Demonstration of sub-3 ps temporal resolution with a superconducting nanowire single-photon detector

Boris Korzh[1,a),b)], Qing-Yuan Zhao[2,b)], Jason P. Allmaras[1,3], Simone Frasca[1], Travis M. Autry[4], Eric A. Bersin[1,2], Andrew D. Beyer[1], Ryan M. Briggs[1], Bruce Bumble[1], Marco Colangelo[2], Garrison M. Crouch[1], Andrew E. Dane[2], Thomas Gerrits[4], Adriana E. Lita[4], Francesco Marsili[1], Galan Moody[4], Cristián Pena[5,6], Edward Ramirez[1], Jake D. Rezac[4], Neil Sinclair[5], Martin J. Stevens[4], Angel E. Velasco[1], Varun B. Verma[4], Emma E. Wollman[1], Si Xie[5], Di Zhu[2], Paul D. Hale[4], Maria Spiropulu[5], Kevin L. Silverman[4], Richard P. Mirin[4], Sae Woo Nam[4], Alexander G. Kozorezov[7], Matthew D. Shaw[1] and Karl K. Berggren[2]

[1]*Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA*

[2]*Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

[3]*Department of Applied Physics, California Institute of Technology, Pasadena, California 91125, USA*

[4]*National Institute of Standards and Technology, Boulder, Colorado 80305, USA*

[5]*Division of Physics, Mathematics and Astronomy, California Institute of Technology, Pasadena, California 91125, USA*

[6]*Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA*

[7]*Department of Physics, Lancaster University, Lancaster, LA1 4YB, UK*

a) Corresponding author: bkorzh@jpl.caltech.edu

b) These authors contributed equally to this work.

Abstract

Improvements in temporal resolution of single photon detectors enable increased data rates and transmission distances for both classical and quantum optical communication systems, higher spatial resolution in laser ranging, and observation of shorter-lived fluorophores in biomedical imaging. In recent years, superconducting nanowire single-photon detectors (SNSPDs) have emerged as the most efficient, time-resolving single-photon counting detectors available in the near infrared, but understanding of the fundamental limits of timing resolution in these devices has been limited due to a lack investigations into the time scales involved in the detection process. We introduce an experimental technique to probe the detection latency in SNSPDs and show that the key to achieving low timing jitter is the use of materials with low latency. By using a specialised niobium nitride (NbN) SNSPD we demonstrate that the system temporal resolution can be as good as 2.6±0.2 ps for visible wavelengths and 4.3±0.2 ps at 1550 nm.

Introduction

Superconducting nanowire single-photon detectors[1] (SNSPDs) have advanced rapidly over the last decade with their capability showcased by achievements of high detection efficiency (>90%),[2,3] fast reset times (<1 ns)[4] and scalability to kilo-pixel arrays.[5] The number of applications utilising SNSPD is continually growing, spanning wavelengths from the ultra-violet[6,7] to mid-infrared,[8] while the near-infrared region has seen the largest number of landmark experiments enabled by SNDPDs, such as the demonstration of a loophole-free Bell test,[9] several long-distance records for quantum key distribution[10–12] and quantum teleportation,[13–15] kilometre-range single photon laser ranging[16] as well as deep-space optical communication.[17] A key detector parameter in all these applications is the temporal resolution[18,19] and while SNSPDs have been demonstrated with timing precision as low as 14.2 ps,[20] the intrinsic timing response in these devices have not been thoroughly studied to date.

Temporal resolution of SNSPDs, commonly referred to as jitter, is quantified by the width of the temporal distribution of output signals with respect to the photon arrival times. This statistical distribution is known as the instrument response function (IRF), and its width is commonly evaluated as the full-width at half-maximum (FWHM). Significant effort has been made in the community to reduce the jitter and to ultimately understand the fundamental limitations. Though timing fluctuations arise from the microscopic physics of the detection mechanism, it is only recently that limited indications of these effects have been observed, such as asymmetry in the IRF, dependence on photon energy and external magnetic field.[21–23] The ability to directly observe the intrinsic timing jitter has been difficult, primarily because jitter in SNSPDs has often been limited by instrumental mechanisms, preventing a systematic experimental study of the intrinsic timing characteristics on different device design parameters.

A typical SNSPD signal is shown in Figure 1a which illustrates the effect of electrical noise jitter on the system, which arises due to amplifier noise-induced shifts of the readout voltage that lead to fluctuations in the time at which the threshold is crossed. Electrical noise jitter can be reduced by engineering devices with higher operating currents and faster rise times, as well as by using lower-noise radio-frequency (RF) cryogenic amplifiers (see Ref.[20] for a review of recent efforts). Despite this progress, a reduction of the noise contribution down to the 1 ps level has so far been missing. In addition, it was recently realised that fluctuations in the longitudinal position of the photon absorption lead to significant changes in the detection latency (defined as the time between the absorption of a photon and the registration of an electrical output pulse). This occurs because the high kinetic inductance of typical nanowire structures renders the speed of RF pulse propagation along the wire to be a small fraction of the speed of light in vacuum.[24] Fortunately, this so-called “longitudinal geometric” jitter can be compensated, at least in part, by using a differential readout scheme with a low-noise amplifier at each end of the nanowire,[25] provided that an impedance matching taper is used to avoid reflections at the end of the nanowire.[24,26]

Results

Low-jitter superconducting nanowire single-photon detector

In this work, we realise 5 μm-long NbN SNSPD devices (illustrated in Figure 2) designed to simultaneously minimise the longitudinal geometric and intrinsic jitter. The electrical noise jitter is also minimised through the use of a cryogenic RF amplifier and by maintaining a low device kinetic inductance, which ensures a fast rise of the readout signal[27] (see Methods for details). Achievement of low system jitter allows us to carry out a comprehensive study, demonstrating significant photon energy, nanowire width, and temperature dependence of the jitter (see Extended Data Figure 1), as well as intrinsic variations in the detection latency. These results make an important step towards probing the fundamental physics of photon detection and have guided us in refining the model of the detection mechanism in SNSPDs.

Measurement of the intrinsic detection timescales

The IRF for a 80 nm-wide device is shown in Figure 2b, demonstrating a clear dependence on the detected wavelength. With 1550 nm light, we observed 4.6±0.2 ps jitter, which is reduced to 3.8±0.3 ps for 775 nm light, at the same bias current (see Figure 3 and Figure 4 for a comparison of a wider range of devices and wavelengths). Fundamental understanding of this effect starts from the consideration of the intrinsic detection latency. Numerical modelling has predicted that variation in photon energy should lead to changes in detection latency in SNSPDs[28–30] but until now it has not been possible to experimentally quantify the time scales involved in order to validate the models. A previous attempt[31] to study this time delay, based on the difference between the high photon flux bolometric response and the single photon response, suffers from a lack of calibration of longitudinal signal propagation effects[25] as well as the change of signal rise time which has been shown to depend on the number of absorbed photons.[32] To overcome these limitations, we implement a straightforward measurement of the relative latency between

single photons of different wavelengths, synchronised in time, in order to measure the time scales involved (see Methods and Extended Data Figure 2 for a description of the experimental setup). Figure 1b shows that the detection signals resulting from 1550 nm photons are delayed by 4.2±0.4 ps relative to 775 nm photons. Reducing the bias current increases both the jitter and the relative detection latency, as illustrated in Figure 1c, hinting at the intrinsic link between the latency and jitter.

Intrinsic jitter mechanism

Figure 3a shows the effect of varying the nanowire width and bias current on the jitter while Figure 3c shows the relative latency change for these parameters. For a fixed bias current, the jitter and latency decreases for narrower nanowires and for higher energy photons. The same trend is evident in the normalised photon count rates (PCR) shown in Figure 3b: for a given bias current, increasing the photon energy or reducing the nanowire width leads to higher internal detection efficiency. The existence of saturated PCR plateaus suggests that the devices are operating with near-unity internal detection efficiency at high bias currents.[2] The link between these metrics can be interpreted qualitatively by introducing time-dependence into the internal detection efficiency and calculating the detection latency as a function of energy retained in the nanowire following the photon absorption, as shown in Figure 1d (see Methods for details). For a uniform nanowire the amount of energy retained in the electronic system fluctuates from shot to shot, due to Fano fluctuations,[33] leading to latency changes, which is the jitter. From Figure 1d it is clear that a normal distribution of energy deposition leads to a positive-skew bell-shaped jitter histogram and for a given nanowire it is possible to reduce the overall jitter by either increasing the photon energy or increasing the bias current as illustrated in Figure 1e. We estimate that we are able to achieve up to 85% of the theoretical maximum depairing current (see Methods), which helps explain the low values of jitter achieved in this work.

Intrinsic jitter dependence on the photon energy

The jitter reduction for increasing photon energies has been studied with the widest nanowire (120 nm), using wavelengths down to 273 nm, as illustrated in Figure 4a and the lowest achieved values of jitter are outlined in Figure 4d. As before, the jitter curves are correlated with the PCR curves (Figure 4b). The change in histogram shape for several wavelengths is seen in Figure 4c and it is clear that a long tail of detection events is present, which is expected to arise primarily due to detections in the tapering sections at the ends of the nanowire (see Figure 2). It is expected that focusing the light onto the straight section of the nanowire will significantly reduce this tail (see Supplementary Note 1 for analysis). We note that even once detections in tapering sections and bends are prevented, the simulations predict that the shape of the tail will not to be purely exponential.

Comparison of different superconducting materials

As discussed earlier, in order to achieve low intrinsic jitter it is important to operate at a high fraction of the depairing current which calls for good nanowire uniformity (note that spatial non-uniformities are themselves a potential source of jitter).[30] We believe that an important factor in achieving the high uniformity, and thus low jitter presented in this work, is the use of RF-bias sputtered NbN films which creates a quasi-amorphous small-crystal structure.[34] Amorphous superconductors such as tungsten silicide (WSi) and molybdenum silicide (MoSi) are popular choices for SNSPD fabrication precisely due to their uniformity and typically saturated internal efficiency, indicating high fractions of depairing current. Unfortunately, this alone is not sufficient for low jitter operation. We characterised the relative latency of WSi devices, using the same low jitter design, which achieve a comparable fraction of the depairing current to the NbN devices, and found that it is a factor of 2-3.5 larger depending on the bias current (see Extended Data Figure 3 for details). Consequently, the jitter is a factor of two higher. Biasing at similar fractions of critical depairing current results in close functional dependencies of latency versus

energy deposition (Figure 1d) for both NbN and WSi (for spatial non-uniformities one can describe their effect by introducing the effective deposited energy fluctuations).[33] Intrinsic timing jitter occurs due to variability (derivative of latency) of sensor latency with respect to fluctuating parameters, such as energy deposition and random spatial variations of material properties. As a result, the FWHM of the IRF is approximately proportional to the derivative of SNSPD latency with respect to energy deposition. We have also verified that MoSi devices have a similarly increased intrinsic jitter compared to NbN (although the relative latency data has not been presented so far).[35]

Discussion

We have demonstrated that when seeking an SNSPD with low intrinsic jitter, the key characteristic to evaluate is the intrinsic latency, even though the details of time- and spatial-evolution of the highly non-equilibrium state following photon absorption are dependent on many individual material parameters. In that respect, the low jitter results presented in this work has been made possible through the use of low latency NbN devices with good uniformity. Our proposed method of characterising the relative latency proves to be a practical way to analyse and compare devices fabricated using different materials and designs, as well as matching the data to simulations. Reproducing the intrinsic relative latency is easier than comparing the IRF to microscopic models, since many sources of instrumental fluctuations and timing offsets can be eliminated. Furthermore, a model can average over transverse coordinate dependence,[29,36] Fano fluctuations,[33] and thermal fluctuations, reducing the complexity and computational demands of generating a full IRF for model validation.

While the current detection model (see Methods) neglects two-dimensional effects (known as transverse geometric jitter)[29] it can be used for qualitative interpretation of the data and we have taken an important step in showing that the calculated relative latency time-scales match the experimental data for narrow nanowires (see Figure 3c), while previous models considerably underestimate the timescales

involved.[30] We have already observed the emergence of transverse geometric jitter for high photon energies with wide nanowires at low bias currents (see Supplementary Note 2) which highlights the need for further development of the detection models with the goal of achieving quantitative agreement with experiment.

Our results show that SNSPDs can outperform any other free-running single-photon detection technology in terms of jitter: we observed a jitter of 2.6±0.2 ps for visible wavelengths, whereas the lowest jitter achieved with Si avalanche diodes is 7.8 ps.[37] In the near-infrared, our results show 4.3±0.2 jitter and the best alternative detectors are indium gallium arsenide (InGaAs) avalanche diodes, that are capable of 50 ps jitter.[38] Until this work, the lowest jitter achieved with SNSPDs was 14.2 ps for NIR wavelengths.[20] We note that although sub-picosecond jitter and single-photon sensitivity is achievable with commercial streak cameras, their temporal dynamic range is extremely limited - deeming them incapable of detecting both fast and slow processes simultaneously, which is possible with the SNSPD detectors achieved in this work.

To put these results into context, one can consider the potential impact on a handful of applications. The combination of an SNSPD with a time-correlated single-photon counting (TCSPC) module results in an optical sampling oscilloscope,[19] where an IRF width of 2.6 ps is equivalent to a signal bandwidth of about 135 GHz - comparable to the fastest available sampling oscilloscopes. We have demonstrated this capability by capturing a single-photon-level 100 GHz optical waveform using the detector developed here.[39] Laser ranging also benefits from SNSPDs with low timing jitter: detection of a single visible photon reflected from an object with 2.6 ps resolution yields a depth precision of 1 mm with a confidence of approximately 90%. Averaging of multiple detections adds further advantage, with a resolution of 10 μm possible after just 10,000 detected events.[19] Figure 5 shows a three-dimensional reconstruction of a small key achieved with a table-top laser-ranging scanner using the detector presented here (see Methods for details), which shows that millimetre-scale depth resolution is achievable with just tens of detected

photons, highlighting the potential impact on laser ranging. While the small active area of the devices results in low photon absorption probability, the techniques shown here could be applied toward practical low-jitter devices with high detection efficiency by using differential readout[25] with impedance matching tapers[24,26] to remove the contribution from longitudinal geometric jitter. For integrated photonics applications,[40] detectors with dimensions similar to those used in this work can be coupled efficiently to a photonic waveguide,[4,41,42] to achieve high on-chip detection efficiency.

The observed saturation of jitter at short wavelengths (see Figure 4a), which is likely due to limits of the characterisation setup as well as the noise floor of our readout scheme, suggests that there is still room for further optimisation of the noise jitter and that the intrinsic jitter could approach the sub-picosecond level as suggested by recent models.[29,30] This optimisation could be achieved using a near quantum-limited amplifier,[43] a superconducting digital readout element such as a nanocryotron,[44] by using an adiabatic taper to impedance match the nanowire to 50 Ω,[24,26] or by engineering SNSPDs with faster rise times.[27] Reaching the sub-picosecond level would also require experimental studies of the effect of transverse geometric jitter as well as the role of spatial inhomogeneities.[30] Our investigation has taken an important step towards better understanding the fundamental origin of timing jitter in SNSPDs and calls for further theoretical work to quantitatively reproduce the experimental data.

Methods

Nanofabrication and screening

The devices were fabricated in the Massachusetts Institute of Technology (MIT) cleanroom, patterned from a NbN film with a nominal thickness of 7 nm. We deposited the film in a reactive sputtering system, at room temperature on a 4 inch silicon wafer with a 300 nm-thick thermal oxide layer.[34] As deposited, the film sheet resistance was $R_{sq} = 340\ \Omega/\square$ at room temperature and the residual resistance ratio was

$\rho_{RRR} = 0.8$. The critical temperature of the film was measured to be $T_c = 8.65$ K. The entire device was patterned with one step of electron-beam lithography. We spun an 80 nm-thick positive-tone e-beam resist (gl2000, diluted to 5%) baked at 180°C for 2 minutes. The e-beam lithography was performed with a 125 kV system. We separated the writing into two steps. The first step exposed the nanowire, which was designed as a coplanar waveguide structure. To expose the gap area of the nanowire with fine edge roughness, the exposure beam current was 500 pA with a step size of 1 nm. The second step was to expose the outline of the electrical contacts. Because the size of the contacts exceeded the size of the maximum writing field, we applied a multi-pass method with a 10 nA beam current and a step size of 4 nm. We developed the chip in O-xylene at 5°C for 30 seconds and a subsequent rinse in isopropanol for another 30 seconds. To transfer the pattern to the NbN film, we performed a CF_4 reactive ion etch at 50 W for 3 minutes. After the etch, resist was applied to protect the device from oxidisation and dicing. We first removed the residual gl2000 layer in 60ºC N-methyl pyrodone for 10 minutes, and then spun a new layer of gl2000 and a 1.5 μm-thick photoresist (Microposit S1813) for protection. After dicing, the S1813 was stripped with acetone while the gl2000 layer was left as a protection layer. It was possible to wire bond though the gl2000 layer directly to the NbN contact pad.

The total length of the inductor was 1.5 mm, corresponding to an inductance of 96 nH assuming a sheet inductance of 64 pH/□, extracted from pulse shape measurements in SNSPD structures fabricated from similar films. This inductor was added in order to avoid detector latching.[45] We applied a hyperbolic curve to the two ends of the short nanowire to minimise current crowding[46] and reduce the length of the taper area.

Overall, 160 devices were fabricated on 20 dies in a single fabrication run. Approximately a third of the devices were screened to measure the switching currents at 0.9 K, and the PCR curves at 1550 nm were measured for about 20 of those screened. This allowed us to select a representative device for each of the four nanowire widths for tests with the low-noise cryogenic amplifier.

Cryogenic setup

The experiment was carried out by using a pulse-tube cryocooler with a ^{4}He sorption refrigerator at a base temperature of 0.9 K. The signal from the SNSPD was amplified with a silicon germanium (SiGe) cryogenic amplifier (Cosmic Microwave, CITLF1) mounted on the 4 K stage of the cryocooler. The amplifier had a nominal gain of 50 dB, a nominal bandwidth of 1.5 GHz with a low frequency cut-off at 1 MHz and a nominal noise temperature of less than 7 K. The SNSPD was biased with a low-noise current source through a resistive bias-T at the input of the amplifier. The amplifier was biased with a supply voltage of 3 V and dissipated approximately 30 mW of power.

Jitter and latency measurement

The jitter was measured in a broad range of wavelengths, primarily using a tuneable titanium-sapphire mode-locked laser capable of pumping an optical parametric oscillator (OPO), configurable for visible and near-infrared operation. The centre wavelength of the laser was set to 800 nm, which resulted in a pulse width of 130 fs at a repetition rate of 76 MHz. The OPO was used to generate 600 nm and 1200 nm light with a pulse duration of 200 fs. Pulses at 400 nm were achieved with second harmonic generation (SHG) using a type-I β-Barium borate (BBO) crystal. 273 nm light was generated using a BBO-based tripler with the pump laser tuned to 820 nm. 532 nm light was generated with SHG of a 1064 nm mode locked laser (0.7 ps nominal pulse width) using a 0.5 mm periodically-poled lithium niobite (PPLN) crystal.

The SNSPD and laser synchronisation signals were acquired simultaneously on a digital real-time oscilloscope. To investigate the effect of the oscilloscope sample rate, we carried out measurements with both 40 and 80 Giga-samples per second, which reduces the lowest jitter from 3.2 ps to 2.6 ps (See Figure 4a compared to Figure 4d). The reason for this reduction is due to an improved reproduction of both the

SNSPD and synchronisation signals. Since the laser synchronisation signal had a rise time of approximately 50 ps, it is actually under-sampled with a rate of 40 Giga-samples per second, resulting in quantisation noise of the acquired waveform. The optimum analogue bandwidth setting of the oscilloscope was found to be 6 GHz, which is sufficiently high for the SNSPD signal which had a rise time (20% to 80%) of approximately 80 ps. Increasing the bandwidth further generates additional noise, which increases the noise jitter. The vertical scale setting of the oscilloscope was also optimised in order to minimise the noise floor of the input whilst limiting saturation effects. The residual quantisation and noise jitter was measured to be 1.5 ps by splitting the electronic synchronisation signal and passing one side through the cryogenic readout chain and running an IRF acquisition with the same settings. Since this value is close to the system jitter achieved, it is likely that even lower jitter could be achieved with further improvements to the measurement setup.

It is difficult to obtain the absolute detection latency due to the complexity of calibrating the exact moment of photon absorption in the nanowire (see main text for discussion). Instead, we investigated the relative detection latency difference between two photon energies. Since the 775 nm and 1550 nm pulses are synchronised in our optical setup and follow the same optical path (see Extended Data Figure 2), an IRF was collected for each bias current with light of both wavelengths without realigning the beam, by simply changing the free-space filter. The delay through the two filters was verified to be the same, by using a 1064 nm pulsed laser for which both filters are transparent. The trigger level was kept the same for the two wavelengths at each bias current, and it was verified that the SNSPD pulse shape was the same for both wavelengths (see Extended Data Figure 4). The latency difference was calculated by measuring the shift of the 1550 nm data relative to 775 nm, taking the peak of the fitted IRF as the reference. Although the IRF does not follow an exponentially modified Gaussian exactly, it provides enough accuracy for extraction of the peak location and FWHM values.

Laser ranging scanner

A proof-of-principle laser-ranging experiment was carried out over a distance of 1 meter (see Extended Data Figure 5 for details of the setup) with a 1064 nm mode-locked laser and a fast scanning mirror. The collected light was focused onto the detector using a lens external to the cryostat, yielding a system efficiency on the order of 0.01%. With 1 mW of laser power, an additional 30 dB of channel attenuation was added to prevent detector saturation effects. The detection events were time-tagged using a Becker and Hickl SPC-150NXX TCSPC module which has an electronic IRF of 3 ps, minimum bin duration of 203 fs and a saturated acquisition rate of 10 Mcps. The system IRF was 6.2 ps (see Extended Data Figure 6). The three-dimensional image of a small key in Figure 5 was achieved using a 200 by 200 scan with a 20 ms dwell time at each location.

Depairing current estimation

We recently demonstrated a practical method of estimating the depairing current in superconducting nanowires based on the bias current dependence of the kinetic inductance.[47] By fabricating resonator structures from the same NbN film used for the detectors in this work we extracted a depairing current of 38.8 μA for a 120 nm-wide nanowire. The switching current of 33 μA for the 120 nm detector corresponds to 85% of the depairing current. The same technique was used for estimation of the depairing current in the WSi device studied (see Extended Data Figure 3).

Detection latency simulations

The qualitative picture of intrinsic jitter can be drawn from very general considerations. Latency is one of the fundamental characteristics of any sensor response.[29,30] In SNSPDs, the monitored state is that of the condensate through its capacity to carry supercurrent without dissipation throughout the evolution of the non-equilibrium state formed during photoresponse. Considering a fully deterministic response, i.e.

$PCR(t, I_B, T_b, E) = \Theta\big(t - \tau_{lat}(I_B, T_b, E)\big)$, when the deposited energy exceeds the detection (cut-off) energy, $E \geq E_{det}(I_B, T_b)$, a detection event is registered at time $\tau_{lat}(I_B, T_b, E)$. Here $\Theta(t)$ is the Heaviside function, and I_B and T_b are the bias current and bath temperature respectively. The latency $\tau_{lat}(I_B, T_b, E)$ as a function of energy E must be:

a) infinite for $E < E_{det}(I_B, T_b)$,

b) monotonically decreasing with increasing E for $E \geq E_{det}(I_B, T_b)$,

c) assumed to be positively curved (consistent with numerical modelling)[29,30],

d) saturating, $\lim_{E\to\infty} \tau_{lat}(I_B, T_b, E) = \tau_\infty$.

Condition a) follows from the definition of the detection threshold, b) from the causality principle, c) from the requirement that an increase of energy away from the detection threshold makes the jitter distribution monotonically narrower, and d) from the fact that superconductivity cannot be broken immediately following the absorption of a photon, even at arbitrarily large photon energies. Regarding a), the authors of[30] use the model of an integrable singularity in τ_{lat} at $E = E_{det}(I_B, T_b)$ consistent with their 1D modelling while the author of[29] argues in favour of a finite value of τ_{lat} at $E = E_{det}(I_B, T_b)$. The exact nature of the behavior at $E = E_{det}(I_B, T_b)$ does not change the qualitative conclusions of this analysis.

The expression for the IRF, $H(t)$ was derived in terms of fluctuations of detector latency for the case of the 1D model with small fluctuations in the form,[30]

$$H(t) = -\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(E(t,I_B,T_b)-\bar{E})^2}{2\sigma^2}\right) \frac{dE(t,I_B,T_b)}{dt}.$$

Here $E(t, I_B, T_b)$ is the single-valued (as follows from a) and b)) solution of equation $t - \tau_{lat}(I_B, T_b, E) = 0$, $\bar{E} = \chi E_\lambda$, E_λ is the photon energy, $\chi < 1$ is the fraction of the photon energy retained by the nanowire after down-conversion, and the standard deviation σ describes the dynamic Fano fluctuations[33] and static spatial non-uniformities of the nanowire. The expression for $H(t)$ combined with a)-d) is sufficient for

the qualitative interpretation of all data depicted in Figure 1, Figure 3 and Figure 4. More quantitative analysis requires precise functional forms of the latency. The latter can be derived solving the standard or generalised time-dependent Ginzburg-Landau equations (TDGL) together with energy balance equations for quasiparticles and phonons. Additional details have been described in recent manuscripts.[29,30] The qualitative analysis of the data following from the general features of the latency and examples of the simulation results for the 80 nm wide SNSPD are described in Supplementary Note 3.

Data availability

The data that support the plots within this paper and other findings of this study are available from the corresponding author upon reasonable request.

References

1. Gol'tsman, G. N. *et al.* Picosecond superconducting single-photon optical detector. *Appl. Phys. Lett.* **79**, 705–707 (2001).

2. Marsili, F. *et al.* Detecting single infrared photons with 93 % system efficiency. *Nat. Photon.* **7**, 210–214 (2013).

3. Reddy, D. V *et al.* Exceeding 95% system efficiency within the telecom C-band in superconducting nanowire single photon detectors. in *Conference on Lasers and Electro-Optics* FF1A.3 (Optical Society of America, 2019).

4. Vetter, A. *et al.* Cavity-Enhanced and Ultrafast Superconducting Single-Photon Detectors. *Nano Lett.* **16**, 7085–7092 (2016).

5. Wollman, E. E. *et al.* A kilopixel array of superconducting nanowire single-photon detectors. *Opt. Express* **27**, 35279–35289 (2019).

6. Slichter, D. H. *et al.* UV-sensitive superconducting nanowire single photon detectors for integration in an ion trap. *Opt. Express* **25**, 8705–8720 (2016).

7. Wollman, E. E. *et al.* UV superconducting nanowire single-photon detectors with high efficiency, low noise, and 4 K operating temperature. *Opt. Express* **25**, 26792–26801 (2017).

8. Chen, L. *et al.* Mid-infrared Laser-Induced Fluorescence with Nanosecond Time Resolution Using a Superconducting Nanowire Single-Photon Detector: New Technology for Molecular Science. *Acc. Chem. Res.* **50**, 1400–1409 (2017).

9. Shalm, L. K. *et al.* Strong Loophole-Free Test of Local Realism. *Phys. Rev. Lett.* **115**, 250402 (2015).

10. Takesue, H. *et al.* Quantum key distribution over a 40-dB channel loss using superconducting single-photon detectors. *Nat. Photon.* **1**, 343–348 (2007).

11. Yin, H. L. *et al.* Measurement-Device-Independent Quantum Key Distribution over a 404 km Optical Fiber. *Phys. Rev. Lett.* **117**, 1–5 (2016).

12. Boaron, A. *et al.* Secure Quantum Key Distribution over 421 km of Optical Fiber. *Phys. Rev. Lett.* **121**, 190502 (2018).

13. Bussières, F. *et al.* Quantum teleportation from a telecom-wavelength photon to a solid-state quantum memory. *Nat. Photon.* **8**, 775–778 (2014).

14. Takesue, H. *et al.* Quantum teleportation over 100 km of fiber using MoSi superconducting nanowire single-photon detectors. *Optica* **2**, 832–835 (2015).

15. Valivarthi, R. *et al.* Quantum teleportation across a metropolitan fibre network. *Nat. Photon.* **10**, 676–680 (2016).

16. McCarthy, A. *et al.* Kilometre-range, high resolution depth imaging using 1560 nm wavelength

single-photon detection. *Opt. Express* **21**, 8904–8915 (2013).

17. Grein, M. E. *et al.* Design of a ground-based optical receiver for the lunar laser communications demonstration. in *International Conference on Space Optical Systems and Applications* (IEEE, 2011).

18. Hadfield, R. H. Single-photon detectors for optical quantum information applications. *Nat. Photon.* **3**, 696–705 (2009).

19. Becker, W. *Advanced Time-Correlated Single Photon Counting Techniques*. (Springer-Verlag, 2005).

20. Wu, J. *et al.* Improving the timing jitter of a superconducting nanowire single-photon detection system. *Appl. Opt.* **56**, 2195–2200 (2017).

21. Sidorova, M. *et al.* Physical mechanisms of timing jitter in photon detection by current-carrying superconducting nanowires. *Phys. Rev. B* **96**, 184504 (2017).

22. Caloz, M. *et al.* High-detection efficiency and low-timing jitter with amorphous superconducting nanowire single-photon detectors. *Appl. Phys. Lett.* **112**, 061103 (2018).

23. Sidorova, M. *et al.* Timing jitter in photon detection by straight superconducting nanowires: Effect of magnetic field and photon flux. *Phys. Rev. B* **98**, 134504 (2018).

24. Zhao, Q.-Y. *et al.* Single-photon imager based on a superconducting nanowire delay line. *Nat. Photon.* **11**, 247–251 (2017).

25. Calandri, N., Zhao, Q. Y., Zhu, D., Dane, A. & Berggren, K. K. Superconducting nanowire detector jitter limited by detector geometry. *Appl. Phys. Lett.* **109**, 152601 (2016).

26. Zhu, D. *et al.* Superconducting nanowire single-photon detector with integrated impedance-matching taper. *Appl. Phys. Lett.* **114**, 042601 (2019).

27. Smirnov, K. V. *et al.* Rise time of voltage pulses in NbN superconducting single photon detectors. *Appl. Phys. Lett.* **109**, 052601 (2016).

28. Vodolazov, D. Y. Single-Photon Detection by a Dirty Current-Carrying Superconducting Strip Based on the Kinetic-Equation Approach. *Phys. Rev. Appl.* **7**, 034014 (2017).

29. Vodolazov, D. Y. Minimal Timing Jitter in Superconducting Nanowire Single-Photon Detectors. *Phys. Rev. Appl.* **11**, 014016 (2019).

30. Allmaras, J. P., Kozorezov, A. G., Korzh, B. A., Berggren, K. K. & Shaw, M. D. Intrinsic Timing Jitter and Latency in Superconducting Nanowire Single-photon Detectors. *Phys. Rev. Appl.* **11**, 034062 (2019).

31. Zhang, J. *et al.* Time delay of resistive-state formation in superconducting stripes excited by single optical photons. *Phys. Rev. B* **67**, 132508 (2003).

32. Cahall, C. *et al.* Multi-photon detection using a conventional superconducting nanowire single-photon detector. *Optica* **4**, 1534–1535 (2017).

33. Kozorezov, A. G. *et al.* Fano fluctuations in superconducting-nanowire single-photon detectors. *Phys. Rev. B* **96**, 054507 (2017).

34. Dane, A. E. *et al.* Bias sputtered NbN and superconducting nanowire devices. *Appl. Phys. Lett.* **111**, 122601 (2017).

35. Caloz, M. *et al.* Intrinsically-limited timing jitter in molybdenum silicide superconducting nanowire single-photon detectors. *J. Appl. Phys.* **126**, 164501 (2019).

36. Wu, H., Gu, C., Cheng, Y. & Hu, X. Vortex-crossing-induced timing jitter of superconducting nanowire single-photon detectors. *Appl. Phys. Lett.* **111**, 062603 (2017).

37. Nolet, F. *et al.* Quenching Circuit and SPAD Integrated in CMOS 65 nm with 7.8 ps FWHM

Single Photon Timing Resolution. *Instruments* **2**, 19 (2018).

38. Amri, E., Boso, G., Korzh, B. & Zbinden, H. Temporal jitter in free-running InGaAs/InP single-photon avalanche detectors. *Opt. Lett.* **41**, 5728–5731 (2016).

39. Wang, X. *et al.* Oscilloscopic Capture of Greater-than-100 GHz, Ultra-low Power Optical Waveforms Enabled by Integrated Electro-Optic Devices. *J. Light. Technol.* Preprint at https://doi.org/10.1109/JLT.2019.29542 (2019).

40. Ferrari, S., Schuck, C. & Pernice, W. Review article Waveguide-integrated superconducting nanowire single-photon detectors. *Nanophotonics* **7**, 1725–1758 (2018).

41. Pernice, W. H. P. *et al.* High-speed and high-efficiency travelling wave single-photon detectors embedded in nanophotonic circuits. *Nat. Commun.* **3**, 1325 (2012).

42. Mohsen K. Akhlaghi, Haig Atikian, Amin Eftekharian, Marko Loncar, and A. H. M. Reduced dark counts in optimized geometries for superconducting nanowire single photon detectors. **20**, 23610–23616 (2012).

43. Ho Eom, B., Day, P. K., LeDuc, H. G. & Zmuidzinas, J. A wideband, low-noise superconducting amplifier with high dynamic range. *Nat. Phys.* **8**, 623–627 (2012).

44. McCaughan, A. N. & Berggren, K. K. A superconducting-nanowire three-terminal electrothermal device. *Nano Lett.* **14**, 5748–5753 (2014).

45. Kerman, A., Yang, J., Molnar, R., Dauler, E. & Berggren, K. Electrothermal feedback in superconducting nanowire single-photon detectors. *Phys. Rev. B* **79**, 100509 (2009).

46. Clem, J. R. & Berggren, K. K. Geometry-dependent critical currents in superconducting nanocircuits. *Phys. Rev. B* **84**, 174510 (2011).

47. Frasca, S. *et al.* Determining the depairing current in superconducting nanowire single-photon

detectors. *Phys. Rev. B* **100**, 054520 (2019).

Acknowledgements

Part of the research was performed at the Jet Propulsion Laboratory (JPL), California Institute of Technology, under contract with the National Aeronautics and Space Administration (NASA). Support for this work was provided in part by the Defence Advanced Research Projects Agency, Defence Sciences Office, through the Detect program and the National Science Foundation under grant number ECCS-1509486. E.A.B., A.E.D., G.M.C. and J.P.A. acknowledge partial support from the NASA Space Technology Research Fellowship program. E.R. acknowledges support from the MARC-U*STAR program. D.Z. acknowledges support from the A*STAR National Science Scholarship. M.S., S.X. and C.P. acknowledge partial and N.S. full support from the Alliance for Quantum Technologies' (AQT) Intelligent Quantum Networks and Technologies (INQNET) research program. M.S., C.P. and S.X. acknowledge partial support from the Department of Energy, High Energy Physics QuantISED program grant, QCCFP (Quantum Communication Channels for Fundamental Physics), award number DE-SC0019219. C.P. acknowledges partial support from the Fermilab's Lederman Fellowship. We thank Peter Day, Billy Putnam, Daniel Santavicca, and Bill Rippard for valuable discussions and loan of measurement equipment as well as JPL and Caltech staff for technical support. The use of trade names is intended to allow the measurements to be appropriately interpreted and does not imply endorsement by the US government, nor does it imply these are necessarily the best available for the purpose used here.

Author contributions

B.K., Q.-Y.Z. conceived and designed the experiments. B.K., Q.-Y.Z., S.F., J.P.A., E.R., E.A.B., M.J.S., T.M.A., G.M., M.C., C.P., N.S., A.E.V., V.B.V., S.X., D.Z. and A.E.D. performed the experiments. B.K., S.F. and J.P.A. analysed the data. J.P.A. carried out the simulations. B.A.K., Q.-Y.Z., S.F., E.A.B., T.G., M.J.S., T.M.A., G.M., M.C., A.D.B., B.B., R.M.B., C.P., N.S., A.E.L., A.E.V., V.B.V., S.X., D.Z., A.E.D., E.E.W., G.M.C., J.P.A., J.R., P.D.H, K.L.S., R.P.M., M.S., S.W.N., F.M., A.G.K., M.D.S and K.K.B contributed materials/analysis tools. B.K., Q.-Y.Z., J.P.A. and M.D.S wrote the paper with input from all authors.

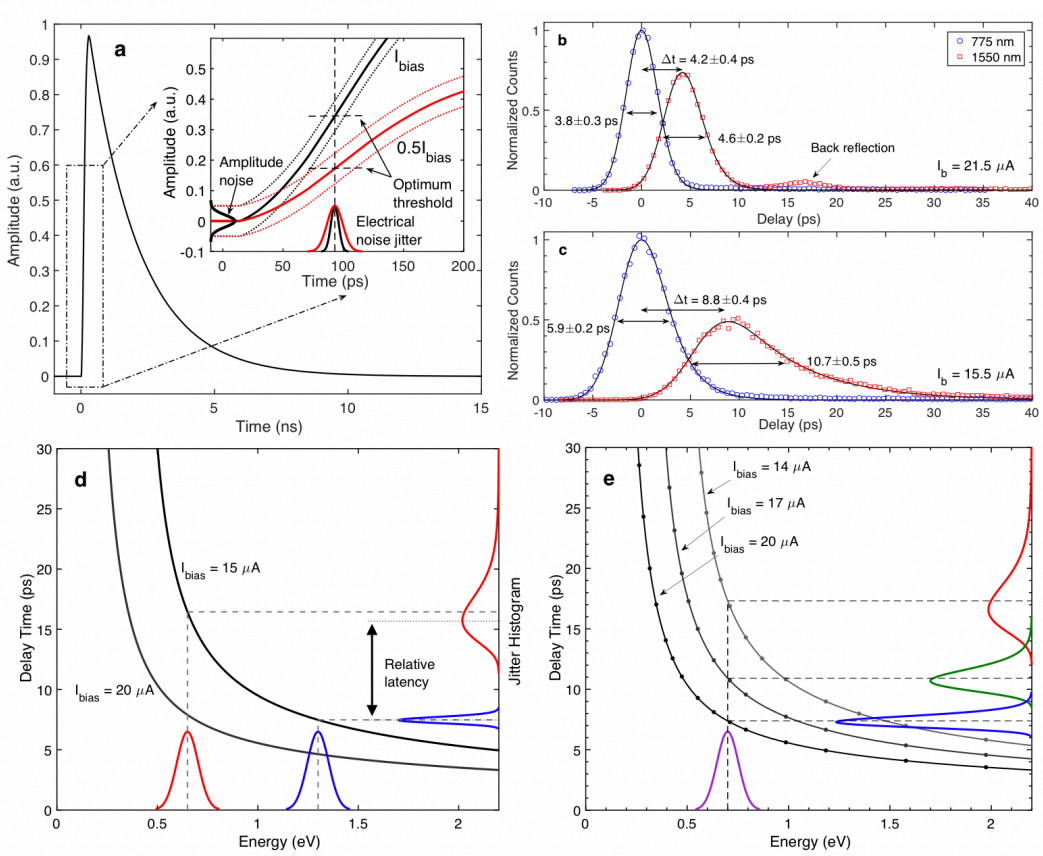


Figure 1

Wavelength and bias current dependence of intrinsic detection latency and jitter as well as the effect of electrical noise jitter. a, A typical SNSPD pulse and effect of slew rate and amplitude noise on the timing jitter. IRF with illumination at 775 nm (blue circles) and 1550 nm (red squares) wavelengths at bias currents of **b,** 21.5 μA and **c,** 15.5 μA for the 80 nm-wide nanowire. The relative latency between the two wavelengths increases at lower bias currents. **b**, The 1550 nm data shows a second peak at a delay of 12.4±0.4 ps, with an amplitude of 6% of the main peak, which is caused by a reflection from the back of the silicon wafer, which is 500 μm thick. This peak was not observed in the 775 nm data, since silicon is not transparent at this wavelength. It will be possible to eliminate this peak in the future by embedding the 1550 nm devices in a broadband optical stack. **d**,**e**, Calculated detection delay as a function of deposited energy in the nanowire with an illustration of the effect of energy fluctuations.

Fluctuations of the energy retained in the nanowire are caused by Fano fluctuations, which leads to an asymmetric distribution of the detection times due to a nonlinear delay curve. Higher energy photons lead to a shorter latency and a narrower jitter distribution. Increasing the bias current reduces the relative latency as well as the jitter.

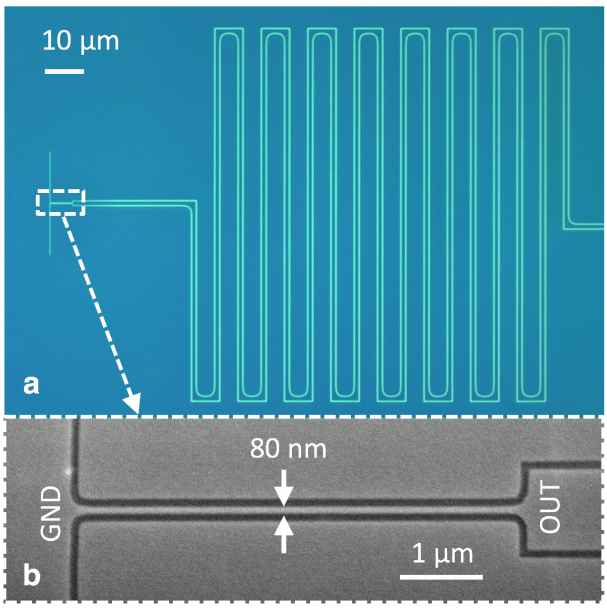


Figure 2

Low-jitter superconducting nanowire single-photon detector. a, Optical micrograph of a representative device, with a small dashed rectangle on the left showing the nanowire region. Darker blue represents the NbN, while the lighter green/blue colour represents the regions removed in the CF_4 etch (see Methods). The large meandering structure on the right is an inductor designed to prevent the short device from latching and is wide enough (1 µm) to prevent it from being a single photon detector itself. **b,** Scanning electron micrograph of the active area. Only the 80 nm-wide nanowire is displayed here, however, four different widths were studied in this work: 60 nm, 80 nm, 100 nm, and 120 nm. The nanowire length of 5 µm length, which is short relative to standard devices, was selected to keep the

predicted longitudinal geometric contribution to the jitter below 1 ps, assuming that the speed of signal propagation in the nanowire is approximately 2% of the speed of light in vacuum.

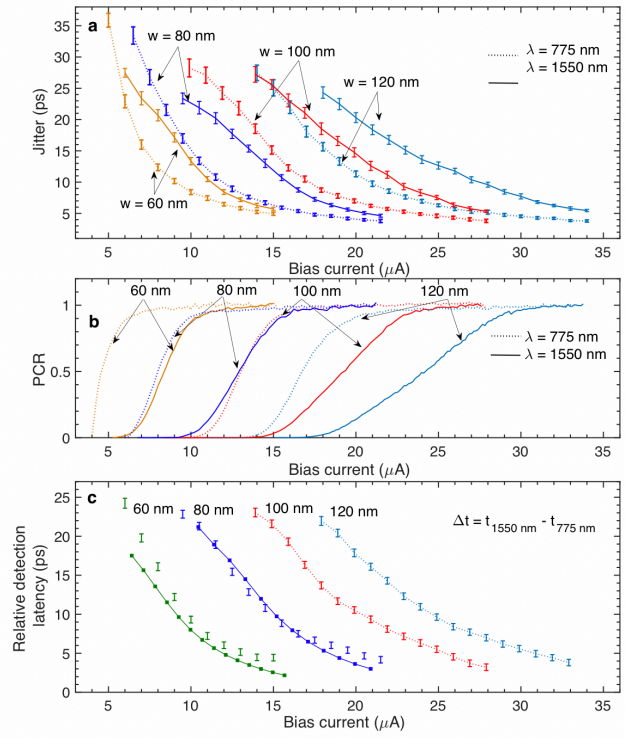


Figure 3

Bias current dependence of the jitter, normalised photon count rate (PCR) and relative detection latency. a, Jitter and **b,** normalised PCR as a function of bias current, for illumination at 775 nm (dashed lines) and 1550 nm (solid lines). Nanowire widths are represented with different colours: 60 nm (orange), 80 nm (blue), 100 nm (red), and 120 nm (pink). The switching currents of the devices at 0.9 K ranged between 16 μA for the 60 nm-wide nanowire and 35 μA for the widest nanowire. **c,** Bias-current dependence of the relative latency between the detection of a 775 nm photon and a 1550 nm photon. Positive values indicate that detections due to 1550 nm photons are delayed relative to 775 nm detections. The solid lines in **c** represent simulated values for the 60 and 80 nm nanowires using the 1D

hotbelt model, which performs well to understand the qualitative behaviour for narrow nanowires. The error bars represent 95% confidence bounds of the data.

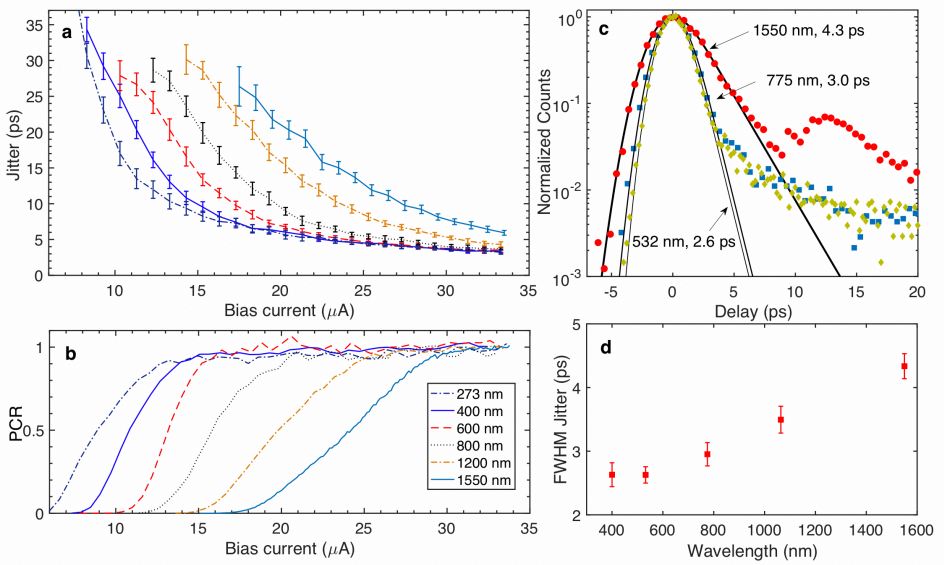


Figure 4

Wavelength dependence of the jitter. a, Jitter and **b,** normalised PCR as a function of bias current for the 120 nm-wide nanowire measured with wavelengths of light between 273 nm and 1550 nm. **c,** Jitter histograms for several wavelengths. It is believed that a significant portion of the tail events is occurring due to detections in the tapering sections of the device, see Supplementary Note 1. The exponentially modified Gaussian function is used for extraction of the FWHM values, but is not expected to fit the true SNSPD IRF even if the tapering detections are eliminated, since the simulations predict a non-Gaussian IRF tail. We observed a change in the delay histogram tail at low bias currents for the ultraviolet light, which can be qualitatively described by a 2D model of detection (see Supplementary Note 2 for details). **d,** Dependence of the FWHM jitter as function of wavelength. We expect that the saturation of the jitter

for wavelengths shorter than 532 nm is due to external instrumental sources. Data in **c,** and **d,** was collected using an 80 GS/s oscilloscope, while data in **a,** was collected with a 40 GS/s scope, leading to a higher noise floor to the jitter (see Methods). The error bars represent 95% confidence bounds of the IRF fit.

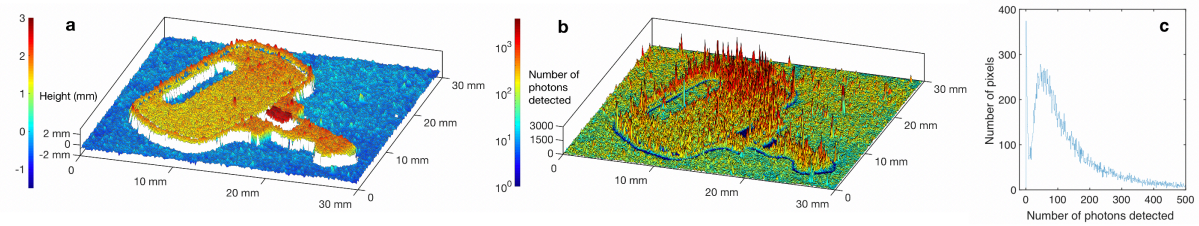


Figure 5

Few-photon scanning laser-ranging with sub-millimetre depth resolution. a, Reconstructed profile of a small key (see Extended Data Figure 5 for a photograph and experimental setup) achieved by detecting and time-tagging reflected photons of a scanned pulsed laser followed by a fit of the IRF to estimate the time of flight at each point of the scan. The dwell time at each position of the scan was fixed to 20 ms meaning that the number of detected photons [shown in **b**] was different at each location. **c,** Distribution of the detected number of photons over all locations of the scan, with the mode of the distribution at around 50 photons. Since no photons were detected at some locations (metallic edges of the key) the IRF fit was conditioned on detecting two or more photons in at least one 1.6 ps bin. The locations where this threshold was not met are represented by gaps in the surface plot in **a**, while all over locations show robust height reconstruction, even revealing some sub-millimetre vertical features such as the step around the edge of the key.

Demonstration of sub-3 ps temporal resolution with a superconducting nanowire single-photon detector

Boris Korzh[1]*, Qing-Yuan Zhao[2], Jason P. Allmaras[1,3], Simone Frasca[1], Travis M. Autry[4], Eric A. Bersin[1,2], Andrew D. Beyer[1], Ryan M. Briggs[1], Bruce Bumble[1], Marco Colangelo[2], Garrison M. Crouch[1], Andrew E. Dane[2], Thomas Gerrits[4], Adriana E. Lita[4], Francesco Marsili[1], Galan Moody[4], Cristián Pena[5,6], Edward Ramirez[1], Jake D. Rezac[4], Neil Sinclair[5], Martin J. Stevens[4], Angel E. Velasco[1], Varun B. Verma[4], Emma E. Wollman[1], Si Xie[5], Di Zhu[2], Paul D. Hale[4], Maria Spiropulu[5], Kevin L. Silverman[4], Richard P. Mirin[4], Sae Woo Nam[4], Alexander G. Kozorezov[7], Matthew D. Shaw[1] and Karl K. Berggren[2]

[1]*Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA*

[2]*Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

[3]*Department of Applied Physics, California Institute of Technology, Pasadena, California 91125, USA*

[4]*National Institute of Standards and Technology, Boulder, Colorado 80305, USA*

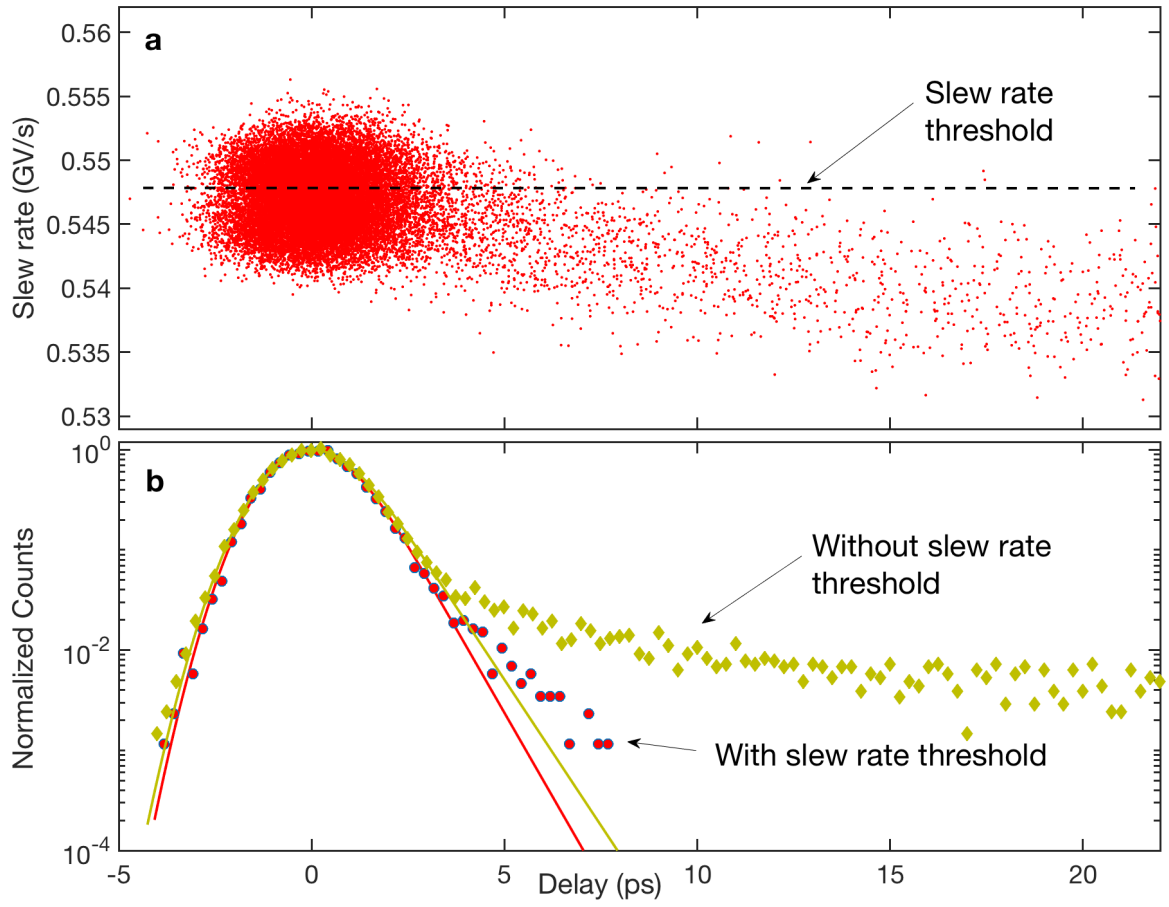
[5]*Division of Physics, Mathematics and Astronomy, California Institute of Technology, Pasadena, California 91125, USA*

[6]*Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA*

[7]*Department of Physics, Lancaster University, Lancaster, LA1 4YB, UK*

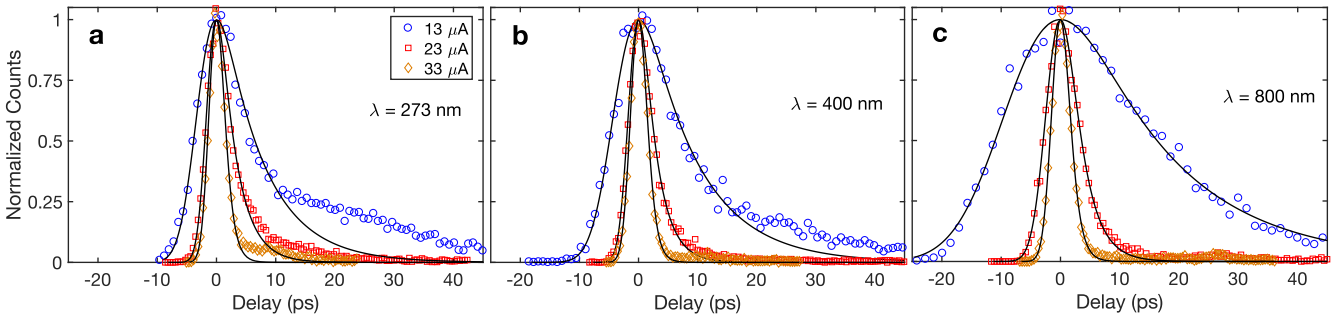
*Correspondence to: bkorzh@jpl.caltech.edu

Supplementary Note 1 | Latency to slew rate correlation. In order to avoid current crowding,[1] the nanowire is connected to the wire terminals with a short tapering section (see Figure 2 in the main manuscript). If a photon detection occurs in this region of widening nanowire, then the overall detection latency will increase (see Figure 3 in main text for effect of nanowire width at a fixed bias current), resulting in a larger number of tail of events in the IRF histogram. As evidence towards this hypothesis we have studied the correlation between the shot-to-shot slew rate of the SNSPD signals and the threshold crossing delay. An additional effect of a detection in a widening section of a nanowire is that during the hotspot expansion, the final resistance will be lower and since the slew rate is proportional to the kinetic inductance (fixed) divided by the hotspot resistance. It means that the slew rate will be lower for detection in the tapering sections. It is therefore possible to use the slew rate as an indicator of whether the detection happened in the tapering sections. Supplementary Figure 1a confirms this since there is a correlation between the slew rate and detection delay – the lower the slew rate, the wider the nanowire section responsible for the detection and hence the larger the delay. In order to infer how the jitter histogram would look like without the effect of the tapering sections we set a threshold in the slew rate such as to reduce the contribution of the tail events. Supplementary Figure 1b shows the resulting IRF with and without the slew rate thresholding, achieving a significant reduction of the tail events. Note that the tail still deviated from exponential, which is expected based on the simulations (see Supplementary Note 3). This analysis indicates that for best performance the light absorption should be concentrated away from tapering sections and bends.



Supplementary Figure 1. a, Correlation between the slew rate of the detection signal and the threshold crossing delay for the 120 nm-wide nanowire biased at the maximum bias current and illuminated with 532 nm light. When excluding all detection below the slew rate threshold indicated in (a) the tail events considerably reduce **b,** indicating that the tapering sections of the nanowire are the main contributing factors to the histogram tail. The exponentially modified Gaussian fit deviates from the experimental data even with the slew rate thresholding, however, it provides an adequate fit for extraction of the FWHM of the distribution.

Supplementary Note 2 | Changes in the IRF tail for short wavelengths. When characterizing the 120 nm-wide nanowire, we noticed that the tail of the distribution for the shortest wavelength light has a significantly different shape, especially at low bias currents. Supplementary Figure 2 shows a comparison of the instrument response functions for three wavelengths (273 nm, 400 nm and 800 nm) for different bias currents. At the lowest bias current of 13 μA, there is a systematic deviation (below 0.25 of the peak) for the 273 nm wavelength data (Supplementary Figure 2a). This deviation decreases at longer wavelengths (Supplementary Figure 2b, c). The origin of this effect can be attributed to differing detection latency for various transverse coordinates of absorption in the nanowire.[2] This effect is known as 'transverse geometric' jitter.



Supplementary Figure 2. Instrument response functions for the 120 nm nanowire biased at 13 μA (circles), 23 μA (squares) and 33 μA (diamonds). Data for the 273 nm (a), 400 nm (b) and 800 nm (c) is shown for comparison.

Supplementary Note 3 | Fitting the experimental data with simulations. The general features a)-d) of the detector latency function discussed in Methods: Detection latency simulations are sufficient for qualitative analysis of the experimental data, but numerical simulation is required to determine the functional form of the latency. To simulate latency curves we numerically solved the system of generalized TDGL equations together with energy balance equations and current continuity equations subject to the appropriate boundary conditions[3]. We used the hotbelt detection model[4–8] and 1D-geometry for simplicity and clarity of presentation, but verified that the results are consistent with the results of the full 2D hotspot detection scenario.[2,7,9,10] The use of the generalized TDGL equations[11] with less stringent validity conditions than the standard TDGL is more appropriate for simulating the suppression of the gap over extended intervals of time when both the superconducting order parameter and the energy gap remain finite. The major parameters of the NbN SNSPD used for this simulation are shown in the Supplementary Table 1. The simulated latency results for the 80 nm SNSPD are shown in Supplementary Figure 3.

It is seen that all the calculated latency curves have features a) to d) described in the Methods of the main manuscript. The bell-shaped curve above horizontal axis in Supplementary Figure 4 schematically shows the normal distribution density of energy deposition into electron system following photon absorption and the right-skewed bell-shaped curve on the right schematically shows the origin and shape of timing jitter distribution originating from fluctuations of detector latency.

i. Shape of jitter distributions. It is easy to see from this figure that because of c) the shape of jitter distribution is distorted Gaussian with long tail at higher delays consistent with data, see Figure 1 of the main manuscript.

ii. Bias current dependence of the jitter and detection latency difference. For any fixed photon energy both latency and jitter FWHM increase with decreasing current, as in Figures 3a and c of the main manuscript.

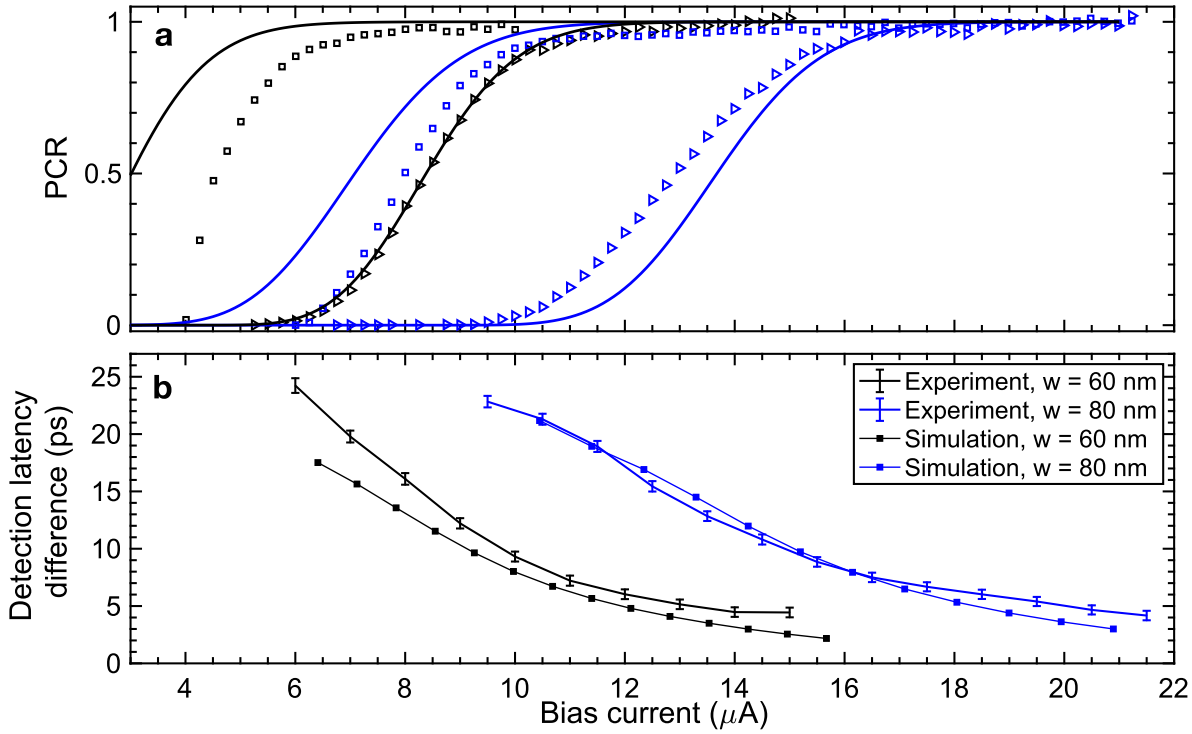
iii. Wavelength dependence of the jitter. For any bias current because of a)-c) both latency and jitter FWHM decrease when photon energy increases as in Figure 4a of the main manuscript.

iv. Width of nanowire dependence. The data in Figure 3 of the main manuscript are easily understood as a result of scaling of current and deposition energy densities with the nanowire width.

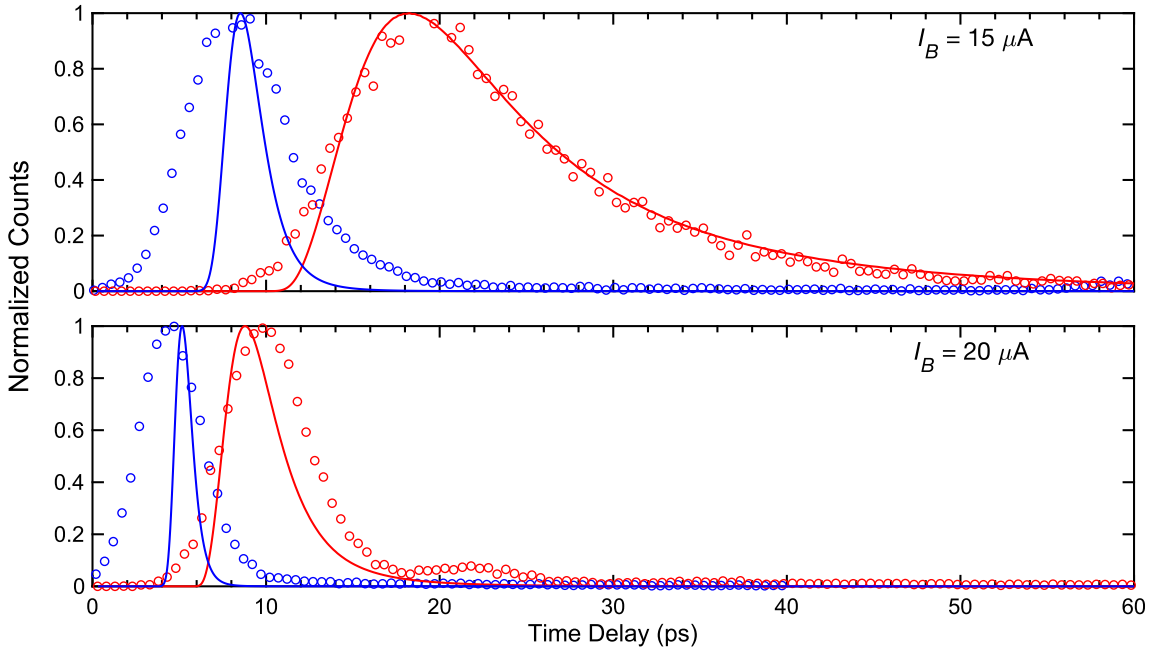
The model used to simulate the detection latency difference shown in Figure 3c of the main manuscript simultaneously predicts the PCR and IRF for different bias conditions and photon energies. With the three fitting parameters $\chi = 0.667$, $\sigma_{1550} = 100$ meV, and $\tau_{ee}(T_c) = 6$ ps, we fit these three key experimental metrics for both the 80 nm and 60 nm wide samples. We note that with these fitting parameters, the high energy tail of the absorbed energy distribution can extend beyond the full photon energy for low photon energies, which is an unphysical result of using the simplified 1D model. If a small contribution of the fluctuations comes from non-uniformities ($\sigma_{1550,\mathrm{n-u}}$) as treated in Ref.[3] such that $\sigma_{1550,\mathrm{Fano}} = 92$ meV and $\sigma_{1550,\mathrm{n-u}} = 40$ meV, this effect is eliminated without changing the quality of the fit. The PCR results are shown for 1550 nm and 775 nm photons in Supplementary Figure 3. The comparison of the IRF for the 80 nm wire is shown in Supplementary Figure 4.

Description	**Symbol**	**Value**
Nanowire width	W	80 nm
Nanowire thickness	d	7 nm
Critical temperature	T_c	8.65 K
Diffusion coefficient	D	0.5 $\mathrm{cm}^2 s^{-1}$
Sheet resistance	R_{sq}	608 Ω/□
Phonon parameter	γ	60
Hotbelt length	L	40 nm
Inelastic electron-electron scattering time at T_c	$\tau_{ee}(T_c)$	6 ps
Electron-phonon coupling time at T_c	$\tau_{ep}(T_c)$	24.7 ps
Phonon escape time	τ_{esc}	20 ps
Mean fraction of the photon energy deposited in the nanowire	χ	0.667
Standard deviation due to Fano fluctuations	$\sigma_{1550,Fano}$	92 meV
Standard deviation due to spatial non-uniformity[3]	$\sigma_{1550,n-u}$	40 meV

Supplementary Table 1. Parameters for simulation of latency in the 80 nm SNSPD.



Supplementary Figure 3 | **a,** Normalized *PCR* for 80 nm (blue symbols) and 60 nm (black symbols) samples for 775 nm light (squares) and 1550 nm light (triangles) compared to simulation (solid lines). **b,** Detection delay difference for 80 nm (blue) and 60 nm (black) samples compared to simulation (squares



Supplementary Figure 4 | Comparison of simulated IRF with experiment for the 80 nm wide nanowire at bias for **a,** 15 μA bias current and **b,** 20 μA bias current. Solid lines indicate the simulated IRF while the symbols correspond to the experimental data. The data in red corresponds to a 1550 nm photon, while blue data corresponds to a 775 nm photon. This illustrates the changes in the relative latency between the two photon energies as well as the change in the IRF width. The agreement with experimental data is not perfect indicating a need for further development of the numerical model.

References

1. Clem, J. R. & Berggren, K. K. Geometry-dependent critical currents in superconducting nanocircuits. *Phys. Rev. B* **84**, 174510 (2011).
2. Vodolazov, D. Y. Minimal Timing Jitter in Superconducting Nanowire Single-Photon Detectors. *Phys. Rev. Appl.* **11**, 014016 (2019).
3. Allmaras, J. P., Kozorezov, A. G., Korzh, B. A., Berggren, K. K. & Shaw, M. D. Intrinsic Timing Jitter and Latency in Superconducting Nanowire Single-photon Detectors. *Phys. Rev. Appl.* **11**, 034062 (2019).
4. Marsili, F. *et al.* Hotspot relaxation dynamics in a current-carrying superconductor. *Phys. Rev. B* **93**, 094518 (2016).
5. Bulaevskii, L. N., Graf, M. J., Batista, C. D. & Kogan, V. G. Vortex-induced dissipation in narrow current-biased thin-film superconducting strips. *Phys. Rev. B* **83**, 144526 (2011).
6. Kozorezov, A. G. *et al.* Quasiparticle recombination in hotspots in superconducting current-carrying nanowires. *Phys. Rev. B* **92**, 064504 (2015).
7. Vodolazov, D. Y. Single-Photon Detection by a Dirty Current-Carrying Superconducting Strip Based on the Kinetic-Equation Approach. *Phys. Rev. Appl.* **7**, 034014 (2017).
8. Kozorezov, A. G. *et al.* Fano fluctuations in superconducting-nanowire single-photon detectors. *Phys. Rev. B* **96**, 054507 (2017).
9. Zotova, A. N. & Vodolazov, D. Y. Photon detection by current-carrying superconducting film: A time-dependent Ginzburg-Landau approach. *Phys. Rev. B* **85**, 024509 (2012).
10. Engel, A., Lonsky, J., Zhang, X. & Schilling, A. Detection Mechanism in SNSPD: Numerical Results of a Conceptually Simple, Yet Powerful Detection Model. *IEEE Trans. Appl. Supercond.* **25**, 2200407 (2015).
11. Kopnin, N. B. *Theory of Nonequilibrium Superconductivity*. (Oxford University Press, 2001).

12. Frasca, S. *et al.* Determining the depairing current in superconducting nanowire single-photon detectors. *Phys. Rev. B* **100**, 054520 (2019).