Ge-on-Si single-photon avalanche diode detectors for short-wave infrared wavelengths

Fiona Thorburn, Xin Yi, Zoe M Greener, Jaroslaw Kirdoda, Ross W Millar, Laura L Huddleston, Douglas J Paul and Gerald S Buller

School of Engineering and Physical Sciences, Institute of Photonics and Quantum Sciences, Heriot-Watt University, Edinburgh EH14 4AS, United Kingdom
James Watt School of Engineering, University of Glasgow, Rankine Building, Oakfield Avenue, Glasgow G12 8LT, United Kingdom
* Author to whom any correspondence should be addressed.
E-mail: f.thorburn@hw.ac.uk
Keywords: single-photon, short-wave infrared, LiDAR, single-photon avalanche diodes (SPADs)

Abstract
Germanium-on-silicon (Ge-on-Si) based single-photon avalanche diodes (SPADs) have recently emerged as a promising detector candidate for ultra-sensitive and picosecond resolution timing measurement of short-wave infrared (SWIR) photons. Many applications benefit from operating in the SWIR spectral range, such as long distance light detection and ranging, however, there are few single-photon detectors exhibiting the high-performance levels obtained by all-silicon SPADs commonly used for single-photon detection at wavelengths <1 µm. This paper first details the advantages of operating at SWIR wavelengths, the current technologies, and associated issues, and describes the potential of Ge-on-Si SPADs as a single-photon detector technology for this wavelength region. The working principles, fabrication and characterisation processes of such devices are subsequently detailed. We review the research in these single-photon detectors and detail the state-of-the-art performance. Finally, the challenges and future opportunities offered by Ge-on-Si SPAD detectors are discussed.

1. Introduction

1.1. Single-photon detector applications
One of the most significant discoveries in modern day science is that electromagnetic radiation is composed of the individual energy packets now known as photons [1]. The idea of the photon was originally postulated by both Planck [2, 3] and Einstein [4] in the very early 20th century, to explain black-body radiation and the photoelectric effect respectively, but the word 'photon' itself was not used until the mid-1920s [5]. Since then, the interest and knowledge relating to both single-photon sources and detectors has grown rapidly, where single-photon detectors are defined by their extreme high sensitivity allowing for the ability to detect these individual energy quanta [6, 7]. In recent decades there has been a dramatic increase in the number of applications which utilise single-photons, thus, this vast number of applications drives the development of high performance single-photon detectors. Initially, biophotonics was perhaps the primary area of interest motivating single-photon detector development, particularly in the visible and near infrared spectral regions [8–11], but arguably the field that has been the major driver over the past 20 years, is that of quantum information [6]. Quantum information applications involve the manipulation and measurement of discrete quantum objects, typically photons, which can be used to encode and communicate information. As a result, this field promises to have a huge technological impact in the coming years. To date, the most developed area of quantum information science is quantum key distribution (QKD) [12–14], which provides a means for verifiably secure communication and prototype QKD systems are now commercially available [15, 16]. More long-term quantum technology applications include the ambitious and challenging goal of linear optical quantum computing (LOQC) which could vastly increase the efficiency of many processing and computational tasks currently carried out by classical methods [6, 17, 18]. This type of advanced quantum
information application places rigorous and demanding requirements on the optical components in the system, including the issues of integrating multiple single-photon detectors, as a result this application still has a number of technological challenges to overcome [17, 19]. This has led to a considerable amount of effort being invested in the development of single-photon detectors and other quantum components in order to make these quantum information technologies a reality.

There are numerous other disciplines besides biophotonics and quantum information science which employ single-photon detection. Examples include, but are not limited to; optical time domain reflectometry for optical fibre characterisation [20]; environmental monitoring and other remote sensing applications [21], light detection and ranging (LiDAR) and 3D imaging including long-range and imaging in turbid media [22–28]. The high detection sensitivity and low temporal jitter (∼100 ps) has meant that single-photon detection is an excellent candidate technology for LiDAR imaging in extreme environments, such as over very long distances [22], through obscurants (high loss, highly scattering environments) [26], or in complex environments composed of multiple scattering surfaces [24].

1.2. Single-photon detection in the short-wave infrared
In many applications it is particularly beneficial to operate at wavelengths in the short-wave infrared (SWIR), ∼1–3 µm, for several key reasons. Firstly, the capability to operate at the low-loss optical fibre telecommunications wavelengths of 1310 and 1550 nm is fundamental for quantum communications and other fibre-based applications which rely on high-speed and ultra-sensitive SWIR single-photon avalanche diode (SPAD) detectors. A second important consideration is that in many free-space applications, the source power is restricted by laser eye-safety thresholds. The eye-safety threshold is considerably higher in the SWIR than the visible. This allows for the use of much higher optical power sources that increase the signal-to-noise in the system, resulting in increased attainable maximum ranges and improved depth resolution [29]. A third advantage of operating in the SWIR is that atmospheric transmission is enhanced because of the decrease in scattering from particles in the atmosphere, thus it is possible to image more clearly through certain obscurants such as smoke, haze and dust [30–32]. Finally, the solar background in the SWIR is significantly reduced [33], effectively decreasing the background noise level in any outdoor system utilising optical detection. These advantages for operating in the SWIR mean that LiDAR and 3D imaging applications are particularly benefited by operating in this spectral range.

1.2.1. SWIR single-photon detection methods
Photomultiplier tubes (PMTs) were the first devices able to detect light at single-photon levels and so initially dominated this market. When a photon is incident on a photocathode, a photoelectron is generated via the photovoltaic effect and then accelerated in a high electric-field to a positively biased dynode within a vacuum chamber. Secondary electrons are emitted at the dynode after the collision with the primary photoelectron, resulting in electron gain. These electrons are then accelerated to the next positively biased dynode and then to other dynodes in the chain, resulting in higher gain with each dynode. Whilst PMTs can operate over a wide wavelength range, high voltages of the order of 1000 V are required to accelerate the electrons sufficiently to achieve the gain required for a detectable photocurrent. Furthermore, they tend to be large in size and composed of mechanically fragile material which is impractical for many applications outside of a controlled laboratory environment, where compact and robustness become more critical factors. In the SWIR there are few available PMTs, one example is commercially available from Hamamatsu Photonics K.K. with model number R5509-73 [34]. The manufacturers quote a typical temporal jitter of the order of 1 ns, and relatively low quantum efficiency ∼2% in the spectral range 900–1600 nm (taken from R5509-73 datasheet [34]). Such PMTs are suitable only for controlled laboratory conditions [7].

Many alternative single-photon detection methods have also been extensively developed including Microchannel plates, superconducting nanowire single-photon detectors (SNSPDs), superconducting transition edge sensors and SPAD detectors. There also exist various other less technologically mature single-photon detection methods including, for example, quantum-dot, semiconductor defect [6] and superconducting tunnel junction [35] based systems. Each method has inherent advantages and disadvantages and the detector of choice is generally dependent on the application in question. SPADs have become one of the most commonly used single-photon detectors on the market. They are composed of doped semiconductor material and the bandgap of the material chosen governs what photon energies can be absorbed and hence the operational design wavelength of the device. Thanks to the impressive ionisation properties of silicon (Si), the development of Si based SPADs meant that they quickly replaced PMTs as the detector of choice for detection in the visible spectral region ∼400–1000 nm, enjoying great success for a range of applications including the testing of quantum foundations [36], quantum communications [37] and 3D imaging [25, 26, 28]. A major advantage of Si SPADs is that they are ideally suited for integration with standard Si complementary metal-oxide-semiconductor (CMOS) processes allowing for the fabrication
of large detector arrays. The result is a manufacturable, low-cost, highly-sensitive detector technology which is useful for a range of time-of-flight (ToF) applications for efficient operation at wavelengths <1 μm. Si-based SPADs, however, have decreased performance at wavelengths approaching 1 μm, hence alternative semiconductor materials are required for the detection of photons wavelengths beyond this. The absorption coefficient of geranium (Ge) extends much greater than 1 μm into the SWIR, therefore all-Ge homojunction avalanche photodiodes (APDs) operating in the Geiger mode were explored in early laboratory-based experiments [38, 39]. The main issue with these devices, however, is that the direct bandgap of Ge is quite narrow, ∼0.8 eV at room temperature. As a result, considerable band-to-band tunnelling occurs, leading to high detector noise which severely limit the sensitivity attainable by these devices. To mitigate this high detector noise issue, these devices are commonly operated at low temperatures, <100 K, and as a result the detector efficiency at longer wavelengths approaching 1550 nm is severely restricted due to the reduced absorption coefficient at such low temperatures [39, 40].

Currently, the most commonly used single-photon detectors operating in the SWIR are SNSPDs and InGaAs/InP based SPADs. SNSPDs can operate through the spectral range spanning the visible to the mid-infrared offering impressive detection efficiency, low noise, high-speed operation, as well as high count rate operation [6, 41–43]. In order to achieve such performance levels, these devices are operated at temperatures in the range 1.5 K–4 K, which both increases the cost of the system and can make them unsuitable for some applications, particularly those requiring operation on mobile platforms. InGaAs/InP SPADs tend to be the detector of choice in the SWIR as they operate within the much more easily attainable Peltier cooled temperature range from 220 K to 255 K being beneficial as the size, weight, power consumption and associated costs are considerably lower than those associated with the cryogenic operation. As a result, InGaAs/InP SPADs have been extensively developed to the point that detection efficiencies of ∼15%–50% in the SWIR have been reported, with low noise and jitter [44–48]. This has, in turn, facilitated long distance QKD [14] and LiDAR [49] application demonstrations as well as high-performance imaging array cameras for single-photon depth imaging applications [50]. Despite the success of InGaAs/InP based SPADs, one major drawback is that the maximum attainable count rates are severely restricted by afterpulsing effects (these effects are explained in detail in the following section), and although progress continues to be made [46, 51, 52], mitigating this issue remains challenging. Furthermore, although lower cost than SNSPDs, InGaAs/InP technology remains relatively expensive when compared to the widely available Si CMOS SPADs used in the visible spectral region.

1.3. Introduction to Ge-on-Si SPADs
There is a distinct lack of practical single-photon detectors which operate at wavelengths >1 μm with the ideal combination of high efficiency, low noise, low afterpulsing effects and near room temperature operation. In order to address many of the drawbacks, a useful alternative to these existing detection technologies is to utilise Ge as an absorber in tandem with a Si multiplication layer. The bandgap of Ge provides efficient absorption at wavelengths throughout the visible and infrared to a cut-off wavelength of approximately 1600 nm at room temperature. The development of dual temperature growth techniques [53] have provided a method of fabricating high-quality Ge layers grown directly on Si thus benefitting from low-cost and well-established integrated Si wafer technology, and providing the potential for integration onto Si-based photonic and electronic layers. As a result of this, they are highly suited for applications such as QKD and LOQC, where detector integration has game-changing technical and cost advantages. These detectors also hold potential for operating in the range of temperatures accessible by Peltier coolers. The purpose of this paper is to review the development of germanium-on-silicon (Ge-on-Si) photodetector technology with a specific focus on Ge-on-Si SPAD detector performance for single-photon applications in the SWIR. The next section discusses the background of Ge-on-Si photodetectors including the work done previously and the methods of characterisation. In section 3 a waveguide-coupled geometry device is discussed, followed by the state-of-the-art technology performance of planar geometry Ge-on-Si SPADs in section 4. Section 5 focuses on the applications of these SPAD detectors and hence details the results of a planar Ge-on-Si SPAD LiDAR investigation. Finally, the perspectives and future direction of the field are discussed justifying the continued research and development of Ge-on-Si SPAD detectors.

2. Background of Ge-on-Si SPADs

2.1. Fundamentals of Ge-on-Si SPADs
This section will discuss the background of Ge-on-Si SPAD detectors in addition to detailing the methods which are used to characterise the operating parameters of such devices. It is important first, to understand the working principle of photodiode detectors. The basis of any photodiode is a p–n or, often, a p–i–n junction. When operated at high voltages near but below the voltage breakdown point (V_{BD}), low intensity
Figure 1. (a) A simplified cross-sectional schematic diagram of an InGaAs/InP SPAD showing the multiplication, charge sheet, graded, absorption, buffer and substrate layers ($i = \text{undoped}$). (b) The energy band diagram of an InGaAs/InP SPAD under reverse bias conditions.

Incident light can initiate a readily detectable avalanche current via impact ionisation. An incident photon, with an energy greater than that of the semiconductor material bandgap, is absorbed resulting in the generation of an electron–hole pair (EHP), the generated free carriers are accelerated towards opposite sides of the junction by the applied reverse biased electric-field. These primary carriers gain sufficient kinetic energy that they can impact ionise a bound electron in the valence band of the material and promote it to the conduction band, leaving behind a free hole in the valence band. These secondary carriers are also accelerated leading to a chain of impact ionisation events creating more EHPs, which leads to further carrier multiplication, hence significant carrier gain and a current multiplication factor. Detectors using this principle of operation are called APDs which are operated at voltage levels below $V_{BD}$ achieving typical multiplication factors generally of the order of 10 when operated with high bandwidths. Although APDs can detect low intensity level light, they are typically unable to provide sufficient gain to detect a single incident photon. In order to achieve single photon sensitivity, the diode must be biased above the avalanche breakdown threshold ($V_{BD}$). Here, the electric-field is sufficiently high that a single incident photon creating an EHP is able to initiate a self-sustaining avalanche process effectively leading to an infinite multiplication factor and with timing precision typically in the picosecond domain. The resulting avalanche current will flow continuously until it is quenched by reducing the diode’s applied voltage to a level below $V_{BD}$. This mode of operation is often referred to as Geiger mode.

Indium gallium arsenide (InGaAs) is highly suited to the absorption of SWIR photons. With a direct bandgap of $\sim0.75$ eV (dependent on the In content), similar to Ge, high noise levels caused by band to band tunnelling are typically observed. To avoid such high electric-fields in the narrow-gap absorber, InGaAs is commonly incorporated with an InP avalanche region resulting in a separate absorption, charge and multiplication (SACM) structure. This SACM design has been widely employed in InGaAs/InP APDs and SPADs and the general structure is shown in the schematic diagram in figure 1(a) alongside the energy band diagram in figure 1(b).

When reverse biased above breakdown, a photon absorbed in the InGaAs layer creates an EHP. The electron is accelerated towards the back $n$-contact as the hole is swept into the InP multiplication region (MR) where it undergoes impact ionisation which leads to the self-sustaining avalanche required to generate macroscopic detectable current. The charge sheet layer is $n$-doped included to control the electric-field from the InP MR to the InGaAs absorption layer. One disadvantage associated with InGaAs/InP based devices is that the holes must overcome a large energy barrier caused by the discontinuity in the valence band between the narrow gap InGaAs absorption layer and the wider gap InP MR. This valence band discontinuity is highlighted in figure 1(b). For this reason, a graded InGaAsP layer which has an intermediate bandgap level,
Figure 2. A simplified cross-sectional schematic diagram of a Ge-on-Si SPAD showing the Ge absorption and Si multiplication layers. (b) The energy band diagram of an InGaAs/InP SPAD under reverse bias conditions.

is inserted between these layers. This layer acts as a step level which assists the transfer of holes to increase the efficiency of hole transit to the multiplication layer. Employing the same type of SACM design, a Ge absorber layer can be incorporated with an Si avalanche region. A schematic diagram and the electronic band structure of such a Ge-on-Si SACM photodetector is displayed in figures 2(a) and (b) respectively.

The measurement of single photons is achieved through the following process: an incident photon is absorbed in the intrinsic (i) Ge absorption layer and creates an EHP. With an applied reverse bias above \( V_{BD} \), the electron is accelerated towards the Ge–Si interface and enters the Si MR. The Si layer is held at a sufficiently high electric-field that the electron can undergo impact ionisation, creating a further EHP. Further impact ionisation events of these multiple carriers can create the self-sustaining, detectable avalanche current. This current can be quenched by reducing the applied voltage to a level below \( V_{BD} \). Compared to the situation in InGaAs/InP SPADs shown in figure 1(b), the photogenerated carrier (an electron in this case) created by absorption in the Ge does not encounter an energy barrier in drifting into the MR, allowing for high efficiency transfer of electrons without the requirement for an additional grading layer. Whilst this is a clear advantage for Ge-on-Si SPADs for detection in the SWIR compared to InGaAs/InP SPADs, it is however, challenging to grow high-quality Ge-on-Si material epitaxially, compared to the more mature InGaAs/InP material system. The challenge for growing Ge-on-Si of sufficiently high-quality for single-photon detection is the 4.2% lattice mismatch which potentially can cause three major problems: a high threading dislocation density in the Ge layer; a high density of misfit dislocations at the Si/Ge interface; and high levels of surface roughness. Both types of dislocations hinder the performance of any resulting detector as dark noise is increased as a result of defect states introduced at the dislocations, whereas high surface roughness reduces the compatibility with Si electronics due to the planar processing required by CMOS technology [53, 54].

In 2002 Loudon et al [55] successfully utilised the principle of separate absorption and avalanche regions to demonstrate a SPAD in which the absorption of 1210 nm wavelength photons took place in Si\(_{0.7}\)Ge\(_{0.3}\)/Si multiple quantum well (MQW) structure and the multiplication occurred in an adjacent Si avalanche region. The schematic diagram of this structure is displayed in figure 3.

It can be seen in figure 3 that the quantum well strained superlattice absorbing layer in this device was necessarily thin, \( \sim 300 \) nm, owing to the strain placing limitations on the critical thickness for a defect-free layer. These material constraints led to a low detection efficiency, less than 0.1%, at \( \lambda = 1210 \) nm. Despite the limited performance of this initial device, this work successfully demonstrated enhanced single-photon detection efficiency (SPDE) at wavelengths >1000 nm by using Ge in a Si based SPAD. Subsequently, the development of dual temperature growth techniques has facilitated the growth of high-quality epitaxial Ge layers on Si substrates with much greater thickness than the critical thickness due to strain [53]. In addition to single-photon detectors, these growth techniques have also made possible the fabrication of both CMOS compatible SACM Ge based waveguide photodiodes [56] and high performance APDs [57, 58].
2.2. Characterisation methods

Generally, the performance of single-photon detectors can be fully quantified in terms of several key parameters. Firstly, SPDE is defined as the probability that a single-photon incident on the detector will result in a registered output count, assuming that no more than one photon is incident at a time, and that the detector has fully reset before the next single photon is incident. Detector noise in a SPAD is characterised as the dark count rate (DCR), which is the rate of events recorded when no light is incident on the device. This rate should be normalised to include any detector dead time. These dark events typically arise from thermally generated EHPs within the device active volume and will critically depend on the device dimensions and operating temperature. The spectral range is the wavelength range over which a detector is operational, and in a SPAD this is generally governed by the constituent semiconductor material which absorbs the incident photons. The noise-equivalent-power (NEP) is a measure of the sensitivity, indicating the optical power required to achieve unity signal to noise. NEP provides a useful figure of merit to make valid comparisons between different detector types. Other potentially critical considerations include temporal jitter, dead time and afterpulsing, which are discussed in greater detail later in this section.

The measurements recorded from planar geometry Ge-on-Si SPADs presented in the following section have been taken using the time-correlated single-photon counting (TCSPC) technique. This technique involves the measurement of electrical pulses generated by the SPAD which may or may not be due to the detection of a photon. A schematic diagram of this experimental set-up is shown in figure 4.

![Figure 4. A cross-sectional schematic diagram of the SPAD design in which the absorption occurred in the SiGe/Si MQW structure and the multiplication occurred in the separate Si avalanche region. The p-type spike doping region is effectively a charge sheet, the purpose is, to locate the high-field region in the avalanche layer. Reprinted with permission from [55] © The Optical Society.](image)

The SPAD was mounted in a variable temperature liquid nitrogen Oxford Instruments cryostat (OptistatDN) which allows for temperature control in the range of 78 K–295 K. Photons were supplied to the SPAD by a supercontinuum pulsed laser (NKT Photonics, SuperK EXTREME EXW-6) which can emit picosecond laser pulses at pre-selected wavelengths in the range 1150–2000 nm with a pulse width of <50 ps [49]. The output from the laser was directed into the 50:50 fibre beam-splitter (BS) by way of a single mode optical fibre. One output from the BS was directed through a calibrated optical attenuator, whilst the other output was directed to a calibrated power meter to allow continuous measurement of the laser power level. After attenuation, an objective lens (not shown) focused the collimated free-space output onto the SPAD active area. A pellicle BS (not shown) was inserted before the focusing lens allowing an imaging channel directed onto a SWIR InGaAs camera (Hamamatsu, C10633) to be used to accurately focus the laser beam onto the SPAD active area. A pellicle BS (not shown) was inserted before the focusing lens allowing an imaging channel directed onto a SWIR InGaAs camera (Hamamatsu, C10633) to be used to accurately focus the laser beam onto the SPAD active area. This set-up allowed the attenuation to be varied to ensure that a calibrated value of less than 0.01 photons per pulse was incident upon the SPAD hence guaranteeing that the statistical probability of >1 photon per pulse striking the SPAD was negligible.

A gated quenching approach was used. The DC voltage supply applied a fixed constant reverse voltage at a level just below $V_{BD}$ to the SPAD. The pulse generator applied a gated voltage to the SPAD which increased the total reverse bias above $V_{BD}$ for a periodic timing gate of ~50 ns. This gated signal was applied in synchronisation with the optically attenuated pulses from the laser, thus the SPAD operated in the Geiger...
mode and a temporal measurement was formed over many pulses of the source. At all times outside the gating pulses, the detector remained below $V_{BD}$. The gated voltage and the constant DC voltage level were combined using a Bias Tee (Tektronix 5530) and the output was subsequently connected to the anode of the SPAD. The master clock controlled the timing of the laser pulse, the gated voltage and also the start signal for the photon counting data acquisition unit, which in these experiments was an Edinburgh Instruments TCC900 photon counting card (PCC). The manufacturers of this PCC estimate an intrinsic timing jitter of the order of $\sim 10$ ps. The SPAD cathode was connected to the PCC’s stop signal and the measurements were taken at a low repetition rate of 10 kHz to reduce the likelihood of afterpulsing. The PCC measures the stream of electrical pulses output by the SPAD which are generated by the incident photons. Each measurement is made with respect to its arrival time relative to the start signal set by the master clock. The events detected by the PCC are therefore recorded in the form of a histogram of photon counts as a function of arrival time from which the SPAD performance parameters can then be extracted. The SPDE is extracted from the histogram with the laser incident on the SPAD whereas the DCR measurements are recorded by measuring the output with no laser. The reliability and repeatability of the time taken for a photon to produce an avalanche and subsequently a detectable output current pulse is vital to SPAD applications in which precise timing is essential, such as LiDAR. The variance of this risetime of the detected event measured over many repeated events is called the timing jitter—it is generally quantified as the full-width half-maximum (FWHM) of the timing histogram recorded. Therefore, the timing bin width was reduced to allow for higher resolution measurements. The NEP parameter gives an indication of the sensitivity of the device and is calculated using equation (1):

$$\text{NEP} = \frac{h \nu}{\text{SPDE}} \sqrt{2\text{DCR}}$$

where $h$ is Planck’s constant and $\nu$ is the frequency of the incident light. As the calculation of the NEP includes both the DCR and the SPDE, it provides a useful and valid means to compare the performance of different types of single-photon detectors which operate at different temperatures and bias levels. The dead time, $\tau$, is the time period following the absorption of a photon, during which the detector cannot reliably register a second photon. Generally, $\tau$ depends on the detector type in question and it is difficult to quantify because often the measured value of $\tau$ is that of the biasing circuit and/or counting electronics, rather than the detector itself. A long dead time can significantly limit the maximum achievable count rates and speed of a SPAD.

Afterpulsing is a phenomenon which occurs after a photon or dark event initiates a large avalanche current to flow through these small area devices. During the avalanche carriers can become trapped in deep energy levels and subsequently released at a later time. If these traps are not fully depopulated before the next application of the gated voltage, they can contribute to an increase in the average dark count level within that gated time period. In addition, the release of trapped carriers is temperature dependent, which further complicates the optimum operation parameters for a SPAD. As the operating temperature is reduced. The DCR is lowered, however, the trap lifetime will lengthen thus increasing the deleterious effects of afterpulsing. At higher temperatures where the DCR is greater, often the effects of afterpulsing are much less evident. Therefore, a compromise must be found between operating the device optimally to maximise the SPDE and count rate and also reduce the thermally-induced DCR and afterpulsing effects. Characterisation of afterpulsing effects can be performed using the time-correlated carrier counting method (TCCCM) which is described in the schematic diagram in figure 5.

The experimental set-up is very similar to that in figure 4, initially the SPAD is reverse biased at a DC ($V_{DC}$) level below $V_{BD}$. A voltage gate, $V_G$, is then applied to the SPAD to take it above $V_{BD}$ into the Geiger
mode exactly synchronous with the incident 1310 nm laser pulse. As can be seen in figure 5, a second gate with the same amplitude $V_G$ is then applied at a later time of $\Delta t$, no light is incident on the SPAD during this second voltage gate. The incident laser pulse at the first gate guarantees that the device breaks down at the beginning of each measurement period for the full gate length and also that the charge flowing is the same during each measurement. Only dark counts are measured over the second gate time period, and so this is a measurement of the trapped states which are released at $\Delta t$ and therefore contribute to the DCR occurring at that time. The delay, $\Delta t$, can be varied within the measurement period before the next laser pulse is incident. From these results the afterpulsing probability can be deduced as a function of $\Delta t$. This can be repeated at different temperatures to analyse the temperature dependence of the afterpulsing lifetime. One common afterpulsing mitigation strategy is to use a sufficiently long time between gating events to allow the traps to become depopulated. Such a strategy, however, results in a long dead time, which can place severe limitations on the count rates possible.

2.3. Mesa geometry Ge-on-Si SPAD

The first Geiger mode operation of Ge-on-Si APDs was reported in the mesa geometry structure by Lu et al [60]. This was a 30 $\mu$m diameter area device which was operated in the gated mode described above and demonstrated single-photon detection at a wavelength of 1310 nm. The maximum SPDE observed was 14% at a temperature of 200 K and the DCRs reported were of the order of several hundred Mcps. It should be noted, however, that during this investigation the light pulses incident on the detector contained one photon per pulse. This has likely led to an overestimation of the calculated SPDE value because at this level the statistical probability of >1 photon per pulse striking the SPAD is no longer negligible. The jitter was also measured and a minimum value of 195 ps was recorded at 200 K. In addition, afterpulsing was characterised by investigating the DCR versus frequency in the range from 1 kHz to 1 MHz, with the authors concluding that the afterpulsing in this Ge-on-Si SPAD is considerably less compared to that for InGaAs/InP based SPADs operating under similar conditions, owing to the very high quality Si constituting the multiplication layer. Following on from the demonstration by Lu et al [60], in 2013, Warburton et al reported a 25 $\mu$m diameter mesa geometry Ge-on-Si SPAD [61] which had a lower maximum SPDE of 4% at 1310 nm but improved DCR with a maximum of 6 Mcps measured at a lower temperature of 100 K. The authors also reported the detection of single-photons at a wavelength of 1550 nm in this paper—the longest wavelength detected for a Si-based device. The geometry of this mesa device is shown in figure 6.

Despite some improvement in the DCR reported in [61] compared to [60], DCRs in the Mcps region tend to be prohibitively high resulting in low sensitivity, leading to an NEP of $1 \times 10^{-14}$ WHz$^{-1/2}$ for this device [61]. In comparison, a commercial InGaAs/InP SPAD has an NEP value of $5 \times 10^{-17}$ WHz$^{-1/2}$ at a temperature of $\sim$230 K and wavelength of 1550 nm (calculated from Micro Photon Devices Sr.i PDM-IR 25 $\mu$m diameter InGaAs/InP SPAD datasheet [62]). In a mesa geometry Ge-on-Si SPAD a high portion of the total DCR is likely due to the exposed sidewalls of these devices which can be seen in figure 6. There are a high density of interface traps at the exposed sidewalls which make a significant contribution to the DCRs. As result, many efforts have been made to investigate different passivation techniques such as SiO$_2$, GeO$_2$, and atomic layer deposition with the aim of mitigating the large contribution from the sidewalls. Optimal passivation for the two different materials, Ge and Si however, remains very challenging.

3. Waveguide-coupled Ge-on-Si SPAD detector

One method to mitigate the high DCR associated with the mesa type geometry employed by Lu et al [60] and Warburton et al [61] is to use a waveguide-coupled arrangement. This was reported by Martinez et al [63]. Previous to this the same research group and others had reported APD operation of waveguide based devices [58]. When operated in Geiger mode the authors demonstrated a waveguide-coupled Ge-on-Si SPAD at a
temperature of 80 K exhibiting an SPDE of 5.27\% at $\lambda = 1310$ nm and a DCR of 534 kcps [63]. The geometry of this waveguide device is presented in figure 7. In this case, light is coupled from the waveguide into the high index Ge absorption region. When biased above $V_{BD}$, the photogenerated electrons flow into the Si MR and are accelerated, thus undergoing impact ionisation to generate a detectable output current. With reference to figure 7, it can be seen that the $p$-charge Si layer extends well beyond the edge boundary of the Ge absorber, this prevents the high electric-field in the Si MR from penetrating significantly into the absorber region. This reduces the number of thermally generated EHPs in the Ge absorber, and so the authors propose that this has effectively resulted in the low DCR performance shown by this device.

4. Planar Ge-on-Si SPADs

An alternative method to mitigate the high DCRs associated with the exposed sidewall issue in a mesa type SPAD is to employ a planar geometry. The effect of this is that the high-field region is confined to the centre of the detector, leaving a much lower electric-field at the exposed sidewall regions and thus eliminating the requirement for optimal passivation of both the absorption and multiplication layers. This type of planar geometry design has previously been employed in the design and fabrication of various SPAD detectors, including InGaAs/InP based SACM SPADs [44]. In the fabrication of these SPADs the $p$–$n$ junction is generally made by directly diffusing the $p$-type dopant zinc into the top InP multiplication layer, using spatial patterning to define the effective active area of the device. This results in a planar structure, using one epitaxial growth, without the need for any etch and regrowth techniques. A similar design cannot be implemented in Ge-on-Si SPADs as the Si multiplication layer cannot be grown directly onto a Ge absorption layer because of the lattice mismatch between the Si and Ge discussed earlier. Epitaxial Ge, however, can be grown on Si using dual temperature growth approaches. By implanting a local buried charge sheet layer over a selective area in the Si layer which forms the multiplication layer, the high-field region can be located in the centre of the device. After growth of the Ge, by use of a top contact layer with an area slightly smaller than the charge sheet layer, a planar geometry SPAD can be created. Unlike InGaAs/InP SPADs, this necessitates a two
growth strategy: firstly, growth of the Si multiplication layer on an $n^+$ Si substrate; then selective ion implantation to create the charge sheet; and then growth of the Ge absorption layer. The effect of planarizing a normal-incidence Ge-on-Si SPAD is presented in figure 8 in which the cross-sectional diagrams of both a mesa and planar device are presented in (a) and (b) respectively. The corresponding simulated electric-field profiles at a bias above $V_{BD}$ for the mesa and planar devices is shown in (c) and (d) respectively.

It can be clearly observed by comparing figures 8(c) and (d), that above $V_{BD}$, the planar geometry device removes the electric-field hot-spots produced in the mesa geometry which lead to premature avalanche breakdown at the edge of the device, reducing the overall effective detection efficiency of the SPAD. The peak electric-field of the planar geometry device is in the centre of the device where the photons are incident. In addition, the planar geometry significantly reduces the electric-field at the sidewalls, hence states at the sidewalls are less likely to contribute to the multiplied dark current and DCR.

4.1. Fabrication
The specific techniques and processes used to fabricate SPADs are very much dependent on the type of SPAD and its constituent materials. Despite the challenges in fabricating planar geometry Ge-on-Si SPADs compared to SPADS in other material systems, high performance Ge-on-Si planar devices have been achieved [59, 64, 65]. The fabrication process undertaken to produce these SPADs will now be described in detail. The SPADs were grown on 150 mm diameter $n^{++}$-doped Si (001) substrates. First, a layer of nominally undoped Si was grown epitaxially by means of a commercial reduced pressure chemical vapour deposition system, to define the MR. Following this, photolithography defined regions for the charge sheet layers of each SPAD device. Implantation of the charge sheets was carried out using boron acceptors at an energy of 10 keV which were then activated with a 30 s anneal at 950 °C using a rapid thermal annealer. This charge sheet layer controls the relative electric-field profiles in the Ge absorber and Si MRs; therefore, various doping levels were implanted in an attempt to identify the optimal parameters. The wafers were then cleaned, which was followed by the growth of a layer of undoped Ge absorber and a 50 nm $p^{++}$ Ge cap for the top Ohmic contact. Following this, a trench etch through the Ge layer was carried out to electrically isolate every SPAD from each another. This electrical isolation is necessary because of the conductive path which is a consequence of the background doping level of the Ge layer. Finally, metal contacts, GeO$_2$ passivation, antireflection coatings and bond-pads were deposited. A microscope image of the top view of a typical planar Ge-on-Si SPAD is shown in figure 9 and the various layers which have been discussed are highlighted in the schematic diagram of the cross section of such a device in figure 8(b).

The next section presents the state-of-the-art performance of planar Ge-on-Si SPADs which have been developed using the fabrication methods presented in this section. Further information relating to the fabrication processes involved for producing Ge-on-Si SPADs can be found in [59, 64, 66, 67].

4.2. Large area Ge-on-Si SPAD detector performance
Using the design processes described in the section 4.1, the fabrication and performance of a 100 µm diameter charge sheet SACM Ge-on-Si SPAD was reported by Vines et al [59]. The device was fabricated with a 1.5 µm layer of nominally undoped Si for the MR and a 1 µm thick Ge absorption layer. The SPAD performance was characterised using the techniques detailed in section 2.2. The PCC acquisition used a
timing bin width of 195 ps for SPDE and DCR measurements. The SPDE at $\lambda = 1310$ nm and DCR as a function of excess bias at a temperature of 125 K demonstrated by this SPAD is shown in figure 10.

An impressive maximum SPDE of 38% was measured, with the DCR reaching a maximum of $\sim 2$ Mcps. This is a considerable improvement over both the Lu et al [60] and Warburton et al [61] mesa geometry normal incidence SPADs. The NEP is also significantly improved when compared to these devices; a minimum value of $1.9 \times 10^{-16}$ WHz$^{-1/2}$ for this 100 $\mu$m diameter device was obtained at a temperature of 78 K. This is 50 fold reduction compared to the mesa geometry device described by Warburton et al [61], despite the much greater device area, and this remarkable improvement is owed to both sizeable increases in the SPDE and a significant reduction in DCR. The timing bin width was reduced to 19.5 ps for measuring the timing jitter to allow for higher resolution measurements. The device recorded a minimum jitter value of 310 ps at an excess bias of 5.5% and temperature of 78 K.

The afterpulsing effects were also characterised by Vines et al [59]. The results obtained were compared to a commercial InGaAs/InP SPAD manufactured by Princeton Lightwave Inc. with model number PGA-246-25, this device has an effective optical diameter of 25 $\mu$m. To provide a valid comparison, the two devices were operated at the same temperature of 125 K, with the excess biases carefully chosen to result in both detectors operating with an SPDE of 17%. The results are shown in figure 11.

It is clear from figure 11 that afterpulsing effects are significantly less at all given delay times for the Ge-on-Si SPAD compared to the InGaAs/InP SPAD. Whilst this is a promising result for the longer-term use of Ge-on-Si SPADs, afterpulsing will require considerable further analysis under a wide range of conditions, particularly at higher temperatures in the range of Peltier cooling. Previously, it has been found that the afterpulsing observed from InGaAs/InP SPADs is a consequence of deep level trap states in the InP multiplication layer [68–70]. Hence, it could be anticipated that the high-quality Si avalanche region in Ge-on-Si SPADs will contain a much lower trap density and may experience much less afterpulsing events than an InGaAs/InP SPAD operated under similar conditions.
Figure 11. The afterpulsing probability as a function of gate delay time for a 25 µm diameter commercial InGaAs/InP based SPAD and a 100 µm diameter planar Ge-on-Si SPAD. Both devices were operating at an excess bias which resulted in an SPDE of 17% at an operating temperature of 125 K. Reproduced from [59]. CC BY 4.0.

Figure 12. The SPDE of a 26 µm diameter planar Ge-on-Si SPAD as a function of temperature and excess bias at a wavelength of 1310 nm [64].

4.3. Small area Ge-on-Si SPAD detector performance
The next stage in the development of planar Ge-on-Si SPADs was the fabrication of smaller area 50 and 26 µm diameter devices with the aim of reducing the DCR and thus improving the efficiency and sensitivity of these SPADs. A similar fabrication process was utilised as described in section 4.1 with a 1.5 µm thick Si multiplication layer and 1 µm Ge absorption layer. However, in this case trenches were established between the SPAD devices by photolithography and a fluorine inductively coupled plasma reactive ion etch (ICP-RIE) was used to etch through the Ge epilayers in order to obtain electrical isolation [64]. Planarization was performed with hydrogen silsesquioxane and plasma enhanced chemical vapour deposition (PECVD). Ti/Al was then evaporated and annealed to construct the back contact to the n++ Si substrate. Photolithography was then used to create holes to the p++ cap, and Ti/Al was deposited for the top contact. Lastly, a SiN anti-reflection coating was applied using PECVD on the surface to minimise reflection losses.

Characterisation of the SPDE, DCR and jitter for a 26 µm diameter device was performed at various temperatures between 78 K and 175 K using the TCSPC method detailed in section 2.2. These results were performed at a wavelength of 1310 nm to provide a valid comparison with the 100 µm diameter devices previously discussed. Again, a PCC timing bin width of 195 ps was used for the SPDE and DCR measurements, this was reduced to 19.5 ps for jitter measuring the jitter. There were 12 × 26 µm diameter SPAD detectors on the chip, six of the devices tested demonstrated very similar performance, the best performing SPAD results are presented here. Figures 12 and 13 present the SPDE and DCR respectively as a function of excess bias at various temperatures.

As can be seen in figure 11, the maximum SPDE measured for this device was 29.4% for an applied excess bias of 7.5% at a temperature of 125 K. This is a reduction in SPDE compared to the larger 100 µm diameter device previously presented which had a maximum value of 38% at 125 K. It is proposed, however, that the decrease may be due to non-uniform lateral efficiency profile in the device when integrated across the
∼10 µm diameter focused laser spot size used. Further investigation is required to confirm if this is indeed the cause, but the SPDE compares well to that of commercially available InGaAs/InP SPADs operating in the SWIR.

Figure 13 shows that the DCR for 125 K remains below 100 kcps at excess biases up to 6.6%—this is a considerable improvement over the 100 µm diameter device. To quantify this, the 100 µm device exhibited a DCR of ∼2 Mcps for an excess bias of 5.4% at 125 K, whereas for the same temperature and bias the 26 µm device recorded a DCR of ∼36.9 kcps. This represents an overall reduction in the DCR by a factor of ∼54 and, very importantly, an improvement in the DCR per unit area by a factor of ∼3.7, clearly demonstrating that both the use of a smaller active area and improved fabrication processes have significantly improved the performance of Ge-on-Si SPADs. As predicted by Fermi–Dirac statistics, the DCR rises with temperature, this is clear in the range of 100 K–175 K but is not true for the 78 K results which are higher than may be expected. It is possible that afterpulsing effects, which, as discussed previously are more prevalent at low temperatures, are a contributing factor to the higher than expected DCR. However, there may be a contribution due to partial freeze-out of dopants in the charge sheet resulting in a re-normalising of the electric-field profile. Further investigation is required to confirm and quantify the extent of this effect as accounting for this DCR increase at 78 K.

Figure 14 presents the jitter values, described as the FWHM of the measured timing histograms, as a function of excess bias and temperature. Generally, the jitter values are seen to reduce for increasing excess bias and are lower for the lowest temperatures. The minimum jitter recorded was 134 ± 10 ps at a temperature of 100 K and excess bias of 5.5%—the timing histogram for this measurement is shown in the inset of figure 14. This value represents a greater than two fold reduction in the temporal jitter compared to the value of 310 ps demonstrated by the 100 µm diameter device. It is expected that jitter will be minimised further with reducing device diameter, due to the restricted lateral extent of the avalanche build-up, as has been found to be the case in Si SPADs [71, 72]. This value of 134 ps, however, is higher than those routinely demonstrated for InGaAs/InP SPADs which have demonstrated jitter of <100 ps [45, 47, 52]. Improved packaging of the future generations of SPAD detectors will also allow for improved jitter performance.

The full results and further details relating to this 26 µm diameter planar Ge-on-Si SPAD characterisation can be found in [64]. In addition to the testing of the 26 µm diameter SPAD, the 50 µm diameter SPADs on this chip were also investigated for comparison. The SPDE, DCR and jitter for a 50 µm diameter SPAD were characterised at 125 K using the same experimental set-up. The full results can be found in [65] but the important conclusions from this investigation are summarised here. With the same applied excess bias and temperature conditions, the 50 µm diameter devices demonstrated DCR values around four times higher than 26 µm diameter devices. This indicates that the DCR is proportional to the device volume and surface contributions are negligible. This is an important finding in fully understanding the operation of these devices which is vital to improving the design and fabrication processes for future devices. For an applied excess bias of 6.5%, the 26 and 50 µm diameter devices exhibited jitters of 157 ± 10 ps and 210 ± 10 ps respectively, providing further evidence that smaller area devices demonstrate improved device jitter.
5. LiDAR using Ge-on-Si SPADs

The 100 μm diameter Ge-on-Si SPAD was investigated for use in a laboratory-based LiDAR imaging system [73]. The SPAD detector was operated in the electrically gated mode described in section 2.2, where it was taken into Geiger mode for 50 ns at the same expected time as that of an incident photon. The experimental set-up was similar to the TCSPC system described in figure 4 with the addition of the targets to be imaged, a schematic diagram of this set-up is shown in figure 15. The SPAD was again housed in the Oxford Instruments cryostat and all measurements were carried out at a temperature of 100 K. In this case, a laser wavelength of 1450 nm was used to illuminate the target and a laser repetition rate of 104 kHz was used, such a high value is allowed by the low afterpulsing rate of this SPAD. As shown in figure 15, after the optical attenuator, the laser was collimated by means of a reflective collimator arrangement and then focused onto the target by a lens, $L_1$. Photons reflected by the target were collected by $L_1$ and directed into the common transmit/receive channel. The BS directed part of the reflected light onto the SPAD, focused by $L_2$.

The TCSPC timing module recorded the time interval between the electrical gate applied coincident to the laser pulse and the single-photons detected by the SPAD. Recording measurements over a large number of laser pulses made it possible to obtain a statistically accurate timing histogram containing the ToF information, in spite of the occurrence of detector noise. To minimise the ambient light contributions to the noise the measurements were performed in a dark conditions. The ToF for the photon to return gives an estimate of the target range, and the rate of photon return estimates the target reflectivity. The distance between the laser and target was $\sim 0.4$ m. The target itself was moveable by means of motorised translation stages whilst the laser beam was kept stationary. Models of both a double decker bus and of a Mini Cooper car were chosen as targets in attempts to analyse the resolution depth imaging capabilities. The maximum average laser power used was 813 and 912 pW for the bus and car respectively. The depth and intensity information could be estimated from the obtained histograms using a cross-correlation algorithm method, this method will not be described further in this summary of the results, but more information can be found in [23, 26, 31]. The key results in this investigation are shown in figure 16. Figures 16(a) and (b) are visible camera photographs of the targets. The depth and intensity profiles recorded for the bus are shown in figures 16(c)–(e) respectively, for the car they are shown in (d) and (f).

Figures 16(c) and (d) show that millimetre depth resolution 3D profiles were achievable at the measurement distance of 0.4 m. Figures 16(e) and (f) clearly show that fine features on the side of the bus and hood of the car are recognisable. The measurements were performed using 300 ms per pixel acquisition times. In addition to these results the authors reported that the acquisition time could potentially be decreased by acquiring only limited information about the target, then reconstructing the data using the restoration of depth and intensity using total variation [74] and the manifold point process [75] algorithms. These algorithms demonstrated fairly good reconstruction of the targets even when up to 75% of the data was missing.

Further to these results, the authors used LiDAR modelling to extrapolate the laboratory performance results to estimate the parameters required in a long-range system and separately through an attenuating medium representative of an obscurant. They predicted that to construct a depth profile of a target at a 1 km range with a LiDAR system based on these type of Ge-on-Si SPADs, there would be a per-pixel acquisition
time of 1 ms for a wavelength of 1310 nm, and similarly 3 ms for 1450 nm. The longer acquisition times at longer wavelengths are due to the decreased SPDE which is a result of reduced absorption at these wavelengths in the Ge absorber. In addition, it was shown that <1 mW of average laser power would suffice for successful single-pixel imaging through a weakly attenuating medium. The average power required increased to the order of a few mW at higher attenuation levels. This is comparable with previous results from a single-photon SWIR LiDAR investigation using alternative InGaAs/InP SPADs, except that this prior work used detectors at higher operating temperatures and used Peltier cooling, as shown in [49]. Overall, the results presented in this paper clearly show the motivation for further research and development of Ge-on-Si SPAD detector arrays for useful applications such as eye-safe 3D depth imaging through obscurants at long range and video frame rates.

6. Summary and future perspectives

6.1. Summary
This paper has discussed the performance of Ge-on-Si SPADs for the detection of timed single-photons in the SWIR spectral region for a range of applications. This detector design could provide a useful alternative and overcome some of the disadvantages associated with the methods which currently dominate at these wavelengths such as InGaAs/InP based SPADs and SNSPDs. The earliest successful demonstrations of these detectors, in 2012–13, used mesa geometry arrangements, but the exposed sidewalls and resulting electric-field hotspots in such devices leads to high DCR values. In 2017, a waveguide-coupled arrangement was reported as a method to mitigate the high DCR associated with the mesa type geometry. This device successfully demonstrated lower DCR values with slight improvement in SPDE compared to the mesa devices. Planar geometry Ge-on-Si SPADs is reported as a method of mitigating this issue. A planar geometry was proposed in 2019 as an alternative method to mitigate the high DCR problem. This planar design has
Figure 17. The NEP as a function of temperature for a 26 µm diameter planar Ge-on-Si SPAD compared to; the 100 µm planar device, a 25 µm diameter mesa device, a Ge SPAD and a waveguide-coupled Ge-on-Si SPAD. All NEP values are calculated for an incident photon wavelength of 1310 nm. This graph has been reproduced from [64].

resulted in notable improvements in the DCR and SPDE performance. This is because the planar design both removes the electric-field hot-spots and maintains the high-field region to the centre of the device where photons are incident. As a result, surface states are less likely to contribute. Furthermore, Ge-on-Si afterpulsing effects were significantly less in a 100 µm diameter Ge-on-Si SPAD compared to an InGaAs/InP device under very similar operating conditions, thus demonstrating a potential further advantage of using Ge-on-Si. This large area device has also been used successfully in a laboratory-based 3D LiDAR imaging investigation, demonstrating the real potential of Ge-on-Si SPADs for use in practical imaging applications. More recently, smaller 26 µm diameter planar SPADs were fabricated and measured in attempts to further reduce the DCR and improve the overall SPAD performance. The NEP is a figure of merit that provides a method of comparing detector sensitivity, hence the minimum NEP values for both the large and small are planar SPADs have been calculated for varying temperature using the SPDE and DCR data in equation (1) and presented as a function of temperature in figure 17. This graph also includes the NEP values calculated for various other Ge-on-Si SPADs and an all-Ge homojunction SPAD for comparison.

The planar geometry devices clearly have superior sensitivity compared to the other devices. The minimum NEP reported on the 100 µm diameter planar Ge-on-Si SPAD was $1.9 \times 10^{-16}$ WHz$^{-1/2}$ at 80 K, this is an impressive improvement of $\sim$50 times smaller than that of the earlier mesa devices. The 26 µm diameter device demonstrated record low DCRs leading to a minimum NEP value of $7.7 \times 10^{-17}$ WHz$^{-1/2}$ at 100 K. Although not reaching the values of $5 \times 10^{-17}$ WHz$^{-1/2}$ shown by the mature technology of commercial InGaAs/InP SPADs in the SWIR at $\sim$230 K [62], the small area Ge-on-Si SPADs have shown considerable improvements compared to the mesa geometry type SPADs and the larger area planar SPAD, paving the way for further advancements in performance.

6.2. Future perspectives
The history and current performance of Ge-on-Si SPADs has been reviewed in this paper demonstrating that Ge-on-Si SPADs are a useful alternative to current technologies for the detection of single-photons in the SWIR. To date, planar geometry small area devices have shown the best performance—demonstrating good single-photon detection characteristics at temperatures below 175 K. There remains, however, several challenges to overcome and improvements to be made in this field.

6.2.1. Optimising the device structure
The logical next step in development is to further reduce the DCR and improve the SPDE. To achieve the former, perhaps the simplest method is to reduce the device active area further, whereas improvements in the latter could be made by increasing the Ge absorber layer thickness. Vines et al [59] estimated that only up to 50% of the incident 1310 nm photons are absorbed by a 1 µm thick Ge absorber layer throughout the temperature range of 78 K–125 K. Calculations using Beer–Lambert’s law predict that doubling the thickness to 2 µm could increase the absorption of photons to greater than 70%, with thicker layers increasing this further. Future device generations incorporating both smaller active areas and an increased absorption layer thickness are likely to exhibit significant improvements in SPAD overall performance.
6.2.2. Operation at longer wavelengths

One of the key goals is to operate these devices with high efficiency at the strategically important wavelength of 1550 nm. The high SPDE values, up to 38% at 1310 nm, afforded by the best performing planar Ge-on-Si SPADs is in part a consequence of the absorption of photons in the Ge occurring over the direct bandgap. Whereas the absorption of longer wavelengths (i.e. >1500 nm) occurs over the much weaker indirect bandgap. The absorption coefficient of the Ge absorber varies with operating temperature and as a result, the cut-off wavelength decreases with temperature. Therefore, the absorption of 1550 nm wavelength photons is significantly reduced at temperatures 78 K–200 K compared with at room temperature values, reducing the Ge-on-Si SPADs efficiency at these temperatures. The authors of [59] investigated how the SPDE changes with both increasing wavelength and temperature, and extrapolated the results to estimate the detection cut-off wavelength, $\lambda_c$. $\lambda_c$ is defined as the wavelength for which the SPDE of the device has fallen to half of that for a wavelength of 1450 nm. These results are summarised in figure 18. In figure 18(a), the normalised measured SPDE values are plotted as a function of wavelength. From this, it can be seen that at the highest temperature of 175 K, $\lambda_c$ has increased to 1495 nm. In figure 18(b), $\lambda_c$ is plotted as a function of temperature. The experimental results are represented by the blue squares and the black line is fitted to these results and extrapolated to show that $\lambda_c$ is estimated to reach 1550 nm at a temperature of 245 K. The authors also calculated the expected cut off wavelength for a device with an increased absorption layer thickness of 2 µm. These estimates are plotted against temperature in figure 18(c) and it can be seen from this graph, that with the absorption layer thickness doubled, $\lambda_c$ is increased to longer wavelengths and reaches 1550 nm at a lower temperature of ~220 K.

A further advantage to operating at this temperature range is that afterpulsing effects are expected to be reduced compared to ~78 K–200 K, as previously discussed in detail in section 2.2. Furthermore, Peltier coolers could be used for temperature control—both simplifying and reducing the overall footprint and energy consumption, resulting in a compact device capable of efficient single-photon detection at wavelengths throughout the SWIR spectral region. By referring to figure 13, however, it can be clearly seen that with the current devices, operation at temperatures >175 K would result in very high DCRs of the order of Mcps. One of the main challenges is therefore to reduce the DCR sufficiently so that the devices operate with high absorption efficiency at Peltier-cooler accessible temperatures.

Another method currently being explored to improve operation at 1550 nm is to use a germanium–tin (GeSn) alloy material as the absorber [76]. Ge is an indirect bandgap semiconductor; alloying Ge with Sn changes the electronic band structure of GeSn and with sufficient Sn concentration the material transitions into a fundamental direct bandgap semiconductor. This process red-shifts the absorption edge, and therefore even small fractions of Sn incorporated in the Ge layer of a SPAD can increase the cut-off wavelength towards the SWIR. The growth and fabrication of GeSn-on-Si devices, however, is challenging due to poor thermal stability of the GeSn alloy and the large lattice mismatch between GeSn and Si but work continues in this field to overcome these issues [76–80].

6.2.3. Further work and conclusions

Other future and ongoing works include afterpulsing characterisation over a wider temperature range and with varying device areas in order to compare with the 100 µm diameter device results presented in
Further laboratory-based LiDAR investigation with these devices will also be performed to compare with that presented in section 5. The record low DCR values reported should be advantageous for use in LiDAR applications, but the decrease in the SPDE may have a detrimental effect, thus it will be very useful to analyse the effects of both. In addition, longer range imaging experiments as well as demonstrating imaging of moving objects would prove the potential of these devices for automotive LiDAR applications.

Further to this, a fully functional 3D LiDAR system would benefit from the development of sufficiently large arrays of SPAD detectors. A linear or 2D SPAD detector array would significantly reduce the total acquisition time resulting in a highly sensitive, low noise, high-speed SWIR camera. Therefore, future work in this field will include the development of SPAD arrays once the individual detector design has been optimised sufficiently. The key advantage of such arrays compared to other material systems for detection in the SWIR, is that this technology would be highly suited for integration with CMOS foundries.

In conclusion, this paper has presented and reviewed the state-of-the-art in Ge-on-Si SPADs for use in the SWIR. Initially, the advantages of carrying out single-photon detection in the SWIR for a range of applications were presented. The principles of operation of a SPAD, as well as the fabrication and characterisation methods of Ge-on-Si SPADs were then presented, in addition to a review of the work carried out to date in this field. Since the first single-photon detection demonstration of a SACM Ge-on-Si SPAD in 2011, the performance of these devices has improved significantly. Despite the relatively immature growth and fabrication technology, planar geometry normal incidence Ge-on-Si SPADs have demonstrated SPDE, DCR, jitter and NEP values which are approaching comparable performance to competitor SPAD detector technology in the SWIR spectral range. A number of methods to improve the current performance at 1310 nm and also allow for efficient single-photon detection at longer wavelengths (focusing mainly on the important wavelength of 1550 nm) have been discussed. The implementation of these modifications could result in a high-performance SWIR SPAD system which would provide an alternative to methods currently used for single-photon detection in this spectral range with advantages including lower afterpulsing effects, lower associated costs and ease of integration with Si based technologies. The results presented and the future perspectives discussed justify the requirement for continued investment and research into the further development of Ge-on-Si based SPAD detectors for use in the SWIR.

Data availability statement

No new data were created or analysed in this study.

Acknowledgments

Royal Academy of Engineering (RF-201819-18-187, CiET2021_123); Innovate UK (17706, 44835); Defence Science and Technology Laboratory (DSTLX-1000092774); Engineering and Physical Sciences Research Council (UK EPSRC; EP/L024020/1, EP/N003446/1, EP/S026428/1, EP/N003225/1, EP/T001011/1, EP/T00097X/1).

ORCID iD

Fiona Thorburn  
https://orcid.org/0000-0001-9342-6446

References

[1] Loudon R 2000 Introduction: the photon The Quantum Theory of Light (Oxford: Oxford University Press) pp 1–2
[2] Planck M 1900 On an improvement of Wien's equation for the spectrum Verh. Dtsch. Phys. Ges. 2 202
[3] Planck M 1900 On the theory of the energy distribution law of the normal spectrum Verh. Dtsch. Phys. Ges. 2 237
[4] Einstein A 1905 On a heuristic point of view concerning the production and transformation of light Ann. Phys. 17 132–48
[5] Lewis G N 1926 The conservation of photons Nature 118 874–5
[6] Hadfield R H 2009 Single-photon detectors for optical quantum information applications Nat. Photon. 3 696–705
[7] Buller G S and Collins R J 2009 Single-photon generation and detection Meas. Sci. Technol. 21 012002
[8] Daniel R, Almog R, Ron A, Belkin S and Diamand Y S 2008 Modeling and measurement of a whole-cell bioluminescent biosensor based on a single photon avalanche diode Biosens. Bioelectron. 24 888–93
[9] Bruschi C, Homulle H, Antolovic I M, Burri S and Charbon E 2019 Single-photon avalanche diode imagers in biophotonics: review and outlook Light Sci. Appl. 8 87
[10] Karami M A and Ansarian M 2017 Neural imaging using single-photon avalanche diodes Basic Clin. Neurosci. 8 19–26
[11] Belliu S, Jackson C and König A 2005 Photon counting sensors for medical and biophotonic applications Laser+Photon. 5 34–38
[12] Collins R J, Amiri R, Fujisawa M, Honjo T, Shimizu K, Tamaki K, Takoaka M, Andersson E, Buller G S and Sasaki M 2016 Experimental transmission of quantum digital signatures over 90 km of installed optical fiber using a differential phase shift quantum key distribution system Opt. Lett. 41 4883–6
[50] Jiang X, Itzler M, O’Donnell K, Entwistle M, Owens M, Slomkowski K and Rangwala S 2015 InP-based single-photon detectors and Geiger-mode APD arrays for quantum communications applications IEEE J. Sel. Top. Quantum Electron. 21 5–16
[51] Warburton R E, Itzler M and Buller G S 2009 Free-running, room temperature operation of an InGaAs/InP single photon avalanche diode Appl. Phys. Lett. 94 071116
[52] Scarcella C, Bosso G, Ruggeri A and Tosi A 2015 InGaAs/InP single-photon detector gated at 1.3 GHz with 1.5% afterpulsing IEEE J. Sel. Top. Quantum Electron. 21 17–22
[53] Michel J, Liu J and Kimerling L C 2010 High-performance Ge-on-Si photodetectors Nat. Photon. 4 527–34
[54] Chen C, Li C, Huang S, Zheng Y, Lai H and Chen S 2012 Epitaxial growth of germanium on silicon for light emitters Int. J. Photoenergy 2012 789605
[55] Loudon A Y, Hiskett P A, Buller G S, Carline R T, Herbert D C, Leong W Y and Rarity J G 2002 Enhancement of the infrared concentration Sci. Rep. 2 38386
[56] Jiang X, Itzler M A, Ben-Michael R and Slomkowski K 2007 InGaAsP–InP avalanche photodiodes for single photon detection J. Sel. Top. Quantum Electron. 13 895–905
[57] Warburton R E et al 2013 Ge-on-Si single-photon avalanche diode detectors: design, modeling, fabrication, and characterization at wavelengths 1310 and 1530 nm IEEE Trans. Electron. Devices 60 3807–13
[58] Micro Photon Devices S.r.l. MPD PDM-IR 900 nm–1700 nm infrared photon detection module (available at: www.micro-photon-devices.com/MPD/media/Datasheet/PDM-IR%20Datasheet%20window.pdf)
[59] Martinez N J D, Derose C T, Brock R W, Starbuck A L, Lentine A L, Trotter D C and Davids P S 2016 High performance Germanium-on-silicon single-photon avalanche diode detectors Nat. Commun. 7 10862
[60] Lu Z, Kang Y, Hu C, Zhou Q, Liu H and Campbell J C 2011 Geiger-mode operation of Ge-on-Si avalanche photodiodes IEEE J. Quantum. Electron. 47 731–5
[61] Warburton R E et al 2013 Ge-on-Si single-photon avalanche diode detectors: design, modeling, fabrication, and characterization at wavelengths 1310 and 1530 nm IEEE Trans. Electron. Devices 60 3807–13
[62] Micro Photon Devices S.r.l. MPD PDM-IR 900 nm–1700 nm infrared photon detection module (available at: www.micro-photon-devices.com/MPD/media/Datasheet/PDM-IR%20Datasheet%20window.pdf)
[63] Martinez N J D, Gehl M, Derose C T, Starbuck A L, Pomerene A T, Lentine A L, Trotter D C and Davids P S 2017 Single photon detection in a waveguide-coupled Ge-on-Si lateral avalanche photodiode Opt. Express 25 16130–9
[64] Ferre-Llinà L et al 2020 High sensitivity Ge-on-Si single-photon avalanche diode detectors Opt. Lett. 45 6406–9
[65] Thorburn F E, Huddleston L L, Kirdoda J, Millar R W, Ferre-Llinà L, Yi X, Paul D J and Buller G S 2020 High efficiency planar geometry germanium-on-silicon single-photon avalanche diode detectors Proc. SPIE Advanced Photon Counting Techniques XIV 11386 113860N
[66] Dumas D C S, Kirdoda J, Millar R W, Vines P, Kuzmenko K, Buller G S and Paul D J 2019 High-efficiency Ge-on-Si SPADs for short-wave infrared Proc. SPIE Optical Components and Materials XVI 10914 10914D-1
[67] Kirdoda J, Dumas D C S, Kuzmenko K, Vines P, Greener Z M, Millar R W, Mirza M M, Buller G S and Paul D J 2019 Geiger mode Ge-on-Si single-photon avalanche diode detectors IEEE J. Sel. Top. Quantum Electron. 25 2807–15
[68] Liu Z, Cong H, Yang F, Li C, Zheng J, Xue C, Zuo Y, Cheng B and Wang Q 2016 Defect-free high Sn-content GeSn on insulator layers grown by rapid melting growth Sci. Rep. 6 38386
[69] Tran H et al 2019 Study of GeSn mid-infrared photodetectors for high frequency applications Front. Mater. 6 278
[70] Eales T D et al 2019 Ge1-xSnx alloys: consequences of band mixing effects for the evolution of the band gap Γ-character with Sn concentration Sci. Rep. 9 14077
[71] Tran H et al 2019 Si-based GeSn photodetectors toward mid-infrared imaging applications ACS Photon. 6 2807–15
[72] Warburton R E, Dumas D C S, Gallacher K E, Jahandar P, MacGregor C, Myronov M and Paul D J 2017 Mid-infrared light emission >3 μm wavelength from tensile strained GeSn microdisks Opt. Express 25 25374–85