OPTOELECTRONICS

Fast silicon photodiodes

There is typically a compromise between speed and efficiency when designing silicon photodiodes. Now, researchers have exploited microstructuring to achieve fast and thin devices that are also efficient.

Michael B. Johnston

How much Internet traffic did you generate today? Perhaps more than you realize given the increasing popularity of streaming audio or video content, 'cloud' data storage and social media. It is estimated that approximately 1 zettabyte (10^{21} bytes) of Internet traffic was transmitted globally last year[1], which is the equivalent of about 360 MB per day per person in the world. Much of the long-distance high-volume Internet traffic is transmitted via near-infrared (NIR) light through optical-fibre waveguides. At the end of the optical fibre, the optical signal is turned into an electrical signal, typically for use in silicon-based integrated circuits. However, at present, most receivers for long-distance optical-fibre communications systems are based on photodiodes made from other semiconductors such as $In_xGa_{1-x}As$ or Ge, which are challenging and costly to integrate with silicon CMOS electronics on a single chip.

Now writing in *Nature Photonics*, Gao and co-workers report a silicon photodiode that utilizes photonic microstructuring to develop a photodiode that is both fast and responsive[2]. As their technology is compatible with current silicon CMOS technology, this work could lead to cheaper and better integrated fibre-optic communications systems.

Fibre optics is an excellent technology for transmitting large volumes of information. The good news is that we are, at present, only exploiting a small fraction of the available bandwidth of light (~200 THz for NIR light) in our communications systems. However, to increase the density of information transfer we need to continually develop technologies to modulate the light at higher frequencies, control dispersion in the optical fibres and be able to detect increasingly shorter pulses of light[3].

The purpose of a photodiode is to turn an optical signal into an electrical signal. The electrical current generated by an ideal photodiode illuminated with light of a particular wavelength is directly proportional to the number of photons hitting the device per second. The more closely the electrical signal can follow changes in the intensity of light, the faster the photodiode. This speed or response time of a diode is important as the faster the photodiode, the more closely spaced in time the 1's and 0's (that is, on's and off's) of a digital signal can be, and hence the faster information can be transmitted.

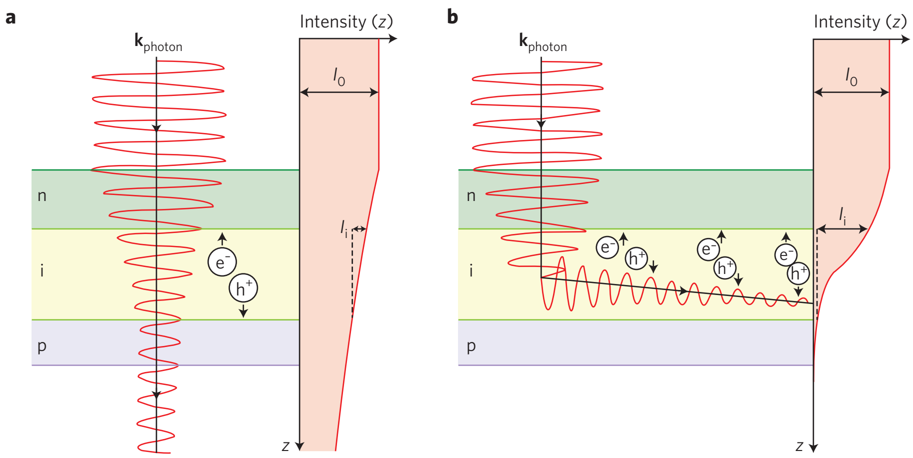

Figure 1 | Schematic of a silicon p–i–n photodiode. The diagram shows propagation of light with intensity I_0 and wavevector $\mathbf{k}_{photon}$ incident normal to the surface of a p–i–n photodiode. Absorption of photons in the high-electric-field i region of the device leads to electron (e^-) and hole (h^+) pair production that rapidly contributes to the photocurrent. I_i is the drop in intensity of light associated with absorption in the i region. **a**, For propagation straight through a conventional thin device only I_i/I_0 of photons are absorbed in the i region, with a significant fraction of photons passing though the active region. **b**, If, however, light can be scattered into a lateral mode, a much higher proportion of photons can be absorbed in the active region, as the lateral extent of a fast p–i–n photodiode is many times the thickness of the i region. For conceptual clarity, changes in refractive index between air and the layers are ignored.

So what limits the speed of a photodiode? Consider the operation of a conventional p–i–n photodiode that consists of a sandwich of thin p-doped, intrinsic (i) and n-doped layers of semiconductors. Light is absorbed according to Beer's law through the p, i and n layers. The transit of photons through the device and the generation of electron–hole pairs occurs on a femtosecond timescale and this does not limit the photodiode speed. Instead, it is the collection of the photogenerated electrons and holes that determines the speed of the diode. In the case of a p–i–n diode, speed is typically limited by diffusion of photogenerated electrons and holes to the high-electric-field 'depletion region' of the device; the 'drift' transit time of electrons and holes under the influence of the electric field in the depletion region; and the capacitance of the device[4].

To minimize the effect of slow diffusion, most light should be absorbed in the high-electric-field i layer, rather than in the n and p layers. In addition, to minimize the drift time and achieve low-voltage operation the i region should be thin. Efficient and high-speed photodiodes can be realized with direct-bandgap semiconductors, such as InGaAs, which have very high absorption coefficients for NIR light. However, silicon is an indirect-bandgap semiconductor and has a very low absorption coefficient for NIR light. For example, NIR light (λ = 850 nm) needs to travel through just 0.14 µm of $In_{0.53}Ga_{0.47}As$ for half the photons to be absorbed compared with 13 µm for silicon. Thus, to create a conventional photodiode from silicon, the i region needs to be 100 times thicker, leading to a long drift transit time and hence much lower speed of operation.

So, is there a way that the speed of silicon diodes can be increased without compromising their efficiency? Gao and co-workers have come up with a photonic solution to this problem. By micropatterning a silicon p–i–n photodiode they were able to redirect normal incident light laterally along the plane of the photodiode. This thereby allowed them to maintain a thin intrinsic region in their photodiode with a high 'vertical' electric field and short drift-time (creating fast photodiodes) while also keeping the absorption, and hence efficiency, of their devices high. Thus, they avoid the compromise between speed and sensitivity by exploiting the larger lateral dimensions of the photodiode to increase the effective absorption of their device (Fig. 1).

There are a number of other methods to achieve silicon-integrated fast photodetectors. Silicon avalanche photodetectors (APDs) utilize a multiplication effect to achieve very high speed and up to single-photon sensitivity[3]. However, these devices typically require higher externally applied voltages than the devices described by Gao and colleagues. An alternate technology is metal–semiconductor–metal (MSM) photodiodes, which consist of interdigitated Schottky contacts on a silicon layer[5]. This MSM technology is highly compatible with CMOS technology, and is relatively high speed owing to low capacitance, however, the sensitivity is much lower than that of APD or p–i–n photodiodes.

So why bother with silicon if direct-gap semiconductors such as InGaAs can already provide us with high-speed, efficient photodiodes? The issue is with compatibility with integrated circuits that are dominated by silicon-based technologies, such as CMOS. Silicon photodiodes integrate well with other silicon-based components, even on single chips. On the other hand, Ge or InGaAs photodetectors are usually built as discrete components, or are combined with CMOS integrated circuits by wafer-bonding or flip-chip techniques that add expensive additional steps in the device fabrication process. However, alternative methods of direct growth of Ge on Si (ref. 6) or III–V nanowires on Si (ref. 7) could lead to more integrated competing technologies.

Probably the biggest drawback of silicon photodiode technologies for optical-fibre communications is the bandgap of silicon. Silicon's bandgap energy is 1.12 eV at room temperature, thereby limiting efficient photodiode operation to wavelengths <1,100 nm. Although 800–900 nm was originally used for optical-fibre communications, it is uncommon now for long-distance links owing to high attenuation of optical fibres at these wavelengths (~3 dB km^{-1} at 850 nm) and less-mature fibre amplifier technologies in this band as compared with low-dispersion 1,300 nm (O band) and low-loss 1,550 nm (C band) wavelength regions. The C band is most commonly used owing to excellent erbium-based fibre amplifiers and low attenuation (~0.2 dB km^{-1}). However, for shorter communication links, such as 'last mile' internet links, device-to-device connections or intranets, where cheap single-chip solutions are needed, sensitive and fast CMOS-integrated silicon photodiodes look most promising. ❐

Michael B. Johnston is in the Department of Physics, University of Oxford, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK.
e-mail: michael.johnston@physics.ox.ac.uk

References

1. *White Paper on The Zettabyte Era: Trends and Analysis* (Cisco, 2016).
2. Gao, Y. *et al. Nat. Photon.* **11,** 301–308 (2017).
3. Caplan, D. O. *J. Opt. Fiber Commun. Rep.* **4,** 255–362 (2007).
4. Sze, S. M. & Ng, K. K. *Physics of Semiconductor Devices* 3rd edn (Wiley, 2016).
5. Eng, P. C., Song, S. & Ping, B. *Nanophotonics* **4,** 277–302 (2015).
6. Michel, J., Liu, J. & Kimerling, L. C. *Nat. Photon.* **4,** 527–534 (2010).
7. Koblmüller, G. & Abstreiter, G. *Phys. Status Solidi RRL* **8,** 11–30 (2014).

METAMATERIALS

A low-energy Cherenkov glow

Hyperbolic metamaterials are shown to enable the emission of Cherenkov radiation from low-energy charged particles travelling at slow speeds. The achievement could lead to new forms of light sources and detectors.

Mário Silveirinha

It is well known that when fast charged particles travel inside a material with a speed exceeding the medium's phase velocity for light, they spontaneously emit electromagnetic radiation at the cost of a reduction in their kinetic energy. The effect is known as Cherenkov radiation and owes its name to Pavel Cherenkov — a Russian physicist who in the mid-1930s was a student working under Sergey Vavilov at the Lebedev Institute in Moscow and was studying the luminescence emitted from liquids under the incidence of highly energetic gamma rays[1]. This breakthrough earned him the Nobel Prize in Physics in 1958 alongside Ilya Frank and Igor Tamm, who helped explain the phenomenon. Often, the Cherenkov effect is associated with the blue glow of the water surrounding the core of nuclear reactors created by highly energetic beta particles released by the radioactive decay of nuclei. In fact, traditionally, the Cherenkov effect has various applications in the context of high-energy physics, for example, to detect cosmic ray events or in particle accelerators. Charged particles also emit radiation when they travel in a vacuum near a grating, an effect discovered by Smith and Purcell[2]. Furthermore, neutral polarizable atoms may spontaneously radiate away part of their kinetic energy when they travel near a smooth planar metal surface[3]. However, all of the above cited phenomena are generally relevant only when the particles travel with ultra-relativistic speeds.

Now, writing in *Nature Photonics*, Fang Liu and co-workers from