Nonlinear optical properties and applications of 2D materials: theoretical and experimental aspects

Abstract: In this review, we survey the recent advances in nonlinear optics and the applications of two-dimensional (2D) materials. We briefly cover the key developments pertaining to research in the nonlinear optics of graphene, the quintessential 2D material. Subsequently, we discuss the linear and nonlinear optical properties of several other 2D layered materials, including transition metal chalcogenides, black phosphorus, hexagonal boron nitride, perovskites, and topological insulators, as well as the recent progress in hybrid nanostructures containing 2D materials, such as composites with dyes, plasmonic particles, 2D crystals, and silicon integrated structures. Finally, we highlight a few representative current applications of 2D materials to photonic and optoelectronic devices.

Keywords: nonlinear optics; 2D materials; numerical algorithms; experimental techniques.

1 Introduction

Materials can exhibit a nonlinear optical response upon interaction with an electric field that is of the order of interatomic fields (10⁵–10⁸ V/m) [1]. Whereas a nonlinear optical response of solids and liquids subjected to strong external DC electric fields was observed by John Kerr all the way back in 1875 [2], the birth of modern nonlinear optics can perhaps be traced to the experimental demonstration of second-harmonic generation (SHG) by Franken et al. [3]. The main development that made this possible has been the demonstration of the first laser [4]. The optical nonlinearity of an optical medium manifests itself through field-dependent variations of optical constants, such as absorption and index of refraction, a phenomenon that is usually accompanied by the generation of new optical frequencies. In this context, the nonlinear optical response of materials can be classified with respect to the mechanisms responsible for the variation of optical constants; some of the most notable types of optical nonlinearity are electronic optical nonlinearity [5], thermally induced optical nonlinearity [6], and external field-induced optical nonlinearity [7–10].

With very few exceptions (e.g. surface SHG), nonlinear optics has been concerned with nonlinear optical interactions that take place in three-dimensional (3D) bulk optical media. This paradigm has changed dramatically with the recent advent of two-dimensional (2D) materials, as these newly discovered materials provide a novel 2D platform to study a multitude of nonlinear optical effects. Thus, research in 2D layered materials (2DLMs) has perhaps started with the discovery of graphene [11], and the unique and remarkable properties of this 2D form of graphite have sparked rapidly growing research interest in the physics and applications of this 2D material. In particular, graphene is beginning to find applications in many key areas of optics and photonics. This might be surprising at first sight as graphene only absorbs about 2.3% of the incident light [12], so it barely interacts with electromagnetic waves. Despite this, its zero band-gap nature, unusually large chemical and electrical tunability, and effects such as Pauli blocking provide unique functionalities for photonic nanodevices. For example, graphene exhibits saturable absorption behavior [13, 14], which can play an
important role in lasing applications. Moreover, properties such as large nonlinear optical response, ultrafast photoexcitation dynamics, high chemical and mechanical stability, and large thermal and optical threshold damage of graphene make it an ideal test-bed for studying nonlinear optics phenomena in 2D physical systems [5–7, 15].

Other classes of recently discovered 2DLMs, namely transition metal chalcogenides (TMCS) such as MoS$_2$, MoSe$_2$, MoTe$_2$, WS$_2$, WSe$_2$, and TiS$_2$, gallium selenide (GaSe), black phosphorus (BP), hexagonal boron nitride (h-BN), and perovskites, have broadened the set of specific properties and functionalities 2D materials possess and consequently have widened the spectrum of their technological applications. The strong optical nonlinearity and ultrafast response of these materials have been successfully employed in all-optical modulators, saturable absorbers (SAs) used in passive mode locking and Q-switching, wavelength converters, and optical limiters [16–19]. The success of these 2D materials in nonlinear optics resides in the fact that they meet several requirements that an ideal nonlinear optical material should fulfill, including large and ultrafast nonlinear optical response, broadband and tunable optical absorption, ultrafast recovery time, large optical and thermal damage threshold, high chemical and mechanical stability, and low fabrication costs.

In this article, we review the recent theoretical and experimental developments pertaining to nonlinear optics in photonic structures containing graphene and other 2D materials. The paper is organized as follows. In the next section, we introduce several basic theoretical concepts concerning linear and nonlinear optical properties of 2D materials. Then, in Section 3, we briefly present several experimental developments pertaining to nonlinear optics in 2D materials. In this area of research. In this section, a summary of the main concepts forming the basis of a theoretical description of linear and nonlinear optical properties of 2D materials is given.

## 2.1 Linear optical properties of 2D materials

As graphene and other 2D materials are physical systems consisting of a single atomic layer, their optical properties are conveniently characterized by surface quantities. For instance, assuming that graphene lies in the $x$-$y$ plane, its linear surface conductivity tensor is generally represented as

\[
\sigma = \sigma_{xx} \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix}.
\]

This means that the components containing the $z$-coordinate vanish, that is $\sigma_{xz}=\sigma_{zx}=\sigma_{yz}=\sigma_{zy}=0$. More generally, considering the magneto-optical effects [20–22], the surface conductivity tensor can be rewritten as

\[
\sigma = \begin{pmatrix} \sigma_L & \sigma_H \\ -\sigma_H & \sigma_L \end{pmatrix},
\]

where the longitudinal conductivity $\sigma_L$ and the Hall conductivity $\sigma_H$ can be determined from Kubo formalism [23].

At room temperature and for frequencies below the mid-infrared (IR) region, the longitudinal conductivity and Hall conductivity can be cast in the Drude model form [24]:

\[
\sigma_L = \sigma \frac{1-i\omega \tau}{(\omega \tau)^2+(1-i\omega \tau)^2},
\]

\[
\sigma_H = \sigma \frac{i \omega \tau}{(\omega \tau)^2+(1-i\omega \tau)^2},
\]

where $\tau=1/(2\Gamma)$ is the scattering time ($\Gamma$ is the plasmon damping rate), $\mu=e\hbar v_f^2/\mu_e$ is the chemical potential ($\hbar$ is the reduced Planck constant, $v_f=10^6$ m/s is the Fermi velocity, and $\mu_e$ is the carrier density), $\omega = eB_y v_f / \mu_e$ is the cyclotron frequency ($e$ is the charge of the electron and $B_y$ is the $z$-component of the magnetic field), and $\sigma$ is given by

\[
\sigma = \frac{2e^2 k_B T}{\pi \hbar^2} \ln \left[ 2 \cosh \left( \frac{\mu^-}{2k_B T} \right) \right].
\]

In the equation above, $T$ is the temperature and $k_B$ is the Boltzmann constant.
For frequencies above the mid-IR region, the general expression [20] for the longitudinal and Hall conductivity cannot be simplified to a Drude model. However, in the case of no magnetostatic bias ($B_z = 0$), the longitudinal conductivity $\sigma_L = \sigma_{\text{intra}} = \sigma_{\text{inter}}$ is due to intraband ($\sigma_{\text{intra}}$) and interband ($\sigma_{\text{inter}}$) contributions. The intraband electron-photon scattering processes can be evaluated as [25, 26]

$$\sigma_{\text{intra}}(\omega) = \frac{2e^2 k_B T}{\pi h} \left[ \frac{\tau}{1 - i\omega \tau} \ln \left( 2 \cosh \left( \frac{\mu_c}{2k_B T} \right) \right) \right].$$ (6)

The interband conductivity originates from direct interband electron transitions. It is usually ignored at room temperatures and for frequencies below the mid-IR region, as it is much smaller than the intraband term. It can be expressed in an integral form as

$$\sigma_{\text{inter}}(\omega) = \frac{e^2}{4\hbar} \int G(\omega) - G(\omega/2) + \frac{i\omega}{\pi} \int G(s) - G(\omega/2) ds \, ds,$$ (7)

where

$$G(s) = \frac{\sinh[hs/(k_B T)]}{\cosh[\mu_c/(k_B T)] + \cosh[hs/(k_B T)]}.$$ (8)

The total surface longitudinal conductivity $\sigma_L$ is illustrated in Figure 1 in the case of $T = 0$ K, $\tau = 0.125/\pi$ ps, and $\mu_c = 0.6$ eV, where $\sigma_0 = e^2/(4\hbar)$ is the universal dynamic conductivity. This figure shows that $\sigma_L$ is similar to that of noble metals for the energy below the mid-IR region ($\lesssim 1$ eV). On the contrary, it suggests that graphene behaves as a dielectric material for frequencies above the mid-IR region.

In many computational electromagnetic methods, it is more convenient to work with bulk rather than surface quantities; therefore, one often introduces bulk equivalents of the surface quantities. In particular, instead of the sheet conductance, one uses a bulk conductivity $\sigma_b = \sigma_L/h_{\text{eff}}$, where $h_{\text{eff}}$ is the effective thickness of the 2D material. Moreover, the electromagnetic properties of 2D materials can alternatively be described by the electric permittivity $\varepsilon$, which is related to the conductivity via the relation:

$$\varepsilon_{\text{r}} = 1 + \frac{i\sigma_b}{\varepsilon_0 \omega h_{\text{eff}}},$$ (9)

The electric permittivity of transition metal dichalcogenide (TMDC) monolayered materials can be described as a superposition of $N$-order Lorentzian functions:

$$\varepsilon_{\text{r}} = 1 + \sum_{k=1}^{N} \frac{f_k}{\omega_k^2 + i\gamma_k^2 + (\omega - \omega_0)^2},$$ (10)

where $f_k$, $\omega_k$, and $\gamma_k$ are the oscillator strength, resonance frequency, and spectral width of the $k$th oscillator, respectively. The values of the model parameters for four TMDC monolayers are determined by fitting Eq. (10) to the experimental data provided in Ref. [27]; the corresponding values of the surface conductance are depicted in Figure 2.

Figure 1: Dispersion of the complex sheet conductance of graphene in the interband and intraband wavelength range. Reproduced from Ref. [26].

Figure 2: Dispersion of the complex sheet conductance of several TMDCs.
Frequency dependence of the real (A) and imaginary (B) parts of the sheet conductance of several TMDC monolayered materials, respectively. Reproduced from Ref. [26].
2.2 Nonlinear optical properties of 2D materials

The crystal lattices of graphene and TMDC monolayers belong to different space symmetry groups, which means that each class of 2D materials requires a separate treatment. The graphene lattice belongs to the $D_{6h}$ point group, meaning that it is a centrosymmetric material; thus, SHG in graphene is a forbidden nonlinear optical process. On the contrary, third-harmonic generation (THG) is an allowed, particularly strong process in graphene \[27 –30\], which makes it a suitable material for nonlinear optical applications. In contrast, TMDC monolayers belong to the $D_{3h}$ point group \[31\], so that in this case SHG is the lowest-order nonlinear optical process.

The nonlinear properties of 2D materials are generally described by a nonlinear surface conductivity tensor $\sigma_s^{(n)}(\Omega, \omega)$, where $\Omega = n\omega$ and $n = 2, 3, \ldots$ For instance, the scalar part of the third-harmonic (TH) surface conductivity tensor $\sigma_s^{(3)}$ of graphene is illustrated in Figure 3A; the corresponding expression is given by

$$
\sigma_s^{(3)}(3\omega, \omega) = \frac{i\sigma_s(h\nu e)^2}{48\pi(h\omega)^{3/2}} \frac{h\omega}{2\mu c} \Omega \omega
$$

(11)

where $T(x) = 17F(x) - 64F(2x) + 45F(3x)$ and $F(x) = \ln|1 + x|/(1 - x) + izH(|x| - 1)$. Using this, the frequency-dependent form of nonlinear current density $J_{d}^{NL}$ can be evaluated as

$$
J_{d}^{NL}(\Omega, \omega) = -i\Omega e \left\{ \sum_{i=2}^{N} \chi^{(n)}(\Omega, \omega)[E(\omega)]^{n-1} \right\} E(\omega),
$$

(12)

where

$$
\chi^{(n)}(\Omega, \omega) = i\sigma_s^{(n)}(\Omega, \omega)/(\epsilon_0 \Omega \hbar c)
$$

(13)

is the $n$-th-order nonlinear susceptibility.

Combining Eqs. (12) and (13), the nonlinear current can be written in terms of the nonlinear conductivity as

$$
J_{d}^{NL}(\Omega, \omega) = \sum_{i=2}^{N} \sigma_s^{(n)}(\Omega, \omega) \left[ E(\omega) \right]^{n}/\hbar c.
$$

(14)

Generally, the surface conductivity $\sigma_s^{(n)}$ is an $(n+1)$-rank tensor. For instance, the TH surface conductivity tensor $\sigma_s^{(3)}$ of graphene has 27 elements. However, there are only eight nonzero components, among which three are independent. As a result, the third-order current density of graphene is evaluated as

$$
J_{d}^{NL}(3\omega, \omega) = \sigma_s^{(3)} \sum_{\vec{m} \vec{n}} \Delta^{(3)}_{\vec{m} \vec{n}} E(\omega) E_{\vec{m}}(\omega) E_{\vec{n}}(\omega),
$$

(15)

where $\Delta^{(3)}_{\vec{m} \vec{n}} = \delta_{\vec{m} \vec{n}}(\delta_{\vec{x} \vec{x}} + \delta_{\vec{y} \vec{y}} + \delta_{\vec{x} \vec{y}}) / (3\hbar c)$, and $\delta_{\vec{m} \vec{n}}$ is the Kronecker delta.

Unlike graphene, TMDC monolayers are noncentrosymmetric; therefore, SHG is allowed \[32, 33\]. Based on the symmetry properties of their point group, $D_{3h}$, it can be shown that their nonlinear susceptibility tensor has only one independent, nonvanishing component \[33\], namely

$$
\chi^{(2)}_{b,0} = \chi^{(2)}_{b,xx} = -\chi^{(2)}_{b,yy} = -\chi^{(2)}_{b,zy} = -\chi^{(2)}_{b,yx},
$$

(16)

where $x$ is the armchair direction of the monolayer and $y$ is the orthogonal zigzag direction. The values of $\chi^{(2)}_{b,0}$ over a small spectral range for WS$_2$, MoS$_2$, and WSe$_2$ can be found in Refs. [33–37] and are given in Figure 3B. The largest values in the corresponding spectral domain are $\chi^{(2)}_{b,WS_2} = 1140$ pm/V, $\chi^{(2)}_{b,MoS_2} = 132$ pm/V, and $\chi^{(2)}_{b,WSe_2} = 67$ pm/V. All of these values are comparable to that of GaAs [37], $\chi^{(2)}_{b,GaAs} = 740$ pm/V, which is a material with strong bulk second-order susceptibility.

3 Computational methods for 2DLMs

The advent of 2D materials has created a large demand for accurate and efficient numerical methods suitable for
3.1 Frequency-domain numerical modeling of 2D materials

The rigorous coupled-wave analysis (RCWA) algorithm, one of the most widely used frequency-domain methods for modeling photonic devices, is a key numerical method to describe the interaction of electromagnetic waves with periodic structures. It has been used extensively and successfully over the last few decades [38–42]. To extend this powerful numerical method to nonlinear optical materials [43–47], a new generalized source (GS)-RCWA hybrid method has been proposed [26, 48]. The basic steps of this method, adapted for 2D materials, are briefly summarized as follows.

**Step 1: Linear simulation at fundamental frequency (FF)**

In the first step, one assumes that there are only linear materials in the computational region and evaluates the frequency-domain near-field distribution using the RCWA method. In RCWA, the permittivity of a periodic structure with periods \((\Lambda_x, \Lambda_y)\) is expanded as a 2D Fourier series,

\[
\varepsilon_r(x, y, z) = \sum_{n_1=\ldots}^{\ldots} \sum_{n_2=\ldots}^{\ldots} \varepsilon_{r, n_1, n_2}(z) e^{i \frac{2\pi n_1 x}{\Lambda_x}} e^{i \frac{2\pi n_2 y}{\Lambda_y}},
\]

where \(n=(n_1, n_2)\). If one assumes that a plane-wave \(E(r, t) = E_0 e^{i(kr - \omega t)}\), where \(k\) is the wavevector, illuminates this periodic structure, the electromagnetic field in the computational region can be expressed as a Fourier series as well, which is truncated at finite numbers \((N_x, N_y)\) of harmonics in the \(x\) and \(y\) directions, respectively. Substituting these truncated Fourier series into Maxwell equations, the original differential equations are cast into an eigenvalue problem:

\[
M(E_x | E_y | H_x | H_y)^T = \kappa (E_x | E_y | H_x | H_y)^T,
\]

where \(M\) is the system-matrix and the symbol \(\kappa\) denotes the complex propagation constant. This equation can be solved by many well-known numerical methods.

So far, we only considered the case of one periodic layer. However, in practical applications, a number of optical structures contain multiple layers of 2D materials. To study such multilayer (ML) structures, a staircase approximation in the \(z\)-direction and boundary conditions for the fields between adjacent layers are employed. To implement the boundary conditions, a so-called scattering matrix (S-matrix) formalism is used to match the mode coefficients of adjacent layers and hence to determine the fields in the entire structure. The final S-matrix relation in linear simulation can be simply expressed as

\[
\begin{pmatrix}
a^+ \\
b^+
\end{pmatrix}
= S^{44}
\begin{pmatrix}
a^- \\
b^-
\end{pmatrix},
\]

where matrix \(S^{44}\) incorporates the upward and downward propagation of the incoming modes \(b^+\) and \(a^-\) in layers \(B\) and \(A\), respectively, and determines the coefficients \(a^+\) and \(b^+\) of the outgoing modes at the \(A/B\) interface. Based on Eqs. (17) to (19), one can easily evaluate the FF field distribution located at the surface of 2D materials.

**Step 2: Nonlinear GS evaluation**

Using the electric field at the FF obtained in Step 1, one evaluates the frequency-domain nonlinear surface current density \(J^NL\) via Eq. (14). In the case of 2D materials, the nonlinear surface current lies in the plane of the material; thus, the value of the normal component of electric field in Eq. (14) is zero.

**Step 3: Linear simulation at higher-order harmonics**

In the nonlinear simulation, the nonlinear surface current density calculated using Eq. (14) is treated as a nonlinear GS, which is used to excite the same structure that was investigated in Step 1. To incorporate this GS into the RCWA method, one assumes that the infinitely thin graphene is placed at the interface between two adjacent bulk layers. Therefore, the nonlinear GS corresponding to graphene can be implemented via field boundary conditions as

\[
\mathbf{n} \times [H^{(\alpha)}(x, y, z) - H^{(\beta)}(x, y, z)]
= J^{NL}(x, y) + J^{NL}(x, y),
\]

where \(A\) and \(B\) denote two adjacent bulk layers and \(z\) indicates the position of the graphene layer. Combining Eq. (19) and Eq. (20), the final S-matrix relation in nonlinear simulations is expressed as

\[
\begin{pmatrix}
a^+ \\
b^+
\end{pmatrix}
= S^{44}
\begin{pmatrix}
a^- \\
b^-
\end{pmatrix},
\]

where matrix \(S^{44}\) incorporates the upward and downward propagation of the incoming modes \(b^+\) and \(a^-\) in layers \(B\) and \(A\), respectively, and determines the coefficients \(a^+\) and \(b^+\) of the outgoing modes at the \(A/B\) interface. Based on Eqs. (17) to (19), one can easily evaluate the FF field distribution located at the surface of 2D materials.
where \( \hat{a} \) and \( \hat{b} \) are the contributions of the nonlinear surface current to the total coefficients \( a^\dagger \) and \( b^\dagger \), respectively. Based on Eq. (21), one can compute the nonlinear outgoing radiated field at the higher-order harmonic.

### 3.2 Time-domain numerical modeling of 2D materials

As we discussed in the preceding subsection, the dispersive, anisotropic, and nonlinear properties of 2D materials can be readily implemented in the frequency-domain numerical method. However, in some cases, time-domain numerical methods, such as the finite-difference time-domain (FDTD) method [49], can be used to improve the computational efficiency, as the optical response in a specific spectral range can be obtained in a single time-domain simulation. Moreover, the FDTD numerical method is more general than the RCWA approach as the structure is not required to be periodic [50–55]. Therefore, in what follows, we briefly outline a time-domain numerical method to model linear and nonlinear properties of 2D materials.

To this end, one can extend the well-known FDTD method to the case of optical structures containing optically nonlinear 2D materials again using nonlinear GS. From the physics of SHG and THG [32], it is known that the nonlinear polarizations, which are the sources of SHG and THG, are entirely determined by the local field at the FF. As a result, and similar to the GS-RCWA method, one can study SHG or THG with only two linear FDTD simulations. In the first linear simulation, the system is excited by a linear source, whereas in the second simulation, the system is excited by nonlinear currents, that is, by a nonlinear GS. The basic steps of the GS-FDTD method are summarized in what follows.

#### Step 1: Linear simulation at FF

In the first linear FDTD simulation, one assumes that there are only linear materials in the computational region and stimulates this linear system with a source at the FF. As a result, one can calculate the time-domain near-field distribution at the FF using a single FDTD simulation. However, as we discussed in Section 2.1, most 2D materials are dispersive. To include these dispersive effects in the auxiliary differential equation (ADE)-FDTD method [56], a general dispersion model is introduced. Using a small set of fitting coefficients, the dispersion models most used in practice, namely Debye, Drude, Lorentz, and modified Lorentz, can be described by a common formula:

\[
\epsilon(\omega) = \epsilon_0 + \sum_{m=1}^{M} \epsilon_m(\omega),
\]

where \( \epsilon_0 \) is the frequency-independent part of the permittivity and \( M \) is the number of dispersion terms:

\[
\epsilon_m(\omega) = \frac{a_m^0 + a_m^1(-i\omega)}{b_m^0 + b_m^1(-i\omega) + b_m^2(-i\omega)^2},
\]

and \( a_m^0, a_m^1, b_m^1, b_m^2, \) and \( b_m^2 \) are dispersion coefficients defining the \( m \)th dispersion term. Based on this general dispersion model, the linear response of 2D materials can be calculated by the ADE-FDTD method whose iteration is given as

\[
E_{n+m} = \epsilon_0 E_n + \alpha_1 \left( \mathbf{C} \mathbf{H}_{n+m} - J_{n+m} \right),
\]

where \( \mathbf{C} \) is a discrete curl operator [57] and \( \alpha_1 \) and \( \alpha_2 \) are iterative coefficients defined as

\[
\alpha_1 = \frac{2 \epsilon_0 - \sigma \Delta t}{2 \epsilon_0 + \sigma \Delta t}, \quad \alpha_2 = \frac{2 \Delta t}{2 \epsilon_0 + \sigma \Delta t},
\]

and

\[
J_{n+m} = \sum_{m=1}^{M} \mathbf{J}_{m}^n + \sum_{m=1}^{M} \left( a_m^0(-i\omega) + a_m^1(-i\omega)^2 \right) E_{n+m}.
\]

#### Step 2: Nonlinear GS evaluation

Different from the first linear simulation, the second linear simulation at higher-order harmonic is excited with a nonlinear GS. Such nonlinear GS for the THG of graphene is evaluated from Eq. (15). To compute the frequency-domain near-field in Eq. (15), the time-domain near-field distribution at FF obtained in Step 1 is first Fourier transformed to the frequency domain. Subsequently, one uses Eq. (15) to evaluate the frequency-domain nonlinear current density. To incorporate these frequency-domain nonlinear current sources into the FDTD simulation, one inverse Fourier transforms them to obtain the corresponding time-domain nonlinear current sources. Note that the number of frequency sampling points should strictly satisfy the Nyquist-Shannon sampling theorem so that the final time-domain nonlinear current source can be calculated accurately via the inverse Fourier transformation.
Step 3: Linear simulation at high-order frequency

In the second linear FDTD simulation, one assumes that there are only linear materials. However, unlike the procedure in Step 1, the second linear FDTD simulation is excited with the time-dependent nonlinear current source obtained at Step 2 rather than an external excitation. In this way, one can accurately calculate the intensity of SHG or THG.

To illustrate the efficiency of the GS-FDTD method, a so-called double-resonance effect in graphene nanostructures [58] is described. As schematically illustrated in Figure 4, the graphene photonic structure consists of a periodic optical grating placed in the $x$-$y$ plane, with the ribbons oriented along the $x$-axis. The period in this example is $\Lambda = 100$ nm, the width of the ribbons is $W = 86$ nm, the chemical potential is $\mu_c = 0.6$ eV, the relaxation time is $\tau = 0.25$ ps, and the temperature is $T = 300$ K.

In the first linear FDTD simulation, graphene grating is illuminated by a plane wave with a Gaussian pulse envelope. This pulse covers the whole range of the FF, namely it extends from 30 to 150 THz. The angles characterizing the incident beam in Figure 4 are $\theta = 0$ and $\phi = 0$. Using the near-field distribution at the FF, the GS-FDTD method has been employed to calculate the nonlinear response of graphene grating; the corresponding results are given in Figure 5. Similar to the linear case, there are a series of strong peaks in Figure 5A; the near-field profiles for the first three peaks are depicted in Figure 5B–D. In addition, these results have been compared to the predictions of the GS-RCWA method [26], as in Figure 5A, showing a good agreement between the two methods.

4 Experimental techniques

It is certainly necessary to characterize the nonlinear optical properties of 2DLMs irrespective of the design and fabrication method of 2DLM-based active photonic devices. Various experimental techniques have been developed so far to characterize the specific nonlinear optical properties of photonic materials. These include various types of Z-scan techniques [59], four-wave mixing (FWM) [60], pump-probe (PP) [61] two-photon-induced fluorescence [62], HG [63–65], and self-phase modulation (SPM) [66]. These approaches can quantitatively reveal the various characteristics of optical nonlinearities in photonic materials. The choice of the experimental technique to be used depends on the desired nonlinear optical parameters and material property. Here we briefly present a few of the well-established nonlinear optical techniques.
4.1 Z-scan technique

Z-scan is a sensitive and simple experimental technique widely used for measuring third and higher-order (odd) nonlinear optical processes, such as nonlinear absorption [imaginary part of $\chi^{(3)}$], nonlinear refraction [real part of $\chi^{(3)}$], and thermally induced optical nonlinearity of the sample. It is a convenient alternative to the FWM technique, which was used mainly for measuring the third-order optical nonlinearity of optical materials. In addition to nonlinear optical constants, it can also provide characteristic features of nonlinear interactions [67, 68].

The basic principle involved in the Z-scan technique is the spatial distortion of the transmitted beam (both amplitude and phase) at the sample exit. It was invented in 1990 by Sheik-Bahae et al. [69], and in the corresponding scheme, the sample was excited by a beam with various pump fluence and collected the transmitted light from the sample (Figure 6A). The transmitted light from the sample was collected in two different modes, namely open aperture and closed aperture Z-scans. In the open aperture Z-scan, all the light was collected using a double convex lens ($L_3$) and monitored the intensity distortion of transmitted beam (PD3). In the closed aperture Z-scan, a small aperture ($I_2$) is placed in front of the detector and provides the phase distortion of the transmitted beam (PD2). The measurements were carried out while scanning ($z$-direction) the sample in the focal plane of a lens $L_2$. The ratio of measurements of signals (PD2 and PD3) to reference photodiodes (PD1) gives the nonlinear transmittance profile of the sample as a function of $z$. Further, the obtained data were analyzed to quantitatively characterize the nonlinear optical responses of the sample. Numerous modifications have been made to the original Z-scan setup for various applications. The complete details of the various Z-scan techniques can be found in the study of De Araujo et al. [59].

Figure 6D gives the Z-scan profiles of 17-layer chemical vapor deposition (CVD) graphene, with red for closed aperture, green for open aperture, and blue for the quotient of the red curve divided by the green curve [68]. The open aperture Z-scan curve shows a peak at the focus, which corresponds to the saturable absorption, whereas the closed aperture/open aperture profile (blue) exhibits the valley followed by a peak, indicating the presence of self-focusing processes in the sample.
4.2 HG setup

HG is a wave mixing process in which two or more waves simultaneously interact in a nonlinear optical medium and generate new frequencies. By measuring the SHG and THG emission signal, one can estimate the corresponding nonlinear optical susceptibilities of the material [34, 63]. Moreover, polarization-resolved and high spatial resolution 2D mapping HG emission measurements provide the sample characteristics and high-contrast images of 2DLMs [33, 60, 73].

The experimental layout for the HG microscopy is shown in Figure 6B. Initially, a pump beam with specific wavelength was focused onto the sample (selectively and tightly) with the help of a CCD camera and microscope objective lens. Further, the generated HG signal from the sample was separated with the help of a set of wavelength dichroic filters (LF, HF, and DM). A variable ND filter was used to control the pump power and a λ/2 plate was used to control the polarization of the pump beam. The spectral profile and magnitude of the HG signal were collected by the ultraviolet (UV)-visible-near-IR (NIR) spectrometers and a detector (e.g. PMT), respectively. The high spatial resolution 2D mapping of HG image of samples can be obtained by the computer-controlled 2D micromover stage. In case of FWM measurements, samples can be obtained by the computer-controlled 2D micromover stage. In case of FWM measurements, the sample is excited by the two nondegenerate pump beams [60]. Figure 6E (a) shows the SHG (780 nm) and THG (520 nm) emission spectra of 2DLM GaTe with two different thicknesses (11 and 57 nm) under pumping with 1560 nm [70]. Figure 6E (b and c) presents the SHG and THG images of 2DLM MoS$_2$ obtained from the 2D spatial mapping HG setup [71].

4.3 PP technique

Ultrafast PP spectroscopy is used for measuring ultrafast events in a photo-induced process that are much faster than the bandwidth of conventional electronic devices. It is widely used to measure, among others, the ultrafast carrier dynamics (UCD) [74, 75] and charge and energy transfer processes [76, 77]. The main principle of the PP technique amounts to triggering the optical process under investigation via a high-intensity pump pulse that interacts with the sample, whereas the dynamics of the process is interrogated using a weak probe beam that can be delayed with respect to the pump beam. Typically, the energy of the probe is several orders of magnitude smaller than that of the pump beam. By measuring the variation in the probe intensity (transmitted or reflected) as a function of the time delay between the pump and probe, one can get information about the decay dynamics of excited carriers, which were initiated by the pump pulse.

The optical layout of a typical PP experimental setup is shown in Figure 6C. The technique can be performed in various ways depending on the specific nature of the investigation. For example, both pump and probe pulses can have the same or different frequencies, known as degenerate or nondegenerate PP techniques, respectively. Transient absorption (TA) spectroscopy is the special case of nondegenerate PP technique in which the probe is a supercontinuum light source. A typical PP decay curve obtained for a 2DLM [e.g. few-layer (FL) MoS$_2$] is shown in Figure 6F [72].

5 Nonlinear optical response of 2DLMs

2DLMs are presently a fascinating research area of modern nonlinear optics chiefly due to their ultrathin layer form, large light-matter interaction, strong nonlinear optical response, ultrabroadband and tunable absorption, strong interlayer coupling, UCD, flexible and effective interaction with other materials, large mechanical strength, high thermal conductivity, and chemical inertness [15, 18, 72, 78–82]. The nonlinear optical response of 2DLMs has found a plethora of applications in photonics and optoelectronics [5, 16, 60, 73, 83–89].

2D materials, such as graphene and beyond graphene, which include TMCs, BP, h-BN, perovskites, topological insulators (TIs), and their hybrid structures, display unique and diverse nonlinear optical response [5, 89–95]. The optical nonlinearity of each 2DLM has its own particular characteristics and can be used to specific photonic applications [96]. Graphene is a widely studied nonlinear optical 2DLM due to its broadband absorption at optical and terahertz (THz) frequencies, which is the result of its nearly zero band-gap nature. Other widely studied 2DLMs are TMCs. These are semiconductor materials with formulas MX$_2$ (MoS$_2$, MoSe$_2$, WS$_2$, WSe$_2$) and MX (GaSe, GaTe) known as TMDCs and transition metal monochalcogenides, respectively. In these chemical formulas, M represents the transition metal, such as Mo, W, Re, and Ga, and X is the chalcogenide element (e.g. S, Se, and Te). TMCs exhibit the typical layer-dependent tunable band-gap spanning from 1 eV for bulk or ML cases to 3 eV for single-layer (SL) configuration [97–100].

Several theoretical and experimental investigations have predicted and proven that TMCs (e.g. MoS$_2$ and WS$_2$)
J.W. You et al.: Nonlinear optical properties and applications of 2D materials

Nonlinear optical response well beyond their lower energy band-gap (~1 eV) [18, 101, 102]. More specifically, BP is a thickness-dependent direct band-gap semiconductor that could be widely tunable from 0.3 eV (bulk) to 2 eV (SL) [103, 104]. It bridges the gap between zero band-gap graphene and relatively wide band-gap TMCs, making it suitable for broadband optoelectronic applications, particularly in the IR and mid-IR regions [96]. h-BN is another 2DLM with a large band-gap of ~6 eV. It finds nonlinear optics applications in UV to NIR regions [105, 106]. Figure 7 illustrates the atomic crystal structure, electronic band structure, and accessible range for nonlinear optics applications of various 2DLMs.

5.1 Graphene nonlinear optics

Since its discovery, graphene has been the generic 2DLM used as test bed for new ideas in the field of physics of 2D systems. The brief overview of graphene nonlinear optics can give a clear insight into the nonlinear optical properties of other emerging 2DLMs. Thus, in 2007, Butscher et al. have theoretically studied UCD in graphene [108]. The first experimental measurements on UCD of photogenerated carriers in epitaxial graphene have been carried out by Dawlaty et al. [109]. It was found that a fast relaxation dynamics of 70 to 120 fs corresponds to the carrier-carrier intraband scattering. A slower decay dynamic of 0.4 to 1.7 ps was attributed to the carrier-phonon scattering and electron-hole recombination processes. The nonlinear absorption in graphene oxide (GO) at visible wavelengths has been revealed by Liu et al. [110]. Owing to the strong nonlinear optical response, UCD, wideband, and large absorption of graphene, in 2009, atomic thin-layer graphene-based SAs have been successfully demonstrated for the ultrashort pulse generation at NIR region by Bao et al. [111] and Sun et al. [15].

The unique photonic properties of graphene, the many nonlinear optical processes such as nonlinear absorption [67], nonlinear refraction [112], third-order nonlinear optical susceptibility, UCD [113], HG [29, 114], high HG (HHG) [115], FWM [60], SPM [66], and optical bistability [116], have been studied extensively by considering several key aspects. These include (i) studies in a broad frequency domain, i.e. from UV to microwave measurements [6, 117–119]; (ii) graphene of various thicknesses was investigated, from SL, bilayer (BL), and FL graphene to ML graphene [13, 63, 82]; (iii) various preparation methods and sample conditions were used: CVD, epitaxial, exfoliated grown graphene, reduction amount of GO [120], defect-induced graphene, coupling and decoupling layers, and medium dependent (e.g. suspended, solvent, polymers, and substrate) [114, 121–123]; (iv) various forms of graphene have been created (e.g. GO, reduced GO (rGO), graphene sheets, and graphene ribbons) [16, 120, 123]; (v) hybrid structures, such as composites with organic dyes, metal structures, graphene integrated with silicon photonics, and 2D crystal heterostructures [81, 124–128]; (vi) various laser parameter excitation, including continuous-wave (CW), nanosecond, picosecond, and femtosecond regimes, at various repetition rates and pump fluence [110, 117, 129, 130]; and (vii) at various temperatures and in external fields of different strengths [119, 131, 132].

Third-order nonlinear optical interactions in graphene have been extensively studied in experiments. Graphene exhibits intensity-dependent nonlinear optical absorption over a broad spectral range [117, 129, 133]. At low pump fluence, it exhibits saturable absorption behavior and switches to optical limiting response at high pump fluence. Figure 8A shows the pump fluence-dependent switchable nonlinear absorption response of electrochemical GO at 800 nm in a femtosecond regime [67]. A tunable modulation depth of graphene saturable absorption induced by varying the thickness of graphene was observed by Bao et al. [111]. Studies on the coupling and decoupling layers of graphene suggested that the nonlinear optical response is no longer proportional to the layer numbers and the superimposition of ML graphene exhibits efficient nonlinear optical interactions [68].

The controlled and enhanced optical limiting effects of GO by IR reduction, chemical reduction, and hydrothermal dehydration have been reported by several
Moreover, diverse mechanisms for optical limiting response in graphene-based materials under excitation with pulses of different widths have been reported by various authors [16, 110, 133, 137]. In the case of short pulses (picosecond and femtosecond), the dominant contribution of two-photon absorption (TPA) was observed, whereas for broader pulses (nanosecond and broader) reverse saturable absorption (RSA) and...
nonlinear scattering (NLS) have been reported. Moreover, Wang et al. have studied the thermal lens effect in graphene and GO by varying the solvent, temperature, and modulation of the CW laser [130]. Importantly, the spectral width of the excitation laser also strongly influences the nonlinear optical interactions in graphene. The non-degenerate TPA-induced optical limiting response was observed in GO using a white light continuum Z-scan [138]. Finally, nonlinear optical effects in various organic solvents have been reported for graphene-based materials [123, 139].

Graphene-based materials exhibit strong nonlinear refraction in a wide range of wavelengths [140]. An SL CVD-grown graphene has nonlinear absorption (β) and nonlinear index of refraction (n₂) coefficients of the order of 10⁻⁶ cm/W and 10⁻⁹ cm²/W, respectively [68]. A giant nonlinear refractive index coefficient of an order of 10⁻⁵ cm²/W was obtained for the FL CVD-grown graphene (Figure 8B) by Zhang et al. [112]. One of the consequences of intensity-dependent and thermally induced n₂ in photonic materials is SPM [130, 141]. Figure 8C shows the purely coherent SPM in FL graphene dispersion reported by Wu et al. [66]. They observed that graphene exhibits broadband SPM from UV (267 nm) to NIR (800 nm) regions due to a broadband, ultrafast, and large third-order nonlinear response. The value of n₂ of an order of 10⁻⁵ cm²/W has been inferred from SPM in rings under CW laser excitation at 532 nm.

Another important application of intensity-dependent refractive index (optical Kerr effect) is the optical bistability, in which an optical structure exhibits two stable outputs for a specific input due to optical nonlinearity and positive feedback. Optical bistability in graphene-based materials has been studied both theoretically and experimentally in different spectral domains [142, 143]. Figure 8D shows the optical bistability in SL and BL graphene placed inside a Fabry-Perot interferometer [116]. The UCD of graphene-based materials has been investigated by several groups in different surrounding media and sample conditions [113, 119, 144]. The results suggested that the characteristic decay time varies from tens to hundreds of femtoseconds; the decay dynamics is chiefly determined by carrier-carrier scattering processes and carrier-phonon scattering [109, 145]. Figure 8E shows the decay dynamics of graphene in various solvents [113]. Short time (τ₁) corresponds to the intraband carrier-carrier scattering and longer time (τ₂) corresponds to the carrier-phonon scattering.

There are many studies on wave-mixing interactions in graphene-based materials, as in Figure 9D, including SHG [86, 146], THG [29, 60, 147, 148], and FWM [60]. Figure 8F–H shows the wave mixing (HG and FWM) emission spectra and imaging of graphene nanosheets, respectively [60]. HHG was also observed in SL and ML graphene [115, 149]. For example, Yoshikawa et al. observed the odd-order harmonics up to ninth in SL graphene (Figure 8I) in the mid-IR region (0.26 eV) at room temperature [115]. It was found that the HHG signal is enhanced much more under excitation with elliptically polarized beams compared to the case in which linearly polarized light is used. Moreover, using the broad range of optical properties of graphene, the integration of graphene photonics with various optical materials, such as various organic dyes [126], metal nanostructures [150], doped GO [151], semiconductor nanostructures [152, 153], semiconductor waveguides [154], and other 2D crystals [125, 155, 156], can lead to large nonlinear optical response and highly functional active photonic devices. The positive synergetic effects were interpreted for the modified and enhanced nonlinear optical response [8, 66, 77, 117, 125, 130]. In the following sections, we will discuss various nonlinear optical properties of several 2DLMs, including saturable absorption, optical limiting, HG, and excited-state decay dynamics.

5.2 Saturable absorption

Saturable absorption is a nonparametric nonlinear optical process in which the absorption of an optical medium decreases with increasing the pump fluence beyond a threshold value [110]. This phenomenon is widely used in ultrashort pulse generation where the CW laser turns to the train of pulses with high peak power [13, 15]. Most 2DLMs exhibit diverse nonlinear absorption upon varying pump fluence [117], in which the total intensity-dependent absorption coefficient, \( \alpha(I) = \alpha_0 + I(1 + I/I_s) + \beta I \), where the first term represents the saturable absorption process, with \( \alpha_0 \) as the linear absorption coefficient and \( I_s \) as the saturation intensity, and the second term describes single or combined effects of TPA, excited-state absorption (ESA), and NLS. For the pure saturable absorption model, \( \beta = 0 \). An efficient saturable absorption material should meet several requirements, such as large optical nonlinearity, high modulation depth, low nonsaturable loss, wide spectral window, and high optical threshold damage [74].

Figure 9 shows the photoexcitation, carrier relaxation, various nonlinear absorptions, and wave mixing processes in graphene-like 2DLMs [94, 103]. The carriers can undergo interband transitions (Figure 9A) from the valence band to the conduction band by resonant excitation [140, 157]. The photoexcited carriers decay through several intraband and interband relaxations and
electron-hole recombination channels (Figure 9B) [111, 117]. A further increase of the input irradiance leads to an increase of carriers accumulated in the conduction band and inhibits a further excitation of carriers, the so-called Pauli blocking phenomenon. Owing to Pauli blocking, upon a further increase of pump fluence, the transmittance increases, thus resulting in the effect of saturable absorption (Figure 9C).

As graphene has gapless energy structure, it has been found that the nonlinear absorption can be switched from saturable absorption to optical limiting by varying the pump fluence. This can be performed over a broad spectral domain and with different temporal widths of the pump pulses [67, 110, 117, 129, 158, 159]. The characteristics of these effects were revealed by the strength of pump fluence [67, 84].

A systematic study of the nonlinear absorption of graphene films containing various numbers of layers has recently been performed (see Figure 10A) [111]. The results showed that the $I_s$ values varied from 0.71 to 0.61 MW/cm$^2$ for a number of layers varying from 3 ± 1 to 10 ± 1, respectively, and a decrease of modulation depth from 66.5% to 6.2% was observed, which was due to the increase of nonsaturable loss caused by scattering from graphene multilayers. Compared to the SL GO, its FL counterpart shows ultrafast decay dynamics and large saturable absorption response [75]. Epitaxial grown SL graphene on SiC substrate shows $I_s$ of 4 GW/cm$^2$ at 800 nm in the femtosecond regime and has large damage threshold, i.e. larger than 300 GW/cm$^2$ [161]. It was also demonstrated that broadband NIR (1150–2400 nm) saturable absorption in ML CVD-grown graphene with laser damage threshold should be larger than 100 GW/cm$^2$ [133]. The lower value of laser damage threshold in ML 2DLMs is likely explained by the larger linear absorption in ML compared to that in SL crystals [162]. The saturable absorption in graphene at microwave frequencies was demonstrated by Zheng et al. [91]. A modulation depth from 4.58% to 12.77% and an $I_s$ from 0.02 to 0.05 MW/cm$^2$ were measured from 96 to 100 GHz.

Saturable absorption in TMDCs of MoS$_2$ [94, 163], MoSe$_2$ [164], WS$_2$ [99], WSe$_2$ [165], ReS$_2$ [166], SnSe$_2$ [167], TiS$_2$ [168], and mixed TMDCs [169] has been reported and showed large and broadband response. The most widely studied TMDCs are MoS$_2$ and WS$_2$. These materials have direct band-gap of 1.8 and 2 eV, respectively, in SL configuration, and indirect band-gap of 1.29 and 1.34 eV, respectively, in their bulk form [18]. The ultrafast saturable absorption in TMDCs was first revealed by Wang et al. in a liquid-phase exfoliated monolayer to ML of MoS$_2$ nanosheets at 800 nm in a femtosecond regime [94]. It was observed that, under similar experimental conditions, MoS$_2$ exhibits significant saturable absorption higher than in graphene (Figure 10B). Indeed, the excitation pump energy of 1.55 eV (800 nm) is lower than the energy gap of SL MoS$_2$ (~1.9 eV); the obtained saturable absorption is ascribed to the dominant contribution from the presence of ML MoS$_2$ in the dispersion.

It has been found that 2D layers of MoS$_2$ [e.g. MoS$_2$/polyvinyl alcohol (PVA)] exhibit saturable absorption response well beyond (~1.29 eV) the bulk band-gap [101, 170]. It was suggested that the origin of saturable absorption in the NIR region is the narrowing of the MoS$_2$
band-gap due to oxygen incorporation into the MoS₂ nanosheets, a prediction confirmed by X-ray photoelectron spectroscopy measurements. With similar procedures, Chen et al. have fabricated an ML WS₂ SAB by the sulfidation growth method and employed it for passively Q-switched mid-IR (2.865 μm) pulsed laser [18]. Another group observed saturable absorption response in ambient stable FL rhenium diselenide (ReSe₂) nanosheets near 1550 nm due to the saturation of edge states, which arisen from a large edge to the surface area of FL ReSe₂ [166]. It also reported the fluence-dependent switchable response from saturable absorption to optical limiting in SL and FL TMDCs (e.g. MoS₂ and WS₂) at visible and NIR frequencies [99, 163].
Due to its layer-dependent band-gap, similar to TMDCs, the optical nonlinear response of BP can be tuned over a broad range of wavelengths from visible to mid-IR region [76, 96, 171]. Studies of size-dependent optical nonlinearity of few to ML BP nanosheets in N-methyl-pyrrolidone at 800 nm in the femtosecond regime suggested that smaller-sized (13–15 layers) BP nanosheets exhibit excellent saturable absorption with ultrafast recovery time attributed to edge states and quantum confinement effects [104]. Hanlon et al. have observed a larger saturable absorption response in FL solvent stabilized BP nanosheets at 515 and 1030 nm in the femtosecond regime, which is larger than that of graphene at the same level of linear transmittance [171]. The NIR and mid-IR saturable absorption of liquid-exfoliated BP nanoplates (BPNPs) in the picosecond and femtosecond regimes was demonstrated by Lu et al. [103]. The excitation energies were well below the energy gap of SL BP (2 eV). Therefore, the observed saturable absorption was attributed to the dominant contribution from the presence of ML BPNPs (band-gap 0.3 eV) in the dispersion.

Defects induce various optical nonlinearities (e.g. self-defocusing observed in FL oxidized BP aqueous solution), in addition to a significant saturable absorption behavior, which is absent in pure BP [172]. The appearance of negative nonlinear refraction is attributed to the thermal lens effect, which is generated by oxygen-induced defects in the BP. Interestingly, anisotropic saturable absorption response (Figure 10C) in BP thin films has been observed by Diao et al. near telecom wavelength along the armchair and zigzag directions [95].

2D TIs, such as Bi\(_2\)Se\(_3\), also possess excellent broadband (optical to microwave) nonlinear optical response [90, 173, 174]. Thus, ML Bi\(_2\)Se\(_3\) nanoplates show large saturable absorption (\(I_s\) of 10.12 GW/cm\(^2\) and high modulation depth of 61.2%) and self-focusing behavior with high nonlinear refractive index of order of 10\(^{-10}\) cm\(^2\)/W at 800 nm in the femtosecond regime [90]. Compared to the ML, FL Bi\(_2\)Se\(_3\) exhibits better saturable absorption response, namely about two orders of magnitude smaller saturation intensity than the ML TI at similar excitation [90, 175]. It was suggested that the superimposition of FL Bi\(_2\)Se\(_3\) TI is a promising candidate for low-threshold, high-efficiency ultrafast laser applications.

Nanohybrid 2D materials can provide certain advantages compared to their pure-phase counterparts, as they possess a wider degree of freedom to modify their photonic properties and tailor them for specific photonic application [92, 117]. Thus, graphene has shown enhancement of saturable absorption response when decorated with metal nanoparticles (e.g. silver nanoparticles and nanowires) [117, 129, 176]. The enhanced nonlinear optical response is attributed to the metal-enhanced local field-induced polarizability and modified electron structures of graphene/metal nanohybrid [117]. On the contrary, using the advantages of broadband and ultrafast dynamics responses in graphene, together with strong light-matter interaction, 2DLMs such as MoS\(_2\), WS\(_2\), and BP can be employed to achieve ultrafast and broadband enhancement of optical nonlinearity of 2DLM nanohybrids [92, 125, 177, 178]. In this context, Jiang et al. have observed the enhancement of saturable absorption response in MoS\(_2\)/graphene nanohybrid (Figure 10D) over a wide range of frequencies from visible to NIR regions in the femtosecond regime [125]. The PP measurements and Z-scan results suggest that the enhanced nonlinear optical response is due to the multiphoton absorption (MPA) channels and increase of carrier density in the nanohybrid.

### 5.3 Optical power limiting

Optical limiting devices are important for protecting sensitive optical detectors and human eyes from hazardous light sources [179]. Optical limiting is a process in which the intensity of light transmitted through an optical material decreases with increasing excitation light fluence beyond a certain threshold value [16]. An efficient optical limiter should possess wide optical window, low optical limiting, and high optical damage threshold [5]. In a photonic material, optical limiting can be induced via several mechanisms [5], including nonlinear absorption [155, 180], nonlinear refraction, and NLS [159].

Nonlinear absorption is a fully electronic process that includes TPA/MPA and ESA, which is also known as RSA [110]. ESA occurs when the excited state has stronger absorption compared to the ground state. Moreover, under resonant and near-resonant excitation, ESA is the dominant process, whereas under nonresonant conditions TPA is the main process. The ESA coefficient increases with pump fluence, whereas the TPA coefficient remains constant [110]. The NLS arises via the onset of thermal lens effects [181], microbubbles, and microplasma scattering inside a sample due to absorption/laser-induced local heat [6]. Thermal lens effects can be observed with high repetition rate pulses or CW lasers [181]. In general, the optical limiting is due to TPA and is observed for shorter pulses (picoseconds and femtoseconds) [117, 129], whereas ESA and NLS are observed in the case of nanosecond pulse excitation [6, 110].

Graphene-based materials show very strong optical limiting response in a broad frequency domain [6, 123, 158].
via TPA [117, 129], MPA [182], ESA [110], and NLS [6]. The TPA (or MPA) process in graphene-like 2DLMs is shown in Figure 9C. At sufficiently high pump fluence, the carriers are excited by TPA, thereby overcoming the Pauli blocking-induced optical limiting [82]. It is worth mentioning that Yang et al. have observed giant TPA coefficients ($\beta_{\text{TPA}}$) of $10 \pm 2 \times 10^4$ and $20 \pm 4 \times 10^5$ cm/GW in Bernal-stacked BL graphene using femtosecond pulses of 780 and 1100 nm, respectively [82]. These studies revealed that at low pump fluence the large saturable absorption masks the weaker TPA, which can be observed only at high pump fluence. Graphene-based materials show different optical limiting in various solvents [6, 123]. Large and efficient optical limiting is observed for low surface tension solvents, which suggests that low surface tension results in large bubble size; hence, effective scattering leads to increased optical limiting [6].

Low-loss electrochemical GO has been demonstrated to possess large nonlinear absorption (7 cm/GW) with high optical limiting modulation depth of 35% at 70% linear transmittance and larger chemical and optical power stability (400 ml/cm$^2$) than other graphene-based materials [67]. Moreover, tunable optical limiting in GO has been achieved by varying the degree of reduction of GO [134–136]. For example, Jiang et al. have studied systematically the nonlinear optical properties of GO in the femtosecond regime by varying the amount of reduction of GO using a laser-assisted and chemical reduction method [135]. As shown in Figure 10E, partially rGO exhibits stronger optical limiting with increasing reduction amount compared to the pristine GO. Increasing the reduction of GO leads to the reduction of $sp^3$ matrix into large and new $sp^2$ domains. The large optical limiting of rGO suggests that the TPA cross-section of $sp^2$ domains is larger than that of the $sp^3$ matrix [134]. In contrast, highly rGO exhibits stronger saturable absorption (at 400 and 800 nm), which is attributed to the large linear absorption of highly rGO [135].

Similar to graphene, TMDCs such as MoS$_2$ [162, 183, 184], MoSe$_2$ [185], WS$_2$ [186], WSe$_2$ [169], and TiS$_2$ [5] exhibit interesting optical limiting response. In addition to the fluence, layer number and excitation wavelength play an important role in the optical limiting effects of TMDCs due to the layer-dependent band-gap [162, 183–185]. For example, Li et al. have measured values of $\beta_{\text{TPA}}$ as large as $7.62 \times 10^4$ cm/GW and large optical damage threshold of 97 GW/cm$^2$ in an SL MoS$_2$ triangular island film at 1030 nm and in the femtosecond regime [162]. In contrast, a saturable absorption behavior was observed for ML MoS$_2$. For similar excitation conditions, a large value of $\beta_{\text{TPA}}$ (1.34 x 10$^5$ cm/GW) and a saturation of the TPA process were observed in one to three layers of WS$_2$ thin films [186]. At 532 nm and in the nanosecond regime, saturable absorption followed by reverse saturable absorption was observed with increasing pump fluence in monolayer to FL of MoS$_2$ array films [184]. Under NIR excitation, TPA occurs rather than saturable absorption due to the larger energy band-gap of monolayer to FL (<4 nm thick) of TMDcs compared to the excitation energy [183, 185]. Very recently, Varma et al. observed a large optical limiting response from one to two layers of TiS$_2$ nanosheets with pumping at 532 nm in the nanosecond regime [5]. The results showed that TiS$_2$ exhibits a large modulation depth of optical limiting response, namely about 77% for linear transmittance of 73% and nonlinear absorption coefficient about 2.5 times larger than that of graphene and C$_{60}$ (Figure 10F). The large optical limiting in TiS$_2$ is attributed to the TPA and induced ESA mechanisms (Figure 10F, inset), which was confirmed by photoacoustic Z-scan measurements.

Other 2DLMs, h-BN (Figure 10G) and BP, also show a strong optical limiting response [93, 187]. Thus, Kumbhakar et al. have measured large TPA (74.84 cm/GW) response of h-BN aqueous solution using nanosecond pulses at 1064 nm wavelength. A saturation (6 MW/cm$^2$) of TPA was observed at large pump fluence, namely 20 times smaller than in GO at 532 nm [106]. The strong TPA is ascribed to the increase in hyperpolarizability upon laser excitation due to the polar behavior between the B and N atoms of h-BN. A large $\beta_{\text{TPA}}$ value of $5 \times 10^4$ cm/GW was observed for CVD-grown h-BN film under femtosecond excitation from 400 to 800 nm and is one order magnitude larger than in the case of SL graphene [105]. Similar to graphene, BP also exhibits saturable absorption followed by optical limiting response upon varying pump fluence [187]. Studies of few to ML BP from UV to NIR showed that the TPA induced the optical limiting response in the picosecond and femtosecond regimes, whereas NLS was observed in the nanosecond regime [187–189]. Moreover, Zheng et al. measured a $\beta_{\text{TPA}}$ of about 45 cm/GW for the BPNPs/PVA composite at 800 nm in the femtosecond regime [187].

Compared to isolated materials, hybrid structures of 2DLMs show large optical limiting response, which can be induced by several mechanisms, including photo-induced charge/energy transfer (PCT/PET) [126, 152, 159, 160, 190–192], strong interlayer interactions [81], metal-enhanced local fields [117], and a mixture of different nonlinear optical responses [128]. Thus, Xu et al. have observed an optical limiting enhancement of graphene by covalent functionalization with porphyrin molecules; this nonlinear effect is larger than in C$_{60}$ [128].

Most heterostructures, including graphene/ZnO, C$_{60}$, dye, oligothiophene, porphyrins, metal nanoparticles [13,
14, 77, 117, 128, 129, 150, 190, 193, 194], rGO/ZnO, ZnPC, PbS [152, 191, 195], MoS2/graphene [156], MoS2/CNT [192], and h-BN/GO [155], are characterized by a large optical limiting response as determined by positive synergy that arises from PET and/or PCT processes. The PCT results in large carrier dynamics, which leads to large RSA or TPA [155, 180], whereas PET improves the TPA or ESA and/or NLS [77, 194]. For example, Figure 10H shows the NLS-induced enhanced optical limiting in donor (positively charged porphyrin)-acceptor (negatively charge GO) ionic complex via PET [77]. In contrast to the usual saturable absorption behavior in MoS2, a semiconductor polymer-MoS2 nanohybrid exhibits a remarkable optical limiting response, which is attributed to the large, efficient PCT between semiconductor polymer and MoS2 nanosheets [180]. Recently, giant TPA (580 cm/GW) in FL MoSe2/GO heterostructures (Figure 10I) was observed by Sharma et al. at 532 nm in the nanosecond regime [81]. The TPA coefficient of heterostructures was found to be about 30 times larger than in GO, which is due to the modification of polarizability of the heterostructure due to strong interlayer coupling and PCT.

5.4 HG

Monolayer to ML 2DLMs are promising photonic materials for ultracompact frequency converters due to their large optical nonlinearity. In particular, graphene [8, 10, 60, 86, 146, 148], h-BN [196], TMCs (including both monochalcogenides and dichalcogenides, such as MoS2, MoSe2, WSe2, GaSe, GaTe, and CdS) [17, 34, 71, 87–199], BP [64, 200–202], perovskites [203–205], and TIs [206, 207] show efficient SHG and THG at atomically thin limit (Figure 11A). These studies showed that 2DLMs have large nonlinear optical susceptibilities, larger than those of commonly used nonlinear crystals (e.g. β-BBO, 2.2 pm/V at 1064 nm; KDP, 0.43 pm/V at 1064 nm; LiNbO3, 27.2 pm/V at 1064 nm).

![Figure 11](https://example.com/figure11.png)

**Figure 11:** HG and imaging of 2DLMs.

(A) Large spectral range (1120–1330 nm) SHG and THG emission from spiral WS2 nanosheets [208]. (B) Layer-dependent SHG and THG emission from MoS2 [71]. (C) THG emission from various graphene layers as a function of pump power [63]. (D) Linear and SHG emission spectra of monolayer (blue) and trilayer (green) MoS2, showing the resonance enhancement at exciton band. (A–C) Excitation bands of MoS2 [33]. (A–D) Reprinted with permission from Refs. [33, 63, 71, 208]. Copyright 2017 Springer Nature. Copyright 2013 American Physical Society. Copyright 2013 and 2017 American Chemical Society. (E) SHG (MoS2, WSe2, and GaSe) [33, 209, 210] and THG (graphene, BP, and MoS2) [29, 202, 211] mapping images of 2D crystals. Reprinted with permission from Refs. [29, 33, 202, 209–211]. Copyright 2013 American Physical Society. Copyright 2014 John Wiley and Sons. Copyright 2013 American Physical Society. Copyright 2017 American Chemical Society. Copyright 2017 AIP Advances licensed under a Creative Commons Attribution (CC BY) license. Copyright 2016 IOP Science under a Creative Commons Attribution (CC BY) license. (F) HHG spectrum of MoS2. Reprinted with permission from Ref. [212]. Copyright 2016 Springer Nature.
at 1064 nm; and quartz, 0.4 pm/V at 1052.8 nm) [32, 213, 214]. In the following sections, we present various recent advances in understanding SHG and THG in 2DLMs.

5.4.1 SHG

SHG from 2DLMs has received much scientific interest due to its potential applications in active photonic nanodevices. Indeed, SHG is forbidden in free-standing or pristine graphene, BP, even-layer TMDCs due to its centrosymmetric property [71]. However, SHG from these materials has been investigated by several researchers both theoretically and experimentally by breaking the inversion symmetry via an external excitation, such as electrical (field/current/charge) [215–217], dopants [218], and structural variations [114, 208, 219], or by placing them onto a substrate [220]. Thus, Dean et al. have demonstrated SHG from exfoliated SL and ML graphene films on oxidized Si(001) substrate at 800 nm in the femtosecond regime [86, 146]. The SHG emission effect was confirmed by the quadratic dependence of SHG intensity ($I_{2\omega}$) upon the pump intensity ($I_{\omega}$), namely $I_{2\omega} \propto I_{\omega}^2$.

The SHG from suspended SL and BL graphene has also been observed and attributed to long-range curvature fluctuations of the suspended graphene. This assumption was validated by the analysis of 2D spatially resolved Raman and SHG mapping [114]. In particular, SHG studies were extensively done on monolayer TMCs due to the inherent noncentrosymmetric nature of these materials and their large optical nonlinearity. Thus, it has been reported that TMCs have large second-order nonlinear susceptibility, more than an order of 1 pm/V [73]. This is several orders of magnitude larger than that of most dielectric materials [17, 34, 73, 100].

The most in-depth studied TMCs are MoS$_2$ [33, 35, 73, 198, 221], MoSe$_2$ [87, 199, 218], WS$_2$ [34, 219], and WSe$_2$ [17, 36, 213, 215, 222, 223]. Thus, Kumar et al. have observed very strong SHG emission from SL MoS$_2$ nanosheets prepared by mechanical and chemical exfoliation methods, with excitation at 810 nm [73]. The estimated effective bulk-like second-order susceptibility, defined as the ratio between the surface second-order susceptibility and the bulk-like second-order susceptibility, defined as the ratio of most dielectric materials [17, 34, 73, 100].

The dependence of SHG on the number of layers has been observed for the TMDCs of spiral WS$_2$ nanosheets due to the broken symmetry from the twisted structures [208]. Stronger SHG has been reported for resonance excitation and excitation at edges of 2DLMs [17, 199, 210, 221]. Figure 11D shows the resonant feature of SHG emission from MoS$_2$, nanosheets at 1560 nm excitation [71]. The results show that odd-layer sheets (1, 3, 5, …, layers) have strong SHG signal, whereas even layers (2, 4, 6, …, layers) of MoS$_2$ exhibit vanishing or very weak SHG emission. It has been found that the SHG emission from an odd number of layers (FL limit) is almost the same for nonresonant SHG emitted photons, whereas in the resonant case (the energy of SHG photons is larger than the band-gap) a significant reduction in SHG (e.g. five layers of MoS$_2$ nanosheets show nearly eight times less SHG than the SL) is observed with increasing the odd number of layers [35].

This dependence is attributed to the reabsorption of SHG photons and interlayer coupling in MoS$_2$, nanosheets [33, 35, 213]. Similar results are observed for h-BN nanosheets under 810 nm excitation [35]. In contrast, the quadratic dependence of SHG on the number of layers has been observed for the TMDCs of spiral WS$_2$ nanosheets due to the broken symmetry from the twisted structures [208].

Interesting nonlinear optical materials belonging to the TMDCs family, such as GaSe [100, 197, 209], GaTe [70], and CdS [225], also exhibit SHG when the inversion symmetry is broken using external sources, including charge, plasmonic hot carriers [215, 217, 223], and specific structures, such as heterostructures, spirals, and pyramids [208, 219, 224].
in the nanosheet [71, 197, 226], unlike the case of TMDCs. In particular, Jie et al. have observed a cubic dependence of SHG with the number of layers, for up to five layers, whereas, when the number of layers was larger than five, a quadratic dependence was observed [209]. Moreover, an ML (nine layers) GaSe-photonic crystal cavity nanostructure showed 75 times larger SHG than that of a monolayer [226]. Similar values of $\chi^{(3)}$, namely 1.15 pm/V at 1560 nm, was observed for FL GaTe nanosheets [70]. The noncentrosymmetric nature of 2D TMCs makes them promising candidates for highly efficient and ultracompact light conversion photonic nanodevices.

5.4.2 THG

All 2DLMs exhibit FWM/THG to various degrees irrespective of their symmetry properties. The use of THG for material characterization is a particularly efficient investigative tool for centrosymmetric materials, including graphene [63], BP [64], and even/ML TMDCs [65]. THG and FWM in graphene-based materials have been investigated both theoretically and experimentally [29, 60, 147, 149]. Thus, it has been revealed that the effective bulk-like third-order nonlinear susceptibility, $\chi^{(3)}$, strongly depends on the excitation wavelength, experimental conditions, and sample preparation method.

Strong THG from mechanically exfoliated SL and FL graphitic films on oxidized Si substrate was demonstrated by Kumar et al. with excitation of 1720 nm wavelength upon normal incidence. The measured values of $\chi^{(3)}$ were of the order of $10^{-16}$ m$^2$/V$^2$ [148]. One order of magnitude larger value of $\chi^{(3)}$ ($10^{-15}$ m$^2$/V$^2$) was measured by Saynatjoki et al. for SL graphene with reference to the substrate (Si/SiO$_2$) at 1550 nm [63]. Similar values were reported by Hendry et al. over the wavelength range of 760 to 840 nm using the FWM technique [60]. Enhanced and broadband electrically tunable THG from graphene samples have also been experimentally demonstrated [131].

There are several theoretical predictions of large THG with values of $\chi^{(3)}$ of the order of $10^{-12}$ m$^2$/V$^2$; the large nonlinearity is attributed to substrate effects, heterostructures, dopants, and strong plasmonic near-field enhancement [147, 227, 228]. Layer-dependent tunable THG from BP nanosheets was reported by several groups [64, 200–202] near telecom wavelength; the measured value of $\chi^{(3)}$ is about $10^{-19}$ m$^2$/V$^2$. Moreover, strong excitonic resonantly enhanced THG from FL BP was observed by Rodrigues et al.; the measured THG emission is three orders of magnitude larger than that of SL graphene under similar experimental conditions [200].

Studies of THG in MoS$_2$ [65, 71, 211, 229, 230], MoSe$_2$, WSe$_2$, WS$_2$ [199, 208], ReS$_2$ [231], GaSe [100, 197], and GaTe [70] have been performed by several groups. In contrast to the experimental conditions, THG exhibits a gradual increase with the number of layers (Figure 11B). The measured value of $\chi^{(3)}$ of monolayer to FL of MoS$_2$ nanosheets in the wavelength range of 1560 to 1950 nm was of the order of $10^{-14}$ m$^2$/V$^2$ [65, 71, 211, 229]. Due to observed discrepancies in the value of $\chi^{(3)}$ of graphene reported by various authors, Saynatjoki et al. and Woodward et al. have measured the THG from SL MoS$_2$ and compared it to that of SL graphene under similar experimental conditions. It was found that the THG efficiency in MoS$_2$ is larger by a factor of three to four than in graphene [71, 211]. Moreover, THG in monolayer to FL of ReS$_2$, GaTe, and GaSe was reported to be larger by one to two orders of magnitude than in SL MoS$_2$; the measured values of $\chi^{(3)}$ are about $10^{-18}$ to $10^{-16}$ m$^2$/V$^2$ [70, 197, 231].

Other class of 2DLMs (perovskites) have also been found to possess a large THG response [203–205]. Thus, Abdelwahab et al. have measured values of $\chi^{(3)}$ of $1.12 \times 10^{-17}$ m$^2$/V$^2$ for 2D lead halide Ruddlesden-Popper perovskites under resonant exciton excitation (at 1675 nm) [203]. It was found that the maximum THG conversion efficiency (0.006%) is more than five orders of magnitude larger than in the case of BP and MoS$_2$. Moreover, it has been shown that there is an optimum number of layers for which one achieves maximum THG (Figure 11C). Also, a quadratic dependence of THG on the number of layers, for less than 15 layers, has been reported in graphene [29, 148], TMCs (e.g. MoS$_2$, WSe$_2$, and GaSe) [71, 197, 232], and BP [64]. A decrease in THG was observed for a large number of layers and attributed to the phase mismatch or depletion of fundamental and/or THG signal by the sample absorption and reflection [29, 60, 64, 71, 231].

A recent progress in the understanding of HG in 2DLMs suggests that these materials provide a particularly fertile platform for the study of HHG [71, 115, 149, 212]. Thus, Figure 11F shows the nonperturbative HHG from SL MoS$_2$ with even and odd orders, up to the 13th harmonics, upon excitation with femtosecond mid-IR pulses (0.3 eV) [212]. It was observed that odd harmonics exhibits a significant enhancement in efficiency per layer compared to the bulk, which is attributed to the electronic correlation effect.

5.4.3 HG enhancement and tunability

Although 2DLMs exhibit much larger HG intensity compared to conventional nonlinear optical crystals in the ultrathin limit [32, 34, 73, 100, 213, 214], the thickness of
2DLMs of just a few atomic layers or less leads to a poor field confinement and weak nonlinear optical interaction with the excitation light. The key research challenge is to enhance the HG efficiency of 2DLMs. Several recent theoretical and experimental investigations demonstrated remarkable enhancement in HG from 2DLMs using resonant pumping [17, 33, 203], applying an external electric field/current/charge [7–10, 36, 215, 216], metasurfaces [233], plasmons [222], integration with silicon waveguides, micro and photonic crystal cavities [87, 226, 234], strain induced [235], and doping with other materials [147, 218]. For example, a giant enhancement of SHG, by about three orders of magnitude, was reported for SL MoSe$_2$ on SiO$_2$/Si under resonant exciton excitation at 4 K temperature by spectral-dependent (~1.7–2.4 eV) SHG measurements [17].

Selective dynamically controlled and enhanced (7000 times) SHG from SL WSe$_2$ placed on gold trenches thinner than 20 nm was reported by Wang et al. [222] (Figure 12A). The observed enhancement is attributed to the enhanced localized optical near-field inside the gold trenches.

Another approach for enhancing the SHG signal is by increasing the local field or length of nonlinear optical interaction between ultrathin 2DLMs and excitation light source by designing novel nanostructures. Thus, by combining the effects of whispering-gallery modes and plasmons, an enhancement of 800-fold edge SHG was achieved for pyramid-shaped ML WS$_2$/plasmonic nanostructure (Figure 12B) [219]. Recently, Gan et al. have achieved SHG from monolayer to FL GaSe nanosheets under microwatt (<10 μW) CW power pumping using a photonic crystal cavity resonant at 1551 nm [226].

The centrosymmetric nature of 2DLMs has been used to dynamically control SHG, which can be used to achieve external source switchable/controlled active photonic devices [223]. It has been suggested that $\chi^{(3)}$ could be tuned from zero to several orders of pm/V at ultrahigh speed [223, 225]. Electrically (charge, current, and field) induced tunable SHG from centrosymmetric 2DLMs (e.g. graphene, rGO, CdS, MoS$_2$, and WSe$_2$) has been reported by several groups [8, 10, 215, 217, 225]. Thus, large tunable (the value of $\chi^{(3)}$ was tuned from 0 to 151 pm/V) and high modulation strength (200% V$^{-1}$) of SHG from CdS nanobelts were obtained by applying a symmetry-breaking external electric field (Figure 12C) [225].

All-optical ultrafast controllable SHG from BL WSe$_2$/metallic nanostructures has been demonstrated by Wen et al. via plasmon-induced hot carriers [223]. This effect was validated by pump (CW, 532 nm)-probe (SHG signal) measurements on BL WSe$_2$ covered with gold nanodisk array of diameter of 70 nm (resonant) and 150 nm (off-resonant) with plasmonic bands at 538 and 772 nm, respectively (Figure 12D). The carrier transfer from the plasmonic nanostructure to BL WSe$_2$ and ultrafast modulation time (<2 ps) were confirmed and evaluated by TA spectroscopy.

Figure 12: Enhancement and tunable SHG from 2DLMs. (A) Schematic of WSe$_2$-gold trench nanostructure (top) and SHG peak intensity of WSe$_2$ on gold trenches and on sapphire (bottom) [222]. The “7000” and “150” denote the normalized SHG enhancement factor under resonant and nonresonant excitation, respectively, where WSe$_2$ on sapphire was used as a reference. (B) (a) Schematic of edge-enhanced SHG response in pyramid-shaped ML WS$_2$. (b) Spatially resolved spectra collected from the edge and center of the P-ML WS$_2$. (Inset) Quadratic dependence of SHG intensity with pump power. (c) SHG signals from P-ML (black) and monolayer (red) WS$_2$ excited with a pulsed laser (810 nm) at a fixed pump power [219]. Reprinted with permission from Refs. [219, 222]. Copyright 2018 American Chemical Society. (C) Schematic of a CdS nanobelt device with the source (S) and drain (D) electrodes (left) [225]. The long axis of the nanobelt (x-axis) is determined to be CdS c-axis, which is perpendicular to the z-axis (or CdS c-axis). The fundamental wave at the frequency of $\omega_0$ is normally incident upon the belt and excites the second-harmonic wave at 2$\omega_0$, which is back scattered. (Right) Voltage dependence of $d_{33}$-SHG signal. (D) Schematic of PP measurement with a CW laser to manipulate the SHG signal (left). (Right) Tunable SHG intensity of BL WSe$_2$-metal nanostructure (diameter 70 nm) with increasing pump power (CW, 532 nm) [223]. Reprinted with permission from Refs. [223, 225]. Copyright 2018 American Chemical Society and Springer Nature.
Dynamically controlled enhanced SHG and THG were also observed in noncentrosymmetric 2DLMs [222].

### 5.5 Excited-state dynamics

The ultrafast spectroscopy (Figure 6C) of 2DLMs provides valuable information about the photoexcited carrier dynamics, which is particularly important to the understanding and characterization of the performance of 2DLM-based photonic devices [74, 94, 161]. For example, graphene exhibits photoexcited UCD with subpicosecond carrier thermalization and few picosecond carrier cooling times [109]. UCD with broadband and large nonlinear optical response in graphene can be used as fast SABs for ultrashort pulse generation over a wide range of wavelengths [15, 91, 118]. It has been found that the UCD in graphene is insensitive to the number of graphene layers (up to multilayers) due to loosely bonded layers of graphene-like planes [111, 236]. However, the carrier dynamics strongly varies with the physical properties of 2DLMs [75, 109], excitation wavelength [237, 238], surrounding medium [113, 144], dopants [239], and sample conditions (e.g. temperature, magnetic field, and electric field) [119, 260, 241]. Typically, GO exhibits a slower dynamics than graphene and rGO; hence, graphene is a superior ultrafast SAB compared to GO [75, 145, 242, 243].

In some cases, photoluminescence (PL) has also been observed from GO with much longer decay time, namely more than 400 ps to several nanoseconds, which arises from smaller \( sp^2 \) bonded clusters or oxygen-related sites within the GO sheet [244, 245]. It has also been found that graphene on substrates is characterized by fast carrier-phonon scattering relaxation time (e.g. 1.4 ps for six-layer graphene on SiC and 2.5 ps for SL graphene on SiO\(_2\)/Si), whereas in the liquid-phase graphene samples the relaxation time is 3.5 to 4.9 ps. The latter relaxation time is larger due to the absence of carrier scattering at the graphene/substrate interface [109, 113, 236]. Very slow cooling photoexcited carrier dynamics (several hundred picoseconds) in graphene was observed by Strait et al. at low substrate temperature of less than 50 K [246].

Many ultrafast studies have been performed on other TMCs, such as MoS\(_2\), WS\(_2\), WSe\(_2\), and CdSe [61, 72, 94, 247, 248], BP [74, 269, 250], and h-BN [121]. A carrier lifetime of 100 to 400 ps was observed for FL MoS\(_2\) nanosheets at room temperature over the \( A \) exciton transition band from 640 to 700 nm with pump injection at 390 and 780 nm [61, 251]. The carrier thermalization and cooling dynamics of FL MoS\(_2\) nanosheets over the \( A \) and \( B \) exciton bands at 77 K were investigated by ultrafast (10 fs) broadband (560–700 nm) pump light [72]. The carrier thermalization of less than 20 fs was due to carrier-carrier and carrier-phonon scattering, whereas carrier cooling takes place at 0.6 ps via phonon emission (Figure 6F). Temperature-dependent carrier dynamics in SL MoS\(_2\) nanosheets shows fast carrier recombination at low temperature (4.5–150 K) and exhibits biexponential decay at high temperature (>150 K) with a long-lived component (50–100 ps) attributed to the exciton-phonon scattering [252].

An ultrafast intraband carrier dynamics of 24 fs was observed in FL BP nanosheets using the degenerate PP technique at 1550 nm [74]. It was found that the UCD response is faster than in other 2DLM (e.g. graphene and MoS\(_2\)), which was attributed to the defect-assisted carrier cooling. Broadband mid-IR (1700–4600 nm) UCD in thin BP nanosheets upon excitation with 792 nm pump shows a fast component of order of 1 ps, which is attributed to the carrier-phonon scattering, whereas a slow decay time of 10 to 100 ps was due to carrier recombination and lattice heating [250]. Both decay times were found to increase with increasing probe energy as it approaches the bandgap. Much larger carrier lifetime (1.3 ns) was observed for ML (30–80 layers) BP at high pump fluence and attributed to the increase in lattice temperature [249]. The carrier dynamics in 2DLMs such as perovskite [253, 254] and TIs [255, 256] was also investigated by various groups.

Ultrafast spectroscopy measurements on 2DLM hybrids revealed various processes, such as PCT [76, 257], PET [77, 258], and interlayer coupling [81] involved in the photonic properties of nanohybrid 2DLMs [126, 128, 152, 155, 180, 190, 193–195]. Thus, Mamidala et al. have reported enhanced optical limiting response in the GO-porphyrin ionic complex via PET from porphyrin to GO [77]. The process of PET was confirmed using time-resolved PL, and TA measurements revealed shortening in decay dynamics and no new bands in the TA spectrum, respectively. Giant enhanced optical limiting response and quenching of excitation bands in MoSe\(_2\)/GO heterostructures were revealed by ultrafast TA spectroscopy and it was found to be due to PCT and trapping at mid states [81]. Ultrafast PCT has been reported by Hong et al. in MoS\(_2\)/WS\(_2\) (hole transfer within 50 fs) and Peng et al. (electron transfer within 470 fs) in MoS\(_2\)/WSe\(_2\) heterostructures [76, 257]. From the measurements of PL and TA spectroscopy, it was inferred that 99% of the photogenerated charge carriers moved from WSe\(_2\) to the MoS\(_2\) nanosheets.

### 6 Nonlinear optical applications of 2D materials

There are numerous applications in the photonics and optoelectronics of nonlinear optical response of 2DLMs,
namely to SAbs [83], passive optical power limiters [5, 16], all-optical switching [84], frequency conversion and HHG [115, 149, 212], THz wave generation [85], parametric amplification [87], bistable optical devices [116, 143], supercontinuum generation [259, 260], electro-optic modulators [36, 217], ultrathin nanoscale light sources [8, 87, 215, 217, 225], and optical imaging [60, 73, 88].

### 6.1 Ultrashort pulse generation

SAbs are widely used optical modulators for the generation of ultrashort laser pulses [261, 262]. The large optical nonlinearity, broadband absorption, ultrafast carrier recovery time, and simple fabrication methods of 2DLMs make these materials a viable alternative to SAb modulators (e.g. dyes and semiconductor SAb). They have been employed in the simple and compact fabrication of ultrafast lasers in a broad spectral range from visible to mid-IR [83, 91, 263–266]. A series of 2DLM SAb modulators have been developed for ultrashort pulse generation in different configurations (Figure 13A). Thus, SAb shown in Figure 13A (a, b, and d) are easy to fabricate and use [15, 111, 268], whereas SAb illustrated in Figure 13A (c, e, and f) are characterized by strong light-matter interaction and high optical damage threshold [267, 269].

Figure 13: Ultrafast laser spectroscopy of 2DLMs.
(A) Schematic illustration of 2DLMs-based SAb. (a) Graphene film coated on fiber pigtail [111], (b) sandwiching graphene-PVA between two fiber connectors with a fiber adapter [15], (c) graphene coated on tapered fiber [267], (d) BP coated on gold mirror [268], (e) 2DLMs coated on D-shaped fiber, and (f) embedded on microfiber [269]. Reprinted with permission from Refs. [15, 111, 267–269]. Copyright 2009 John Wiley and Sons. Copyright 2010 American Chemical Society. Copyright 2016 Springer Nature. Copyright 2014 American Chemical Society. (B) Typical experimental setup for ultrafast fiber (a) and bulk lasers (b) [270, 271]. WDM, wavelength division multiplexer; LD, laser diode; F1 and F2, convex lenses; M1 to M3, concave mirrors with the same radius of curvature of 10 cm; OC, output coupler; graphene SAM, graphene SAb mirror. Reprinted with permission from Refs. [270, 271]. Copyright 2012 and 2015 Optical Society of America. (C) Optical output spectra and pulse width of 2DLM-based pulsed lasers. (a) Broadband tunable output spectra near telecom wavelength from graphene Q-switched fiber laser [263]. (b) Visible output spectrum of WS, Q-switched fiber laser (inset, with a dBm logarithmic scale) [83]. Reprinted with permission from Refs. [83, 263]. Copyright 2011 AIP Publishing. Copyright 2016 Royal Society of Chemistry. Ultrashort pulse generation with a graphene mode-locked (c) [261] and WS, Q-switched laser (d) [262]. Reprinted with permission from Refs. [261, 262]. Copyright 2017 Optical Society of America.
Typical configurations of widely fabricated ultrafast solid-state bulk and fiber lasers based on 2DLM SAbs are shown in Figure 13B [270, 271]. Graphene-based SAbs have been used for ultrashort to microsecond pulse generation from NIR to mid-IR lasers [15, 91, 118, 243, 261]. For example, Figure 13C (a) shows the tunable output spectra from Q-switched graphene SAb fiber laser near telecom wavelength [263]. It has also been reported that graphene can serve as SAb for compact pulsed lasers in the THz and microwave regions [91, 265]. Using graphene SAb, ultrashort pulses with width of sub-20 fs (Figure 13C, c) mode-locked solid-state laser have been demonstrated by Canbaz et al. [261].

The strong light-matter interaction and layer-dependent tunable optical nonlinearity of TMDCs and BP have also been examined intensively in connection to ultrashort pulse generation [83, 96]. The band gap at visible frequencies of TMDCs is suitable for efficient SAbs at optical frequencies for high-power compact visible pulsed lasers [83]. Figure 13C (b) shows the output spectrum of an all-optical Q-switched fiber laser based on the WS2 SAb. Using fiber-taper WS, SAb, the large spectral width (114 nm) and pulse duration as short as 67 fs have been achieved (Figure 13C, d) in a mode-locked erbium-doped fiber [262].

SAbs made of TMDCs have also been used for ultrafast lasers well beyond their lower energy gap due to the aforementioned mechanisms [170, 264]. Importantly, BP SAb is a potential candidate for telecom and mid-IR lasers [96]. Thus, it has been successfully used in bulk and fiber lasers for the generation of ultrashort pulses and high-power lasers [272–274]. For example, reflective-type ML BP gold film SAb has been demonstrated for Q-switched fiber laser in the spectral range of 1.03 to 2.72 μm [96]. Recently, Su et al. used FL BP SAbs for high-power solid-state laser with a pulse width of 272 fs at a repetition rate of 63.3 MHz with an average power of 0.82 W at a central wavelength of 10534 nm [274]. Moreover, TIs and perovskites have also been demonstrated for highly efficient SAbs (large modulation depth and low saturation intensity) in a wide wavelength range for the generation of ultrashort pulses [275–277].

As another important application, SAbs made of 2DLM heterostructures were successfully used for the highly efficient ultrashort pulse generation with improved performance of large modulation depth, ultralarge pulse energy, and ultranarrow pulse width [92, 125, 178]. For example, graphene/BP heterostructures have been used to generate stable trains of pulse with energy of 267.5 nJ in Q-switching operation and narrow pulses with a width of 148 fs in passive mode-locking operation, whose energy is 21 times larger and pulse width is 83% shorter than those achievable using BP SAb [177].

### 6.2 Imaging and characterization of 2D crystals

In addition to measuring nonlinear optical constants, the wave mixing ability of 2DLMs has been widely used as a tool for all-optical imaging [60, 86] and accurate characterization of 2DLMs, including crystal uniformity [29], crystallographic properties [35, 148], crystalline nature [221], grain boundaries [198], number of layers [86], layer stacking order [196, 224], chemical dopant [218], and electronic structure [33]. By probing the polarization of HG waves, one could identify the crystallographic axis [64, 201], crystalline symmetry [33, 86, 100, 146], and polarity of crystalline domains [198, 221] of 2DLMs.

Figure 14A shows the polar plot of SHG emission from SL MoS2 at a fixed position as a function of sample angle. The clear sixfold pattern indicates the threefold rotational symmetry property of the MoS2 2D crystal [33]. Interestingly, the electronic structure of 2DLMs changes at the crystal edges, which enables to probe the crystalline boundaries based on HG measurements [198, 219, 229, 278]. It is worth mentioning that Yin et al. have observed a strong edge nonlinear optical response in MoS2 and demonstrated a large area detection of structural discontinuities or edges of the crystalline grains (Figure 14B, i) [198]. By probing the spatial mapping of polarization resolved SHG emission, the orientation of individual crystal domains in the samples was also observed (Figure 14B, ii).

The optical nonlinearity of 2DLMs is about two to three orders of magnitude larger than that of substrates such as silicon/SiO2 and gold [29, 60, 63]. It enables to achieve high-contrast visualization of 2DLMs (Figure 11E) based on nonlinear wave mixing, including SHG [33, 210], THG [29, 202, 211], and FWM [60]. Moreover, the strong variation of the nonlinear optical response of 2D crystals with the number of layers allowed one to probe the sample thickness (Figure 14C, b), a technique that has poor imaging resolution in the linear optical regime [60, 201, 209].

Heterostructures formed by stacking homolayer or heterolayer 2D crystals have excellent nonlinear optical properties and can be used in highly functional devices. However, the stacking order (orientation) plays a key role in the performance of heterostructures [224, 279]. Thus, SHG is very sensitive to the crystalline symmetry and is an efficient probe for the stacking orientation of 2D crystals in BL and ML crystals [100, 224]. For example, Figure 14D shows the stacking order probed in homolayer BL GaSe using SHG imaging [100]. Thus, the stacking configuration of ε-GaSe ensures that it is noncentrosymmetric and belongs to the D3h group, whereas β-GaSe is...
centrosymmetric and belongs to the $D_{1d}^1$ group. The SHG images of GaSe with a twist angle of $\phi = 0$ (Figure 14D, b) and $\phi = 60^\circ$ (Figure 14D, e) clearly indicates the $\epsilon$ and $\beta$ stacking orders, respectively.

### 6.3 Giant nonlinear enhancement based on double plasmonic resonance

In this section, we present a recently proposed mechanism to enhance THG in graphene nanoribbons (GNRs) based on a so-called double-resonance plasmon effect [58]. The geometrical structure of graphene gratings used to illustrate this idea is schematically presented in Figure 15. It represents a 1D grating with period $A$ and width of graphene ribbons $W$. Graphene structures are assumed to be located at $z = z_s$ and placed on a substrate with relative permittivity $\epsilon_s$ (for specificity, assumed to be glass, $\epsilon_s = 2.25$).

The mechanism of THG enhancement can be understood as follows. By varying the width of graphene ribbons, one can engineer the spectral resonances of the grating to ensure that the resonance wavelength of the fundamental plasmon coincides with the wavelength of...
the incoming beam, whereas the resonance wavelength of one of the higher-order plasmons is exactly a third of the resonance wavelength of the fundamental plasmon mode. Under these circumstances, the diffraction grating will be efficiently excited at the FF, which will lead to a strong field enhancement at this frequency and radiate effectively at the TH, as there is a plasmon resonance at this wavelength, too. In effect, such a diffraction grating would act as a highly effective receiver at the FF and a strong emitter (efficient antenna) at the TH.

Using the numerical method GS-RCWA described in Section 3.1, the dispersion map of the linear optical response of graphene ribbons is calculated and presented in Figure 16. The results suggest that it is indeed possible to engineer an optical diffraction grating with the desired property, because it can be seen that there are certain values of the graphene ribbon width $W$, for which a graphene plasmon mode exists at both FF and TH wavelengths. To be more specific, for $W=85$ nm, graphene diffraction grating supports a (fundamental) plasmon mode at the FF corresponding to $\lambda_{\text{FF}}=9.03 \, \mu\text{m}$ and a third-order plasmon mode at $\lambda_{\text{TH}}=\lambda_{\text{FF}}/3=3.01 \, \mu\text{m}$.

The dispersion map of the nonlinear optical response of the diffraction grating is presented in Figure 16 together with the dependence of the TH intensity on the width of graphene ribbons, determined for the resonance wavelengths of the fundamental plasmon, which is shown in the inset of this figure. An important result illustrated by this figure is that the excitation of graphene localized plasmons at the FF induces a large increase of the intensity at the TH via local field enhancement, as it is clear from the location of the spectral resonance bands in Figure 16.

More importantly, the plot shown in Figure 16B (inset) proves that a further enhancement of the TH intensity occurs when the double-resonance condition holds. Thus, the plot in Figure 16B (inset) was determined by choosing the wavelength of the incident beam to be equal to the resonance wavelength of the fundamental plasmon and varying the width of the ribbons. It can be seen that a maximum intensity of the TH is achieved when $W=85$ nm, that is for the width at which there is a plasmon at both FF and TH. Importantly, from the results in Figure 16A, one finds that the double-resonance condition is fulfilled when graphene ribbons are in relative close proximity to their nearest neighbors. Therefore, although the excitation of localized surface plasmons on graphene ribbons plays the major role in the observed enhancement of the TH intensity, the optical coupling between neighboring ribbons and other diffractive effects could affect as well the optical response of the graphene structure.

6.4 Polarization control using stacked 2D metamaterials

In this subsection, a novel graphene nanostructure to manipulate the polarization state of THz waves is introduced [280]. The structure is composed of two crossed graphene gratings that are separated by an insulator spacer. The schematic diagram of the photonic nanostructure is illustrated in Figure 17. The unit of this crossed GNR-insulator-GNR
A (GIG) metamaterial is depicted in Figure 17 (inset), where the periods of the unit cell are indicated by parameters $L_1$ and $L_2$, whereas the widths of the bottom and top ribbons are represented by parameters $W_1$ and $W_2$, respectively.

To study the optical response of this crossed graphene polarizer, a THz wave carrying a modulated Gaussian signal is used to stimulate this linear system, the frequency band of this Gaussian pulse ranging from 30 to 80 THz. The polarization angle of the incident plane wave is indicated by the $\alpha$ symbol in Figure 17. In this example, this polarization angle $\alpha$ varies from 0 to $\pi/2$. Moreover, the geometrical parameters are $H=100$ nm, $L_1=L_2=200$ nm, and $W_1=W_2=150$ nm. In this example, the value of temperature is 300 K, the relaxation time of graphene is 0.2 ps, and its Fermi energy is 0.2 eV. Based on the ADE-FDTD method [56], the linear response of this crossed graphene polarizer is studied and the results are given in Figure 18.

There are two clear resonance peaks at different wavelengths (Figure 18), although the bottom GNRs have the same size as the top one. The main reason is that they are in contact with different media. Generally, the larger the permittivity of its adjacent medium is, the higher is the resonance wavelength of graphene gratings. For this reason, the resonance wavelength of the bottom GNRs is higher than that of the top GNRs.

The optical response of this crossed graphene polarizer is polarization dependent as shown in Figure 18. For instance, when the incident polarization angle $\alpha=0$, the absorption spectrum (Figure 18, red dashed line) suggests that the interaction between the incident wave and the top GNRs is the strongest at 41 $\mu$m, yet the bottom GNR layer is almost transparent at this wavelength. The opposite situation holds when $\alpha=\pi/2$. This behavior proves that each layer of GNRs can be regarded as a polarization selective metasurface, which is transparent to a parallel electric field but strongly coupled to a perpendicular one. This polarization-dependent feature of crossed graphene gratings can be used to miniaturize the size of traditional polarization polarizers [280].

The polarization-selective feature of a properly designed GIG metamaterial can be used to develop a novel polarization converter [280] as schematically shown in Figure 17. To study the linear and nonlinear response of this polarization converter, an incident plane wave carrying a sinusoidal signal is used.

The physical mechanism of the polarization conversion in the polarization converter is illustrated in Figure 19, where the linearly polarized incident wave with a polarization angle $\alpha=\pi/4$ is decomposed into two orthogonal linearly polarized waves. The electric field component of the vertical wave is polarized along the $y$-axis and the electric field component of the horizontal wave is polarized along the $x$-axis. Based on the polarization-selective feature of the graphene polarization converter, the vertical linearly polarization wave $E_y$ will
pass through the top layer of GNRs without any interaction but react with the bottom layer of GNRs very efficiently as shown in Figure 19B. In contrast, the bottom layer of GNRs is transparent to the linearly polarization wave Ex but strongly interacts with the top layer of GNRs.

Considering this polarization-selective feature, an optical path difference exists between the two waves propagating in the graphene nanostructure, and its value depends on the thickness d of the insulator spacer. For this reason, there is a phase difference \( \Delta \phi \) between the two waves, too. By tuning this phase difference, the output wave can be engineered to have a desired polarization state, including the right- and left-hand circular polarization, as schematically illustrated in Figure 19C.

An analytical formula for the phase difference \( \Delta \phi \) can be derived as [280]

\[
\Delta \phi(z) = \begin{cases} 
0, & z < z_b, \\
\frac{4\pi n}{\lambda_m} (z - z_b), & z_b < z < z_t, \\
\frac{4\pi n}{\lambda_m} d, & z > z_t.
\end{cases}
\] (27)

where \( \lambda_m = \lambda / m \), with m as the order of the nonlinear process (\( m = 3 \) for THG).

For instance, when \( \lambda_T = 23 \mu m \), the analytical dependence of the phased difference \( \Delta \phi(z) \) on the thicknesses \( d \) is given in Figure 20, where the analytical results are calculated from Eq. (27) and the numerical results are evaluated using the FDTD method. It was assumed that the bottom GNR layer was at \( z_b = 0 \) and the top one at \( z_t = d \). These analytical results show that the phased difference \( \Delta \phi = 0 \) in the transmission region as predicted by Eq. (27). In the insulator region, \( \Delta \phi \) linearly increases with \( z \) as predicted by Eq. (27). In the reflection region, the phase difference remains constant as suggested by Eq. (27). These analytical results prove that the linearly polarized incident wave can be converted in reflection to a wave with arbitrary polarization by tuning the thickness of the spacer or the resonance wavelength.
The results in Figure 21 show that the phase difference $\phi$ in the reflection region is proportional to the higher-order number $m$. Thus, the nonlinearity of graphene can be used to further improve the efficiency of the polarization conversion. To validate this idea, the nonlinear response of this graphene polarization converter was simulated using a GS-FDTD code; the relevant results are presented in Figure 21. Compared to the linear results in Figure 20, the nonlinear optical response of the GIG structure shows that the phase difference in TH case is three times larger than that at the FF for the same ratio $\eta$. This feature is particularly useful to further miniaturize the size of the graphene polarization converter. In addition, the results in Figure 21 show that the phase difference $\phi$ can fully cover $[0, 2\pi]$ when the ratio $\eta$ is up to 0.12.

7 Conclusions

In this review, we have presented past and present research in the nonlinear optics of photonic structures made of graphene and other 2D materials. This is a rapidly growing field of science as it has been generating tremendous interest in many new and exciting optical phenomena and a plethora of important emerging applications in active photonic nanodevices, nonlinear integrated photonics, optoelectronics, surface science, and biomedicine. In particular, we have discussed the main linear and nonlinear optical properties of graphene and other 2D materials, the computational methods used to model such materials, the main nanofabrication methods for 2D materials and experimental techniques for their characterization, as well as the exciting technological applications and active photonic devices that rely on 2D materials. Bringing together in this way in the same place the main tools employed in research in the nonlinear optics of 2D materials, namely theoretical modeling, computer simulations, nanofabrication methods, and experimental techniques, makes it easy for the interested reader to form an informed opinion and readily gain valuable knowledge about this dynamic and exciting field of science.

The pace at which research in 2D materials has been expanding makes us believe that many exciting new physics pertaining to the nonlinear optics of 2D materials are yet to be discovered. These impending future developments will be spurred not only by the possibility to discover new 2D materials but also by our increased ability to integrate these materials in novel photonic nanodevices with new or improved functionality. Of course, however, the most exciting developments will be those that we presently do not anticipate.

Acknowledgments: J.W. You and N.C. Panoiu acknowledge support from the European Research Council/ERC Grant Agreement No. ERC-2014-CoG-648328, Funder Id: http://dx.doi.org/10.13039/100010663 and S.R. Bongu and Q. Bao acknowledge support from the National Natural Science Foundation of China (NSFC) (grant 61875139), Shenzhen Nanshan District Pilotage Team Program (grant LHTD20170006) and the Australian Research Council (DP140101501, IH150100006, FT150100450, and CE170100039).

References


