Intrinsic Timing Jitter and Latency in Superconducting Nanowire Single-photon Detectors

Intrinsic Timing Jitter and Latency in Superconducting Nanowire Single-photon Detectors

J.P. Allmaras,[1,*,†] A.G. Kozorezov,[2] B.A. Korzh,[1] K.K. Berggren,[3] and M.D. Shaw[1]

[1]*Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA*
[2]*Department of Physics, Lancaster University, Lancaster LA1 4YB, United Kingdom*
[3]*Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02138, USA*

We analyze the origin of the intrinsic timing jitter in superconducting nanowire single-photon detectors in terms of fluctuations in the latency of the detector response, which is determined by the microscopic physics of the photon-detection process. We demonstrate that fluctuations in the physical parameters, which determine the latency, give rise to the intrinsic timing jitter. We develop a general description of latency by introducing the explicit time dependence of the internal detection efficiency. By considering the dynamic Fano fluctuations together with static spatial inhomogeneities, we study the details of the connection between latency and timing jitter. We develop both a simple phenomenological model and a more general microscopic model of detector latency and timing jitter based on the solution of the generalized time-dependent Ginzburg-Landau equations for the 1D hotbelt geometry. While the analytical model is sufficient for qualitative interpretation of recent data, the general approach establishes the framework for a more quantitative analysis of detector latency and the fundamental limits of intrinsic timing jitter. These theoretical advances can be used to interpret the results of recent experiments measuring the dependence of detection latency and timing jitter on photon energy to the few-picosecond level.

I. INTRODUCTION

When an incident photon is absorbed by a current-carrying superconducting nanowire, the superconductivity is locally suppressed in a nonequilibrium region known as a hotspot [1]. The nonequilibrium dynamics of this hotspot are a topic of broad interest in superconducting detectors, but precise modeling of the physical process remains an open topic of research. While there have been intense experimental [2–8] and theoretical [9–16] efforts to understand the details of the detection mechanism in superconducting nanowire single-photon detectors (SNSPDs; sometimes referred to as SSPDs in the literature), there is still debate over the most appropriate model for understanding this nonequilibrium process in different regimes of photon energy, bias current, and temperature. Considerable effort has been focused on understanding the internal efficiency of nanowire detectors as a means of validating detection models, but less attention has been given to the timing properties predicted by these models.

The primary timescale determining the detector response is its latency. By definition, the latency of a detector, τ_{lat}, is the time interval between the photoabsorption event and the detector click. During the latency interval, the system evolves through the initial quasiequilibration within the excited volume followed by its expansion, cooling down, and gradual suppression of superconductivity. Finally, if the deposited energy exceeds the detection threshold, instability in a current-carrying nanowire develops and initiates a normal domain and detector click. An important property of SNSPD systems, in practice, is the timing uncertainty associated with each detection event, also known as the timing jitter. There are numerous sources of timing jitter in SNSPD systems. It is now understood that the principal contributions to the timing jitter come from electrical and amplifier noise [17,18], longitudinal geometric jitter due to the finite propagation speed of microwave signals along the length of the nanowire [19,20], timing jitter induced by local inhomogeneities in the nanowire [21,22], and intrinsic timing jitter originating from the microscopic physics of the detection process itself. In a theoretical study, the transverse geometric jitter was investigated by considering the variation in detection latency as a function of the transverse location of photon absorption across the nanowire [23]. An

---

_____
*jallmara@caltech.edu
†Also Department of Applied Physics, California Institute of Technology.

experimental study of the jitter associated with meandered [24] and straight [25] nanowires found asymmetry in the jitter profile, which was attributed to intrinsic effects, and more recently, an increase in timing jitter was measured in straight niobium nitride (NbN) nanowires with increasing magnetic field [26] which was qualitatively explained by the hotspot model [15]. The same model is used to study the effects of transverse position dependence on timing jitter [27]. Experimental progress through the use of specialized short structures and low noise cryogenic amplifiers has enabled the measurement of record low timing jitter in NbN nanowires, as low as 2.7 ps FWHM [28]. The photon energy and temperature dependence of timing jitter observed in these measurements suggests that intrinsic effects, which derive from the physics of the photon-detection process, are relevant. The reduction in instrumental sources of timing jitter to the few-picosecond level has enabled a new type of experiment where the latency difference between photons of two energies is measured directly [28]. By observing the latency distribution as a function of the physical parameters of the device, it is now possible to experimentally validate microscopic models of the photon-detection process more directly. As timing jitter is a critical parameter in many applications, a detailed microscopic theory is an essential tool for engineering higher time-resolution detectors in the future.

We begin by introducing and discussing the general relationship between latency and timing jitter in SNSPDs in Sec. II, and develop a simple phenomenological theory. In the presence of dynamic Fano fluctuations and static spatial nonuniformities, studying and understanding the latency distribution of the detector is sufficient to predict the intrinsic jitter which would be observed in an experiment. We make use of general analytical properties of the detector latency to demonstrate that the non-Gaussian instrument response function observed in recent experiments [28] can be attributed to the positive curvature of the latency vs energy relationship. We analyze the simplified hotbelt-detection model, its prediction capabilities, and its limitations. In Sec. III, we derive and discuss the detector latency using the generalized time-dependent Ginzburg-Landau (TDGL) equations, together with energy balance and current continuity equations. The developed formalism is valid for the general case of 2D geometry. In this paper we analyze the details of solutions for the simpler 1D hotbelt model and demonstrate that the 1D model captures the majority of the experimental observations for narrower wires. Section IV presents a general discussion of the results and directions for future study.

II. LATENCY OF SNSPD RESPONSE AND TIMING JITTER

While the intrinsic latency of photon detection is an important parameter for understanding the performance

of a device, it is challenging to measure directly because of the picosecond timescales involved. While the absolute latency of an SNSPD has still not been characterized directly, recent experiments [28] have measured the latency difference (relative latency) between detection events for photons of two different energies (1550 and 775 nm wavelengths) generated in the same optical path. These experiments were made possible by using a high switching current device and a low-noise cryogenic amplifier to reduce the impact of instrumental noise-induced jitter. Furthermore, the length of the active region of the device was reduced to 5 μm, which decreases the effect of timing jitter associated with the propagation delay of electromagnetic signals. Given an estimated transmission speed of approximately 6 μm/ps [20], this geometric timing-jitter is below 1 ps, and can be neglected. Measurements of the relative detector latency add a valuable piece of information about the timescales of photon detection, providing an extra functional characteristic to test specific models of detection and performance in application. Both dynamic and static fluctuations in the detector will affect the detector latency and manifest themselves in the shape of timing jitter distributions. Therefore, understanding the latency of a detector and its fluctuations is key to predicting its timing jitter.

The timescales of a typical photodetection event in an SNSPD are shown schematically in Fig. 1. For clarity of discussion, we separate the photon-detection process into four stages: (a) the down-conversion cascade, (b) the nondissipative suppression of superconductivity, (c) nucleation and growth of the normal domain, and (d) the diversion of current to the amplifier chain, and the recovery of superconductivity and supercurrent flow. Let us consider a photon of energy E_λ absorbed in a nanowire at an initial instance of time $t = 0$ resulting in the generation of a single electron-hole pair in a metal. The typical energy of the electron and hole of the pair is $E_\lambda/2$. Due to the high energy of the initial excitations (≥ 100 meV) relative to the Fermi level they rapidly relax converting the deposited energy mostly into energetic phonon excitations [16,29] described as the phonon bubble. This happens over the timescale τ_d of a few tens of femtoseconds. The variance $\overline{(\delta\tau_d^2)}$ of this time is due to the distribution of initial energies of the primary *e-h* pair. It is of the order of the scattering time for electron or hole emitting a single phonon with a frequency close to the Debye frequency, therefore, $|\delta\tau_d| \ll \tau_d$. Because τ_d is so small, the duration of this time interval plays no role in any subsequent discussions. It is convenient to consider the phonon bubble as a highly nonequilibrium initial distribution of elementary excitations (predominantly phonons, but with a small number of electronic excitations) with a total energy E_λ and radius $2\sqrt{D\tau_d}$, where D is a characteristic electron diffusion coefficient [16,29]. This is the initial state, from which the evolution of nonequilibrium distribution of interacting phonons and electrons starts and proceeds
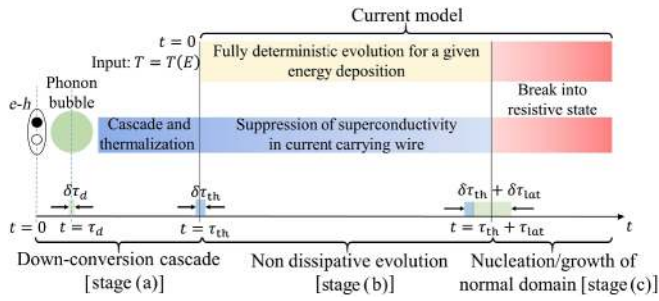
FIG. 1. Schematic of SNSPD detection stages contributing to latency and intrinsic timing jitter. We denote the down-conversion cascade as stage (a), the nondissipative suppression of superconductivity as stage (b), and the nucleation and growth of the normal domain as stage (c).

as an energy down-conversion cascade with rapid multiplication of electron and phonon numbers [15,16]. The down-conversion cascade ends at $t \approx \tau_{\text{th}}$ with the electrons and phonons thermalized at a certain temperature. The estimate of τ_{th} in [15] is in the range of 0.36 and 0.57 ps for WSi and NbN, respectively, for a 1.3-eV photon. This was derived under the assumption that diffusive expansion can be neglected for this down-conversion time. The estimates above serve as an indication of the order of magnitude of this process, and must be corrected to account for the material parameters, diffusion, and density of excitations in the excited volume (photon energy). Accounting for these changes results in a τ_{th} on the order of a few ps for 1550-nm phonons (0.8 eV). For the typical case of a NbN SNSPD, this is smaller (much smaller) than the measured relative latencies, let alone the absolute latency, and $\tau_{\text{th}} \ll \tau_{\text{lat}}$ as in Fig. 1. Similarly, the contribution to the timing jitter from the cascade stage (a) is much smaller than the latency fluctuations over the nondissipative stage (b). The notable exception indicated by the experimental results [28] is when SNSPD is biased close to the depairing critical current when detector latency decreases and comes closer to τ_{th}. In this situation, fluctuations in duration of the stage (a) might contribute to the ultimate limits for the intrinsic timing jitter and will be discussed in more detail in Sec. IV.

The main venue for the current paper is, therefore, the study of the detector latency and its fluctuations over the stage (b) as the dominant source of the intrinsic timing jitter. We expect stage (c) to play only a minor role in contributing to the jitter of the device. Once a normal domain is nucleated at the transition between stages (b) and (c) its growth is a nearly uniform 1D process. For a given bias current, this growth is governed by the balance of Joule heating, diffusion along the length of the wire, and coupling to the substrate. Once the normal domain is formed, this growth is independent of the initial conditions, which formed the domain. There is experimental evidence for this

conclusion based on the photon-energy independence of pulse slew rate recently measured [28]. While a previous work considered timing jitter, which originates during this normal domain growth [22], this process is associated with the presence of nonuniformities along the length of the wire and we exclude this contribution from our definition of intrinsic jitter.

Summarizing, we discuss the model where a thermalized distribution of electrons and phonons (either in the form of a hotbelt or hotspot [15]) at time τ_{th}, is considered as the initial state for the subsequent evolution of the superconducting system in a current-carrying nanowire (top of Fig. 1). For each individual photon count, this state is characterized by an input energy E, which denotes the amount of energy added to the superconducting system as a combination of electronic and phonon excitations in quasiequilibrium at a temperature T. The subsequent evolution of the system is considered as fully deterministic with no other random factors affecting the onset of the resistive transition. We call the period from τ_{th} to $\tau_{\text{th}} + \tau_{\text{lat}}$ in the notation of Fig. 1 the detector latency. It is close to the absolute latency of the detector provided that $\tau_{\text{th}} \ll \tau_{\text{lat}}$. The timing jitter in this picture is associated with the fact that τ_{lat} is a function of input energy E, which fluctuates, depending on the actual energy loss from the individual cascade, and hence from one click to another. The effect of static spatial nonuniformities on latency also can be described in the same way as discussed in detail in the next subsection. All other random sources of timing jitter, for example, those associated with generation of phase-slip lines, vortex-antivortex pair unbinding and escape from the hotspot area, vortex entry from the edges either due to thermal activation or quantum tunneling and vortices moving across the wire are neglected. We expect these effects to be important for spatially uniform wires in the absence of Fano fluctuations or biasing close to the critical depairing current.

We consider a single-photon detection regime and introduce the normalized-efficiency single-photon counting rate, PCR, keeping its explicit dependence on time according to

$$\text{PCR}(t, y, I_B, T_b, B, E) = \Theta\left[t - \tau_{\text{lat}}(y, I_B, T_b, B, E)\right], \quad (1)$$

where $\tau_{\text{lat}}(y, I_B, T_b, B, E)$ is the SNSPD latency, depending on the transverse coordinate y, bias current I_B, bath temperature T_b, external magnetic field B, and energy absorbed by the superconductor electronic system and lattice E. $\Theta(t)$ is the Heaviside function. In this form, the PCR defines the probability of a detector click within the time interval $[0, t]$ due to the absorption of energy E. The energy E is less than the photon energy E_λ due to energy losses during the down-conversion cascade. These losses come from the escape of high-energy phonons from the superconducting film to the substrate and potentially from coupling to

plasmon modes in the superconductor. The details of this energy-loss mechanism are not important in the following analysis. In this paper we do not explicitly discuss the magnetic field dependence, omitting the dependence of B. This form implicitly assumes perfect locality in the description of the photon-absorption site and deterministic evolution of the superconductor after photon absorption. Fluctuations are introduced by multiplying Eq. (1) by the normalized Gaussian probability of energy deposition E,

$$P(E) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(E-\bar{E})^2}{2\sigma^2}}, \qquad (2)$$

with mean value $\bar{E}$ and standard deviation σ to obtain

$$\begin{aligned} &\text{PCR}(t, y, I_B, T_b, E_\lambda) \\ &= \int_0^{E_\lambda} dE P(E) \Theta[I_B - I_{\text{det}}(y, T_b, E)] \\ &\quad \Theta[t - \tau_{\text{lat}}(y, I_B, T_b, E)]. \qquad (3) \end{aligned}$$

This is the general expression that is valid for both hotbelt- (HB) and hotspot- (HS) detection models discussed in the literature [6,9,10,14–16]. We also introduce the photon counting rate averaged over the normalized distribution, $p(y)$ of transverse coordinates of the absorption sites y,

$$\begin{aligned} &\text{PCR}(t, I_B, T_b, E_\lambda) \\ &= \frac{1}{W} \int_{W/2}^{W/2} dy\, p(y) \int_0^{E_\lambda} dE P(E) \Theta[I_B - I_{\text{det}}(y, T_b, E)] \\ &\quad \Theta[t - \tau_{\text{lat}}(y, I_B, T_b, E)] \qquad (4) \end{aligned}$$

for wire width W, which is useful for describing the HS model. The first Heaviside function $\Theta[I_B - I_{\text{det}}(y, T_b, E)]$ in the integrand ensures that an ideal detector clicks with 100% probability once the bias current exceeds the detection current I_{det} for a photon absorption at the site with transverse coordinate y. The second Heaviside function $\Theta[t - \tau_{\text{lat}}(y, I_B, T_b, E)]$ allows the click to occur only after a deterministic interval of time, the detector response latency τ_{lat}, has elapsed. During this time following the absorption of a photon, a strongly nonequilibrium state of interacting quasiparticles and phonons evolves in time and space, suppressing the superconducting order parameter. At a certain stage, the superconducting state may become unstable. In the HS scenario [11,15], either vortices enter from the edge of the wire or vortex-antivortex pairs are unbound inside the hotspot. In the HB scenario, phase-slip lines are generated. If the bias current exceeds a certain threshold, which we call the detection current, energy dissipation in the current-carrying nanowire results in the nucleation and growth of a normal domain through multiple vortex crossings or the formation of phase-slip lines.

A. Role of spatial nonuniformities

The standard deviation $\sigma = \sqrt{\sigma_{\text{Fano}}^2 + \sigma_{\text{nu}}^2}$ describes the combined effect of Fano fluctuations (variance σ_{Fano}^2) [16] and spatial nonuniformity of the wire (variance σ_{nu}^2). Spatial nonuniformity is assumed to originate from the spatial variation of parameters such as the local geometry (thickness of the wire) and local material parameters (critical temperature, density of states, and electron diffusivity). The use of the standard deviation σ in the form of the quadrature of the two statistically independent contributions as an approximation is justified, because the impact of the local nonuniformity can be described by fluctuations of the initial temperature in the excited volume for a fixed energy deposited in the nanowire. This in turn is formally equivalent to a homogeneous medium where local temperature fluctuations originate from fluctuating energy depositions with the appropriate variance. Below, we introduce a simple model to take into account the effect of spatial nonuniformity.

The Fano fluctuations originate during the energy downconversion cascade due to the energy loss from the thin nanowire film. While the Fano fluctuations are dynamic in nature, the spatial inhomogeneity of the wire is static in time. To derive the expression for σ_{nu}, we write down the energy conservation law [15] in the form

$$\mathcal{E}_e(T_i) + \mathcal{E}_{\text{ph}}(T_i) = \mathcal{E}_e(T_b) + \mathcal{E}_{\text{ph}}(T_b) + E, \qquad (5)$$

where $\mathcal{E}_e(T_i)$ and $\mathcal{E}_{\text{ph}}(T_i)$ are the energies for the equilibrated distributions of electrons and phonons at temperature T_i some short time following photon absorption and E is the energy gained by the system following the downconversion process. For temperatures exceeding T_c after the cascade, the energy of the system can be expressed as $\mathcal{E}_e(T) + \mathcal{E}_{\text{ph}}(T) = \frac{1}{2}\bar{T}_c V_i \left[C_e(\bar{T}_c)^2 + \frac{1}{2} C_{\text{ph}}(\bar{T}_c)^4 \right]$, where the top bar designates the mean value. In the spatially nonuniform nanowire we have $V_i = \bar{V}_i + \delta V_i$, $C_e(\bar{T}_c) = \bar{C}_e + \delta C_e$ and $C_{\text{ph}}(\bar{T}_c) = \bar{C}_{\text{ph}} + \delta C_{\text{ph}}$ and δV_i, δC_e, δC_{ph} describe spatial fluctuations with zero mean values. Here V_i is the initial volume and C_e and C_{ph} are electron and phonon specific heat capacities, respectively. For the HS model [15], $V_i = 4\xi_c^2 d$, where $\xi_c^2 = \hbar D/k_B T_c$, and d is the nanowire thickness. For the HB model, $V_i = LWd$, where L is the length of a rectangular hotbelt along the nanowire of width W. Fluctuations in the initial volume of HS originate from both the intrinsic parameters, the diffusion coefficient D, and the critical temperature T_c, and also due to varying thickness of the nanowire. In the HB model, fluctuations are of purely geometric origin. Spatial fluctuations will result in fluctuations of the temperature of the initial quasiequilibrated state even if deposited energy is exactly the same for all absorption sites. Note that if the spatial fluctuations are characterized by a sufficiently small length scale $r \ll L, W$, their effects may be strong in the HS case but will be substantially self-averaged and weakened

for the HB scenario. For small-scale spatial fluctuations $r \ll 2\xi_c$ their effect is substantially weakened also for the HS scenario.

For Fano fluctuations, the fluctuation in deposited energy, δE, results in the fluctuation of the initial temperature, $\delta E = V_i \left[\bar{C}_e(T) + \bar{C}_{\mathrm{ph}}(T) \right] \delta T$. The respective variance is $\sigma_{\mathrm{Fano}}^2 = \overline{(\delta nE)^2} = \left(T/\bar{T}_c \right)^2 V_i^2 \left[\bar{C}_e + \bar{C}_{\mathrm{ph}} \left(T/\bar{T}_c \right)^2 \right]^2 \overline{(\delta T)^2}$. In order to characterize the effect of spatial nonuniformities we introduce the effective fluctuations of deposited energy, δE_{nu}. In the effective homogeneous medium we connect them to δT_{nu} according to $\delta E_{\mathrm{nu}} = V_i \left[\bar{C}_e(T) + \bar{C}_{\mathrm{ph}}(T) \right] \delta T_{\mathrm{nu}}$. Correspondingly we obtain $\sigma_{\mathrm{nu}}^2 = \overline{(\delta E_{\mathrm{nu}})^2} = \left(T/\bar{T}_c \right)^2 V_i^2 \left[\bar{C}_e + \bar{C}_{\mathrm{ph}} \left(T/\bar{T}_c \right)^2 \right]^2 \overline{(\delta T_{\mathrm{nu}})^2}$. The connection between δT_{nu} and δV_i, δC_e and δC_{ph} is obtained from solving $\delta\{ V_i \left[C_e(x^2 - x_b^2) + 1/2 C_{\mathrm{ph}}(x^4 - x_b^4) \right] \} = 0$ with fixed energy deposition, E ($\delta E = 0$), so that $x \equiv T(E)/\bar{T}_c + \delta T_{\mathrm{nu}}$. Fluctuations in δC_e, δC_{ph}, and δV_i may be considered as statistically independent, the first two being proportional to fluctuations in densities of states for electrons and phonons, respectively, while fluctuations of the excited volume contain independent contributions from elastic scattering (diffusion coefficient) and wire thickness and width. Therefore, after straightforward calculations we obtain, for the variance describing the spatial nonuniformity effects, a simple expression

$$
\sigma_{\mathrm{nu}}^2 = E^2 \left[\overline{\left(\frac{\delta V_i}{V_i} \right)^2} + \overline{\left(\frac{\delta C_e}{C_e} \right)^2} \nu^2 + \overline{\left(\frac{\delta C_{\mathrm{ph}}}{C_{\mathrm{ph}}} \right)^2} (1 - \nu)^2 \right]
$$
$$
= \Xi^2 E^2, \tag{6}
$$

where $\nu = \left(x^2 - x_b^2 \right) / \left[x^2 - x_b^2 + 4(\pi^2/5\gamma) \left(x^4 - x_b^4 \right) \right]$ accounts for the fraction of energy in the electronic system, $x_b = T_b/\bar{T}_c$. For $\nu \to 0$ or $\nu \to 1$ we have the standard deviation describing the nonuniformity effect scaling as the linear power of energy, $\sigma_{\mathrm{nu}} \sim E$. It may become sublinear in E only when $\overline{(\delta C_e/C_e)^2} \nu^2 \gg \overline{(\delta V_i/V_i)^2} + \overline{\left(\delta C_{\mathrm{ph}}/C_{\mathrm{ph}} \right)^2} (1 - \nu)^2$ with $\nu < 1$. This is an unlikely situation, because strong disorder in nanowire affects both electron and phonon properties, besides the remaining factor $\overline{(\delta V_i/V_i)^2}$ is expected to be significant. For $\overline{(\delta V_i/V_i)^2} \approx \overline{\left(\delta C_{\mathrm{ph}}/C_{\mathrm{ph}} \right)^2} \approx \overline{(\delta C_e/C_e)^2}$ variation of $\Xi(E, \gamma)$, where $\gamma = 8\pi^2/5 \left(\bar{C}_e/\bar{C}_{\mathrm{ph}} \right)$, does not exceed 5% for 0.1 eV $< E < 3$ eV and for the range $10 < \gamma < 100$, and thus we take $\sigma_{\mathrm{nu}} \sim E$.

The main contribution to Fano fluctuations comes from the down-conversion phonons of the first few generations, when almost all the photon energy has been transferred to phonons of high (approximately Debye) energy. Correspondingly $\sigma_{\mathrm{Fano}}^2 = F_{\mathrm{eff}} \varepsilon E$, where $F_{\mathrm{eff}} = \left[\Omega_D/\Delta(0) \right] \left(l_{\mathrm{ph},D}/d \right) f \left(l_{\mathrm{ph},D}/d \right)$, and $\varepsilon = 1.75 \Delta(0)$. Here F_{eff} is the effective Fano factor, and $\Delta(0)$ is the zero

temperature and zero current order parameter value, $l_{\mathrm{ph},D}/d$ is the ratio of the Debye phonon mean free path to the thickness of the film, and the function $f \left(l_{\mathrm{ph},D}/d \right) < 1$ describes phonon survival chance to reach the escape interface(s) and get transmitted into the substrate [16].

The total variance of energy fluctuations is given by the sum of the Fano and nonuniformity variances

$$
\sigma^2 = \sigma_{\mathrm{Fano}}^2 + \sigma_{\mathrm{nu}}^2 = F_{\mathrm{eff}} \varepsilon E_\lambda + \chi^2 \Xi^2 E_\lambda^2
$$
$$
= \sigma_{\mathrm{Fano}}^2(E_\lambda) \left(1 + E_\lambda/E_{\lambda_0} \right), \tag{7}
$$

where χ is the mean fraction of the photon energy deposited in the nanowire and E_{λ_0} is the reference photon energy for which Fano variance equals the one due to nonuniformity, $E_{\lambda_0} = F_{\mathrm{eff}} \varepsilon / \chi^2 \Xi^2$. Therefore, if experiments cover a broad range of wavelengths, Fano fluctuations dominate the jitter at the long-wavelength end of the spectrum $\lambda > \lambda_0$, while spatial nonuniformities become more important for shorter wavelengths $\lambda < \lambda_0$. This is the typical situation in single-photon-superconducting tunnel-junction detectors, where the Fano-fluctuation limit of spectral resolution is reached for energies below the hard ultraviolet, and spatial inhomogeneities become important at higher energies [30].

B. Timing jitter: hotspot model

The time-dependent photon counting rate $\mathrm{PCR}(t, y, I_B, T_b, E_\lambda)$ gives the probability of a detector click due to the absorption of a single photon within the time interval $[0, t]$. Correspondingly, $H(t, y, I_B, T_b, E_\lambda)dt = [d\mathrm{PCR}(t, y, I_B, T_b, E_\lambda)/dt]dt$ is the probability of a single-photon detection process within the time interval $[t, t + dt]$, where $H(t, y, I_B, T_b, E_\lambda) = d\mathrm{PCR}(t, y, I_B, T_b, E_\lambda)/dt$ is the click probability density, also known as the instrument response function (IRF) for single-photon absorption events. To calculate the IRF, which is often the observed characteristic, we define the quantity $E(t, y, I_B, T_b)$, which represents such an energy deposition to the superconductor, which corresponds to detector latency equal to t. This amount of energy is given by the single-valued solution of

$$
t - \tau_{\mathrm{lat}}(y, I_B, T_b, E) = 0. \tag{8}
$$

Single valuedness follows from the fact that the detector latency can only be a monotonically decreasing function of the energy deposition, $\partial \tau_{\mathrm{lat}}(y, I_B, T_b, E)/\partial E < 0$. Using the definition of $E(t, y, I_B, T_b)$, we rewrite expression (3) in the form

$$
\mathrm{PCR}(t, y, I_B, T_b, E_\lambda) = \int_0^{E_\lambda} dE P(E) \Theta[I_B
$$
$$
- I_{\mathrm{det}}(y, T_b, E)] \Theta \left[E - E(t, y, I_B, T_b) \right]
$$
$$
= \int_{E(t,y,I_B,T_b)}^{E_\lambda} dE P(E) \Theta \left[I_B - I_{\mathrm{det}}(y, T_b, E) \right]. \tag{9}
$$

Differentiating (9) we obtain the IRF in the form

$$
H(t,y,I_B,T_b) = -P[E(t,y,I_B,T_b)]\Theta\{I_B \\
- I_{\text{det}}[y,T_b,E(t,y,I_B,T_b)]\}\frac{\partial E(t,y,I_B,T_b)}{\partial t}.
$$
(10)

The single valuedness of the solution (8) and negative derivative of latency as a function of energy ensures that the derivative $\partial E(t,y,I_B,T_b)/\partial t$ in expression (10) is negative. A positive derivative implies the unphysical possibility of the PCR decreasing with time.

The expression (10) for the IRF is exact under the assumptions of our model. Defining $\bar{t} = \tau_{\text{lat}}(y,I_B,T_b,\bar{E})$ and $\bar{E} = E(\bar{t},y,I_B,T_b)$, we transform Eq. (10) to obtain

$$
H(t,y,I_B,T_b) \\
= -\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{\left[E(t,y,I_B,T_b) - E(\bar{t},y,I_B,T_b)\right]^2}{2\sigma^2}\right\} \\
\times \Theta\{I_B - I_{\text{det}}[y,T_b,E(t,y,I_B,T_b)]\}\frac{\partial E(t,y,I_B,T_b)}{\partial t}.
$$
(11)

The derived IRF as seen from Eq. (11) is non-Gaussian. This is the result of the nonlinearity of the function $E(t,y,I_B,T_b)$ versus time t. By definition the timing jitter is FWHM of the histogram (11).

For small standard deviation σ, we may derive the dominant term in the approximation to $H(t,y,I_B,T_b)$. Taking a series expansion up to linear terms in $(t - \bar{t})$, $E(t,y,I_B,T_b) - E(\bar{t},y,I_B,T_b) = \partial E(t,y,I_B,T_b)/\partial t|_{t=\bar{t}}(t-\bar{t}) = [\partial \tau_{\text{lat}}(y,I_B,T_b,E)/\partial E|_{E=\bar{E}}]^{-1}(t-\bar{t})$ and substituting the result into Eq. (11) yields

$$
H(t,y,I_B,T_b) \\
= \frac{1}{\sqrt{2\pi}\sigma_j(y,I_B,T_b,\bar{E})}\exp\left[-\frac{(t-\bar{t})^2}{2\sigma_j{}^2(y,I_B,T_b,\bar{E})}\right] \\
\times \Theta\left[I_B - I_{\text{det}}(y,T_b,\bar{E})\right],
$$
(12)

where $\sigma_j(y,I_B,T_b,\bar{E}) = \sigma|\partial \tau_{\text{lat}}(y,I_B,T_b,\bar{E})/\partial \bar{E}|$. From Eq. (12), it follows that the shape of the approximate IRF is a Gaussian. Defining its FWHM as the Gaussian part of the contribution to the timing jitter Υ_G, we have

$$
\Upsilon_G \approx 2.355\sigma_j(y,I_B,T_b,\bar{E}) \\
\approx 2.355\sigma\left|\frac{\partial \tau_{\text{lat}}(y,I_B,T_b,\bar{E})}{\partial \bar{E}}\right|_{\bar{E}=\chi E_\lambda}.
$$
(13)

Averaging the histogram over the transverse coordinate of photon absorption, we derive an approximate result for the mean histogram

$$
H(t,I_B,T_b) = \frac{1}{W}\int_{-W/2}^{W/2}dy\,p(y)\frac{1}{\sqrt{2\pi}\sigma_j(y,I_B,T_b,\bar{E})} \\
\times \exp\left[-\frac{(t-\bar{t})^2}{2\sigma_j{}^2(y,I_B,T_b,\bar{E})}\right]\Theta\left[I_B - I_{\text{det}}(y,T_b,\bar{E})\right].
$$
(14)

This averaging results in a distortion of the ideal local Gaussian IRF.

It is important to emphasize that the detector latency and its derivative $|\partial \tau_{\text{lat}}(y,I_B,T_b,E)/\partial E|$ may both exhibit a singularity. They are finite if

$$
I_B > I_{\text{det}}(y,T_b,E) \quad \text{or} \quad E > E_{\text{det}}(y,I_B,T_b),
$$
(15)

where $E_{\text{det}}(y,I_B,T_b)$ is the detection (or cutoff) energy. By definition of the detection current or detection energy, $\tau_{\text{lat}}(y,I_B,T_b,E) = \infty$ for either $I_B \leq I_{\text{det}}(y,T_b,E)$ or $E \leq E_{\text{det}}(y,I_B,T_b)$ and becomes finite at $I_B > I_{\text{det}}(y,T_b,E)$ or $E > E_{\text{det}}(y,I_B,T_b)$. Therefore, at $I_B = I_{\text{det}}(y,T_b,E)$ or $E = E_{\text{det}}(y,I_B,T_b)$, the detector latency and its derivative either diverge or exhibit discontinuity. Vodolazov [27] argues that due to the dynamical nature of the hotspot there exists a trade-off between the current-dependent time window for the destruction of superconductivity within the expanding hotspot and the current-dependent order parameter relaxation time, which provides a maximum finite value for the latency. We believe that for a complicated nonequilibrium system such as a current-carrying nanowire with an expanding hotspot or hotbelt with slowly varying temperature profiles, any trade-off must include the details of the energy-dissipation profiles. The ultimate test of the actual functional behavior of the detector latency is the microscopic theory. Here we hypothesize that a weak (integrable) singularity is present in the latency in the vicinity of I_{det} and E_{det}. This assumption is useful for the analysis of IRF functions, but because the IRF functions are integrals, the presence of a weak singularity is not important. In Sec. III we confirm the presence of weak singularities in latency vs energy for the range of bias currents for 1D hotbelt geometry by direct simulations.

The important general feature of detector latency as a function of photon energy is its positive curvature. This is intuitively sensible, because the increase in energy due to photon absorption cannot result in an instantaneous break of superconductivity. Although the initial temperature of the quasiequilibrated distributions of electrons and phonons increases with deposited energy, the latency can only asymptotically approach a small but nonzero value. As a result, the rate of decrease in the latency slows down at higher energies. This asymptotic behavior combined with the singularity at the detection energy leads to a general positive curvature of the latency as a function of

photon energy. In Sec. III, we examine the detector latency by solving the generalized TDGL equations and explicitly demonstrate the accuracy of this statement. The same conclusion is reached in [27] using the standard TDGL model in two dimensions. An important consequence of the positive curvature of the latency function can be immediately seen from Eqs. (8) and (11): for latency time increasing beyond $\bar{t}$, further reduction of $E(t, y, I_B, T_b)$ in the argument of the exponential function in Eq. (11) becomes more gradual, resulting in the formation of an extended tail of the IRF.

The presence of the singularity at $I_{\text{det}}(y, T_b, E)$ results in the general trend of the detector latency τ_{lat} and jitter Υ increasing when the bias current decreases, $I_B \to I_{\text{det}}$, for all other parameters fixed and assuming noiseless amplification. If the bias current and other parameters except photon energy are fixed, then both the detector latency and jitter $\left[\sim \left|\partial \tau_{\text{lat}}(y, I_B, T_b, E)/\partial \bar{E}\right|_{\bar{E}=\chi E_\lambda}\right]$ monotonically increase as the photon energy decreases to $E_{\text{det}}(y, I_b, T_b)$, due to the latency being a monotonically decreasing function of energy deposition with positive curvature, as is observed in experiments [28].

It is worth emphasizing that our formalism is general. In particular, the influence of magnetic field on timing jitter can be examined once the dependencies of latency and detection current and energy on B are reinstated in all expressions of this subsection. In this way, we can reproduce the results of recent work [26]. The asymmetry of the jitter histogram as shown by our main result, Eq. (11), is connected with the positive curvature of latency vs energy deposition. Long-delay tail occurs naturally already in zero magnetic field. In the 2D hotspot scenario there appears the component depending on transverse coordinate y both in the Gaussian part of the histogram and in the tail. In increased magnetic fields it becomes more prominent due to transformation of the profiles of the detection energy, detection current, and latency in the magnetic field following exactly the same argument as discussed by the authors of Ref. [26]. The difference in our approach, however, is that (i) the histogram derived within the deterministic model is not an exponentially modified Gaussian, and (ii) Fano fluctuations affect both the Gaussian part and the tail. Another recent work [27] studies contributions to intrinsic timing jitter in the hotspot scenario originating from the dependence of detector latency (delay time) on the transverse coordinates of absorption sites and also with the account of Fano fluctuations. The model of intrinsic timing jitter used in this work is the same as our approach. The IRF functions for the 2D geometry can be obtained and analyzed within our general formalism.

C. Timing jitter: hotbelt model

In the hotbelt scenario, $I_{\text{det}}(y, T_b, E)$ does not depend on y. Solving $I_B - I_{\text{det}}(T_b, E^*) = 0$ for $E^* = E_{\text{det}}(I_B, T_b)$ and

replacing $\Theta[I_B - I_{\text{det}}(T_b, E)]$ by $\Theta[E - E_{\text{det}}(I_B, T_b)]$, we obtain

$$\text{PCR}(t, I_B, T_b, E_\lambda)$$
$$= \int_{E_{\text{det}}(I_B, T_b)}^{E_\lambda} dEP(E)\Theta[t - \tau_{\text{lat}}(I_B, T_b, E)]. \quad (16)$$

Normalized PCRs in the familiar form [16] are obtained from Eqs. (3–4) by taking the limit $t \to \infty$ and $\Theta[t - \tau_{\text{lat}}(I_B, T_b, E)] \to 1$. This results in

$$\text{PCR}(I_B, T_b, \chi, E_\lambda) = \frac{1}{2}\left\{\text{erf}\left[\frac{\chi E_\lambda - E_{\text{det}}(I_B, T_b)}{\sqrt{\sigma}}\right] \right. $$
$$\left. + \text{erf}\left[\frac{E_\lambda(1-\chi)}{\sqrt{\sigma}}\right]\right\}, \quad (17)$$

where $E_{\text{det}}(I_B, T_b)$ is determined from the energy conservation law. This conservation law can be written

$$\mathcal{E}_e\left[\left(1 - I_B^{2/3}\right)^{1/2}\right] + \mathcal{E}_{\text{ph}}\left[\left(1 - I_B^{2/3}\right)^{1/2}\right]$$
$$= \mathcal{E}_e(T_b) + \mathcal{E}_{\text{ph}}(T_b) + E_{\text{det}}(I_B, T_b), \quad (18)$$

where I_B is in units of the zero-temperature critical depairing current. The solution of Eq. (18) determines the threshold for energy deposition into a current-carrying superconductor with initial temperature T_b such that the critical current of the superconductor heated by this energy deposition becomes equal to the bias current I_B, using for the former the Bardeen relation. Similarly, for a given energy deposition E the solution of Eq. (18) for current defines the detection current, $I_{\text{det}}(E, T_b)$. Substituting Eq. (18) into Eq. (17) yields the explicit expression for PCR $(I_B, T_b, \chi, E_\lambda)$. For the hotbelt model, neglecting the y dependence of the detector latency and detection current and simplifying Eq. (11), we arrive at

$$H(t) = -\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{[E(t, I_B, T_b) - \bar{E}]^2}{2\sigma^2}\right\}$$
$$\times \frac{dE(t, I_B, T_b)}{dt}. \quad (19)$$

The approximate (Gaussian) IRF is similarly obtained from the general result of the hotspot model, acquiring the forms of Eq. (12) and (13) where the detector latency, detection current, and variance are independent of the transverse coordinate y.

D. Understanding intrinsic jitter in SNSPDs using the simplified hotbelt model

The phenomenological model described above is sufficient to understand and qualitatively interpret the results of

recent experiments [28] based on the analytical properties of the latency function, its singularities, and its monotonic decrease and positive curvature with increasing energy deposition. Such an analysis becomes especially straightforward for the hotbelt model, where the analytical forms for the PCR can be used for fitting and for the initial selection of material parameters. However, a more predictive model requires a more accurate determination of the latency of the detector for an arbitrary deposition energy, which is needed to interpret experimental results of the relative detection latency between a pair of photons with different energies [28]. This is sensitive to the detection model, and requires a more sophisticated approach and more extensive numerical analysis. The practical value of the above analysis is that we only need to study the detector latency and fluctuations in order to generate the IRF.

In this subsection we use the simplified hotbelt model. It is the easiest way to introduce fluctuations and discuss the shape of the photon counting rates as a function of the device parameters [16], and it leads to useful qualitative insights into the physics of the detector latency and the intrinsic timing jitter. In Sec. III, we develop a more advanced hotbelt model based on the solution of the 1D generalized time-dependent Ginzburg-Landau equations together with current continuity and thermal-balance equations. We leave the analysis of detector latency in the context of a fully general hotspot model based on the generalized 2D TDGL equations for future work.

The main results of the simplified hotbelt model are described above by Eqs. (17), (18), and (19). To illustrate the predictions of the model, we consider the specific case of an NbN SNSPD described in [28]: $W = 80$ nm, $d = 7$ nm, critical temperature $T_c = 8.65$ K, $D = 0.5$ cm^2/s, $R_{sq} = 608$ Ω/sq, and $N(0) = 1/(2e^2 DR_{sq}d) = 14.7 \times 10^{21}$ cm^{-3}/eV. Following [31], we estimate the depairing current at zero temperature for the nanowire using $I_{dep}(0) = 1.491 N(0)e[\Delta(0)]^{3/2}(D/\hbar)^{1/2} Wd$, arriving at $I_{dep}(0) \approx 25.8$ μA, where $\Delta(0)$ is the superconducting order parameter at zero temperature. The parameter γ is estimated to be 60 based on the acoustic properties of NbN [32]. Figure 2 shows the detection energy and detection current for a hotbelt with $L = 40$ nm calculated from Eq. (18). Using the calculations of the detection energy from Eq. (18) shown in Fig. 2(a) and expression (17), we can calculate PCR curves.

Figure 3 shows the simulated PCR curves for 1550 and 775 nm photons at $T_b = 1$ K, which closely follow the data presented in [28] if we use $\chi = 0.40$ and $\sigma_{Fano} = 0.1$ eV. This value of χ can be justified as follows: energy loss from the quasiequilibrated quasiparticles and phonons making up the hotbelt occurs due to the out-diffusion of quasiparticles and the escape of phonons. Cooling of the hotbelt and the order-parameter suppression takes a substantial part of the detector-latency interval, terminating with the SNSPD entering the resistive state. During this
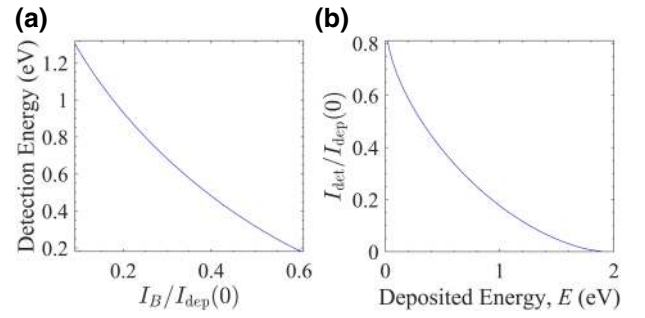

FIG. 2. (a) Detection energy E_{det} vs bias current I_b normalized by the depairing critical current at zero temperature $I_{dep}(0)$. The detection energy at a given I_b defines the energy deposition E above which the detector generates a click. (b) Normalized detection current $I_{det}/I_{dep}(0)$ vs deposited energy E. The detection current defines the current above which the detector generates a click for a given amount of deposited energy. Results are calculated using the hotbelt model for bath temperature $T_b = 1$ K and the material parameters described in the text.

stage, the cooling rate is determined by a combination of the phonon escape rate τ_{esc}^{-1} and the thermal conductivity, both of which are independent of the bias current and the initial state determined by the photon energy. Therefore, the energy loss from electrons and phonons is determined by the ratio $\tau_{lat}(I_B, T_b, \bar{E})/\min\{\tau_{esc}, LW/4D\}$; the larger this ratio is, the more energy is lost. The energy of the quasiparticles and phonons in the hotbelt is a monotonically decreasing function of this ratio. When $\tau_{lat}(I_B, T_b, \bar{E})/\min\{\tau_{esc}, LW/4D\} \ll 1$, there is no loss of photon energy. In the opposite case, i.e., $\tau_{lat}(I_B, T_b, \bar{E})/\min\{\tau_{esc}, LW/4D\} \gg 1$, the loss is substantial. Approximately half the photon energy is lost when $b\tau_{lat}(I_B, T_b, \bar{E})/\min\{\tau_{esc}, LW/4D\} = 1$, where $b \leq 1$ is a numerical factor of order unity accounting for the fact that only the first part of the latency interval is dissipationless until the superconducting current flow becomes unstable. In typical situations, $\tau_{esc} = 4d/\eta c$ [33] where $d \geq 5$ nm, the phonon transmission coefficient through the escape interface $\eta \sim 0.3$, and mean sound velocity $c \sim$
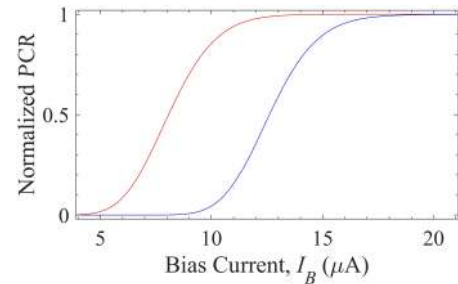

FIG. 3. Normalized PCR for photon wavelengths of 1550 (blue) and 775 nm (red) at $T_b = 1$ K using the simplified HB model.

5×10^5 cm/s, we find $\tau_{\text{esc}} \sim 15$ ps and $LW/4D \sim 30$ ps for $L = 40$ nm and $W = 80$ nm, and the detector latency is either shorter or close to min $\{\tau_{\text{esc}}, LW/4D\}$. Under these conditions we expect the middle of the interval corrected for volume expansion beyond $W \times L \times d$ to be a good representative value and $\chi < 0.5$.

To model the jitter, we need to use the approximation for the detector latency $\tau_{\text{lat}} (I_B, T_b, E)$, which must be found from more advanced detection models as explained in the previous sections. Here we exploit its singular character at $E_{\text{det}} (I_B, T_b)$ or at $I_{\text{det}} (E, T_b)$ and its monotonic decrease as $E \to \infty$, choosing the forms

$$\tau_{\text{lat}} (I_B, T_b, E)$$
$$= \tau_{\text{lat}} (I_B, T_b, E_0) \left[\frac{E_0 - E_{\text{det}} (I_B, T_b)}{E - E_{\text{det}} (I_B, T_b)} \right]^\alpha \quad (20)$$

and

$$\tau_{\text{lat}} (I_B, T_b, E)$$
$$= \tau_{\text{lat}} (I_{\text{sw}}, T_b, E) \left[\frac{I_{\text{sw}} - I_{\text{det}} (E, T_b)}{I_B - I_{\text{det}} (E, T_b)} \right]^\alpha. \quad (21)$$

Here $E_0 = 0.8$ eV is the reference energy of a 1550-nm photon and I_{sw} is the switching current of the SNSPD. Equation (21) reflects the singularity in the latency when

the current approaches the detection current for a deposited energy E. The exponents in Eqs. (20) and (21) are the same. When $E \to E_{\text{det}} (I_B, T_b)$, we have $I_B - I_{\text{det}} (E, T_b) \approx \partial I_{\text{det}} / \partial E \big|_{E_{\text{det}}(I_B, T_b)} [E - E_{\text{det}} (I_B, T_b)]$ neglecting higher-order terms. The derivative $\partial I_{\text{det}} / \partial E$ is continuous [see Fig. 2(b)] and Eq. (21) exhibits the same singularity as Eq. (20). Combining Eqs. (20) and (21), we obtain

$$\tau_{\text{lat}} (I_B, T_b, E) = \tau_{\text{lat}} (I_{\text{sw}}, T_b, E_0)$$
$$\times \left[\frac{I_{\text{sw}} - I_{\text{det}} (E_0, T_b)}{I_B - I_{\text{det}} (E_0, T_b)} \right]^\alpha \left[\frac{E_0 - E_{\text{det}} (I_B, T_b)}{E - E_{\text{det}} (I_B, T_b)} \right]^\alpha. \quad (22)$$

For simplicity, we assume that the exponential α does not depend explicitly on I_B and T_b, and also that $\lim_{E \to \infty} \tau_{\text{lat}} (I_B, T_b, E) = 0$, i.e., that the detector responds instantaneously to an infinite energy deposition. Here, $\tau_{\text{lat}} (I_{\text{sw}}, T_b, E_0)$ is the latency at some reference energy, in this case corresponding to $\lambda = 1550$ nm, and at $I_B = I_{\text{sw}}$. Solving Eq. (8) we obtain

$$E (t) = E_{\text{det}} (I_B, T_b) + [E_0 - E_{\text{det}} (I_B, T_b)]$$
$$\times \left[\frac{I_{\text{sw}} - I_{\text{det}} (E_0, T_b)}{I_B - I_{\text{det}} (E_0, T_b)} \right] \left[\frac{\tau_{\text{lat}} (I_{\text{sw}}, T_b, E_0)}{t} \right]^{1/\alpha}. \quad (23)$$

Substituting this result into Eq. (19) we finally obtain

$$H (t) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left(- \frac{\left\{ E_{\text{det}} (I_B, T_b) + [E_0 - E_{\text{det}} (I_B, T_b)] \left[\frac{I_{\text{sw}} - I_{\text{det}}(E_0, T_b)}{I_B - I_{\text{det}}(E_0, T_b)} \right] \left[\frac{\tau_{\text{lat}}(I_{\text{sw}}, T_b, E_0)}{t} \right]^{1/\alpha} - \chi E_\lambda \right\}^2}{2\sigma^2} \right)$$
$$\times \frac{[E_0 - E_{\text{det}} (I_B, T_b)]}{\alpha t} \left[\frac{I_{\text{sw}} - I_{\text{det}} (E_0, T_b)}{I_B - I_{\text{det}} (E_0, T_b)} \right] \left[\frac{\tau_{\text{lat}} (I_{\text{sw}}, T_b, E_0)}{t} \right]^{1/\alpha}. \quad (24)$$

The expression (24) can now be used for the analysis of latency difference and IRF. For illustration, we enter the parameters for an 80-nm-wide NbN SNSPD with $I_{\text{sw}} = 21$ μA, $\tau_{\text{lat}} (I_{\text{sw}}, T_b, E_0) = 7$ ps, and $\alpha = 0.6$. Figure 4 shows the calculated IRFs for several different currents for 1550- and 775-nm photons at $T_b = 1$ K. Figure 5(a) shows the latency difference for pairs of 1550- and 775-nm photons extracted from the simulations shown in Fig. 4. Figure 5(b) presents the results for the FWHM of the IRF calculated with the use of Eq. (22)–(24).

The simulated curves in Figs. 3–5 are qualitatively consistent with the results of recent experiments [28], providing support to the conclusion that intrinsic jitter was observed in this experiment. It is remarkable that
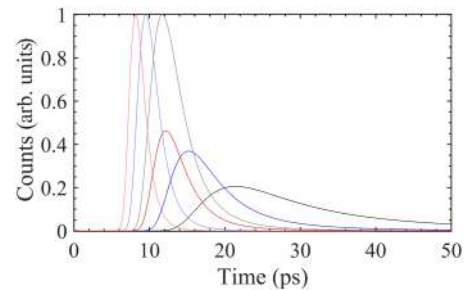

FIG. 4. Instrument response functions for 1550 nm (dark) and 775 nm (light) photons at bias currents of 14 μA (red), 17 μA (blue), and 20 μA (black). The critical depairing current at zero temperature is 25.8 μA.
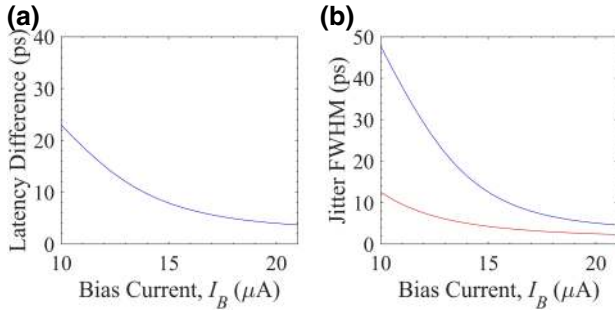
(a)

(b)

FIG. 5.   (a) Relative latency vs bias current for 1550/775 nm photons; (b) Timing jitter vs bias current for 775 nm (red) and 1550 nm (blue) incident photons.

incorporation of the general features of detector latency alone, namely singularities at the detection current and detection energy, together with monotonic positive curvature variation with energy deposition, is sufficient to qualitatively reproduce the intrinsic timing jitter. The model succeeds in reproducing all the primary experimental features: strong dependence of the jitter on bias current, similar functional behavior of the latency difference and timing jitter, and transformation of the IRF distributions with the bias current. In addition, it allows for satisfactory fitting of the observed PCR vs bias-current curves for different wavelengths. Even the magnitudes of the observed latency difference, timing jitter, and counting rates fall in the same range with reasonable accuracy. Nonetheless, the results of the simulations are sensitive to the exact functional dependence of $\tau_{\text{lat}}(I_B, T_b, E)$, the magnitudes of $\tau_{\text{lat}}(I_{\text{sw}}, T_b, E_0)$ and α, and the possibility to model detector latency in the form of Eq. (22). The primary disadvantage of the simple hotbelt theory is its inherent incapability, because of its phenomenological character, to derive the detector latency from the basic principles of nonequilibrium superconductivity.

III. TIME-DEPENDENT GINZBURG-LANDAU MODEL OF DETECTOR LATENCY

The simple phenomenological hotbelt model of latency and timing jitter described above is insufficient to provide a quantitatively predictive picture of the latency characteristics of SNSPDs. To rigorously describe the detector latency, we must use a more advanced technique, which includes the simulation of the evolving nonequilibrium superconducting state in combination with numerical modeling. To analyze the evolution of the nonequilibrium state caused by photon absorption in a superconducting nanowire and understand the characteristics of detector latency in SNSPDs, we use the generalized time-dependent Ginzburg-Landau model.

We limit our generalized TDGL analysis to one dimension for numerical simplicity, but expect the results to

be representative of the main behavior of narrow NbN nanowires. Given the recent observation of relative latency on the order of 5–25 ps [28], the criterion for describing the crossover between 1D and 2D detection must be re-evaluated because these measurements imply a delay time, which is comparable to the characteristic diffusion time to establish a hotbelt $\tau_{D,W} \simeq w^2/16D - w^2/4D \simeq 8$ ps $-$ 32 ps for $W = 80$ nm and $D = 0.5$ cm^2/s. Therefore, the 1D regime is of interest because the transition between 1D and 2D detection behavior is not well understood. The generalized TDGL equations are fully capable of describing the detector response in the 2D HS scenario and in an external magnetic field. The use of the generalized equations is not expected to change the qualitative conclusions reached by Refs. [26,27], respectively, regarding these effects. However, the quantitative difference is substantial, and such a correction is necessary to describe the relative latency results of Ref. [28]. The use of the 1D model is acceptable for making this comparison because the absolute latency of detection does not change substantially in moving from 1D to 2D for the standard TDGL equations once scaling for changes in the detection energy for a given bias current. In this sense, the 1D model is informative about the timescales of detection even if it neglects some of the finer details of detector performance.

A. Generalized TDGL formulation

Different microscopic models can be compared and validated by predicting the detector latency and comparing it to experiments. At present, the most advanced microscopic model of an SNSPD uses electrothermal equations coupled to TDGL equations describing the superconducting order parameter in a 2D system [15]. In this formulation, the energy balance equations in the electron and phonon systems take the form

$$\frac{d}{dt}\left[\frac{\pi^2 k_B^2 N(0) T_e^2}{3} - E_c \mathcal{E}_s\left(T_e, |\Delta|\right)\right]$$

$$= \nabla k_s \nabla T_e - \frac{2\pi^2 k_B^2 N(0)}{15\tau_{ep}(T_c)} \frac{T_e^5 - T_{\text{ph}}^5}{T_c^3} + \vec{j} \cdot \vec{E} \quad (25)$$

$$\frac{dT_{\text{ph}}^4}{dt} = \frac{\gamma}{2\pi^2 \tau_{ep}(T_c)} \frac{T_e^5 - T_{\text{ph}}^5}{T_c} - \frac{T_{\text{ph}}^4 - T_b^4}{\tau_{\text{esc}}}, \quad (26)$$

where k_s is the thermal conductivity of the electron system in the superconducting state

$$k_s = \frac{2D\pi^2 k_B^2 N(0) T_e}{3}\left[1 - \frac{6}{\pi^2}\int_0^{|\Delta|/k_B T_e}\frac{x^2 e^x dx}{(e^x + 1)^2}\right] \quad (27)$$

and Joule heating is included in the electron energy equation as the dot product of the total current density $\vec{j}$

and electric field $\vec{E}$. The energy gain of the electron system due to transitioning to the superconducting state $\mathcal{E}_s$ is given by

$$
\mathcal{E}_s = \int_0^{|\Delta|/k_B T_e} \tilde{\epsilon} n_{\tilde{\epsilon}} d\tilde{\epsilon}
$$

$$
- \int_{|\Delta|/k_B T_e}^{\infty} \tilde{\epsilon} \left[\frac{\tilde{\epsilon}}{\sqrt{\tilde{\epsilon}^2 - \left(\frac{|\Delta|}{k_B T_e}\right)^2}} - 1 \right] n_{\tilde{\epsilon}} d\tilde{\epsilon}
$$

$$
+ \left(\frac{|\Delta|}{2k_B T_c}\right)^2 \left[\frac{1}{2} + \ln\left(\frac{\Delta(0)}{|\Delta|}\right)\right]. \quad (28)
$$

The superconducting order parameter at zero temperature is given by $\Delta(0) = 1.764 \, k_B T_c$, $\tilde{\epsilon} = \epsilon/k_B T_c$, and $n_{\tilde{\epsilon}}$ in Eq. (28) is the Fermi distribution with the electron temperature in units of T_c. The total current density $\vec{j} = \vec{j}_n + \vec{j}_s$ is the sum of the normal current

$$
\vec{j}_n = -\sigma_n \nabla \varphi \quad (29)
$$

and the supercurrent, for which we use an expression derived for the limit of small superfluid momentum, which turns out to be a close approximation to the general Usadel result [15]

$$
\vec{j}_s = \vec{j}_s^{Us} \simeq \frac{\pi \sigma_n}{2e\hbar} |\Delta| \tanh\left(\frac{|\Delta|}{2k_B T_e}\right) \hbar \left(\nabla\phi - \frac{2e}{\hbar c} A\right), \quad (30)
$$

where φ is the electrostatic potential and σ_n is the conductivity in the normal state. The order parameter $\Delta = |\Delta| e^{i\phi}$ is described by the dynamics of the TDGL equations. It is commonly known that the standard TDGL equations are only valid when the temperature is close to T_c and deviations from equilibrium are small. The equations also only apply in the limiting case of a gapless superconductor satisfying $|\Delta| \tau_{s\_mag} \ll \hbar$, where τ_{s_mag} is the spin-flip scattering time. In what follows, we use a more general version of the TDGL equations with less stringent requirements [34,35]. For a dirty superconductor with strong impurity scattering, the assumptions of slow variations in time and space are no different from those of the standard TDGL. Derivation of the generalized TDGL equations does not require the strong limitations necessary for gapless superconductivity. Correspondingly, the generalized TDGL equations better suit simulating the suppression of the gap over extended intervals of time when both the superconducting order parameter and the energy gap remain finite. This is exactly the case for the order-parameter dynamics following the absorption of a photon. The precise details of the order-parameter evolution are required for reliable description of temporal properties of SNSPDs such as the latency and timing jitter.

The generalized TDGL equations can be written as the time dependent partial differential equations

$$
\frac{\pi \hbar}{8 k_B T_c} \left[\varrho(T_e) \frac{\partial}{\partial t} |\Delta| + \frac{i |\Delta|}{\varrho(T_e)} \frac{\partial}{\partial t} \phi + \frac{2ie |\Delta|}{\varrho(T_e) \hbar} \varphi \right]
$$

$$
= \xi_{\text{mod}}(T_e)^2 \left[\nabla + i \left(\nabla\phi - \frac{2e}{\hbar c} A \right) \right]^2 |\Delta|
$$

$$
+ \left[1 - \frac{T_e}{T_c} - \frac{|\Delta|^2}{\Delta_{\text{mod}}^2(T_e)} \right] |\Delta|
$$

$$
+ i \frac{(\nabla \cdot \vec{j}_s^{Us} - \nabla \cdot \vec{j}_s^{GL})}{|\Delta|} \frac{\hbar e D}{\sigma_n \sqrt{2} \sqrt{1 + T_e/T_c}}, \quad (31)
$$

where the parameter $\varrho(T_e) = \sqrt{1 + 4|\Delta|^2 \tau_{sc}(T_e)^2/\hbar^2}$ modifies the rates of phase and magnitude evolution of the TDGL equations. The terms $\xi_{\text{mod}}(T_e)^2 = \pi\hbar D/\left(4\sqrt{2}k_B T_c \sqrt{1 + T_e/T_c}\right)$ and $\Delta_{\text{mod}}^2(T_e) = [\Delta(0) \tanh\left(1.74\sqrt{T_c/T_e - 1}\right)]^2/(1 - T_e/T_c)$ are modified as suggested in Ref. [15] in order to closely match the correct temperature dependencies well below T_c. The generalized TDGL equations break the symmetry between the evolving phase and magnitude of the order parameter because the relaxation of the order-parameter magnitude is controlled by a different process than the relaxation of its phase. Consequently, they cannot be written in the standard TDGL form. Inelastic scattering incorporates both electron-electron and electron-phonon interactions according to $\tau_{sc}(T_e) = 1/\left[1/\tau_{ee}(T_e) + 1/\tau_{ep}(T_e)\right]$, where $\tau_{ee}(T_e)$ and $\tau_{ep}(T_e)$ are the electron-electron and electron-phonon inelastic scattering times respectively. The temperature dependence of these scattering rates is defined by $\tau_{ee}(T_e) = \tau_{ee}(T_c) T_c/T_e$ and $\tau_{ep}(T_e) = \tau_{ep}(T_c)(T_c/T_e)^3$. The final term of Eq. (31) enforces the conservation of the Usadel supercurrent in the stationary state following Ref. [15]. Conservation of total current density, $\nabla \cdot \vec{j} = 0$, is enforced with an additional equation

$$
\sigma_n \nabla^2 \varphi = \nabla \cdot \vec{j}_s. \quad (32)
$$

The boundary conditions at the ends of the nanowire are defined by $T_e = T_c$, $T_{\text{ph}} = T_b$, $|\Delta| = 0$, $\vec{j}_s|_n = 0$, and $\vec{j}_n|_n = I/Wd$. The introduction of current to the simulation domain through $\vec{j}_n$ is numerically easier than through $\vec{j}_s$, but leads to runaway Joule heating of the superconductor. This is solved by limiting Joule heating to the domain of interest in the center of the nanowire and enhancing cooling through electron-phonon coupling near the nanowire edges. Demagnetization effects are negligible in thin and narrow nanowires, so the vector potential A is neglected in the absence of a magnetic field. The system of equations described by Eqs. (25), (26), (31), (32) is solved numerically in one dimension. The system is first allowed to

evolve to a stationary-state configuration for a fixed bias current. Once the system has stabilized, a fixed amount of energy is added to the electron and phonon systems in a hotbelt of length L such that $T_e = T_{\mathrm{ph}} \neq T_b$. This excitation serves as the initial conditions for the system, and the resulting evolution models the response of the nanowire to an excitation of known energy.

For simplicity, we consider a constant bias current in our model. In electrothermal simulations, SNSPDs are typically modeled as a variable resistor in series with an inductor representing the kinetic inductance of the superconductor [36]. This circuit is then coupled to an external readout circuit used to record the electrical pulse generated during a detection event. In an SNSPD, current is diverted from the nanowire to the readout path once a significant potential forms across the hotspot. The electric potential rises quickly after the first phase slip or vortex crossing occurs, but remains small until that point. Therefore, the current will not be diverted until multiple phase slips have occurred and the growth of the normal domain is dominated by the thermal balance of Joule heating and cooling through diffusion or phonon coupling to the substrate. In the 1D hotbelt model at bias currents high enough for photodetection, the first phase-slip event leads to runaway Joule heating. This further justifies the use of a constant bias current for determining the relative latency characteristics within this simplified model because the latency is primarily determined by the time for the first phase slip to occur. This behavior is unlike the case of the 2D model where vortices can cross the nanowire without nucleating a normal domain at low bias currents [15], but we leave the analysis of the generalized TDGL equations within a 2D framework for the future. The choice of constant bias current is further supported by experimental evidence that photon energy does not affect the shape of the electrical signal generated during a photodetection event [28].

B. Detector latency

Calculations are performed for the same 80-nm-wide nanowire parameters listed in Sec. II D at a substrate temperature $T_b = 2$ K. This choice allows for easier numerical computation compared to 1 K and is expected to be representative of experimental results at 1 K based on the temperature dependence measurements of Ref. [28]. The parameter defining the electron-electron inelastic scattering time $\tau_{ee}(T_c)$ is taken as an unknown parameter, while τ_{esc} is 20 ps and $\tau_{ep}(T_c)$ is 24.7 ps based on measurements of $\tau_{ep}(T_c)$ of 16 ps for NbN with a T_c of 10 K scaled by an inverse cubic power law [15,37]. An initial hotbelt length of $L = 40$ nm is used for all simulations. The exact choice of $\tau_{ep}(T_c)$ and τ_{esc} does not change the qualitative nature of the results and has only a small quantitative impact. Once the temperature of the hotbelt is elevated to the initial temperature governed by the photon energy, the
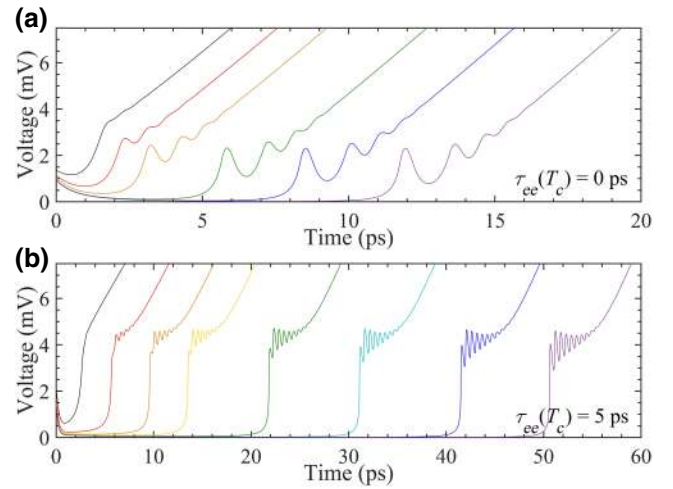

FIG. 6. Voltage curves for a bias current of 12.35 μA for (a) $\tau_{ee}(T_c)$ of 0 ps and (b) $\tau_{ee}(T_c)$ of 5 ps. For $\tau_{ee}(T_c) = 0$ ps, the deposited energies are 0.6316 eV (black), 0.4737 eV (red), 0.3947 eV (orange), 0.3454 eV (green), 0.3392 eV (blue), and 0.3385 eV (violet) while for $\tau_{ee}(T_c) = 5$ ps, the deposited energies are 1.5789 eV (black), 0.9868 eV (red), 0.7895 eV (orange), 0.7105 eV (yellow), 0.6447 eV (green), 0.6217 eV (teal), 0.6143 eV (blue), and 0.6127 eV (violet). The voltage is initially greater than zero due to the temperature dependence of the definition of the Usadel supercurrent (30) and the instantaneous increase in T_e due to the photon energy deposition. The oscillations are due to individual phase slips in the 1D superconductor.

system evolves until either a normal domain forms or the photon-induced excitation relaxes back to the superconducting state. Simulated voltage traces are shown in Fig. 6 for a bias current of 12.35 μA and $\tau_{ee}(T_c)$ of 0 ps and 5 ps under different energy depositions. Lower-energy depositions that are above the detection energy lead to longer waiting times before a voltage pulse appears. For $\tau_{ee}(T_c)$ of 0 ps, the shape of the voltage traces resembles that of the standard TDGL equations in 2D (see Fig. 8 of [15]) while a value of $\tau_{ee}(T_c)$ of 5 ps leads to qualitatively different behavior. This change is a result of the use of the generalized TDGL equations where the magnitude and phase of the order parameter evolve at different rates, as modified by the parameter $\varrho(T_e)$. The slower evolution of the magnitude of the order parameter increases the latency of the detection process compared to the limit of the standard TDGL equations with $\tau_{ee}(T_c)$ of 0 ps. During this slow suppression of the magnitude, the phase rapidly adjusts to maintain the supercurrent flowing through the HB region. After some time, the order parameter is suppressed sufficiently in the HB region that no reconfiguration of phase is able to maintain the supercurrent, at which point the electric potential rapidly appears. The voltage ripples, which appear after this initial jump in potential are individual phase slips, which occur during the final suppression of the order parameter. These phase slips have an oscillation

frequency faster than that of the $\tau_{ee}(T_c) = 0$ ps limit due to the faster phase evolution, which occurs in the generalized TDGL equations with $\varrho(T_e) > 1$.

By selecting a fixed voltage threshold, characteristic latency curves are determined as a function of energy deposited in the superconducting system. In typical experiments, this threshold is chosen to maximize the slew rate of the rising edge of the pulse in order to minimize the effect of electrical noise-induced timing jitter [28]. In our model, the choice is less crucial because a given bias current leads to normal domain growth at a fixed rate once the first phase slips nucleate this normal domain. As a result, the slopes of the voltage curves are approximately the same and a shift in the threshold leads to a latency offset, but no change in the relative latency characteristics for different amounts of deposited energy. The hotbelt TDGL model predicts a monotonically decreasing positive curvature latency as photon energy increases, as predicted by the phenomenological arguments of Sec. II D. The expression (22), which was used to describe the latency within the framework of the simplified HB model, only approximately captures the behavior of the 1D TDGL model. While other analytical forms can provide better fits to the TDGL results, we use interpolation between closely spaced simulation points in order to faithfully capture the behavior of the model near the singularity. This is necessary for evaluating the latency of each energy to derive the instrument response function for the deposited energy distribution.

It is worth noticing that the detection energy curve for the simple HB model of Sec. II D, which was derived from Eq. (5) and shown in Fig. 2(a), is approximately even with the blue curve of Fig. 7(b), but is significantly below the red and black curves. This is understandable because the HB model of Sec. II D uses the instantaneous critical current of the hotbelt section of the nanowire while the generalized TDGL model accounts for the finite suppression time of the order parameter. During this suppression, energy diffuses away from the HB along the length of the nanowire and couples to the phonon system and substrate. Due to these energy losses, more energy is required to break superconductivity within the TDGL model with $\varrho > 1$ compared to the simple HB model. Furthermore, the detection energy increases as the electron-electron scattering parameter $\tau_{ee}(T_c)$ increases. At temperatures around or below T_c, as experienced in the hotbelt model, the magnitude of the parameter $\varrho(T_e)$ is dictated by $\tau_{ee}(T_c)$. This parameter has the dominant influence on the latency compared to other parameters such as τ_{esc} or $\tau_{ep}(T_c)$. Increasing $\tau_{ee}(T_c)$ leads to slower suppression of the order parameter and significantly longer detector latency. This increase in delay time leads to additional dissipation of energy so the detection energy for a given bias current increases, as seen in Fig. 7(b).

The fitting parameters χ, σ_{Fano}, and σ_{nu} are chosen in order to achieve the best possible match to the measured
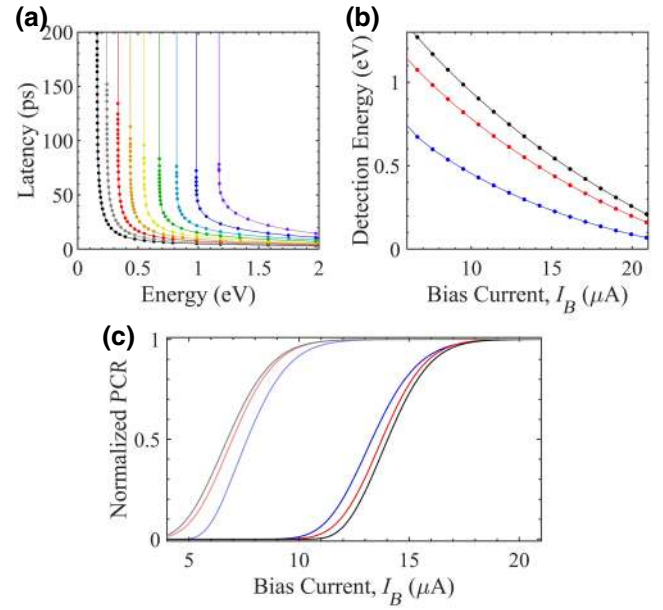

FIG. 7. (a) Latency vs energy simulation results for $\tau_{ee}(T_c)$ of 5 ps for bias currents of 5.7 μA (violet), 7.6 μA (blue), 9.5 μA (teal), 11.4 μA (green), 13.3 μA (yellow), 15.2 μA (orange), 17.1 μA (red), 19.0 μA (grey), and 20.9 μA (black). The associated interpolation fits (lines) are shown for each bias current. (b) Detection energy vs bias current for $\tau_{ee}(T_c)$ of 0 ps (blue), 5 ps (red), and 10 ps (black). (c) Model PCR curves for 775 nm (light) and 1550 nm photons for $\tau_{ee}(T_c)$ of 0 ps (blue), 5 ps (red), and 10 ps (black). The experimental PCR curves of Ref. [28] reasonably match the blue curves.

PCR curves of Ref. [28] as shown in Fig. 7. The values of χ are 0.37, 0.65, and 0.79, σ_{Fano} are 0.064, 0.088, and 0.104 eV, and σ_{nu} are 0.024, 0.040, and 0.048 eV for $\tau_{ee}(T_c)$ of 0, 5, and 10 ps, respectively. When evaluating the distribution of energies for a given photon energy, the Gaussian distribution with a mean energy fraction χ and standard deviation σ_{Fano} has a high-energy cutoff at E_λ. This distribution is convolved with a Gaussian distribution with standard deviation σ_{nu} to give the final energy distribution used to evaluate the PCR response and IRF. For $\tau_{ee}(T_c)$ of 10 ps, the high-energy cutoff is extended beyond E_λ to maintain a smooth PCR response and IRF, enabling a comparison to the lower values of $\tau_{ee}(T_c)$.

C. Instrument response function

The interpolated latency function can be used to generate the detector IRF according to Eq. (11) as shown in Fig. 8(a). The FWHM of the jitter profile is extracted and shown in Fig. 8(b) for 1550 and 775 nm photons for different values of the parameter $\tau_{ee}(T_c)$. In the presence of energy fluctuations, the increase in $\tau_{ee}(T_c)$ and corresponding increase in the detector latency leads to an increase in the timing jitter. As can be seen in Fig. 8(b),
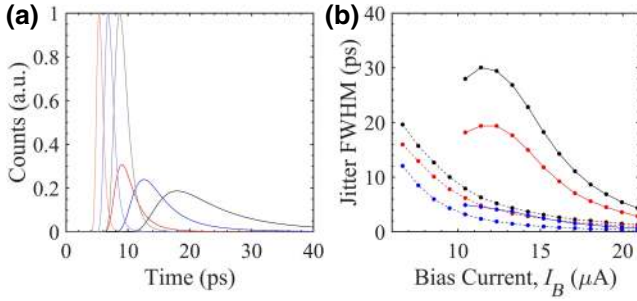
FIG. 8. (a) IRF for bias currents of 15.2 μA (black), 17.6 μA (blue), and 20.0 μA (red) for $\tau_{ee}(T_c)$ of 5 ps for 775-nm (light) and 1550-nm (dark) photon wavelengths. For each bias current, the 775-nm histogram is normalized to a unit maximum while the corresponding 1550-nm histogram is subsequently normalized to have the same area. (b) Jitter FWHM vs bias current for 1550-nm (solid) and 775-nm (dashed) photon energies. The results are shown for $\tau_{ee}(T_c)$ values of 0 ps (blue), 5 ps (red), and 10 ps (black). The case of $\tau_{ee}(T_c)$ reduced to 0 ps corresponds to the standard TDGL formulation, which follows from Eq. (31) in the limit $\tau_{ee} \to 0$.

the generalized TDGL equations show significantly different quantitative behavior when compared to the standard TDGL equations represented by $\tau_{ee}(T_c) = 0$. This correction is necessary to simulate intrinsic jitter on the same scale as observed experimentally [28].

The instrument response function is often fit with an exponentially modified Gaussian (EMG) profile to account for a notable tail observed experimentally [24,28,38]. The histograms generated using the generalized TDGL hotbelt formulation are not strictly defined by an EMG, but share the same characteristics of a mostly Gaussian profile with a tail at longer latency times. As a comparison, we fit our simulation IRFs with an EMG distribution and extract the Gaussian FWHM (2.355σ) and exponential ($1/\lambda$) contribution as done in Ref. [28]. Figure 9 shows the relative contribution of the Gaussian and exponential components
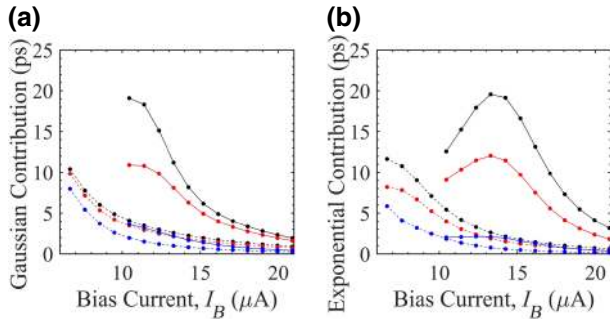

FIG. 9. Gaussian (a) and exponential (b) contributions to the timing jitter as a function of bias current for 1550-nm (solid) and 775-nm (dashed) photon energies when fitting the IRF with an exponential modified Gaussian function. The results are shown for $\tau_{ee}(T_c)$ values of 0 ps (blue), 5 ps (red), and 10 ps (black).
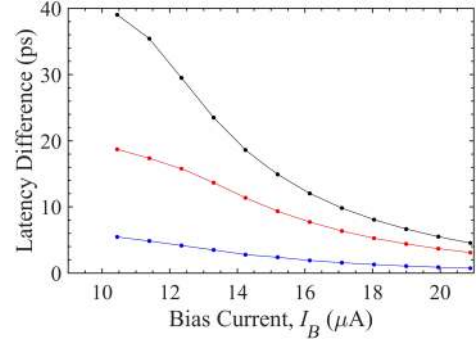

FIG. 10. Latency difference between photons of 1550- and 775-nm wavelength. Results are shown for $\tau_{ee}(T_c)$ values of 0 ps (blue), 5 ps (red), and 10 ps (black). Increasing $\tau_{ee}(T_c)$ leads to larger relative latency difference.

of the fit. The contributions are similar, with the exponential portion contributing slightly more to the timing jitter. This is qualitatively consistent with the experimental findings of Ref. [28].

D. Photon pair latency difference

Recent experiments [28] show that the latency difference between pairs of photons with 1550- and 775-nm wavelengths ranges from 5 to 25 ps depending on the bias current. The latency difference is extracted from simulated IRFs by calculating the time difference between the maxima of the histograms of photon energies of 1550 and 775 nm. Figure 10 shows the extracted latency difference for various bias currents and values of the parameter $\tau_{ee}(T_c)$. The latency difference shows a decreasing trend as the bias current increases, which is consistent with experiments. Only by moving to the generalized TDGL formulation with nonzero $\tau_{ee}(T_c)$ is it possible to find latency differences on the order of 5 to 25 ps, as measured experimentally [28]. The curves in Figs. 8(b), 9, and 10 indicate slowing of the increase of timing jitter and latency with bias current decreasing close to the onset of detection. Nonmonotonic behavior is even observed for the jitter FWHM for values of $\tau_{ee}(T_c) > 0$ ps. Within the discussed model this is connected with the decreasing likelihood of the large energy depositions needed for detector response at small bias currents.

IV. DISCUSSION

In an SNSPD, the four main stages governing the detector response are (a) the initial equilibration in the system of interacting electronic excitations and phonons; (b) a nondissipative stage over which the system evolves with the order parameter being gradually suppressed until the superconducting state becomes incapable of supporting the current through the wire as a supercurrent. In this stage, instability results in the generation of phase-slip

lines or vortices, depending on the symmetry of current flow through the region of suppressed order parameter (i.e., HB or HS regime); (c) a dissipative stage resulting in the nucleation and expansion of the normal domain; and (d) current diversion into the readout and relaxation of the normal domain. Stages (a) to (c) contribute to the intrinsic timing jitter of SNSPDs.

During stage (a) both Fano fluctuations and spatial nonuniformities cast their imprints onto the duration of stage (b). To derive a simple expression for $\delta\tau_{th}$ at the end of stage (a), we write $1/\tau_{th}(T_e) = 1/\tau_{ep}(T_e) + 1/\tau_{ee}(T_e)$ incorporating both electron-phonon and electron-electron scattering. The mean electron (and phonon) temperature after thermalization is found from the energy conservation law (5). Due to Fano fluctuations, the amount of energy deposited into the film fluctuates, which causes fluctuations in T_e. Differentiating (5) we have

$$\frac{\delta E}{\mathcal{E}_0 V_{init}} = \frac{\delta T_e}{T_c} \frac{\pi^2}{3} \frac{T_e}{T_c} \left[\frac{1}{2} + \frac{4\pi^2}{5\gamma} \left(\frac{T_e}{T_c} \right)^2 \right]. \quad (33)$$

After straightforward calculations, we obtain the standard deviation $\sqrt{\overline{(\delta T_e)^2}}$ of the form

$$\frac{\sqrt{\overline{(\delta T_e)^2}}}{T_c} = \frac{\sigma}{\mathcal{E}_0 V_{init}} \left\{ \frac{\pi^2}{3} \frac{T_e}{T_c} \left[\frac{1}{2} + \frac{4\pi^2}{5\gamma} \left(\frac{T_e}{T_c} \right)^2 \right] \right\}^{-1} \quad (34)$$

and correspondingly

$$\sqrt{\overline{(\delta\tau_{th})^2}} = \left| \frac{d\tau_{th}}{dT_e} \right| \sqrt{\overline{(\delta T_e)^2}}$$
$$= 3\tau_{ep}(T_c) \frac{\sigma}{\mathcal{E}_0 V_{init}} \frac{1 + \beta/3x^2}{(1 + \beta/x^2)^2} \left\{ \frac{\pi^2}{3} x^5 \left[\frac{1}{2} + \frac{4\pi^2}{5\gamma} x^2 \right] \right\}^{-1}, \quad (35)$$

where $\beta = \tau_{ep}(T_c)/\tau_{ee}(T_c)$ and $x = T_e/T_c$. The contribution to the timing jitter becomes $\Upsilon_{th} \approx 2.355 \sqrt{\overline{(\delta\tau_{th})^2}}$ and
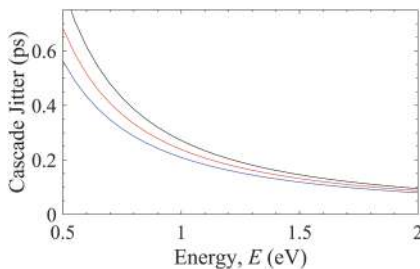

FIG. 11.   Down-conversion cascade timing jitter estimate for NbN. Calculations use the initial hotspot volume $V_{init} = 2\xi_c \times 2\xi_c d$ [15] and $\beta = 0.1$ (black), 1 (red), and 2 (blue). Larger β reduces the cascade jitter $\Upsilon_{th}(E)$. As the photon energy increases, the cascade jitter decreases.

is shown in Fig. 11. As seen from this estimate, the cascade contribution Υ_{th} for typical NbN material parameters is in the subpicosecond regime (≤ 0.3 ps for 1550-nm photons). The only likely situation when the cascade duration may start contributing to the total latency $\tau_{th} + \tau_{lat}$, and the cascade jitter to the total jitter, is when biasing the SNSPD at large currents close to the switching (critical depairing) current. In this case, τ_{lat} is expected to decrease with current, while the thermalization process at temperatures exceeding the critical temperature of a superconductor is practically independent of the current value. The main contribution to Fano variance comes from losing phonons of the first few generations, where most of the deposited energy is in the form of phonons and their numbers are small [16]. In contrast, the total amount of energy loss is accumulated over the whole latency time of the detector. The Fano fluctuations reflect the stochastic nature of phonon loss from the film. The variance remains almost unchanged as the cascade enters the later stages of thermalization when approaching τ_{th} (and beyond). This is due to both the multiplication of phonon numbers in subsequent generations, so that the relative fluctuations for larger numbers become progressively smaller, and the gradual energy flow from phonons to electronic excitations while nonequilibrium state cools down. For exactly the same reason, the cascade jitter is expected to be the same for both the HS and HB detection scenarios.

To minimize the intrinsic timing jitter, the variances σ_{Fano}^2 and σ_{nu}^2 must be reduced as much as possible. Perfecting the technologies of superconducting thin-film growth and nanowire fabrication are the principal ways to minimize the effects of spatial nonuniformities. Their contribution to energy fluctuations relative to Fano fluctuations is expected to increase as the photon energy extends into the ultraviolet region. Reducing the variance of Fano fluctuations generally requires reducing the escape of high-frequency phonons, including Debye phonons and the next generation of down-converted phonons. One way of implementing this is by increasing the thickness of the wire, making it as large as possible with respect to characteristic phonon mean free path $d > l_{p-c}$. Another approach is to control the acoustic properties of the escape interfaces between the nanowire and the substrate and the nanowire and the dielectric passivation layer. The latter may not work for acoustically soft metal films on rigid substrates, but may be of interest for acoustically fast NbN films, where higher-frequency phonons may not escape if they face a gap in the phonon spectrum across the interface. It is important to emphasize that when the escape of high-frequency phonons of the first generations of the down-conversion cascade has been greatly reduced or eliminated, lower-frequency phonons, $\hbar\omega < 2\Delta$, may stay decoupled from the condensate for longer than the duration of the latency and/or equilibration. Fluctuations in their number occur due to the energy partition between electrons

and phonons and are described by the standard Fano variance $\sigma^2 = F \varepsilon E_\lambda$, where $F \sim 0.2$ [39,40]. Thus, in the ideal case the Fano-fluctuation effect can be reduced down to that described by the standard deviation $\sigma_{\text{Fano}} \sim \sqrt{F \epsilon E_\lambda}$. This is nearly a factor of 4 smaller than the σ_{Fano} for 1550 nm used for simulations in Figs. 3–5 and 8–10. Such a dramatic decrease in the strength of fluctuations will move the intrinsic timing jitter (neglecting spatial inhomogeneities) into the subpicosecond range. Even if such ideal performance cannot be easily achieved, the potential for improvement is significant, and further work is needed in acoustic matching and the control of interface properties and phonon fluxes including internal phonon bottlenecking in acoustically soft metal films on rigid substrates and with rigid passivation layers [41].

Stage (b) comprises a longer part of the detector latency compared to the down-conversion, during which the cooling of the affected area due to phonon escape and thermal conduction is accompanied by the suppression of the order parameter, resulting in an instability of the superconducting state and transition into the resistive state. The control parameters, which govern the behavior of this stage, describe the energy exchange between quasiparticles and phonons τ_{ep}, the exchange and loss of phonons to the thermal bath τ_{esc} [important in energy balance over the duration of stage (b) and controlling the fluctuations in stage (a)], the thermal conductivity in the electronic system k_s, and the energy relaxation time, τ_{sc} (T_e).

As evidenced by our analysis and pointed out in Sec. III, the strongest potential impact on reducing the latency and jitter is to make the suppression of superconductivity as fast as possible. We modeled the presence of an extra channel of inelastic scattering by adding the electron-electron scattering in parallel with the electron-phonon interaction. However, in a disordered 2D metal film, the energy transfer in the diffusion channel is small, making the energy relaxation in electron-electron collisions slower than the phase relaxation time [42]. The thermalization time was reported as 7 ps in NbN [43], which is consistent with the prediction of the Altshuler-Aronov formula [44]. The relatively weak effect of electron-electron scattering was incorporated in the form of an electron-electron collision integral in the kinetic theory of photon detection [15]. However, it is known that in disordered mesoscopic metal wires [45], dilute magnetic impurities less than 1 ppm can cause both anomalous decoherence and fast energy exchange. Whether this is the case for the material used in recent experiments [28], making it closer to optimum in terms of intrinsic jitter performance, is not immediately clear. Nonetheless, the prospect of improving detector latency and timing jitter by introducing a controlled amount of spin-flip scattering, affecting the energy relaxation time, deserves further attention.

Nonmonotonic behavior of the intrinsic timing jitter with increasing bias current has been predicted in both

the MoSi [46] and NbN [26] systems in the probabilistic detection regime. Within the deterministic 1D generalized TDGL model, nonmonotonic behavior is observed for the lowest detection efficiencies as shown in Fig. 8 for τ_{ee} (T_c) of 5 and 10 ps. This behavior is a result of the Gaussian distribution of energies deposited in the superconductor and the nature of the singularity, which occurs at the detection energy for a given bias current. As the mean photon energy drops well below the detection energy for a given bias current, higher-energy depositions are excluded from the energy distribution, which leads to a smaller distribution of allowable energies. Depending on the exact shape of the latency curves at different bias currents, this can result in a decrease in jitter FWHM as the bias current decreases. While nonmonotonic behavior is not reported in Ref. [28], measurements were not shown for bias currents far into the probabilistic regime, and the noise contribution to the jitter was not subtracted from the measured jitter, so it is premature to conclude if nonmonotonic behavior is a universal feature of the intrinsic jitter of nanowire devices.

In the 1D generalized TDGL model described in Sec. III, the effect of the transverse coordinate dependence on detector latency is neglected. This simplification is partially supported by recent measurements [28]. When the IRF was measured for nanowire widths ranging from 60 to 120 nm, the timing jitter ranged from approximately 25 ps to 5 ps independent of the width, once the bias current was scaled according to the detection energy. This suggests that the timing differences due to vortex crossing and the varying geometry of phase-slip lines as a function of transverse absorption coordinate may be much smaller than those due to energy fluctuations as described above. If timing jitter due to transverse coordinate dependence were to dominate, the intrinsic jitter would be expected to show a more significant dependence on nanowire width, because the vortex transit times would change. Furthermore, the estimates of the timing jitter due to vortex crossing based on a model of vortex entry from the nanowire edge suggest that this effect is small compared to the measured intrinsic jitter [23]. Note that this work neglected the effect of vortex-antivortex unbinding due to detections near the center of the nanowire, which will alter the expected IRF. In the recent work [27], the transverse coordinate dependence of timing jitter was studied using the standard TDGL equations, which includes the vortex-antivortex unbinding mechanism, and the FWHM was found to be < 5 ps for the energies and bias currents reported, which is significantly smaller than those measured experimentally [28]. For all data available at this time, as evidenced by the period of voltage oscillations (see our Fig. 6, Fig. 8 of [15]), and Fig. 6 of [27], vortex crossing times appear to be too short in comparison with the observed latencies of the detector to be dominant. Therefore, the contribution of these effects to timing jitter must be of lesser importance, at least at bias currents well below the critical depairing current.

The measurement of detection latency differences on the order of 25 to 5 ps between 1550 and 775 nm photons further supports the usefulness of a 1D model, because the absolute latency must be of the same magnitude. The characteristic diffusion time to establish a hotbelt based on the parameters of NbN used in Sec. III is $\tau_{D,W} \simeq 8 - 32$ ps for $W = 80$ nm, which is comparable to the measured latency. The generalized TDGL model captures this behavior through the introduction of the parameter ϱ, which modifies the characteristic time of the variation of the order parameter under the standard TDGL model, $\tau_{|\Delta|}$. Under this modification, the condition for quasi-1D-detection, $\varrho \tau_{|\Delta|} \simeq \tau_{D,W}$, can be satisfied even when calculated according to the standard TDGL formation suggests that a 1D model is inappropriate because $\tau_{|\Delta|} \ll \tau_{D,W}$. As the bias current approaches I_{sw}, the latency falls below the characteristic diffusion time and the 1D approximation becomes less appropriate. In this regime, the effects of transverse-coordinate dependence may become important. More advanced modeling using the 2D generalized TDGL formulation will provide information about the impact of the transverse coordinate of absorption relative to energy fluctuations. The inclusion of the full 2D dynamics is also expected to change the behavior of the longer latency tail observed experimentally and through modeling. Detailing the nature of this non-Gaussian behavior may be another way to experimentally probe the details of the detection mechanism.

Within our 1D model, stage (c) is less significant. Due to the uniform nature of the current flow within the cross section, the normal domain expands at the same rate for excitations of different energy at the same bias current. In the 2D case, the initial stage of the growth of the normal domain may contribute to timing jitter because the expansion depends on the initial coordinate of detection. In this case, the resulting timing jitter would be correlated with the transverse-coordinate dependence caused by vortex nucleation and motion. While not important within the model of intrinsic jitter, the rate of normal domain growth plays a role in an experimental system where electrical noise contributes to timing jitter. The slew rate of the rising edge of the electrical signal from a detection event determines the timing jitter associated with the electrical noise in the system. Maximizing the rate at which the normal domain grows and diverts current [stages (c) and (d)] can reduce the electrical-noise contribution to the total timing jitter.

Comparing the results of Sec. III to recent experiments [28], we find a semiquantitative match for a value of $\tau_{ee}(T_e) \sim 5 - 7$ ps, which is consistent with the value of thermalization time reported in experiment [43], but smaller than that predicted by theory [44]. The detection latency difference between pairs of 1550 and 775 nm photons shows the same monotonically decreasing behavior with increasing bias current and similarly shows an inflection within the range of bias currents that do not saturate the internal detection efficiency. The qualitative behavior of the IRF FWHM also matches the experiment for the same values of the inelastic scattering time. In both experiment [28] and the model, the IRF FWHM for 1550-nm photons decreases with an inflection point occurring at a bias current just below the current where internal efficiency saturates. For 775-nm photons, no inflection point is observed. The simple model predicts a somewhat higher exponential contribution to the total jitter compared to the Gaussian contribution than found experimentally, but the qualitative behavior of the two components each match experiment. Despite the adopted 1D approximation, the generalized TDGL model simultaneously captures the PCR, jitter FWHM, and relative latency behavior observed for the 80-nm sample measured in Ref. [28]. The use of the 1D model has limitations. It is well known that the detection process for most SNSPDs is dependent on the transverse coordinate of the absorption site [5]. Such a mechanism is needed to qualitatively understand the detector response in a magnetic field [5,26]. However, at the present time, the model used to understand this behavior has only been applied qualitatively. When extended to two dimensions, the generalized TDGL approach will not only describe the cross-over from 1D to 2D geometry, but more importantly, is expected to be fully capable of reproducing the qualitative conclusions mentioned above, but within a framework which more accurately captures the appropriate timescale of detection. Given the intimate connection between detector latency and timing jitter, this is a necessary step for generating a detection model which is capable of reproducing all experimentally observed features of SNSPDs.

V. CONCLUSION

We demonstrate how the introduction of detector latency in the presence of Fano fluctuations and spatial inhomogeneities accurately reproduces the qualitative features of the intrinsic timing jitter recently measured in SNSPDs. In the characteristic latency vs energy curve, the presence of a singularity at the detection energy combined with monotonic scaling and positive curvature leads to a non-Gaussian IRF with an extended tail at longer delay times, which has been observed in multiple experiments. Within the framework of the generalized TDGL model, the inelastic scattering time plays a dominant role in determining the detector latency and timing jitter. The addition of this contribution to the standard TDGL model is necessary to reproduce the detector latency observed in experiment. By engineering materials with smaller $\tau_{sc}(T_e)$, it may be possible to reduce this component of the intrinsic jitter in the future. The structural features of detector IRFs such as FWHM, asymmetry characteristics, and latency differences between pairs of photons with different energies offer a new means of studying the detection mechanism in SNSPDs in more detail going forward.

ACKNOWLEDGMENTS

[1] G. N. Gol'tsman, O. Okunev, G. Chulkova, A. Lipatov, A. Semenov, K. Smirnov, B. Voronov, A. Dzardanov, C. Williams, and R. Sobolewski, Picosecond superconducting single-photon optical detector, Appl. Phys. Lett. **79,** 705 (2001).

[2] A. Engel, K. Inderbitzin, A. Schilling, R. Lusche, A. Semenov, H.-W. Hübers, D. Henrich, M. Hofherr, K. Il'in, and M. Siegel, Temperature-dependence of detection efficiency in NbN and TaN SNSPD, IEEE Trans. Appl. Supercond. **23,** 2300505 (2013).

[3] J. J. Renema, R. Gaudio, Q. Wang, Z. Zhou, A. Gaggero, F. Mattioli, R. Leoni, D. Sahin, M. J. A. de Dood, A. Fiore, and M. P. van Exter, Experimental Test of Theories of the Detection Mechanism in a Nanowire Superconducting Single Photon Detector, Phys. Rev. Lett. **112,** 117604 (2014).

[4] R. Lusche, A. Semenov, K. Ilin, M. Siegel, Y. Korneeva, A. Trifonov, A. Korneev, G. Goltsman, D. Vodolazov, and H.-W. Hübers, Effect of the wire width on the intrinsic detection efficiency of superconducting-nanowire single-photon detectors, J. Appl. Phys. **116,** 043906 (2014).

[5] J. J. Renema, R. J. Rengelink, I. Komen, Q. Wang, R. Gaudio, K. P. M. op't Hoog, Z. Zhou, D. Sahin, A. Fiore, P. Kes, J. Aarts, M. P. van Exter, M. J. A. de Dood, and E. F. C. Driessen, The effect of magnetic field on the intrinsic detection efficiency of superconducting single-photon detectors, Appl. Phys. Lett. **106,** 092602 (2015).

[6] F. Marsili, M. J. Stevens, A. Kozorezov, V. B. Verma, C. Lambert, J. A. Stern, R. D. Horansky, S. Dyer, S. Duff, D. P. Pappas, A. E. Lita, M. D. Shaw, R. P. Mirin, and S. W. Nam, Hotspot relaxation dynamics in a current-carrying superconductor, Phys. Rev. B **93,** 094518 (2016).

[7] R. Gaudio, J. J. Renema, Z. Zhou, V. B. Verma, A. E. Lita, J. Shainline, M. J. Stevens, R. P. Mirin, S. W. Nam, M. P. van Exter, M. J. A. de Dood, and A. Fiore, Experimental investigation of the detection mechanism in WSi nanowire superconducting single photon detectors, Appl. Phys. Lett. **109,** 031101 (2016).

[8] M. Caloz, B. Korzh, N. Timoney, M. Weiss, S. Gariglio, R. J. Warburton, C. Schönenberger, J. Renema, H. Zbinden, and F. Bussières, Optically probing the detection mechanism in a molybdenum silicide superconducting nanowire single-photon detector, Appl. Phys. Lett. **110,** 083106 (2017).

[9] L. N. Bulaevskii, M. J. Graf, and V. G. Kogan, Vortex-assisted photon counts and their magnetic field dependence in single-photon superconducting detectors, Phys. Rev. B **85,** 014505 (2012).

[10] A. N. Zotova and D. Yu, Vodolazov, Intrinsic detection efficiency of superconducting nanowire single photon detector in the modified hot spot model, Supercond. Sci. Technol. **27,** 125001 (2014).

[11] A. Engel, J. Lonsky, X. Zhang, and A. Schilling, Detection mechanism in SNSPD: Numerical results of a conceptually simple, yet powerful detection model, IEEE Trans. Appl. Supercond. **25,** 1 (2015).

[12] A. Engel, J. J. Renema, K. Il'in, and A. Semenov, Detection mechanism of superconducting nanowire single-photon detectors, Supercond. Sci. Technol. **28,** 114003 (2015).

[13] D. Yu. Vodolazov, Yu. P. Korneeva, A. V. Semenov, A. A. Korneev, and G. N. Goltsman, Vortex-assisted mechanism of photon counting in a superconducting nanowire single-photon detector revealed by external magnetic field, Phys. Rev. B **92,** 104503 (2015).

[14] A. G. Kozorezov, C. Lambert, F. Marsili, M. J. Stevens, V. B. Verma, J. A. Stern, R. Horansky, S. Dyer, S. Duff, D. P. Pappas, A. Lita, M. D. Shaw, R. P. Mirin, and S. W. Nam, Quasiparticle recombination in hotspots in superconducting current-carrying nanowires, Phys. Rev. B **92,** 064504 (2015).

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

[16] A. G. Kozorezov, C. Lambert, F. Marsili, M. J. Stevens, V. B. Verma, J. P. Allmaras, M. D. Shaw, R. P. Mirin, and S. W. Nam, Fano fluctuations in superconducting-nanowire single-photon detectors, Phys. Rev. B **96,** 054507 (2017).

[17] Q. Zhao, L. Zhang, T. Jia, L. Kang, W. Xu, J. Chen, and P. Wu, Intrinsic timing jitter of superconducting nanowire single-photon detectors, Appl. Phys. B **104,** 673 (2011).

[18] J. Wu, L. You, S. Chen, H. Li, Y. He, C. Lv, Z. Wang, and X. Xie, Improving the timing jitter of a superconducting nanowire single-photon detection system, Appl. Opt. **56,** 2195 (2017).

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

[20] 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, Single-photon imager based on a superconducting nanowire delay line, Nat. Photonics **11,** 247 (2017).

[21] J. A. O'Connor, M. G. Tanner, C. M. Natarajan, G. S. Buller, R. J. Warburton, S. Miki, Z. Wang, S. W. Nam, and R. H. Hadfield, Spatial dependence of output pulse delay in a niobium nitride nanowire superconducting single-photon detector, Appl. Phys. Lett. **98,** 201116 (2011).

[22] Y. Cheng, C. Gu, and X. Hu, Inhomogeneity-induced timing jitter of superconducting nanowire single-photon detectors, Appl. Phys. Lett. **111,** 062604 (2017).

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

[24] M. Sidorova, A. Semenov, H.-W. Hübers, I. Charaev, A. Kuzmin, S. Doerner, and M. Siegel, Physical mechanisms of timing jitter in photon detection by current-carrying superconducting nanowires, Phys. Rev. B **96**, 184504 (2017).

[25] M. Sidorova, A. Semenov, A. Kuzmin, I. Charaev, S. Doerner, and M. Siegel, Intrinsic jitter in photon detection by straight superconducting nanowires, IEEE Trans. Appl. Supercond. **28**, 1 (2018).

[26] M. Sidorova, A. Semenov, H.-W. Hubers, A. Kuzmin, S. Doerner, K. Ilin, M. Siegel, I. Charaev, and D. Yu. Vodolazov, Timing jitter in photon detection by straight superconducting nanowires: Effect of magnetic field and photon flux, Phys. Rev. B **98**, 134504 (2018).

[27] D. Yu. Vodolazov, Minimal Timing Jitter in Superconducting Nanowire Single-Photon Detectors, Phys. Rev. Appl. **11**, 014016 (2019).

[28] B. A. Korzh *et al.*, Demonstrating Sub-3ps Temporal Resolution in a Superconducting Nanowire Single-photon Detector, arXiv:1804.06839.

[29] A. G. Kozorezov, A. F. Volkov, J. K. Wigmore, A. Peacock, A. Poelaert, and R. den Hartog, Quasiparticle-phonon downconversion in nonequilibrium superconductors, Phys. Rev. B **61**, 11807 (2000).

[30] A. G. Kozorezov, J. K. Wigmore, D. Martin, P. Verhoeve, and A. Peacock, Electron energy down-conversion in thin superconducting films, Phys. Rev. B **75**, 094513 (2007).

[31] M. Yu. Kupriyanov and V. F. Lukichev, Temperature dependence of pair-breaking current in superconductors, Sov. J. Low Temp. Phys. **6**, 210 (1980).

[32] Y. Zou, X. Wang, T. Chen, X. Li, X. Qi, D. Welch, P. Zhu, B. Liu, T. Cui, and B. Li, Hexagonal-structured ε-NbN: Ultra-incompressibility, high shear rigidity, and a possible hard superconducting material, Sci. Rep. **5**, 10811 (2015).

[33] S. B. Kaplan, Acoustic matching of superconducting films to substrates, J. Low Temp. Phys. **37**, 343 (1979).

[34] R. J. Watts-Tobin, Y. Krähenbühl, and L. Kramer, Nonequilibrium theory of dirty, current-carrying superconductors: Phase-slip oscillators in narrow filaments near T_c, J. Low Temp. Phys. **42**, 459 (1981).

[35] N. B. Kopnin, *Theory of Nonequilibrium Superconductivity* (Oxford University Press, Oxford, 2001).

[36] J. K. W. Yang, A. J. Kerman, E. A. Dauler, V. Anant, K. M. Rosfjord, and K. K. Berggren, Modeling the electrical and thermal response of superconducting nanowire single-photon detectors, IEEE Trans. Appl. Supercond. **17**, 581 (2007).

[37] A. D. Semenov, R. S. Nebosis, Yu. P. Gousev, M. A. Heusinger, and K. F. Renk, Analysis of the nonequilibrium photoresponse of superconducting films to pulsed radiation by use of a two-temperature model, Phys. Rev. B **52**, 581 (1995).

[38] F. Najafi, Thesis, School Massachusetts Institute of Technology, 2015.

[39] M. Kurakado, Possibility of high resolution detectors using superconducting tunnel junctions, Nucl. Instrum. Methods Phys. Res **196**, 275 (1982).

[40] N. Rando, A. Peacock, A. van Dordrecht, C. Foden, R. Engelhardt, B. G. Taylor, P. Gare, J. Lumley, and C. Pereira, The properties of niobium superconducting tunneling junctions as X-ray detectors, Nucl. Instrum. Methods Phys. Res **313**, 173 (1992).

[41] M. Sidorova, A. Semenov, A. Korneev, G. Chulkova, Y. Korneeva, M. Mikhailov, A. Devizenko, A. Kozorezov, and G. Goltsman, Electron-phonon relaxation time in ultrathin tungsten silicon film, arXiv:1607.07321.

[42] B. L. Altshuler and A. G. Aronov, in *Electron-electron Interactions in Disordered Systems*, edited by A. L. Efros and M. Pollak (Elsevier Science, Amsterdam, 1985).

[43] K. S. Il'in, I. I. Milostnaya, A. A. Verevkin, G. N. Gol'tsman, E. M. Gershenzon, and R. Sobolewski, Ultimate quantum efficiency of a superconducting hot-electron photodetector, Appl. Phys. Lett. **73**, 3938 (1998).

[44] Y. Korneeva, I. Florya, S. Vdovichev, M. Moshkova, N. Simonov, N. Kaurova, A. Korneev, and G. Goltsman, Comparison of hot spot formation in NbN and MoN thin superconducting films after photon absorption, IEEE Trans. Appl. Supercond. **27**, 1 (2017).

[45] F. Pierre, A. B. Gougam, A. Anthore, H. Pothier, D. Esteve, and N. O. Birge, Dephasing of electrons in mesoscopic metal wires, Phys. Rev. B **68**, 085413 (2003).

[46] M. Caloz, M. Perrenoud, C. Autebert, B. Korzh, M. Weiss, C. Schonenberger, R. J. Warburton, H. Zbinden, and F. Bussieres, High-detection efficiency and low-timing jitter with amorphous superconducting nanowire single-photon detectors, Appl. Phys. Lett. **112**, 061103 (2018).