3.1 Introduction

Two-dimensional (2D) materials are atomically thin films originally derived from layered crystals such as graphite, hexagonal boron nitride (h-BN), and the family of transition metal dichalcogenides (TMDs, such as MoS$_2$, WSe$_2$, MoTe$_2$, and others). Atomic planes in such crystals are weakly stacked on each other by van der Waals forces so that they can be easily isolated, leaving no dangling bonds. This is in distinct contrast to their counterpart, quasi-low-dimensional semiconductors, which are produced by thinning down conventional bulk or epitaxial crystals. The lack of dangling bonds at the interfaces and surfaces of 2D materials enables new devices with unprecedented performance.

The merits of 2D materials are not limited to the absence of dangling bonds. They also show a high degree of mechanical stability, as well as unique electronic and optoelectronic properties. This makes 2D materials highly suitable for a wide range of applications, from high performance transistors to extremely sensitive photodetectors and sensors. In addition, the few-atom thickness of many of these novel devices and systems and the low temperatures required during the device fabrication allow their seamless integration with conventional silicon electronics. It is possible to fabricate many of these devices on top of a fully fabricated silicon CMOS wafer without degrading the Si transistors underneath, bringing new functionality to the silicon chip. This integration process can be repeated numerous times to build complex 3D systems.

This chapter provides an overview of the technology and advantages of the heterogeneous integration of various 2D materials-based devices with a standard
Si platform. Some of the system-level examples that will be discussed include chemical and infrared sensors, large area electronics, and optical communication systems. Section 3.1 describes the advantages of wide bandgap MoS$_2$ and other TMDs over conventional semiconductors for aggressive scaling of the transistors channel length as well as for ultra-low power applications. Section 3.2 summarizes the research on graphene-based infrared sensors and the methods for building such sensors on top of conventional Si-CMOS readout chips. Section 3.3 focuses on the heterogeneous integration of 2D materials with Si nanophotonics, while Sect. 3.4 discusses different approaches on how 2D materials can be used as chemical or biological sensors.

3.2 Scaling and Integration of MoS$_2$ Transistors

3.2.1 MoS$_2$ Transistors for Ultimate Scaling and Power Gating

As the channel length of transistors has shrunk over the years, short-channel effects have become a major limiting factor to further the transistor miniaturization. Current state-of-the-art silicon-based transistors at the 14-nm technology node have channel lengths of around 20 nm, and several technological reasons are compromising further reductions in the channel length. In addition to the inherent difficulties of high-resolution lithography, the direct source-drain tunneling is expected to become a very significant fraction of the off-state current in sub-10-nm silicon transistors, dominating in this way the standby power. Therefore, new transistor structures that reduce the direct source-drain tunneling are needed to achieve further reductions in the transistor channel length. Transistors based on high mobility III–V materials [1, 2], nanowire field-effect transistors (FETs) [3, 4], internal gain FETs [5, 6] (such as negative capacitance devices), and tunnel FETs are among those that have been considered to date. More recently, layered 2D semiconducting crystals of transition metal dichalcogenides (TMDs), such as molybdenum disulfide (MoS$_2$) and tungsten diselenide (WSe$_2$), have also been proposed to enable aggressive miniaturization of FETs [7–10]. In addition to the reduced direct source-drain tunneling current possible in these wide-bandgap materials, the atomically thin body of these novel semiconductor materials is expected to improve the transport properties in the channel thanks to the lack of dangling bonds. Some studies have reported, for example, that single and few layer MoS$_2$ could potentially outperform ultrashort channel and ultrathin body silicon at similar thicknesses [11].

Moreover, the atomically thin body thickness of TMDs also improves the gate modulation efficiency. This can be seen in their characteristic scaling length ($\lambda = \sqrt{\frac{t_{\text{semi}}}{\varepsilon_{\text{ox}} t_{\text{ox}}, t_{\text{semi}}}}$), which determines important short channel effects such as the drain-induced barrier lowering (DIBL) and the degradation of the subthreshold swing (SS). In particular, MoS$_2$ has low dielectric constant $\varepsilon = 4 - 7$ [12, 13] and an atomically thin body ($t_{\text{semi}} = 0.7$ nm × number of layers) which facilitate the decrease of $\lambda$ while its relatively high bandgap energy (1.85 eV for a monolayer) and high effective mass allow for a high on/off current ratio ($I_{\text{on}}/I_{\text{off}}$) via reduction
of direct source–drain tunneling. These features make MoS$_2$ in particular, and wide-bandgap 2D semiconductors in general, highly desirable for low-power subthreshold electronics.

Ni et al. [14] used first principles quantum transport investigations to predict that monolayer MoS$_2$ FETs would show good performance at sub-10-nm channel lengths and also display small SS values, comparable to the current best sub-10-nm silicon FETs. In addition, its large bandgap makes MoS$_2$ an excellent semiconductor for low power applications, while its ability to form atomically thin films allows excellent electrostatic gate control over the FET channel.

To experimentally demonstrate and benchmark MoS$_2$ transistors with channel lengths below 10 nm, two important challenges need to be overcome. Firstly, a suitable lithography technology is required; secondly, a low-contact resistance is needed for the source and drain to prevent the channel resistance from dominating the device behavior.

Liu et al. [15] demonstrated channel length scaling in MoS$_2$ FETs from 2 $\mu$m down to 50 nm in devices built with a 300-nm SiO$_2$ gate dielectric. Despite the thick dielectric oxide layer used in this study, short channel effects were limited for channel lengths as low as 100 nm. However, devices with channels below 100 nm showed a high off current and DIBL (Fig. 3.1).

![Fig. 3.1](image_url)

**Fig. 3.1** (a, b) Transfer and output characteristics of a 12-nm layer of MoS$_2$ with a channel length of 50 nm. (c, d) Channel length dependence of the current on/off ratio and DIBL for MoS$_2$ devices with channel thickness of 5 and 12 nm. Liu et al. [15]
Fig. 3.2 (a) Schematic cross section of a short channel double-gate (DG) MoS$_2$ FET with graphene source/drain contacts. (b) AFM images showing 10, 15, 20 nm graphene slits that define the channel length. (c) Transfer characteristics ($I_d$–$V_g$) for a 15-nm 4-layer DG-MoS$_2$ FET with $SS_{\text{min}} = 90$ mV/dec and $I_{\text{off}} < 10$ pA. Nourbakhsh et al. [9]

Fig. 3.3 (a) SEM image of MoS$_2$ channel lengths ranging from 10 to 80 nm after deposition of Ni contacts. (b, c) Output and transfer characteristics of the 10-nm nominal channel length MoS$_2$ FET built on a 7.5-nm HfO$_2$ gate dielectric. Yang et al. [10]

Similar to Si and III-V FETs, reducing the channel length of MoS$_2$ FETs toward the sub-10 nm regime requires state-of-the-art high-k dielectric thin films to be used in the place of SiO$_2$ dielectrics. Nourbakhsh et al. [9] demonstrated a MoS$_2$ FET with a channel length as low as 15 nm, using graphene as the immediate source/drain contacts and 10 nm HfO$_2$ as the gate dielectric. As shown in Fig. 3.2, short channel effects were limited in this device, which showed high on/off ratio of $10^6$ and an $SS_{\text{min}}$ of 90 mV/dec.

This performance indicated that further scaling to a sub-10-nm channel length might be possible. In a different attempt Yang et al. [10] successfully reduced the MoS$_2$ channel length to 10 nm, in a device with a Ni source/drain contacts and a 7.5-nm HfO$_2$ gate dielectric. The device maintained its low off-current to about 100 pA/µm (Fig. 3.3).

Another approach for aggressive scaling is to extend the channels of transistors in the third dimension, including nanowire gate-all-around (NW GAA) FETs [3, 4], finFETs [16], etc. A surface free of dangling bonds and a low-temperature synthesis method also make MoS$_2$ a promising candidate as the channel material in finFETs.
Chen et al. have demonstrated a CMOS-compatible process for few-layer MoS$_2$/Si hybrid finFETs with improved on-current and good threshold voltage ($V_t$) matching [17]. In subsequent work, the same group improved the process and realized 4-nm-thick ultrathin body MoS$_2$ finFETs in the sub-5 nm technology node with reduced contact resistance and good $V_t$ control with back-gate biasing [18].

In all of the aforementioned devices, standard lithography, including e-beam techniques, was used to define the source/drain electrodes in the MoS$_2$ FETs. Realizing ultra-short channels in MoS$_2$ transistors using lithography can be challenging. Electron beam lithography can potentially provide sub-10-nm patterning resolution; however, it has a low throughput and it is difficult to control at these dimensions. Alternatively, Nourbakhsh et al. [19] used a directed self-assembly (DSA) of block copolymers (BCPs) to push MoS$_2$ channel lengths to the sub-10 nm regime. Unlike conventional lithography methods, DSA-BCP is a bottom-up approach in which smaller building block molecules associate with each other in a coordinated fashion to form more complex supramolecules. Using this fabrication approach to MoS$_2$ FETs, a MoS$_2$ layer was patterned with metallic and semiconducting phases to achieve channel lengths as low as 7.5 nm.

The stable metallic phase of MoS$_2$ can be achieved by chemically treating the semiconducting phase with $n$-butyllithium solution. As shown in Fig. 3.4, the MoS$_2$ channel was first patterned with BCP, then a chemical treatment was used to convert the MoS$_2$ film to a chain of alternating metallic and semiconducting MoS$_2$. The semiconducting regions, 7.5 nm across, acted as the FET channel and the metallic portions acted as the immediate source/drain contacts. This method produced a chain of MoS$_2$ FETs with a record-low channel length of 7.5 nm. This device structure permitted experimental probing of the transport properties of MoS$_2$ in the sub-10-nm channel length regime for the first time. As predicted, MoS$_2$ FETs demonstrated superior subthreshold characteristics with lower off-currents.

![Fig. 3.4](image-url)

Fig. 3.4  (a) SEM image showing lines of BCP (polystyrene-b-dimethylsiloxane) with a 15 nm pitch formed on a MoS$_2$ film contacted by a pair of Au electrodes. (b) Schematic of short channel FET comprising a semiconducting (2H) MoS$_2$ channel contacted to two adjacent metallic (1 T') MoS$_2$ regions that form internal source/drain contacts. (c) $I_d$-$V_g$ of the final MoS$_2$ device (after semiconductor to metallic MoS$_2$ phase transition) the chain transistor was composed of six MoS$_2$ FETs having a channel length of 7.5 nm. Nourbakhsh et al. [19]
than devices based on Si and III-V materials, at the same channel lengths. This MoS2 composite transistor with six FETs in series possessed an off-state current of 100 pA/μm and an $I_{on}/I_{off}$ ratio greater than $10^5$. Modeling of the resulting current-voltage characteristics revealed that the metallic/semiconducting MoS2 junction had a low resistance of 75 Ω μm. These experimental results reveal the remarkable potential of 2D MoS2 for future developments of sub-10 nm technology. Although the structure studied by Nourbaksh et al. was composed of a chain of short channel transistors rather than an individual device, the same short channel effects that occur in single transistors were also active in this series of transistors, because of metallic regions present between any two devices in the chain.

An alternative approach for self-aligned MoS2 transistors was recently demonstrated by English et al. [20] In this work, an MoS2 FET with a self-aligned 10 nm top gate was fabricated by using a self-passivated Al2O3 layer around an Al gate electrode as the gate dielectric (Fig. 3.5). This allowed for a decrease of the ungated regions to 10 nm.

To reduce the gate length of these devices even further and probe the ultimate limit of scaling, a nanotube-gated MoS2 FET was demonstrated by Desai et al. [21] In their work, a metallic single-wall carbon nanotube (SWCNT) with a diameter of 1 nm was used as the gate electrode enabling a physical gate length down to 1 nm to be achieved (see Fig. 3.6). However, because of the fringing electric field induced by the SWCNTs, the effective channel length in the off-state, calculated by simulation, was 3.9 nm. This ultrashort channel MoS2 FET showed excellent switching characteristics with a subthreshold swing of 65 mV/dec (Fig. 3.6). In this device structure, the SWCNT gate underlapped the source/drain electrodes by some hundreds of nanometers, which caused an extremely large access region. To decrease this resistance, the ungated regions were electrostatically doped by the Si back-gate during electrical measurements.

These initial experimental results show the great promise of MoS2 devices to push Moore’s law beyond the scaling limits of silicon. In addition, as all of these devices can be fabricated at low temperature (<400 °C) on top of a fully fabricated Si wafer, they can provide a high performance transistor interposer to be used in 3D chip architectures to drive memory planes, interconnects or power-gate the entire chip. However, before these complex circuits can be fabricated on this novel technology, it is necessary to improve the material, device, and circuit yield. For these novel fabrication technologies more-robust circuit topologies are being investigated as described in the following section.

### 3.2.2 Designing Complex Circuits with Immature Technologies

Despite its promising material characteristics and initial device performance, MoS2-based circuits to date have been limited to a single or few-transistors [22–25] due to the many challenges associated with the uniformity and yield control in both material growth and device technology. To significantly improve the fabrication
Fig. 3.5 Fabrication steps of the self-aligned top-gate MoS$_2$ FET with a minimum 10 nm gate length. (a) Deposition of Al gate electrode with seed layer, (b) formation of self-passivated Al$_2$O$_3$ gate dielectric, (c) self-aligned Au source and drain. (d) SEM image (top view) of the device. (e) Colorized cross-sectional TEM of the device with the channel length of 10 nm. (f) TLM structure used to extract mobility and contact resistance. English et al. [20]

yield and be able to increase the complexity of these circuits, it is necessary to follow an integrated approach where all the main steps to make functional MoS$_2$ circuits, including material synthesis, device technology, compact modeling, circuit design, layout automation, chip fabrication, and circuit measurement are connected (Fig. 3.7) [25–28]. Such an end-to-end design flow allows for the rapid optimization of the material synthesis, device technology, and circuit layout/design to ensure maximum system yield.
A main challenge for MoS$_2$ circuit development is that MoS$_2$ is natively n-type, as the material is subject to sulfur vacancies during the CVD growth process resulting in intrinsic electron concentration. Therefore, most MoS$_2$ transistors reported in literature are $n$-MOS with negative threshold voltage [22, 23]. Although tremendous efforts have been made to realize $p$-MOS with MoS$_2$ [29] and new 2D semiconductors have been discovered with improved ambipolar conduction [30, 31], $n$-type MoS$_2$ has so far been the most advanced technology for large-scale integrated circuits. Despite chemical doping of MoS$_2$ in both n type and p type in a controlled manner is key to realize a CMOS technology, it is beyond the scope of this chapter. Nevertheless, enhancement mode FETs with positive threshold voltage are necessary to cascade logic circuits and complete a stand-alone system. Furthermore, FETs made of new materials such as MoS$_2$ often suffer from large process variations and hysteresis, which result in significant performance degradation and poor long-term reliability. This problem should be addressed in both device technology and circuit design.
Fig. 3.7 Design flow for MoS$_2$-based electronics. The flow connects material growth, device technology, compact modeling, circuit design, layout generation, chip fabrication and circuit’s measurement. This speeds up the iteration between different development stages of new material technology. The interactions between different developments stages are indicated as dashed arrow [26].

Most 2D transistors today are fabricated with a gate-last process. Dramatic improvement can be achieved by making the gate structure first, and depositing and annealing the gate dielectric before transferring large area, single layer MoS$_2$. This technology, when compared with the conventional gate-last technology, greatly improves the device uniformity. Statistical study of the transistor performance shows that gate-first devices have an average $V_t$ of 2.41 V with a standard deviation of 0.17 V while those from gate-last process have an average $V_t$ of -4.20 V and a much larger standard deviation of 1.75 V [26].

To demonstrate the high uniformity of the new technology, various combinational (NAND, NOR, AND, OR, XOR, XNOR) and sequential (latches, edge-triggered registers) digital circuit as well as switch capacitor voltage regulators based on MoS$_2$ were fabricated on a silicon wafer [25, 26]. Figure 3.8 shows the performance of an edge-triggered register, which is designed using two latches in a master–slave configuration. When the clock is low, the master (connected to the input) is transparent while the slave is in hold mode. When the CLK turns high, the master is in hold mode while the slave latch is transparent. Thus, the output of the edge-triggered register will capture the input value at the positive edge of the CLK. The measurement results of the designed circuits fit well with the simulation, indicating the great promise of our technology and CAD flow for realizing large-scale complex MoS$_2$ systems. The complex four-stage integrated circuits evidence the robustness and scalability of the technology.

The flow as shown in Fig. 3.7 provides an end-to-end guidance to implement functional MoS$_2$ circuits from scratch. However, the availability of more complex circuits is still limited by the poor circuit yield. In the case of MoS$_2$, this yield is limited by variations in the MoS$_2$ material quality, non-uniform doping level across
the sample and scattering from impurities and nucleation centers. Thus, in order to develop a stable technology using MoS₂, the sources of variability in the fabrication process need to be understood and captured at the modeling stage and counteracted at the circuit design stage. Moreover, a scalable and flexible yield model, which can quickly evaluate fabrication process and designs as the number of transistors scale, is highly valuable.

Several factors have been identified to be critical to reduce the variation in MoS₂ devices. They include: uniformity of material, cleanness of transfer process, robustness of passivation technology and immunity to threshold voltage change in the designed circuits. Figure 3.9 shows that the devices made of high quality CVD MoS₂ (Fig. 3.9a, c, e, g) have tighter distribution of current, lower off current, higher on current, compared with low quality CVD MoS₂ devices with nucleation particles from growth phase (Fig. 3.9b, d, f, h) [26, 27]. A statistic compact model has been developed to capture the corner performances of the devices for a certain technology as shown in red line in Fig. 3.9c, d, allowing capture variation of devices in circuit design. Very importantly, clean high-quality sample results in very small threshold variation ($\Delta V_{t0} = 36$ mV) compared with low quality sample ($\Delta V_{t0} = 0.17$ V) [26, 27].

To be able to fabricate large MoS₂ circuits on a silicon chip, it is important to develop yield models for those complex logic circuits. This was done by simulating the noise margin (NM) of the MoS₂ transistors based on their key parameters [26, 27]. The simulated NM of zero-$V_{GS}$ inverters with different threshold voltages for
Fig. 3.9 Performance of the transistor arrays with (a) high (sample A) and (b) low (sample B) quality MoS₂ in linear (c, d) and log (e, f) scales. The red solid lines are the compact models capturing the slow and fast corners of devices on each sample. Threshold voltage distribution of sample A (g) and sample B (h). Both threshold voltage distributions are fitted using Gaussian distribution with average value of $V_{t0}$ and standard deviation value of $\Delta V_{t0}$. The fitted parameters for both samples are shown in the images [26, 27].

Fig. 3.10 (a) Color map of noise margin/VDD for the zero-VGS inverter as a function of various threshold voltage values for top ($V_{t1}$) and bottom ($V_{t2}$) transistors. The probability of a given noise margin is the joint probability of ($V_{t1}$, $V_{t2}$) pair, e.g., for the probability of the highlighted point, $\rho(NM) = \rho(V_{t1} = 1.7 \text{ V}) \times \rho(V_{t2} = 1.8 \text{ V})$. (b) Yield model for different circuit complexities for both pseudo CMOS and zero-VGS design, with various $V_{t0}$ and constant $V_{t0}$ of 1.8 V. Both topologies show close-to-unit yield for wide range of $V_{t0}$ and $\Delta V_{t0}$ [26, 27].

Top and bottom transistors are summarized in the color map shown in Fig. 3.10. The dark blue region indicates NM < 0, that is circuit failure. By combining the NM plot with the statistical distribution of $V_{t}$’s in the MoS₂ technology, it is possible to determine the circuit yield as a function of transistor count ($N$). All transistors in the circuit are assumed to be independent from each other, and their threshold voltages are sampled according to the sample statistical distribution (with measured $V_{t0}$ and $\Delta V_{t0}$). The simulation results for both zero-VGS and pseudo CMOS [32] design are summarized in Fig. 3.10b. As expected, for both topologies, the yield decreases with increasing the circuit complexity ($N$) and global and local variations.
Fig. 3.11 Experimental and simulated performance of circuits fabricated on two different samples (A, C) with high quality material. No circuit failure is observed across those two samples, indicating the high performance and robustness of the technology. The good match between the measurement and simulation demonstrates the power of the variation-aware design platform [26, 27]. (a) Zero-VGS inverters; (b) Psuedo-CMOS inverters; (c) NAND gates; (d) XNOR gates.

Figure 3.10b shows the yield for different standard deviation of the threshold voltage (ΔVt0 from 0.05 V to 0.3 V) for both designs, with Vt0 of 1.8 V. For ΔVt0 smaller than 0.15 V, both designs show close-to-unity yield and pseudo CMOS design is more robust to larger ΔVt0. Figure 3.11 depicts the simulation and the measurement results of part of the combination circuits and from inverters from two samples (A and C) made with high quality material. No circuit failure is observed across these two samples, indicating the high performance and robustness of the technology. The good match between the measurement and simulation evidences the power of the proposed variation-aware design platform when designing large area complex circuits with new material systems such as MoS2.
3.3 2D Materials for IR Detectors on a Si Platform

Infrared (IR) sensing technologies, originally used mainly for night vision, surveillance, and remote controlling in military applications, have gradually shifted to applications for civilian use, including medical, industry, earth resources, and automotive. For example, medical diagnostics can be assisted by IR thermography, in which IR scanning is used to detect spatial temperature abnormality and identify in this way cancers and other trauma [33]. The use of high resolution IR imagers and/or spectrographs on aircrafts and satellites has also allowed the chemical mapping of the surface of earth, with important applications in mineral search [34], as well as geological and environmental survey [35], and inspection of natural hazards and disasters [36]. However, in spite of the large number of applications where IR sensing is key, its use is still limited by the high cost of high performance IR detectors.

The bandgap of 2D materials spans from 0 electron volts (eV) in the case of monolayer graphene, all the way to around 5 eV in the case of hexagonal boron nitride (Fig. 3.12). The energy gaps of graphene and black phosphorus, in particular, lie in the mid- to far-infrared range, and exhibit abundant novel light–matter interaction phenomena, such as thermoelectric effect, photothermoelectric effect, and various other optoelectronic effects related to hot carrier dynamics. The use of these novel phenomena could lead to ultrasensitive and/or fast-response IR detectors with the potential to outperform today’s state-of-the-art IR detection technologies.

An IR imaging system is usually made of a large number of detectors, which need to be routed to readout integrated circuits (ROICs) that amplify the signals and implement analog-to-digital conversion. The integration of the detector arrays and the ROICs has traditionally been a major challenge in IR system design, due to material integration and noise issues. As a result, in a conventional IR imaging

![Fig. 3.12 Summary of band gaps ($E_{bg}$ in eV) of various 2D materials and corresponding wavelength each material is capable of detecting](image-url)
system, the detector array and the ROIC are usually fabricated on separate wafers and then integrated in a hybrid fashion. However, recent advances of wafer-scale synthesis of graphene and other 2D materials, the easiness of transferring 2D materials onto arbitrary substrates and their low-temperature fabrication technology make it possible to monolithically integrate 2D material-based IR detectors with CMOS integrated circuits. This is expected to reduce the noise and the cost for these systems.

In this section, we will first summarize the current research on 2D material-based IR detectors, and then we will focus on graphene thermopiles as an example to investigate their potential as compared to the state-of-the-art mainstream technologies. Finally, we will discuss the possibility and benefits of monolithically integrating 2D material image arrays and CMOS integrated circuits.

### 3.3.1 2D Material for Infrared Detectors

Due to its linear electronic dispersion relation, monolayer graphene can show ultra-high mobility (up to 200,000 cm²/Vs [37–39]), which makes graphene an excellent candidate for high-speed electronic or optoelectronic applications. Moreover, the interband optical absorption of suspended, near-intrinsic monolayer graphene is $\pi \alpha \approx 2.3\%$ for a wide range of incident photon energy, determined by the fine structure constant, $\alpha = e^2/hc \approx 1/137$ [40]. As for heavily doped graphene, the interband transition is forbidden by Pauli blocking, whereas the intraband, free-carrier absorption could lead to a resonant absorption, called surface plasmon polariton, in which free electrons and holes in graphene vibrate in-plane with respect to incident light. The resonant absorption can be enhanced efficiently by spatially confining the electromagnetic wave excitation [41–45]. Table 3.1 summarized the key performance metrics of different 2D material infrared detectors.

### 3.3.2 Graphene Thermopiles

Figure 3.13 shows the basic structure of a widely used graphene photodetector design. This device, made of a sheet of graphene with dual split-backgates, develops a photovoltage across electrodes M1–M2 as a function of the voltage applied to the backgates. A photovoltage is measured for laser illumination wavelengths of 0.83 μm, 1.55 μm, and 10.6 μm, respectively [64]. The fact that graphene shows a clear response even at 10.6 μm wavelength at which graphene absorption should be very limited due to Pauli blocking [65, 66] indicates that most of the light absorption is happening not in the graphene itself but in the substrate underneath, while the graphene devices are using the thermoelectric effect to convert the temperature rise in the substrate to a voltage difference [47, 64, 67]. This voltage is described by $V_{TE} = \Delta S \cdot \Delta T$, where $\Delta S$ is the difference of the Seebeck coefficient between the
### Table 3.1 Summary of performances of various 2D material-based infrared detectors

<table>
<thead>
<tr>
<th>Device structure</th>
<th>Wavelength</th>
<th>Responsivity</th>
<th>Specific detectivity</th>
<th>Response time</th>
<th>Operating temperature</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gr. p–n junct. PTE</td>
<td>Visible, NIR</td>
<td>10 mA/W</td>
<td>–</td>
<td>50 fs</td>
<td>Room T</td>
<td>[46–48]</td>
</tr>
<tr>
<td>Gr.-metal plasmonic</td>
<td>MIR</td>
<td>0.5 V/W</td>
<td>–</td>
<td>60 ns</td>
<td>Room T</td>
<td>[49]</td>
</tr>
<tr>
<td>Gr. thermopile</td>
<td>MIR</td>
<td>10 V/W</td>
<td>$8 \times 10^8$ cm Hz$^{1/2}$ W$^{-1}$</td>
<td>23 ms</td>
<td>Room T</td>
<td>[50]</td>
</tr>
<tr>
<td>Biased gr. bolometer</td>
<td>Visible, NIR</td>
<td>0.2 mA/W</td>
<td>–</td>
<td>–</td>
<td>Room T</td>
<td>[51, 52]</td>
</tr>
<tr>
<td>Dual-gated 2L gr. bolometer</td>
<td>MIR</td>
<td>$10^5$ V/W</td>
<td>$3 \times 10^{10}$ cm Hz$^{1/2}$ W$^{-1}$</td>
<td>&lt;1 ns</td>
<td>&lt;10 K</td>
<td>[52]</td>
</tr>
<tr>
<td>Gr./QDs photo-gating</td>
<td>Visible NIR</td>
<td>$10^8$ A/W</td>
<td>$7 \times 10^{13}$ cm Hz$^{1/2}$ W$^{-1}$</td>
<td>10 ms</td>
<td>Room T</td>
<td>[53, 54]</td>
</tr>
<tr>
<td>Gr./insulator/gr. photo-gating</td>
<td>Visible, NIR, MIR</td>
<td>&gt;1 A/W</td>
<td>$5 \times 10^7$ cm Hz$^{1/2}$ W$^{-1}$</td>
<td>1 s</td>
<td>Room T</td>
<td>[55]</td>
</tr>
<tr>
<td>Gr./insulator/gr. thermionic</td>
<td>NIR</td>
<td>0.01–0.2 mA/W</td>
<td>–</td>
<td>10 fs</td>
<td>Room T</td>
<td>[56, 57]</td>
</tr>
<tr>
<td>Gr./semi. PV</td>
<td>Visible, NIR</td>
<td>0.7 A/W</td>
<td>$5 \times 10^{13}$ cm Hz$^{1/2}$ W$^{-1}$</td>
<td>0.3 ms</td>
<td>Room T</td>
<td>[58, 59]</td>
</tr>
<tr>
<td>Gr./pyroelectric</td>
<td>MIR</td>
<td>0.27 mA/W</td>
<td>$6 \times 10^4$ cm Hz$^{1/2}$ W$^{-1}$</td>
<td>20 ms</td>
<td>Room T</td>
<td>[60]</td>
</tr>
<tr>
<td>Gr./NEMS resonator</td>
<td>MIR</td>
<td>–</td>
<td>$1 \times 10^{10}$ cm Hz$^{1/2}$ W$^{-1}$</td>
<td>3 ms</td>
<td>Room T</td>
<td>[61]</td>
</tr>
<tr>
<td>BP photoconductor</td>
<td>Visible, NIR</td>
<td>5 mA/W</td>
<td>–</td>
<td>40 μs</td>
<td>Room T</td>
<td>[62, 63]</td>
</tr>
</tbody>
</table>

$p$- and n-region of graphene, and $\Delta T$ is the temperature difference between the graphene $p$–$n$ junction and the metal contacts. Because the Seebeck coefficient of graphene has an “S” shaped dependence on the Fermi level, the split-gate sweeping maps as shown in Fig. 3.13c, d, e indicate a non-monotonic six-lobed pattern, which is not true if the photovoltaic effect dominated the photocurrent generation. In a related work, the spectral response of graphene on SiO$_2$ was measured and it was found to match the absorption spectrum of SiO$_2$ very well in the mid-IR. However, the detectivity of these devices was only $10^2$ cm Hz$^{1/2}$ s$^{-1}$. By optimizing the device structure, a follow-up study achieved a detectivity of $10^5$ cm Hz$^{1/2}$ s$^{-1}$ [50]. The observed three orders of magnitude improvement in performance was attributed to two structural changes. First, the absorption in the 8–12 μm spectral range was increased by >40%, as compared to less than 1% absorption in the previous work, thanks to the use of an optimized IR absorption layer made of a SiO$_2$/Si$_3$N$_4$/SiO$_2$ combination instead of a single dielectric layer, which allowed strong resonant absorption at 9 and 11.5 μm, respectively. Secondly, the deposition of the SiO$_2$/Si$_3$N$_4$/SiO$_2$ tri-layer absorber was done with plasma enhanced chemical vapor deposition (PECVD) under an optimized high-frequency
Fig. 3.13 Graphene split-gate thermopile with supported substrate. (a) Schematic and (b) microscopic image of the device. M1 and M2 are metal contacts to graphene, and G1 and G2 are split gates that electrostatically dope the graphene channel to form a p-n junction. (c–e) Photovoltage ($V_{PH}$) as a function of the split gate voltages $V_{G1}$ and $V_{G2}$, for different wavelength of incident light. (c) 10.6 μm, (d) 0.83 μm, and (e) 1.55 μm. The six-lobe feature indicates that the photoresponse is dominated by thermoelectric effect in graphene [64].

to low-frequency plasma ratio to make the thin film stress-free. This allowed for the fabrication of a free-standing absorber membrane after undercutting the silicon underneath with XeF$_2$ isotropic etching. When measured in vacuum (less than $10^{-2}$ torr), both thermal conduction and convection in the vertical direction were efficiently attenuated. As a result, the temperature generated by the IR irradiance at the absorber is maximized.

Multiple graphene photodetectors can be combined into the thermopile shown in Fig. 3.14a, composed of an infrared absorber that is suspended from the substrate, a series of thermal arms that connect the absorber and the surrounding, with interleaved $p$- and $n$- type graphene channels on top. The schematic of the structure and the key geometrical parameters are shown in Fig. 3.14b. The graphene channels
Fig. 3.14  Graphene thermopile with suspended IR absorber. (a) Schematic of a graphene thermopile. The red and blue regions indicate the p-type and n-type region of graphene, and the square in the center is the dielectric absorber. The whole structure is suspended on the substrate to reduce the thermal conductance in the vertical direction. (b) A geometrical abstraction of the graphene thermopile with geometrical parameters listed below.

are parallel in terms of temperature gradient, but connected in series electrically. When IR radiation is present, the IR absorber (dielectric multilayer thin film) is heated up, which can then be probed electrically by the graphene p–n junctions due to the thermoelectric effect. The specific detectivity ($D^*$), considering the Johnson–Nyquist noise, can be expressed as

$$D^* = \frac{N_j \Delta S \Delta T}{P_{in} \sqrt{\frac{\Delta S}{\rho^2 D}}} \times \sqrt{D^2_{abs} \Delta f} = \left( \frac{\alpha_{abs}}{t} \right) \cdot \left( \frac{\Delta S}{\rho^2 D} \right) \cdot \left( \frac{1}{\kappa_{th}} \right) \cdot \left( \frac{L^{1/2} D_{abs}}{N_j^{1/2} W^{1/2}} \right) \sqrt{ \frac{1}{32k_BT} } \tag{3.1}$$

Here the first term ($\alpha_{abs}/t$) is absorbance per thickness, indicating the capability of IR absorption of the absorber; the second term ($\Delta S/\rho^2 D$), with $\rho_{2D}$ the 2D resistivity of graphene, is determined by the electrical and thermoelectric properties of the sensing material; the third term ($1/\kappa_{th}$) indicates the quality of thermal isolation, with $\kappa_{th}$ denoting the thermal conductivity of the absorber; and the fourth term is made of geometrical parameters in the parallel direction, where $N_j$ is the total number of graphene $p$–$n$ junctions at the hot spot. The response time of such a device can be written as

$$\tau = R_{th} \cdot C_{th} = \frac{1}{2} \left( \frac{c_V}{\kappa_{th}} \right) \left( \frac{L D^2_{abs}}{N_j W} \right) \tag{3.2}$$

with the heat capacitance $C_{th}$, and the specific heat capacity $c_V$. Notice that ($D^{*2}/\tau$) is independent of the lateral geometries.
$$\frac{(D^*)^2}{\tau} = \left( \frac{\alpha_{\text{abs}}}{t} \right) \cdot \left( \frac{\Delta S^2}{\rho_{2D}} \right) \cdot \left( \frac{1}{k_{\text{th}cV}} \right) \cdot \frac{1}{16k_B T} \tag{3.3}$$

with light–matter interaction factor \((\alpha_{\text{abs}}/t)\), the figure of merits of the thermoelectric material \((\text{FOM} = \Delta S^2/\rho_{2D})\) and the thermal transport factor \((1/k_{\text{th}cV})\).

According to Eq. 3.3, the thermoelectric figure of merit FOM plays a significant role in thermopile IR detectors. In order to benchmark graphene-based thermoelectric detectors with respect to the other material systems, the Seebeck coefficient and the FOM are plotted as a function of resistivity in Fig. 3.15. Note that the FOM for today’s standard CVD graphene on SiO\(_2\), with the average mobility of 2000 cm\(^2\) V\(^{-1}\) s\(^{-1}\), can already outperform the performance of any thermopiles made with metals and most of TE materials. The use of higher quality graphene and properly passivating the dangling bonds on the substrate with hexagonal BN could make the FOM two orders of magnitude higher than that of all the other material systems. Also, the FOM of TMDs in 2D form is higher than in their 3D counterparts, which also shows great potential for thermal detection and other thermoelectric applications.

Note from Eq. 3.3 that the optical absorption per thickness \((\alpha_{\text{abs}}/t)\) and the thermal transport factor \((1/k_{\text{th}cV})\) of a graphene thermopile can be improved through properly engineering the IR absorber. Current work used the inherit strong absorption of SiO\(_2\)/SiN layers \([50, 64, 68]\). It is also possible to use metal-based metamaterial absorbers in the mid-infrared range \([69–71]\). The surface plasmon polariton (SPP) in graphene \([41, 45, 72, 73]\) and phonon polariton (PP) in hBN \([74, 75]\) have been studied recently. With appropriate engineering of the light field to enhance the light–matter interaction, it is possible to achieve strong resonant absorption based on these mechanisms with negligible loss.

Figure 3.16 compares the specific detectivity \((D^*)\) and the response time \((\tau)\) of graphene thermopiles with different types of state-of-the-art thermal detector technologies, including bolometers (VO\(_x\), CMOS-MEMS, etc.) \([6–8]\), thermopiles \([9, 10]\) (poly-Si, Al, thermoelectric materials, etc.), and pyroelectric devices (PZT and other piezoelectric materials). Note that photon detectors, such as mercury cadmium telluride (MCT) photoconductors, were not included in this comparison, because they would suffer from high dark current and consequently high noise at room temperature \([76]\). Here we exploit the dimensionless relation from Eq. 3.2 to represent each technology node of graphene thermopiles. Although current graphene thermopile technology is still not as good as the state-of-the-art thermal detectors, the performance is predicted to be competitive, or even better than today’s state-of-the-art technologies. For example, a 100-fold improvement in the FOM = \(\Delta S^2/\rho\) could be achieved by encapsulating large area, high-quality CVD graphene with hexagonal boron nitride, which would increase the mobility and Seebeck coefficient to 100,000 cm\(^2\) V\(^{-1}\) s\(^{-1}\) and 200 \(\mu\)V/K, respectively. Furthermore, the absorber can be thinned down to 10 nm with good mechanical stability, and 100% perfect absorption can be achieved through nano-photonic structures, which would make graphene thermopiles better than any existing bolometers.
Fig. 3.15 (a) Seebeck coefficient and (b) thermoelectric figures of merit as a function of resistivity for various materials.
3.3.3 Heterogeneous Integration of Graphene and Silicon Integrated Circuits for Thermal Imaging Application

One important application of IR detection technology is to image the black-body radiation of ordinary objects, as in night vision goggles, automotive imaging systems, surveillance cameras, and temperature-control systems. A 2D imaging system could be implemented by either spatially modulating the optical path to map the pixelated 2D object information into a time series of signal, called 2D scanning, or directly mapping the pixelated 2D object into a 2D array of image sensors at the focal plane, called real-time imaging. Figure 3.17a shows a scanning IR imaging system using a single graphene thermopile IR detector. With this system, a black-body source at 472 K could be imaged, as shown in Fig. 3.17b, c.

A real-time image sensor is made of a focal plane array (FPA), that is a 2D array of IR detectors at the focal plane of the optical system, and a readout integrated
Fig. 3.17  Scanning thermal imaging system based on a single graphene thermopile. (a) Schematics of the scanning imaging setup. (b) “MIT” log aperture to be measured. (c) Scanning thermal image of the 472 K black-body source passing through the “MIT” log aperture

Circuit (ROIC) that controls the operation of the FPAs and converts the analog signals from the FPA into digital signals that can be transmitted to and processed by a microprocessor. The integration between the FPA and the ROIC could be discrete, hybrid, or monolithic. The discrete integration is impractical because the number of connections between the FPA and the ROIC is on the order of thousands or even millions, which could hardly be routed through a PCB board. At present, most of the mainstream IR image sensors have separate FPA and ROIC chips, and integrate them through flip chip bonding. This limits the pixel size, increases the system noise, and reduces the speed of the device. The monolithic integration of the FPA and the ROIC could solve these challenges, however the thermal budget of the ROIC demands that the deposition temperature of the sensing material has to be lower than 400 °C, which degrades the material quality of most 3D semiconductors used in these applications.

Given that 2D materials can be synthesized by chemical vapor deposition on a metal foil, and wet-transferred onto any substrates, it is possible to integrate 2D material-based IR FPAs with CMOS ROICs monolithically without sacrificing the performance of either component. The back-end-of-line process of graphene thermopile FPAs fabricated directly on CMOS ROIC chips is schematically shown in Fig. 3.18a and the optical images of the completed graphene thermopile-CMOS image sensor are shown in Fig. 3.18b, c. This hybrid graphene/Si system enables new opportunities for low cost, high performance IR detection.
Fig. 3.18 Graphene thermopile/CMOS monolithic integration. (a) Back-end-of-line process of the graphene thermopile FPA fabricated directly onto a CMOS ROIC chip. PASS passivation, ILD interlayer dielectric, M6 Metal 6 (Cu), M5 Metal 5 (Cu), blue PECVD SiO2, grey line is graphene. The left and right images are the side and top view of the pixel area, respectively. (b) Photograph of the CMOS ROIC chip. (c) Microscopic images of the completed Graphene thermopile FPA/CMOS ROIC chip

3.4 Heterogeneous Integration of 2D Materials with Si Nanophotonics

Graphene can serve as core materials combined with Si or other conventional semiconductor photonic structures for ultrafast photodetection and optical modulation on a CMOS platform for high speed, low power optical interconnects. Early demonstrated graphene optoelectronics normally rely on a vertical incident configuration to couple light into graphene, therefore resulting in a low absorption coefficient of 2.3% due to the atomic thickness of graphene [77]. Although those devices show promising properties of strong electro-absorptive effect [65, 66] and high-speed photodetection [78, 79], the low total absorption in those devices does not meet the requirement of practical applications. To tackle this issue, nanophotonic devices, including waveguides and cavities, can squeeze optical field in a subwavelength volume, providing a promising platform to enhance light–matter interaction in 2D materials.
Fig. 3.19 Finite element simulation (COMSOL) of the waveguide mode indicates that the evanescently coupled bilayer graphene on top of the waveguide introduce a loss of $\sim 0.085 \text{ dB/}\mu\text{m}$. The absorption coefficient suggests that, for example, a 70-\mu m-long overlapping of graphene with waveguide will absorb $\sim 70\%$ (6 dB) of the incoming light incident from the waveguide, which is consistent with the experimental results and much greater than the 0.1 dB absorption in the normal-incidence configuration.

When 2D materials couple to a nanophotonic cavity, enhanced light–matter interaction of graphene in a cavity can enable spectrally selective, order-of-magnitude enhancement in optical absorption and intensity modulation in cavity reflection. Using a temporal coupled mode theory that considers the absorption of graphene coupling to the cavity, Gan et al. showed that the cavity reflection and the absorption strongly depend on the intrinsic cavity loss $\kappa_c$ without 2D materials and the excess loss introduced by absorption of 2D materials $\kappa_{2D}$ in the cavity. The attenuation ratio for the reflectivities with and without 2D materials indicates that the reflectivity of the 2D materials-cavity system can achieve a contrast of 20 dB with a tuning of $\kappa_c/\kappa_{2D}$ in the range of 0.1–1. On the contrary, the on-resonance absorbance of 2D materials has a maximum value of $\eta$, which is the out-coupling efficiency of the cavity field to the environment. The maximum occurs when the two decay rates satisfy $\kappa_{2D} = \kappa_c$, i.e. in the critical coupling condition when excess loss of 2D materials equals to the intrinsic cavity loss.

Figure 3.19a shows an experimental implementation of a coupled graphene-cavity system. The cavity was created by drilling a planar photonic crystal (PPC) air-hole lattice in a suspended silicon membrane. Linear three-missing-hole (L3) defect in the middle of the lattice forms the cavity area. Heterogeneously integrated graphene on top of the membrane can couple to the optical evanescent field and alter the cavity quality factor and out-coupling radiation intensity. Both experimental and calculation results show that such an evanescently coupled graphene-cavity system results in a loss ratio of $\kappa_c/\kappa_{2D} \sim 0.1$, indicating a maximum modulation contrast in reflection can reach more than 20 dB if graphene becomes completely transparent. In this cavity, the out-coupling of the cavity field to the free space radiation mode is symmetric. Therefore, the maximum $\eta$ is 50% if the waveguide-cavity coupling is
responsible for the entire intrinsic loss of the cavity. For cavities with traveling-wave resonant modes, such as ring resonators [80], the $\eta_{\text{maximum}}$ is 100%, enabling 100% absorption in 2D materials.

The enhanced light–matter interaction via optical resonators is inherently narrowband due to the narrow resonant bandwidth of the resonators. On the contrary, coupling 2D materials with a single mode bus waveguide can enhance their interaction with light across a broad spectrum due to the large extension of the interaction length. Figure 3.20b shows a schematic of a graphene layer deposited on top of a silicon waveguide.

### 3.4.1 High-Speed Graphene Electro-Optic Modulators

Due to the unique linear dispersed band structure [81], graphene exhibits uniform absorption in the spectral ranges from visible to mid-infrared [77]. In addition, the absorption of graphene varies by tuning its Fermi energy ($E_F$) via an electrostatic gate voltage, as schematically shown in Fig. 3.20a. When $E_F$ is tuned away from
the Dirac point by more than half of the photon energy $\omega/2$, the interband transitions become forbidden by Pauli blocking, reducing the graphene absorption [65, 66].

Heterogeneously integrated graphene E–O modulators first appear in a structure consisting of a silicon waveguide and $\sim 50$-μm-long graphene overlapping [82], as shown in Fig. 3.20b. The enhanced total absorption of the graphene layer on top of the waveguide results in enhanced electro-absorptive effects, enabling broadband intensity modulation for wavelengths from 1.35 to 1.6 μm with 3 dB modulation depths. The 3 dB cut-off speed of this device is about 1 GHz. An improved dual-layer graphene capacitor structure on waveguide shows improved modulation depths of 6 dB [83] and 16 dB [84] with low insertion loss of 3 dB, demonstrating the scalability to acquire higher modulation depth with stacking of multiple single graphene layers.

In these waveguide-based E–O modulators, the footprint of the devices is intrinsically limited by the coupling strength of the waveguide and the graphene layers, which commonly requires a device area of 50 μm$^2$. To reduce the footprint of the E–O modulator, therefore reducing the switching energy and increasing the operation speed, a resonator-coupled graphene structure is desired. When graphene strongly absorbs the optical field in a cavity, i.e., $\kappa_c/\kappa_{cg} \ll 1$, the cavity reflection $R$ is strongly attenuated. While the Fermi level in graphene increases (decreases) due to electrostatic gating, the optical absorption in graphene reduces ($\kappa_c/\kappa_{cg}$ increases). Therefore the reflection $R$ of the cavity recovers. This modulation mechanism can in principle provide more than 20 dB modulation depths for a cavity area around 0.5 μm$^2$.

We discuss an example of a multilayer graphene/boron nitride heterostructure integrated a PPC L3 cavity. In this architecture, mutually gated dual-layer graphene parallel capacitor could provide high doping strength while operating at high speed due to high carrier mobility of graphene [85]. Figure 3.20c shows the schematic of the high-speed graphene E–O modulators consisting of a dual-layer graphene capacitor and a PPC nanocavity [86]. A BN/Graphene/BN/Graphene/BN five-layer stack was built by the van der Waals (vdW) assembly technique and then transferred onto a quartz substrate [85]. The two graphene sheets were positioned as crossed stripes in order to be contacted individually. The metal contacts to the graphene employ a one-dimensional edge contact technique. In this encapsulated dual-layer graphene structure, each one of the graphene sheets can supply gate voltage to each other.

The device showed in Fig. 3.20c exhibits a maximum modulation depth of 3.2 dB with a high-speed cut-off frequency around 1.2 GHz. The operation frequency response of the device indicates a RC-limited time constant of the dual-layer graphene capacitor, as deduced by the impedance measurement of the device. In this device, the graphene capacitor has an area of $\sim 100$ μm$^2$ and a capacitance of 320 fF. The switching energy of this device is approximately 1 pJ/bit. For L3 PPC cavities, the overlap between the resonant mode and the graphene capacitor has an area of only $\sim 0.5$ μm$^2$, corresponding to the three-missing-hole defect region. The graphene capacitor could therefore be reduced in size to match this cavity area to lower the capacitance by approximately 200 times, which should reduce the switching energy to 5 fJ/bit and increase the 3 dB cut-off frequency to 70 GHz.
The cavity bandwidth in this work exceeds 600 GHz for a $Q$ value of 300, i.e., it would be possible to obtain a relatively large modulation contrast without the need for particularly high-$Q$ cavities, as is required in silicon carrier-depletion (injection) modulators. This broader bandwidth would also improve temperature stability, which is a limiting factor in the carrier-modulation of Si modulators [87–89].

The electro-absorptive effect in graphene can not only induce tunable loss of the cavity, but also the cavity coupling efficiency to external optical modes, as the device demonstrated by Phare et al. [90] Fig. 3.20d shows the dual-layer graphene capacitor covering a silicon nitride ring resonator. The ring resonator is intentionally designed to be under-coupled to a bus waveguide when the graphene introduces loss to the cavity. While the absorption in graphene reduces, the resonant field becomes stronger and the coupling of the resonator to the waveguide increases. The tuning of both cavity internal field and out-coupling to the waveguide gives extra tuning of the light transmission, resulting a modulation depth of 15 dB with 10 V voltage swing. The speed of modulator exceeds 30 GHz due to reduced capacitance area of $\sim 45 \mu\text{m}^2$. Other resonator-coupled graphene E–O modulators have also shown both high modulation depth and small device footprint, including graphene-silicon ring resonator structures [91, 92], and silicon Mach-Zehnder interferometers-integrated structures [93].

### 3.4.2 On-Chip Graphene Photodetectors

On-chip graphene photodetectors could enable on-chip optical interconnects at unprecedented speeds. Although early demonstration of graphene-based photodetector for these applications has shown promising operation speed and broadband response [78, 79], the responsivity of the photodetector is low due to weak absorption in graphene in a vertical incident configuration. To address this, graphene has been integrated with nanocavities [94], microcavities [95], and plasmon resonant structures [49, 72]. In these examples, a range of responsivity of 10–30 mA/W is possible, with a trade-off for limited spectral response due to the narrowband enhancement of the resonant structures. To implemented broadband photodetectors, heterogeneous integrated graphene with waveguides is most promising for broadband spectral response while enhancing the photoresponse of the detector.

Here, we describe a waveguide-coupled graphene photodetector [96, 97], as schematically illustrated in Fig. 3.21a. The silicon waveguides were fabricated in an SOI wafer with a CMOS compatible process. The chip was planarized by backfilling with a thick SiO$_2$ layer, followed by chemical mechanical polishing to reach the top silicon surface. Multilayered hBN/SLG/hBN stack were assembled onto the photonic chip using van der Waals (vdW) assembly [85]. The single layer graphene (SLG) channel spans 40 $\mu\text{m}$ of the waveguide, inducing $\sim 2.2$ dB absorption, consistent with simulation results. One-dimensional edge contacts (Fig. 3.21b) to the encapsulated graphene layer is applied to hBN/SLG/hBN stack and the drain
Fig. 3.21 (a) Schematic of the hBN/SLG/hBN photodetector on a buried silicon waveguide. (b) Cross section view of the side-contacted hBN/SLG/hBN multilayer stack. (c) Optical microscope image of the as-fabricated device. (d) Responsivity mapping as a function of gate-source $V_{GS}$ and drain-source $V_{DS}$ voltages. (e) High-speed response of the graphene photodetector. The dashed line shows the fitting to the experiment results with a RC low-pass filter model. 12 Gbit s$^{-1}$ optical data link test of the device, showing a clear eye opening.

electrode is positioned only 200 nm from the waveguide to induce a pn junction near the optical mode [98]. Figure 3.21c shows the completed structure.

Figure 3.21d shows the photoresponse of the detector via illuminating the waveguide with a continuous-wave (c.w.) laser at 25 $\mu$W. The responsivity is defined as the ratio of the short-circuit photocurrent ($I_{ph}$) to the optical power $P_{in}$ in the waveguide, $R = I_{ph}/P_{in}$. Since the metal-induced pn junction near the drain contact serves to separate the photoexcited carrier and generate photocurrent, the doping profile of the pn junction is critical to achieve maximum responsivity of the graphene photodetector. A polymer electrolyte (poly(ethyleneoxide) and LiClO$_4$) layer covering the entire chip serves to independently tune the graphene Fermi level and electric field across the waveguide mode [99]. The responsivity of the detector as a function of $V_{GS}$ and $V_{DS}$ shows a six-fold pattern in the photocurrent, which qualitatively matches the behavior of the photothermoelectric (PTE) effect [47]. The photocurrent reaches a maximum of 0.36 A/W at $V_{GS} = 2$ V and $V_{DS} = 1.2$ V.

Figure 3.21e shows the operation speed of the photodetector measured in the frequency range from 1 to 50 GHz, indicating a 3-dB cut-off frequency at
42 GHz. To gauge the viability of the waveguide-integrated graphene photodetector in realistic optical applications, an optical data transmission was performed. A pulsed pattern generator with a maximum 12 Gbit/s internal electrical bit stream modulated a 1550 nm CW laser via an electro-optic modulator, which was launched into the waveguide-graphene detector. The output electrical data stream from the graphene detector was amplified and sent to a wide-band oscilloscope to obtain an eye diagram. As shown in the inset of Fig. 3.21e, a clear eye-opening diagram at 12 Gbit/s was obtained.

To date, waveguide-integrated graphene photodetectors have been extensively studied, showing possible integration in CMOS-compatible processes [98]. In addition, large-scale epitaxial graphene samples integrated with waveguides can achieve photodetection with a data rate of 50 GHz [100], and high-responsivity by integrating with a silicon slot waveguide [101]. Owing to the broadband absorption of graphene, waveguide-integrated graphene heterostructure also enables photodetectors for mid-infrared wavelengths [102]. For other 2D materials, black phosphorus (BP)-based photodetectors have drawn great attention due to its small bandgap that is promising for telecommunication and mid-IR wavelength ranges [103, 104]. Heterogeneously integrated BP-silicon photodetectors have shown up to 6 A/W responsivity with more than 3 GHz speed.

In addition to E–O modulators and photodetectors, other on-chip devices including all-optical modulators [105], mode-locked ultrafast laser [106], thermo-optic modulators [107], light-emitting diodes [108], and single photon sources [109] could be integrated on a silicon platform in the future, thanks to 2D materials such as transition metal dichalcogenides (TMDs) [110], black phosphorus [104, 111], and superconducting 2D niobium diselenide [112] (NbSe2). The seamless heterogeneous integration enabled by the low processing temperatures of 2D materials will allow an efficient optical interconnect with conventional silicon CMOS technology for inter- and intra-chip communication, back-end deposited silicon photonics [113], mid-infrared photonics [114], and also flexible photonics [115].

### 3.5 2D Material-Based Chemical and Biological Sensors

Chemical and biological sensors are ubiquitous in every aspect of our lives—from health care and environmental monitoring all the way to transportation and industry. Sensors translate a physical, chemical, or biological input signal, for example a gas concentration, into an electrical signal such as a change in voltage or current. In order to benchmark the performance of this transduction process, different metrics exist. One of the most important parameters is the sensitivity of a sensor, which is defined as proportionality factor between input signal and electrical output signal. A sensor’s selectivity is given by its ability to respond to only one input signal and to be insensitive to all other relevant environmental changes. Another important characteristic is the detection limit, which marks the lower bound of what can
be measured with confidence. Lastly, sensors exhibit a finite *response time* and *recovery time*. They are defined as the delay between a step increase or decrease of the input signal, respectively, and the output reaching a steady state [116, 117].

The integration of chemical and biological sensors with silicon chips could play an important role in the goal of extending the capabilities of the traditional CMOS platform. However this heterogeneous integration has traditionally been difficult due to the instability of SiO$_2$ dielectrics in the presence of many analytes of interest, as well as the large thickness of the interconnect and back-end layers in state-of-the-art silicon chips, which increase the distance between the chip surface and the transistors. 2D materials such as graphene or molybdenum disulfide could change this, and make the seamless integration of chemical and biological sensors with silicon chips a reality. These materials have excellent physical and chemical properties that make them perfectly suited as sensor materials. They are atomically thin, resulting in the highest possible surface to volume ratio and making them extremely sensitive to environmental changes. Furthermore, 2D materials have intrinsically low electronic noise, are biocompatible and mechanically strong and, as mentioned in other sections of these chapter, they can be seamlessly integrated with Silicon microsystems.

This section is organized into two parts. The first half gives an overview of gas and chemical sensors based on 2D materials. The second half focuses on biological sensors made of these materials. The aim of this section is to highlight a few promising sensor concepts rather than to give detailed account of all the extensive research work. In addition, it focuses on sensor solutions that can be eventually integrated with CMOS electronics.

### 3.5.1 Gas and Chemical Sensors

Gas and chemical sensors are used, for example, in our homes to detect carbon monoxide, in security applications to identify explosives or in cars to monitor exhaust and lower hazardous emissions. Sensors can exploit different physical phenomena to detect and quantify a certain gas or chemical, for example changes in the optical, magnetic, or electrical properties of the sensor material. Electrical gas sensors are arguably one of the most common types. They use changes in sensor resistance or capacitance in a variety of materials such as semiconductors, metal thin film, metal oxides, and polymers to facilitate the sensing [118].

Recently, 2D materials have attracted a large amount of interest as gas or chemical sensor due to their high sensitivities and low electronic noise. Both of these properties are rooted in their atomic thickness which yields a very large surface to volume ratio. This makes 2D material very susceptible to the environment with molecules adsorbing on the surface which leads to a charge transfer and dopes the channel with electrons or holes. This increase in carrier concentration strongly alters the channel resistance of the device [119].
In the following paragraphs, a few promising examples of gas sensors using graphene and other 2D materials are given. For a more in-depth discussion, the interested reader is referred to comprehensive literature review papers by Yuan and Shi and Vergheese et al. [120, 121] covering graphene-based gas sensors as well as another review paper by Vergheese et al. [119] which focuses on gas sensors based on 2D materials (in addition to graphene).

In 2007 Schedin et al. demonstrated the first graphene gas sensor based on a simple Hall bar structure. In this device, gas molecules adsorb on the surface of graphene which dope the material and change its electrical properties. Based on an optimized design and the intrinsically low noise level of graphene, the group was able to detect individual molecule ad- and desorption events which is the ultimate sensitivity limit [122]. More recently, Chen et al. simplified the sensor design to a two-terminal chemically sensitive resistor, or chemiresistor, and added UV-illumination for continuous cleaning of the graphene surface. This way, they were able to demonstrate a detection limit as low as 0.16 parts-per-trillion for nitric oxide gas molecules [123].

Another way to sense gases using graphene is to exploit the influence of specific analytes on the low frequency noise level of graphene [124]. Using a back-gated transistor configuration, Rumyantsev et al. found that besides changing the channel resistance, certain gases introduce a Lorentzian noise bulge at a distinctive center frequency (see Fig. 3.22) that differs for each gas and is reproducible across samples. This adds another sensing parameter to chemically sensitive field effect transistors, or chemFETs, which holds the promise of achieving better sensor selectivity through recognizing a distinctive signature of Lorentzian center frequency and change in channel resistance.

Reduced graphene oxide is a graphite-derived material that can be easily solution-processed. Fowler et al. used it to create gas sensors on a chip with an

![Fig. 3.22](image) (a) Scanning electron microscope (SEM) image and (b) Noise spectral density $S_i/\nu f$ multiplied by frequency $f$ versus frequency $f$ for the device in open air and under the influence of various solvent vapors. Different vapors induce noise with different characteristic frequencies $f_c$. The difference in the frequency $f_c$ is sufficient for reliable identification of different gases with the same graphene transistor [124]
integrated heater for temperature control by spin coating a graphene oxide solution on the chip and reducing it with anhydrous hydrazine. They found the device to have a detection limit of 52 ppb to 2,4-dinitrotoluene (DNT), which is a molecule related to explosive detection. Furthermore, their temperature experiments showed that a substrate temperature of 149 °C helps to decrease the response and recovery time significantly in comparison to room temperature. The benefit however comes as the expense of lowered sensitivity [125]. More recently, reduced graphene oxide sensors were also successfully fabricated by inkjet printing [126, 127].

2D material based gas and chemical sensor are not just limited to graphene. With the exploration of other 2D materials like transition metal dichalcogenides, viable sensors were also shown with molybdenum disulfide (MoS2) [128–131] and phospherene [132] on top of a silicon wafer. Late et al. analyzed the response of MoS2 chemFETs to ammonia and nitrogen dioxide gas while exploring the influence of light illumination, gate bias and number of atomic layers. They found that five-layer MoS2 devices were more sensitive to gas analytes than their bilayer counterparts. Furthermore, applying a gate bias helps to increase the device sensitivity by electrostatically changing the carrier concentration in the channel to an optimal point [129]. In a related study by Perkins et al. the authors demonstrated single-layer MoS2 chemiresistors that are sensitive to different organic compounds such as triethylamine (TEA) or acetone with a detection limit of 10 ppb and 500 ppm, respectively. They also found that the response to these organic vapors is often complementary in polarity to carbon nanotube sensors [130]. Another way to tailor the selectivity of MoS2 to a specific gas is by grafting functional groups onto its surface. Kim et al. demonstrated this approach by attaching a thiolated ligand called mercaptoundecanoic acid (MUA) to chemiresistor devices. The sensors were built with interdigitated electrodes on a layer of vacuum-filtrated MoS2 flakes with or without this functionalization that results in different responses towards volatile organic compounds (see Fig. 3.23a). In particular, the untreated chemiresistors showed an increase in channel resistance being exposed to oxygen-functionalized

![Fig. 3.23](image-url) (a) Schematic of MoS2 exfoliation and functionalization with MUA. The chemiresistors were built by vacuum filtration and placed in a test chamber that was exposed to different organic volatile compounds (VOC). (b) Relative change in channel resistance of untreated and MUA-treated MoS2 chemiresistors as a result of exposure to different VOCs [131]
VOCs while the MUA treated samples showed a decreasing resistance as shown in Fig. 3.23b. Furthermore, the group showed a detection limit of these sensors below 1 ppm. Considering different available thiolated ligands, this work demonstrates the potential of targeted MoS₂ functionalization to create sensor arrays that are sensitive to a variety of gaseous compounds [131].

### 3.5.2 Biological Sensors

Biological sensors are another important subgroup of sensors, which is mainly used in medicine to determine blood sugar levels, identify bacteria, cancer cells, and other analytes of interest in the laboratory. There is a large variety of biosensor technologies, although many use optical detection through labeling with fluorescent dyes or electrochemical detection techniques using functionalized electrode surfaces [133, 134].

Recently, 2D materials such as graphene have been extensively explored as biosensors. Besides their low intrinsic noise and large surface to volume ratio, 2D materials are especially compelling to use as biosensors because of their chemical inertness and easier handling compared to other new materials such as carbon nanotubes. The paragraphs below summarize some of the approaches to use graphene, MoS₂, and other 2D materials in monitoring cells or pH levels and detect DNA, biomolecules and other proteins. Most of these devices were fabricated on top of silicon wafers and could easily be integrated with silicon electronic chips. This could eventually enable small, inexpensive diagnostic tests for laboratory and personal use. For a more detailed literature review, the interested reader is referred to more comprehensive review papers by Liu et al. and Moldvan et al. [135].

Being able to sense pH levels is important in biology and medicine, for example for cell monitoring or diagnosing disease in humans [133]. Ang et al. demonstrated that graphene electrolyte gated FETs (EGFETs) can be used to measure pH levels. In particular, they found that the hydroxyl (OH⁻) and hydronium (H₃O⁺) change the channel carrier concentration through capacitive coupling and hence change the neutrality point of the chemFET transfer characteristics with a sensitivity of 99 mV/pH [136]. In a similar work with monolayer graphene, Ohno et al. were able to improve the sensitivity down to 25 mV/pH [137].

In the realm of biological cell detection, Huang et al. demonstrated a graphene chemiresistor that can sense *E. coli* bacteria by functionalizing the graphene surface with anti-*E. coli* antibodies. This approach is much simpler and faster than traditional solution-based techniques for detecting bacteria. Huang’s device was able to detect bacteria concentrations as low as 10 colony-forming cells per milliliter (cfu/mL) and also showed high selectivity against other bacteria [138]. Another interesting graphene biosensor application was highlighted by Ang et al. who fabricated a microfluidic channel equipped with protein-functionalized graphene chemFETs that could detect malaria infected red blood cells. The group showed the selective capture of infected blood cells to a transistor channel, with a much
simpler setup compared to traditional optical detection methods. This capture event changes the channel conductance when in contact with a cell due to capacitive charging. By analyzing the dwell time and magnitude of the conductivity change, they could also distinguish between two different stages of the infected blood cells [139]. Graphene transistors can also be used to monitor the action potential of cells for neural prosthesis application [140]. Hess et al., for example, demonstrated graphene EG-FET arrays that are very biocompatible and are able to monitor the action potential of cells grown on top of the channel [140]. In their experiments, the graphene arrays proved to be more chemical resistant and exhibit far higher gate sensitivities than traditional silicon FET arrays. Lastly, Jiang et al. fabricated hemin-functionalized graphene chemFETs with integrated microfluidic delivery system to sense nitric oxide (NO), an important by-product of many cell reactions. With their devices, they demonstrated nanomolar NO sensitivity with sensor areas of 0.25–1 μm², which is similarly sensitive but much smaller than state-of-the-art electrochemical sensors. Hence, this approach paves the way to a specially resolved cell monitoring [141].

DNA sensors are becoming a vital tool for genetic screening and pathogen detection that help diagnose diseases. However, the current technology for DNA sensing relies on marking it with fluorescent dyes and optically detecting it. Recently, 2D materials have been intensively studied for this application due to their easy functionalization and electronic integration, which could lead to compact biochips that can detect DNA in real time. In 2008, Mohanty and Berry used reduced graphene oxide (rGO) functionalized with single-stranded DNA to build highly sensitive and selective DNA chem-FETs that changed between 60–200% in channel conductance after DNA hybridization [142]. A slightly different design by Stine et al. improved the selectivity of this sensor concept by introducing a second reference transistor without functionalization that could compensate for sensor drift and non-specific binding [143]. In an effort to make graphene DNA sensors more scalable, Ping et al. recently reported on graphene transistor arrays with 52 devices functionalized with single-stranded DNA as depicted in Fig. 3.24a. The sensors measure DNA concentration by a positive shift Dirac voltage, see Fig. 3.24b, which is ascribed to an increase in positive carrier concentration induced by the negatively charged phosphate groups of the target DNA molecules. In their experiments, the transistor arrays had a device yield of over 90%, good reproducibility and could achieve a detection limit of down to 1 fM for 60mer DNA strands. Furthermore, the sensor selectivity was tested by analyzing the response to DNA strands that had one or two mismatching base pairs either at the center or the end of the stand. The group found that a mismatch at the end or the center of a sequence resulted in a 20% reduction or 90% reduction in Dirac voltage shift, respectively, which highlights the good selectivity of the sensor to even small mismatches [144].

Detecting proteins and biomolecule such as glucose or immunoglobulin has also been demonstrated with 2D materials by using different kinds of functionalization [145–149]. Hunag et al., for example, used graphene chemFETs decorated with glucose oxidase to sense glucose in solution down to a level of 100 μM which is comparable to common electrochemical sensors. Using glutamic dehydrogenase as
immobilized enzyme, they could furthermore measure glutamate with a detection limit of 5 μM [145, 147]. A group around Mao et al. chose a different approach to detect immunoglobulin G (IgG). They build reduced graphene oxide chemiresistors and deposited gold nanoparticles by a combination of electrospraying and electrostatic assembly that were decorated with matching antibodies.

After exposing the devices to an IgG solution for 1 h, washing and drying the chemiresistors, the group was able to show statistically significant changes in channel resistance at IgG concentrations as low as 2 ng/mL while maintaining a good selectivity against other types of immunoglobulin [147]. Using exfoliated MoS2, Lee et al. recently demonstrated a prostate specific antigen biosensor which is an important tool to detect prostate cancer. The group functionalized the MoS2 transistors by using PSA antibodies that nonspecifically physisorb on the sensor surface. Compared to the pristine devices this leads to an increased baseline off-current of the transistor due to doping of the positively charged antibodies. Once the biosensors are exposed to a PSA solution, the negatively charged PSA antigens selectively bind to their respective antibodies, which decreases the transistor off-current again and is illustrated in Fig. 3.25. With this approach, the researchers were able to demonstrate a detection limit of 1 pg/mL which is three orders of magnitude below the clinical cut-off level [149].

Many of the aforementioned biosensors are based on electrolyte gated field effect transistors (EGFETs). To better understand the variations that are associated with building these arrays, large-scale sensor systems based on graphene EGFETs have been fabricated [150]. These systems contain up to 256 individual devices that can be analyzed by DC-measurements within minutes, which enables meaningful statistical analysis. Given the low thermal budget involved in the fabrication of these sensor arrays, their integration with fully fabricated silicon chips is expected to be imminent, and very similar to what has already happened with infrared detectors.
3.6 Conclusion

2D materials, thanks to their atomically thin nature and diverse electronic properties, are promising candidates for a variety of applications. Semiconducting MoS$_2$ is, for example, an ideal channel material to both provide power gating in future Si chips and replace Si in sub-10 nm technology. At the same time, graphene, a gapless semiconductor, can serve as a core material combined with a Si photonic structure for fast photodetection and optical modulation on a CMOS platform. The transferable nature of atomically thin 2D materials and the low thermal budget of their processing technology allows seamless integration of these and other 2D devices, circuits and systems on a Si platform to develop complex 3D systems. Such 3D integration will significantly increase the performance of future Si microsystems, enabling exciting new opportunities for hybrid systems.

References

10. L. Yang, R.T. Lee, S.P. Rao, W. Tsai, P.D. Ye, 10 nm nominal channel length MoS2 FETs with EOT 2.5 nm and 0.52 mA/μm drain current, in *2015 73rd Annual Device Research Conference (DRC)* (2015), pp. 237–238


73. T. Low, P. Avouris, Graphene plasmonics for terahertz to mid-infrared applications. ACS Nano 8, 1086–1101 (2014)


100. D. Schall, D. Neumaier, M. Mohts, B. Chmielak, J. Bolten, C. Porschatis, et al., 50 GBit/s photodetectors based on wafer-scale graphene for integrated silicon photonic communication systems. ACS Photonics 1, 781–784 (2014)
133. X. Zhang, H. Ju, J. Wang, Electrochemical sensors, biosensors and their biomedical applications (Elsevier, New York, 2008)


144. J. Ping, Scalable production of high sensitivity, label-free DNA biosensors based on back-gated graphene field effect transistors. ACS Nano 10, 8700–8704 (2016)


150. C. Mackin, T. Palacios, Large-scale sensor systems based on graphene electrolyte-gated field-effect transistors. Analyst 141, 2704–2711 (2016)