Ge-on-Si Single-Photon Avalanche Diode Detectors: Design, Modeling, Fabrication, and Characterization at Wavelengths 1310 and 1550 nm

Ryan E. Warburton, Giuseppe Intermite, Maksym Myronov, Phil Allred, David R. Leadley, Kevin Gallacher, Douglas J. Paul, Senior Member, IEEE, Neil J. Pilgrim, Leon J. M. Lever, Zoran Ikonic, Robert W. Kelsall, Edgar Huante-Cerón, Andrew P. Knights, and Gerald S. Buller

Abstract—The design, modeling, fabrication, and characterization of single-photon avalanche diode detectors with an epitaxial Ge absorption region grown directly on Si are presented. At 100 K, a single-photon detection efficiency of 4% at 1310 nm wavelength was measured with a dark count rate of ~6 megacounts/s, resulting in the lowest reported noise-equivalent power for a Ge-on-Si single-photon avalanche diode detector (1 $\times$ 10$^{-14}$ WHz$^{-1/2}$). The first report of 1550 nm wavelength detection efficiency measurements with such a device is presented. A jitter of 300 ps was measured, and preliminary tests on after-pulsing showed only a small increase (a factor of 2) in the normalized dark count rate when the gating frequency was increased from 1 kHz to 1 MHz. These initial results suggest that optimized devices integrated on Si substrates could potentially provide performance comparable to or better than that of many commercially available discrete technologies.

Index Terms—Detector, germanium on silicon, single-photon avalanche diode, single-photon counting.

I. INTRODUCTION

NUMBER of emerging application areas require the technological development of efficient, low-noise, picosecond-response, single-photon detection at near-infrared wavelengths. While Si single-photon avalanche diodes (SPADs) perform effectively at wavelengths below 1000 nm, efficient detection at wavelengths greater than this remains more problematic despite the commercial availability of some single-photon detectors. Applications such as quantum key distribution [1], time-of-flight ranging [2], and remote gas sensing [3] all drive the requirement for efficient, low-noise, and high-sensitivity infrared single-photon detectors. There has been increasing interest in quantum photonic circuitry using Si platforms [4] which further requires monolithic SPAD technology that can be integrated onto a Si substrate at high density. Particular demand is placed on single-photon detectors efficient at the low-loss optical fiber wavelength regions (1310 and 1550 nm) and operational at or near room temperature. Commercially available infrared detectors suffer from various drawbacks that limit their sensitivity and practicality: gated operation, long dead-time, high dark count rates (DCRs), afterpulsing, and low operating temperatures.

InGaAs/InP avalanche photodiodes (APDs) operated above breakdown for single-photon detection were first reported in 1996 [5], and, subsequently, custom-designed InGaAs/InP SPADs were fabricated and characterized [6]. More recently, novel designs that incorporate negative feedback to quench the avalanche current passively have been realized [7]. The effects of after-pulsing, where dark events are induced by the slow release of carriers trapped during previous avalanche events, remain a significant operational issue. The after-pulsing probability can be reduced by decreasing the charge passing through the device during the avalanche event by using electrical gating techniques, such that the detector is active only for a short window around the expected photon arrival time [8], [9]. For applications such as laser ranging and time-resolved photoluminescence, however, a long temporal detection window may be required for more comprehensive measurements. For this reason, free-running InGaAs/InP SPADs have been investigated, although the overall detection efficiency is reduced in this approach [7], [10].

Recently, superconducting nanowire single-photon detectors have been demonstrated with impressive performance in terms of low DCRs, fast timing jitter, and short recovery times, which led to high maximum count rates [11]. The main disadvantages are their operating temperature of 4 K [2] and below, and the small active areas that necessitate potentially lossy optical fiber coupling.
Ge has good absorption properties at wavelengths up to \( \sim 1600 \text{ nm at } 300 \text{ K} \) [12], and some work has been performed on all-Ge SPADs [13]. The high-field multiplication layer was necessarily of narrow-gap Ge, resulting in large DCRs due to band-to-band tunneling. After-pulsing was also a serious issue in these homojunction devices [14]. A SPAD that combines a Ge absorption layer and a larger gap Si multiplication layer potentially offers low-noise operation across the telecommunications wavebands, and may offer a device less affected by after-pulsing as discussed later in this paper. Such a Si-substrate-based infrared detector geometry also opens up significant potential for on-chip integration with other Si photonics components operational at the low-loss windows at 1300 or 1550 nm wavelengths [15]. Previous work demonstrated an improvement in single-photon detection efficiency (SPDE) at infrared wavelengths in a SPAD utilizing a strained SiGe/Si multiple quantum well (MQW) absorbing layer, but the necessarily low fraction of Ge and narrow thickness of MQW layer resulted in low efficiency above 1000 nm wavelength—i.e., only \(<0.01\%\) SPDE at \( \lambda =1210 \text{ nm} \) was reported [16].

More recently, techniques for growing Ge directly on Si have improved such that thick (\(>1 \mu\text{m}\)) all-Ge layers have been grown on Si substrates. Most techniques involve growing a seed layer (\(\sim 25–100 \text{ nm}\)) of Ge on Si at low temperature (\(\sim 350–400 \text{ °C}\)) to prevent strain relaxation, which manifests as undesirable island growth. Because of the large lattice mismatch between Si and Ge (4.2%), the buffer layer will contain a high threading dislocation density (TDD), typically \(10^8–10^9 \text{ cm}^{-2}\) [17]. After the seed layer, the thick Ge layer can be grown at \(\sim 650 \text{ °C}\). The structure is then annealed (800–900 °C, several cycles) to reduce the TDD to typically \(10^6–10^7 \text{ cm}^{-2}\) [17]. As a result of these advances in growth techniques, Ge integration in Si photonics has produced such integrated devices as high-gain APDs [18], p–i–n detectors [19], and metal–semiconductor–metal detectors [20]. The demands of effective single-photon detection using Ge-on-Si heterostructures place particularly challenging constraints on growth and fabrication. Recently, some reports of Ge-on-Si SPAD characterization have appeared in the literature, most notably by Lu et al. [21], where SPDE of up to 14% were reported at 1310 nm. These devices, however, had a very high DCR (\(>10^8 \text{ Hz}\)), where it is likely that the device has insufficient time to fully recover its bias level before another dark count is triggered, resulting in an underestimate of the DCR for a given bias. This is further evidenced by a notable decrease in the DCR as the gating frequency was increased from 1 kHz to 100 kHz—exactly opposite to that expected with after-pulsing (see later) but consistent with lack of voltage recovery. In [21], the SPDE was measured with an incident photon flux of 1 photon per pulse, thus giving a high probability of multiphoton pulses incident on the device, potentially leading to an overestimation of SPDE. Another report of a Ge-on-Si device claiming single-photon sensitivity was published by Aminian et al. [22], although the detection efficiency in the Geiger mode was measured only by analysis of the photocurrent above breakdown, which cannot be regarded as a robust single-photon counting characterization method.

II. DEVICE DESIGN, MODELING, AND PROCESSING

In this paper, single-photon detection by custom-designed prototype Ge-on-Si SPADs, illuminated by \(<1\) photon/pulse photon flux, is reported at 1310 and 1550 nm wavelengths in accordance with accepted techniques of time-correlated single-photon counting, as described, for example, by Becker [23]. Separate absorption, charge, and multiplication (SACM) structures were fabricated, using a Si multiplication layer and a Ge absorber, as shown in Fig. 1. Infrared photons are absorbed in the Ge absorption layer and create electron–hole pairs. The electrons are accelerated toward the Ge–Si interface and, once inside the high-field Si multiplication region, may undergo impact ionization. If the electric field is held above the avalanche breakdown threshold, further impact ionization of both holes and electrons can create a self-sustaining avalanche current, which is readily detectable. This avalanche current can be extinguished only by taking the device below the breakdown field, which can be achieved by use of various quenching approaches [24].

The device is similar to the APD structure reported by Kang et al. [18]; however, to achieve SPAD performance, significant modifications to the design are required. The material was designed with a thicker multiplication region and a different doping concentration for the charge sheet layer to enable SPAD performance. This increased layer thickness also increases the voltage difference between punch-through (when the electric field extends into the Ge absorber) and the avalanche breakdown voltage. This voltage difference becomes important when operating the devices at lower temperatures since the breakdown voltage decreases as the temperature is reduced while the punch-through voltage remains almost unchanged because it depends only on the doping of the charge sheet layer. If there was only a small difference between punch-through and breakdown, at lower temperatures the device might reach breakdown before punch-through has occurred, significantly reducing the SPDE at the design wavelengths of 1310 and 1550 nm.
Before growth, the structures were modeled in two dimensions in terms of their electric field profile using Silvaco ATLAS. One of the main considerations was the correct doping of the charge sheet layer. In these SACM structures, under reverse bias, the presence of the charge sheet must ensure that the electric field in the narrow-bandgap Ge absorber remains low (to avoid tunneling) and well below its avalanche breakdown field of (∼100 kV/cm), and that the field in the multiplication layer is greater than the breakdown field (∼300 kV/cm) to provide impact ionization. For these reasons, three different wafers with different doping densities of the charge sheet layer were simulated and then grown. Fig. 2 shows the simulated electric field at 95% of the breakdown voltage for doping densities of 1 × 10^17, 2 × 10^17, and 5 × 10^17 cm⁻³ in the charge sheet. With a doping density of 1 × 10^17 cm⁻³ (black line in Fig. 2), the field in the Ge is too high—tunneling and impact ionization would increase the dark current and hence DCR. At 5 × 10^17 cm⁻³ doping (blue dotted line), too much of the electric field is dropped across the charge sheet and consequently the Ge is not depleted, leading to much reduced detection efficiency at wavelengths greater than 1000 nm. A doping density of 2 × 10^17 cm⁻³ (red dashed line), however, shows that the whole intrinsic Ge layer is depleted with a modest electric field such that photogenerated carriers will drift into the multiplication region.

The structures were grown by reduced-pressure chemical vapor deposition on highly doped n⁺⁺ Si substrates. Intrinsic Si of 1 μm thickness was grown to form the multiplication/avalanche region. Three different wafers, each containing a different dopant level in the charge sheet (1, 2, and 5 × 10^17 cm⁻³), were grown to allow for growth tolerances, although device simulations showed that a charge sheet doping of 2 × 10^17 cm⁻³ had the best performance as previously mentioned. A low-temperature Ge seed layer of 50 nm was grown followed by a high-temperature growth of intrinsic Ge. These Ge layers are not intentionally doped, and the low growth temperature for the seed layer prevents any significant dopant migration from the Si charge sheet. Finally, a high-concentration boron-doped (~5 × 10^19 cm⁻³) p⁺⁺ Ge layer was grown. Cylindrical mesas, ranging from 25 to 500 μm in diameter, were then defined by photolithography and etched anisotropically by an inductively coupled plasma reactive ion etch using fluorine-based chemistry (SF₆/C₄F₈), resulting in near-vertical sidewalls [25]. Ni was chosen for the contacts, as it is known to form the lowest electrical resistivity phases for silicides and germanides [26]. A 20 nm layer of Ni was deposited to form the top contact and then annealed at 325 °C. The structure was passivated and planarized with Si₃N₄. Via holes were etched to allow interconnects and bond pads to the contacts using 1.2 μm of deposited Al.

III. CHARACTERIZATION AND RESULTS

Prior to characterization, promising devices were selected through the analysis of their current–voltage (I–V) characteristics. An example I–V is shown in Fig. 3 of a 25 μm-diameter device (2 × 10^17 cm⁻³ doping in the charge sheet) at different temperatures. An I–V trace can give an indication of the device suitability from dark current considerations, particularly the multiplied dark current characteristic. From Fig. 3, it is clear that, as the temperature is decreased from room temperature to 100 K, breakdown occurs at lower voltages and the onset of breakdown becomes progressively sharper. As the temperature is reduced, the characteristics demonstrate a lowering of dark current just prior to breakdown—previously observed in other semiconductor material systems as a good indication of lower noise photon-counting behavior at biases above breakdown [27]. The inset of Fig. 3 also shows the photocurrent for light at wavelengths of 1310 and 1550 nm.

To characterize the devices in terms of single-photon detection capability, the setup shown in Fig. 4 was used. The devices were mounted in a liquid nitrogen cryostat, enabling stable temperature tuning between 77 and 300 K. Pulsed picosecond semiconductor laser diodes were used to test the detection efficiency at both 1310 and 1550 nm. These were coupled into a single-mode fiber (SMF-28) and into a 50/50 fiber splitter: one output was used to constantly monitor the optical power, and the other output passed through...
connections, and those with colored lines are free-space optical connections. Those with dotted lines are fiber-optic connections. The connections denoted by black lines are electrical connections. Those with colored lines are free-space optical connections.

A calibrated programmable optical attenuator (HP 8156A) with up to 60 dB of attenuation available. Prior to characterization, the optical power level arriving at the device under test was measured and compared with the reading on a calibrated fiber-coupled power meter. This enabled direct monitoring of the power level arriving at the device once the optical pulse had passed through the optical attenuator, thereby ensuring that a photon flux of <0.1 photons per pulse (on average) was incident on the device. For characterization purposes, the devices were operated in gated mode [6]. A dc bias of a few Volts below the breakdown voltage $V_{bd}$ was applied, and an electrical pulse, typically tens of nanoseconds duration, was used to bias the device above $V_{bd}$ into the so-called Geiger mode of operation. The breakdown voltage was defined as the voltage at which the current reached 1 $\mu$A. When an avalanche was initiated (by either a dark count or a photogenerated event), the avalanche current persisted until the end of the gate when the voltage was brought back below $V_{bd}$. In the future, active quenching circuits could be employed to quench the avalanche current more rapidly, thereby minimizing the charge flow per event and thus reducing the probability of trapped carriers that cause after-pulsing [24].

The output pulse from the device was split to enable oscilloscope traces to be recorded while simultaneously providing the stop signal for the photon-counting card (Edinburgh Instruments TCC900). The start signal was provided by a master clock that has three outputs: one for the TCC900, another for the laser driver, and the third for the gate generator. For each operating condition, two photon-counting histograms were recorded: one in completely dark conditions, and one with an attenuated laser pulse coincident with the gate on the detector. The DCR and SPDE can be extracted from these histograms by summing the counts within a certain region of interest. Since the gate-on and gate-off times result in the device not being at a constant bias for the whole gate duration, only a portion from the center of the histograms was considered—this portion has a flat background level showing that the bias was stabilized and gives a true value for both SPDE and DCR. Other methods, such as using a photon counter over a predetermined gate period, may underestimate DCR and overestimate the SPDE (since it is difficult to ascertain the effects of after-pulsing with a gated photon counter which does not possess the necessary timing resolution).

Fig. 5 presents the bias-dependent SPDE and DCR of a 25-$\mu$m-diameter device operated at 100 K. This device had a dark current of $\sim$0.5 nA when measured at 95 % of $V_{bd}$.

The SPDE depends mainly on the photon absorption probability in the depleted Ge absorption region, the probability of electron drift into the Si multiplication region, and the avalanche triggering probability in the Si multiplication region. While the first two phenomena are unlikely to be strongly dependent on the excess bias, the avalanche triggering probability increases linearly with excess bias, before becoming saturated. In Fig. 5, it can be observed that the SPDE increases linearly with excess bias due to the increasing field within the device. At 10% excess bias, an SPDE of 4% at $\lambda = 1310$ nm was recorded, which shows a significant improvement when compared with strained SiGe/Si MQW structures (only 0.001% at $\lambda = 1210$ nm) [16]. These results show good agreement with the SPDE obtained with a commercially available, planar-geometry, homojunction Ge APD operated in the Geiger mode, where an SPDE of between 4% and 30% was reported at $\lambda = 1310$ nm and a temperature of 77 K [28].

These prototype mesa devices exhibited a high DCR which limited the operating temperatures to between 100 and 150 K. Additionally, the high levels of DCR restricted the maximum excess bias applied. As shown in Fig. 5, a DCR of $10^5$–$10^7$ Hz was obtained at 100 K, which is similar to those obtained using commercially available Ge APDs as SPADs. These commercial Ge homojunction APDs were, however, optimized planar devices, whereas our prototype devices are of mesa geometry and thus suffer from the deleterious effects caused by the high density of surface states at the sidewalls, which is known to increase DCR. After plotting the dark current density for devices of different sizes, the dark current appeared to scale with device perimeter, and not area, demonstrating that surface effects dominate the dark current and hence the DCR.
in this sample set. Compared to InGaAs/InP planar SPADs, the DCR is several orders of magnitude higher: DCRs of $10^2-10^3$ Hz are achievable at this temperature [6]. The high DCR is likely to be caused primarily by the surface effects from the mesa geometry. The present Ge-on-Si results indicate that the use of different passivation techniques for mesa sidewall surface states, as well as planar device geometries which are known to reduce DCR, should result in Ge-on-Si devices with significantly higher performance than all demonstrated Ge homojunction APDs. It is also worth noting that reducing the TDD may also decrease the dark current, further improving the DCR [29].

The exponential increase in DCR with increasing excess bias does demonstrate that our devices recover fully before the subsequent gate period. The present results are therefore encouraging that optimization can substantially improve the performance.

Based on these single-photon measurements, the noise-equivalent power (NEP) was also calculated, from

$$\text{NEP} = \frac{h \nu}{\text{SPDE}} \sqrt{2 \text{DCR}}. \quad (1)$$

Hence, lower NEPs are characteristic of more sensitive detectors. Overall, the NEP was similar across the range of excess voltages measured. However, 4% SPDE and an NEP of $1 \times 10^{-14}$ WHz$^{-1/2}$ at a wavelength of 1310 nm compares well with other work where NEPs of $\sim 1.6 \times 10^{-14}$ and $4 \times 10^{-15}$ WHz$^{-1/2}$ were reported using commercially available planar all-Ge APDs operated in the Geiger mode at a temperature of 77 K [13], [28]. Although not stated explicitly by Lu et al. [21], we can infer that the NEP of those Ge-on-Si devices was on the order of $3 \times 10^{-14}$ WHz$^{-1/2}$ at $\lambda = 1300$ nm (from the quoted SPDE and DCR) at a higher temperature of 200 K. However, there still remains a performance gap between these results on Ge-containing SPADs and those for InGaAs/InP SPADs where NEPs of less than $1 \times 10^{-17}$ WHz$^{-1/2}$ at $\lambda = 1550$ nm have been reported at a temperature of 193 K [6], [7].

Jitter was investigated at various excess bias levels. The measured jitter was a convolution of the laser pulse width ($\sim 50$ ps), the detector response, and the contribution from the rest of the acquisition system. The minimum jitter at full-width half-maximum (FWHM) is shown in Fig. 6 and was measured to be 300 ps at 10% excess bias. With the same experimental setup, we had previously measured jitter of less than 80 ps with alternative low-jitter all-Si SPAD detectors, and hence we can assume that the overall measured jitter is dominated by the detector contribution. This value shows good agreement with the jitter measured previously ($\sim 100-350$ ps, $\lambda = 1310$ nm, laser pulse width $= 40$ ps) for homojunction Ge APDs operated in the Geiger mode [28]. Higher excess biases will reduce this value in future optimized devices.

The use of histograms for characterization provides some information on after-pulsing, as well as jitter. If the background levels are the same for both dark and light measurements (as in Fig. 6), it is an indication that the detector is operating in a regime with negligible after-pulsing.

Although the time-correlated carrier counting method provides a more comprehensive analysis of after-pulsing [30], we used the simpler method of increasing the gating frequency while observing the impact on the DCR. A slight increase (by nearly a factor of 2) in the normalized DCR was observed when increasing the gating frequency from 1 kHz to 1 MHz at a temperature of 150 K, as shown in Fig. 7. This behavior was also observed using two different Ge homojunction APDs operated in the Geiger mode [28]. Although InGaAs/InP SPADs have lower DCR at these temperatures, such devices demonstrate a rapid increase in the DCR due to after-pulsing at frequencies above 100 kHz, perhaps highlighting an advantage of Ge-on-Si devices [6].

In InGaAs/InP SPADs, it was found that the traps contributing to after-pulsing were located within the high-field InP multiplication region [31], [32]. Cova et al. [31] further postulated that only traps within the high-field region are of concern, since traps within the low-field InGaAs layer do not initiate an after-pulse [31]. Similarly, in a Ge-on-Si SPAD the avalanche is electron-initiated, and if a photon is absorbed in the Ge, then the photogenerated hole will drift to the top contact adjacent to the Ge layers. When the photogenerated electron passes into the multiplication region, both electrons and holes will undergo impact ionization. The holes will drift out of the Si layers toward the Ge absorber and to the top contact, while the electrons will drift to the Si substrate contact. If holes are trapped within the Ge layer and are later released, they cannot contribute to after-pulsing since they drift directly to the top contact and do not enter the multiplication region. Hence, electrons and/or holes that are trapped in the Si multiplication region are the only carriers that are likely to initiate an after-pulse. Generally, the after-pulsing within all-Si SPADs is very low at higher temperatures [33], and much lower than exhibited in all-Ge SPADs [14], which may explain the apparently small effects of after-pulsing in these Ge-on-Si devices. Since the multiplication region is in Si, the Ge-on-Si SACM SPAD may prove to be a superior device in terms of after-pulsing than alternative SPADs operating at this wavelength band.
Fig. 7. DCR versus gating frequency for a 25-μm-diameter device at a temperature of 150 K demonstrating the dependence of the normalized DCR with the gate frequency.

The performance at the longer wavelength of 1550 nm was measured on a device at 125 K. An SPDE of ∼0.15% at 6% excess bias was measured, resulting in an NEP of 5 × 10^{-12} WHz^{-1/2} with a jitter of 420 ps (FWHM). As Ge is cooled, its bandgap increases and hence the absorption at longer wavelengths decreases rapidly (assuming ∼0.2% tensile strain in the Ge layer, due to thermal expansion mismatch, the direct bandgap at 125 K is 0.84 eV, and hence the 1550-nm (0.8-eV) photons lie outside the main absorption edge [12]. For this reason, we see a factor of >10 decrease in the SPDE between 1310 and 1550 nm as previously observed with a Ge homojunction APD [28]. As previously mentioned, we expect that process optimization and planar device geometry will reduce the DCR to a level that will allow higher temperature operation, which will significantly improve the SPDE at 1550 nm.

IV. CONCLUSION

In summary, prototype epitaxial Ge-on-Si SPAD mesa geometry devices were designed, modeled, grown, fabricated, and characterized. The performance at λ = ∼1310 nm was comparable to that of the best all-Ge APDs previously reported, for Geiger mode operation. The efficiency at 1550 nm was, however, a factor of >10 lower due to the reduced absorption coefficient at low temperatures. DCRs of ∼10^{6}−10^{7} were observed—values which we anticipate can be significantly reduced by optimization of the mesa sidewall passivation or by use of a planar device geometry. At 100 K, the NEP was measured to be 1 × 10^{-14} WHz^{-1/2} (at λ = 1310 nm) and was constant over the range of measured excess biases. The total measured jitter at FWHM varied with excess bias as expected, but was as low as 300 ps. The contribution of after-pulsing to the DCR at higher gating frequencies was found to be low, indicating a potential advantage over the InGaAs/InP materials system.

ACKNOWLEDGMENT

The authors would like to thank the staff of the James Watt Nanofabrication Centre at the University of Glasgow for help in fabricating the devices reported in this paper.

REFERENCES


