manufacturing have led to drastically falling PV module prices and an oversupplied market. Ironically, these factors combined have made it difficult for emerging PV technologies to enter the market. Currently, a new technology needs to meet both <US\$0.40 W^{-1} module prices and an insured guarantee of a 25-year useful product life to compete with the incumbent silicon module technology.

If, in coming years, demand for solar were to increase faster than manufacturing of incumbent technologies could support (for example, due to an imperative to quickly reduce carbon emissions), this could change the situation and create a shortage in the market and an opportunity for emerging technologies. Continued PV R&D to create technology diversity is useful as an insurance policy to address future energy crises, enable new applications as they become technologically feasible and prepare for a wide range of scenarios that may require flexibility in the scaling of manufacturing capacity to meet demand cycles.

The threat of climate change combined with the recent advances in clean-energy technology development has motivated nations in the developed and developing worlds to come together to invest in clean energy. A decade ago, PV was written off by many as being too expensive to be useful, but the progress in that decade now positions PV to be a solid contributor going forward.

As we enter this new era, in the next decade we expect that increasing pressure to improve energy security, reduce local air pollution and avoid global climate change will enable electricity generated from solar to grow from ~1% to >10% of our electricity base and the fraction of total energy supplied by solar to grow even more. Understanding both the opportunity for massive growth (Fig. 2a) and the reality that PV is already supplying a substantial fraction of new electricity-generating capacity (Fig. 2c) will help researchers identify priorities for the coming decade. The photonics research community can play a critical role — by refining and broadening the ways in which photons can be harvested — in enabling this transition. ❐

Sarah Kurtz and Nancy Haegel are at the National Renewable Energy Laboratory, Golden, Colorado 80401, USA. Ronald Sinton is at Sinton Instruments, Boulder, Colorado 80301, USA Robert Margolis is at the National Renewable Energy Laboratory, Washington DC 20024, USA.
e-mail: sarah.kurtz@nrel.gov; nancy.haegel@nrel.gov; ron@sintoninstruments.com; robert.margolis@nrel.gov

References

1. Mints, P. *Photovoltaic Manufacturer Shipments: Capacity, Price & Revenues 2015/2016* (SPV Market Research, 2016).
2. Feldman, D., Boff, D. & Margolis, R. *Q2/Q3 2016 Solar Industry Update, NREL Technical Report NREL/PR-6A20-67246* (US Department of Energy, 2016); www.nrel.gov/docs/fy17osti/67246.pdf
3. Mints, P. *SPV Market Research Solar Flare Report — Issue 4* (August, 2016).
4. Barbose, G. & Darghouth, N. *Tracking the Sun IX: The Installed Price of Residential and Non-Residential Photovoltaic Systems in the United States* (Lawrence Berkeley National Laboratory, 2016).
5. US Energy Information Administration *Electricity Data Browser* go.nature.com/2gxrvBK (accessed 16 October 2016).
6. Loftus, P. J. *et al. WIREs Climate Change* **6,** 93–112 (2015).
7. Kurtz, S. *et al. Nat. Photon.* **10,** 141–142 (2016).
8. Smith, D. *et al.* Silicon solar cells with total area efficiency above 25%. In *43rd IEEE Photovoltaic Specialists' Conf. (PVSC)* 3351–3355 (IEEE, 2016).
9. Roberts, C. SunPower solar panel again raises the bar in efficiency: technology leader achieves 24.1 percent, a new world record. *PR Newswire* (27 June 2016); go.nature.com/2fkt01d
10. Burst, J. M. *et al. Nat. Energy* **1,** 16055 (2016).
11. Zhao, Y. *et al. Nat. Energy* **1,** 16067 (2016).
12. US Energy Information Administration (EIA) *International Energy Statistics* http://go.nature.com/2f0pLRr (accessed 17 October 2016).
13. Renewable Energy Policy Network for the 21st Century (REN21) *Renewables 2016 Global Status Report* (2016); http://go.nature.com/2e3PUKx

---

Communications expands its space

Joseph M. Kahn and David A. B. Miller

Optical communication systems have traditionally sent the most information possible through a few spatial channels to minimize cost and maximize density. Energy constraints now compel systems at the longest and shortest distances to employ a new strategy of using more spatial channels, each carrying less data.

Optical communications has been remarkably successful in supporting the explosive growth of information technology over the past three decades. A key strategy has been to force ever-larger amounts of information over a small number of spatial channels, whether they be modes in fibres or beams in free space. However, limits on energy make this approach suboptimal at the longest distances and preclude it altogether at the shortest, threatening to slow the advance of information technology. The solution in both cases is to employ more spatial channels, while transmitting less information in each.

Today's transport systems carry tens of terabits per second over thousands of kilometres in each single-mode fibre (SMF), forming the backbone of the global Internet. Cost is minimized by maximizing the bit rate per transceiver and the aggregate throughput per fibre. Each spatial channel — each SMF — is fully exploited by multiplexing data signals into all available physical dimensions or degrees of freedom, including numerous wavelengths, two polarizations and two field quadratures. Transceivers now employ coherent detection and digital signal processing (DSP)-based compensation of dispersion and polarization effects to optimize transmission and will soon compensate even the Kerr nonlinearity of the silica fibre[1]. As a result, the throughput per fibre is now approaching information-theoretic capacity limits imposed by amplifier noise and Kerr nonlinearity[2], creating a looming 'capacity crunch'[3].

Accommodating a projected 30-fold traffic growth over the next decade will require space-division multiplexing[4] (SDM) to create more spatial channels for multiplexing information. SDM can be realized by employing multimode fibre (MMF), multicore fibre (MCF), or simply multiple SMFs. Multiple SMFs or uncoupled-core MCFs are the most straightforward options, as MMF requires multi-input multi-output (MIMO) DSP to compensate for modal coupling and dispersion. However, MMFs may simplify the scaling of optical switching components[5] and mitigate the Kerr nonlinearity to increase information capacity per mode[6].

SDM will solve the capacity crunch, but transport systems still face a cost and

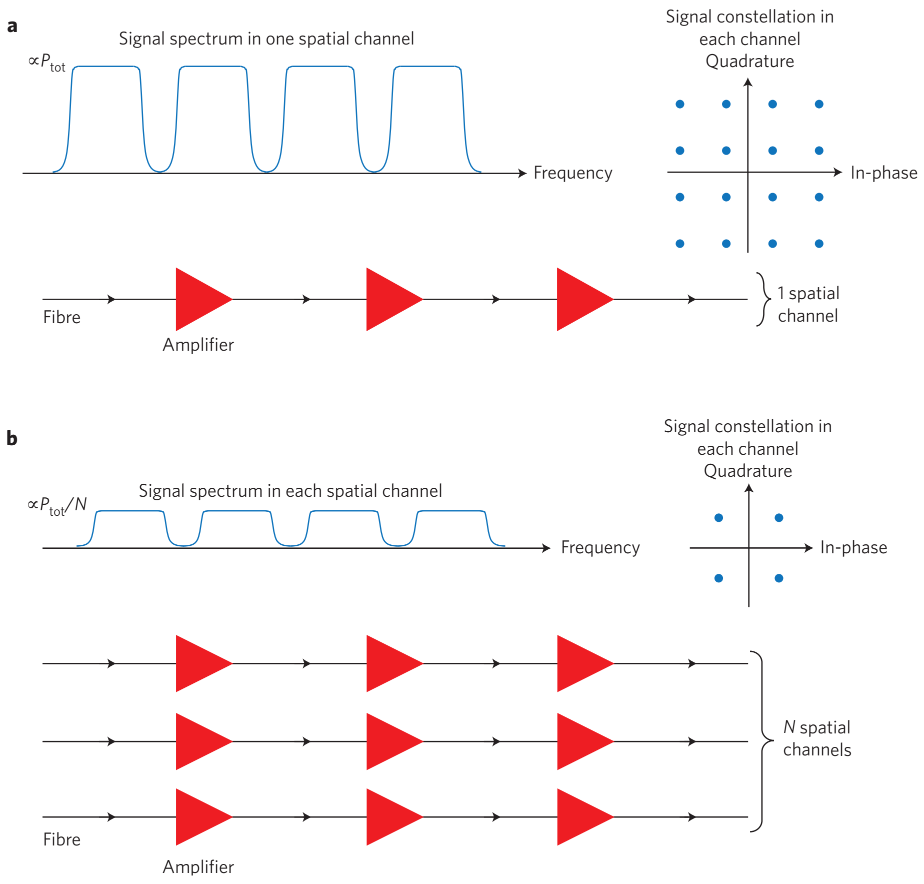

Figure 1 | Ultra-long-haul submarine transport system designs. **a**, Using one spatial channel with high power and high-order modulation. **b**, Using N spatial channels ($N > 1$) with low power and low-order modulation. The design in **b** may yield higher capacity when total signal power over all spatial channels P_{tot} is constrained, because capacity scales linearly with N but only logarithmically with signal power per spatial channel P_{tot}/N. In the example shown, **a** yields $\log_2 16 = 4$ bits per symbol, whereas **b** yields $3 \times \log_2 4 = 6$ bits per symbol.

energy crunch that must be addressed without compromising performance. One solution lies in dense integration of arrays of transceivers and other subsystems. Hybrid integration can combine 'best-in-breed' components, including glass passives, thin-film lithium niobate modulators, III–V semiconductor or doped glass gain media, and silicon circuits. Another solution lies in sharing synchronization functions across wavelengths or spatial channels, provided they all propagate together from source to destination[5]. Transceiver power consumption can be reduced from hundreds to tens of picojoules per bit, particularly in shorter links requiring fewer DSP operations.

Submarine transport cables, up to tens of thousands of kilometres long, connect continents and coastal cities. Like their terrestrial counterparts, to minimize cost and hardware requirements, submarine systems have evolved to exploit each spatial channel aggressively. Submarine cables use SMFs with large modal effective area to minimize Kerr nonlinearity, which enables higher launched signal power to maximize data throughput per fibre. This traditional approach can be extended by replacing segments of the SMF by MMF, propagating signals in the fundamental mode, which has a larger effective area than SMF[7].

Long submarine cables must now exploit spatial channels in an entirely different way to continue scaling up the throughput per cable. Feed voltage constraints in cables longer than about 5,000 km limit the electrical power available to amplifiers, thus limiting the optical power propagating in the cable. As a result, the traditional approach of maximizing launched power per spatial channel can be suboptimal. Suppose a total signal power P_{tot} is divided among N spatial channels. The information-theoretic capacity scales linearly in N and only logarithmically in the signal power per channel, P_{tot}/N. Capacity may be maximized by increasing N and transmitting at a lower rate in each spatial channel. This principle was understood as early as 1973[8] and brought to light again recently[9]. Figure 1 shows an example in which dividing the power among $N = 3$ spatial channels increases the bit rate by a factor of 1.5. This approach was recently used to transmit 105.1 Tb s^{-1} over 14,350 km through 12-core MCF[10], representing a 2.9-fold increase in throughput over one SMF using the same amplifier pump power.

Take to the skies

Free-space optical (FSO) links have long been considered promising for airborne or space communications. FSO systems exploit a huge, unregulated bandwidth, while short optical wavelengths allow the use of narrow beams that can carry significant energy over long distances. As bit rate requirements have increased, FSO links have become a preferred technology over radio links for satellites and deep-space probes.

Providing Internet access to homes and businesses via fibre requires huge infrastructure investment, posing economic challenges even in developed nations. FSO systems offer an alternative solution and may soon provide access in remote areas or developing nations. By 2018, a planned constellation of up to 12 satellites in medium Earth orbit, interconnected by FSO links, will provide 72 FSO links at 100 Gb s^{-1} to the ground[11]. An experimental system has demonstrated interconnection of unmanned aerial vehicles (UAVs) at 20 km altitude by using FSO links up to 300 km long to form a backbone network for wireless access[12].

Wavelength-division multiplexing (WDM) is readily employed in these long-range FSO systems. SDM methods based on multimode beams cannot be used, however. Considering diffraction, the maximum number of spatial channels with efficient coupling in an FSO link is approximately[13] $N_{max} \approx (A_T A_R)/(\lambda^2 L^2)$, assuming transmitter and receiver areas A_T and A_R, separation L and optical wavelength λ. Estimates of N_{max} less than unity are obtained even with optimistic parameters for satellite systems (transmitter and receiver diameters $D_T = D_R = 1$ m, $L = 1{,}000$ km, $\lambda = 1{,}550$ nm, $N_{max} \approx 0.3$), or UAV systems ($D_T = D_R = 0.1$ m, $L = 10$ km, $\lambda = 1{,}550$ nm, $N_{max} \approx 0.3$). An estimate of N_{max} less than unity indicates that the receiver can capture only a fraction of the transmitted beam, and cannot separate additional spatial modes. In shorter-range FSO systems, values of $N_{\max}$ above unity may be obtained. Numerous spatial channels can be exploited fully using any complete set of modes, such as parallel Gaussian beams, for example. Nevertheless, in FSO systems at any length scale, beams between different transmitter–receiver pairs may propagate in parallel or even cross with negligible interference, a form of SDM that exploits the many spatial channels available in free space.

The data centre challenge

Data centres, each containing up to tens of thousands of servers supporting cloud applications and web sites, represent about 1.5% of humanity's electrical energy consumption[14], a share that is growing. Each bit sent to a data centre may induce the servers to exchange many thousands of bits to support complex algorithms. Server racks are interconnected in a multi-tiered network of electrical packet switches and optical links, with network throughput as large as 1 Pb s^{-1} (ref. 15). As the switches scale to higher port speeds and port counts, the optical links must achieve higher bit rates and densities under tight cost and power constraints.

Optical links between server racks are evolving to use fewer spatial channels to satisfy these requirements. While past links used up to 10 parallel MMFs, new designs use SMFs carrying up to eight wavelengths. The new links achieve bit rates up to 100 Gb s^{-1} per wavelength using multilevel intensity modulation with direct detection, an approach yielding poor optical power efficiency that is difficult to scale further. Future data centre links should exploit all four dimensions of the optical field (polarizations and quadratures), using coherent detection to improve receiver sensitivity by about 20 dB. Approaches such as optical phase locking[16] may reduce energy consumption to about 10 pJ b^{-1}.

Multi-tiered data centre networks consume significant energy in electrical switching matrices and in (de)serializing and synchronizing bit streams at numerous switch and transceiver interfaces. Nearer-term solutions include increasing switch port counts to reduce the number of network tiers, and integrating transceivers with switches to reduce the number of interfaces. Hybrid electrical–optical switching[17] may offer a longer-term solution. It combines electrical switching for short data packets with optical switching for connections lasting at least tens of milliseconds. The optical switches, based on microelectromechanical systems, exploit free space to achieve interconnection with minimal crosstalk, like the satellite- and UAV-based FSO systems described above.

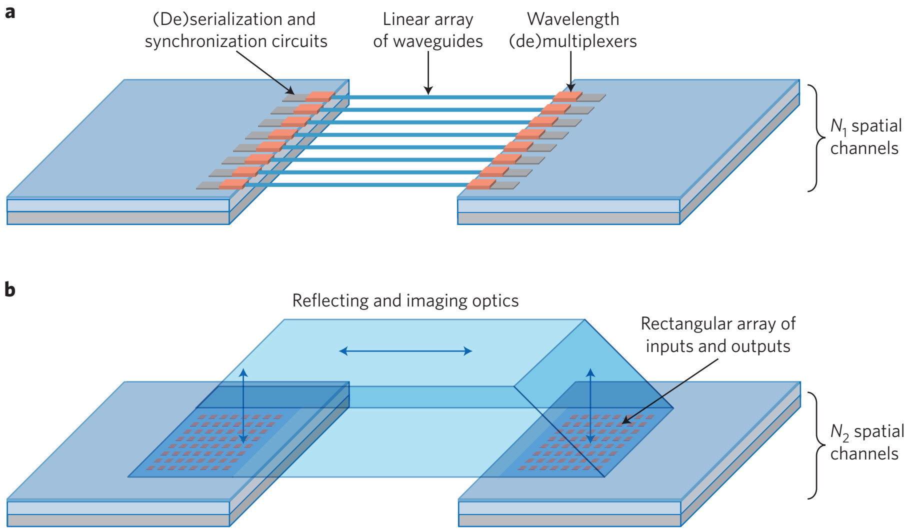

Figure 2 | Short-reach chip-to-chip interconnect designs. **a**, Using N_1 spatial channels with high serial bit rates and multiple wavelengths. **b**, Using N_2 spatial channels ($N_2 >> N_1$) with low serial bit rates and one wavelength. The design in **b** may consume lower power because it avoids (de)serialization and synchronization.

Short-reach interconnects — in Internet routers, data centre server racks and the servers themselves — represent a major fraction of the total energy consumption of these digital systems, and are a major impediment to scaling their density and performance. Overcoming these interconnect bottlenecks is critical for the continued growth of all information technology.

Electrical interconnects consume energy in the charging and discharging of wires. The energy cost to communicate between a gate and its near neighbours is comparable to the energy required to switch the gate's logic state[13]. Over longer distances, however, interconnection energy dominates over logic energy.

Optical interconnects can reduce energy consumption, enabling higher throughput density. One concept for a chip-to-chip optical interconnect, shown in Fig. 2a, combines SDM, WDM and time-division multiplexing (TDM). A ribbon of waveguides or fibres provides multiple spatial channels, each carrying multiple wavelengths. Each wavelength carries a high-speed signal time-multiplexed from many parallel low-speed electrical signals. Such an approach can achieve an energy consumption of just a few picojoules per bit, lower than comparable short electrical interconnects[13], but is difficult to scale to much lower energy and higher throughput density. While optoelectronic devices consume a few picojoules per bit, prospects exist for scaling them to levels of femtojoules per bit[13]. A more intractable problem is that TDM from the chip's native interconnect speed, typically about 2 Gb s^{-1} to serial speeds of tens of gigabits per second, requires high-speed receivers and (de)serialization and clock synchronization circuits, which dissipate several picojoules per bit. This TDM is required because of a limited number of wavelength channels, and because the linear array of waveguides provides a limited number of spatial channels N_1. For example, a 1 cm^2 chip can support at most $N_1 = 1{,}000$ spatial channels per edge, assuming the waveguides, WDM and TDM components can be packed with 10 μm pitch.

A radical solution for reducing energy per bit while increasing throughput density uses imaging optics to create a massive number of parallel spatial channels for FSO communication between chips. Transmitting in each channel at a low speed may obviate the need for high-speed receiver and (de)serialization and synchronization circuits, reducing energy consumption substantially[13]. This approach, shown in Fig. 2b, uses a rectangular array of optical inputs and outputs to provide N_2 spatial dimensions, where $N_2 >> N_1$. For example, a 1 cm^2 chip with 10 μm input/output pitch could provide $N_2 = 10^6$ spatial channels. (The diffraction limit $N_{\max} \approx (A_T A_R)/(\lambda^2 L^2)$ yields $N_{\max}$ of at least 10^6 for chip-to-chip spacing up to 12 cm, assuming a symmetric single-lens imaging system, $A_T = A_R = 1\ \mathrm{cm}^2$ and $\lambda = 1{,}550$ nm.) Without using any

WDM, and modulation at only 2 Gb s^{-1}, this could provide a throughput of up to $10^6 \times 2$ Gb s^{-1} = 2 Pb s^{-1}. This approach relies on a waveguide-based interposer that maps between the optical inputs and outputs, shown in Fig. 2b, and detectors and modulators (or emitters) located near data registers on the chip[13].

Optical communications must support the continued growth of information technology in the coming decade by providing orders of magnitude higher throughput, while achieving unprecedented density, energy and cost per bit. Emerging integration technologies will help achieve these goals. Spatial multiplexing will also play an indispensable role. ❒

Joseph M. Kahn and David A. B. Miller are at the Edward L. Ginzton Laboratory, Stanford University, 348 Via Pueblo Mall, Stanford, California 94305, USA. e-mail: jmk@ee.stanford.edu; dabm@ee.stanford.edu

References

1. Ip, E. & Kahn, J. M. *J. Lightwave Technol.* **28,** 502–519 (2010).
2. Essiambre, R., Kramer, G., Winzer, P. J., Foschini, G. J. & Goebel, B. *J. Lightwave Technol.* **28,** 662–701 (2010).
3. Chraplyvy, A. R. The coming capacity crunch. In *Proc. European Conf. Optical Commun.* 1.0.2 (IEEE, 2009).
4. Richardson, D. J., Fini, J. M. & Nelson, L. E. *Nat. Photon.* **7,** 354–362 (2013).
5. Arık, S. Ö., Ho, K.-P. & Kahn, J. M. *Opt. Express* **22,** 29868–29887 (2014).
6. Antonelli, C., Mecozzi, A., Golani, O. & Shtaif, M. Inter-modal nonlinear interference in SDM systems and its impact on information capacity. In *2016 IEEE Photonics Soc. Summer Topical Meeting Series* 10–11 (IEEE, 2016).
7. Downie, J. D. *et al.* Quasi-single-mode transmission for long-haul and submarine optical communications. In *Conf. Lasers Electro-Optics* SM4F.6 (OSA, 2016).
8. Caves, C. M. & Drummond, P. D. *Rev. Mod. Phys.* **66,** 481–537 (1994).
9. Li, G., Bai, N., Zhao, N. & Xia, C. *Adv. Opt. Photon.* **6,** 413–487 (2014).
10. Turukhin, A. *et al.* 105.1 Tb/s power-efficient transmission over 14,350 km using a 12-core fiber. In *Optical Fiber Commun. Conf.* Th4C.1 (OSA, 2016).
11. http://laserlightcomms.com/halo_network.php
12. Hempel, J. Inside Facebook's ambitious plan to connect the whole world. *Wired* (19 January 2016); go.nature.com/2gwzoaV
13. Miller, D. A. B. Preprint at https://arxiv.org/abs/1609.05510 (2016).
14. Heddeghem, W. V. *et al. Comput. Commun.* **50,** 64–76 (2014).
15. Singh, A. *et al.* Jupiter rising: a decade of Clos topologies and centralized control in Google's data center network. In *SIGCOMM '15* 183–197 (2015).
16. Park, H.-C. *et al. Opt. Express* **20,** B197–B203 (2012).
17. Farrington, N. *et al.* Helios: a hybrid electrical/optical switch architecture for modular data centers. In *SIGCOMM '10* 339–350 (2010).

Unrelenting plasmons

Antonio I. Fernández-Domínguez, Francisco J. García-Vidal and Luis Martín-Moreno

Worldwide research efforts on plasmonics and metamaterials have been growing exponentially for the past ten years. Will this course hold true over the next decade?

Following a brief historic introduction to plasmons, their useful properties and early applications, we highlight some of the key advances in the field over the past decade. We then discuss new directions for the future, such as the use of 2D materials and strong coupling phenomena, which are likely to shape the field over the next ten years.

For centuries, metals were employed in optical applications only as mirrors and gratings. New vistas opened up in the late 1970s and early 1980s with the discovery of surface-enhanced Raman scattering and the use of surface plasmon (SP) resonances for sensing. However, it was not until the 1990s, with the appearance of accurate and reliable nanofabrication techniques, that plasmonics blossomed[1]. Initially, the attention focused on the exploitation of SPs (collective electronic oscillations at the surface of metals) for sensing, subwavelength waveguiding and extraordinary optical transmission[2]. Since then, the scientific and technological interest in SPs has expanded. Correspondingly, as illustrated in Fig. 1, the number of publications in the field has increased in a steady exponential fashion for more than two decades, and the momentum driving plasmonics research looks set to continue.

In a simplified picture, there are five distinctive characteristics that make SPs attractive: their ability to concentrate light beyond the diffraction limit; their ability to modify the local density of photonic states; their ultrafast response; their environmental sensitivity; and their flexibility in design. The main factor limiting their use is the high optical absorption inherent to metals. Therefore, the quest for minimizing dissipative damping has been a crucial driving force for plasmonics.

Beyond noble metals

The endeavour of improving plasmonic performance, most notably by reduced optical loss, has been accompanied by the pursuit of materials that outperform noble metals in specific optical functionalities. Different options, such as aluminium, metallic alloys and heavily doped semiconductors, have all been investigated and have advanced our understanding and broadened the catalogue of available plasmonic platforms. However, noble metals may still have more to offer. Recently, it has been shown that their plasmonic characteristics can be significantly improved using standard surface science fabrication protocols[3].

The advent of graphene has inspired research into 2D materials defined by their atomic thickness, some of which are very promising for applications in both the terahertz and mid-infrared regimes. Doped graphene supports SPs in these frequency ranges, featuring out-of-plane decay lengths and in-plane wavelengths that can be three orders of magnitude smaller than free-space radiation. This fact, together with their large electrical tunability, makes graphene plasmons excellent candidates for resonators and sensors. Despite the inherent absorptive character of graphene, there are reasons for being optimistic about the prospects of reducing damping in graphene-based plasmonics. For instance, the encapsulation of graphene within boron nitride (BN) films has increased SP propagation lengths by one order of magnitude[4]. Moreover, graphene is not the only member of the set of 2D plasmonic materials. Apart from BN, the interest in other media, such as semiconducting MoS$_2$ or black phosphorus, has been mounting over the past year or so. We envisage that plasmonic materials of atomic thickness will be a very active and fruitful research area over the next ten years.

Sensing

The efficiency of SP-based phenomena in metallic platforms has been pushed to unprecedented limits, opening the way for a myriad of applications. For example, through the implementation of nanoscale