Simulation of a high-efficiency and low-jitter nanostructured silicon single-photon avalanche diode

JIAN MA,† MING ZHOU,† ZONGFU YU, XIAO JIANG, YIJIE HUO, KAI ZANG, JUN ZHANG, JAMES S. HARRIS, GE JIN, QIANG ZHANG, AND JIAN-WEI PAN

1Department of Modern Physics and National Laboratory for Physical Sciences at Microscale, Shanghai Branch, University of Science and Technology of China, Hefei, Anhui 230026, China
2CAS Center for Excellence and Synergetic Innovation Center in Quantum Information and Quantum Physics, Shanghai Branch, University of Science and Technology of China, Hefei, Anhui 230026, China
3Department of Electrical and Computer Engineering, University of Wisconsin, Madison, Wisconsin 53706, USA
4Department of Electrical Engineering, Stanford University, Stanford, California 94305, USA
5e-mail: qiangzh@ustc.edu.cn
6e-mail: pan@ustc.edu.cn

Received 24 August 2015; accepted 18 October 2015 (Doc. ID 248289); published 18 November 2015

Silicon single-photon avalanche diodes (SPADs) are core devices for single-photon detection in the visible and the near-infrared wavelength range and are widely used in many fields such as astronomy, biology, lidar, quantum optics, and quantum information. Due to limitations in their structural design and fabrication, however, the key parameters of detection efficiency and timing jitter cannot be optimized simultaneously. Here, we propose a nanostructured silicon SPAD that achieves high detection efficiency with excellent timing jitter over a broad spectral range. Our optical and electrical simulations show significant performance enhancement compared to conventional silicon SPAD devices. This nanostructured device can be easily fabricated and is thus well suited for practical applications. © 2015 Optical Society of America

OCIS codes: (040.1345) Avalanche photodiodes (APDs); (040.5570) Quantum detectors; (310.6628) Subwavelength structures, nanostructures.

http://dx.doi.org/10.1364/OPTICA.2.000974

1. INTRODUCTION

Silicon single-photon avalanche diodes (SPADs) operating in Geiger mode have become standard devices to detect ultraweak optical signals in many fields such as astronomy [1,2], biology [3–5], lidar [6,7], quantum optics [8,9], and quantum information [10]. Compared with the photomultiplier tube [11], for single-photon detection, the Si SPAD has higher detection efficiency and a lower dark count rate, and does not require a high-voltage operation.

Currently, there are two primary types of structure for Si SPADs, the Slik structure and the thin depletion layer structure. The representative devices based on Slik are single-photon counting modules (SPCMs) produced by PerkinElmer (now Excelitas). The detection efficiency of a typical device is higher than 50% for the spectral range from 600 to 800 nm, and its peak efficiency can be around 70% [12]. This performance benefits from a thick depletion layer (20–25 μm) [13], which guarantees adequate absorption of a single photon. However, the thick depletion layer induces a large timing jitter; their typical time resolution full-width at half-maximum (FWHM) is about 400 ps [12].

For the thin depletion layer structure, the thickness of the depletion region is around 1 μm [14] and its timing resolution can be reduced down to 35 ps FWHM at room temperature [15]. However, the thin depletion layer may result in inadequate absorption of a single photon. The representative devices based on the narrow depletion layer are PDM photon counting modules produced by MPD. The peak detection efficiency of these devices is blue-shifted compared with SPCMs, and the detection efficiency at 800 nm is about 15% [15].

One may ask an interesting question: whether it is possible to combine the excellent performance of high detection efficiency and low timing jitter in the same SPAD device. The solutions to this question have been previously investigated, e.g., using a resonant cavity to enhance the efficiency without increasing the thickness [16]. The cavity enhanced structure, however, reduces a small spectral bandwidth of around a few nanometers, which severely limits its use in practice.

In this paper, we propose a nanostructured silicon SPAD to solve such a problem. The nanostructured SPAD achieves high detection efficiency with excellent timing jitter over a broad spectral...
range. The enhancement mechanism is based on the same principles as light trapping enhancement used in solar cells [17–19]. Note that there have been some efforts to fabricate a nanostructure on the surface of a photodiode to increase its quantum efficiency [20], where the main effect of that nanostructured surface is anti-reflection. And in order to achieve high quantum efficiency, a thick absorption region is needed, which results in poor jitter performance.

In the following sections, we first describe the new device structure and its enhancement mechanism. Then we provide optical and electric simulations to demonstrate the performance improvement of the new device. Finally, we briefly discuss the feasibility of device fabrication.

2. OPTICAL MODELING

The photon detection efficiency \( P_d \) depends on three parameters, i.e., the photon absorption efficiency \( P_a \), carrier collection efficiency \( P_c \), and avalanche probability \( P_b \). In this section, we analyze the light absorption in the semiconductor film of a SPAD, and show how to enhance the absorption efficiency using a nanostructure. For a thin film with a thickness of \( d \), the single-pass absorption is given by

\[
P_a = 1 - e^{-\alpha d},
\]

where \( \alpha \) is the absorption coefficient. We assume that there is no reflection of light at the interface between the semiconductor film and the air. For silicon, the thickness \( d \) has to be tens or even hundreds of micrometers in order to achieve adequate absorption. Such thickness, however, severely limits other parameters of the SPAD. Here we aim to enhance the photon absorption over a broad spectral range while reducing the required thickness. The same challenge arises in the development of solar cells, where the aim is to reduce the material cost of silicon by using thinner film. For this purpose, a light-trapping technique has been developed to improve the light absorption in solar cells. It allows a film to absorb much more light by using judiciously designed nanostructures to scatter light such that light propagates over a longer path in the film. A variety of structures has been proposed and implemented. It has been shown that the upper limit of light absorption by a film is [17,21,22]

\[
P_a = 1 - \frac{1}{1 + 4n^2 \alpha d},
\]

where \( n \) is the refractive index of absorptive material, i.e., silicon in our case. To see the effect of the absorption enhancement, we can take the limit of a very thin thickness such that \( \alpha d \ll 1 \); the absorption calculated from Eq. (2) is \( 4n^2 \) times larger than that from Eq. (1). Based on light trapping, we design a highly efficient SPAD with enhanced light absorption over a broad spectral range.

The structure is based on a thin film with nanocone gratings coated on both sides as shown in Fig. 1(a). The top nanocone grating serves the purpose of broadband antireflection. The periodicity of the grating is 400 nm. The base diameter and the height of the nanocone are 400 and 800 nm, respectively. The choices of parameters are optimized for efficient antireflection; the height and base of the nanocones were adjusted to make sure the transmission is as high as possible. On the other hand, the bottom grating aims to scatter the light strongly toward the lateral direction; the height and the base were adjusted to ensure the absorption almost reaches the theoretical limit. As is known, the breakdown probability in the depletion region is not constant. In order to increase the detection efficiency, besides increasing the absorption efficiency, we also optimize the absorption distribution in the SPAD so that more photons are absorbed in the region where breakdown probability is high. For this purpose, the periodicity of the grating is chosen as 800 nm, which optimizes the scattering efficiency for near-infrared (NIR) light. The base diameter and the height of the nanocones are 750 and 250 nm, respectively. The combination of the effective antireflection and strong scattering is critically important in reaching the upper limit of the light absorption [23]. We choose \( \text{Si}_3\text{N}_4 \) as the material for the top nanocone gratings; \( \text{Si}_3\text{N}_4 \) has been widely used as an antireflection and passivation layer for silicon photodiodes because of its transparency and stability. The bottom nanocone grating consists of silicon. Apart from the nanocone gratings,

![Fig. 1. Structure of two silicon single-photon avalanche diodes (SPADs). (a) Nanostructured SPAD. There are silicon nitride nanocone gratings on the surface, and a silicon nanocone grating on the bottom. For the upper gratings, the period is 400 nm, and the base diameter and the height of the nanocones are 400 and 800 nm, respectively. For the lower gratings, the period is 800 nm, and the base diameter and the height of the nanocones are 750 and 250 nm, respectively. On the bottom of the device there is a silver layer with a thickness of 200 nm. Between the silver layer and the lower gratings there is a silicon oxide spacer layer with a thickness of 2000 nm. (b) Conventional flat-film SPAD. Compared with the nanostructured SPAD, all the gratings are removed, and an additional antireflection layer of silicon nitride with a thickness of 100 nm is placed on the top instead. The dimensions of other layers are the same as the nanostructured SPAD.](image-url)
the SPAD is designed as a typical PIN structure as shown Fig. 1(a), with p-layer of 300 nm thickness, i-layer of 1 μm thickness, and n-layer of 300 nm thickness; both the p-type region and the n-type region are heavily doped. On the backside of the device there is a silver mirror with 200 nm thickness, and a silicon oxide layer with 2000 nm thickness is inserted between the lower gratings and the silver layer to reduce the absorption by the mirror. If the silver mirror is quite close to the nanocones, the absorption in the silver mirror will be enhanced due to the strong field in the tip of the nanocones. The optical absorption is simulated using S4 [24], which is based on the rigorous coupled wave analysis (RCWA) method [25,26].

To verify the enhancement of our device, a conventional SPAD with flat thin film as shown in Fig. 1(b) is also simulated for comparison. The upper gratings are replaced by an antireflection layer with 100 nm thickness, and the lower gratings are removed. The silicon PIN structure is the same as the nanostructured SPAD. There are also a silver mirror and a silicon oxide spacer on the backside of the film. For a fair comparison, a Si$_3$N$_4$ antireflection layer with a thickness of 100 nm is used on the top of the PIN. The light absorption in the PIN junction in both structures is calculated and shown in Fig. 2(a). For the flat thin-film structure, the absorption is around 60% in the short wavelength range. In the NIR region, however, the absorption falls below 20%, indicating poor efficiency for the SPAD. In great contrast, the nanostructured SPAD maintains high absorption around 80% from visible to NIR regions.

3. ELECTRIC MODELING

Based on the simulation results of the absorption efficiency, we can further characterize the parameters of interest for single-photon detection, e.g., detection efficiency and timing jitter, through the simulations in Geiger mode.

For the SPADs as shown in Fig. 1 there are three regions that can absorb incoming photons, i.e., the upper p-type region, the middle i-type region, and the lower n-type region. In the different regions, the probability that photo-generated carriers reach the depletion region, i.e., $P_c$, is also different. When a photon is absorbed in the depletion region, an avalanche is initiated immediately so that $P_c$ is close to 1. Since the depletion region is mainly in the intrinsic layer, the lifetime of photo-generated carriers is longer than the avalanche process such that the recombination effect can be ignored in this region [27]. When a photon is absorbed in the p-type region or the n-type region, due to the absence of high electric field the photo-generated minority carriers move randomly before reaching the depletion region. This process has an effect on detection efficiency and timing jitter as discussed previously [27]. In this paper, we use the 1D random walk model to simulate such a process. The diffusion coefficient in the simulation is taken from the literature [28]. As the boundary conditions of the model, we assume that random walk starts when the first electron–hole pair is generated in the neutral region; meanwhile because of the high quality of the Si$_3$N$_4$/Si interface, we ignore the recombination at the Si$_3$N$_4$/Si interface.

As the last step for the evaluation of detection efficiency, we calculate the avalanche probability $P_b$, i.e., the probability that an electron–hole pair triggers a self-sustaining avalanche, by using the random path length (RPL) model [29]. In the RPL model, the random ionization path length of electrons and holes can be described by their probability density functions, which are used as inputs. For electrons, the probability density function is

$$h_e(ξ) = \begin{cases} 0, & ξ ≤ d_e' \\ α_e \exp(-α_e(ξ - d_e')), & ξ > d_e', \end{cases}$$

where $α_e$ is the enabled ionization coefficient of electrons, and $d_e'$ is the dead space length, $d_e' = E_{th}/q · f$, where $E_{th}$ is the ionization threshold energy of electrons and $f$ is the electric field in the depletion region. The parameters of $α_e$ and $d_e'$ are calculated using the values of the local ionization coefficient and $E_{th}$ taken from the literature [30,31]. The probability density function of holes can be obtained similarly. In such a way, the RPL model can be effectively simulated using the Monte Carlo method [32]. For each trial, when an avalanche breakdown occurs, a detection event and threshold crossing time of the avalanche current are recorded. After enough turns of trials, $P_b$ can be obtained by calculating the ratio of the number of recorded events to the number of trials.

The results of detection efficiency simulation for the two SPADs are shown in Fig. 2(b). Both SPADs are biased with the same excess voltage $V_{ex} = 4$ V; the nanostructured SPAD exhibits much higher detection efficiency than the conventional SPAD at any wavelength. Particularly in the NIR range, the
improvement of the detection efficiency is significant, which is similar to the trend as shown in Fig. 2(a).

Timing jitter is another key parameter of the SPAD to characterize the time uncertainty between the photon absorption and avalanche detection. It can be calculated as

$$\sigma = \sqrt{(t_b^2) - \langle t_b^2 \rangle}, \quad (4)$$

where $t_b$ is defined as the time for an avalanche current reaching the threshold. For the SPADs, the timing jitter is mainly attributed to the time dispersion of the photo-generated carriers in the neutral region to reach the high field region, and the intrinsic randomness in the avalanche process as well [28,33]. Previous studies have revealed that the peak of the timing jitter histogram is coming from the photons absorbed in the depletion region, which undergo an avalanche growth process directly, and the tail of the timing jitter histogram is coming from the diffusion process in the neutral region. The contributions of timing jitter from the avalanche growth process (the peak of timing jitter histogram) can be modeled by a Gaussian function [27]. In our nanostructured SPAD, the absorption in the depletion region has been maximized for infrared wavelengths, most photons are absorbed in the depletion region, and the timing jitter will mainly be determined by these photons. Because the width of the $i$ layer in the nanostructured SPAD is the same as that of the corresponding layer in the flat SPAD, and both the nanostructured SPAD and the flat SPAD bias with the same excess voltage, as a rough estimate, the nanostructured SPAD can have the same timing jitter performance as the flat SPAD. In order to evaluate the timing jitter performance further, first we use the random walk method to evaluate the contribution of time dispersion in the neutral region. Second, the intrinsic randomness in the avalanche process includes the contributions from two parts, i.e., the avalanche buildup process and the propagation process [28]. Here, a low threshold current of 0.2 mA is used to detect avalanche occurrence, such that the avalanche process is still confined around the seed point. In this scenario, the avalanche buildup process plays an important part in the intrinsic randomness [28]. The time uncertainty contribution due to the avalanche buildup process is extracted directly from the simulation of the RPL model.

As shown in Fig. 3, different from the superconductor single-photon detector [34], the nanostructured SPAD has a non-Gaussian timing jitter, with a FWHM of timing jitter of 15 ps, and a full width at tenth maximum (FWTM) of 35 ps. In comparison with the conventional SPAD, the nanostructured SPAD has a less perceptible diffusion tail. This phenomenon can be explained by the fact that the widths of both the $p$-type and $n$-type regions are small and the time dispersion induced by the diffusion process in this region is small.

Figure 4 shows the timing jitter of the two SPADs as a function of wavelength with $V_{ex} = 4$ V. For the short wavelength photon, the timing jitter is larger than for the long wavelength photon; we attribute this to the fact that a large part of the short wavelength photon will be absorbed in the top $p$-type region, and before triggering an avalanche process the carrier will undergo a diffusion process in the $p$-type region, and this will induce an additional timing uncertainty. The overlap of two lines indicates that the timing jitter of the nanocone SPAD can be as low as the flat thin-film SPAD. The inset in Fig. 4 exhibits that the timing jitter characteristics of the nanocone SPAD decrease as the bias voltage increases.

Combining the simulation results of the detection efficiency and timing jitter, one can conclude that the nanostructured SPAD with a depletion width of 1 μm presents the advantages of high efficiency and low jitter simultaneously. Therefore, nanostructure is an effective approach to solve the aforementioned coexistence problem of SPAD performance.

With the help of light absorption enhancement, the nanostructured SPAD achieves high absorption efficiency with a very thin depletion layer. By increasing the thickness of the depletion layer, such high absorption efficiency can also be achieved using the conventional flat-film SPAD. However, as the thickness of the depletion layer increases, the jitter performance of the SPAD significantly decreases [35]. To verify such an effect, we compare the jitter of the two SPADs at the same absorption efficiency. The depletion layer thickness of the flat-film SPAD is increased to keep the same absorption efficiency.

The timing jitter results are shown in Fig. 5, at the wavelength of 900 nm. The large difference between the two lines in Fig. 5 shows a remarkable improvement in the timing jitter using the nanostructured SPAD. For instance, with a depletion layer thickness of 1 μm, the nanostructured SPAD has a detection efficiency of 32%. For the flat-film SPAD, in order to achieve the same efficiency the thickness of the depletion layer has to be increased.

![Fig. 3. Timing response of nanostructured SPAD at 800 nm, biased with excess bias voltage $V_{ex} = 4$ V.](image1)

![Fig. 4. Timing jitter versus wavelength of two SPADs. Both SPADs are biased with the same excess bias voltage $V_{ex} = 4$ V. The inset shows the excess voltage dependence of timing jitter at 900 nm for the nanocone SPAD.](image2)
from 1 to 5.8 μm. As a result of such thickness, the timing jitter increases with one order of magnitude.

By increasing the thickness of the depletion layer, we can also increase the absorption efficiency of the nanostructured SPAD. However, unlike the flat-film structure, the nanostructured SPAD does not require an extremely thick depletion layer in order to obtain high absorption efficiency. Considering a nanostructured SPAD with a depletion layer of 3 μm thickness, a p-layer of 100 nm thickness, and an n-layer of 100 nm thickness, when the nanostructured SPAD is biased with excess voltage $V_{ex} = 11 \text{ V}$, $P_d$ is higher than 80% in the range from 400 to 1000 nm as shown in Fig. 6(a). $P_d$ reaches around 60% at the wavelength of 900 nm as shown in Fig. 6(b), which is much higher than that in the case of 1 μm thickness as shown in Fig. 2(b). The flat-film SPAD with the same layer parameters is also simulated for comparison. Without nanostructure the detection efficiency of the flat-film SPAD drops to around 20% at the same wavelength. Similarly, we calculate the timing jitter performance of this thick nanostructured SPAD. In Fig. 7, the timing jitter as a function of excess bias voltage is plotted, and here we assume the wavelength of the photon is 900 nm. With $V_{ex} = 11 \text{ V}$, the timing jitter is only increased to 20 ps. On the other hand, for the flat-film structure SPAD, the thickness of the depletion layer has to be 12 μm to achieve a detection efficiency of 60%, indicating a much larger timing jitter.

We remark that although the timing jitter for the nanostructured SPAD is similar to that for the flat SPAD, this does not mean that our electrical modeling is not necessary. When the light is trapped in the nanocone waveguide, it propagates in the 2D dimension nanostructure. It can be absorbed in any point of the 2D structure, and whenever it is absorbed, an electron–hole pair is generated and an avalanche starts. Compared to a plain Si APD, the nanostructured SPAD has a different absorption distribution profile. The light transportation can be ignored due to the large inherent jitter of carrier collection processes and avalanche buildup times. On the other hand, the role that the electronic process plays remains unclear, which calls for further investigation.

In addition to better device performance, this nanostructure is easy to fabricate. A silicon nitride nanocone structure can be made by applying a self-assembled nickel nanoparticle and dry etching [36]. The double-sided nanocone structure can be processed sequentially, and one may even directly process on a thin-film silicon wafer [37]. After nanocone etching, the device backside can be coated with oxide and metal to form the final devices.

4. CONCLUSIONS

In summary, we have proposed and theoretically simulated a nanostructured SPAD that has the remarkable performance of high detection efficiency over a broad spectral range and low timing jitter at the same time. Our approach effectively solves the
problem of coexistence of high efficiency and low jitter for silicon SPADs. Moreover, by optimizing the structure, the detection efficiency of the nanostructured SPAD could be as high as the conventional thick silicon SPAD, particularly in the NIR range, while maintaining a pretty low timing jitter. Such a nanostructured SPAD is well suited for many practical applications requiring high efficiency and low timing jitter.

Funding. Chinese Academy of Sciences (CAS); Chinese National Fundamental Research Program (2011CB921300, 2013CB336800); National Natural Science Foundation of China (NSFC).

Acknowledgment. The authors thank Xintao Bi, Jin Wang, and Jinrong Wang for providing the workstation for computation.

†These authors contributed equally to this work.

REFERENCES