Minimal Timing Jitter in Superconducting Nanowire Single-Photon Detectors

D. Yu. Vodolazov*

Institute for Physics of Microstructures, Russian Academy of Sciences, 603950, Nizhny Novgorod, GSP-105, Russia and Department of Physics, Moscow State University of Education, 29 Malaya Pirogovskaya St, 119435 Moscow, Russia

(Received 12 July 2018; revised manuscript received 19 October 2018; published 9 January 2019; corrected 14 January 2019)

Using a two-temperature model coupled with the modified time-dependent Ginzburg-Landau equation, we calculate the delay time τ_d in the appearance of a growing normal domain in the current-biased superconducting strip after absorption of the single photon. We demonstrate that τ_d depends on the place in the strip where the photon is absorbed and monotonically decreases with an increasing current. We argue that the variation of τ_d (timing jitter), connected either with position-dependent response or Fano fluctuations, could be as small as the lowest relaxation time of the superconducting order parameter, approximately $\hbar/k_B T_c$ (T_c is the critical temperature of the superconductor), when the current approaches the depairing current.

DOI: 10.1103/PhysRevApplied.11.014016

I. INTRODUCTION

In a superconducting nanowire single-photon detector (SNSPD), absorption of a single photon in the visible or infrared range with energy E_ν leads to the appearance of a voltage pulse at a relatively large transport current in the superconducting strip. Experiments demonstrate that there is a finite delay time τ_d in the appearance of the voltage response [1] and, moreover, there is a variance in τ_d (called a timing jitter) that depends on the material or bias current [2–6]. The origin of the timing jitter may come from the electronics, read-out system, or the finite length of the superconducting strip (geometrical jitter [3]), but it also may have an intrinsic origin connected with the dynamics of the superconducting order parameter Δ in response to the current-carrying superconducting strip on the absorbed photon. Indeed, the photon heats electrons (theoretical estimations show that locally the electron temperature may well exceed the critical temperature of the superconductor [7,8]) but because of the finite relaxation time of the magnitude of Δ ($\tau_{|\Delta|}$), the superconductivity is not destroyed instantly. This effect is well known, for example, from the study of the time response of a superconducting bridge/stripe on the supercritical current pulse (current pulse with an amplitude exceeding critical current) [10–14]. In those works, it was found that a finite delay time is strongly reduced with an increasing current pulse amplitude, and a qualitatively similar result was found in experiments with SNSPD [1,4–6].

In SNSPD, the timing jitter could be connected with a position-dependent response [7,15–18], when the photon absorbed at different sites across the strip produces the voltage signal at different detection (critical) currents $I_{\mathrm{det}}(y)$ (y is a coordinate across the strip). Then, in accordance with the results of Refs. [10–14], one may expect different delay times at fixed current I: $\tau_d(I/I_{\mathrm{det}}(y))$, depending on where in the superconductor the photon is absorbed.

The additional mechanism of timing jitter in SNSPD comes from the so-called Fano fluctuations [8,9] (loss of part of the energy of the photon due to the fluctuating nature of the escape of nonequilibrium Debye phonons to the substrate) or local variations of material parameters of the superconducting strip (mean free path, local T_c, or thickness of the strip). Because local detection current $I_{\mathrm{det}}(y)$ depends on the energy E deposited to the electron/phonon system (it determines how strong electrons and phonons are heated) and on the material parameters, at fixed current, the ratio I/I_{det} varies from one absorption event to the next and it produces the variance in the delay time.

In this paper, based on the two-temperature model coupled with a modified time-dependent Ginzburg-Landau equation and current continuity equation [7], we calculate the position-dependent delay time in SNSPD in both the absence and presence of Fano fluctuations and study how τ_d depends on the current and deposited energy. The effect of Fano fluctuations in our model is taken into account via the introduction of probability $P(E)$ to deposit energy $E < E_\nu$ to the electron/phonon systems of a superconductor [8,9]. The effect of variations of material parameters may be considered in a similar way [9] and we do not study them explicitly. We define the delay time τ_d as the time needed for the formation of the growing normal domain after the absorption of the photon somewhere in the superconducting strip. We find that τ_d is drastically

*vodolazov@ipmras.ru

reduced as the current approaches the depairing current and timing jitter may be as small as $\hbar/k_BT_c$ (approximately 0.8 ps for a superconductor with $T_c = 10$ K). We also show that the considered model with a position-dependent response predicts a stronger deviation of the dependence of photon counts on the delay time [in the literature, it is called the probability density function (PDF) [4] or instrument response function (IRF) [6,9]] from the Gaussian-like distribution than the hot belt model predicts [9]. We argue that it occurs due to the photons absorbed near the edge of the strip that give the largest delay time.

The structure of the paper is the following. In Sec. II, we introduce our theoretical model. In Sec. III, we present our results on the dependence of the position-dependent τ_d on the current in the absence of Fano fluctuations and, in Sec. IV, we include the effect of Fano fluctuations and calculate the energy dependence of τ_d. In Sec. V, we discuss the value of delay time and timing jitter at low currents, when the intrinsic detection efficiency of the detector is much smaller than unity and, in Sec. VI, we relate our results with existing experiments and theoretical works.

II. MODEL

The main assumption of our model implies that, at any moment of time, the distribution function of electrons and phonons could be described by Fermi-Dirac and Bose-Einstein functions with local temperatures of electrons T_e and phonons T_p that are different from the bath temperature T. In Ref. [7], it was shown that this assumption is approximately valid in rather dirty (with diffusion coefficient $D \lesssim 0.5$ cm^2/s) thin superconducting films and energy of photon $E_\nu \gtrsim 1$ eV. In this model, the temporal and space evolution of T_e and T_p are governed by heat conductance and energy balance equations [Eqs. (30) and (31), respectively, in [7]]. These equations are coupled to the time-dependent Ginzburg-Landau (TDGL) equation for the superconducting order parameter Δ [Eq. (36) in Ref. [7]], which is modified to take into account the correct temperature dependence of coherence length, superconducting order parameter, and critical (depairing) current at temperatures far below T_c. Together with these equations, one also has to solve the current continuity equation—Eq. (37) in Ref. [7].

In this model, the absorbed photon is modeled by instant local heating of both electrons and phonons up to $T_e = T_p = T_{\rm init}$ in the area $2\xi_c \times 2\xi_c$, the so-called "initial hot spot" [7], where $T_{\rm init}$ should be determined from the energy conservation law [see Eq. (23) in Ref. [7], with replacement w^2d by $6.25\xi_c^2 d$, where the numerical coefficient 6.25 arises from our choice of grid and step $\delta x = \delta y = 0.5\xi_c$]. Here, $\xi_c = (\hbar D/k_BT_c)^{1/2} \sim \xi_0 = (\hbar D/1.76k_BT_c)^{1/2}$ is convenient in the numerical calculation length scale, the initial hot spot is placed at $x = L/2$, and different transverse coordinates $y \in (0, w)$ (L is the length of the strip and w is its width). In Ref. [7], the eligibility and limitation of this choice of initial condition on the basis of the kinetic equations approach was discussed.

In numerical calculations, we use the parameters of a typical NbN strip: $w = 20\xi_c \simeq 130$ nm ($\xi_c = 6.4$ nm), thickness $d = 4$ nm, and $T_c = 10$ K. Important parameters $\gamma = 10$ and $\tau_0 = 900$ ps, which stay in front of the electron-phonon and phonon-electron collision integrals in kinetic equations [see Eqs. (3), (4), (6), (7), (30), and (31) in Ref. [7]], control the corresponding electron-phonon τ_{ep} and phonon-electron τ_{pe} inelastic relaxation times. We also use $L = 4w = 80\xi_c$; $\tau_{\rm esc} = 0.05\tau_0$ (the escape time of nonequilibrium phonons to the substrate); and the boundary conditions for Δ, T_e, and electrostatic potential φ in the x and y directions as in Ref. [7].

Strictly speaking, the TDGL equation is derived at temperatures close to T_c, and it is quantitatively valid when $|\Delta| < k_BT_e$ [19,20]. Note that, in the hot spot area, the local temperature satisfies this condition (at least at the initial stage of its time evolution) and to the moment of the appearance of the first vortices does not strongly deviate from T_c (see, for example, Fig. 8 in Ref. [7]). Secondly, we also do calculations at temperatures close to T_c (at $T/T_c = 0.9$ and 0.95) and do not find any qualitative difference with the results found at lower T. Namely, when the initial hot spot appears in the central part of the strip, the vortex-antivortex nucleate in that place and move in opposite directions, while when it appears near the edge, the vortex enters the strip when $I > I_{\rm det}(y)$. The only quantitative difference is that at $T \sim T_c$, the delay time becomes much longer (which favors energy leakage of the photon's energy to the substrate) and, because of the lower absolute value of the detection current, the normal domain grows much slower or even does not appear in the strip (depending on the choice of $\tau_{\rm esc}$ and T). As a result, the order parameter relaxes in the hot spot area to the equilibrium value after the passage of several vortices (antivortices) across the strip without the appearance of a large voltage signal.

Here, we do not consider fluctuation-assisted photon counting at $I < I_{\rm det}$, which is connected with the penetration of vortices via the energy barrier formed near the hot spot (so we are working strictly in the so-called "deterministic regime" [4]). Therefore, the delay time is not defined at $I < I_{\rm det}(y)$ and it is finite at $I \geq I_{\rm det}(y)$ (see Sec. III).

III. POSITION-DEPENDENT DELAY TIME

As discussed in the Introduction, the position-dependent response in SNSPD should lead to different delay times in the appearance of the voltage depending on where in the strip the photon is absorbed. To illustrate this effect, in Fig. 1 we show position-dependent $I_{\rm det}$ found from the model described in Sec. II. At fixed bias current $I > I_{\rm det}^{\rm max}$, the ratio $I/I_{\rm det}(y)$ depends on the coordinate y, where the

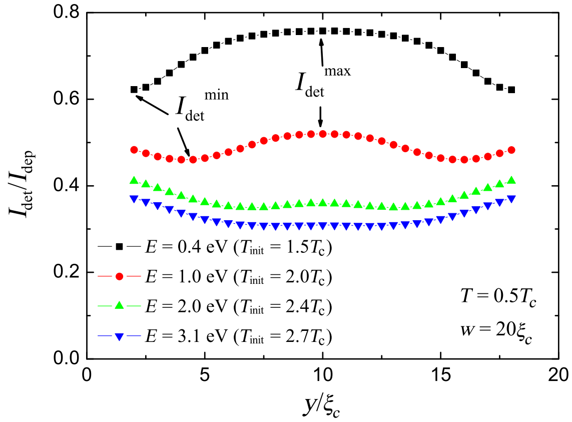

FIG. 1. (a) Position-dependent detection current for different energies E deposited to electrons and phonons after absorption of the photon. Results are obtained with the help of a two-temperature model combined with a modified time-dependent Ginzburg-Landau equation (see Sec. II or Ref. [7] for more details).

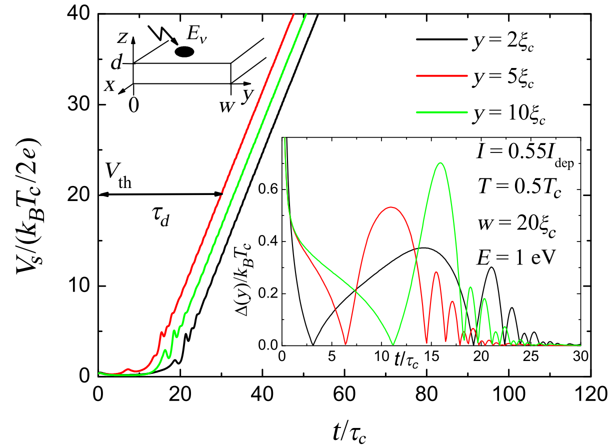

FIG. 2. Time dependence of the voltage drop along the superconducting strip (time is normalized in units of $\tau_c = \hbar/k_B T_c$). The initial hot spot appears at $t = 0$ in different places ($y = 2, 5, 10\xi_c$) across the strip with width $20\xi_c$. The definition of τ_d is seen in the figure. In the right inset, we show the time dependence of Δ in the center of the initial hot spot. Deposited energy to electron/phonon systems $E = 1$ eV corresponds to the initial temperature $T_{\text{init}} = 2T_c$ (see Sec. II). In the left inset, we show the geometry of the strip.

photon is absorbed, across the strip. Therefore, using the results of Refs. [10–14] on the current-dependent $\tau_d(I/I_c)$, one may suppose that τ_d should depend on the coordinate and it should decrease with increasing current.

The theoretical definition of the delay time in some respect is a tricky question. It could be defined as the time needed for vanishing of superconductivity in the whole sample (as it was done, for example, in Refs. [10–14] when the zero-dimensional (0D) problem was studied) or as the time needed for nucleation of the first vortex (this definition was used in Ref. [14] when a two-dimensional (2D) superconducting bridge with a nonuniform current distribution was studied). In our problem, it is useful to define the delay time as the time needed for the nucleation of the normal domain around the place where the photon is absorbed. Indeed, in the modern SNSPD, only relatively large voltages could be detected, which assumes that the large part of the superconducting strip has to go to the normal state.

In Fig. 2, we show the time dependence of the voltage response following from the used model (the photon is absorbed at $t = 0$). One can see that, depending on the site where the photon is absorbed, there is a different delay time in the appearance of large (increasing linearly in time) voltage, connected with the appearance of the growing normal domain. From these dependencies, it is clear that the variance in τ_d does not depend on the choice of threshold voltage V_{th} (if it is large enough) and, in the following, we choose $V_{\text{th}} = 20V_c = 20k_B T_c/2e$ for the evaluation of τ_d.

In Figs. 3(a) and 3(b), we present the calculated position and current-dependent τ_d. Delay time is minimal in the place where $I_{\text{det}}(y)$ is minimal [compare Fig. 3(a) with Fig. 1] and τ_d monotonically decreases with the current increase [at a fixed position of the hot spot—see Fig. 3(b)]. Neither of these results is surprising and they are expected, as we discuss above. The main difference with Refs. [10–14] is that we are looking not for the suppression of superconductivity (uniform or local one, via nucleation of vortex/vortices) but for the nucleation of the growing normal domain. From Fig. 2, it is clear that these definitions are not the same. For example, at $y = 2\xi_c$ (the photon is absorbed near the edge of the strip), the vortex appears earlier than the vortex/antivortex pair nucleates at $y = 10\xi_c$ (the photon is absorbed in the center of the strip) because I_{det} in that place is smaller, but the normal domain appears earlier in the last case due to the shorter time needed to cross the strip by the vortex and antivortex than by a single vortex (only after that does the normal domain appear and expand along the strip, leading to a large voltage response). Moreover, in the considered model, the appearance of the vortices does not obligatorily lead to the appearance of the normal domain when the bias current is close to the retrapping current (see the discussion in Ref. [7]). Therefore, in our problem, τ_d is not only the function of ratio $I/I_{\text{det}}(y)$, but it may also depend on the location of the initial hot spot [even if $I/I_{\text{det}}(y)$ are the same]. For example, at $y = 2\xi_c$ and $y \sim 7\xi_c$, the detection currents are the same (see Fig. 1 in the case of $E = 1$ eV), but τ_d are different [see Fig. 3(a)].

In contrast to the problem with a supercritical current pulse [10–14], τ_d does not diverge as $I \to I_{\text{det}}(y)$ [see Fig. 3(b)]. This fact is connected with the dynamic nature of the hot spot. After absorption of the photon, the hot spot expands and the electronic temperature T_e inside it

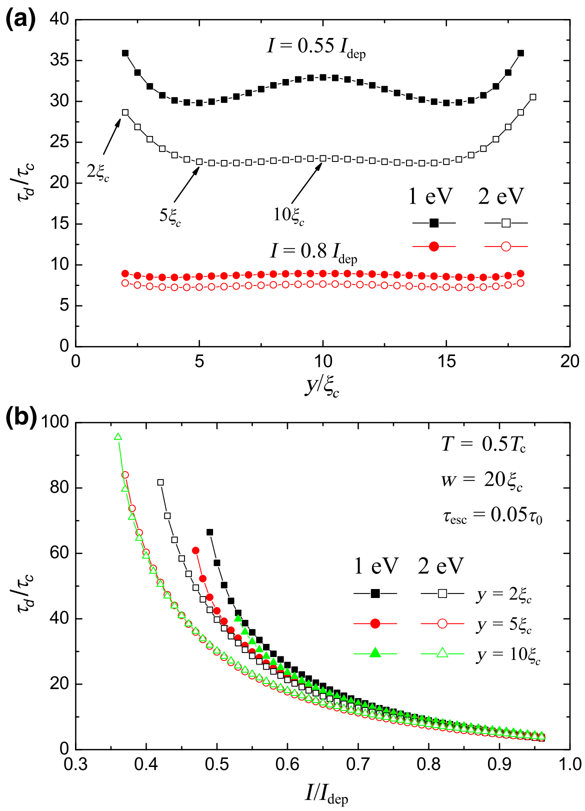

FIG. 3. (a) Position-dependent τ_d at different currents and deposited energies ($E = 1$ eV $\rightarrow T_{\rm init} = 2T_c$ and $E = 2$ eV $\rightarrow T_{\rm init} = 2.4T_c$). (b) Dependence of τ_d on current for three positions of the initial hot spot $y = 2, 5, 10\xi_c$ and two deposited energies 1 and 2 eV.

decreases. As a result, there is an "optimal" moment of time $t_{\rm opt}$ when the hot spot reaches an "optimal" size with "optimal" $T_e \sim T_c$ (for a given deposited energy E), which provides the minimal time-dependent detection current (do not confuse it with $I_{\rm det}^{\rm min}$ in Fig. 1). At $t > t_{\rm opt}$, $I_{\rm det}(t)$ increases despite the larger size of the hot spot due to smaller T_e and a larger superconducting order parameter inside the hot spot. So, in comparison with the problem studied in Refs. [10–14], we have a time-dependent critical (detection) current of the strip with an expanding hot spot. It is equal to $I_{\rm dep}$ (for the strip with no defects) at $t = 0$ and $t = \infty$ and it reaches the minimal value at $t = t_{\rm opt}$. Therefore, for any applied current $I_{\rm det}(t_{\rm opt}) < I < I_{\rm dep}$, there is a finite time window when the superconducting state can be destroyed. This time window could be associated with the current-dependent lifetime of the hot spot that reaches a maximal value at $I = I_{\rm dep}$ and it is equal to zero at $I = I_{\rm det}(t_{\rm opt})$. However, this picture is valid only when $\tau_{|\Delta|} = 0$. In reality, $\tau_{|\Delta|}$ depends on the ratio $I/I_{\rm det}(t)$ and it is the minimal one when $I_{\rm det}(t) = I_{\rm det}(t_{\rm opt})$, but the time window at $I = I_{\rm det}(t_{\rm opt})$ is equal to zero. The trade-off between a current-dependent time window and current-dependent $\tau_{|\Delta|}$ provides a maximal *finite* delay time in this problem and minimal detection current, whose coordinate dependence is shown in Fig. 1.

From Figs. 3(a) and 3(b), one can see that with an increasing current, the variance in delay time decreases and it approaches $\hbar/k_BT_c$ as the current goes to the depairing current. In the next section, we discuss how Fano fluctuations affect this result.

IV. DELAY TIME IN PRESENCE OF FANO FLUCTUATIONS

In this section, we consider the effect of Fano fluctuations on delay time and timing jitter. We follow Ref. [9] and introduce a normalized probability of energy deposition E both to electron and phonon systems after the absorption of the photon:

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

In this model, it is assumed that part of the energy of the photon $E_\nu - E$ is lost due to fluctuations in the escape of nonequilibrium Debye phonons to the substrate [8,9] and the most probable deposited energy is equal to $E = \bar{E} < E_\nu$. In Fig. 4, we show the effect of relatively large Fano fluctuations on the dependence of the IDE [15] [IDE $\sim$ photon count rate (PCR)] on current. The main visible result is that IDE(I) becomes more smooth and qualitatively resembles the PCR found in the absence of a

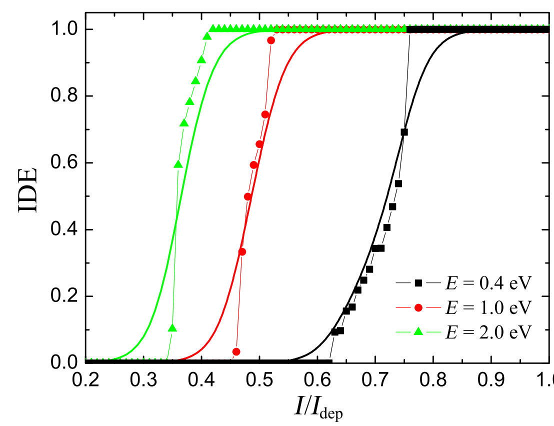

FIG. 4. Dependence of intrinsic detection efficiency (IDE) on the current for three values of deposited energy 0.4, 1, and 2 eV. Symbols are obtained using the results shown in Fig. 1 and assuming an equal probability of photon absorption across the width of the strip with no fluctuations in the deposited energy. Solid curves are obtained in the presence of both position-dependent $I_{\rm det}(y)$ and Fano fluctuations, which provide local fluctuations of $I_{\rm det}(y)$. In our model, instead of an error function [8], we use a simplified expression for local detection efficiency $\mathrm{LDE}(y) = 0.5 \cdot (1 + \tanh\{[I - I_{\rm det}(y)]/dI\})$ with control parameter $dI = 0.05 I_{\rm dep}$ [$\mathrm{IDE}(I) = \int_0^w \mathrm{LDE}(y)dy/w$] to show qualitatively the effect of Fano fluctuations. When Fano fluctuations are absent ($dI = 0$), one obtaines the curves with symbols.

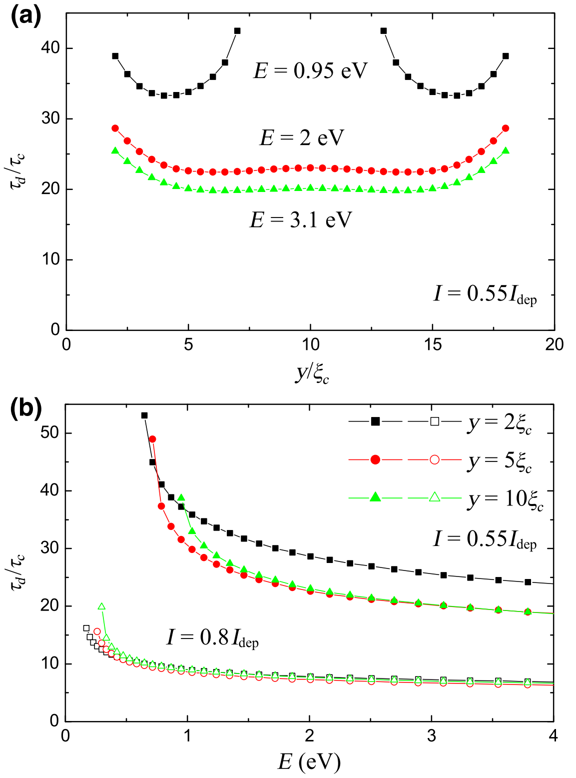

FIG. 5. (a) Position-dependent τ_d for different deposited energies E and $I = 0.55 I_{\text{dep}}$. At $E = 0.95$ eV, the central part of the strip does not "detect" photons (the absorbed photon does not "produce" vortices and the normal domain does not appear). (b) Dependence of τ_d on deposited energy for three photon absorption sites $y = 2, 5, 10\xi_c$ and two values of the current.

position-dependent response (see, for example, Fig. 4 in Ref. [8]).

In Fig. 5(a), we show the position-dependent delay time for different E and, in Fig. 5(b), we demonstrate the energy dependence of τ_d. Based on these results and Eq. (1), we calculate and plot in Fig. 6 the local probability to observe some τ_d in the case of absorption of the photon in the center of the strip $P(\tau_d, y = 10\xi_c)$ [for this purpose, we convert dependence $\tau_d(E)$ to $E(\tau_d)$ and insert it into Eq. (1)]. In calculations, we use $\bar{E} = 1.5, 2.5$ eV and $\sigma = 0.1\bar{E}$. One can see that with an increasing current (at fixed $\bar{E}$) or $\bar{E}$ (at fixed current), the function $P(\tau_d, y)$ tends to a Gaussian-like form. This result follows from the nearly linear dependence $\tau_d(E) \simeq a + bE$ at large currents and E in the finite range of energies 2σ, which, together with Eq. (1), automatically leads to a Gaussian-like dependence. Because nonlinearity is stronger at smaller E (large τ_d) dependence, $P(\tau_d, y)$ is not Gaussian-like at large τ_d, similar to the results found in Refs. [4,5,9,21].

Now, we can combine this result with position-dependent τ_d. We calculate $P(\tau_d, y)$ at each discrete point of our numerical grid, integrate it over y, and assume equal probability for photon absorption across the strip. In this way, we find $P(\tau_d) = \int P(\tau_d, y)dy$, which is proportional

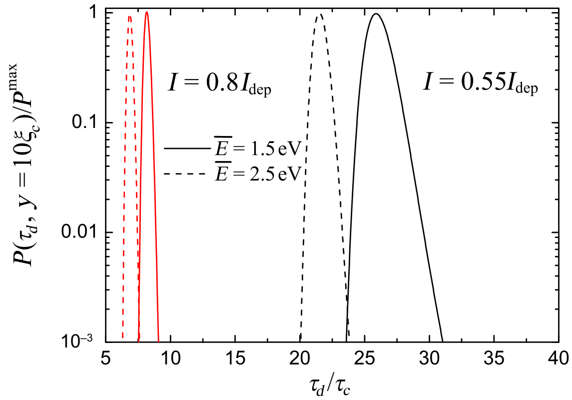

FIG. 6. Local normalized probability to have delay time τ_d at different currents and different $\bar{E} \sim E_v$ (solid curves show $\bar{E} = 1.5$ eV and dashed curves show $\bar{E} = 2.5$ eV) for the photon absorbed in the center of the strip ($y = 10\xi_c$). In calculations, we take $\sigma = 0.1\bar{E}$.

to the experimentally found probability density function [4], instrument response function [6,9], or dependence of photon counts on delay time; see Fig. 7. Local $P(\tau_d, y)$ at any y has a shape similar to the one shown in Fig. 6 but centered at different τ_d. The contribution from the near-edge regions, which give the largest τ_d, provides on dependence $P(\tau_d)$ some kind of "shoulder" at relatively small current $I = 0.55 I_{\text{dep}}$ ("oscillations" on the shoulder visible for $\bar{E} = 2.5$ eV have an artificial origin and are connected with the finite step $\delta y = 0.5\xi_c$ used in numerical calculations), while at a relatively large current the shoulder practically disappears. Visibility of the shoulder depends on the parameter σ in Eq. (1) and, with its increase, the shoulder smears out, leading to a shape of $P(\tau_d)$ that is qualitatively similar to the one shown in Fig. 6, while with its decrease the shoulder becomes more pronounced.

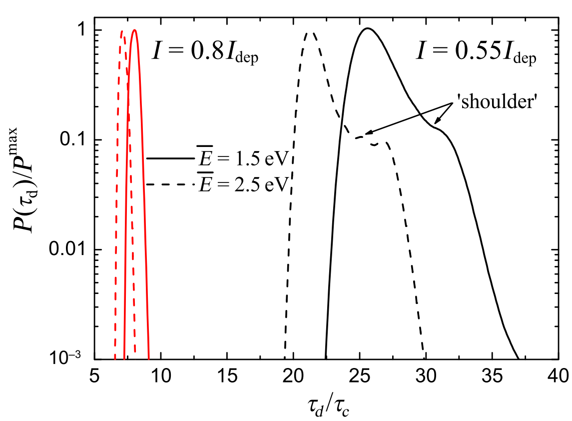

FIG. 7. Normalized probability to have delay time τ_d (parameters are the same as in Fig. 6). At a relatively low current ($I = 0.55 I_{\text{dep}}$), there is a "shoulder" on dependence $P(\tau_d)$ connected with the contribution of photons absorbed in near-edge regions of the strip that provide large τ_d.

From Figs. 5(b) and 7, it follows that timing jitter in the presence of both position dependent-response and relatively large Fano fluctuations ($\sigma = 0.1\bar{E}$) still could be about $\hbar/k_B T_c$ (when deposited energy $\bar{E} > 1$ eV) as the current approaches the depairing current. Physically, it is connected with a relatively short delay time when $I/I_{\text{det}} \gtrsim 1.8$ (see Sec. VI), which is the case for our parameters (see Fig. 1) as $I \to I_{\text{dep}}$.

V. JITTER AT LOW DETECTION EFFICIENCY

So far, we consider delay time and timing jitter at currents exceeding $I_{\text{det}}^{\max}$ (see Fig. 1) when intrinsic detection efficiency reaches unity (or photon count rate, system detection efficiency reaches the plateau or saturates at a relatively large current—see Fig. 4). At $I > I_{\text{det}}^{\max}$, both τ_d and timing jitter decrease with increasing current. What can one expect at low currents $I \gtrsim I_{\text{det}}^{\min}$ when IDE $\ll 1$?

In the model with a position-dependent response and no Fano fluctuation, the detector stops to operate at $I < I_{\text{det}}^{\min}$. At a current slightly exceeding $I_{\text{det}}^{\min}$, only the part of the strip where $I > I_{\text{det}}(y)$ detects photons and it is clear that position-dependent timing jitter in this case should be small. To illustrate it in Fig. 8, we show τ_d at different currents just above $I_{\text{det}}^{\min}$.

In the presence of Fano fluctuations, $I_{\text{det}}^{\min}$ varies from one act of photon's absorption to the next because of the variation of the deposited energy E. However, the maximal deposited energy cannot exceed the energy of the photon E_ν and, hence, there is a minimal $I_{\text{det}}^{\min}$ that corresponds to $E = E_\nu$. The same situation occurs with the variation of material parameters—in the "weakest" place of the strip, I_{det} reaches the minimal value when $E = E_\nu$. Therefore, in the framework of the used model, we expect that at low currents the timing jitter decreases (while delay time increases) with the decrease of the current.

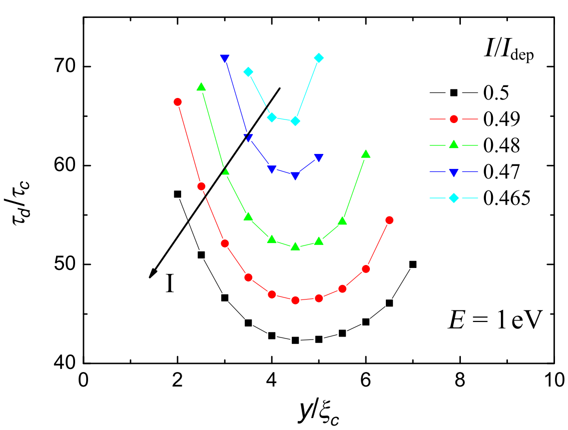

FIG. 8. Position-dependent delay time at currents close to $I_{\text{det}}^{\min} \simeq 0.460 I_{\text{dep}}$. We present results only for the left half of the strip; in the right half, $\tau_d(y)$ is symmetric.

VI. DISCUSSION

We do not compare quantitatively our numerical results with available experiments on the dependence of timing jitter on the current and energy of the photon [4–6] because we believe that the theoretical model used may give at most a semiquantitatively correct result. The assumption of the model used (complete thermalization in an electron system) is fulfilled only partially due to relatively large inelastic electron-electron relaxation τ_{ee} for electrons with energy of about $|\Delta|$ above the Fermi level. As a result, the electron distribution function deviates from the Fermi-Dirac distribution with effective temperature T_e, which should affect τ_d. For example, in Ref. [14], two limiting cases were considered: complete thermalization of electrons (a quasiequilibrium model in the notations of Ref. [14]) and a nonthermalized distribution function (nonthermal model). It was found that different (but qualitatively similar) dependencies $\tau_d(I)$ and τ_d are shorter in the case of thermalized electrons (compare Figs. 3 and 6 in Ref. [14]). Therefore, we make only a semiquantitative comparison of our results with available experiments.

In Ref. [6], the monotonous decay of the timing jitter with current is found for a wide range of E_ν and widths of NbN strips (a similar effect was found for MoSi meanders in Ref. [5]). According to our result, this effect is connected with a decreasing delay time as the current increases—the effect comes from the current-dependent relaxation time of $|\Delta|$ [10–14]. Because τ_d is a function of the ratio I/I_{det} and I_{det} decreases with increasing E_ν, the delay time and timing jitter are smaller at a fixed current for larger E_ν; this effect was observed in Ref. [6]. Estimation of the depairing current for a 100-nm-wide strip from Ref. [6] gives us $I_{\text{dep}} \simeq 45\ \mu$A ($T = 0.9$ K). It means that the experimental critical current for this strip ($I_c \simeq 28\ \mu$A) is about $0.62\, I_{\text{dep}}$ and, therefore, the timing jitter does not reach its minimal (from a theoretical point of view) possible value $\sim \hbar/k_B T_c \sim 1$ ps for that NbN strip with $T_c = 8.65$ K (in Ref. [6], the minimal experimental timing jitter is about 3 ps). Sheet resistance for MoSi meanders is not present in Ref. [5] and we cannot estimate the depairing current for studied structures. Because T_c in MoSi is smaller than in NbN, we expect larger timing jitter.

In Refs. [4,5], nonmonotonous dependence of jitter on current is observed in the range of currents where IDE is smaller than unity. As we discuss in Sec. V, the decrease of jitter at relatively small currents could be connected with a decreasing active area of the detector and/or contribution to photon counts only by photons with the largest deposited energy E_ν. It is not clear if this effect exists in Ref. [6] because timing jitter is not presented for the currents where IDE $\ll 1$.

The presence of the shoulder on the dependence of photon counts on τ_d is a fingerprint of a position-dependent response. This shoulder, qualitatively similar to the one marked in Fig. 7, can be recognized in the supplementary Fig. 8(a) of Ref. [6], while in Refs. [4,5] it appears to be absent. We have to stress that the existence of the shoulder depends on the probability of photon absorption across the strip and, hence, on the wavelength of the photon and its polarization. The shoulder is most visible when photon absorption does not depend on the coordinate, as we assume in our calculations. From another side, relatively strong Fano fluctuations ($\sigma > 0.2E_\nu$ for our parameters) may wash out this feature. Even in this case, however, the position-dependent response could be revealed in the experiment with an external magnetic field, where it leads to an increasing width of dependence $P(\tau_d)$, and, hence, the timing jitter, while no shoulder is seen (see Fig. 3 in Ref. [21]). We believe that in a similar way the shoulder feature also should be rather sensitive to an applied magnetic field.

The main *qualitative* difference of our results with the theoretical results found in Ref. [9] for the timing jitter and delay time is the presence of the shoulder on the dependence of photon counts on delay time. This difference is not surprising because, in Ref. [9], the position-dependent response was not studied. There are also two quantitative differences with the model from Ref. [9]: (i) we do not have a coefficient in front of the time derivative $\partial|\Delta|/\partial t$ in the TDGL equation that is proportional to inelastic τ_{ep} and/or τ_{ee} [see Eq. (31) in Ref. [9]] and (ii) in our model, the maximal delay time is finite, which is connected with the dynamic nature of the hot spot. The coefficient in front of $\partial|\Delta|/\partial t$ appears due to the nonequilibrium effect connected with the variation of $|\Delta|$ in time and leads to a relatively long relaxation time of $|\Delta|$ [10,12,14,19]. In the form used in Ref. [9], it is valid in the condition that the delay time be much larger than $\min\{\tau_{ep}, \tau_{ee}\}$. When this is not the case (as in Ref. [6] at large currents), its usage overestimates the delay time as it was first discussed in Ref. [10] (see Fig. 5 there). In the problem with a response on the supercritical current pulse already at $I/I_c \gtrsim 1.8$, the delay time practically does not depend on τ_{ep} as can be seen from Figs. 3 and 6 in Ref. [14]. In our model, this nonequilibrium effect is included via the term $\partial(E_0\mathcal{E}_s(T_e, |\Delta|)/\partial t$ [see Eq. (30) in Ref. [7]] which is equivalent to the term $\sim \partial|\Delta|^2/\partial t$ in Eq. (6) of Ref. [14] as $T \to T_c$. Moreover, our model automatically takes into account that there is no effect of finite τ_{ep} on τ_d in the case of a strong external driving force (proportional in this problem to $I/I_{\rm det}$).

The delay time and position-dependent timing jitter are also calculated in Ref. [18], where the authors used the model from Ref. [16]. The disadvantage of this model is connected with the assumption that vortices enter the strip via the edge of the strip even when the hot spot is located far from the edge. This assumption comes from the expression used in Ref. [16] for the energy barrier for vortex entry, which is obtained in the framework of the London model with a spatially uniform superconducting order parameter for a straight strip with no hot spot. If one considers the spatial variation of Δ (using, for example, the Ginzburg-Landau, Usadel, or Eilenberger theories), one immediately finds that the vortex nucleates in the place where the superconductivity is maximally suppressed and the supervelocity reaches the maximal value. For the straight strip with no hot spot, the London model gives the correct answer (up to some numerical coefficient) for the energy barrier because Δ is suppressed at the edge and the supervelocity, together with the superconducting current density, is maximal there. In the case when the hot spot is located close to the edge of the strip, the supervelocity is also maximal at the edge (while the superconducting current density is maximal in another place) and the vortex enters via the edge [15]. However, when the hot spot is located far from the edge, the supervelocity is maximal inside the hot spot (Δ is minimal there) and the vortices (vortex/antivortex pair) nucleate inside the hot spot [15]. From some point of view, the vortex itself is a good illustration of this phenomenon. In the center of the vortex $\Delta = 0$, the supervelocity diverges and the superconducting current density is equal to zero. Indirect confirmation of vortex nucleation inside the hot spot also comes from the recent experiment [22] where single photon detection with IDE ~ 1 is observed in several micron-wide NbN strips that cannot be explained by vortex penetration via the edge.

In this work, we do not study fluctuation-assisted vortex penetration in/out of a hot spot and consider only the so-called deterministic regime [4]. Thermoactivated vortex penetration may give a contribution to timing jitter at low currents (when IDE $\ll 1$) because, as a result of fluctuations, vortices may enter the superconducting area even at $I < I_{\rm det}^{\rm min}$ and it is a purely probabalistic process, leading to some distribution of delay times and, hence, timing jitter. However, to calculate this contribution to timing jitter, one should know the value of the energy barrier for vortex penetration δF and how it depends on the current because $\delta F(I)$ enters into the probability for vortex penetration as approximately $\exp(-\delta F(I)/k_BT)$. This problem is not trivial and cannot be solved in the London model when the current is close to $I_{\rm det}$ (for discussion, see Ref. [15]), which makes such calculations rather difficult.

VII. CONCLUSION

In the framework of a two-temperature model combined with a modified time-dependent Ginzburg-Landau equation, we find the following:

(i) delay time and variation of delay time (timing jitter) in SNSPD connected either with position-dependent response or Fano fluctuations monotonically decreases with an increasing current when $I > I_{\rm det}^{\rm max}$ and, at the current close to the depairing current, timing jitter may be about $\hbar/k_BT_c$. The effect is connected with the fast decrease of the relaxation time of the superconducting order parameter at large currents. At fixed current, the delay time and timing jitter are smaller for photons with larger energy due to a larger ratio I/I_{det}.

(ii) Fano fluctuations and the nonlinear dependence of $\tau_d(E)$ provide a non-Gaussian dependence of photon counts on delay time, most pronounced at larger τ_d. The position-dependent response leads to the appearance of the shoulder on this dependence connected with a contribution from the photons absorbed in the near-edge area of the strip. The shoulder decreases with the current and it is maximal in the case of coordinate-independent photon absorption across the strip.

ACKNOWLEDGMENTS

The work was supported by the Russian Science Foundation, Grant No. 17-72-30036 (in part concerning study of delay time at large currents) and by the Russian Foundation for Basic Research, Grant No. 18-29-20100 (in part concerning study of delay time at small currents).

[1] J. Zhang, W. Sysz, A. Pearlman, A. Verevkin, R. Sobolewski, O. Okunev, G. Chulkova, and G. N. Gol'tsman, Time delay of resistive-state formation in superconducting stripes excited by single optical photons, Phys. Rev. B **67**, 132508 (2003).

[2] C. M. Natarajan, M. G. Tanner, and R. H. Hadfield, Superconducting nanowire single-photon detectors: Physics and applications, Supercond. Sci. Technol. **25**, 063001 (2012).

[3] 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).

[4] M. Sidorova, A. Semenov, H.-W. Hubers, 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).

[5] M. Caloz, M. Perrenoud, C. Autebert, B. Korzh, M. Weiss, Ch. 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).

[6] B. A. Korzh, Q-Y. Zhao, S. Frasca, J. P. Allmaras, T. M. Autry, E. A. Bersin, M. Colangelo, G. M. Crouch, A. E. Dane, T. Gerrits, F. Marsili, G. Moody, E. Ramirez, J. D. Rezac, M. J. Stevens, E. E. Wollman, D. Zhu, P. D. Hale, K. L. Silverman, R. P. Mirin, S. W. Nam, M. D. Shaw, and K. K. Berggren, Demonstrating sub-3 ps temporal resolution in a superconducting nanowire single-photon detector, arXiv:1804.06839.

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

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

[9] J. P. Allmaras, B. A. Korzh, M. D. Shaw, and A. G. Kozorezov, Intrinsic timing jitter and latency in superconducting single photon nanowire detectors, arXiv:1805.00130.

[10] M. Tinkham, in *Nonequilibrium Superconductivity, Phonons, and Kapitza Boundaries, Proceedings of NATO Advanced Study Institutes*, edited by K. E. Gray (Plenum, New York, 1981), p. 231.

[11] J. A. Pals and J. Wolter, Measurement of the order parameter relaxation in superconducting Al strips, Phys. Lett. A **70**, 150 (1979).

[12] A. Geier and G. Schön, Response of a superconductor to a supercritical current pulse, J. Low Temp. Phys. **46**, 151 (1982).

[13] F. S. Jelila, J. P. Maneval, F. R. Ladan, F. Chibane, A. Marie-de-Ficquelmont, L. Mechin, J. C. Villegier, M. Aprili, and J. Lesueur, Time of Nucleation of Phase-Slip Centers in $YBa_2Cu_3O_7$ Superconducting Bridges, Phys. Rev. Lett. **81**, 1933 (1998).

[14] D. Yu. Vodolazov and F. M. Peeters, Temporary cooling of quasiparticles and delay in voltage response of superconducting bridges after abruptly switching on the supercritical current, Phys. Rev. B **90**, 094504 (2014).

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

[16] 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**, 2200407 (2015).

[17] J. J. Renema, Q. Wang, R. Gaudio, I. Komen, K. op't Hoog, D. Sahin, A. Schilling, M. P. van Exter, A. Fiore, A. Engel, and J. A. de Dood, Position-dependent local detection efficiency in nanowire superconducting single-photon detector, Nano Lett. **15**, 4541 (2015).

[18] H. Wu, Ch. Gu, Yu. Cheng, and X. Hu, Vortex-crossing-induced timing jitter in superconducting nanowire single-photon detectors, Appl. Phys. Lett. **111**, 062603 (2017).

[19] 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).

[20] A. Schmid, in *Nonequilibrium Superconductivity, Phonons, and Kapitza Boundaries*, edited by K. E. Gray (Plenum Press, New York, 1981), p. 423).

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

[22] Y. Korneeva, D. Yu. Vodolazov, A. V. Semenov, I. Florya, N. Simonov, E. Baeva, A. A. Korneev, G. N. Goltsman, and T. M. Klapwijk, Optical Single-photon Detection in Micrometer-scale NbN Bridges, Phys. Rev. Applied **9**, 064037 (2018).

Correction: The last word in the title contained a typographical error and has been fixed.