}}$ , where  $Z_{\text{load}}$  is limited to  $50\ \Omega$  in standard r.f. electronics. Here, we break this limit by interfacing the  $50\ \Omega$  load and the SNSPD using an integrated superconducting transmission line taper. The taper is a transformer that effectively loads the SNSPD with high impedance without latching. It increases the amplitude of the detector output while preserving the fast rising edge. Using a taper with a starting width of  $500\ \text{nm}$ , we experimentally observed a  $3.6\times$  higher pulse amplitude,  $3.7\times$  faster slew rate, and  $25.1\ \text{ps}$  smaller timing jitter. The results match our numerical simulation, which incorporates both the hotspot dynamics in the SNSPD and the distributed nature in the transmission line taper. The taper studied here may become a useful tool to interface high-impedance superconducting nanowire devices to conventional low-impedance circuits.

The superconducting nanowire single-photon detector (SNSPD) is the leading single-photon detection technology at infrared wavelengths.<sup>1,2</sup> With exceptional performance,<sup>3–7</sup> it has played an essential role in various applications, especially quantum information science<sup>8,9</sup> and deep-space optical communication.<sup>10</sup>

A common problem with SNSPDs is their low output voltage and signal-to-noise ratio (SNR), which has been a limiting factor in detector timing jitter.<sup>11</sup> A simple lumped-circuit model dictates that the output voltage from the nanowire cannot exceed  $I_B \times Z_{\text{load}}$ , where  $I_B$  is the bias current and  $Z_{\text{load}}$  is the load impedance.<sup>12</sup>  $I_B$  is limited by the nanowire’s switching current at the  $\mu\text{A}$  range.  $Z_{\text{load}}$  is set by the input impedance of the coaxial cable and r.f. electronics, which is conventionally  $50\ \Omega$ . To improve readout SNR, significant progress has been made in developing cryogenic amplifiers with low noise, dissipation and cost, e.g. using silicon germanium and gallium arsenide transistors.<sup>13–16</sup> Digital readout circuits built directly from superconducting electronics, such as nanocryotrons<sup>17</sup> and single flux quantum (SFQ) circuits,<sup>18,19</sup> have also been demonstrated. These integrated superconducting circuits are low-noise and scalable, but usually require additional biasing and suffer from leakage current and crosstalk.

An alternative approach to increase the output signal is to increase  $Z_{\text{load}}$ . Compared to a standard  $50\ \Omega$  load, a high-impedance load is often more desirable—it not only increases the detector output, but also enables direct mapping of hotspot resistance and photon num-

ber/energy resolution.<sup>20–22</sup> However, high-impedance loading is difficult to achieve in practice. The lack of high-impedance coaxial cables makes it necessary to place the high-impedance amplifiers close to the detectors (at the low-temperature stage), which imposes a more stringent power budget. More importantly, even if a high-impedance amplifier is available,<sup>15</sup> loading a standard SNSPD directly with high impedance can lead to latching.<sup>12</sup>

In this work, without the need of high-impedance cryogenic amplifiers or any active circuit elements, we break the  $I_B \times 50\ \Omega$  limit by using an integrated superconducting transmission line taper. The taper gradually transforms its characteristic impedance from  $\text{k}\Omega$  to  $50\ \Omega$ , which effectively loads the SNSPD with a  $\text{k}\Omega$  impedance without latching. We designed the taper to be a coplanar waveguide (CPW) and fabricated it from the same superconducting thin film as the SNSPD. Using a taper with a starting width of  $500\ \text{nm}$  and nominal passband from  $200\ \text{MHz}$ , we experimentally observed  $3.6\times$  higher output voltage and no added noise compared to the non-tapered reference device. This voltage gain is equivalent to a  $11\ \text{dB}$  passive, dissipation-free cryogenic amplifier. Despite its large inductance, the taper preserves the detector’s fast rising edge, resulting in an increased slew rate and reduced timing jitter (from  $48.9\ \text{ps}$  to  $23.8\ \text{ps}$ ). The integrated impedance taper demonstrated here is useful for interfacing high-impedance nanowire-based devices to conventional low-impedance components, such as memory, and electrical or optical modulators.

Figure 1(a) shows a circuit model of a conventional SNSPD readout circuit, where the detector is modeled as a kinetic inductor  $L_K$  in series with a time-dependent

<sup>a)</sup>Electronic mail: berggren@mit.edu

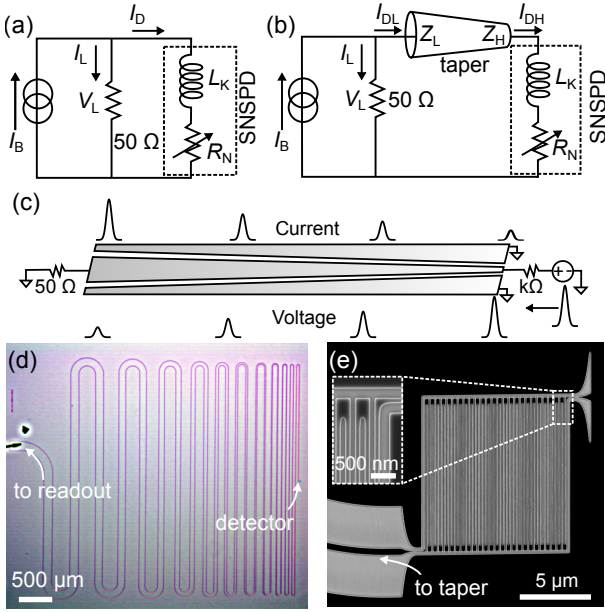


FIG. 1. Circuit diagram and micrographs of the tapered SNSPD readout. (a) A circuit diagram of a conventional SNSPD readout. (b) A circuit diagram of a tapered readout. The taper loads the SNSPD with high impedance while interfacing at its other end to the readout electronics at  $50\ \Omega$ , resulting in a larger output voltage. (c) Schematic diagram of a co-planar waveguide transmission line taper. When an electrical pulse is launched from the high-impedance end, its voltage drops but current increases while traveling towards the low-impedance end. (d) An optical micrograph of the integrated transmission line taper. Light area: NbN; red outlines: substrate. (e) A scanning electron micrograph of the SNSPD. Dark area: NbN; light area: substrate.

variable resistor  $R_N$ . When an incident photon triggers the detector,  $R_N$  switches from 0 to  $\approx k\Omega$  within  $\approx 100$ s of ps and diverts the bias current to the load. The evolution of  $R_N$  is determined by the non-linear electrothermal feedback in the detector.<sup>12,23</sup> The currents from the bias source ( $I_B$ ), in the nanowire ( $I_D$ ), and to the load ( $I_L$ ) simply follow Kirchhoff's law  $I_L = I_B - I_D$ . The maximum  $I_L$  is therefore limited to  $I_B$ , corresponding to the case where  $R_N$  pushes all of the current out of the nanowire ( $I_D = 0$ ). The output voltage on the load thus can not exceed  $I_B \times 50\ \Omega$ . In practice, due to the electro-thermal feedback,<sup>12</sup>  $I_D$  usually has some remainder, which depends on the bias current, kinetic inductance, and thermal constants of the materials.

Figure 1(b) shows a simplified circuit diagram for the tapered readout. The taper is inserted between the SNSPD and load, with a low impedance  $Z_L = 50\ \Omega$  on the load end and a high impedance  $Z_H$  on the detector end. In our implementation, the taper consists of a continuous nanowire transmission line without any dissipative elements. The taper is high-pass—it works as a transformer at high frequency but acts as an inductor at low frequency. When an incident photon triggers the

SNSPD,  $R_N$  switches on and pushes the current away from the nanowire. Instead of diverting the current directly to the  $50\ \Omega$  load as in the conventional readout, the SNSPD injects current to the taper at  $Z_H$ . As the electrical pulse travels towards the low impedance end, its current amplitude increases while the voltage amplitude drops, with a ratio that satisfies the change of impedance (Fig. 1(c)). Assuming an ideal broadband transformer with perfect impedance matching and power transmission, the current leaving the low-impedance end (to the load)  $\Delta I_{DL}$  is related to the current injected to the high-impedance end  $\Delta I_{DH}$  by  $\Delta I_{DL}^2 Z_L = \Delta I_{DH}^2 Z_H$ . In our transmission line taper, this relation is valid only at high frequency (passband of the taper), which dominates the rising edge of the detector pulse. In the extreme case where the SNSPD pushes all the bias current out, i.e.,  $\Delta I_{DH} = I_B$ , the current diverted to the load can be as large as  $I_L = \Delta I_{DL} = I_B \sqrt{Z_H/Z_L}$ , corresponding to an output voltage of  $V_L^{\text{taper}} = I_B \times 50\ \Omega \sqrt{Z_H/Z_L}$  and an effective voltage gain of  $\sqrt{Z_H/Z_L}$  with respect to the conventional readout. In practice, when terminated with the high-impedance taper, the SNSPD latches, leaving the residual current at the hotspot current  $I_{ss}$ , then resets through reflection from the taper. As we will show later, the actual voltage gain is always less than  $\sqrt{Z_H/Z_L}$  due to the electro-thermal feedback and limited taper bandwidth.

Figure 2(d) shows an optical micrograph of a fabricated SNSPD with a meandered transmission-line taper. The bright area is NbN, and the red area is the substrate, where the NbN was etched away. The NbN was sputtered at room temperature on a silicon substrate with a 300 nm thick thermal oxide layer.<sup>24</sup> The film had a critical temperature of 8.1 K and room-temperature sheet resistance of  $342\ \Omega/\text{sq}$ . The sheet inductance was estimated to be 80 pH/sq by fitting the falling edge of the output pulse from a reference detector. The nanowire fabrication process is described in Ref.<sup>4,25</sup>. The taper was made from a CPW with a fixed gap size of  $3\ \mu\text{m}$ , and a varying center conductor width from  $135\ \mu\text{m}$  ( $50\ \Omega$ ) to  $500\ \text{nm}$  ( $1.7\ \text{k}\Omega$ ). Its left/wide end is wire bonded to an external circuit board, and the right/narrow end connects to the SNSPD through a  $1\ \mu\text{m}$ -long hyperbolic taper. The SNSPD was 100 nm wide, densely packed with a 50% fill factor, and spanned a rectangular area of  $11\ \mu\text{m} \times 10\ \mu\text{m}$  (see Fig. 1(e)). A 200 nm gap surrounded the detector region to reduce the proximity effect in fabrication. On the same chip, we also fabricated non-tapered detectors as references.

The taper was designed to be a 5672-section cascaded transformer with a lower cut-off frequency of 200 MHz and a total electrical length of 851 mm, following the Klopfenstein profile.<sup>26</sup> The physical length was 77.9 mm due to the slow phase velocity of the superconducting transmission line, and the total inductance was 1.410  $\mu\text{H}$  (see SI for the simulated S parameters). This length was chosen so that the maximum reflection in the pass-

band ( $\geq 200$  MHz) would not exceed  $-20$  dB. The total electrical length is calculated as  $l_e = Ac/2\pi f_{co}$ , where  $f_{co}$  is the nominal cut-off frequency and  $c$  is the speed of light in vacuum.  $A$  is a factor that determines the maximum reflection in the passband and is calculated as  $\cosh(A) = \rho_0/\rho_{pb}$ , where  $\rho_0$  is the initial reflection coefficient (i.e., without taper) and  $\rho_{pb}$  is the maximum passband reflection coefficient (taken as 0.1 here). For design convenience, we followed Klopfenstein's original approach and took  $\rho_0 = 0.5 \ln(Z_H/Z_L)$  instead of  $(Z_H - Z_L)/(Z_H + Z_L)$ .<sup>26</sup>

The detectors were measured at 1.3 K in a closed-cycle cryostat. Both the bias circuit and readout electronics were at room temperature. The output signal of the detectors were amplified using a 2.5 GHz, 25 dB gain low-noise amplifier (RF BAY LNA-2500), and a 3 dB attenuator was inserted before the amplifier to reduce reflection and prevent latching. The output pulses from the amplifier were then acquired by a 6 GHz real-time oscilloscope (Lecroy 760Zi). The detector chip was flood illuminated using attenuated sub-ps pulsed lasers at 1550 nm (FPL-02CCF) through an optical fiber (SMF-28e). The laser pulses were split into two arms, one to a variable attenuator then to the cryostat, and the other to a fast photodiode (Thorlabs DET08CFC) as timing references. Since the distance between the non-tapered detector and tapered detector ( $\approx 5$  mm) was much less than the distance between the detector chip and fiber tip ( $\approx 10$  cm), we expect the difference in photon arrival time to be  $< 1$  ps. Both the tapered and non-tapered detectors had a switching current of  $30 \mu\text{A}$ , and were biased at  $27.5 \mu\text{A}$  throughout the measurement.

Figure 2(a) shows the measured pulse shapes from the reference and tapered detectors. The amplifier gain was removed to better compare with simulations. To avoid phase distortion in reconstructing the unamplified pulses, we used a weighted gain,  $\bar{G} = \int df PSD(f)G(f) / \int df PSD(f)$ , where  $PSD(f)$  is the power spectral density of the pulse, and  $G(f)$  is the measured system gain spectrum (see SI for details).  $\bar{G}$  was calculated to be 20.5 dB. As shown in Fig. 2(a), we observed a voltage gain of 3.6 and an extra delay of 2.8 ns from the tapered device compared to the reference device (by aligning the electrical pulses to the optical references). This voltage enhancement is equivalent to a passive, excessive-noise-free 11 dB amplifier.

We simulated the tapered detector using a SPICE model that incorporates both the hotspot dynamics in the SNSPD and the distributed nature of the transmission line taper.<sup>27,28</sup> The simulation was implemented in LTspice, a free electrical circuit modeling software. The SPICE model for SNSPD was developed by Berggren et al.<sup>27</sup> based on the phenomenological hotspot velocity model by Kerman et al.<sup>12</sup> The taper was simulated as cascaded lossless transmission lines (down-sampled to 300 sections),<sup>28</sup> and each section was implemented using the LTRA model in LTspice with different length, inductance, and capacitance settings.

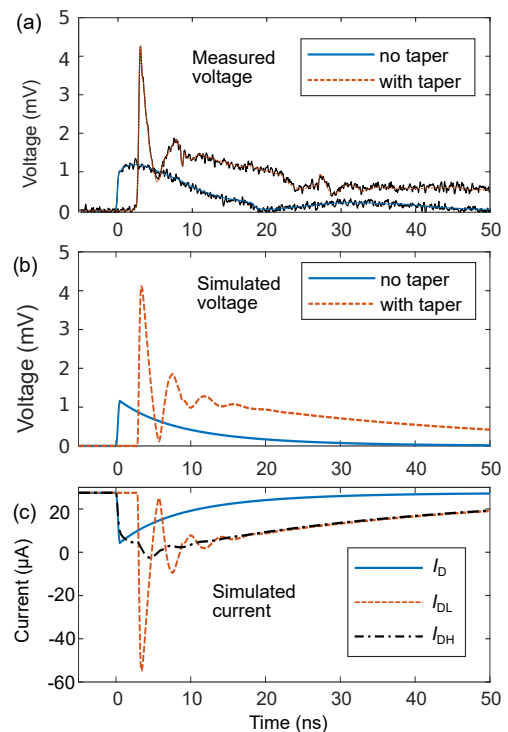


FIG. 2. Measured detector pulses and comparison with SPICE simulation. (a) Measured voltage with amplifier gain removed. The black traces are single-shot waveforms, and the superimposed colored lines are averaged waveforms. (b) Simulated output voltages for both the tapered and non-tapered detectors. (c) Simulated current distributions.  $I_D$  is the current in the SNSPD for a non-tapered readout.  $I_{DL}$  and  $I_{DH}$  are the currents on the low-impedance and high-impedance ends of the tapers in the tapered detector, respectively.

Figure 2(b) plots the simulated load voltages, showing a voltage gain of 3.5 and a delay of 2.8 ns, as compared to the measured gain of 3.6 and delay of 2.8 ns. The subsequent peaks in the output voltage are spaced by  $\approx 4.2$  ns, which should correspond to the round trip time in the taper. The single-trip delay of the taper calculated from the reflection peaks (2.1 ns) is shorter than the delay between the tapered detector and reference detector (2.82 ns), because the hotspot grows for a longer period of time and to a larger resistance in the tapered detector, as can be seen in the simulated currents.

Figure 2(c) shows the simulated currents. For the tapered detector, the current in the nanowire ( $I_{DH}$ ) first drops at a similar rate as the non-tapered case ( $I_D$ ), then enters an intermediate plateau due to latching.<sup>12</sup> A similar current plateau and latching behavior are often observed when loading an SNSPD with a  $k\Omega$  resistor. However, at  $\approx 4.6$  ns,  $I_{DL}$  drops again and kicks the detector out of the latching state. The drop in current is from the reflection in the transmission line taper. Alternatively, it can be interpreted as the distributed capacitors in the transmission line drawing current from the SNSPD. After a few oscillations (high frequency), the current in the

detector recovers with an  $\tau = L/R$  exponential time constant (low frequency). Here,  $L$  is the total inductance of the SNSPD and the taper (at low frequency, the taper behaves as an inductor), and  $R$  is  $50\ \Omega$ . The simulated currents at the high- and low-impedance ends of the taper follow our intuitive understanding on how a transformer works. In this particular detector and taper design, the maximum counting rate (estimated as  $1/3\tau$ ) decreases from 40.3 MHz ( $L = 414\ \text{nH}$ ) for the non-tapered detector to 9.1 MHz ( $L = 1.824\ \mu\text{H}$ ) for the tapered detector. We would like to point out that the added inductance, and consequently slower maximum counting rate, are currently the major drawbacks of the tapered readout. A study of the trade-off between the gain factor, taper impedance, bandwidth, and inductance can be found in the SI.

The impedance taper amplifies the output pulse without sacrificing the fast rising edge, resulting in a faster slew rate. Figure 3(a) compares the averaged rising edges of the detector pulses from the reference and tapered detectors (with amplifier gain). The sampling rate was 40 GS/s. As shown in Fig. 3(b), the maximum slew rates ( $dV/dt$ ) were  $39\ \mu\text{V/ps}$  for the reference detector, but  $143\ \mu\text{V/ps}$  (3.7 times faster) for the tapered detector.

The slew rate directly impacts the electrical noise contribution on the timing jitter, usually referred as noise jitter,  $\sigma_{\text{noise}}$ .<sup>11,29</sup> We sampled the background electrical noise on the oscilloscope for both detectors by measuring the voltage at 400 ps before the rising edge of the pulses. The noise followed a Gaussian distribution and had a standard deviation of  $559\ \mu\text{V}$  and  $547\ \mu\text{V}$  for the reference and tapered detector, respectively (see SI for details). Taking their respective fastest slew rates, we calculated that the reference detector would have a standard deviation  $\sigma_{\text{noise}}$  of 14.3 ps, and the tapered detector would have a  $\sigma_{\text{noise}}$  of 3.8 ps.

We measured the jitter of the detectors following the procedure described in Ref.<sup>30</sup>. The discrimination levels for time tagging were set to voltages with the fastest slew rates. Figure 3(c) shows the instrument response function (IRF) of the reference and tapered detectors at 1550 nm illumination wavelength. With the impedance taper, the full-width half-maximum (FWHM) jitter reduced from 48.9 ps to 23.8 ps. We fitted the IRF using an exponentially modified Gaussian distribution,<sup>4</sup> and found  $\sigma = 16.8\ \text{ps}$  and  $1/\lambda = 17.4\ \text{ps}$  for the reference detector, and  $\sigma = 6.5\ \text{ps}$  and  $1/\lambda = 13.6\ \text{ps}$  for the tapered detector. Here,  $\sigma$  is the standard deviation of the normal distribution,  $\lambda$  is the exponential decay rate. The detectors showed similar jitter reduction at 1064 nm, where both detectors operated on the saturation plateau (see SI for the photon count rate vs. bias current curves). The FWHM jitter reduced from 47.0 ps ( $\sigma = 16.4\ \text{ps}$ ,  $1/\lambda = 15.9\ \text{ps}$ ) to 22.4 ps ( $\sigma = 6.2\ \text{ps}$ ,  $1/\lambda = 12.5\ \text{ps}$ ). We observed a leading edge tail in IRF for the tapered detector. It is likely due to the counting events from the taper or the transition region between the taper and the detector. This effect could be reduced if the optical mode is

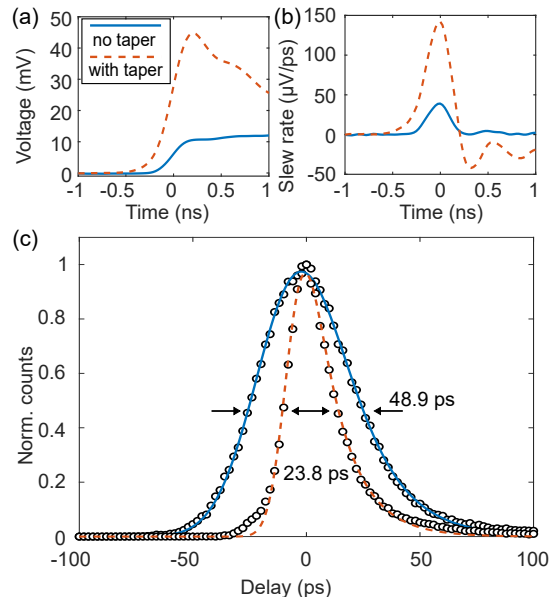


FIG. 3. Experimental observation on the reduction of timing jitter as a result of faster slew rate. (a) Averaged rising edges of the detector pulses from the tapered and non-tapered detectors (amplifier gain not removed); (b) corresponding slew rate calculated as  $dV/dt$ . (c) The measured FWHM timing jitter reduced from 48.9 ps to 23.8 ps with the tapered readout at 1550 nm.

focused in the center of the detector active area, through self-aligned fiber coupling or focusing lenses.<sup>31,32</sup>

As a final remark, we have treated the SNSPD as a lumped element in this paper, because the nanowire was closely meandered and had a dispersion similar to an ideal inductor at the frequency of interest.<sup>33</sup> Despite this choice, multi-photon absorption would generate a different hotspot resistance than the single-photon events.<sup>20,34</sup> The impedance taper provides an effective  $k\Omega$  load, and may thus allow direct discrimination of hotspot resistance and hence photon numbers. In another scheme, where the nanowire is sparse or integrated into a transmission line,<sup>28</sup> the taper can serve as an impedance-matched readout and has been used to resolve photon location and photon numbers.<sup>25,35</sup> We expect the integrated taper to become a widely used tool for matching high-impedance nanowire-based devices to low-impedance systems.

Part of this work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. Support for this work was provided in part by the JPL Strategic University Research Partnerships program, DARPA Defense Sciences Office through the DETECT program, and National Science Foundation grants under contract No. ECCS-1509486 (MIT) and No. ECCS-1509253 (UNF). D. Z. is supported by the National Science Scholarship from A\*STAR, Singapore.

- <sup>1</sup>R. Hadfield, *Nat. Photonics* **3**, 696 (2009).
- <sup>2</sup>C. M. Natarajan, M. G. Tanner, and R. H. Hadfield, *Supercond. Sci. Technol.* **25**, 063001 (2012).
- <sup>3</sup>F. Marsili, V. B. Verma, J. A. Stern, S. Harrington, a. E. Lita, T. Gerrits, I. Vayshenker, B. Baek, M. D. Shaw, R. P. Mirin, and S. W. Nam, *Nat. Photonics* **7**, 210 (2013).
- <sup>4</sup>B. A. Korzh, Q.-Y. Zhao, S. Frasca, J. P. Allmaras, T. M. Autry, E. A. Bersin, M. Colangelo, G. M. Crouch, A. E. Dane, T. Gerrits, F. Marsili, G. Moody, E. Ramirez, J. D. Rezac, M. J. Stevens, E. E. Wollman, D. Zhu, P. D. Hale, K. L. Silverman, R. P. Mirin, S. W. Nam, M. D. Shaw, and K. K. Berggren, (2018), arXiv:1804.06839.
- <sup>5</sup>E. E. Wollman, V. B. Verma, A. D. Beyer, R. M. Briggs, B. Korzh, J. P. Allmaras, F. Marsili, A. E. Lita, R. P. Mirin, S. W. Nam, and M. D. Shaw, *Opt. Express* **25**, 26792 (2017).
- <sup>6</sup>A. Vetter, S. Ferrari, P. Rath, R. Alaee, O. Kahl, V. Kovalyuk, S. Diewald, G. N. Goltsman, A. Korneev, C. Rockstuhl, and W. H. P. Pernice, *Nano Lett.* **16**, 7085 (2016).
- <sup>7</sup>F. Marsili, F. Bellei, F. Najafi, A. E. Dane, E. a. Dauler, R. J. Molnar, and K. K. Berggren, *Nano Lett.* **12**, 4799 (2012).
- <sup>8</sup>H.-L. Yin, T.-Y. Chen, Z.-W. Yu, H. Liu, L.-X. You, Y.-H. Zhou, S.-J. Chen, Y. Mao, M.-Q. Huang, W.-J. Zhang, H. Chen, M. J. Li, D. Nolan, F. Zhou, X. Jiang, Z. Wang, Q. Zhang, X.-B. Wang, and J.-W. Pan, *Phys. Rev. Lett.* **117**, 190501 (2016).
- <sup>9</sup>X. Qiang, X. Zhou, J. Wang, C. M. Wilkes, T. Loke, S. O’Gara, L. Kling, G. D. Marshall, R. Santagati, T. C. Ralph, J. B. Wang, J. L. O’Brien, M. G. Thompson, and J. C. F. Matthews, *Nat. Photonics* **12**, 534 (2018).
- <sup>10</sup>M. E. Grein, A. J. Kerman, E. A. Dauler, M. M. Willis, B. Romkey, R. J. Molnar, B. S. Robinson, D. V. Murphy, and D. M. Boroson, in *Proc. SPIE*, Vol. 9492 (2015) p. 949208.
- <sup>11</sup>J. Wu, L. You, S. Chen, H. Li, Y. He, C. Lv, Z. Wang, and X. Xie, *Appl. Opt.* **56**, 2195 (2017).
- <sup>12</sup>A. J. Kerman, J. K. W. Yang, R. J. Molnar, E. A. Dauler, and K. K. Berggren, *Phys. Rev. B* **79**, 100509 (2009).
- <sup>13</sup>J. C. Bardin, P. Ravindran, S.-W. Chang, C. Mohamed, R. Kumar, J. A. Stern, M. D. Shaw, D. Russell, F. Marsili, G. Resta, and W. H. Farr (International Society for Optics and Photonics, 2014) p. 911404.
- <sup>14</sup>H. Mani and P. Mausekopf, in *2014 11th Int. Work. Low Temp. Electron.* (IEEE, 2014) pp. 17–20.
- <sup>15</sup>C. Wan, Z. Jiang, L. Kang, and P. Wu, *IEEE Trans. Appl. Supercond.* **27**, 1 (2017).
- <sup>16</sup>C. Cahall, D. J. Gauthier, and J. Kim, *Rev. Sci. Instrum.* **89**, 063117 (2018).
- <sup>17</sup>A. N. McCaughan and K. K. Berggren, *Nano Lett.* **14**, 5748 (2014).
- <sup>18</sup>H. Terai, S. Miki, T. Yamashita, K. Makise, and Z. Wang, *Appl. Phys. Lett.* **97**, 112510 (2010).
- <sup>19</sup>T. Ortlepp, M. Hofherr, L. Fritzsche, S. Engert, K. Ilin, D. Rall, H. Toepfer, H.-G. Meyer, and M. Siegel, *Opt. Express* **19**, 18593 (2011).
- <sup>20</sup>M. Bell, A. Antipov, B. Karasik, A. Sergeev, V. Mitin, and A. Verevkin, *IEEE Trans. Appl. Supercond.* **17**, 289 (2007).
- <sup>21</sup>J. Kitaygorsky, S. Dorenbos, E. Reiger, R. Schouten, V. Zwiller, and R. Sobolewski, *IEEE Trans. Appl. Supercond.* **19**, 346 (2009).
- <sup>22</sup>S. Jahanmirinejad and A. Fiore, *Opt. Express* **20**, 5017 (2012).
- <sup>23</sup>J. K. W. Yang, A. J. Kerman, E. A. Dauler, V. Anant, K. M. Rosfjord, and K. K. Berggren, *IEEE Trans. Appl. Supercond.* **17**, 558 (2007).
- <sup>24</sup>A. E. Dane, A. N. McCaughan, D. Zhu, Q. Zhao, C.-S. Kim, N. Calandri, A. Agarwal, F. Bellei, and K. K. Berggren, *Appl. Phys. Lett.* **111**, 122601 (2017).
- <sup>25</sup>Q.-Y. Zhao, D. Zhu, N. Calandri, A. E. Dane, A. N. McCaughan, F. Bellei, H.-Z. Wang, D. F. Santavicca, and K. K. Berggren, *Nat. Photonics* **11**, 247 (2017).
- <sup>26</sup>R. W. Klopfenstein, *Proc. IRE* **44**, 31 (1956).
- <sup>27</sup>K. K. Berggren, Q.-Y. Zhao, N. Abebe, M. Chen, P. Ravindran, A. McCaughan, and J. C. Bardin, *Supercond. Sci. Technol.* **31**, 055010 (2018).
- <sup>28</sup>Q.-Y. Zhao, D. F. Santavicca, D. Zhu, B. Noble, and K. K. Berggren, *Appl. Phys. Lett.* **113**, 082601 (2018).
- <sup>29</sup>Q. Zhao, L. Zhang, T. Jia, L. Kang, W. Xu, J. Chen, and P. Wu, *Appl. Phys. B* **104**, 673 (2011).
- <sup>30</sup>F. Najafi, F. Marsili, E. Dauler, R. J. Molnar, and K. K. Berggren, *Appl. Phys. Lett.* **100**, 152602 (2012).
- <sup>31</sup>A. J. Miller, A. E. Lita, B. Calkins, I. Vayshenker, S. M. Gruber, and S. W. Nam, *Opt. Express* **19**, 9102 (2011).
- <sup>32</sup>F. Bellei, A. P. Cartwright, A. N. McCaughan, A. E. Dane, F. Najafi, Q. Zhao, and K. K. Berggren, *Opt. Express* **24**, 3248 (2016).
- <sup>33</sup>D. F. Santavicca, J. K. Adams, L. E. Grant, A. N. Mccaughan, and K. K. Berggren, *J. Appl. Phys.* **119**, 234302 (2016).
- <sup>34</sup>C. Cahall, K. L. Nicolich, N. T. Islam, G. P. Lafyatis, A. J. Miller, D. J. Gauthier, and J. Kim, *Optica* **4**, 1534 (2017).
- <sup>35</sup>D. Zhu, Q.-Y. Zhao, H. Choi, T.-J. Lu, A. E. Dane, D. Englund, and K. K. Berggren, *Nat. Nanotechnol.* **13**, 596 (2018).

## Supplementary Information

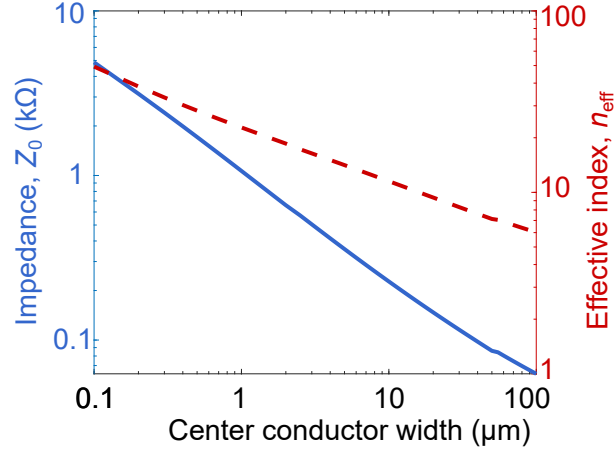


FIG. S1. Simulated characteristic impedance ( $Z_0$ ) and effective index ( $n_{\text{eff}}$ ) of the superconducting NbN coplanar waveguide. The gap size of the coplanar waveguide is fixed at 3  $\mu\text{m}$ , and the NbN film has a sheet kinetic inductance of 80 pH/sq. The substrate is 300 nm  $\text{SiO}_2$  on 500  $\mu\text{m}$  intrinsic Si.

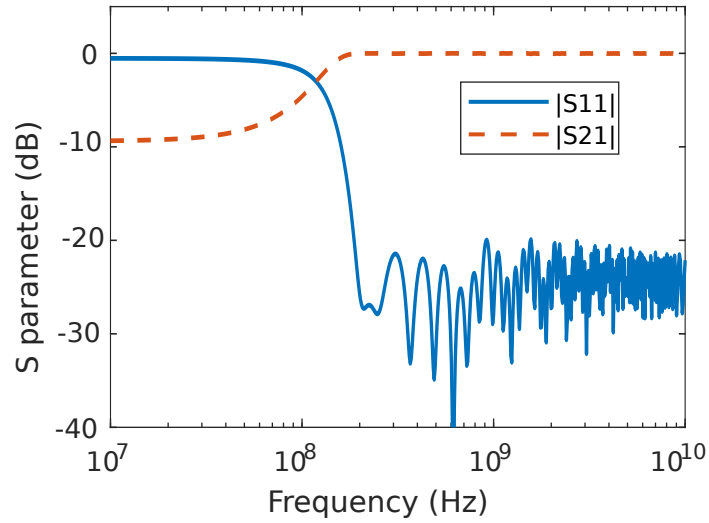


FIG. S2. SPICE simulated S-parameter of the impedance taper. In the simulation, the taper is discretized and down-sampled to 300 sections, and each section is implemented using the LTRA model in LTspice. The taper is terminated with impedance matched resistive load on both ends.

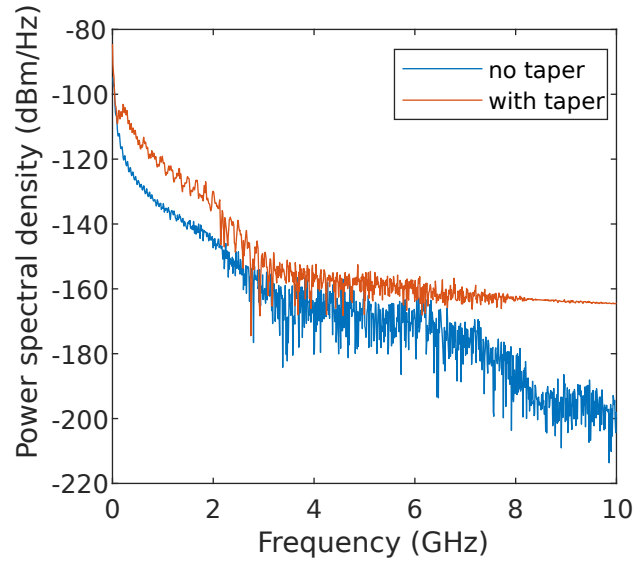


FIG. S3. Power spectral density of the output pulses from tapered and reference detectors. The power spectral density was calculated by taking the Fourier transform of the averaged detector pulses acquired on the oscilloscope. The detector pulses were amplified through a 2.5 GHz amplifier (see main text for the details on measurement setup). The sampling rate was 40 GS/s and the bandwidth was 3 GHz.

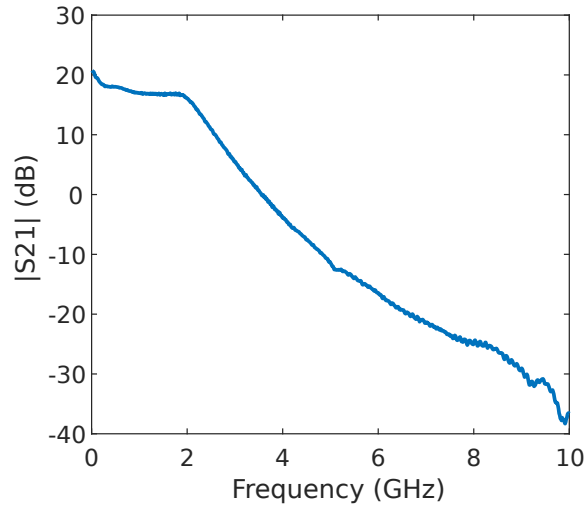


FIG. S4. System gain characterization. The transmission coefficient ( $S_{21}$ ) was measured using a vector network analyzer (Keysight N5224A) from the device under test up to the input port of the oscilloscope, including the cryocable, bias Tee, 3 dB attenuator, and low noise amplifier (LNA2500).



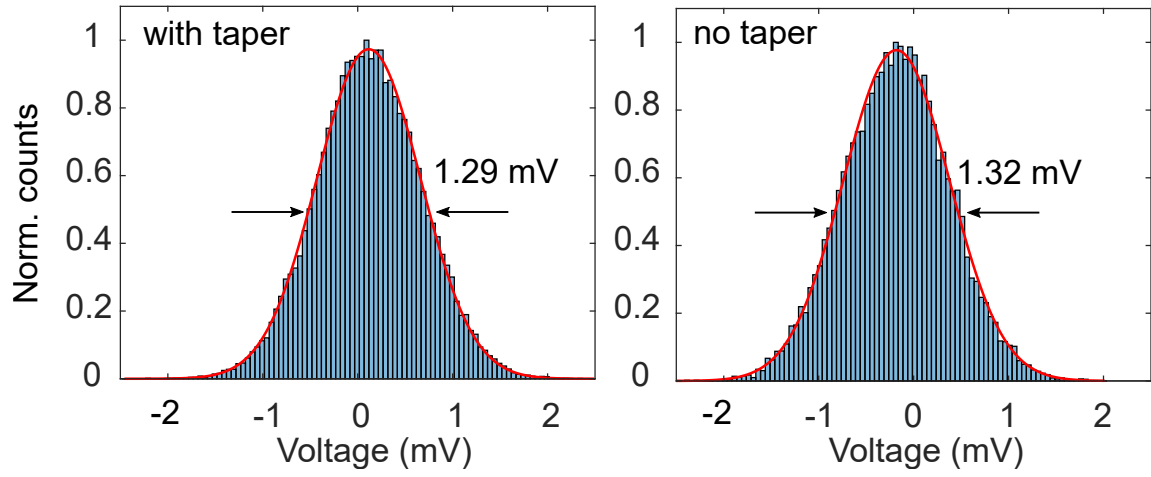


FIG. S5. Measured noise floor from the tapered and non-tapered device. Compared to the reference detector, no added noise was observed from the tapered device. The noise voltage was sampled on the oscilloscope at 400 ps before the rising edge of detector pulses. The sampling rate of the oscilloscope was 40 GS/s and the bandwidth was set to 3 GHz.

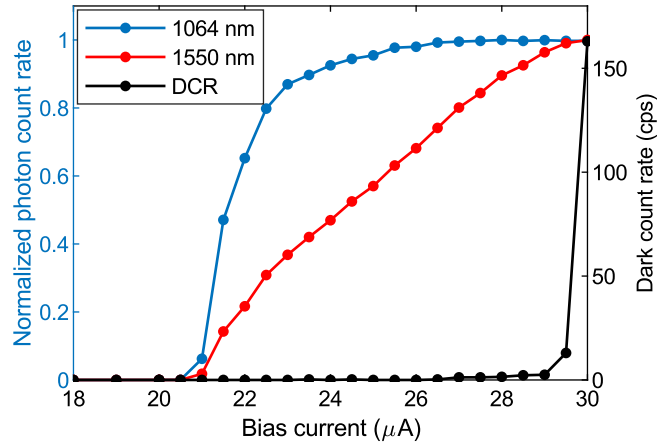


FIG. S6. Normalized photon count rate (PCR, left axis) and dark count rate (DCR, right axis) as a function of bias current. At 1064 nm, the detector showed saturated internal quantum efficiency. All the pulse shape and jitter measurement reported in work were measured at a bias current of 27.5  $\mu\text{A}$ .

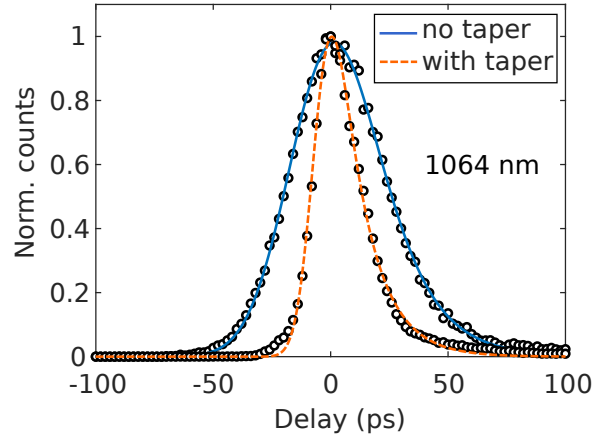


FIG. S7. Instrument response function (IRF) under 1064 nm illumination. The bias current was kept at  $27.5 \mu\text{A}$ . For the reference detector,  $\sigma=16.40 \text{ ps}$ ,  $\lambda = 15.9 \text{ ps}$ ,  $\text{FWHM}=47.0 \text{ ps}$ . For the tapered detector,  $\sigma = 6.2 \text{ ps}$ ,  $1/\lambda = 12.5 \text{ ps}$ ,  $\text{FWHM}=22.4 \text{ ps}$ .

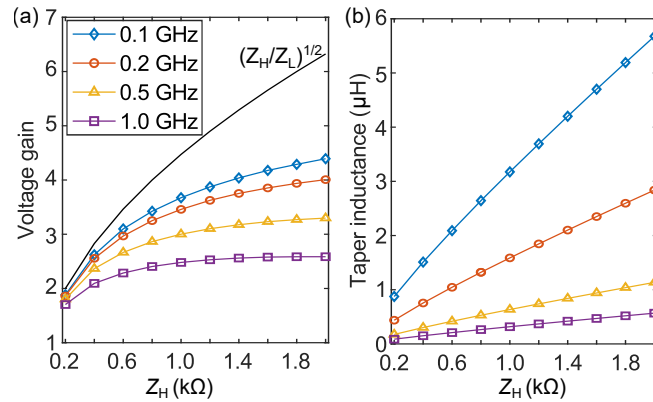


FIG. S8. Simulated gain factor (the ratio between the maximum voltages of tapered and non-tapered detector) and taper inductance as a function of the starting input impedance ( $Z_H$ ) under different nominal cut-off frequencies. Higher input impedance and smaller cut-off frequencies produce higher output voltages (a), but result in larger inductance (b), and hence longer rest time. The non-tapered SNSPD had a bias current of  $27.5 \mu\text{A}$ , an inductance of  $414 \text{ nH}$ , and a maximum output voltage of  $1.1 \text{ mV}$  ( $84\%I_B \times 50 \Omega$ ). Each simulated taper had 300 sections and the electrical length was designed to be  $l_e = \text{arccosh}(\rho_0/\rho_{pb})c/2\pi f_{co}$ , where  $\rho_0 = 0.5 \ln(Z_H/Z_L)$ , and  $\rho_{pb}=0.1$ , as described in the main text.  $Z_L$  was set to  $50 \Omega$ .