Waveguide-integrated, plasmonic enhanced graphene photodetectors

J. E. Muench\textsuperscript{1}, A. Ruocco\textsuperscript{1}, M. A. Giambra\textsuperscript{2}, V. Miseikis\textsuperscript{2,3,4}, D. Zhang\textsuperscript{1}, J. Wang\textsuperscript{3}, H. F. Y. Watson\textsuperscript{1}, G. C. Park\textsuperscript{1}, S. Akhavan\textsuperscript{1}, V. Soriano\textsuperscript{2}, M. Midrio\textsuperscript{5}, A. Tomadin\textsuperscript{6}, C. Coletti\textsuperscript{3,4}, M. Romagnoli\textsuperscript{2}, A. C. Ferrari\textsuperscript{1} and I. Goykhman\textsuperscript{7}

\textsuperscript{1}Cambridge Graphene Centre, University of Cambridge, Cambridge CB3 0FA, UK
\textsuperscript{2}Consorzio Nazionale per le Telecomunicazioni, Photonic Networks and Technologies National Laboratory, 56124 Pisa, Italy
\textsuperscript{3}Center for Nanotechnology Innovation @ NEST, Istituto Italiano di Tecnologia, 56127 Pisa, Italy
\textsuperscript{4}Graphene Labs, Istituto Italiano di Tecnologia, 16163 Genova, Italy
\textsuperscript{5}Consorzio Nazionale per le Telecomunicazioni, University of Udine, 33100 Udine, Italy
\textsuperscript{6}Dipartimento di Fisica, Universitá di Pisa, Largo Bruno Pontecorvo 3, 56127 Pisa, Italy and
\textsuperscript{7}Micro- and Nanoelectronics Research Center, Technion, Haifa 320000, Israel

We present a micrometer scale, on-chip integrated, plasmonic enhanced graphene photodetector (GPD) for telecom wavelengths operating at zero dark current. The GPD is designed and optimized to directly generate a photovoltage and has an external responsivity $\sim 12.2$V/W with a 3dB bandwidth $\sim 42$GHz. We utilize Au split-gates with $\lambda = 1$00 nm gap to electrostatically create a p-n junction and simultaneously guide a surface plasmon polariton gap-mode. This increases light-graphene interaction and optical absorption and results in an increased electronic temperature and steeper temperature gradient across the GPD channel. This paves the way to compact, on-chip integrated, power-efficient graphene based photodetectors for receivers in tele and datacom modules.

The ever-growing demand for global data traffic\cite{1} is driving the development of next generation communication standards\cite{2,3}. The increasing numbers of connected devices\cite{4,5}, the need for new functionalities, and the development of high-performance computing\cite{6,7} require optical communication systems performing at higher speeds, with improved energy-efficiency, whilst maintaining scalability and cost-effective manufacturing. Si photonics\cite{7–9} offers the prospect of dense (nanoscale) interconnects\cite{11} relying on mature, low-cost (based on complementary metal-oxide-semiconductor (CMOS) fabrication processes) manufacturing\cite{8,9}, making it one of the key technologies for short-reach ($<$10km) optical interconnects\cite{11} beyond currently employed lithium niobate\cite{12} and indium phosphate\cite{13}.

A variety of functionalities have been developed and demonstrated in Si photonics for local optical interconnects\cite{11}. Electro-optic modulators based on carrier-depletion (phase-modulation) in Si\cite{14,15} or the Franz-Keldysh effect\cite{16} (amplitude-modulation) in strained Si-Ge\cite{17,18} encode information into optical signals at telecom wavelengths ($\lambda = 1.3\text{-}1.6$μm). On the receiver side, Ge\cite{19} or bonded III-V\cite{20,21} photodetectors (PD) are needed for optical-to-electrical signal conversion, since the telecom photon energies are not sufficient for direct (band-to-band) photodetection in Si\cite{22}.

On-chip integrated Ge PDs\cite{23–27} are standard components in Si photonics foundries\cite{8,9,22}. Their external responsivities (in A/W), $R_I = I_{ph}/P_{in}$, where $I_{ph}$ is the photocurrent and $P_{in}$ is the incident optical power, can exceed 1A/W\cite{8,22} and their bandwidth can reach 60GHz\cite{23,24}. Following the development of high temperature ($> 600$°C)\cite{19} heterogeneous integration of Ge-on-Si using epitaxial growth and cyclic thermal annealing\cite{13,28,29}, the concentration of defects and threading dislocations in Ge epilayers and at Si/Ge interfaces can be reduced\cite{19}, resulting in low ($< 10$A/W\cite{27}) dark current in waveguide integrated Ge p-i-n photodiodes\cite{24,27}. However, Ge-on-Si integration is a complex process\cite{14–27}, as the lattice mismatch between Si and Ge\cite{19}, ion implantation\cite{23,27}, thermal budget (i.e. thermal energy transfer to the wafer) management\cite{22}, and the non-planarity of Ge layers\cite{22} require dedicated solutions during device fabrication\cite{8}.

The charge carrier mobility $\mu$ in Si and the dislocations and defects in grown\cite{20} or evaporated\cite{30} Ge layers set intrinsic limitations that prevent further improvements to the operation speed of Ge PDs without compromising $R_I$\cite{20,22}. These shortcomings, together with the spectrally limited operation regime (band edge in Ge~1.57μm\cite{22}, which can be extended to~1.62μm\cite{31} at the expense of $R_I$), and the incompatibility of Ge epitaxy for monolithic integration with other material platforms, such as SiN, are amongst the main limitations for Ge PDs\cite{8}. Thus, novel solutions for PDs, integrated with Si photonics, at telecom bands are needed.

Graphene is a promising candidate for on-chip integrated photonics\cite{32,52}. The advantages of single-layer graphene (SLG) for photonics stem from its superior optoelectronic properties\cite{53}. These include high-speed (>200GHz\cite{54}) operation\cite{57}, broadband (ultraviolet to far-infrared) absorption\cite{56–58}, efficient optical modulation (electro-optical index change $\Delta n_{eff} > 10^{-3}$)\cite{32–38}, CMOS compatibility\cite{41,50} and integrability\cite{32,51,52} with different on-chip photonics platforms, such as silicon-on-insulator (SOI)\cite{32} and SiN\cite{50}. In the case of waveguide-integrated graphene PDs (GPDs)\cite{10,52}, high speeds up to 128GHz\cite{32}, wafer-scale integration\cite{48}
and $R_l \approx 0.4-0.5\,\text{A/W}$ were reported. GPDs can offer broadband detection across multiple telecommunication channels (O-band~1.31\,\text{µm} to U-band~1.65\,\text{µm}) and direct generation of photovoltage\cite{43,47,50,51}. The latter opens up the possibility of building GPDs without the noise contribution of dark current\cite{43,47} and the requirements of the read-out electronics\cite{62}. The dominating effect for a given GPD depends on device configuration, design geometry, and way of operation\cite{64,71}. For telecom applications, where high-speed (tens GHz) operation is one of the key requirements\cite{32}, PV, PTE and PB are typically considered for waveguide-integrated GPDs\cite{32}, taking advantage of the ultra-fast (\text{fs-ps}) carrier dynamics in SLG\cite{72,73}.

PTE is ideal for PD operation in a voltage mode. In optically illuminated SLG, electron-electron scattering drives the formation of a ‘hot’ (optically excited)-carrier distribution, described by the Fermi-Dirac function\cite{74}, within <50fs\cite{72}. This can remain at elevated temperatures $T_e$, well above the lattice temperature $T_l$, over~2-4ps time scales\cite{72}, before reaching thermal equilibrium via phonons interaction\cite{72,75,76}. In this hyperthermal state, a photovoltage $V_{ph}$ is generated by a thermoelectric current as for the Seebeck effect\cite{60}, if both a temperature and chemical potential gradient are present in the SLG channel. The sign and magnitude of $V_{ph}$ depend on the Seebeck coefficient ($S$), i.e. the proportionality constant between temperature change and induced photovoltage\cite{77}, and $T_e$ profile in the SLG channel:\cite{66}

$$V_{ph} = \int S(x) \cdot \nabla T_e(x) \, dx$$ \hspace{1cm} (1)

where $x$ is the coordinate along the channel from drain to source, and $S$ is given by Mott’s formula\cite{64,66,77}:

$$S(x) = -\frac{\pi^2}{3e} \frac{k_b T_e}{\sigma(x)} \frac{\mu_e}{\mu_i}$$ \hspace{1cm} (2)

with $\sigma(x)$ the conductivity, $k_b$ the Boltzmann constant, $e$ the electron charge and $\mu_i$, the chemical potential ($\mu_e = E_F$ at $T_e = 0$\cite{74}, with $E_F$ the Fermi energy).

PTE-GPDs have been reported in vertically-illuminated\cite{60,78,81} and waveguide-integrated\cite{43,47} configurations. The latter used SLG flakes prepared by micromechanical cleavage (MC) of graphite\cite{52}, with typical device length of tens of $\mu$m\cite{43,47}. They have external voltage responsivities, defined as $R_v = V_{ph}/P_m$, up to \sim 0.47\,\text{V/W}\cite{43} (at zero bias) with speeds up to 65\,\text{GHz}\cite{43}. Depending on PTE-GPD design configuration and the requirements of the read-out electronics (i.e. output photo-signal to be measured as current or voltage), the responsivity can be characterized in terms of $R_l$ or $R_v$. The photovoltage generated by the Seebeck effect is associated with a thermoelectric current across the PD by a Ohmic relation\cite{43,47,62}, $I_{ph} = V_{ph}/R$, with $R$ the resistance.

To increase $R_v$ for PTE-GPDs, Eq\,1 suggests two strategies: 1) maximize $S$; 2) maximize the $T_e$ gradient profile in the SLG channel. The former is related to $\mu$ via Eq\,2 and the Drude conductivity\cite{77}, $\sigma = e\mu n$, where $n$ is the charge carrier concentration. Thus, $S$ can be improved by using high-mobility SLG, e.g. encapsulating SLG in hBN\cite{83,85}, using single-crystals\cite{83,86}, or large (tens $\mu$m) domain-size\cite{87}, in combination with a transfer processes that avoid contamination\cite{84,88}, strain\cite{88}, and cracks\cite{89}. Ref.\,32 suggested that $\mu > 10^4 \text{cm}^2\text{V}^{-1}\text{s}^{-1}$ could enable $R_v > 100\,\text{V/W}$. The $T_e$ gradient can be increased by creating a spatially confined, localized, heat source\cite{45} generated by enhanced optical absorption in SLG over compact (< 10\,\µm) device lengths\cite{50,51}. This could be achieved by integrating plasmonic nanostructures\cite{90-95}. Sub-wavelength plasmonic confinement and associated enhancement of near-field light-matter interaction were used to boost $R_i$ in Si-plasmonic PDs\cite{90,91}, plasmonic-Ge PDs\cite{92}, plasmonic decorated GPDs for free-space\cite{93,95} and waveguide-integrated\cite{43,50,52} configurations. Refs.\,50,52 reported plasmonic enhanced on-chip GPDs based on PV\cite{50,52} and PB\cite{50,52} with $R_i \approx 0.5\,\text{A/W}$ and bandwidth~\sim 110\,\text{GHz} at 1.55\,\µm for source-drain bias <1V.

Here, we report compact (~0.5-4\,µm), PTE-based, waveguide-integrated, plasmonic-enhanced GPDs for telecom wavelengths with $R_v \sim 12.2\,\text{V/W}$ at zero source-drain bias and zero dark current, with a 3dB cutoff frequency~\sim 42\,GHz. To the best of our knowledge, this is the largest $R_v$ to date for waveguide-integrated GPDs operating in voltage mode. We use SLG grown by chemical vapor deposition (CVD) and transferred onto low-loss (~1dB/cm) planarized (i.e. fully-embedded in polished cladding\cite{7}) SiN waveguides with a semi-dry (i.e. combining wet de-lamination from the growth substrate with dry lamination onto the target substrate) transfer\cite{86}. Unlike previous PTE GPDs exploiting non-scalable MC SLG\cite{45,47}, our design relies on Au split-gates to electrostatically create a p-n junction in the SLG channel, as well as to guide a confined SPP waveguide mode. By leveraging optical field enhancement and plasmonic confinement in the gap, we increase light-SLG interaction and optical absorption in the p-n junction region, resulting in a confined electrons heat source, compact device length, and increased $R_v$. This combines high-performance (large $R_v$, high-speed, bias-free, compact, direct $V_{ph}$ read-out) PTE GPDs in the telecom range with scalable fabrication, paving the way for graphene integrated receivers for next-generation transceivers.
The design of our GPD is schematically shown in Fig. 1a, b. It comprises a SLG channel on a SiN waveguide supporting a transverse-electric (TE, in-plane) polarized fundamental waveguide mode. Two Au gates are placed above the channel, separated from the SLG by an Al₂O₃ dielectric spacer and centrally aligned with respect to the waveguide. When this split-gate structure is DC (direct current) biased, it forms a p-n junction, Fig. 1a, and creates a S profile in the SLG channel, as for Eq. 2. When an on-chip guided light signal reaches the PD area, it is evanescently coupled from the SiN waveguide to the split-gate, which acts as SPP waveguide, Fig. 1b. The plasmonic guiding with light confinement in the gap (width \( w_{\text{gap}} \sim 100\,\text{nm} \)) leads to enhanced optical absorption and a localized hot electron distribution with a \( T_e \) gradient in the p-n junction (gap) region. The coupling efficiency, \( P_{\text{out}}/P_{\text{in}} \), where \( P_{\text{out}} \) is the power transferred between two optical components, from photonic to plasmonic waveguide mode can be optimized by tailoring \( w_{\text{gap}} \) and dielectric spacer thickness (\( t_{\text{ox}} \)).

To optimize the cross section parameters at \( \lambda=1.55\,\mu\text{m} \), we perform optical simulations using a commercial finite difference solver tool (Numerical MODE). After selecting the fundamental gap plasmon mode for a given design and \( \lambda \), we extract the optical electric field distribution in the SLG channel to model the absorbed power density that generates the hot carrier distribution as time-averaged electric power dissipation density \( \rho_e \), which we refer to as Joule heat source (\( J \)) hereafter. After normalization to an input power of 1\( \mu \)W, this is used in the heat equation \( \frac{\partial T(x)}{\partial t} - \kappa(x) \left[ \frac{\partial^2 T(x)}{\partial x^2} + \frac{1}{\xi^2} \Delta T(x) \right] = J(x) \) (3)

where \( \Delta T_e(x) = T_e(x) - T_i \) is the local temperature fluctuation, \( \xi \) is the cooling length (see Methods) and \( \kappa_e(x) \) is the electronic thermal conductivity (see Methods). Eq. 3 gives the \( T_e(x) \) profile along the SLG channel. The \( S(x) \) profile from Eq. 2 is used in Eq. 3 to obtain \( V_{\text{ph}} \). The device parameters are chosen to maximize \( V_{\text{ph}} \).

A second aspect of device design concerns the coupling between the dielectric and plasmonic waveguides, as well as the positioning and width of the SLG channel along the split-gate. Taper-assisted butt-coupling (end-to-end alignment) was reported to yield the lowest insertion loss (< 0.6 dB) for the transition from optical to plasmonic modes. However, since evanescent coupling (lateral or vertical alignment) provides simpler fabrication and greater flexibility for the placement of devices on top of integrated optical circuits, we use this here, Fig. 1b-c. To obtain the largest \( R_V \), the electric field distribution along the propagation directions needs to be considered. Light absorption in SLG or in the plasmonic structure along the device leads to an exponential decay of optical power. Thus, the increase in \( T_e \) follows the same decay. The resulting photovoltage drop at different points along the device results in an averaged potential difference between source and drain contacts. To optimize \( R_V \), a compact (<10\( \mu \)m) device with optimized peak absorption and minimal drop-off is preferable. We thus per-
form finite-difference time domain (FDTD) simulations in Lumerical FDTD (see Methods). The co-existence of plasmonic and dielectric waveguide leads to oscillating power exchange between both structures\cite{100, 102}. For the highest coupling efficiency, the vertical distance between these waveguides is typically $> 150\text{nm}$\cite{100, 101}, exploiting interference between quasi-even and quasi-odd eigenmodes\cite{100, 102}. In our design, we keep this separation small (tens nm) to ensure overlap between SLG and gap plasmon mode, to avoid a long ($> 5\mu m$) coupling length\cite{101}, and to create a sharper concentration of power close to the front of the plasmonic structure. The SLG channel is placed accordingly, after a short ($\sim 1\mu m$) taper at the front of the SPP waveguide to reduce mode mismatch.

Fig.2 summarizes the fabrication process of our GPDs. Planarized SiN waveguides, Fig.2a, (260nm high, width 0.8-1.5$\mu m$) on 15$\mu m$ SiO$_2$ are fabricated as follows. The SiN layer is first deposited by low-pressure (LP) CVD. The SiN photonic waveguides are then defined by electron beam lithography (EBL) and reactive ion etching. For surface planarization, a 1.6$\mu m$ thick boron-phosphorus tetraethyl orthosilicate (BPTEOS) layer is deposited as top cladding and subsequently etched to a final thickness $\sim 20\text{nm}$ on top of the SiN waveguides, avoiding chemical mechanical polishing. SLG is grown on pre-patterned, electropolished Cu with Cr nucleation sites as for Ref.\cite{86}. After an initial annealing in argon (10mins), SLG growth is initiated at 25mbar with argon, hydrogen, and methane flowing at 900, 100, and 1 standard cubic centimeters per minute (sccm), respectively. After growth, SLG single crystals are placed onto the
A laser writer is used to define an etch mask to prevent air breakdown in the gap between gate contacts on monic split gates, Fig. 2e. To encapsulate the device and electron beam evaporation are used to fabricate the plasmonic mirror (ALD), Fig. 2d. An additional EBL step and 30 nm AlO$_3$ is used, thermal evaporation and lift-off in acetone, Fig. 2c.

Fig. 2b. Next, contacts are defined by another EBL step. Stripping the PMMA etch mask, thermal evaporation, and lift-off through acetone, Fig. 2b. The quality of SLG is monitored by Raman spectroscopy at all critical points during the fabrication process, using a Renishaw InVia equipped with a 50x objective (numerical aperture NA=0.75) at 514.5 nm with power below 0.5 mW to exclude heating effects and risk of damage. Representative spectra of SLG on Cu (after removal of Cu background photoluminescence) and after transfer onto the waveguide, and after complete device fabrication, are shown in Fig. 4. The absence of a D peak confirms negligible defects are introduced during fabrication. The 2D peaks are single-Lorentzian, confirming the presence of SLG. On Cu, the position and full width at half-maximum of the G peak are $\text{Pos}(G) \sim 1590$ cm$^{-1}$ and FWHM(G) $\sim 8$ cm$^{-1}$. The position of the 2D peak, $\text{Pos}(2D)$, is $\sim 2721$ cm$^{-1}$ with FWHM(2D) $\sim 27$ cm$^{-1}$. The 2D to G peak intensity, $I(2D)/I(G)$, and area, $A(2D)/A(G)$, are $\sim 1$ and $\sim 3.2$. After transfer, $\text{Pos}(G) \sim 1590$ cm$^{-1}$, FWHM(G) $\sim 10$ cm$^{-1}$, $\text{Pos}(2D) \sim 2690$ cm$^{-1}$, FWHM(2D) $\sim 28$ cm$^{-1}$, $I(2D)/I(G) \sim 3.2$ and $A(2D)/A(G) \sim 8.6$. This corresponds to $\sim 250$ meV doping and a carrier concentration $\sim 4 \times 10^{12}$ cm$^{-2}$. After the final encapsulation, $\text{Pos}(G) \sim 1590$ cm$^{-1}$, FWHM(G) $\sim 9$ cm$^{-1}$, $\text{Pos}(2D) \sim 2689$ cm$^{-1}$, FWHM(2D) $\sim 30$ cm$^{-1}$, $I(2D)/I(G) \sim 2.2$ and $A(2D)/A(G) \sim 7.6$, indicating $\sim 350$ meV ($n \sim 7 \times 10^{12}$ cm$^{-2}$) doping.

To determine the DC operating point, we perform electrical characterizations by sweeping the split-gate voltages ($V_{\text{Gate 1}}, V_{\text{Gate 2}}$) while measuring the device current $I_{DS}$ under a constant source-drain bias $V_{DS}=1$ mV, using DC probes on micromanipulators and two source measure units. To record the static photoresponse, we add two fibre probes and couple continuous wave (CW) transverse-electric (TE) polarized light at 1500–1580 nm from a tunable laser into the SiN waveguide via an optical fibre and a grating coupler (GC). While $V_{ph}$ is recorded across the unbiased ($V_{DS} = 0$ V) channel as a function of $V_{\text{Gate 1}}$ and $V_{\text{Gate 2}}$, using a lock-in amplifier under internal modulation (square wave, ON-OFF) of the laser with 200 Hz, we monitor the transmission with a second fibre positioned over the output GC and connected to an external InGaAs power meter to ensure constant $P_{in}$.

Fig. 5 plots the $R$ map of a typical device as a function of $V_{\text{Gate 1}}, V_{\text{Gate 2}}$. This shows a four-fold pattern, corresponding to the four doping constellations (p-n, n-p, n-n, p-p) in the SLG channel for different combinations of gate voltages. The map is symmetric with a maximum $R \sim 9$ kΩ at the crossing of the charge neutrality point (CNP), between -4 and -5 V. This corresponds to n-doping of the unbiased SLG channel with $n \sim 7 \times 10^{12}$ cm$^{-2}$ ($\sim 350$ meV). $R$ has contributions from channel ($R_{ch}$) and contact ($R_c$) resistances. $R_{ch}$ includes a fixed contribution from ungated SLG regions and a gate-dependent contribution from channel segments underneath the split-
gates. The gate-dependent variation in $R$ in Fig. 5(a) suggests $R_{ch}$ as the dominant factor. This is consistent with our contact resistivity ($<1\,\,\Omega\mu m$) for CVD SLG and the calculated $R$ based on channel geometry and sheet resistance obtained from independent four-terminal measurements on reference Hall bars. From these we also extract an average $\mu \sim 2000\,\,cm^2/V\,s$ from linear fits of the conductivity via $\sigma = \mu |dV_{Gate}/C_{ox}|$, where $C_{ox}$ is the top gate capacitance. Fig. 5(b) is a bi-directional gate sweep of the conductivity, indicating low hysteresis and charge-trapping in the Au/Al$_2$O$_3$/SLG gate capacitor.

Fig. 6(a) is a $V_{ph}$ map of a typical device at $P_{in} \sim 100\,\,\mu W$ inside the GPD. The plot exhibits a six-fold pattern with higher response for bipolar (p-n, n-p) junctions and a weaker one with sign-crossing along the diagonal ($V_{Gate\,1} = V_{Gate\,2}$) for the unipolar (n-n, p-p) junctions. When the GPD is operated at zero $V_{DS}$, this indicates a PTE-dominated photodetection as the two sign changes in $V_{ph}$ along a single-gate sweep line (e.g. $V_{Gate\,2}$ = const.) reflect the two sign changes of the $S$ gradient across the junction, arising from the non-monotonic dependence of $S$ on $\mu_{ph}$. The measured photoresponse is in good agreement with the calculated one in Fig. 6(b). We observed a similar behavior on > 12 devices of different sizes across 5 chips, the shortest being 500nm in the light propagation direction for a footprint ~3$\,\,\mu m^2$. For all devices, we got a maximum $V_{ph}$ close to the CNP where $S$ is largest, with a gradual drop-off at higher doping.

To calculate $R_V$, we first estimate the optical power inside our GPDs by taking into account: a) combined loss ($\sim 9.6\,\,dB$ at peak transmission) of waveguide propagation and fiber-to-waveguide coupling (wavelength-dependent, following the response envelope of the GC); b) 3dB power reduction from the input laser modulation (square wave, ON-OFF) with a 50% duty cycle; c) 3dB power splitting in the Y-branches and their ~0.2dB losses. We get $R_V \sim 12.2\,\,V/W$, the largest reported so far for waveguide-integrated GPDs in voltage mode at zero-bias.

Fig. 6(b) plots the $R_V$ wavelength dependence, showing a broadband (1.50-1.58$\,\,\mu m$) photoresponse covering the entire C-band (1.53-1.565$\,\,\mu m$) and beyond. The error bars indicate variations in the wavelength-dependent coupling loss (thus $P_{in}$), estimated as standard deviation from transmission measurements on > 10 reference waveguides. We attribute the gradual increase in $R_V$ with increasing wavelengths to improved coupling efficiency from dielectric to plasmonic waveguide.

Fig. 6(c) is the $V_{ph}$ power dependence at 1.575$\,\,\mu m$ for optical power levels comparable to those required by receivers used in 100GBs$^{-1}$ links. The linear response indicates a power-independent $R_V$.

To highlight our GPDs’ behavior as voltage sources, when a signal is generated, we place two devices back to back on the same waveguide and connect them in series. This modified design, Fig. 7(a), consists of two SLG channels gated from the same split-gate/SPP waveguide. By connecting the drain pad of one channel to the source of the other through a metal lead crossing the waveguide behind the active region of the devices, Fig. 7(b), we measure both GPDs individually, as well as combined. Fig. 7(c) is a false color SEM of the active region of both detectors. Since each GPD is designed to maximally absorb over the device length, the power rapidly decays along the propagation direction after the first GPD. We thus
place the second device~1µm from the first.

Figs[7d,e] plot the photovoltage maps for both GPDs at $P_{in} \sim 70µW$. The GPD closer to the input GC (A) absorbs most of the light and has the six-fold pattern typical of PTE, Fig[7d]. The photoresponse of the second GPD (B) is weaker, due to light absorption in SLG and metal, Fig[7d,e]. A six-fold pattern is not observed, due to photocurrents generated in the junctions between gated and ungated sections at either end of the SLG channel. Figs[7f,g] are photovoltage maps of the combination of both GPDs. The response in series is in Fig[7f], while the sum of the individual responses in Fig[7g]. The two plots are good agreement, confirming that $V_{ph,A+B} = V_{ph,A} + V_{ph,B}$. Thus, in order to increase $R_{ph}$ in long (tens µm) PTE-GPDs with absorption only in the SLG channel, one could avoid the reduction in photosignal due to the decrease in $\Delta T_e$ along the device and instead add the voltages generated in different sections by subdividing the channel into several shorter devices and connecting them as cascaded GPDs. To minimize the length of metal leads for contacting and connecting individual devices, this configuration could comprise individually gated devices with alternating p-n junctions to form a meandered structure. This would ideally be implemented with transparent gates, such as indium tin oxide, or a second SLG at a distance far enough from the channel, to avoid additional losses.

To evaluate the frequency response we use the optical heterodyne set-up in Fig[8a], combing optical signals at different frequencies. The channel is contacted with an RF probe in G-S configuration. The output of our tunable laser source is combined with that of a fixed-wavelength laser diode (Thorlabs SFL1550P) and the GPDs’ response to the amplitude beating at the difference frequency is monitored with an electrical spectrum analyzer (ESA, Agilent PSX N9030A). The two source
outputs are combined in a 50:50 fibre coupler. We monitor the signal stability (i.e. output power and the position of the difference frequency) using a reference PD and an oscilloscope. Prior to coupling the combined signal into the SiN waveguide, we use an erbium-doped fibre amplifier (EDFA, Keyopsys CEFA-C-HG) to raise the optical power to 15dBm (×30mW) to increase the output signal detected at the ESA. To overcome the signal reduction due to impedance mismatch between device $R$ and the 50Ω of the measurement equipment, we use an additional low noise amplifier (LNA) between GPD and ESA. In order to distinguish between the frequency response of our GPDs and that of the LNA and the remaining measurement equipment, the response of the latter two to a low power (-80dBm) input from a 50GHz signal generator is recorded for calibration.

Fig 8b plots the calibrated response (black squares) to the beating signal at different frequencies, while the split-gate is biased to set an operating point in the p-n junction regime resulting in the largest photoresponse under CW illumination in Fig 8. The response stays within 3dB of the low-frequency (1GHz) reference power until 40GHz, the limit of our measurement set-up.

To determine the cut-off, we therefore modify the set-up (Fig 8c) to perform impulse response measurements, where the response to ultra-short (×150fs) optical pulses is monitored with an oscilloscope. For excitation, we use the idler of an optical parametric oscillator, pumped by a Ti:Sa mode-locked laser at 1.55μm, attenuated in free-space prior to coupling into the optical fibre. Fig 8d is the measured impulse response at the same operating point as for the heterodyne measurements. We obtain a pulse duration, assuming a Gaussian pulse shape, $\Delta t \sim 11$ps.

For Gaussian-shaped pulses, the time-bandwidth product is $\sim 0.44\frac{\Delta t}{\text{BW}}$. From this we estimate $f_{3\text{dB}} \sim 40$GHz. The fast Fourier transform of the pulse is in Fig 8b (red circles) after calibration. The trace is in good agreement with the heterodyne response and drops below -3dB at $\sim 42$GHz, showing high-speed operation on par with current Ge PDs, consistent with other reports of high-speed MC-SLG-based PTE GPDs [45–47]. However, this bandwidth is the highest reported so far for PTE-based on-chip GPDs made from CVD SLG.

In summary, we reported waveguide-integrated plasmonic enhanced GPDs with an external responsivity $\sim 12.2$V/W, a -3dB cut-off $\sim 42$GHz, and small ($\sim 3-20\mu m^2$) device footprints, using CVD SLG on SiN. We exploited the integration of an SPP waveguide.
with a SLG p-n junction to enhance light-SLG interaction and create a confined electron heat-source to obtain a strong, PTE-dominated photoresponse. This paves the way to power-efficient receivers for optoelectronic links.

We acknowledge funding from EU Graphene Flagship, ERC Grant Hetero2D, EPSRC Grants EP/K01711X/1, EP/K017144/1, EP/N010345/1, and EP/L016087/1.

**METHODS**

**Device design and modeling**

Plasmonic gap width, SLG placement, and thickness of metal and oxide structures are the key parameters to be optimized to achieve maximum photovoltage. To do so, we build a device model in the layout environment of a commercial eigenmode solver (Lumerical MODE Solutions). In order to model SLG in the optical solver and subsequent calculations consistently, we use a volumetric permittivity material model, in which SLG is described as cuboid with finite thickness \( t = 0.34 \text{nm} \) and in-plane \( (\varepsilon_{\parallel}) \) and out-of-plane \( (\varepsilon_{\perp}) \) permittivity are defined as independent tensor elements. To calculate the in-plane relative permittivity for SLG, we use \[ \varepsilon(\omega) = \varepsilon'(\omega) + i\varepsilon''(\omega) = \varepsilon_r + \frac{i\sigma(\omega)}{\varepsilon_0 \omega l}. \] where \( \varepsilon'(\omega) \) and \( \varepsilon''(\omega) \) are the real and imaginary part of the relative permittivity \( \varepsilon(\omega) \), \( \sigma(\omega) \) is the SLG optical conductivity, \( \omega \) is the angular frequency, \( \varepsilon_r \) is the background relative permittivity (whose frequency-dependence is ignored in the small \( (\lambda = 1.5-1.6 \mu\text{m}) \) wavelength range under consideration), and \( \varepsilon_0 \) is the permittivity of free space. \( \sigma(\omega) \) is obtained by linear-response in the random-phase approximation, and contains terms describing intraband and interband transitions. While the former can be evaluated analytically, the interband term requires a numerical solution. The out-of-plane component of the dielectric tensor matches \( \varepsilon_r \).

We then run the eigenmode solver, using the relative permittivity function for SLG, and select the fundamental (gap plasmon) mode for further processing. We export the simulation mesh grid positions, electric \( E \) and magnetic \( H \) data, effective refractive index \( n_{\text{eff}} = \beta/k_0 \), where \( \beta \) is the propagation constant of the mode and \( k_0 \) is the free space wavevector, and all relevant geometric parameters such as gap and gate width and gate oxide thickness to complete the rest of our modeling.

A crucial intermediate step requires the determination of the \( T_e(x) \) profile in the SLG channel. The first step establishes the operating regime. As discussed in Refs.\[114, 115\], the energy delivered to the electronic carrier distribution by pumping SLG with a pulsed laser can
of all thermodynamic and transport parameters in Eqs. 1-3 can be solved to linear order in the local $T_e(x)$ fluctuation, evaluating all thermodynamic and transport parameters at $T_e(x) = T_l$. In particular, a single $\mu_e$ is established, following photoexcitation, by electron-electron interactions across the valence and conduction bands, and the thermal conductivity, calculated from the Wiedemann-Franz law $\kappa_e = \pi^2 k_B^2 T_e \sigma / (3e^2)$, is uniform in space.

The intended operation of our GPD is under CW or, during data reception, quasi-CW (i.e. the pulse duration exceeds the cooling time of the hot-carrier distribution) illumination with low $P_{in} < 0.1 mW$ at in-plane incidence. Furthermore, a linear dependence of $V_{ph}$ on $P_{in}$, such as in Fig. 8, is observed [62]. We thus conclude that our device operates in the “weak heating” regime.

In order to evaluate Eq. 3 we need to specify the cooling length $\xi$ and the light absorption heat source. In principle, $\xi$, which describes the energy transfer from the electronic system to the optical phonons [63], depends on $T_e(x)$ and $n$. However, under the “weak heating” regime, the $T_e(x)$ dependence can be neglected, and $n$ is uniform in the device (with opposite sign in the two regions of the p-n junction) when the photoresponse is maximal. For these reasons, in our calculations, we assume a constant $\xi = 1 \mu m$, consistent with experimental values [47, 78]. $J(x)$ is calculated from the simulated electric field data as the time average of the electric power dissipation density [97, 98]:

$$J(x) = \frac{1}{2} \omega \epsilon'' |E|^2 = \frac{1}{2} \sigma |E| \cdot E^*$$  (5)

In order to relate $J(x)$ to physically meaningful quantities, we integrate the normal component of the time-averaged Poynting vector over the simulation region and normalize it to a given input power.

After solving Eq. 3 for the $T_e(x)$ fluctuation profile $\Delta T_e(x)$, we take its derivative with respect to $x$ and obtain the second factor of the integrand in Eq. 1. Figs. 9a-c compare the absorption heat source, the resulting $T_e(x)$ profile, and its derivative for a representative GPD in presence of a plasmonic split-gate at different heights over the SLG channel, to an unperturbed fundamental dielectric waveguide mode, where the p-n junction is generated by a transparent (at the chosen $\lambda$) gate, such as a split-gate made from a second SLG at a separation large enough to avoid additional optical losses.

To model $S$ along the channel, we assume that the structure is gated to achieve the maximum $S$ below the gates, as for Eq. 2 (opposite in sign but equal in magnitude for the p-n case) and approximate the gap region with a linear interpolation between the two. Combining both factors in Eq. 1 and computing the integral yields

FIG. 9. a) Joule heat source, b) $T_e$ profile, c) $\Delta T_e$ along SLG channel for a representative GPD with plasmonic gates at different heights above the channel compared to gates transparent at 1.55 $\mu m$.

be sufficiently high to result in $T_e(x) > 100K$. When modeling photoexcited SLG under these conditions, one has to take into account the $T_e(x)$ dependence of all thermodynamic and transport parameters in Eqs. 1-3, i.e. $\mu_e, \sigma(x), \kappa_e(x)$, resulting in a nonlinear system of coupled equations.

The contrasting case, the “weak heating” regime, is characterized by $|\Delta T_e(x)| \ll T_l$. Under this con-
$V_{ph}$ or $R_V$ (if divided by $P_{in}$) as figure of merit to assess different cross section designs:

$$R_V = \frac{|V_{ph}|}{P_{in}} = \frac{1}{P_{in}} \left| \int S(x) \frac{dT_e(x)}{dx} \, dx \right|$$

(6)

To calculate $R_V$ or photovoltage maps as in Fig. 6b, we first generate the $S$ profile for all voltage combinations and then proceed via Eqs. 6 or 1.

We then perform a sweep of gap width, SLG position, gate oxide thickness, gate contact height as a function of $\mu$ in SLG and $\varepsilon_r$. As for Fig. 9, shorter distances between SLG channel and SPP waveguide yield larger signals. Furthermore, the SPP waveguide width affects the expected photovoltage in two ways. 1) narrower confinement improves the field strength at the SLG channel and $R_V$ at the cost of higher propagation losses; 2) the un-gated SLG at the center of the device with not maximum $S$ shrinks with smaller gaps. The device parameters are then chosen based on these trends, taking into account fabrication complexity, robustness to processing-induced deviations (e.g. suppression of fundamental gap plasmon mode below a certain gap width), reliability (e.g. oxide break-down) and outcomes of FDTD simulations on coupling and propagation.

Extracting $E$ and $H$ of the fundamental mode in the hybrid region is sufficient for the design of the device cross-section. However, it does not capture the field distribution along the device, since optical losses, transition from dielectric to plasmonic waveguide, and power exchange between different modes that co-exist in the hybrid region remain unaccounted for. To find a good combination of SLG position and width within the constraints of the cross-section design, targeting a $T_e$ distribution with maximal derivative across the PD, but minimal along the propagation direction over the device length, we perform FDTD of the transition between SiN waveguide and GPD. We construct a device model in the same way as for the eigenmode analysis and adjust the source settings to 1.5-1.6 $\mu$m. We launch the fundamental quasi-TE mode of the SiN waveguide towards the GPD and use frequency domain field monitors (FDFM) with various orientations (parallel and perpendicular to the propagation direction) to track the field and power profiles at different points. Figs. 10d,e plot the field at 3 cross-sections of a representative GPD. We see transitions from injected mode in dielectric waveguide to field distribution resembling the fundamental gap plasmon.

Figs. 10d,e display the electric field intensity from a vertical FDMD monitor along the center line of the GPD, at two wavelengths. As expected for this non-adiabatic transition with fast decrease (<1 $\mu$m) of the taper cross-section down to the target gap size, scattering and reflections at the start of the hybrid region reduce the power at the GPD, but the desired sharp intensity profile over length scales that match the fabricated SLG channel widths is achieved. Consequently, we place the SLG channel 100nm after the SPP structure has reached its final gap width. The comparison of field intensities for 1.5 and 1.6 $\mu$m on the same color scale in Figs. 10d,e reveals a larger peak intensity and a longer interaction length for the latter, which indicates improved coupling efficiency at larger $\lambda$, as for the wavelength-depended $R_V$ in Fig. 6.
REFERENCES

* E-mail: acf26@eng.cam.ac.uk


[31] D. Schall, D. Neumaier, M. Mohsin, B. Chmielak, J....


[94] Z. Fang, Z. Liu, Y. Wang, P. M. Ajayan, P. Nordlander, and N. J. Halas, Nano Lett. 12, 3808 (2012).


[112] G. W. Hanson, J. Appl. Phys. 103, 064302 (2008); ibid. 113, 029902 (2013).

