Simulation of Silicon Waveguide Single-Photon Avalanche Detectors for Integrated Quantum Photonics

Salih Yanikgonul [ORCID], Victor Leong [ORCID], Jun Rong Ong, Ching Eng Png, and Leonid Krivitsky

Abstract—Integrated quantum photonics is recognized as a key enabling technology on the road towards scalable quantum networking schemes. However, many state-of-the-art integrated quantum photonics demonstrations still require the coupling of light to external photodetectors. On-chip silicon single-photon avalanche diodes (SPADs) provide a viable solution as they can be seamlessly integrated with photonic components, and operated with high efficiencies and low dark counts at temperatures achievable with thermoelectric cooling. In this paper, we report the design and simulation of silicon waveguide-based SPADs on a silicon-on-insulator platform for visible wavelengths, focusing on two device families with different doping configurations: p-n$^+$ and p-i-n$^+$. We calculate the photon detection efficiency (PDE) and timing jitter at an input wavelength of 640 nm by simulating the avalanche process using a 2-D Monte Carlo method, as well as the dark count rate (DCR) at 243 K and 300 K. For our simulated parameters, the optimal p-i-n$^+$ SPADs show the best device performance, with a saturated PDE of 52.4 $\pm$ 0.6% at a reverse bias voltage of 31.5 V, full-width-half-max (FWHM) timing jitter of 10 ps, and a DCR of $<$5 counts per second at 243 K.

Index Terms—Photodetectors, optoelectronic and photonic sensors, photonic integrated circuits, silicon photonics.

I. INTRODUCTION

QUANTUM information technologies have been rapidly developing in recent years, and efforts are shifting from conceptual laboratory demonstrations to scalable real-world devices [1]. Chip-scale photonics devices are important candidates for implementing key features of a future quantum internet, but many recent demonstrations still require the coupling of light to external single-photon detectors [2], [3]. Major improvements in device footprint and scalability could be achieved if these photodetectors reside on the same chip and couple directly to the photonic waveguides [4].

Superconducting nanowire single-photon detectors (SNSPDs) are a state-of-the-art solution, featuring waveguide integrability, near-unity quantum efficiencies, low dark count rate of a few counts per second (cps), and low timing jitter down to $<$20 ps [5], [6]. However, they require cryogenic operating temperatures of a few degrees Kelvin, which is expensive and prohibitive for large-scale deployment.

A practical alternative can be found in single-photon avalanche diodes (SPADs), which are typically reverse biased beyond the breakdown voltage. In this so-called Geiger mode, a single incident photon can trigger a macroscopic avalanche current via a cascade of impact ionization processes. In contrast to SNSPDs, SPADs typically only require thermoelectric cooling and can even operate at room temperature [7], [8]. Moreover, SPADs can benefit from mature complementary metal-oxide semiconductor (CMOS) fabrication technologies [9], making them a promising candidate for scalable manufacturing.

Waveguide-coupled photodetectors have gained increasing attention in recent years; compared to traditional free-space devices, waveguide devices can achieve higher speeds and responsivities, and can also be integrated with photonics circuits and platforms [10], [11]. To date, reports of waveguide-coupled SPADs have been limited to operation at infrared wavelengths [11], [12]. However, many relevant quantum systems, including trapped ions [13] and color centers in diamond [14], operate in the visible spectrum, which makes efficient, low-noise SPADs for visible wavelengths highly desirable. Such devices would also find numerous applications in other important technologies, including LIDAR [15], non-line-of-sight imaging [16], and fluorescence medical imaging [17].

In this paper, we extend our recent work on the design and simulation of silicon waveguide-coupled SPADs for visible light operation, where we used a 2D Monte Carlo simulator to obtain the photon detection efficiency (PDE) and timing jitter, and studied the effect of different waveguide dimensions and doping concentrations [18]. Here we perform an in-depth study of different doping configurations, focusing on two device families: p-n$^+$ and p-i-n$^+$. In addition to the PDE and timing jitter, we also analyze the expected dark count rate (DCR) at room temperature
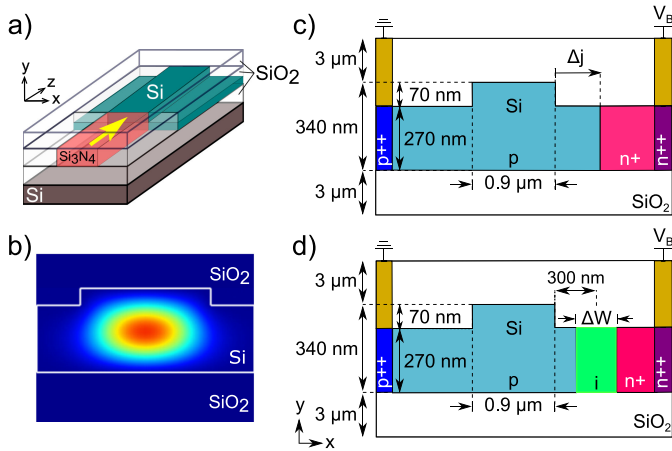
Fig. 1. (a) SPAD structure, consisting of a Si rib waveguide end-fire coupled to an input Si_3N_4 waveguide, (b) optical mode profile at 640 nm for the fundamental (quasi-)TE mode, (c) p-n$^+$ doping configuration with the junction placed at a distance Δj from the right edge of the waveguide core, (d) p-i-n$^+$ doping configuration with an intrinsic region width ΔW. The cross section is constant along the length of the waveguide. Images are not drawn to scale.

and at $-30\,°C$ (243 K), which is a typical operating temperature achievable by Peltier coolers.

Many details regarding the basic SPAD geometry and simulation procedure can be found in ref. [18] and are not repeated here; instead we provide the essential points and highlight the improvements we have made on our previous work.

II. WAVEGUIDE-COUPLED SPAD DESIGNS

A. Device Geometry

The SPAD structure is shown in Fig. 1. It is based on a silicon-on-insulator (SOI) platform, and consists of a 16 μm long silicon (Si) rib waveguide with an absorption of >99% at 640 nm. Input light is end-fire coupled from an input silicon nitride (Si_3N_4) rectangular waveguide, which has high transmittivity at visible wavelengths [9]. We choose this input coupling geometry over a phase-matched interlayer transition, commonly used in integrated photodetectors for infrared wavelengths [12], [19], as the latter is difficult to achieve due to the large difference in refractive indices for Si ($n = 3.8$) and Si_3N_4 ($n = 2.1$). An input coupling efficiency of >90% at the Si/Si_3N_4 interface is obtained using 3D Finite Difference Time Domain (FDTD) simulations (Lumerical).

The structures are cladded with 3 μm of silicon dioxide (SiO_2) above and below. In this study, we fixed the waveguide core width and height at 900 nm and 340 nm respectively, with a shallow etch giving a rib height of 270 nm. A relatively wide waveguide width allows for greater flexibility in the diode junction design, as well as affording a greater tolerance for the fabrication of the junction.

Electrical connections to the device would be made via metal electrodes deposited on top of heavily-doped p^{++} and n^{++} regions at the far ends of the device (along the x axis).

B. Doping Configurations

Our previous simulation study of p-n$^+$ SPADs [18] showed that increasing waveguide core widths (up to 900 nm) could lead
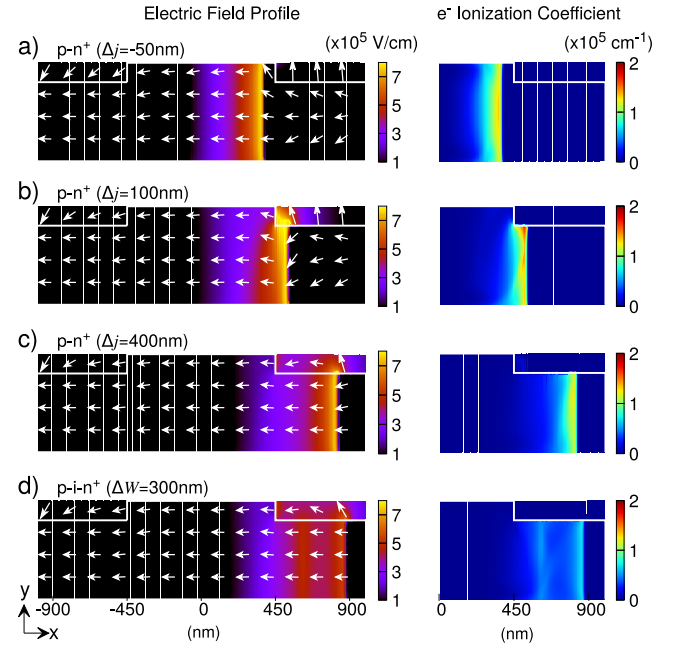

Fig. 2. Electric field profiles (left) and electron ionization coefficients (right) at reverse bias voltages V_B where the photon detection efficiency (PDE) saturates. (a) p-n$^+$ SPAD with $\Delta j = -50$ nm at $V_B = 18.5$ V, (b) p-n$^+$ SPAD with $\Delta j = 100$ nm at $V_B = 23.7$ V, (c) p-n$^+$ SPAD with $\Delta j = 400$ nm at $V_B = 26.5$ V, and (d) p-i-n$^+$ SPAD with $\Delta W = 300$ nm at $V_B = 28.2$ V. For the electron ionization coefficients, only half the waveguide ($x > 0$) is shown as their values are negligible in the other half.

to a higher PDE, as charge carriers can travel a larger distance over which avalanche multiplication can occur. Here, we vary the placement of the p-n$^+$ junction, and investigate the hypothesis that increasing the displacement Δj of the junction beyond the edge of the waveguide core region (Fig. 1(c)) would also enhance this effective distance, and hence the PDE.

Another observation was that impact ionization was most efficient in a narrow region where the highest electric fields are concentrated (similar to Figs. 2(a)–(c)). Widening this high-field region could enhance the PDE, and is achievable by introducing an intrinsic region between the p- and n$^+$-doped areas (Fig. 1(d)). However, doing so would also lower the peak electric field strength (Fig. 2(d)), which could in turn decrease the impact ionization efficiency. Here we explore the effectiveness of such p-i-n$^+$ devices, and attempt to find the optimum width of the intrinsic region ΔW, centered at 300 nm from the edge of the waveguide core.

For both device families, we maintain a constant geometry and doping profile along the length of the waveguide. In this study, we choose a n$^+$ (p) doping concentration of 1×10^{19} (2×10^{17}) dopants/cm^3, and a lightly p-doped intrinsic region with 1×10^{15} dopants/cm^3.

III. SIMULATION METHOD

A. DC Electrical Analysis

For each set of device dimensions and doping configurations, we perform a DC electrical analysis (ATLAS, Silvaco Inc.) by applying a reverse bias voltage V_B across the device electrodes.

For each device, the cathode and anode are placed equidistant from the center of the Si waveguide, with a minimum n^+ region width of 45 nm. We thus obtain the electric field $\mathbf{F}(\mathbf{r})$, ionization coefficients, and other parameters dependent on the 2D position vector $\mathbf{r}$ in the $x - y$ plane; these are required for the Monte Carlo simulation of the avalanche process. Further details can be found in ref. [18]. The breakdown voltage is also identified as the reverse bias voltage V_B at which the device current increases sharply and crosses a threshold of 0.01 mA.

B. 2-D Monte Carlo Simulator

In comparison to deterministic techniques [20], Monte Carlo simulators are well-suited for analyzing SPAD performance, as they can evaluate the timing jitter by modeling the stochastic nature of the impact ionization and avalanche buildup processes. For applications such as quantum key distribution (QKD) [21] and LIDAR [22], low timing jitter is critical to the overall system performance.

In this work, we adapt the 2D Monte Carlo simulator detailed in ref. [18]. Briefly, a random path length (RPL) model is used to simulate the avalanche multiplication process [23]–[25]. Each simulation run starts with a photon absorption which creates an electron-hole pair. At each time step of interval Δt_{rpl}, each charge carrier is accelerated by the electric field and, depending on the ionization coefficients, probabilistically causes an impact ionization after traversing a random path length. This creates further electron-hole pairs, which can then undergo further impact ionizations and eventually lead to a self-sustaining avalanche. Charge carriers are lost when they exit the device boundaries; we note that unlike in ref. [18], the Monte Carlo simulation in this work considers the entire device area (the whole of the p, i, and n^+ regions) and is not restricted to the waveguide core region.

A successful detection event results if the device current reaches a detection threshold I_{det}. Treating the success and failure outcomes as a binomial distribution, the PDE is then the fraction of successful detection events over all simulation runs, with an uncertainty given by the standard deviation (s.d.). The distribution of avalanche times (i.e. time between photon absorption and reaching I_{det}) yields the timing jitter.

1) Diffusion in Quasi-Neutral Regions: The SPAD can be divided into a depletion region and quasi-neutral regions depending on the electric field strength. In the depletion region, the dominant charge carrier transport process is the drift force due to the strong electric fields, and the RPL model applies. However, in the quasi-neutral regions where electric fields are weak, impact ionization can be neglected, and a diffusion model which combines random walks (driven by Brownian motion) and the electric drift force is more suitable. Similar to ref. [18], we use a threshold field to define the quasi-neutral region, i.e. $|\mathbf{F}(\mathbf{r})| < F_{\text{thr}} = 1 \times 10^5$ V/cm, which is on the same order as the breakdown field in silicon [26].

We use the fundamental (quasi-)TE mode profile (Fig. 1(b)) as a probability density map to determine the location where the initial electron-hole pair is injected for each simulation run. If the injection occurs in the quasi-neutral regions, charge carrier transport is simulated using the diffusion model; if the charge carrier crosses over to the depletion region, the simulation continues under the RPL model.

2) Device Current Via Shockley-Ramo's Theorem: Ref. [18] calculates the device current using a 1D approximation of Ramo's theorem, which only considers the motion of charge carriers in one direction. However, this would not be suitable here given our SPAD designs and more complex electric field profiles. As such, we use the generalized Shockley-Ramo's current theorem [27], [28], where each charge carrier i at position $\mathbf{r}_i$ contributes to the device current I induced on the cathode via:

$$I = \sum_i q_i \cdot \mathbf{v}_i(\mathbf{r}_i) \cdot \mathbf{F}_0(\mathbf{r}_i) \tag{1}$$

where q_i is the charge, $\mathbf{v}_i(\mathbf{r}_i)$ is the instantaneous velocity, and $\mathbf{F}_0(\mathbf{r}_i)$ is a weighting electric field calculated in a similar way to $\mathbf{F}(\mathbf{r})$, but under these modified conditions: (i) the cathode is at unit potential, while the anode is grounded; (ii) all charges (including space charges) are removed, i.e. the waveguide is undoped [29].

C. Dark Count Rate

Even in the absence of light, free charge carriers may be generated, which can probabilistically trigger avalanche events and result in dark counts. Due to the high electric fields in SPADs, the most relevant carrier generation mechanisms are thermal excitation enhanced by trap-assisted tunneling (TAT), and band-to-band tunneling (BTBT).

We quantify the dark noise by calculating the DCR $R_D(T)$ via [30]:

$$R_D(T) = L \cdot \iint P_{\text{trig}}(\mathbf{r}) \cdot (G_{\text{TAT}}(\mathbf{r}, T) + G_{\text{BTBT}}(\mathbf{r}, T))\, \mathrm{d}\mathbf{r} \tag{2}$$

where T is the temperature in Kelvin, $L = 16$ μm is the SPAD length, $P_{\text{trig}}(\mathbf{r})$ is the avalanche triggering probability, and $G_{\text{TAT}}(\mathbf{r}, T)$, $G_{\text{BTBT}}(\mathbf{r}, T)$ are the net generation rates of charge carriers (per unit volume) of their respective mechanisms.

1) Trap-Assisted Tunneling: The thermal generation rate of carriers can be obtained from the Shockley-Read-Hall (SRH) model, modified to account for TAT [31], [32]:

$$G_{\text{TAT}}(\mathbf{r}, T) = \frac{n_i(T)}{\tau_g(\mathbf{r}, T)} \tag{3}$$

where $n_i(T)$ is the intrinsic carrier concentration and $\tau_g(\mathbf{r}, T)$ is the electron-hole pair generation lifetime, which can be expressed in terms of the recombination lifetime $\tau_r(\mathbf{r}, T)$ [33]:

$$\tau_g(\mathbf{r}, T) = \frac{\tau_r(\mathbf{r}, T) \cdot e^{|E_t - E_i|/k_B T}}{1 + \Gamma(\mathbf{F}(\mathbf{r}), T)} \tag{4}$$

where the exponential term describes the main temperature dependence in TAT, and the field effect function $\Gamma(\mathbf{F}(\mathbf{r}), T)$ describes the effect of electric fields. E_t and E_i are the energy levels of the recombination centers (assumed to be equal to that of traps at the Si/SiO$_2$ interface [34]) and the intrinsic Fermi level, respectively, and k_B is the Boltzmann constant.

TABLE I
SIMULATION PARAMETERS

| Name	Symbol	Value	Reference
Electric field threshold	F_{thr}	1×10^5 V cm^{-1}	[26]
Detection current threshold	I_{det}	20 µA	-
RPL time step	Δt_{rpl}	10 fs	-
No. of simulations per parameter set	-	> 6000	-
BTBT parameter	B_A	4×10^{14} cm$^{-0.5}$V$^{-2.5}$s^{-1}	[35]
	B_Γ	2.5	[35]
Recombination energy	E_t-E_i	0.25 eV	[34]
Temperature dependent parameters at 300 (243) K:			
– intrinsic carrier concentration	n_i	9.70×10^9 (2.95×10^7) cm^{-3}	[37]
– Recombination lifetime	τ_r	7.0 (7.8) ns	[34], [38]
– BTBT parameter	B_B	1.90 $(1.94) \times 10^7$ V cm^{-1}	[35], [39]
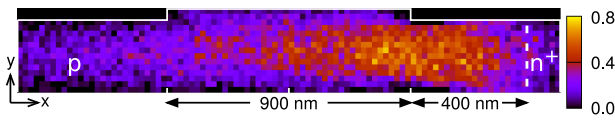
Fig. 3. Avalanche triggering probability $P_{\text{trig}}(\mathbf{r})$ for a p-n$^+$ SPAD with $\Delta j = 400$ nm at $V_B = 16.5$ V. Each 20×20 nm pixel shows the probability of an initial photo-generated electron-hole pair injected within that pixel resulting in a successful detection event. The dashed line indicates the junction position. The results are obtained over >40 k Monte Carlo simulation runs, resulting in an error of $<10\%$ per pixel.

The field effect function $\Gamma(\mathbf{F}(\mathbf{r}), T)$ is:

$$\Gamma(\mathbf{F}(\mathbf{r}), T) = 2\sqrt{3\pi} \cdot \frac{|\mathbf{F}(\mathbf{r})|}{F_\Gamma(T)} \cdot \exp\left(\left(\frac{|\mathbf{F}(\mathbf{r})|}{F_\Gamma(T)}\right)^2\right) \quad (5)$$

in which

$$F_\Gamma(T) = \frac{\sqrt{24 m_t^*(k_B T)^3}}{q\hbar} \quad (6)$$

where q is the electron charge, and $m_t^* = 0.25\, m_0$ is the effective electron tunneling mass, with m_0 being the electron rest mass [35].

2) Band-to-Band Tunneling: The BTBT mechanism has been shown to be important at electric field strengths above 7×10^5 V/cm, where band-bending is sufficiently strong to allow significant tunneling of electrons from the valence band to the conduction band [35]. This rate can be expressed as:

$$G_{\text{BTBT}}(\mathbf{r}, T) = B_A \cdot |\mathbf{F}(\mathbf{r})|^{B_\Gamma} \cdot \exp\left(\frac{-B_B(T)}{|\mathbf{F}(\mathbf{r})|}\right) \quad (7)$$

where B_A, B_B, and B_Γ are model parameters; we use values based on ref [35].

The values of the parameters used in our calculations are listed in Table I, and further details of their derivation can be found in the Appendix.

3) Avalanche Triggering Probability: To obtain the avalanche triggering probability $P_{\text{trig}}(\mathbf{r})$ for each device, we perform >40 k Monte Carlo simulation runs, with photon absorption positions distributed uniformly across the device. A representative map of $P_{\text{trig}}(\mathbf{r})$ is shown in Fig. 3.

IV. SIMULATOR OPTIMIZATION

The Monte Carlo simulations can become computationally expensive due to the need to keep track of and model individual charge carriers, especially when the number of charge carriers grows exponentially during the avalanche process. If we would use the same simulation parameters in our previous work [18] to model one SPAD at a given bias V_B, our simulator (implemented in Python) would require $\sim$24 k CPU-hours on two sets of 12-core CPUs (Intel Xeon E5-2690 v3). Such a high computation cost would limit the variety of SPAD designs we can feasibly study.

Thus, we first use a representative device (p-n$^+$ SPAD with $\Delta j = -50$ nm, at $V_B = 21.5$ V) to perform a series of preliminary studies to optimize the simulation parameters: the detection threshold I_{det}, RPL time step Δt_{rpl}, and number of simulation runs per parameter set. We aim to reduce computation time without sacrificing the simulation accuracy.

A. Detection Current Threshold

A reasonable discriminator threshold in experimental SPAD characterization setups is $I_{\text{det}} = 0.2$ mA [36], a value we used previously [18]. However, it may not be necessary to simulate the multiplication of charge carriers up to that point as the avalanche process might already have passed a self-sustaining threshold at a lower current. On the other hand, a very low I_{det} would overestimate the PDE by falsely identifying small avalanches that would not be self-sustaining, and underestimate the timing jitter by not simulating the full avalanche.

By varying I_{det} while fixing $\Delta t_{\text{rpl}} = 1$ fs with 2 k simulation runs per I_{det} value (Fig. 4(a)), we conclude that we can lower I_{det} to 20 µA without significant deviations in PDE or timing jitter.

B. RPL Time Step

A larger RPL time step Δt_{rpl} would speed up simulations, but reduces time resolution and hence accuracy. A suitable choice would be just short enough such that the charge carrier environment does not change too significantly between each step, even in the high-field regions with large field gradients.

We vary Δt_{rpl} while fixing $I_{\text{det}} = 20$ µA with 2 k simulation runs per Δt_{rpl} value (Fig. 4(b)). We choose $\Delta t_{\text{rpl}} = 10$ fs as an optimal value; for larger time steps, PDE begins to deviate significantly compared to the previous value of $\Delta t_{\text{rpl}} = 1$ fs.

C. Number of Simulation Runs

We analyze the PDE over an increasing number of simulation runs for $\Delta t_{\text{rpl}} = 10$ fs and $I_{\text{det}} = 20$ µA, and observe that the PDE converges to a stable value after several thousand runs. We choose to perform at least 6k runs per parameter set to reduce the relative uncertainty to $\sim$1%.

Compared to the previous simulation parameters (i.e. $\Delta t_{\text{rpl}} = 1$ fs, $I_{\text{det}} = 0.2$ mA, 18k runs), our optimized values ($\Delta t_{\text{rpl}} = 10$ fs, $I_{\text{det}} = 20$ µA, 6k runs) require only $\sim$90 CPU-hours per set, indicating an improved timing performance by a factor of $\sim$270.
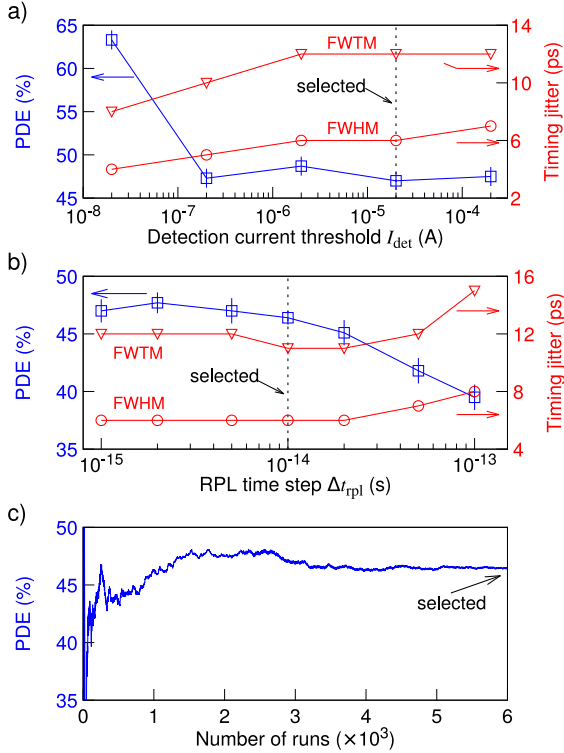
Fig. 4. Optimization study of varying simulator parameters and their effects on the photon detection efficiency (PDE) and timing jitter (full-width-half-max (FWHM) and full-width-tenth-max (FWTM)), for a p-n$^+$ SPAD with Δj = -50 nm at V_B = 21.5 V. (a) Varying I_det with Δt_rpl = 1 fs, 2 k simulation runs per I_det value. (b) Varying Δt_rpl with I_det = 20 μA, 2 k simulation runs per Δt_rpl value. (c) Convergence of PDE for Δt_rpl = 10 fs and I_det = 20 μA after several thousand runs. Error bars for PDE indicate 1 s.d. uncertainty. Selected parameters for subsequent simulations in this paper are marked.

V. SIMULATION RESULTS AND DISCUSSION

A. Photon Detection Efficiency and Timing Jitter

We simulate each device at increasing reverse bias voltages V_B, starting from just above its breakdown voltage. For all devices, PDE increases with V_B and reaches a saturation level (representative plots shown in Fig. 5(a)). We define the saturated bias voltage V_sat as the lowest V_B value where the obtained PDE values within a ± 1 V range agree within their 1 s.d. uncertainty; the PDE at V_sat is then the saturated PDE.

The distribution of avalanche times is generally asymmetric, especially for p-i-n$^+$ SPADs with $\Delta W > 600$ nm (see Fig. 5(b)). Long tails in the timing distribution can adversely affect applications requiring high timing accuracies, e.g. satellite-based quantum communications [40]. Therefore, we present the full-width-half-maximum (FWHM) and full-width-tenth-maximum (FWTM) timing jitter, both extracted from timing histograms with 1 ps bin size, to better describe the timing performance of the SPADs. In general, timing jitter does not vary significantly with V_B, except when V_B is near the breakdown voltage.

1) p-n$^+$ SPADs: For p-n$^+$ SPADs, we observe a general trend of PDE increasing with the junction displacement Δj (Fig. 6(a)). If the junction is placed further away from the waveguide core, charge carriers injected after a photon absorption in the core region travel a longer distance and can undergo more
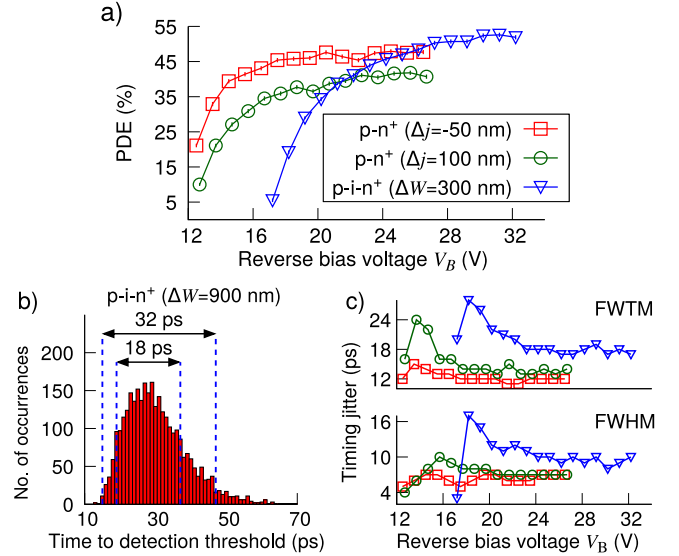

Fig. 5. (a) Simulated PDE at varying reverse bias voltages V_B for representative devices, showing the saturation behavior as V_B increases. Error bars indicating 1 s.d. uncertainty are much smaller than the symbol size. (b) Distribution of simulated avalanche times (i.e. time between photon absorption and reaching the detection threshold I_det) for a p-i-n$^+$ SPAD with $\Delta W = 900$ nm at $V_B = 41$ V. Histogram bin size is 1 ps. The full-width-half-max (FWHM) and full-width-tenth-max (FWTM) timing jitter values are indicated. (c) FWHM and FWTM timing jitter performance for the same devices in (a). (a) and (c) share the same legend.
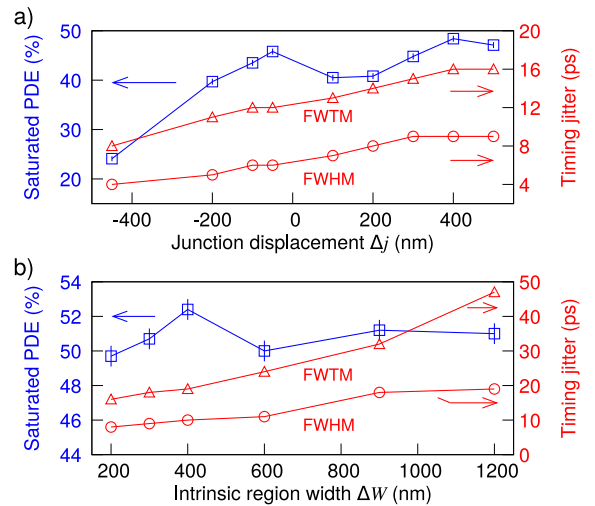

Fig. 6. Saturated PDE and timing jitter for various (a) p-n$^+$ and (b) p-i-n$^+$ SPADs. Both full-width-half-max (FWHM) and full-width-tenth-max (FWTM) timing jitter values are shown. Error bars for PDE indicate 1 s.d. uncertainty.

impact ionizations, thus increasing the likelihood of a successful avalanche. The stochastic avalanche process taking place over a larger distance would also explain the increasing timing jitter at higher Δj. However, Δj being too large would weaken the electric field strength in the waveguide core, which would lead to more charge carriers being lost at the waveguide boundaries due to random walk; this may explain the slight drop in PDE for $\Delta j > 400$ nm.

The observed drop in PDE for $\Delta j = 100$ nm is due to an "edge effect": when the junction is placed in close proximity
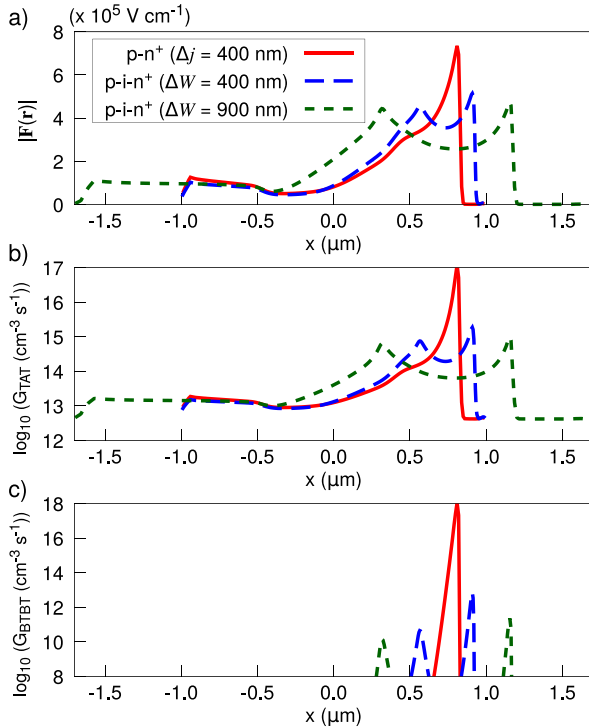
Fig. 7. (a) Electric field strength $|\mathbf{F}(\mathbf{r})|$, (b) trap-assisted tunneling (TAT) generation rates, and (c) band-to-band tunneling (BTBT) generation rates at the waveguide mid-height (170 nm from the bottom) for SPADs with high saturated PDEs: p-n$^+$ SPAD with $\Delta j = 400$ nm, and p-i-n$^+$ SPADs with $\Delta W = 400$ nm and 900 nm, at reverse bias voltages beyond where their PDE has already saturated ($V_B = 31.5$ V, 34.5 V, and 53 V, respectively).
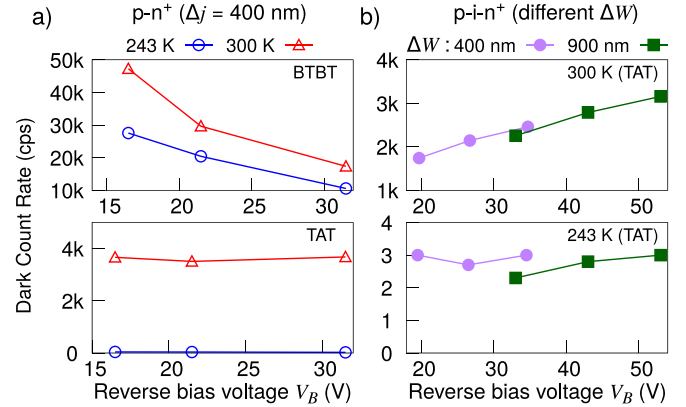

Fig. 8. Dark count rate (DCR) contributions due to TAT and BTBT mechanisms at varying reverse bias V_B and temperatures. (a) p-n$^+$ SPAD with $\Delta j = 400$ nm. (b) p-i-n$^+$ SPADs with $\Delta W = 400$ nm and 900 nm. The contribution from BTBT is negligible, thus only TAT is shown here.

to the waveguide rib edge, we observe a narrowing of the effective impact ionization region where ionization coefficients are high (Fig. 2(b)), which leads to a lower PDE.

The highest saturated PDE obtained for p-n$^+$ SPADs is $48.4 \pm 0.6\%$ at $V_B = 26.5$ V for $\Delta j = 400$ nm, with a FWHM timing jitter of 9 ps.

2) p-i-n$^+$ SPADs: For p-i-n$^+$ SPADs, the widening of the high-field region has led to a higher PDE than for p-n$^+$ devices (Fig. 6(b)). Besides the increased efficiency of impact ionizations, this can also be explained by a lower loss rate of charge carriers under the diffusion model ($<5\%$ for p-i-n$^+$, $\sim 10\%$ for p-n$^+$), which follows a photon absorption event in the quasi-neutral regions. The slight drop in PDE for $\Delta W = 600$ nm is also likely due to an edge effect, as the p-i transition edge corresponds to the position of the waveguide rib edge; we do not otherwise find an obvious dependence of the PDE on the intrinsic region width, although timing jitter increases with ΔW.

Based on our analysis, we conclude that the optimum performance is obtained for $\Delta W = 400$ nm, which gives a saturated PDE of $52.4 \pm 0.6\%$ at $V_B = 31.5$ V and a FWHM timing jitter of 10 ps.

B. Dark Count Rate

We also evaluate the dark noise performance of the SPADs, focusing on devices which display high saturated PDE: p-n$^+$ SPAD with $\Delta j = 400$ nm and p-i-n$^+$ SPADs with $\Delta W = 400$ nm

and 900 nm. We calculate the DCR at 243 K, which is in a typical SPAD operating regime readily achieved with thermoelectric cooling, as well as at 300 K to explore the feasibility of room temperature operation.

For our simulated parameters, BTBT shows a greater sensitivity to peak electric field strength than TAT (Fig. 7). In p-n$^+$ SPADs, where the peak fields are high, BTBT is the dominant dark carrier generation mechanism. As the bias V_B increases, the depletion region widens, leading to a decrease in the peak field strength and hence the overall DCR, while the TAT contribution stays relatively constant (Fig. 8(a)). At an operating bias of $V_B = 31.5$ V (which is above the saturated bias), the DCR is 11 kcps and 21 kcps at 243 K and 300 K, respectively.

In p-i-n$^+$ SPADs, due to wider high-field regions with lower peak fields, BTBT becomes negligible compared to TAT. As such, DCR generally increases with V_B, and shows a steeper dependence on temperature (~ 1000 - fold drop between 300 K and 243 K). We observe that while SPADs with wider intrinsic region widths ΔW had lower dark carrier generation rates per unit volume, this was offset by the larger device volume, and could lead to higher DCR compared to narrower ΔW.

Overall, dark count performance for p-i-n$^+$ SPADs is significantly better compared to p-n$^+$ devices, with observed DCR of <4 kcps at 300 K and <5 cps at 243 K (Fig. 8(b)), even at V_B beyond the saturated bias.

VI. Conclusion

In conclusion, we have simulated waveguide-based silicon SPADs for visible wavelengths, studying both p-n$^+$ and p-i-n$^+$ doping profiles. For our simulated parameters, p-i-n$^+$ SPADs outperform p-n$^+$ devices in terms of PDE and DCR; we identify the optimum device as a p-i-n$^+$ SPAD with $\Delta W = 400$ nm, with a saturated PDE of $52.4 \pm 0.6\%$ at a bias of $V_B = 31.5$ V, FWHM timing jitter of 10 ps, and DCR < 5 cps at 243 K. This is an improvement over our previous study, where the highest PDE obtained was 45% [18].

The PDE is slightly lower than typical free-space SPAD modules with PDEs of up to $\sim 70\%$ [41]; however, our waveguide

devices can offer superior timing performance and dark noise compared to available commercial devices (jitter ~ 35 ps, DCR < 25 cps). We note that even at room temperature, the DCR of a few kcps is acceptable for certain important technologies including LIDAR [42] due to the use of temporal gating, thus indicating the potential applicability of our waveguide SPADs.

Our simulation methods can also be further extended to study other device geometries (e.g. trapezoid waveguides), doping profiles (e.g. p^+-i-p-n^+) and materials (e.g. Ge-on-Si SPADs for near-infrared wavelengths). We are currently fabricating SPAD devices according to the designs reported here, and will compare the measured device performance with our simulation results in the near future.

APPENDIX

A. Trap-Assisted Tunneling

1) Intrinsic Carrier Concentration: We calculate the intrinsic carrier concentration $n_i(T)$ in silicon via [37]:

$$n_i(T) = 5.29 \times 10^{19} \cdot (T/300)^{2.54} \cdot \exp(-6726/T) \quad (8)$$

2) Effective Recombination Lifetime: The effective recombination lifetime $\tau_r(T)$ was measured to be 7 ns at room temperature for an undoped Si rib waveguide device with similar sub-μm dimensions [38]. To obtain a suitable value at 243 K, we analyze the temperature dependence of $\tau_r(T)$: for low-level injection in p-type silicon, $\tau_r(T)$ can be approximated as the electron recombination lifetime [33], i.e.:

$$\tau_r(T) \approx \frac{1}{\sigma_e \cdot \nu_e(T) \cdot N_t} \quad (9)$$

where σ_e is the electron capture cross section, $\nu_e(T)$ is the mean thermal velocity of electrons, and N_t is the trap density. The trap density N_t is assumed to be temperature-independent, while for traps at Si/SiO$_2$ interface with E_t - $E_i = 0.25$ eV, σ_e has been shown to be relatively constant over our relevant temperature range (243–300 K) [43]. Thus, the temperature dependence comes only from $\nu_e(T) \propto \sqrt{T}$, and we obtain

$$\tau_r(243K) = \tau_r(300K) \cdot \sqrt{300/243} \quad (10)$$

B. Band-to-Band Tunneling

Values for B_A, B_B and B_Γ at room temperature are given in ref. [35]. Both B_A and B_Γ are nominally temperature-insensitive, while $B_B(T) \propto [E_g(T)]^{3/2}$, where $E_g(T)$ is the Si bandgap energy [39]:

$$E_g(T) = A + BT + CT^2 \quad (11)$$

in which $A = 1.1785$ eV, $B = -9.025 \times 10^{-5}$ eV/K, and $C = -3.05 \times 10^{-7}$ eV/K^2, for 150 K $\leq T \leq$ 300 K. We thus obtain:

$$B_B(243K) = B_B(300K) \cdot \left(\frac{E_g(243K)}{E_g(300K)}\right)^{(3/2)} \quad (12)$$

REFERENCES

[1] S. Wehner, D. Elkouss, and R. Hanson, "Quantum internet: A vision for the road ahead," *Science*, vol. 362, no. 6412, 2018, Art. no. eaam9288.

[2] X. Qiang *et al.*, "Large-scale silicon quantum photonics implementing arbitrary two-qubit processing," *Nature Photon.*, vol. 12, no. 9, pp. 534–539, Aug. 2018.

[3] P. Sibson *et al.*, "Chip-based quantum key distribution," *Nature Commun.*, vol. 8, Feb. 2017, Art. no. 13984.

[4] J. L. O'Brien, A. Furusawa, and J. Vučković, "Photonic quantum technologies," *Nature Photon.*, vol. 3, no. 12, pp. 687–695, Dec. 2009.

[5] F. Najafi *et al.*, "On-chip detection of non-classical light by scalable integration of single-photon detectors," *Nature Commun.*, vol. 6, 2015, Art. no. 5873.

[6] M. K. Akhlaghi, E. Schelew, and J. F. Young, "Waveguide integrated superconducting single-photon detectors implemented as near-perfect absorbers of coherent radiation," *Nature Commun.*, vol. 6, no. 1, Sep. 2015, Art. no. 8233.

[7] R. E. Warburton, M. Itzler, and G. S. Buller, "Free-running, room temperature operation of an InGaAs/InP single-photon avalanche diode," *Appl. Phys. Lett.*, vol. 94, no. 7, 2009, Art. no. 071116.

[8] Y. Liang *et al.*, "Room-temperature single-photon detection with 1.5-ghz gated InGaAs/InP avalanche photodiode," *IEEE Photon. Technol. Lett.*, vol. 29, no. 1, pp. 142–145, Jan. 2017.

[9] W. D. Sacher, Y. Huang, G.-Q. Lo, and J. K. Poon, "Multilayer silicon nitride-on-silicon integrated photonic platforms and devices," *J. Lightw. Technol.*, vol. 33, no. 4, pp. 901–910, Feb. 2015.

[10] Z. Huang *et al.*, "25 gbps low-voltage waveguide si–ge avalanche photodiode," *Optica*, vol. 3, no. 8, pp. 793–798, 2016.

[11] N. J. D. Martinez *et al.*, "High performance waveguide-coupled Ge-on-Si linear mode avalanche photodiodes," *Opt. Express*, vol. 24, no. 17, pp. 19 072–19 081, Aug. 2016.

[12] N. J. Martinez *et al.*, "Single photon detection in a waveguide-coupled Ge-on-Si lateral avalanche photodiode," *Opt. Express*, vol. 25, no. 14, pp. 16 130–16 139, 2017.

[13] K. K. Mehta *et al.*, "Integrated optical addressing of an ion qubit," *Nature Nanotechnol.*, vol. 11, no. 12, pp. 1066–1070, 2016.

[14] A. Sipahigil *et al.*, "An integrated diamond nanophotonics platform for quantum-optical networks," *Science*, vol. 354, no. 6314, pp. 847–850, 2016.

[15] B. Du *et al.*, "High-speed photon-counting laser ranging for broad range of distances," *Sci. Rep.*, vol. 8, no. 1, Mar. 2018, Art. no. 4198.

[16] M. O'Toole, D. B. Lindell, and G. Wetzstein, "Confocal non-line-of-sight imaging based on the light-cone transform," *Nature*, vol. 555, no. 7696, pp. 338–341, Mar. 2018.

[17] H. A. R. Homulle *et al.*, "Compact solid-state CMOS single-photon detector array for in vivo NIR fluorescence lifetime oncology measurements," *Biomed. Opt. Express*, vol. 7, no. 5, p. 1797, Apr. 2016.

[18] S. Yanikgonul, V. Leong, J. R. Ong, C. E. Png, and L. Krivitsky, "2D Monte Carlo simulation of a silicon waveguide-based single-photon avalanche diode for visible wavelengths," *Opt. Express*, vol. 26, no. 12, May 2018, Art. no. 15232.

[19] H. Chen *et al.*, "100-Gbps RZ data reception in 67-GHz Si-contacted germanium waveguide p-i-n photodetectors," *J. Lightw. Technol.*, vol. 35, no. 4, pp. 722–726, Feb. 2017.

[20] A. Gulinatti *et al.*, "Modeling photon detection efficiency and temporal response of single photon avalanche diodes," *Proc. SPIE*, vol. 7355, 2009, Art. no. 73550X.

[21] R. J. Collins, R. Hadfield, V. Fernandez, S. W. Nam, and G. S. Buller, "Low timing jitter detector for gigahertz quantum key distribution," *Electron. Lett.*, vol. 43, no. 3, pp. 180–182, 2007.

[22] G. Buller and A. Wallace, "Ranging and three-dimensional imaging using time-correlated single-photon counting and point-by-point acquisition," *IEEE J. Sel. Topics Quantum Electron.*, vol. 13, no. 4, pp. 1006–1015, Jul./Aug. 2007.

[23] A. Ingargiola *et al.*, "Avalanche buildup and propagation effects on photon-timing jitter in Si-SPAD with non-uniform electric field," *Proc. SPIE*, vol. 7320, 2009, Art. no. 73200K.

[24] J. Ma *et al.*, "Simulation of a high-efficiency and low-jitter nanostructured silicon single-photon avalanche diode," *Optica*, vol. 2, no. 11, pp. 974–979, 2015.

[25] D. S. Ong, K. F. Li, G. J. Rees, J. P. David, and P. N. Robson, "A simple model to determine multiplication and noise in avalanche photodiodes," *J. Appl. Phys.*, vol. 83, no. 6, pp. 3426–3428, 1998.

[26] I. Wêgrzecka *et al.*, "Design and properties of silicon avalanche photodiodes," *Opto-Electron. Rev.*, vol. 12, pp. 95–104, 2004.

[27] W. Shockley, "Currents to conductors induced by a moving point charge," *J. Appl. Phys.*, vol. 9, no. 10, pp. 635–636, Oct. 1938.

[28] S. Ramo, "Currents induced by electron motion," *Proc. IRE*, vol. 27, no. 9, pp. 584–585, Sep. 1939.

[29] Z. He, "Review of the Shockley–Ramo theorem and its application in semiconductor gamma-ray detectors," *Nucl. Instruments Methods Phys. Res., Sect. A*, vol. 463, no. 1/2, pp. 250–267, May 2001.

[30] Y. Xu, P. Xiang, X. Xie, and Y. Huang, "A new modeling and simulation method for important statistical performance prediction of single photon avalanche diode detectors," *Semicond. Sci. Technol.*, vol. 31, no. 6, 2016, Art. no. 065024.

[31] W. Kindt and H. V. Zeijl, "Modelling and fabrication of geiger mode avalanche photodiodes," *IEEE Trans. Nucl. Sci.*, vol. 45, no. 3, pp. 715–719, Jun. 1998.

[32] G. Hurkx, H. de Graaff, W. Kloosterman, and M. Knuvers, "A new analytical diode model including tunneling and avalanche breakdown," *IEEE Trans. Electron Devices*, vol. 39, no. 9, pp. 2090–2098, Sep. 1992.

[33] D. Schroder, "Carrier lifetimes in silicon," *IEEE Trans. Electron Devices*, vol. 44, no. 1, pp. 160–170, Jan. 1997.

[34] C. H. Ling and Z. Y. Cheng, "An improved analysis for the determination of trap levels in silicon from laser microwave photoconductive decay measurements," *Appl. Phys. Lett.*, vol. 71, no. 22, pp. 3218–3220, Dec. 1997.

[35] G. Hurkx, D. Klaassen, and M. Knuvers, "A new recombination model for device simulation including tunneling," *IEEE Trans. Electron Devices*, vol. 39, no. 2, pp. 331–338, Feb. 1992.

[36] A. Spinelli and A. Lacaita, "Physics and numerical simulation of single photon avalanche diodes," *IEEE Trans. Electron Devices*, vol. 44, no. 11, pp. 1931–1943, Nov. 1997.

[37] K. Misiakos and D. Tsamakis, "Accurate measurements of the silicon intrinsic carrier density from 78 to 340 k," *J. Appl. Phys.*, vol. 74, no. 5, pp. 3293–3297, Sep. 1993.

[38] S. Park *et al.*, "Influence of carrier lifetime on performance of silicon p-i-n variable optical attenuators fabricated on submicrometer rib waveguides," *Opt. Express*, vol. 18, no. 11, May 2010, Art. no. 11282.

[39] W. Bludau, A. Onton, and W. Heinke, "Temperature dependence of the band gap of silicon," *J. Appl. Phys.*, vol. 45, no. 4, pp. 1846–1848, Apr. 1974.

[40] C. Agnesi *et al.*, "Sub-ns timing accuracy for satellite quantum communications," *J. Opt. Soc. Am. B*, vol. 36, no. 3, pp. B59–B64, Mar. 2019.

[41] M. Eisaman, J. Fan, A. Migdall, and S. V. Polyakov, "Invited review article: Single-photon sources and detectors," *Rev. Sci. Instruments*, vol. 82, no. 7, 2011, Art. no. 071101.

[42] I. Takai *et al.*, "Single-photon avalanche diode with enhanced NIR-sensitivity for automotive lidar systems," *Sensors*, vol. 16, no. 4, 2016, Art. no. 459.

[43] D. Garetto *et al.*, "Analysis of defect capture cross sections using non-radiative multiphonon-assisted trapping model," *Solid-State Electron.*, vol. 71, pp. 74–79, May 2012.

Salih Yanikgonul received the B.S. degree in electrical and electronics engineering from Bogazici University, Istanbul, Turkey, in 2014. He is currently working toward the Ph.D. degree in electrical and electronics engineering with Nanyang Technological University and Institute of Materials Research and Engineering, A*STAR, Singapore. His research interests include the engineering of single-photon avalanche detectors, optical spectroscopy of diamond color centers, and quantum optics with metamaterials.

His awards and honors include the Institute of Physics Singapore Outstanding Poster Award in 2019, the Singapore International Graduate Award, and Republic of Turkey Prime Ministry Outstanding Achievement Scholarship.

Victor Leong received the Ph.D. degree in physics from the National University of Singapore, Singapore, in 2016. He was with the Data Storage Institute, A*STAR, Singapore from 2016 to 2018. He is currently a Research Scientist with the Institute of Materials Research and Engineering, A*STAR, Singapore, working on diamond-based quantum emitters and quantum photonics devices.

Jun Rong Ong received the Ph.D. degree in electrical engineering—photonics from the University of California, San Diego, San Diego, CA, USA, in 2014. He is currently with the Electronics and Photonics Group, Institute of High Performance Computing, A*STAR, Singapore as a Research Scientist. His research interests include design and optimization of photonic devices and systems for applications in communications, sensing and computing.

Ching Eng Png received the Ph.D. degree in silicon photonics from Surrey University, Guildford, U.K., in 2004, and the executive MBA degrees from INSEAD, Fontainebleau, France, and Tsinghua University, Beijing, China, in 2014. He also received the Innovative Business Leadership Program from the Massachusetts Institute of Technology Sloan School, Cambridge, MA, USA, in 2013.

From 1999 to 2000, he was with Agilent Technologies. Since 2005, he has been with the Institute of High Performance Computing, A*STAR, Singapore. He co-founded Optic2Connect Pte Ltd. to commercialize optical simulations developed at IHPC and are funded by Skolkovo Foundation and Spring Singapore.

Leonid Krivitsky received the Ph.D. degree in quantum optics from the Department of Physics, Lomonosov Moscow State University, Moscow, Russia, in 2005. He is currently with the Institute of Materials Research and Engineering, A*STAR, Singapore. His research interests include quantum and nonlinear optics, quantum information processing, single photon technology and quantum metrology.