ARTICLE   OPEN

Check for updates

Abnormal nonlinear optical responses on the surface of topological materials

Haowei Xu[1], Hua Wang[1] and Ju Li [ID][1,2 ✉]

The nonlinear optical (NLO) responses of topological materials are under active research. Most previous works studied the surface and bulk NLO responses separately. Here we develop a generic Green's function framework to investigate the surface and bulk NLO responses together. We reveal that the topological surface can behave disparately from the bulk under light illumination. Remarkably, the photocurrents on the surface can flow in opposite directions to those in the bulk interior, and the light-induced spin current on the surface can be orders of magnitude stronger than its bulk counterpart on a per-volume basis. We also study the responses under inhomogeneous field and higher-order NLO effect, which are all distinct on the surface. These anomalous surface responses suggest that light can be a valuable tool for probing the surface states of topological materials. Besides, the surface effects should be prudently considered when investigating the optical properties of topological materials.

INTRODUCTION

In recent years, nonlinear optical (NLO) effects such as the bulk photovoltaic (BPV) effect have attracted substantial interest, owing to their potential applications in e.g., photodetection[1–4], energy harvesting[5–10], and material characterization[11–14]. The interplay between topology and NLO properties is particularly intriguing. Certain NLO effects are closely related to topological quantities such as the Berry curvature and quantum metric tensors, thus the NLO responses can be utilized as a probe of these bulk topological properties[11–13,15–19]. On the other hand, the topological nature can boost the NLO responses[2,20,21], thus the efficiency of applications such as photodetection could be enhanced by using topological materials.

Regarding NLO properties in topological materials, most works hitherto studied the bulk[1–4,11–29] and surface[30–36] responses separately. Only a few works studied the NLO responses of the surface and bulk together. Actually, even in normal materials, the surface and bulk responses can be substantially different. A typical example is that surface naturally breaks the inversion symmetry, which forbids even-order responses. Hence even-order responses are always allowed on the surfaces, even if the bulk possesses inversion symmetry. In addition, in topological insulators, when the Fermi level is within the bulk bandgap, some NLO responses are zero in the bulk when the light frequency is below the bulk bandgap. In contrast, the surface always has non-zero responses due to the gapless topological surface states. These considerations necessitate careful inspections of the surface effects in NLO processes. Typically, a slab model with finite thickness is used to study the surface effects. This approximate model omits a few essential interactions: the surface electrons should interact with all the bulk electrons, and the bulk should be (nearly) infinite in depth in many experimental scenarios. Besides these concerns pertinent to surface effects, previous studies on the NLO effects are mostly based on the non-interacting single-particle framework (independent particle approximation), and the many-body effects are ignored. Indeed, many-body effects such as electron-phonon

coupling[37,38], excitonic effect[39,40], and strong correlations[41] can greatly influence the optical responses.

In this work, we develop a generic many-body framework for computing the NLO effects based on the Green's function formalism, which can naturally incorporate various many-body effects. In the Supplementary Note 2.4, we study strong electron-electron correlations as an example to showcase the merit of the Green's function framework. Besides second-order responses such as BPV, our Green's function formalism can be systematically and conveniently extended to handle higher-order NLO effects and inhomogeneous light fields. We apply our framework to the surface states of topological materials, whose Green's functions are obtained with the iterative Green's function (IGF) method[42,43]. This approach enables a rigorous treatment of the surface-bulk interactions. We take type-II Weyl semimetal T_d-WTe$_2$[44,45] as an example. The bulk of T_d-WTe$_2$ is non-centrosymmetric and gapless, which is similar to the surface from a symmetry or bandgap point of view. However, the NLO responses on the surface are distinct from those in the bulk interior. Specifically, the BPV charge current on the surface and in the bulk can flow in opposite directions (Fig. 1). This striking behavior demonstrates that surface effects in topological materials can be significant. We clarify that this is mainly a topological effect and can be absent in normal materials. Also, the surface bulk spin photovoltaic (BSPV)[46,47] conductivity is colossal and can be larger than its bulk counterpart by a factor of 10 on a per-volume basis. Hence the surface of topological materials can be efficient platforms for spintronics applications, particularly when 2D materials or nanoscale thin films are required. In addition, we show that the responses under inhomogeneous fields and higher-order NLO effects are all distinct between the surface and the bulk. These anomalous responses on the surfaces indicate that the NLO effects can be utilized to probe the surface atomic structure. On the other hand, the surface responses should be prudently considered when investigating the NLO properties of topological materials. This is particularly important when the penetration depth of the light is shallow or when the material is thin, like less than 10^2 monolayers

[1]Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [2]Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. ✉email: liju@mit.edu
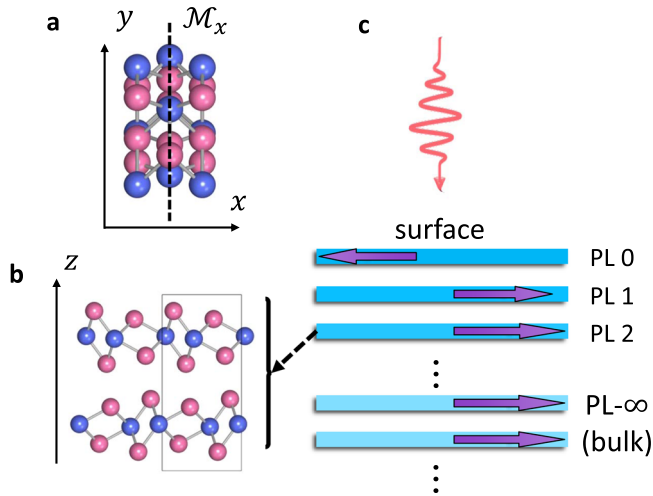
Fig. 1 Illustration of T$_d$-WTe$_2$ under light illumination. a, b Atomic structure of T$_d$-WTe$_2$. The dashed line in **a** denotes the mirror-x symmetry $\mathcal{M}_x$ of T$_d$-WTe$_2$. **c** A sketch of materials under light illumination. The surface and bulk principal layer (PL) corresponds to PL-0 and PL-∞, respectively. The optical responses of the surface and the bulk can be distinct (indicated by the purple arrows). For T$_d$-WTe$_2$, each PL consists of two WTe$_2$ layer without inversion symmetry as shown in **b**.

(~50 nm) thick. Given the semiconductor industry has generally moved to sub-10 nm technology nodes, now is an opportune time that such surface effects are addressed theoretically.

RESULTS

General theory

To calculate the responses under light, one needs the thermo-dynamic and quantum average of an observable θ, which can be formulated as[48]

$$\langle\theta\rangle = -i\int\frac{d^d\mathbf{k}}{(2\pi)^d}\frac{dE}{2\pi}\text{Tr}\{\theta G^<(\mathbf{k},E)\}. \tag{1}$$

Here θ can be various observables, such as charge current ($\theta = -e\mathbf{v}$ with $\mathbf{v}$ as the velocity operator), spin current [$\theta = \frac{1}{2}(\mathbf{vs}+\mathbf{sv})$ with $\mathbf{s}$ as the spin operator], etc. $\int\frac{d^d\mathbf{k}}{(2\pi)^d}$ indicates the integration over the Brillouin zone in d-dimension. Tr is the trace operation. $G^<(\mathbf{k},E)$ is the lesser Green's function and plays a role of energy-resolved distribution function. In non-interacting systems, one has $[G_0^<(\mathbf{k},E)]_{mn} = 2\pi i\delta_{mn}f_m\delta(E-E_m)$, where m and n are band indices, and f_m and E_m are the occupation number and energy of band m at wavevector $\mathbf{k}$, respectively. In this case, one has $-i\int\frac{dE}{2\pi}[G_0^<(\mathbf{k},E)]_{mn} = \delta_{mn}f_m$ and $\langle\theta\rangle = \int\frac{d^d\mathbf{k}}{(2\pi)^d}\sum_m\theta_{mm}f_m$, which is the usual thermal average of $\langle\theta\rangle$. When the electrons have interactions with phonon, defects, other electrons, etc., $G^<(\mathbf{k},E)$ usually does not have a simple expression, but it can be obtained perturbatively with e.g., Feynman diagrams or non-perturbatively with other methods[48] [see Supplementary Note 2.4 for an example]. At equilibrium, the expectation of some observables, such as the charge current, should be zero. However, light can drive the system out of equilibrium, resulting in nonzero $\langle\theta\rangle$. Specifically, the interaction with light leads to a change in the lesser Green's function $\delta G^<$, which is dependent on the electric field $\mathcal{E}$. Then perturbatively one has $\langle\theta\rangle = A\mathcal{E}+B\mathcal{E}^2+C\mathcal{E}^3\ldots$, where A, B and C correspond to the first-, second- and third-order optical response functions.

In the following, we use WTe$_2$ in its T$_d$ phase (space group Pmn2$_1$, no. 31) as an example to study the NLO response on the

surface and in the bulk of topological materials. The atomic structure of T$_d$-WTe$_2$ is shown in Fig. 1. The lattice constants are $a = 3.48$ Å, $b = 6.28$ Å and $c = 14.0$ Å. We define a Cartesian coordinate with x, y and z along the crystallographic a, b and c directions, respectively. T$_d$-WTe$_2$ is non-centrosymmetric, but has a mirror-x symmetry $\mathcal{M}_x$ (dashed line in Fig. 1a). The unit cell of T$_d$-WTe$_2$ consists of two WTe$_2$ layers stacked along z direction (Fig. 1b), which is used as a principal layer (PL). Each PL has interactions with other PLs, and these interactions can be included in the equilibrium Green's function G_{PL}^0 using the IGF method[42,43]. Electrons on each PL also have interactions with phonons, defects, etc. These interactions are implicitly represented by a phenom-enological electron lifetime τ, which is taken to be a uniform value of 0.2 ps throughout this work unless explicitly stated. Under the current framework, the influence of τ is incorporated in the non-interacting Green's function as $G_0(E) = [E-H_0+\frac{i\hbar}{\tau}]^{-1}$, where H_0 is the non-interacting single-particle Hamiltonian (Supplementary Note 3). The choice of $\tau = 0.2$ ps is based on experimental results and should be a conservative value[49–53]. In Supplementary Note 5.1, we show how NLO responses vary with τ, and demonstrate that the main conclusions of our work remain valid for a wide range of τ. In order to demonstrate how our Green's function formalism can incorporate other many-body interactions, we also artificially add a Hubbard $U = 3$ eV term on the d orbitals of W atoms, and the Green's functions of this strongly correlated system are calculated with the dynamical mean field theory (DMFT)[54–56]. We find that even with the artificial Hubbard U term, the main conclusions of our work remain valid (Supplementary Note 2.4).

The Green's function of each PL from the surface into the bulk can be obtained with the IGF method, and the PL-resolved responses, defined as the responses localized on each PL, can be obtained by putting G_{PL}^0 in the Green's function formalism (Methods). Specifically, the surface and the bulk correspond to PL-0 and PL-∞, whose Green's functions are denoted as G_{PL-0}^0 and $G_{PL-\infty}^0$, respectively (Fig. 1 PL-x denotes the x-th PL). In practice, G_{PL}^0 converges with PL $\gtrsim 20$, and we use G_{PL-50}^0 as the bulk Green's function ($G_{PL-\infty}^0 \simeq G_{PL-50}^0$). We fix the Fermi level E_F of T$_d$-WTe$_2$ so that it is charge neutral (no electron/hole doping) throughout this paper (Supplementary Fig. 13), and we study the optical responses in the mid-infrared regime ($\omega = 0.1 \sim 0.5$ eV). The spin-orbit coupling (SOC) is included for both the bulk and the surface.

Bulk photovoltaic effect

BPV effect indicates that in non-centrosymmetric materials, a DC charge current can be generated under photon illumination without any external bias voltage. The BPV effect can be expressed as

$$j^a = \sigma_{bc}^a(0;\omega,-\omega)\mathcal{E}^b(\omega)\mathcal{E}^c(-\omega), \tag{2}$$

where j^a is the current, while a and b/c are the directions of the current and the electric field, respectively. Here $\sigma_{bc}^a(0;\omega,-\omega)$ is the BPV conductivity and can be expressed as (Supplementary Note 2.1)

$$\sigma_{bc}^a(0;\omega,-\omega) = -\frac{ie^3}{\omega^2 S}\sum_{\mathbf{k}}\int\frac{dE}{2\pi}\text{Tr}\{v^a G^<(E)\},$$

$$G^< = G_0^r(E)v^b G_0^r(E+\omega)v^c G_0^<(E)$$
$$+ G_0^r(E)v^b G_0^<(E+\omega)v^c G_0^a(E)$$
$$+ G_0^<(E)v^b G_0^a(E+\omega)v^c G_0^a(E)$$
$$+ (b\leftrightarrow c, \omega\leftrightarrow-\omega). \tag{3}$$

One can see that the response operator θ in Eq. (1) has been taken as $-e\mathbf{v}$, which is the charge–current operator. G_0^r, G_0^a and $G_0^<$ are the retarded, advanced, and lesser Green's function of the system without light illumination, which are calculated with IGF. Here we assume that the light field is uniform. The $\mathbf{k}$ arguments in Eq. (3)
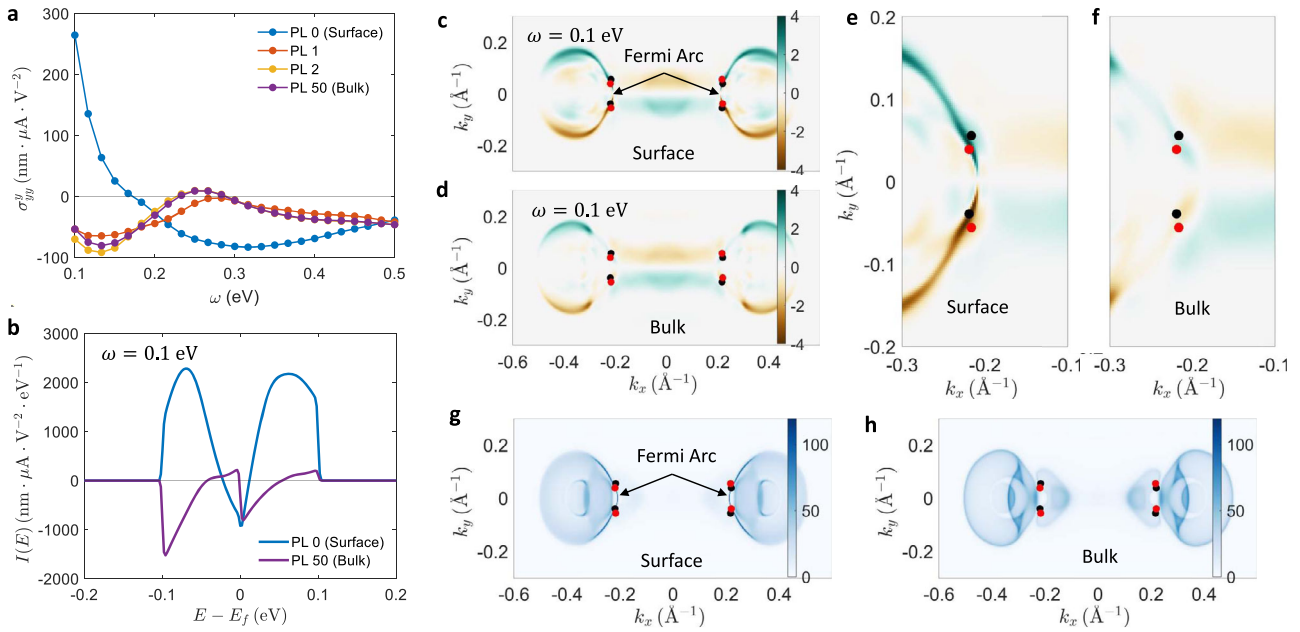
Fig. 2 BPV conductivity of T_d-WTe$_2$. a Layer resolved BPV conductivity for PLs from the surface to the bulk. **b** Energy-resolved contribution to the BPV conductivity for the surface and bulk PL for light frequency $\omega = 0.1$ eV. The energy-resolved contribution is defined as $I(E) \equiv -\frac{ie^3}{\omega^2 S} \sum_{\mathbf{k}} \text{Tr}\{v^a G^<(E)\}$. **c, d** $\mathbf{k}$-resolved contribution, defined as $I(\mathbf{k}, E) \equiv -\frac{ie^3}{\omega^2 S} \text{Tr}\{v^a G^<(\mathbf{k}, E)\}$ at $\omega = 0.1$ eV and $E = E_F$ for (**c**) surface PL and (**d**) bulk PL. **e, f** Same as (**c, d**), but zoomed-in around the Fermi arc. **g, h** Spectrum function $A(\mathbf{k}, E)$ at $E = E_F$ for (**e**) surface and (**f**) bulk PL. The Fermi arcs are labelled by the arrows in (**c, e**). In (**c–h**) the black and red dots mark the locations of Weyl points with $+1$ and -1 chirality, respectively.

are omitted for simplicity. S is the area of the unit cell. ($b \leftrightarrow c, \omega \leftrightarrow -\omega$) indicates the simultaneous exchange of b, c and $+\omega, -\omega$, which symmetrizes $\mathcal{E}^b$ and $\mathcal{E}^c$.

The layer-resolved BPV conductivities of T_d-WTe$_2$ under linearly polarized light with y-polarization are shown in Fig. 2a. We show $\sigma(0; \omega, -\omega)$ for $\omega > 0.1$ eV because Eq. (3) has a spurious divergence as $\omega \to 0$, due to the $\frac{1}{\omega^2}$ factor. It is clear that the surface layer (PL-0) has distinct responses from the other PLs. The responses on the first (PL-1) and second (PL-2) are also different from those in the bulk, despite that the differences are less significant than those in the case of PL-0. This implies that the surface effects can penetrate three PLs, with a total thickness of around 5 nm. Remarkably, in some frequency region σ_{PL-0} has the opposite sign to σ_{PL-1} and $\sigma_{PL-\infty}$, indicating that under photon illumination, the local charge current would flow in opposite directions on PL-0. This is counter-intuitive, as PL-0 is directly attached to PL-1. Moreover, in some frequency regions $\sigma_{PL-\infty}$ is close to zero, while σ_{PL-0} has a finite value, thus the current would flow mostly on the surface. The BPV conductivities under circularly polarized light are also distinct for the surface and the bulk (Supplementary Fig. 14). The BPV responses are closely related with geometric and topological properties[15–17]. Thus, the layer-resolved NLO responses can be harnessed to probe and characterize the topological surface states. Another interesting observation is that the counter-propagating currents (Fig. 1) may lead to a magnetic field $\mathbf{B}$ between PL-0 and PL-1, whose magnitude can be estimated using Ampère–Maxwell law as $\mathbf{B}(\mathbf{r}) \approx \frac{1}{2} \mu_0 \mathbf{j} \times \hat{\mathbf{d}}$, where μ_0 is the vacuum permittivity, $\mathbf{j}$ is the current density on a single PL, $\mathbf{r}$ is the spatial location, while $\hat{\mathbf{d}}$ is a unit vector that points from the PL to $\mathbf{r}$, and is perpendicular to the PL plane. Since $j = \sigma \mathcal{E}^2$, the magnitude of the magnetic field is $|B| \approx \frac{1}{2} \mu_0 |j| = \frac{1}{2} \mu_0 \sigma \mathcal{E}^2 \sim 6.3 \times 10^{-4} \mathcal{E}^2$ [T], where σ is taken as 100 nm·μA/V^2, and $\mathcal{E}$ is in the unit of MV/cm. Note that this magnetic field $\mathbf{B}$ has a pure orbital-magnetic origin, and the spin contribution is not

included. Specifically, an electric field of 1 MV/cm can generate a detectable interlayer magnetic field of 6 Gauss. In experiments, an electric field of 1 MV/cm is readily available and can be below the material damage threshold if pulsed lasers are used.

Next, we argue that the surface effects described above are mainly topological effects. First, we look at the $\mathbf{k}$, E-resolved contribution to σ in Eq. (3). The energy spectrum $I(E) \equiv -\frac{ie^3}{\omega^2 S} \sum_{\mathbf{k}} \text{Tr}\{v^a G^<(E)\}$ for $\omega = 0.1$ eV is plotted in Fig. 2b, where one can see that $I_{PL-0}(E)$ is generally different from $I_{PL-\infty}(E)$. Besides, although $G^<(E)$ should be nonzero for a wide range of E, $I(E)$ is nonzero only when $|E| \lesssim \omega$. This indicates that the shift current is essentially a resonant interband process: light with frequency ω can assist the electrons to transit from $(\mathbf{k}, E)$ to $(\mathbf{k}, E + \omega)$, but due to the Pauli exclusion principle, such a transition is allowed only when $f_{FD}(E)[1 - f_{FD}(E \pm \omega)] > 0$, leading to $|E| \lesssim \omega$. Here f_{FD} is the Fermi–Dirac distribution. The $\mathbf{k}$-resolved contribution defined as $I(\mathbf{k}, E) \equiv -\frac{ie^3}{\omega^2 S} \text{Tr}\{v^a G^<(\mathbf{k}, E)\}$ reveals more detailed information on the surface effects. $I(\mathbf{k}, E)$ for $\omega = 0.1$ eV at the Fermi level $E = E_F$ is plotted in Fig. 2c, d. The difference between $I_{PL-0}(\mathbf{k}, E_F)$ and $I_{PL-\infty}(\mathbf{k}, E_F)$ is also significant, which can be inferred from the spectrum function defined as $A(\mathbf{k}, E) \equiv i \text{Tr}\{G_0^r(\mathbf{k}, E) - G_0^a(\mathbf{k}, E)\}$ (Fig. 2g, h). One can see that the difference between $I_{PL-0}(\mathbf{k}, E_F)$ and $I_{PL-\infty}(\mathbf{k}, E_F)$ lies largely in the region where the surface Fermi arc is located (labelled in Fig. 2c, g). This indicates that the difference between σ_{PL-0} and $\sigma_{PL-\infty}$ is mainly from the topological surface states (see zoom-in plots around the Fermi arc in Fig. 2e, f). We have also examined $I(\mathbf{k}, E)$ for $E \neq E_F$, and found that it is generally different for PL-0 and PL-50 (see e.g., Supplementary Fig. 19), although after the integration over the Brillouin zone, $I(E)$ can be close in some E-regions, such as $E \sim E_F + 0.8$ eV (Fig. 2b). Note that in Fig. 2c–f there are no odd parities for $+y$ and $-y$, i.e., $I(\mathbf{k}, E) \neq I(\mathcal{M}_y \mathbf{k}, E)$, where $\mathcal{M}_y \mathbf{k}$ is the mirror-y image of $\mathbf{k}$.
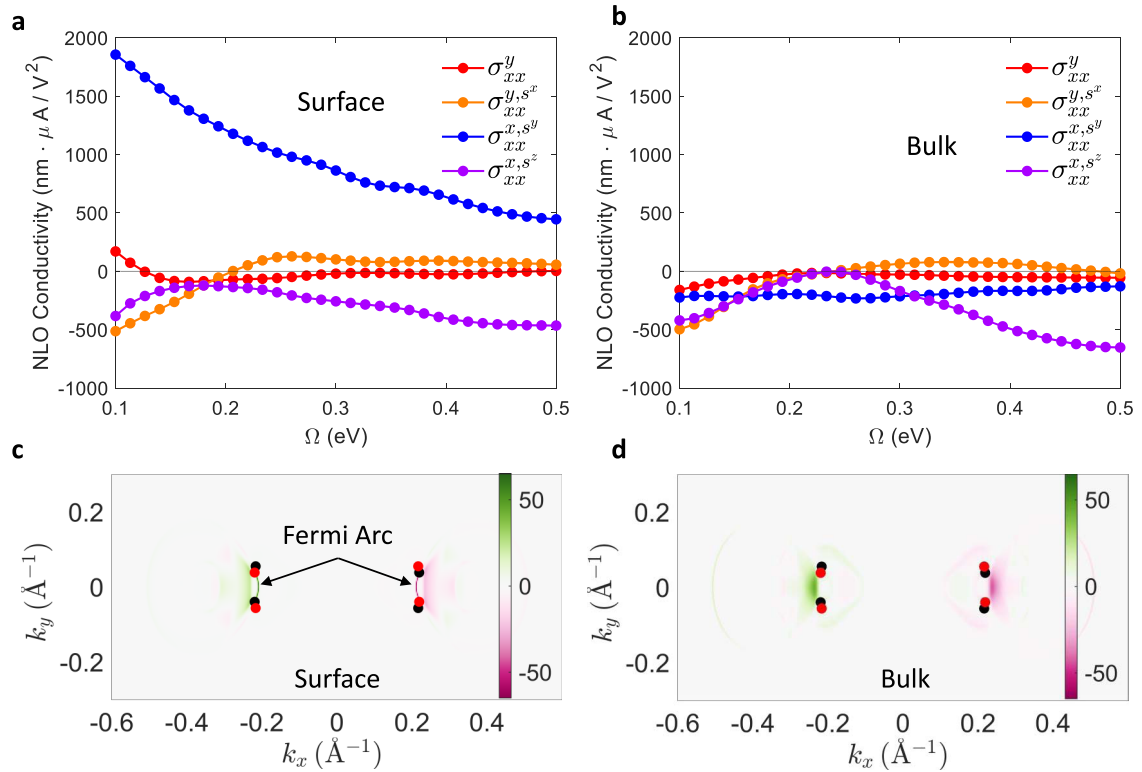
Fig. 3 BSPV conductivity for T_d-WTe$_2$. Nonzero elements of the conductivity tensor are shown for the (**a**) surface PL and (**b**) bulk PL. **c, d** Spin density of states, defined as $D_{s^y}(\mathbf{k}, E) = i\mathrm{Tr}\{s^y[G_0^r(\mathbf{k}, E) - G_0^a(\mathbf{k}, E)]\}$ at $E = E_f$ for (**c**) surface PL and (**d**) bulk PL. In (**c, d**) the black and red dots to mark the locations of Weyl points with $+1$ and -1 chirality, respectively.

The surface responses can be influenced by non-topological surface effects as well, such as symmetry effect and surface spin-orbit coupling. The symmetry effect is the most prominent when the bulk has inversion symmetry. In this case, the second-order responses are only present on the surface and are forbidden in bulk. We also studied Au as an example to show how spin-orbit coupling can influence nonlinear optical response. As spin-orbit energy in Au is relatively large (around 0.5 eV per atom), it makes a non-negligible difference in the NLO responses for ω in the infrared range. To further distinguish the topological surface effects from trivial surface effects, we artificially strain T_d-WTe$_2$ so that it becomes topologically trivial. In this case, the differences between surface and bulk NLO are much less significant, indicating that the trivial surface effects, including symmetry effect and surface spin-orbit coupling, are minor in the case of T_d-WTe$_2$. Besides, we investigate topologically trivial materials such as 2H-MoS$_2$ and find that their σ_{PL-0} and $\sigma_{PL-\infty}$ are generally close to each other (Supplementary Note 4). This again suggests the essential role of topological band effects. Finally, we would like to remark that topological surface effects have a significant impact on the NLO response, but it does not necessarily lead to opposite currents on the surface and in the bulk for all topological materials under all conditions (see Supplementary Fig. 18 for an example).

Bulk spin photovoltaic effect

Electrons have both charge and spin. When electrons move under light illumination, their charge degree of freedom leads to a DC charge current, which is the BPV effect discussed in the previous section. Concurrently, the spin degree of freedom leads to a DC spin current, which is called the bulk spin photovoltaic (BSPV) effect[46,47]. BPV and BSPV are cousin processes and have similar physical origins. The microscopic mechanism of the BSPV can be explained in the following way: when electrons are pumped into the conduction bands by light, the spin-current operator $j^{a,s^i} =$

$\frac{1}{2}(v^a s^i + s^i v^a)$ for electrons on $+\mathbf{k}$ and $-\mathbf{k}$ would not cancel each other, and a net spin current can be generated. Here s^i is the spin operator. However, since spin is an axial vector, BPV and BSPV have very different selection rules under symmetry operations. This can be harnessed for the generation of pure spin current——the spin current is allowed by symmetry, while the charge current is forbidden[46]. As each electron carries a charge of e and spin of $\frac{\hbar}{2}$, one may expect that in the sense of equivalating $\frac{\hbar}{2} = |e|$, BPV and BSPV should have similar magnitude. However, on the surface of topological materials such as T_d-WTe$_2$, BSPV can be stronger than BPV by a factor of 10, as we will show below. This makes the surfaces of topological materials ideal platforms for spintronics applications.

For spin current traveling in direction a with spin polarization i, we set the response operator θ in Eq. (1) as $j^{a,s^i} = \frac{1}{2}(v^a s^i + s^i v^a)$. The BSPV conductivity $\sigma_{bc}^{a,s^i}(0; \omega, -\omega)$ has a similar expression to the BPV conductivity σ_{bc}^a in Eq. (3) (Supplementary Note 2.1). Note that we divide σ_{bc}^{a,s^i} by $\frac{\hbar}{2e}$ so that it has the same unit as σ_{bc}^a. The mirror symmetry $\mathcal{M}_x$ of T_d-WTe$_2$ forbids some elements of the B(S)PV conductivity tensor, such as σ_{xx}^x and σ_{xx}^{x,s^x}. This is because polar vectors such as $\mathcal{E}_x, v_x$ flip sign under $\mathcal{M}_x$, while axial vectors such as s^x do not. Under linearly polarized light polarized in x-direction, the nonzero B(S)PV tensors are shown in Fig. 3a, b (we do not consider the current along the out-of-plane z-direction). A prominent feature is that σ_{xx}^{x,s^y} on the surface is almost ten times larger than its bulk counterpart, and is also ten times larger than other components such as σ_{xx}^y. Indeed, hundreds of nm·µA/V^2 are typical values for B(S)PV conductivities in typical topological materials[29,46], while σ_{xx}^{x,s^y} on PL-0 is as large as 2000 nm·µA/V^2 at $\omega = 0.1$ eV. This indicates that the spin current generation is exceptionally efficient on the surface of T_d-WTe$_2$. Such a strong spin current comes partially from the Rashba spin-orbit coupling on the surfaces. In Fig. 3c we plot the spin density
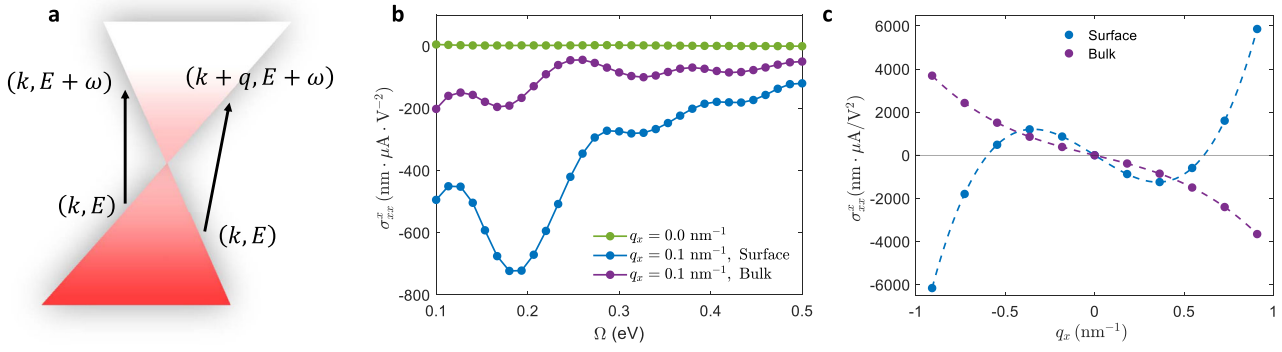
Fig. 4 BPV under inhomogeneous light field. a An illustration of the electron transitions under homogeneous (left arrow) and inhomogeneous (right arrow) field. **b** BPV conductivity under homogeneous ($\mathbf{q} = 0$) and in inhomogeneous ($q_x \neq 0$) field. **c** The relationship between $\sigma^x_{xx}(\omega = 0.1\text{ eV}; q_x)$ and q_x for surface (blue curve) and bulk (purple curve) PL. The dashed curves are fittings of the solid dots with cubic functions.

of states of s^y (equivalent to the expectation value of s^y), defined as $D_{s^y}(\mathbf{k}, E) \equiv i\text{Tr}\left\{ s^y\left[G^r_0(\mathbf{k}, E) - G^a_0(\mathbf{k}, E) \right] \right\}$ at $E = E_F$. One can see that spin-y polarization is stronger around the surface Fermi arc, while for **k**-points distant from the Fermi arc, the spin-y polarization can be weaker. This suggests that BSPV is not solely determined by the spin polarization, but requires synergy with other factors, such as band velocities.

Inhomogeneous light field

In the visible or infrared range, the photon wavevector **q** is much smaller than the size of the Brillouin zone, thus one usually sets **q** = 0 when studying light-matter interactions. In other words, the light field is assumed to be homogeneous. However, the electromagnetic wave can be strongly inhomogeneous in the case of e.g., the Laguerre–Gaussian beam[57] with nonzero angular momentum. Subwavelength-scale variation can also be induced by plasmonic, polaritonic interactions[58,59]. In these situations, the spatial variation of the light field is strong, and **q** should not be neglected[60]. In addition to these practical considerations, finite **q** is also conceptually important, as it breaks certain spatial symmetries, and thus fundamentally alters the selection rules on optical process.

Our Green's function formalism provides a convenient way for incorporating the finite **q** effect. Here we take the second-order BPV as an example. Under an inhomogeneous and oscillating electric field $\mathcal{E}(\omega, \mathbf{q})$ with frequency ω and wavevector **q**, one has a spatially homogeneous and temporally static current, which is

$$j^a = \sigma^a_{bc}(0; \omega, -\omega; \mathbf{q}, -\mathbf{q})\mathcal{E}^b(\omega, \mathbf{q})\mathcal{E}^c(-\omega, -\mathbf{q}), \quad (4)$$

where $\sigma^a_{bc}(0; \omega, -\omega; \mathbf{q}, -\mathbf{q})$ is the conductivity and can be calculated with the Green's function formalism (Supplementary Note 2.2). We will use $\sigma^a_{bc}(\omega, \mathbf{q}) \equiv \sigma^a_{bc}(0; \omega, -\omega; \mathbf{q}, -\mathbf{q})$ as a shorthand in the following. Intuitively, an electron in the state $(\mathbf{k}, E)$ can (virtually) absorbs momentum **q** and energy ω from the inhomogeneous light field, jump to another state $(\mathbf{k} + \mathbf{q}, E + \omega)$, then return **q** and ω to the light field and finally jump back to its original state (Fig. 4a). During this process, the electron may displace in real space, resulting in a charge current. Notably, **q** identifies a preferred (or unique) direction for such a process, which could break certain spatial symmetries, including inversion, mirror, and rotation symmetries.

T_d-WTe$_2$ has mirror symmetry $\mathcal{M}_x$. As a result, $\sigma^x_{xx}(\omega, \mathbf{q})$ must be zero when **q** = 0 since it flips sign under $\mathcal{M}_x$. This is different from σ^y_{yy} studied in the previous sections, which can be non-zero even when **q** = 0. In contrast, if $q_x \neq 0$, then $\sigma^x_{xx}(\omega, \mathbf{q})$ can be nonzero, which is vividly illustrated in Fig. 4b. As a comparison, we keep $q_x = 0$ and vary q_y, and $\sigma^x_{xx}(\omega, \mathbf{q})$ remains zero as expected

(Supplementary Fig. 17), since q_y cannot break $\mathcal{M}_x$. In Fig. 4c, we show how $\sigma^x_{xx}(\omega = 0.1\text{ eV}; \mathbf{q})$ varies with q_x. Again, one can see that PLs on the surface and in the bulk have distinct behavior. $\sigma^x_{xx}(\omega, \mathbf{q})$ of the bulk PLs shows a cubic and monotonic relationship with q_x. In contrast, $\sigma^x_{xx}(\omega, \mathbf{q})$ on the surface is non-monotonic with **q**, and reverses direction at around $|q_x| \approx 0.6\text{ nm}^{-1}$. This is somewhat surprising since $|q_x|$ determines the extent to which $\mathcal{M}_x$ is broken, and one might expect that $\sigma^x_{xx}(\omega, \mathbf{q})$ should increase monotonically with q_x. The non-monotonic behavior can be understood intuitively in the following way. Two competing factors contribute to the magnitude of the NLO responses: (A) when $|\mathbf{q}|$ is large, spatial inhomogeneity of the light field is stronger. As a result, different electrons may have stronger tendencies to jump in the same (rather than the opposite) directions, which could result in a larger $\sigma^x_{xx}(\omega, \mathbf{q})$ when $|q_x|$ is large; (B) when $|\mathbf{q}|$ is large, the wavefunction overlap between $(\mathbf{k}, E)$ and $(\mathbf{k} + \mathbf{q}, E + \omega)$ may be smaller, and the probability (transition rate) for a certain electron to jump between $(\mathbf{k}, E)$ to $(\mathbf{k} + \mathbf{q}, E + \omega)$ may be smaller. This would result in a smaller $\sigma^x_{xx}(\omega, \mathbf{q})$ when $|q_x|$ is large. The competition between factors (A) and (B) can result in a non-monotonic relationship between $\sigma^x_{xx}(\omega, \mathbf{q})$ and q_x. In Supplementary Note 2.2, we give a more quantitative explanation of this effect.

Finally, we would like to remark that besides the inhomogeneous light field, a finite **q** can also arise from inhomogeneous materials, which can be induced by e.g., strain gradient[61] or heterostructures. Inhomogeneous light fields and materials should conceptually lead to similar results, although methodologically, the effect of inhomogeneous materials should be incorporated differently.

Higher-order response

Until now, we have been discussing the second-order optical responses, which scales as $\mathcal{E}^2$. In recent years, higher-order responses have also attracted great interest[34,62–65], and efficient high-order responses have been demonstrated separately on the surface[34] and in the bulk[65] of topological materials. The higher-order responses can be distinct on the surface and in the bulk interior as well. Here we consider the third-order response, which can be characterized by the conductivity $\sigma^a_{bcd}(\omega_b + \omega_c + \omega_d; \omega_b, \omega_c, \omega_d)$ – that is, three electric fields $\mathcal{E}_b(\omega_b), \mathcal{E}_c(\omega_c), \mathcal{E}_d(\omega_d)$ with polarizations (b, c, d) and frequencies $(\omega_b, \omega_c, \omega_d)$ are coupled, and a current along direction a with frequency $\omega_b + \omega_c + \omega_d$ is generated. The detailed formula to calculate $\sigma^a_{bcd}(\omega_b, \omega_c, \omega_d)$ can be found in Supplementary Note 2.3.

Here we consider a simple but typical example of the third-order effect. When one applies two laser beams with

a

b
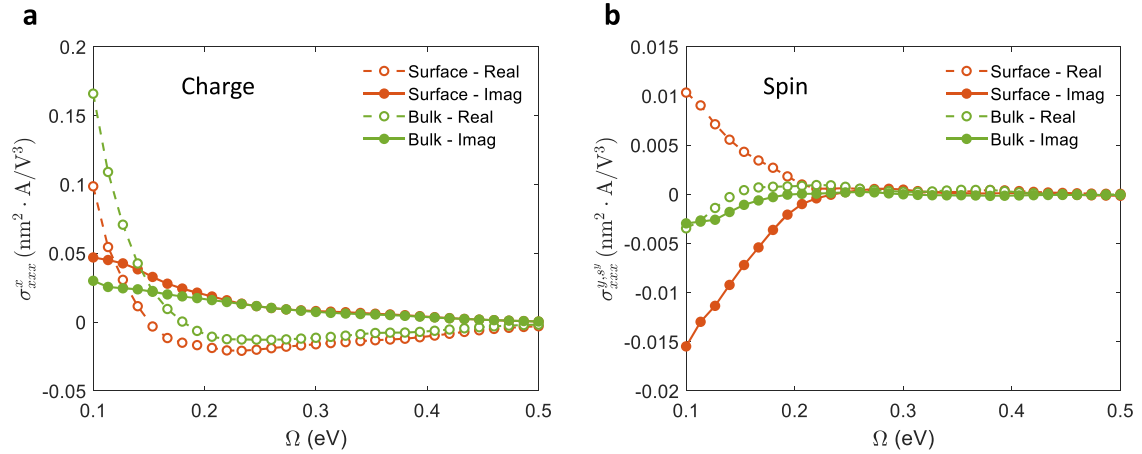
Fig. 5 **Third-order conductivity of T_d-WTe$_2$. a** Charge current and **b** spin current. Red and green curves are for surface and bulk responses, respectively. Solid and dashed curves are the imaginary and real parts of the response functions, respectively.

$\mathcal{E}_b = 2\tilde{\mathcal{E}}_b \cos(\omega t + \phi_b)$ and $\mathcal{E}_c = 2\tilde{\mathcal{E}}_c \cos(2\omega t + \phi_c)$, a static current j^a can be generated as

$$j^a = \tilde{\mathcal{E}}_b^2 \tilde{\mathcal{E}}_c \left[\sigma_{bbc}^a(0;\omega,\omega,-2\omega) e^{i(2\phi_b-\phi_c)} + \sigma_{bbc}^a(0;-\omega,-\omega,2\omega) e^{-i(2\phi_b-\phi_c)} \right]$$
$$= \tilde{\mathcal{E}}_b^2 \tilde{\mathcal{E}}_c \text{Re}\left[\sigma_{bbc}^a(0;\omega,\omega,-2\omega) e^{i(2\phi_b-\phi_c)} \right],$$

$$(5)$$

where we used $\sigma_{bbc}^a(0;\omega,\omega,-2\omega) = \left[\sigma_{bbc}^a(0;-\omega,-\omega,2\omega)\right]^*$, which is a consequence of the time-reversal symmetry. We computed $\sigma_{xxx}^x(0;\omega,\omega,-2\omega)$ with the Green's function formalism, and the results are shown in Fig. 5a, where one can see that the third-order responses are distinct for the surface and the bulk as well. Moreover, $\sigma_{bbc}^a(0;\omega,\omega,-2\omega) \equiv e^{i\psi}\left|\sigma_{bbc}^a(0;\omega,\omega,-2\omega)\right|$ is a complex number, yielding $j^a = \tilde{\mathcal{E}}_b^2 \tilde{\mathcal{E}}_c \left|\sigma_{bbc}^a(0;\omega,\omega,-2\omega)\right| \cos(2\phi_b - \phi_c + \psi)$. Interestingly, since spin flips its direction under time-reversal symmetry, the spin current conductivity (Fig. 5b) satisfies $\sigma_{bbc}^{a,s^i}(0;\omega,\omega,-2\omega) = -[\sigma_{bbc}^{a,s^i}(0;-\omega,-\omega,2\omega)]^*$, and the spin current obeys $j^{a,s^i} \propto \sin(2\phi_b - \phi_c + \psi_{s^i})$, where ψ_{s^i} is the phase factor for the spin current. Therefore, the magnitude of both charge and spin current can be controlled by the phase difference $2\phi_b - \phi_c$. Furthermore, pure spin current without accompanying charge current can be generated when $2\phi_b - \phi_c + \psi = n\pi + \frac{\pi}{2}$, $n \in \mathbb{Z}$. These effects belong to the so-called two-color quantum interference control[66–69].

DISCUSSION

In this work, we used T_d-WTe$_2$ as an example to study the surface effects in NLO processes. The reason we choose T_d-WTe$_2$ is that the bulk of T_d-WTe$_2$ is non-centrosymmetric and semi-metallic, which is similar to the surface from a symmetry or bandgap point of view. Our results indicate that even in this case, the surface responses can still be drastically different from those in the bulk lattice. We clarify that this is mainly a topological effect, as topologically trivial materials do not give such drastic differences. This point is verified with both topologically trivial T_d-WTe$_2$ and 2H-MoS$_2$ (Supplementary Note 4).

As discussed before, in materials with inversion symmetry or nonzero bandgap, the difference between surface and bulk can be even more dramatic. In centrosymmetric materials, B(S)PV, as a second-order effect, is forbidden in the bulk interior. Therefore, under light illumination, the currents are purely on the surfaces, where the inversion symmetry is broken. As for topological insulators, the bulk has a finite bandgap and thus has no responses under light with below-bandgap frequencies (when the Fermi level

is in the bulk bandgap). However, the surface states are gapless and can have responses under light with (in principle) arbitrarily low frequencies. These features are illustrated using Bi$_2$Se$_3$ as an example (Supplementary Fig. 12). Note that in Bi$_2$Se$_3$, the BPV conductivities on the top and bottom surfaces are opposite, thus the total current should be zero if the top and bulk surfaces are under the same light field. However, light with above-bulk-bandgap frequencies may not reach the bottom surface if the Bi$_2$Se$_3$ is thick enough. In this case, the total current comes solely from the top surface and can be directly used to probe the surface states.

In conclusion, we have developed a generic Green's function framework for calculating the NLO properties, which can incorporate many-body effects beyond the single-particle approximation. As an example, we study the effect of strong electron-electron correlations represented by a Hubbard U term. In future works, we will use the Green's function framework to study NLO effects in other many-body systems. In this work, the Green's function framework is used to study the surfaces of topological materials, and it is found that under light illumination, the surface can behave distinctly from the bulk. Therefore, light can be used to probe the surface properties. On the other hand, when investigating the NLO properties of topological materials, the surface effects should be carefully considered, particularly when the light penetration depth is shallow, or the material has nanoscale dimensions. For example, in topological (semi-)metals, only tens of PLs (10–100 nm in thickness) may be active under light, and the surface effect can make a significant difference in the total responses (see Supplementary Note 6.1 for a quantitative estimation). This is important when using light as a probe of topological properties: the desired bulk properties may be obscured by surface effects, and the experimental results may deviate from theoretical predictions if the surface effects are not properly considered. Also, when searching for materials with large NLO responses for e.g., photodetection or energy harvesting purposes, it may be insufficient to look at only the bulk properties. The true responses can be obtained only when the surface and the bulk are both considered, and the surface can act either positively or negatively to the total response.

METHODS

Green's function formalism

Electrons in solid-state systems have interactions with e.g., phonons, defects, and other electrons. We call the electron system with these internal interactions the *base* system and assume that we have full knowledge of these base systems––we have their Green's functions G_0 in hand, which can either be rigorously calculated or be approximately obtained with e.g., perturbative expansions. External fields such as light illumination are treated as additional interactions, which can drive the

system out of equilibrium. Such external interactions are described by the self-energy terms Σ. The Green's functions with both internal and external interactions are denoted with G, which can be obtained from G_0 and Σ using Dyson's equation. Regarding the light-matter interaction, one has[70,71] (Supplementary Note 1)

$$G^< = \left[\sum_{n=0}^{\infty}(G_0^r \Sigma^r)^n\right] G_0^< \left[\sum_{m=0}^{\infty}(\Sigma^a G_0^a)^m\right], \quad (6)$$

where $G_0^r/G_0^a/G_0^<$ are the retarded/advanced/lesser Green's function of the base system. $\Sigma^r(\mathbf{q}, \omega) = \Sigma^a(\mathbf{q}, \omega) = ie\mathbf{v} \cdot \frac{\mathcal{E}}{\omega} e^{j(\mathbf{q}\cdot\mathbf{r}+\omega t)}$ is the retarded/advanced self-energy, where $\mathbf{v}$ is the velocity matrix, while $\mathcal{E}$ is the electric field with wavevector $\mathbf{q}$ and frequency ω. Equations (1, 6) provide a systematic and convenient approach for computing the optical responses to an arbitrary order: for the N-th order optical response, one simply picks up all terms with $m + n = N$ in Eq. (6). The $\mathbf{k}$ and E arguments are omitted in Eq. (6) for simplicity. Note that after each self-energy term $\Sigma^{r/a}(\mathbf{q}, \omega)$, the arguments $\mathbf{k}$ and E should be shifted by $\mathbf{q}$ and ω, respectively, corresponding to the momentum and energy conservation. When the base system is non-interacting, Eqs. (1) and (6) can be reduced to the common single-particle formula. This is verified theoretically and computationally in the Supplementary Note 1.3 and 1.4.

Ab initio calculations

The ab initio calculations are based on density functional theory (DFT)[72,73] as implemented in the Vienna ab initio simulation package (VASP)[74,75]. Generalized gradient approximation (GGA) in the form of Perdew-Burke-Ernzerhof (PBE)[76] is used to treat the exchange-correlation interactions. Core and valence electrons are treated by projector augmented wave (PAW) method[77] and plane-wave basis functions, respectively. For DFT calculations of T_d-WTe$_2$, the first Brillouin zone is sampled by a $15 \times 9 \times 3$ $\mathbf{k}$-mesh. Then a tight-binding (TB) Hamiltonian is built from DFT results using the Wannier90 package[78] and is used to calculate the NLO responses within the Green's function framework. For the localized NLO responses on each PL, the BZ integration in Eq. (1) is carried out by $\mathbf{k}$-mesh sampling with $\int \frac{d\mathbf{k}}{(2\pi)^2} = \frac{1}{S}\sum_{\mathbf{k}} w_{\mathbf{k}}$, where S is the area of the 2D unit cell on each PL and $w_{\mathbf{k}}$ is the weight factor. For the NLO conductivity of common bulk materials, the $\mathbf{k}$-mesh sampling should be $\int \frac{d\mathbf{k}}{(2\pi)^3} = \frac{1}{V}\sum_{\mathbf{k}} w_{\mathbf{k}}$, where V is the volume of the 3D unit cell. Thus, the conductivities shown in this work differ from the common definition of conductivities by a factor L, which is the thickness of the unit cell. For the $\mathbf{k}$-integration in Eq. (1), a $\mathbf{k}$-mesh of 128×64 is used for T_d-WTe$_2$, while for the E-integration, a trapezoidal method is used with an energy interval of 10 meV. The convergence in both $\mathbf{k}$ and E integrations is tested.

DATA AVAILABILITY

The authors declare that the main data supporting the findings of this study are available within the article and its Supplementary Information files.

CODE AVAILABILITY

The data that support the findings within this paper and the MATLAB code for calculating the shift and circular current conductivity are available from the corresponding authors upon reasonable request.

REFERENCES

1. Ma, J. et al. Nonlinear photoresponse of type-II Weyl semimetals. *Nat. Mater.* **18**, 476–481 (2019).
2. Xu, H. et al. Colossal switchable photocurrents in topological Janus transition metal dichalcogenides. *npj Comput. Mater.* **7**, 1–9 (2021).
3. Osterhoudt, G. B. et al. Colossal mid-infrared bulk photovoltaic effect in a type-I Weyl semimetal. *Nat. Mater.* **18**, 471–475 (2019).
4. Zhang, Y. & Fu, L. Terahertz detection based on nonlinear Hall effect without magnetic field. *Proc. Natl. Acad. Sci.* **118** (2021).
5. Qin, M., Yao, K. & Liang, Y. C. High efficient photovoltaics in nanoscaled ferroelectric thin films. *Appl. Phys. Lett.* **93**, 122904 (2008).
6. Choi, T., Lee, S., Choi, Y. J., Kiryukhin, V. & Cheong, S. W. Switchable ferroelectric diode and photovoltaic effect in BiFeO3. *Science* **324**, 63–66 (2009).
7. Yang, S. Y. et al. Above-bandgap voltages from ferroelectric photovoltaic devices. *Nat. Nanotechnol.* **5**, 143–147 (2010).
8. Daranciang, D. et al. Ultrafast photovoltaic response in ferroelectric nanolayers. *Phys. Rev. Lett.* **108**, 087601 (2012).
9. Grinberg, I. et al. Perovskite oxides for visible-light-absorbing ferroelectric and photovoltaic materials. *Nature* **503**, 509–512 (2013).
10. Bhatnagar, A., Roy Chaudhuri, A., Heon Kim, Y., Hesse, D. & Alexe, M. Role of domain walls in the abnormal photovoltaic effect in BiFeO 3. *Nat. Commun.* **4**, 1–8 (2013).
11. Xu, S.-Y. et al. Electrically switchable Berry curvature dipole in the monolayer topological insulator WTe2. *Nat. Phys.* **14**, 900–906 (2018).
12. Ma, Q. et al. Direct optical detection of Weyl fermion chirality in a topological semimetal. *Nat. Phys.* **13**, 842–847 (2017).
13. Xu, S.-Y. et al. Spontaneous gyrotropic electronic order in a transition-metal dichalcogenide. *Nature* **578**, 545–549 (2020).
14. Kaplan, D., Holder, T. & Yan, B. Nonvanishing subgap photocurrent as a probe of lifetime effects. *Phys. Rev. Lett.* **125**, 227401 (2020).
15. Ahn, J., Guo, G.-Y. & Nagaosa, N. Low-frequency divergence and quantum geometry of the bulk photovoltaic effect in topological semimetals. *Phys. Rev. X* **10**, 041041 (2020).
16. Holder, T., Kaplan, D. & Yan, B. Consequences of time-reversal-symmetry breaking in the light-matter interaction: berry curvature, quantum metric, and diabatic motion. *Phys. Rev. Res* **2**, 033100 (2020).
17. Watanabe, H. & Yanase, Y. Chiral photocurrent in parity-violating magnet and enhanced response in topological antiferromagnet. *Phys. Rev. X* **11**, 011001 (2021).
18. Liu, J., Xia, F., Xiao, D., García de Abajo, F. J. & Sun, D. Semimetals for high-performance photodetection. *Nat. Mater.* **19**, 830–837 (2020).
19. Ma, Q., Grushin, A. G. & Burch, K. S. Topology and geometry under the nonlinear electromagnetic spotlight. *Nat. Mater.* **20**, 1601–1614 (2021).
20. Tan, L. Z. & Rappe, A. M. Enhancement of the bulk photovoltaic effect in topological insulators. *Phys. Rev. Lett.* **116**, 237402 (2016).
21. Xu, H., Zhou, J., Wang, H. & Li, J. Giant photonic response of Mexican-Hat topological semiconductors for mid-infrared to terahertz applications. *J. Phys. Chem. Lett.* **11**, 6119–6126 (2020).
22. De Juan, F., Grushin, A. G., Morimoto, T. & Moore, J. E. Quantized circular photogalvanic effect in Weyl semimetals. *Nat. Commun.* **8**, 1–7 (2017).
23. Chan, C.-K., Lindner, N. H., Refael, G. & Lee, P. A. Photocurrents in Weyl semimetals. *Phys. Rev. B* **95**, 041104 (2017).
24. Ni, Z. et al. Linear and nonlinear optical responses in the chiral multifold semimetal RhSi. *npj Quantum Mater.* **5**, 96 (2020).
25. Ni, Z. et al. Giant topological longitudinal circular photo-galvanic effect in the chiral multifold semimetal CoSi. *Nat. Commun.* **12**, 154 (2021).
26. Morimoto, T. & Nagaosa, N. Topological nature of nonlinear optical effects in solids. *Sci. Adv.* **2**, e1501524 (2016).
27. Fei, R., Song, W. & Yang, L. Giant linearly-polarized photogalvanic effect and second harmonic generation in two-dimensional axion insulators. *Phys. Rev. B* **102**, 035440 (2020).
28. Wu, L. et al. Giant anisotropic nonlinear optical response in transition metal monopnictide Weyl semimetals. *Nat. Phys.* **13**, 350–355 (2017).
29. Xu, Q. et al. Comprehensive scan for nonmagnetic Weyl semimetals with nonlinear optical response. *npj Comput. Mater.* **6**, 1–7 (2020).
30. Hosur, P. Circular photogalvanic effect on topological insulator surfaces: berry-curvature-dependent response. *Phys. Rev. B* **83**, 035309 (2011).
31. Braun, L. et al. Ultrafast photocurrents at the surface of the three-dimensional topological insulator Bi2Se3. *Nat. Commun.* **7**, 13259 (2016).
32. Chang, G. et al. Unconventional photocurrents from surface fermi arcs in topological chiral semimetals. *Phys. Rev. Lett.* **124**, 166404 (2020).
33. Kim, K. W., Morimoto, T. & Nagaosa, N. Shift charge and spin photocurrents in Dirac surface states of topological insulator. *Phys. Rev. B* **95**, 035134 (2017).
34. Giorgianni, F. et al. Strong nonlinear terahertz response induced by Dirac surface states in Bi2Se3 topological insulator. *Nat. Commun.* **7**, 11421 (2016).
35. Hsieh, D. et al. Nonlinear optical probe of tunable surface electrons on a topological insulator. *Phys. Rev. Lett.* **106**, 057401 (2011).
36. Junck, A., Refael, G. & Von Oppen, F. Photocurrent response of topological insulator surface states. *Phys. Rev. B* **88**, 075144 (2013).
37. Dai, Z., Schankler, A. M., Gao, L., Tan, L. Z. & Rappe, A. M. Phonon-assisted ballistic current from first-principles calculations. *Phys. Rev. Lett.* **126**, 177403 (2021).
38. Ishizuka, H. & Nagaosa, N. Theory of bulk photovoltaic effect in Anderson insulator. *Proc. Natl. Acad. Sci* **118**, 2021 (2021).
39. Chan, Y.-H., Qiu, D. Y., Jornada, F. Hda & Louie, S. G. Giant exciton-enhanced shift currents and direct current conduction with subbandgap photo excitations produced by many-electron interactions. *Proc. Natl Acad. Sci.* **118**, 2021 (2021).

40. Trolle, M. L., Seifert, G. & Pedersen, T. G. Theory of excitonic second-harmonic generation in monolayer <span class. *Phys. Rev. B* **89**, 235410 (2014).

41. Kaneko, T., Sun, Z., Murakami, Y., Golež, D. & Millis, A. J. Bulk photovoltaic effect driven by collective excitations in a correlated insulator. *Phys. Rev. Lett.* **127**, 127402 (2020).

42. Sancho, M. P. L., Sancho, J. M. L. & Rubio, J. Quick iterative scheme for the calculation of transfer matrices: application to Mo (100). *J. Phys. F. Met. Phys.* **14**, 1205 (1984).

43. Sancho, M. P. L., Sancho, J. M. L., Sancho, J. M. L. & Rubio, J. Highly convergent schemes for the calculation of bulk and surface Green functions. *J. Phys. F. Met. Phys.* **15**, 851 (1985).

44. Soluyanov, A. A. et al. Type-II Weyl semimetals. *Nature* **527**, 495–498 (2015).

45. Li, P. et al. Evidence for topological type-II Weyl semimetal WTe2. *Nat. Commun.* **8**, 2150 (2017).

46. Xu, H., Wang, H., Zhou, J. & Li, J. Pure spin photocurrent in non-centrosymmetric crystals: bulk spin photovoltaic effect. *Nat. Commun.* **12**, 4330 (2021).

47. Mu, X., Pan, Y. & Zhou, J. Pure bulk orbital and spin photocurrent in two-dimensional ferroelectric materials. *npj Comput. Mater.* **7**, 61 (2021).

48. Mahan, G. D. *Many-Particle Physics.* (Springer US, 2000).

49. Wang, Q. et al. Room-temperature nanoseconds spin relaxation in WTe2 and MoTe2 thin films. *Adv. Sci.* **5**, 1700912 (2018).

50. Liang, T. et al. Ultrahigh mobility and giant magnetoresistance in the Dirac semimetal Cd3As2. *Nat. Mater.* **14**, 280–284 (2015).

51. Shekhar, C. et al. Extremely large magnetoresistance and ultrahigh mobility in the topological Weyl semimetal candidate NbP. *Nat. Phys.* **11**, 645–649 (2015).

52. Wang, H., Zhang, C. & Rana, F. Surface recombination limited lifetimes of photoexcited carriers in few-layer transition metal dichalcogenide MoS2. *Nano Lett.* **15**, 8204–8210 (2015).

53. Niehues, I. et al. Strain control of exciton-phonon coupling in atomically thin semiconductors. *Nano Lett.* **18**, 1751–1757 (2018).

54. Georges, A., Kotliar, G., Krauth, W. & Rozenberg, M. J. Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions. *Rev. Mod. Phys.* **68**, 13 (1996).

55. Kotliar, G. et al. Electronic structure calculations with dynamical mean-field theory. *Rev. Mod. Phys.* **78**, 865–951 (2006).

56. Singh, V. et al. DMFTwDFT: an open-source code combining Dynamical Mean Field Theory with various density functional theory packages. *Comput. Phys. Commun.* **261**, 107778 (2021).

57. Allen, L., Beijersbergen, M. W., Spreeuw, R. J. C. & Woerdman, J. P. Orbital angular momentum of light and the transformation of Laguerre-Gaussian laser modes. *Phys. Rev. A* **45**, 8185 (1992).

58. Lezec, H. J. et al. Beaming light from a subwavelength aperture. *Science* **297**, 820–822 (2002).

59. Barnes, W. L., Dereux, A. & Ebbesen, T. W. Surface plasmon subwavelength optics. *Nature* **424**, 824–830 (2003).

60. Ji, Z. et al. Spatially dispersive circular photogalvanic effect in a Weyl semimetal. *Science* **18**, 955–962 (2019).

61. Feng, J., Qian, X., Huang, C.-W. & Li, J. Strain-engineered artificial atom as a broad-spectrum solar energy funnel. *Nat. Photonics* **6**, 866–872 (2012).

62. Hafez, H. A. et al. Extremely efficient terahertz high-harmonic generation in graphene by hot Dirac fermions. *Nature* **561**, 507–511 (2018).

63. Tancogne-Dejean, N. & Rubio, A. Atomic-like high-harmonic generation from two-dimensional materials. *Sci. Adv.* **4**, eaao5207 (2018).

64. Fregoso, B. M., Muniz, R. A. & Sipe, J. E. Jerk current: a novel bulk photovoltaic effect. *Phys. Rev. Lett.* **121**, 176604 (2018).

65. Cheng, B. et al. Efficient terahertz harmonic generation with coherent acceleration of electrons in the dirac semimetal <math xmlns. *Phys. Rev. Lett.* **124**, 117402 (2020).

66. Atanasov, R., Haché, A., Hughes, J. L. P., Driel, H. Mvan & Sipe, J. E. Coherent control of photocurrent generation in bulk semiconductors. *Phys. Rev. Lett.* **76**, 1703 (1996).

67. Haché, A. et al. Observation of coherently controlled photocurrent in unbiased, Bulk GaAs. *Phys. Rev. Lett.* **78**, 306 (1997).

68. Stevens, M. J. et al. Quantum interference control of ballistic pure spin currents in semiconductors. *Phys. Rev. Lett* **90**, 136603 (2003).

69. Zhao, H., Loren, E. J., van Driel, H. M. & Smirl, A. L. Coherence control of hall charge and spin currents. *Phys. Rev. Lett.* **96**, 246601 (2006).

70. Parker, D. E., Morimoto, T., Orenstein, J. & Moore, J. E. Diagrammatic approach to nonlinear optical response with application to Weyl semimetals. *Phys. Rev. B* **99**, 045121 (2019).

71. João, S. M. & Lopes, J. M. V. P. Basis-independent spectral methods for nonlinear optical response in arbitrary tight-binding models. *J. Phys. Condens. Matter* **32**, 125901 (2019).

72. Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **136**, B864–B871 (1964).

73. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133–A1138 (1965).

74. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).

75. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).

76. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).

77. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).

78. Mostofi, A. A. et al. An updated version of wannier90: a tool for obtaining maximally-localised Wannier functions. *Comput. Phys. Commun.* **185**, 2309–2310 (2014).

AUTHOR CONTRIBUTIONS

H.X. and J.L. conceived the idea and designed the project. H.X. derived the theories. H.X. performed the calculations and wrote the paper with the help of H.W. and J.L. J.L. supervised the project. All authors analyzed the data and contributed to the discussions of the results.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-022-00782-y.

Correspondence and requests for materials should be addressed to Ju Li.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Materials

to

Abnormal Nonlinear Optical Responses on the Surface of Topological Materials

Haowei Xu[1], Hua Wang[1], and Ju Li[1,2]

[1]Department of Nuclear Science and Engineering, Massachusetts Institute of Technology,

Cambridge, Massachusetts 02139, USA

[1]Department of Material Science and Engineering, Massachusetts Institute of Technology,

Cambridge, Massachusetts 02139, USA

Supplementary Note 1   Green's function: general framework

In this section we demonstrate the general framework of using the Green's functions to calculate linear and nonlinear optical responses. The derivation is based on Dyson's equation and is equivalent to the diagrammatic approach based on Feynman's diagrams. We also theoretically and computationally show that the Green's function formalism can be reduced to the widely-used single-particle formulae in non-interacting systems.

Supplementary Note 1.1   From the density matrix to the lesser Green's function

The first issue is how we can get the thermal and quantum average of a certain observable j in equilibrium (or in steady state) from Green's functions. In the language of density matrix, one has the

average as

$$\langle\theta\rangle = \int [\mathrm{d}k]\mathrm{Tr}\,\{\rho_k\theta\} \tag{1}$$

where ρ is the density matrix. Within the Green's function formalism, the lesser Green's function $G^<(k, E)$ plays a similar role to ρ, and one has

$$\langle\theta\rangle = -\mathrm{i}\int [\mathrm{d}k]\frac{\mathrm{d}E}{2\pi}\mathrm{Tr}\,\{\theta G^<(k, E)\} \tag{2}$$

One can see that $G^<(k, E)$ is something like an "energy resolved" distribution function. This point can be understood from the definition of $G^<$, which is

$$G^<(x_1, x_2) = \mathrm{i}\langle\psi^\dagger(x_2)\psi(x_1)\rangle \tag{3}$$

where $x = (r, t)$ and ψ is the field operators. This is somewhat similar to the definition of the density matrix, which is $\rho_{mn} = \langle c_m^\dagger c_n\rangle$. One can also look at the explicit formula for $G^<(k, E)$

$$\begin{aligned}
G_0^<(k, E) &= 2\pi\mathrm{i}\rho_k\delta(E - H_k), & \text{non-interacting system} \\
G^<(k, E) &= \mathrm{i}n_{\mathrm{FD}}(E)A(k, E), & \text{interacting system}
\end{aligned} \tag{4}$$

In the non-interacting case, ρ_k is the non-interacting density matrix ($\rho_{cc} = 0$ and $\rho_{vv} = 1$ for unoccupied and occupied bands, respectively), while H_k is the Hamiltonian. Thanks to the delta function, the integration over E in Eq. (2) directly leads to Eq. (1). In the interacting case, n_{FD} is the Fermi-Dirac distribution function, while $A(k, E)$ is the spectral function, which is essentially the probability that the electron is "in state k and with energy E" Note that in multi-band systems, ρ_k, H_k and $A(k, E)$ are all matrices.

Supplementary Note 1.2   Perturbative solution to Dyson's equation

The next step is, of course, to obtain the lesser Green's function $G^<$ under external fields. This can be achieved with Dyson's equation. Relevant to the responses under light illumination, one needs

$$\begin{aligned}
G^\mathrm{r} &= G_0^\mathrm{r}[1 + \Sigma^\mathrm{r}G^\mathrm{r}] \tag{5} \\
G^\mathrm{a} &= G_0^\mathrm{a}[1 + \Sigma^\mathrm{a}G^\mathrm{a}] \tag{6} \\
G^< &= (1 + G^\mathrm{r}\Sigma^\mathrm{r})G_0^<(1 + \Sigma^\mathrm{a}G^\mathrm{a}) + G^\mathrm{r}\Sigma^<G^\mathrm{a} \tag{7}
\end{aligned}$$

where G^{r} and G^{a} are retarded and advanced Green's function, respectively. Σ^{r}, Σ^{a} and $\Sigma^{<}$ are the retarded, advanced, and lesser self-energy, respectively. The self-energies come from the interaction with the light field and can be treated as perturbations here. Quantities with subscript 0 indicates their values in the "base" system under no light illumination. We assume that we have full knowledge about these base systems. As discussed in the main text, there can be interactions in the base system, but these interactions are handled before the interaction described by Σ is included. For example, the interaction with impurities, phonons, etc. can be included in G_0. G_0 can be obtained perturbatively with e.g., Feynman's diagram, or some other non-perturbative approaches. In the following we will show how strong-strong electron correlation can be incorporated in G_0 as an example.

Pertinent to the problem of optical responses, the electron interaction with electric field can be represented with $\Sigma^{\mathrm{r}} = \Sigma^{\mathrm{a}} = -ev \cdot A$ and $\Sigma^{<} = 0$,[1] where v is the velocity matrix, while A is the vector potential. Then one can perturbatively solve Eqs. (5-7), yielding

$$G^{\mathrm{r}} = G_0^{\mathrm{r}} \sum_{n=0}^{\infty} (\Sigma^{\mathrm{r}} G_0^{\mathrm{r}})^n \tag{8}$$

$$G^{\mathrm{a}} = G_0^{\mathrm{a}} \sum_{m=0}^{\infty} (\Sigma^{\mathrm{a}} G_0^{\mathrm{a}})^m \tag{9}$$

$$G^{<} = \left[\sum_{n=0}^{\infty} (G_0^{\mathrm{r}} \Sigma^{\mathrm{r}})^n \right] G_0^{<} \left[\sum_{m=0}^{\infty} (\Sigma^{\mathrm{a}} G_0^{\mathrm{a}})^m \right] \tag{10}$$

To get the N-th order response, one just need to pick up terms with $m + n = N$ in Eq. (10), and the first order is

$$G_1^{<} = G_0^{\mathrm{r}} \Sigma^{\mathrm{r}} G_0^{<} + G_0^{<} \Sigma^{\mathrm{a}} G_0^{\mathrm{a}} \tag{11}$$

the second order

$$G_2^{<} = G_0^{\mathrm{r}} \Sigma^{\mathrm{r}} G_0^{\mathrm{r}} \Sigma^{\mathrm{r}} G_0^{<} + G_0^{\mathrm{r}} \Sigma^{\mathrm{r}} G_0^{<} \Sigma^{\mathrm{a}} G_0^{\mathrm{a}} + G_0^{<} \Sigma^{\mathrm{a}} G_0^{\mathrm{a}} \Sigma^{\mathrm{a}} G_0^{\mathrm{a}} \tag{12}$$

and the third order

$$G_3^{<} = G_0^{\mathrm{r}} \Sigma^{\mathrm{r}} G_0^{\mathrm{r}} \Sigma^{\mathrm{r}} G_0^{\mathrm{r}} \Sigma^{\mathrm{r}} G_0^{<} + G_0^{\mathrm{r}} \Sigma^{\mathrm{r}} G_0^{\mathrm{r}} \Sigma^{\mathrm{r}} G_0^{<} \Sigma^{\mathrm{a}} G_0^{\mathrm{a}}$$
$$+ G_0^{\mathrm{r}} \Sigma^{\mathrm{r}} G_0^{<} \Sigma^{\mathrm{a}} G_0^{\mathrm{a}} \Sigma^{\mathrm{a}} G_0^{\mathrm{a}} + G_0^{<} \Sigma^{\mathrm{a}} G_0^{\mathrm{a}} \Sigma^{\mathrm{a}} G_0^{\mathrm{a}} \Sigma^{\mathrm{a}} G_0^{\mathrm{a}} \tag{13}$$

Even higher order responses can be obtained similarly.

Here one needs to be careful with the interpretation of the multiplications of Green's functions. If

---

[1] Here we assume that all photons are external, provided by external sources. The photons do not have their own dynamics, so their propagators are simply 1. They interact with the electrons through $-ev \cdot A$. Since v is a Hermitian matrix with $v = v^{\dagger}$, one has $\Sigma^{<} \propto \Sigma^{\mathrm{r}} - \Sigma^{\mathrm{a}} = \Sigma^{\mathrm{r}} - [\Sigma^{\mathrm{r}}]^{\dagger} = 0$. If we only consider $-ev \cdot A$, then $\Sigma^{\mathrm{r}} = \Sigma^{\mathrm{a}} = -ev \cdot A$ is the only contribution in the self-energy. There are no other ways to draw inequivalent Feynman diagrams that contribute to the self-energy, since photons have to be external, while electrons have to be internal in the current framework. In e.g., a tight binding model, the photons can couple with electrons through high-order terms, such as $\frac{\partial^2 H}{\partial k_a \partial k_b} A_a A_b$, which arise from high-order band dispersion, and may also contribute to the self-energy. Here we do not consider these contributions, but it is straightforward to include these terms.

one uses real space and real time domain, then $C = AB$ (A, B, C are Green's functions) actually means

$$C(r_1 - r_3, t_1 - t_3) \equiv \int \mathrm{d}t_2 \mathrm{d}r_2 A(r_1 - r_2, t_1 - t_2) B(r_2 - r_3, t_2 - t_3) \tag{14}$$

which is somewhat like an "inner product" between A and B. If one transform to reciprocal k and frequency ω domain, then $C = AB$ is simply matrix multiplications without any integrations any more (convolution theorem). The k arguments of all Green's function are the same as long as we assume a uniform/homegeneous system. However, there can be a subtle shift in frequency ω arguments, if the perturbation $\Sigma^{\mathrm{r}} = \Sigma^{\mathrm{a}}$ is oscillating in time.

The frequency shift can be illustrated in real time domain. Here we take $S = A\Sigma^1 B\Sigma^2 C$ as an example. We assume that

$$\Sigma^1(t - t') = \Sigma(\Omega_1)e^{\mathrm{i}\Omega_1 t}\delta(t - t'),$$

and

$$A(t - t') = \int \mathrm{d}\omega A(\omega)e^{-\mathrm{i}\omega(t - t')}.$$

Then $S = A\Sigma^1 B\Sigma^2 C$ really means

$$S(t - t') = \int \mathrm{d}t_1 \mathrm{d}t_2 \mathrm{d}t_3 \mathrm{d}t_4 A(t - t_1)\Sigma^1(t_1)\delta(t_1 - t_2)B(t_2 - t_3)\Sigma^2(t_3)\delta(t_3 - t_4)C(t_4 - t') \tag{15}$$

Due to the time-translation symmetry, we can set $t' = 0$ above. The integration over t_2 and t_4 can be easily carried out with the delta functions. Then

$$S(E) = \int \mathrm{d}t e^{\mathrm{i}Et} S(t)$$

$$= \int \mathrm{d}t e^{\mathrm{i}Et} \int \mathrm{d}t_1 \mathrm{d}t_3 A(t - t_1)\Sigma^1(t_1)B(t_1 - t_3)\Sigma^2(t_3)C(t_3)$$

$$= \int \mathrm{d}t \mathrm{d}t_1 \mathrm{d}t_3 e^{\mathrm{i}Et} \int \mathrm{d}\omega_A A(\omega_A)e^{-\mathrm{i}\omega_A(t-t_1)}\Sigma^1(\Omega_1)e^{\mathrm{i}\Omega_1 t_1} \int \mathrm{d}\omega_B B(\omega_B)e^{-\mathrm{i}\omega_B(t_1-t_3)}\Sigma^2(\Omega_2)e^{\mathrm{i}\Omega_2 t_3} \int \mathrm{d}\omega_C C(\omega_C)e^{-\mathrm{i}\omega_C t_3}$$

$$= \int \mathrm{d}\omega_A \mathrm{d}\omega_B \mathrm{d}\omega_B A(\omega_A)\Sigma^1(\Omega_1)B(\omega_B)\Sigma^2(\Omega_2)C(\omega_C) \int \mathrm{d}t e^{\mathrm{i}(E-\omega_A)t} \int \mathrm{d}t_1 e^{\mathrm{i}(\omega_A + \Omega_1 - \omega_B)t_1} \int \mathrm{d}t_3 e^{\mathrm{i}(\omega_B + \Omega_2 - \omega_C)t_1}$$

$$= \int \mathrm{d}\omega_A \mathrm{d}\omega_B \mathrm{d}\omega_B A(\omega_A)\Sigma^1(\Omega_1)B(\omega_B)\Sigma^2(\Omega_2)C(\omega_C)\delta(E - \omega_A)\delta(\omega_A + \Omega_1 - \omega_B)\delta(\omega_B + \Omega_2 - \omega_C)$$

$$= A(E)\Sigma^1(\Omega_1)B(E + \Omega_1)\Sigma^2(\Omega_2)C(E + \Omega_1 + \Omega_2)$$

$$\tag{16}$$

One can see that in the frequency domain, the argument of the Green's function should shift by Ω after each $\Sigma(\Omega)$. This point is actually obvious if one looks at the Feynman diagram: the energy and momentum should be conserved at each vertex, which leads to the shift in energy and momentum. Note that if the small but finite momentum of photons is taken into consideration, then the k arguments should shift as well.

Supplementary Note 1.3   Responses functions in non-interacting systems

In this section, we deal with the simplest system, where the Hamiltonian H_k of the base system can be diagonalized, with eigenstates $|nk\rangle$ and eigenenergy E_{nk}. Both the Hamiltonian and the Green's functions are diagonal in the basis of $|nk\rangle$. The interaction with external electric field is treated as perturbations. The Green's function approach can be reduced to the widely used single particle formulae.

In this non-interacting case, one has non-interacting Green's functions as[2]

$$
[G_0^{\mathrm{r}}(\omega)]_{mn} = \frac{\delta_{mn}}{\omega - E_n + \mathrm{i}\delta} \tag{17}
$$

$$
[G_0^{\mathrm{a}}(\omega)]_{mn} = \frac{\delta_{mn}}{\omega - E_n - \mathrm{i}\delta} \tag{18}
$$

$$
[G_0^{<}(\omega)]_{mn} = 2\pi \mathrm{i}\delta_{mn} f_n \delta(\omega - E_n) \tag{19}
$$

where f_n is occupation number of band n.

Supplementary Note 1.3.1   Linear responses

For the linear response, Eq. (11) leads to

$$
\langle \theta \rangle_1 = -\mathrm{i} \int [\mathrm{d}k] \frac{\mathrm{d}E}{2\pi} \mathrm{Tr} \left\{ \theta \left(G_0^{\mathrm{r}} \Sigma^{\mathrm{r}} G_0^{<} + G_0^{<} \Sigma^{\mathrm{a}} G_0^{\mathrm{a}} \right) \right\}
$$

$$
= \int [\mathrm{d}k] \mathrm{d}E \sum_{mn} \left\{ \theta_{mn} \frac{1}{E - E_n + \mathrm{i}\delta} \Sigma_{nm}^1(\Omega_1) f_m \delta(E + \Omega_1 - E_m) + \theta_{mn} f_n \delta(E - E_n) \Sigma_{nm}^1(\Omega_1) \frac{1}{E + \Omega_1 - E_m - \mathrm{i}\delta} \right\}
$$

$$
= \int [\mathrm{d}k] \left\{ f_m \theta_{mn} \frac{1}{E_{mn} - \Omega_1 + \mathrm{i}\delta} \Sigma_{nm}^1(\Omega_1) + f_n \theta_{mn} \Sigma_{nm}^1(\Omega_1) \frac{1}{E_{nm} + \Omega_1 - \mathrm{i}\delta} \right\}
$$

$$
= \int [\mathrm{d}k] \frac{f_{mn} \theta_{mn} \Sigma_{nm}^1(\Omega_1)}{E_{mn} - \Omega_1 + \mathrm{i}\delta} \tag{20}
$$

where $f_{mn} = f_m - f_n$, $E_{mn} = E_m - E_n$, and $\Sigma_{mn}^1(\Omega_1) = -\frac{e}{\mathrm{i}\Omega_1} v_{mn} \cdot \mathcal{E}$ is the self-energy in velocity gauge. This is exactly what we obtained before from linear response theory.

---

[2]The k indices are omitted.

Supplementary Note 1.3.2   Second-order response

From Eq. (12) one can see that there are three terms contributing to the second-order response,

$$
\begin{aligned}
\langle\theta\rangle_2 = -\mathrm{i}\int[\mathrm{d}k]\frac{\mathrm{d}E}{2\pi}\mathrm{Tr}\{&\theta[G_0^{\mathrm{r}}(E)\Sigma^{\mathrm{r}}(\Omega_1)G_0^{\mathrm{r}}(E+\Omega_1)\Sigma^{\mathrm{r}}(\Omega_2)G_0^{<}(E+\Omega_1+\Omega_2)\\
&+G_0^{<}(E)\Sigma^{\mathrm{a}}(\Omega_1)G_0^{\mathrm{a}}(E+\Omega_1)\Sigma^{\mathrm{a}}(\Omega_2)G_0^{\mathrm{a}}(E+\Omega_1+\Omega_2)\\
&+G_0^{\mathrm{r}}(E)\Sigma^{\mathrm{r}}(\Omega_1)G_0^{<}(E+\Omega_1)\Sigma^{\mathrm{a}}(\Omega_2)G_0^{\mathrm{a}}(E+\Omega_1+\Omega_2)]\}
\end{aligned}
\tag{21}
$$

The evaluations of these three terms are similar, and we take the first term as an example.

$$
\begin{aligned}
\langle\theta\rangle_2^{\mathrm{(I)}} &= -\mathrm{i}\int[\mathrm{d}k]\frac{\mathrm{d}E}{2\pi}\mathrm{Tr}\left\{\theta G_0^{\mathrm{r}}(E)\Sigma^{\mathrm{r}}(\Omega_1)G_0^{\mathrm{r}}(E+\Omega_1)\Sigma^{\mathrm{r}}(\Omega_2)G_0^{<}(E+\Omega_1+\Omega_2)\right\}\\
&= -\mathrm{i}\int[\mathrm{d}k]\frac{\mathrm{d}E}{2\pi}\sum_{mnl}\theta_{mn}\frac{1}{E-E_n+\mathrm{i}\delta}\Sigma_{nl}^1(\Omega_1)\frac{1}{E+\Omega_1-E_l+\mathrm{i}\delta}\Sigma_{lm}^2(\Omega_2)2\pi\mathrm{i}f_m\delta(E+\Omega_1+\Omega_2-E_m)\\
&= \int[\mathrm{d}k]\sum_{mnl}\theta_{mn}\Sigma_{nl}^1(\Omega_1)\Sigma_{lm}^2(\Omega_2)\frac{f_m}{(E_{mn}-\Omega_{12}+\mathrm{i}\delta)(E_{ml}-\Omega_2+\mathrm{i}\delta)}
\end{aligned}
\tag{22}
$$

The other two terms can be obtained similarly. For static response, one has $\Omega_1 = -\Omega_2 = \Omega$ and $\Omega_{12} = \Omega_1+\Omega_2 = 0$, and we finally have

$$
\begin{aligned}
\langle\theta\rangle_2 &= \langle\theta\rangle_2^{\mathrm{(I)}} + \langle\theta\rangle_2^{\mathrm{(II)}} + \langle\theta\rangle_2^{\mathrm{(III)}}\\
&= \int[\mathrm{d}k]\sum_{mnl}\theta_{mn}\Sigma_{nl}^1(\Omega_1)\Sigma_{lm}^2(\Omega_2)\\
&\quad\times\left\{\frac{f_m}{(E_{mn}+\mathrm{i}\delta)(E_{ml}+\Omega+\mathrm{i}\delta)}+\frac{f_n}{(E_{nm}-\mathrm{i}\delta)(E_{nl}+\Omega-\mathrm{i}\delta)}+\frac{f_l}{(E_{ln}-\Omega+\mathrm{i}\delta)(E_{lm}-\Omega-\mathrm{i}\delta)}\right\}
\end{aligned}
\tag{23}
$$

Here the third term in the bracket can be rewritten as

$$
\begin{aligned}
\frac{f_l}{(E_{ln}-\Omega+\mathrm{i}\delta)(E_{lm}-\Omega-\mathrm{i}\delta)} &= -\frac{f_l}{E_{ml}+E_{ln}+2\mathrm{i}\delta}\left(\frac{1}{E_{ln}-\Omega+\mathrm{i}\delta}+\frac{1}{E_{ml}+\Omega+\mathrm{i}\delta}\right)\\
&= -\frac{f_l}{E_{mn}+2\mathrm{i}\delta}\left(\frac{1}{E_{ln}-\Omega+\mathrm{i}\delta}+\frac{1}{E_{ml}+\Omega+\mathrm{i}\delta}\right)
\end{aligned}
\tag{24}
$$

If we ignore the difference between $2\mathrm{i}\delta$ and $\mathrm{i}\delta$,[3] then Eq. (23) becomes

$$
\begin{aligned}
\langle\theta\rangle_2 &= \int[\mathrm{d}k]\sum_{mnl}\frac{\theta_{mn}\Sigma_{nl}^1(\Omega_1)\Sigma_{lm}^2(\Omega_2)}{E_{mn}+\mathrm{i}\delta}\\
&\quad\times\left\{\frac{f_m}{E_{ml}+\Omega+\mathrm{i}\delta}-\frac{f_n}{E_{nl}+\Omega-\mathrm{i}\delta}-\frac{f_l}{E_{ln}-\Omega+\mathrm{i}\delta}-\frac{f_l}{E_{ml}+\Omega+\mathrm{i}\delta}\right\}\\
&= \int[\mathrm{d}k]\sum_{mnl}\frac{\theta_{mn}}{E_{mn}+\mathrm{i}\delta}\left\{\frac{f_{ml}\Sigma_{nl}^1(\Omega_1)\Sigma_{lm}^2(\Omega_2)}{E_{ml}+\Omega+\mathrm{i}\delta}-\frac{f_{nl}\Sigma_{nl}^1(\Omega_1)\Sigma_{lm}^2(\Omega_2)}{E_{nl}+\Omega-\mathrm{i}\delta}\right\}
\end{aligned}
\tag{25}
$$

---

[3]This would make a difference at frequencies $\Omega\sim\delta$

The contribution from a similar term with exchanged $\Omega_1 \leftrightarrow \Omega_2$, $\Sigma^1 \leftrightarrow \Sigma^2$ should be added, and one has

$$\langle\theta\rangle_2 = \int [\mathrm{d}k] \sum_{mnl} \frac{\theta_{mn}}{E_{mn}+\mathrm{i}\delta} \left\{ \frac{f_{ml}\Sigma_{nl}^1(\Omega_1)\Sigma_{lm}^2(\Omega_2)}{E_{ml}+\Omega+\mathrm{i}\delta} - \frac{f_{nl}\Sigma_{nl}^1(\Omega_1)\Sigma_{lm}^2(\Omega_2)}{E_{nl}+\Omega-\mathrm{i}\delta} + \frac{f_{ml}\Sigma_{nl}^2(\Omega_2)\Sigma_{lm}^1(\Omega_1)}{E_{ml}-\Omega+\mathrm{i}\delta} - \frac{f_{nl}\Sigma_{nl}^2(\Omega_2)\Sigma_{lm}^1(\Omega_1)}{E_{nl}-\Omega-\mathrm{i}\delta} \right\}$$

$$= \int [\mathrm{d}k] \sum_{mnl} \frac{\theta_{mn}}{E_{mn}+\mathrm{i}\delta} \left\{ \frac{f_{ml}\Sigma_{nl}^2(\Omega_2)\Sigma_{lm}^1(\Omega_1)}{E_{ml}-\Omega+\mathrm{i}\delta} - \frac{f_{ln}\Sigma_{nl}^1(\Omega_1)\Sigma_{lm}^2(\Omega_2)}{E_{ln}-\Omega+\mathrm{i}\delta} + \frac{f_{ml}\Sigma_{nl}^1(\Omega_1)\Sigma_{lm}^2(\Omega_2)}{E_{ml}+\Omega+\mathrm{i}\delta} - \frac{f_{ln}\Sigma_{nl}^2(\Omega_2)\Sigma_{lm}^1(\Omega_1)}{E_{ln}+\Omega+\mathrm{i}\delta} \right\}$$

(26)

On the second line, the first two terms are exactly what we have obtained with the single-particle quadratic response theory, and the second two terms are also obtained with the exchange of $\Omega_1 \leftrightarrow \Omega_2$, $\Sigma^1 \leftrightarrow \Sigma^2$.

Supplementary Note 1.4   Computational benchmarks

In this section we show some computational benchmarks of the Green's function approach described above. Results from the Green's function approach are compared against results from the single-particle formula. We use monolayer MoS_2 as an example. The single MoS_2 atom layer has no neighbors, so the electrons can be regarded as truly "free" with well defined intralayer Hamiltonian H_{00} [4] and $G_0^r(\omega) = \frac{1}{\omega - H_{00} + \mathrm{i}\delta}$. In this case, ideally the Green's function approach and the single-particle approach should give exactly the same result[5], as theoretically proved in Sec. Supplementary Note 1.3. This is numerically verified for the bulk photovoltaic (BPV) conductivities σ_{yy}^y and σ_{xy}^x, which correspond to the shift and the circular current, respectively. The results from the two approaches are shown in Fig. 1, which agree well with each other.

We have also benchmarked our Green's function formalism with calculations in previous works by other researchers. We use GeS monolayer studied in Ref. [1] as an example. The BPV conductivities of GeS monolayer calculated with our Green's function formalism is shown in Figure 2, in satisfactory agreement with Figure 2 in Ref. [1], which was calculated with the single-particle formulae.

---

[4]The detailed definition of H_{00} can be found in Sec. Supplementary Note 3.

[5]except for the smearing factor as described around Eq. (24)
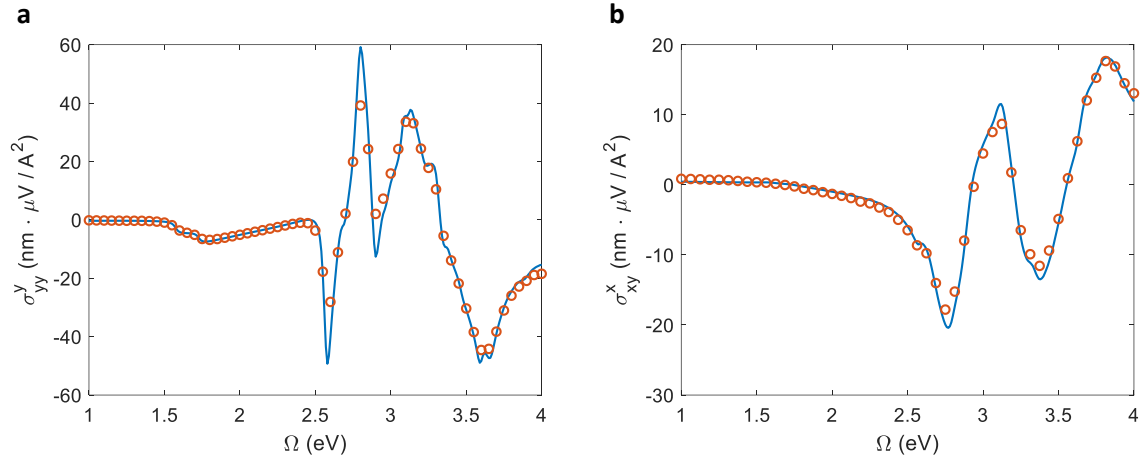
Supplementary Figure 1: (a) BPV conductivity of monolayer MoS_2 under linearly polarized light from Green's function approach (red circles) and single particle formula (blue solid curves). (b) Similar to (a), but under circularly polarized light.
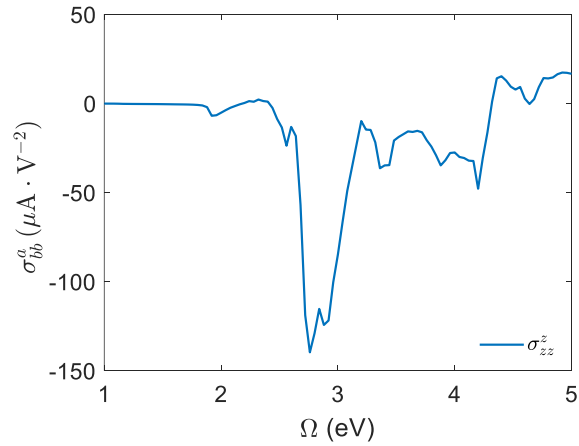
Supplementary Figure 2: BPV conductivity of monolayer GeS, calculated with the Green's function approach. The results are in good agreement with those in Ref. [1] (Figure 2 therein)

Supplementary Note 2   Applications of the Green's function framework

In this section we discuss how the Green's function framework can be applied to a variety of nonlinear optical effects.

Supplementary Note 2.1   Bulk (spin) photovoltaic effect

As discussed in the main text, the bulk (spin) photovoltaic effect [B(S)PV] indicates that under light illumination, a DC charge (spin) current can be generated in non-centrosymmetric materials. Here we first assume that the light field is homogeneous, and is varying in time as $\cos(\omega t + \phi) = \frac{e^{i(\omega t + \phi)} + e^{-i(\omega t + \phi)}}{2}$. The two self-energy term in Eq. (12) are $ie\frac{\mathcal{E}}{\omega}e^{i(\omega t + \phi)}$ and $ie\frac{\mathcal{E}}{-\omega}e^{-i(\omega t - \phi)}$, respectively, so that the current can be static in time.

For BPV effect, the response this $j^a = -ev^a$, which is the charge current, and one as the conductivity as

$$\sigma_{bc}^a(\omega, -\omega) = \frac{\langle j^a \rangle}{\mathcal{E}^b \mathcal{E}^c} = -\frac{ie^3}{\omega^2} \int [dk] \int \frac{dE}{2\pi} \text{Tr}\left\{v^a G^<\right\} \tag{27}$$

with

$$\begin{aligned}
G^< = {} & G_0^r(E)v^b G_0^r(E+\omega)v^c G_0^<(E) + G_0^r(E)v^b G_0^<(E+\omega)v^c G_0^a(E) \\
& + G_0^<(E)v^b G_0^a(E+\omega)v^c G_0^a(E) + (b \leftrightarrow c, +\omega \leftrightarrow -\omega)
\end{aligned} \tag{28}$$

For the spin current, the response should be $j^{a,s^i} = \frac{1}{2}(v^a s^i + s^i v^a)$, where s^i is the spin operator. As discussed in the main text, we divide the BSPV conductivity by $\frac{\hbar}{2e}$, so that it has the same unit as the BPV conductivity. In this case, the BSPV conductivity is

$$\sigma_{bc}^{a,s^i} = \frac{\langle j^{a,s^i} \rangle}{\mathcal{E}^b \mathcal{E}^c} = -\frac{ie^3}{\omega^2} \int [dk] \int \frac{dE}{2\pi} \text{Tr}\left\{\frac{1}{2}(v^a \sigma^i + \sigma^i v^a) G^<\right\} \tag{29}$$

where $G^<$ is the same as that in Eq. (36), while $\sigma^i = \frac{2s^i}{\hbar}$ is the dimensionless spin operator.

Supplementary Note 2.2   Inhomogeneous fields

The Green's function formalism can be easily generalized to deal with inhomogeneous external field characterized by the wavevector q. In this case the self energy should be

$$\Sigma(q,\omega) = \tilde{\Sigma} e^{\mathrm{i}(qr - \omega t)} \tag{30}$$

As a result, the self-energy will lead to a shift not only in the frequency argument, as shown around Eq. (16), but also in the wavevector argument. As an example, one has

$$G^{\mathrm{r}} v G^{\mathrm{r}} v G^{<} \equiv G^{\mathrm{r}}(k,E) v(k,k+q) G^{\mathrm{r}}(k+q,E-\omega) v(k+q,k) G^{<}(k,E) \tag{31}$$

where $[v(k,k+q)]_{mn} = \langle m,k|v|n,k+q\rangle$ is the velocity operator between states with wavevector k and $k+q$. In order to obtain $\langle m,k|v|n,k+q\rangle$, it is more convenient to diagonalize H_{00} and go to the so-called Hamiltonian gauge. In the Hamiltonian gauge, $|mk\rangle$ and $|nk\rangle$ are the eigenstates of H_{00} with eigenvalues E_m and E_n. For small q, one has

$$|m,k+q\rangle = |mk\rangle + q\nabla_k|m,k\rangle,$$

and the derivative of wavefunction $\nabla_k|m,k\rangle$ can be obtained from perturbation theory,

$$\nabla_k|m,k\rangle = \sum_{n\neq m} \frac{\langle nk|\nabla_k H|mk\rangle}{E_m - E_n}|nk\rangle \tag{32}$$

Except for the shift in wavevector argument as shown in Eq. (31), the BPV conductivity under light with wavevector q is the same as that in Eq. (27).

In the main text, we demonstrate that the NLO responses such as BPV conductivities can vary with q in a non-monotonic fashion, and give an intuitive and qualitative explanation. Here we explain this effect in a more quantitative way. To illustrate some typical features more intuitively, we use the single particle wavefunctions $|m,k\rangle$. The optical responses are mainly interband transitions, and the interband transition probability is determined by the transition dipole $v_{mn}(k,q) \equiv \langle m,k|v|n,k+q\rangle$, where v is the velocity matrix. The second-order response is a three-band process and involves three bands m, n, and l. Quantitatively, we need to look at

$$p^{(3)}_{abc}(m,n,l;k;q) \equiv \langle m,k|v_a|n,k\rangle\langle n,k|v_b|l,k+q\rangle\langle l,k+q|v_c|m,k\rangle \tag{33}$$

Here a, b, c can be x, y, z. Generally speaking, $p^{(3)}_{abc}$ determines the interband transition probability.
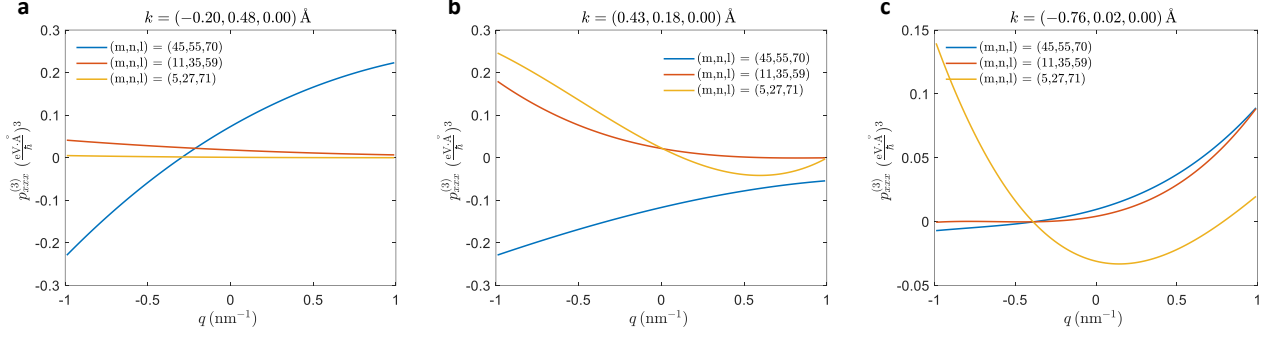
Supplementary Figure 3: $p_{xxx}^{(3)}(m, n, l; k, q)$ for randomly chosen k and (m, n, l) as a function of q. One can see that $p_{xxx}^{(3)}(m, n, l; k; q)$ can vary with q in a non-monotonic fashion, and can change sign.

When $p_{abc}^{(3)}$ is large, an electron with wavevector k can contribute to a larger optical response by doing (virtual) interband transitions between bands (m, n, l). When calculating the conductivity $\sigma_{bc}^{a}(\omega, q)$, one needs to sum $p_{abc}^{(3)}$ over all possible (m, n, l) pairs and all k-points in the first Brillouin zone. This is essentially "summing over all possible electron transitions". $p_{abc}^{(3)}$ with different $(m, n, l; k)$ arguments can have different behavior as a function of q. In Figure 3 we show $p_{xxx}^{(3)}$ of T_d-WTe$_2$ for three randomly chosen (m, n, l) at three randomly chosen k-points. One can see that indeed $p_{xxx}^{(3)}$ can change non-monotonically with q, and can have a sign reversal as q varies. This is a quantitative reason why the BPV conductivities can show a non-monotonic behavior with q with a sign reversal.

Moreover, we can investigate the the scattering of electrons with photons characterized by wavevector q. These scattering are desirable in the NLO processes, since they assist the interband transitions of electrons, which are the origins of the NLO responses. The scattering rates are sensitively dependent on q. In our theory, these scatterings are implicitly but rigorously incorporated by the interband transition matrix $\langle m, k | v | n, k + q \rangle$. Generally speaking, the scatterings with photons are faster when the magnitude of $\langle m, k | v | n, k + q \rangle$ is larger. The electron-photon scattering can be understood in a more intuitive way with Fermi's golden rule in the single particle picture. The coupling between the electron and the photon is described by the Hamiltonian $H = -ev \cdot A = i\frac{e}{\omega} v \cdot E$, where A is the vector potential, E is the electric field, while ω is the frequency of the photon. Leaving the constant factor $i\frac{eE}{\omega}$ aside, one can see the transition rate (or the scattering rate) between two states $|m, k\rangle$ and $|n, k+q\rangle$ is determined by the transition dipole matrix $M_{mn}(k, q) \equiv |\langle m, k | v | n, k + q \rangle|^2$. In Figure 4 we plot $M_{mn}(k, q)$ in the first Brillouin zone for some randomly chosen m, n and q. One can see that $M_{mn}(k, q)$ is dependent on all the four parameters $(m, n; k, q)$. These complexities are fully incorporated in our calculations.
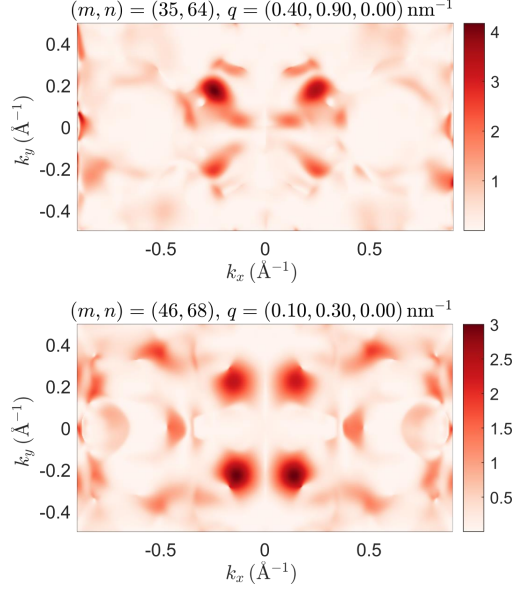
Supplementary Figure 4: Interband transition dipole in the first Brillouin zone, defined as $M_{mn}(k,q) \equiv |\langle m,k|v|n,k+q\rangle|^2$ for randomly selected q points and randomly selected band indices (m,n).

Supplementary Note 2.3   Second-harmonics generation

Previous studies [2] claim that the susceptibility χ of second-harmonic generation (SHG) can be obtained from nonlinear photoconductivity,

$$\chi^a_{bc}(2\omega;\omega,\omega) = \mathrm{i}\frac{\sigma^a_{bc}(2\omega;\omega,\omega)}{2\epsilon_0\omega} \tag{34}$$

where ϵ_0 is the vacuum permittivity, $\sigma^a_{bc}(2\omega;\omega,\omega)$ is the nonlinear conductivity, which can be obtained from the Green's function framework as

$$\sigma^a_{bc}(2\omega;\omega,\omega) = -\frac{ie^3}{\omega^2}\int[dk]\int\frac{dE}{2\pi}\mathrm{Tr}\left\{v^a G^<\right\} \tag{35}$$

with

$$G^< = G^r_0(E)v^b G^r_0(E+\omega)v^c G^<_0(E+2\omega) + G^r_0(E)v^b G^<_0(E+\omega)v^c G^a_0(E+2\omega)$$
$$+ G^<_0(E)v^b G^a_0(E+\omega)v^c G^a_0(E+2\omega) + (b \leftrightarrow c, +\omega \leftrightarrow -\omega) \tag{36}$$

We have calculated χ^y_{yy} for surface and bulk principal layers of $\mathrm{T_d\text{-}WTe_2}$, and the results are shown in Figure 5
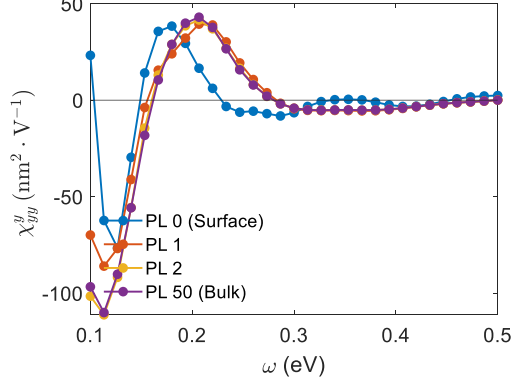
Supplementary Figure 5: Layer-resolved susceptibility of second-harmonics generation of T_d-WTe$_2$

Supplementary Note 2.4   Third order responses

Under three electric fields $\mathcal{E}_b(\omega_b) = \tilde{\mathcal{E}}_b e^{i\omega_b t}$, $\mathcal{E}_c(\omega_c) = \tilde{\mathcal{E}}_c e^{i\omega_c t}$ and $\mathcal{E}_d(\omega_d) = \tilde{\mathcal{E}}_d e^{i\omega_d t}$, a charge current $j^a = \sigma^a_{bcd}\tilde{\mathcal{E}}_b\tilde{\mathcal{E}}_c\tilde{\mathcal{E}}_d e^{i(\omega_b+\omega_c+\omega_c)t}$ can be generated. The conductivity can be calculated with

$$\sigma^a_{bcd} = \frac{\langle j^a \rangle}{\tilde{\mathcal{E}}_b\tilde{\mathcal{E}}_c\tilde{\mathcal{E}}_d} = \frac{e^4}{\omega_b\omega_c\omega_d}\int [dk]\int \frac{dE}{2\pi}\mathrm{Tr}\left\{v^a G^<\right\} \tag{37}$$

where $G^<$ should be obtained according to Eq. (13). Note that one needs to symmetrize over v^b, v^c, v^d and $\omega_b, \omega_c, \omega_d$ in the formula above.

Supplementary Note 2.5   Strong electron-electron correlations

In the main text we mention that the Green's functions in the presence of many-body interactions can be calculated perturbatively with e.g., Feynman's diagrams, or non-perturbatively with other approaches. Here we consider strong electron-electron correlations beyond the density functional theory (DFT) framework as an example, and demonstrate how our Green's function formalism can incorporate many-body interactions.

We artificially add a Hubbard $U = 3\,\mathrm{eV}$ term on the d-orbitals of W atoms, and the Green's functions of this strongly correlated system are calculated with the dynamical mean field theory (DMFT), which is a non-perturbative approach [3, 4]. Note that usually people do not add Hubbard U terms for WTe$_2$. Here we artificially add this Hubbard U term just to showcase the capability of our Green's function formalism.
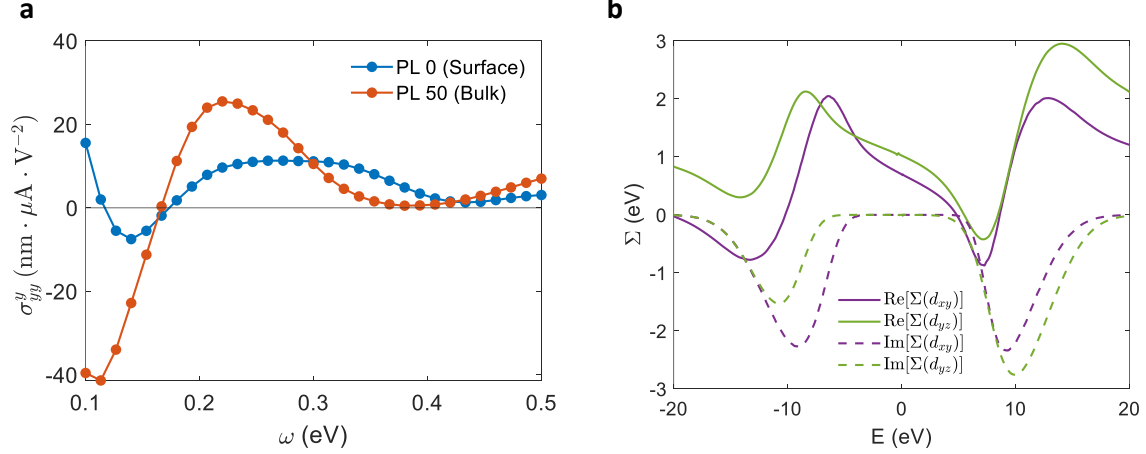
Supplementary Figure 6: (a) BPV conductivities of T_d-WTe$_2$ with an artificial Hubbard interaction of $U = 3\,\text{eV}$ for the d orbitals of W atoms. The red and blue curves are the conductivities for surface and bulk PLs, respectively. (b) The local self-energy of two selected d orbitals of the W atoms on the real frequency axis.

The Hamiltonian of T_d-WTe$_2$ with Hubbard U interaction is

$$H = \sum_{ij,\sigma}(t_{ij}^{\sigma}a_{i,\sigma}^{\dagger}a_{j,\sigma} + h.c.) + U\sum_{d}n_{d\uparrow}n_{d\downarrow} \tag{38}$$

Here $a_{i,\sigma}^{\dagger}$ and $a_{i,\sigma}$ are creation and annihilation operators of orbital i with spin σ. $\sigma = \uparrow$ or $\downarrow$ indicate spin up and down states, respectively. t_{ij}^{σ} is the hopping amplitude, which can be obtained with VASP+Wannier90 calculations (see Methods section in the main text for details). $h.c.$ indicates Hermitian conjugate. $n_{d\sigma}$ is the number operate of the d orbitals of W atoms. The Hubbard U term $U\sum_{d}n_{d\uparrow}n_{d\downarrow}$ introduces strong electron-electron correlations. We treat the Hubbard U term with DMFT, from which we obtain the self-energy and the Green's functions. Finally, the Green's functions are used to calculate the NLO responses.

The self-energy $\Sigma(d)$ of d orbitals of W atoms on the real frequency axis are shown in Figure 6(b). We only show d_{xy} and d_{yz} orbitals for better visualization. The BPV conductivities on the surface and in the bulk of T_d-WTe$_2$ are shown in Figure 6(a). One can see that in the presence of the artificial Hubbard U term, the surface and the bulk still have distinct NLO responses for a wide range of light frequency ω.

Supplementary Note 3   Surface states

In this section we consider the surface states. Different from the non-interacting system described in Sec. Supplementary Note 1.3, the surface electrons are not *free*, as they interact with bulk electrons. In other words, there is no simple single-particle Hamiltonian that can describe the surface states. We treat the surface states interacting with the bulk states as the "base" system, whose Green's functions G_0^r, G_0^a, and $G_0^<$ can be obtained with iterative Green's function method. Then, the optical responses can be calculated with the Green's function framework described above.

Here we need a set of basis functions for the surface states, which can be the Bloch waves built from atomic orbitals localized on the surfaces, as we will show below. Then the surfaces Green's function G, velocity matrix v, and the self-energy $\Sigma = -ev \cdot A$ are all represented in this basis functions set. Note that these basis functions need not to be the energy eigenstates (and in general cases they are not the eigenstates).

Supplementary Note 3.1   Iterative Green's function method

An iterative method for obtaining the surface Green's function was introduced in Refs. [5, 6]. First, we define a principal layer as the smallest group of neighbouring atomic planes such that only nearest-neighbour interactions exist between principal layers (in other words, the principal layer should be thick enough). The surface+bulk system is just a semi-infinite stacking of the principle layers. We label the principle layers by n, and $n = 0$ indicates the surface layer, while $n \to \infty$ indicates bulk layers (Fig. 7). The in-plane wavevector $k_\parallel$ is a good quantum number on the surface. The basis wavefunctions can be built as

$$|\psi_n^\alpha(k_\parallel)\rangle = \frac{1}{\sqrt{N_\parallel}} \sum_{R_\parallel} e^{ik_\parallel \cdot R_\parallel} |\phi_n^\alpha(R_\parallel)\rangle$$
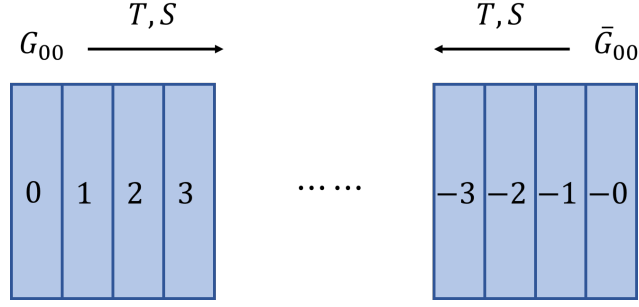
where $N_\parallel$ and $R_\parallel$ are the number and index of unit cells on the surfaces, respectively. $|\phi_n^\alpha(R_\parallel)\rangle$ is the α-th orbital in the $R_\parallel$-th unit cell on the n-the principal layer. In the following we will omit $k_\parallel$.

The the layer-resolved Hamiltonian and Green's functions are

$$[H_{mn}]_{\alpha\beta} = \langle \psi_m^\alpha | H | \psi_n^\beta \rangle \tag{39}$$

$$[G_{mn}(\omega)]_{\alpha\beta} = \langle \psi_m^\alpha | (\omega - H + i\delta)^{-1} | \psi_n^\beta \rangle \tag{40}$$

where H is the Hamiltonian of the surface+bulk system. The iteration for obtaining the surface Green's

Supplementary Figure 7: An illustration of the stacking of the principal layers. The Green's functions on the two surfaces (top and bottom) are denoted by G_{00} and $\bar{G}_{00}$. Transfer matrices T, S and $\bar{T}, \bar{S}$ can be used to obtain Green's function for interior layers, from two opposite directions.

function $G_s \equiv G_{00}$ is

$$\alpha_i = \alpha_{i-1}(\omega - \varepsilon_{i-1})^{-1}\alpha_{i-1}$$

$$\beta_i = \beta_{i-1}(\omega - \varepsilon_{i-1})^{-1}\beta_{i-1}$$

$$\varepsilon_i = \varepsilon_{i-1} + \alpha_{i-1}(\omega - \varepsilon_{i-1})^{-1}\beta_{i-1} + \beta_{i-1}(\omega - \varepsilon_{i-1})^{-1}\alpha_{i-1} \tag{41}$$

$$\varepsilon_i^s = \varepsilon_{i-1}^s + \alpha_{i-1}(\omega - \varepsilon_{i-1})^{-1}\beta_{i-1}$$

$$\bar{\varepsilon}_i^s = \bar{\varepsilon}_{i-1}^s + \beta_{i-1}(\omega - \varepsilon_{i-1})^{-1}\alpha_{i-1}$$

with the initialization $\varepsilon_0 = \bar{\varepsilon}_0 = \bar{\varepsilon}_0^s = H_{00}$, $\alpha_0 = H_{01}$, $\beta_0 = H_{01}^\dagger$. Iteration of Eq. (41) should be converged until $\varepsilon_n \simeq \varepsilon_{n-1}$ $\varepsilon_n^s \simeq \varepsilon_{n-1}^s$, and $\bar{\varepsilon}_n^s \simeq \bar{\varepsilon}_{n-1}^s$. The surface and bulk retarded Green's function $G_s^r(\omega)$ and $G_b^r(\omega)$ can be obtained as

$$G_s^r(\omega) = (\omega - \varepsilon_n^s + i\delta)^{-1} \tag{42}$$

$$\bar{G}_s^r(\omega) = (\omega - \bar{\varepsilon}_n^s + i\delta)^{-1} \tag{43}$$

$$G_b^r(\omega) = (\omega - \varepsilon_n + i\delta)^{-1} \tag{44}$$

Where $\bar{G}_s^r \equiv \bar{G}_{00}$ is the Green's function of the bottom surface. Then the surface spectrum function, advanced and lesser surface Green's functions can be obtained with

$$G^a(k, \omega) = [G^r(k, \omega)]^\dagger$$

$$G^<(k, \omega) = i n_F(\omega) A(k, \omega) \tag{45}$$

$$A(k, \omega) = i [G^r(k, \omega) - G^a(k, \omega)]$$

Supplementary Note 3.2   Green's function for interior layers

In addition to the Green's function for the surface layer (the outmost layer), we are also interested in the Green's function for interior layers, which can be obtained with the transfer matrices

$$T = G_{00}H_{01}^\dagger$$
$$S = H_{01}G_{00}$$

(46)

and the Green's function of other principle layers can be obtained with (Fig. 7)

$$G_{n,0} = G_{00}H_{01}^\dagger G_{n-1,0} = TG_{n-1,0}$$
$$G_{0,n} = G_{0,n-1}H_{01}G_{00} = G_{0,n-1}S$$
$$G_{n,n} = G_{00} + G_{00}H_{01}^\dagger G_{n-1,n-1}H_{01}G_{00} = G_{00} + TG_{n-1,n-1}S$$

(47)

It is also possible to start from the bottom surface, whose Green's function is $\bar{G}_{00} = \bar{G}_s$, and the define the transfer matrix in the reverse direction, which are

$$\bar{T} = \bar{G}_{00}H_{01}$$
$$\bar{S} = H_{01}^\dagger \bar{G}_{00}$$

(48)

and in the reverse direction,

$$\bar{G}_{-n,0} = \bar{G}_{00}H_{01}\bar{G}_{-n+1,0} = \bar{T}\bar{G}_{-n+1,0}$$
$$\bar{G}_{0,-n} = \bar{G}_{0,-n+1}H_{01}^\dagger \bar{G}_{00} = \bar{G}_{0,n-1}\bar{S}$$
$$\bar{G}_{-n,-n} = \bar{G}_{00} + \bar{G}_{00}H_{01}\bar{G}_{-n+1,-n+1}H_{01}^\dagger \bar{G}_{00} = \bar{G}_{00} + \bar{T}\bar{G}_{-n+1,-n+1}\bar{S}$$

(49)

Note that $G_b = \lim_{n\to\infty} G_{n,n} = \lim_{n\to\infty} \bar{G}_{-n,-n}$. This completes the set of Green's functions that we may be interested in.

Supplementary Note 4   Trivial surface effects

The surface is in a different environment than the bulk, and there should be trivial surface effects, which could also influence the NLO responses on the surface. In this section, we argue that these trivial surface effects are much less significant than the topological effects.
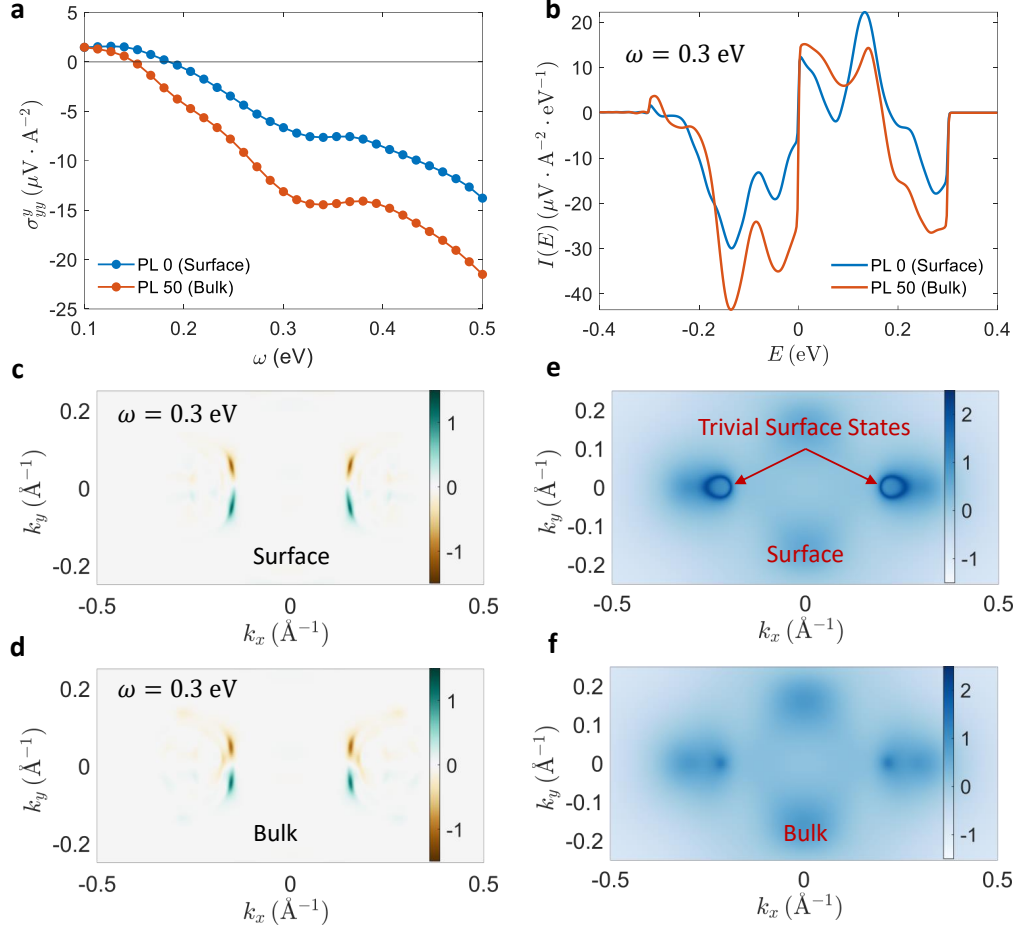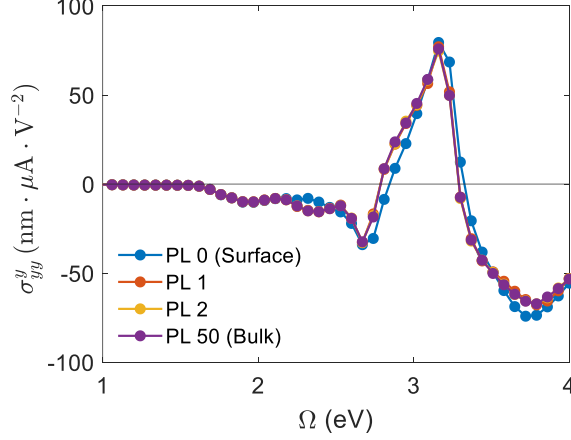
Supplementary Figure 8: BPV conductivity of strained T_d-WTe$_2$. The strained T_d-WTe$_2$ is topologically trivial. (a) layer-resolved BPV conductivity for PLs from the surface to the bulk. (b) energy-resolve contribution to the BPV conductivity for the surface and bulk PL for light frequency $\omega = 0.3$ eV. The energy-resolved contribution is defined as $I(E) \equiv -\frac{ie^3}{\omega^2 S}\sum_k \text{Tr}\{v^a G^<(E)\}$. (c, d) k-resolved contribution, defined as $I(k, E) \equiv -\frac{ie^3}{\omega^2 S}\text{Tr}\{v^a G^<(k, E)\}$ at $\omega = 0.3$ eV and $E = E_F$ for surface (c) PL and (d) bulk PL. (e, f) Spectrum function $A(k, E)$ at $E = E_F$ for surface (e) PL and (f) bulk PL. The spectrums in (e, f) are plotted in logarithm scale to give better visualization. Since the strained T_d-WTe$_2$ is topologically trivial, there are no Weyl points or Fermi arcs. But there are still topologically trivial surface states, which are marked in (e)

Supplementary Note 4.1   Strained and topologically trivial T_d-WTe$_2$

To distinguish the topological effects from the trivial surface effects, we artificially strain T_d-WTe$_2$ along a and c axis (8 % strain), so that the Weyl points merge and annihilate, and the system becomes topologically trivial [7]. Then we calculate surface and bulk NLO properties of this topologically trivial T_d-WTe$_2$. Note that applying bi-axial strain does not change any symmetry conditions for either the surface or the bulk. Hence, the trivial surface effects should be similar for topologically trivial (strained) and non-trivial (un-strained) T_d-WTe$_2$. By comparing topologically trivial and non-trivial T_d-WTe$_2$, the topological effects and trivial surface effects can be better distinguished.

The BPV conductivities of the trivial T_d-WTe$_2$ are shown in Figure 8. One can see that the

18

Supplementary Figure 9: The shift current conductivity σ_{yy}^{y} of bulk 2H-MoS$_2$. Two atomic layers of MoS$_2$ are used as a principal layer. The response on the surface is almost the same as that in the bulk.

differences between surface and bulk are much less significant, as compared with that in the case of topological non-trivial T$_d$-WTe$_2$ (Figure 2 in the main text). The surface and bulk spectrum functions are shown in Figure 8(e, f), and one can see there are still topologically trivial surface states, as marked in Figure 8(e). However, these trivial surface states cannot lead to a big difference in the total NLO responses. From the k-resolved contribution [Figure 8(c, d)], one can see that the trivial surface states have minor contributions to the total responses. Another remarkable feature is that, when T$_d$-WTe$_2$ is topologically trivial, the BPV conductivities have much smaller magnitude (on the order of $10 \, \text{nm} \cdot \mu\text{A/V}^2$) than that of the topologically non-trivial T$_d$-WTe$_2$ (on the order of $100 \, \text{nm} \cdot \mu\text{A/V}^2$).

Supplementary Note 4.2   Bulk 2H-MoS$_2$

We can also look at bulk 2H-MoS$_2$, which is topologically trivial, as an example to illustrate the surface effects in topologically trivial materials. Results are shown in Figure 9. One can see that for bulk 2H-MoS$_2$ the NLO responses of the surface are almost the same as that of the bulk. This again suggests that the trivial surface effects can be small.

Supplementary Note 4.3   Spin-Orbit Coupling in Au

Spin-orbit coupling (SOC) can influence the BPV responses as well. Here we take Au as an example, whose SOC is relatively strong since Au is a heavy element. We have calculated the BPV conductivity σ_{xx}^{x} on the surface PL of Au with and without SOC, and the results are shown in Figure 10 (due to the inversion symmetry, the conductivities are zero in the bulk of Au). One can see that the influence of

19
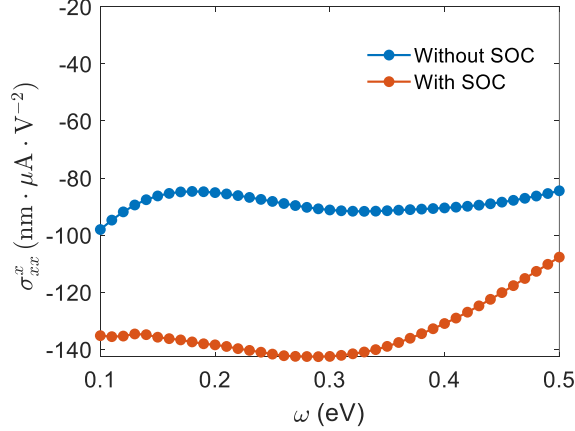
Supplementary Figure 10: BPV conductivity on the surface of Au. SOC is considered (not considered) for the red (blue) curve.

SOC is indeed non-negligible. This is because the SOC energy of Au is around $0.5\,\mathrm{eV}$ per atom, thus its influence in the electronic band structure is relatively large when we study responses in the frequency range of $\omega < 0.5\,\mathrm{eV}$.

Supplementary Note 5   Comments on the phenomenological carrier lifetime

In the main text we use a uniform carrier lifetime of $\tau = 0.2\,\mathrm{ps}$, which is chosen based on experimental results. In reality, the carrier lifetime may deviate from this value. In this section, we show our main conclusions are robust for a wide range of τ. Besides, the relationship between NLO conductivities σ and the carrier lifetime τ give us some clues on the underlying mechanism of the NLO responses.

Supplementary Note 5.1   Relationship between NLO responses and the carrier lifetime

In Figure 11 we show how the NLO charge and conductivities vary with τ. Generally speaking, the NLO conductivities get enhanced when the carrier lifetime is longer. This is reasonable: longer carrier lifetime indicates weaker dissipations in the system, and thus the carriers can have stronger responses under light illumination. Remarkably, the main conclusions of our work hold true for the wide range of carrier lifetime studied here. The main conclusions include: (1) the surface and the
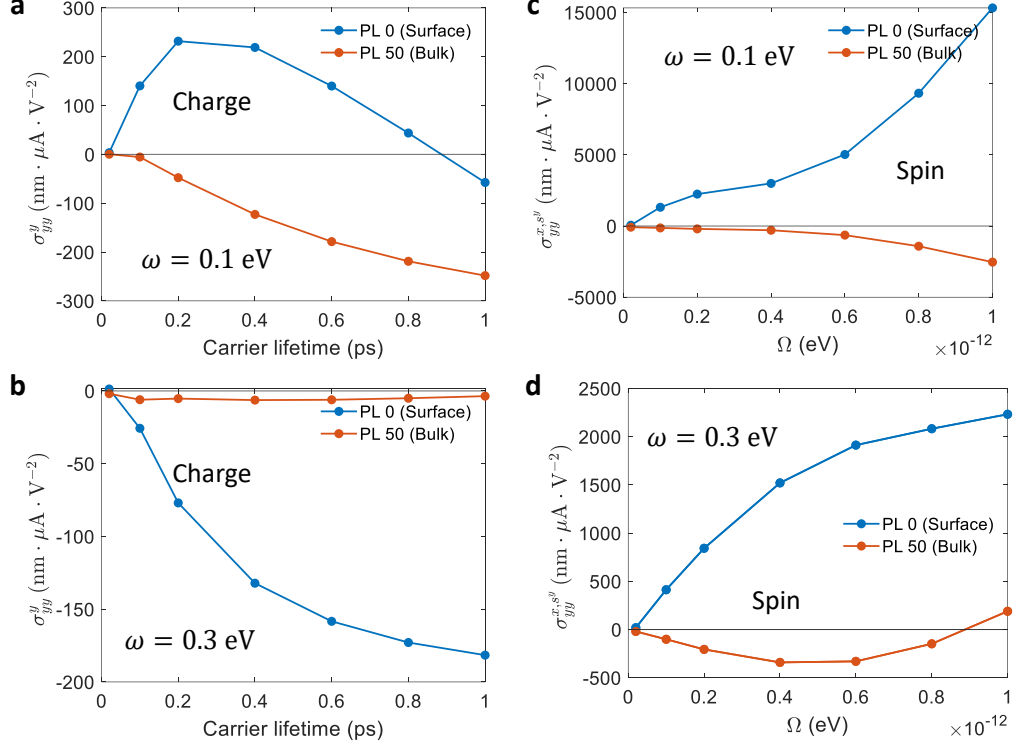
Supplementary Figure 11: NLO charge (a, b) and spin (c, d) conductivities as a function of carrier lifetime τ. (a, c) are for light frequency $\omega = 0.1$ eV, while (b, d) are for $\omega = 0.3$ eV

bulk can have distinct (even opposite) responses under light for certain light frequencies; (2) the layer-resolved nonlinear spin conductivity on the surface of T_d-WTe$_2$ can be orders of magnitude larger than that in the bulk.

In the discussions above, we use a uniform phenomenological carrier lifetime. This is usually called the "constant relaxation time approximation", which can greatly simplify calculations and is widely used in works studying NLO effects (e.g., Refs.[8, 9]). In reality, the carrier lifetime should be dependent on the mode of the carrier, including the wavevector k and the energy E of the carrier. Actually, when we add a Hubbard U term for the d orbitals of W atoms, there would be a non-uniform self-energy for the d orbitals of W. The non-uniform self-energy is mathematically equivalent to a mode-dependent lifetime (lifetime is equivalent to a purely imaginary self-energy $\Sigma = i\frac{\hbar}{\tau}$). From Figure 6 we can see that despite of the mode-dependent self-energy, the responses on the surface and in the bulk of T_d-WTe$_2$ are still distinct. This again suggests that the main conclusions of our work are solid and robust.

Supplementary Note 5.2   Underlying mechanisms of the nonlinear photocurrents

In the (non-interacting) single-particle formalism, the nonlinear photocurrents can be classified into different microscopic mechanism, each of which has a simple scaling relationship with τ. For example, (1) the shift current has conductivities $\propto \tau^0$ (do not depend on τ); (2) the injection current has conductivities $\propto \tau$ (see e.g., Refs. [10, 11]); (3) the quantum nonlinear Hall current has conductivities $\propto \tau$ [12]; and (4) the nonlinear Drude current has conductivities $\propto \tau^2$ (see e.g., Ref. [13]). In contrast, in Figure 11 there are no simple scaling relationships between conductivities σ and τ. There are two main reasons. (A) In our Green's function formalism, the electrons are not free, due to interactions with electrons on neighboring layers. As a result, the single-particle picture does not hold here. Intuitively, the nonlinear current calculated in our work cannot be attributed to a single mechanism but should be a combination of different mechanisms described above, as we will discuss below. (B) Even in the single particle formalism, the simple scaling relationship described above requires that the frequency ω of the light satisfies $\omega \gg 1/\tau$ [14]. However, in our work we focus on infrared light with frequencies $\hbar\omega \in [0.1, 0.5]$ eV. This is comparable with the electron lifetime in Figure 11, where one has $\hbar/\tau \in [0.004, 0.2]$ eV. Thus, the simple scaling relationships do not hold true.

Besides, in the single-particle formalism the second order NLO responses are closely connected to some topological or geometrical properties [10]. While our Green's function formalism is equivalent to the single-particle formalism in non-interacting systems, the topological and geometrical features are relatively obscured in the Green's function formalism, and it is not easy to distinguish the underlying mechanisms (e.g., shift current or injection current). However, by comparing the relationship between σ and τ shown in Figure 11 and the the scaling relationship described above, we can get some clues on the underlying mechanisms. For the charge current under linearly polarized light, we believe there should be shift current + nonlinear Drude mechanism in play. The shift current is generally an interband transition process: the wavefunction centers of the electrons and holes are different, leading to an electric dipole upon photon absorption. On the other hand, the Drude mechanism is essentially an intra-band process, which exists only in (semi-)metallic systems. For the spin current, the injection mechanism could also come into play. The injection mechanism comes from the fact that the electron and holes have different velocities, and that the coherent k and $-k$ excitations are imbalanced, leading to k and $-k$ asymmetry in steady-state population and a net current. Therefore, the currents shown in the current work should have multiple underlying mechanisms [11].

Supplementary Note 6 Other supplementary notes and figures

Supplementary Note 6.1 Estimation of the surface contribution to the total responses

In the main text we argue that the surface contribution to the total NLO responses can be significant. Here we do a rough quantitative estimation. From our ab initio calculations, we find the absorption coefficient of T_d-WTe$_2$ to be on the order of 10^5 cm^{-1} with light frequency $\omega \in [0.1, 0.5]$ eV. This indicates that the penetration depth is on the order of 100 nm. From Figure 2a in the main text, we can see that at $\omega = 0.1$ eV, the layer-resolved conductivity on the surface is about $250\,\text{nm} \cdot \mu\text{A/V}^2$, while that in the bulk is about $-50\,\text{nm} \cdot \mu\text{A/V}^2$. On the other hand, one principal layer (PL) is about 1.4 nm thick, therefore there are around 70 PLs within the 100 nm penetration depth. If we arbitrarily take the total current as the summation of all currents generated on these 70 PLs, then the surface PL would contribution to a fraction of $\frac{250}{70 \times 50} \approx 7$ %. In other words, the single surface PL makes a 7 % difference in the total current generated under light, which is not negligible in many situations (note that the response on the surface layer is opposite to that in the bulk).

Also, we would like to remark that the total current measured in experiments is also influenced by many practical factors, such as how the electrodes are connected to the samples. If the electrodes are directly attached on the surface, then the contributions from the surface may play a more significant role in the total current measured.

Moreover, when nanoscale thin films are used, the contribution from the surface can be even more significant. For example, if a 10 nm thick film is used, then the contribution from the surface may exceed 50 %.

Supplementary Note 6.2 NLO responses of Bi$_2$Se$_3$

As discussed in the main text, Bi$_2$Se$_3$, a prototype of topological insulator, is centrosymmetric in bulk, and has a bulk bandgap ~ 0.3 eV. Thus second order NLO responses would be zero in the bulk of Bi$_2$Se$_3$. These arguments are verified by our calculations (Figure 12).

Supplementary Note 6.3 Other supplementary figures for T_d-WTe$_2$

In this section we show other supplementary figures for T_d-WTe$_2$ (Figure 13-20).
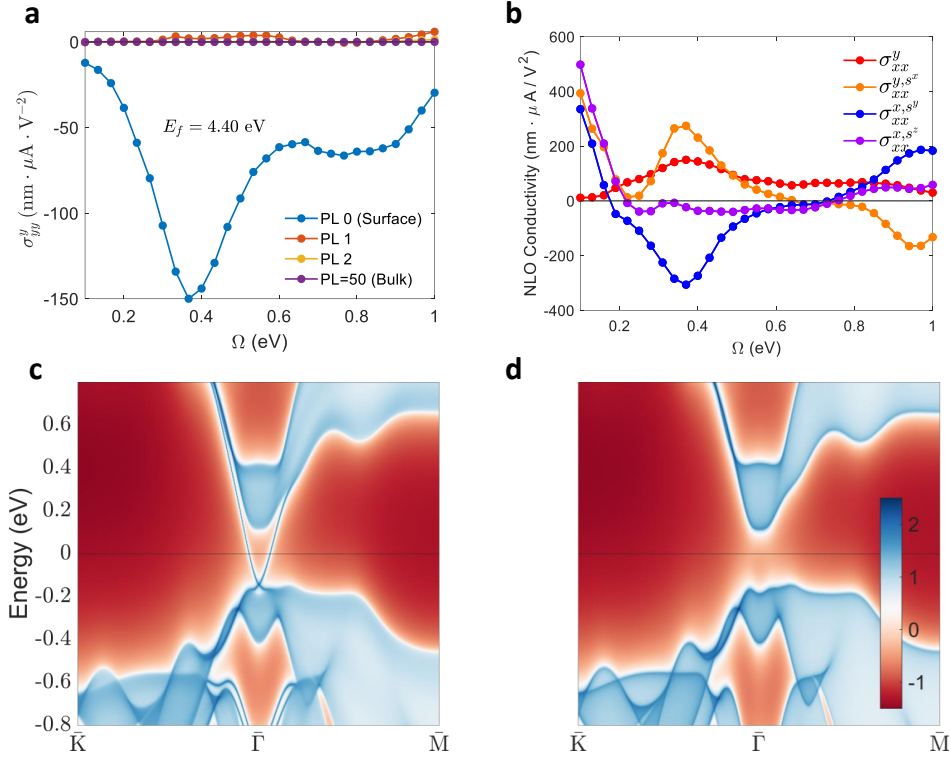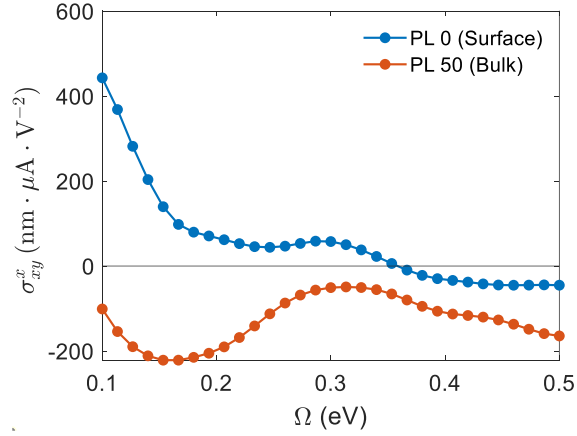
Supplementary Figure 12: (a) the shift current conductivity σ_{yy}^y of Bi$_2$Se$_3$ for PLs from the surface into the bulk. Three quintuple layers of Bi$_2$Se$_3$ are used as one principal layer. Since Bi$_2$Se$_3$ has inversion symmetry in the bulk, σ_{yy}^y is nonzero only on the surface. While in the bulk, σ_{yy}^y is unanimously zero. (b) bulk (spin) photovoltaic conductivities on the surface PL of Bi$_2$Se$_3$. (c, d) the spectrum function along high symmetry lines for (c) surface PL and (d) bulk PL. The horizontal line indicates the fermi level used in the calculations.
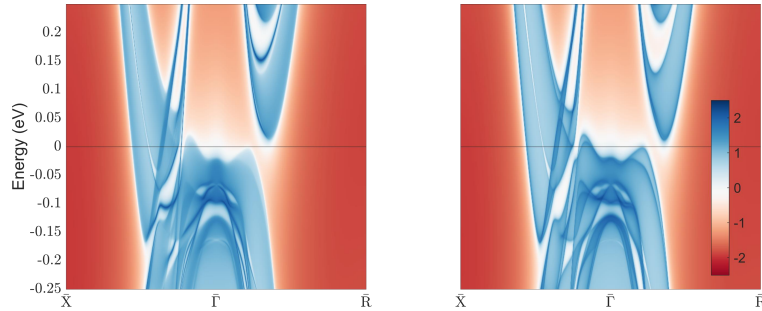
Supplementary Figure 13: The spectrum function along high symmetry lines in the Brillouin zone for (a) the surface PL and (b) the bulk PL of T$_d$-WTe$_2$. The surface states in the bulk bandgap is clearly seen in (a). The horizontal line indicates the Fermi level used in the main text.
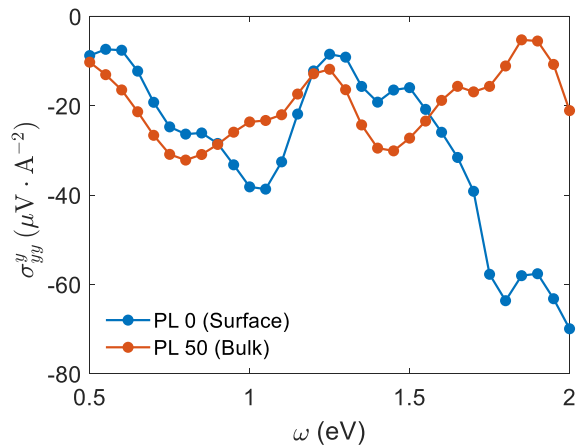
Supplementary Figure 14: The circular current conductivity σ_{xy}^x of WTe$_2$ for PLs on the surface and in the bulk. Similar to the shift current conductivity shown in the main text, the circular current on the surface can also be opposite to that in the bulk.
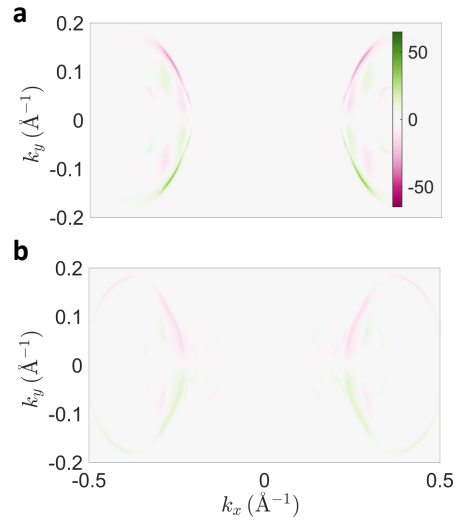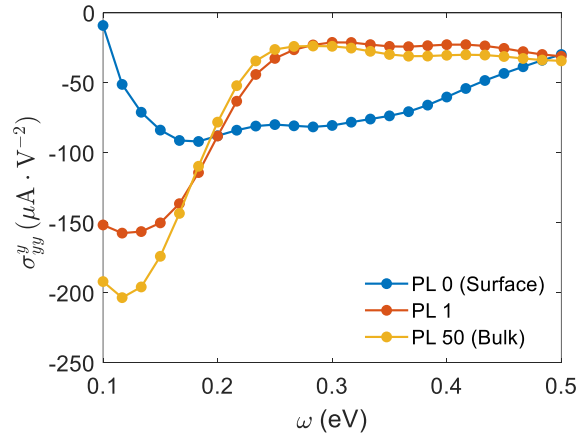

**Supplementary Figure 15:** The spin-x polarization on (a) surface PL and (b) bulk PL.

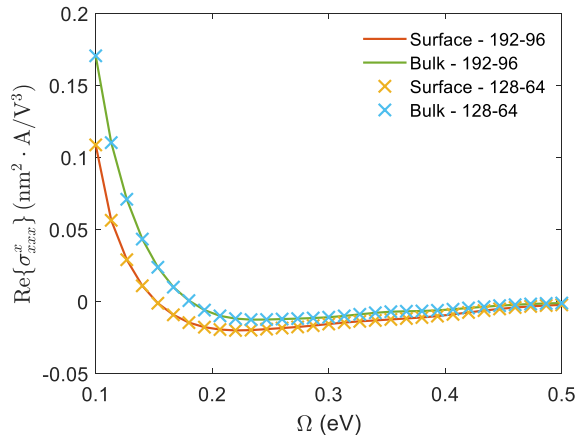**Supplementary Figure 16:** BPV conductivities of T$_d$-WTe$_2$ for light frequency above 0.5 eV.
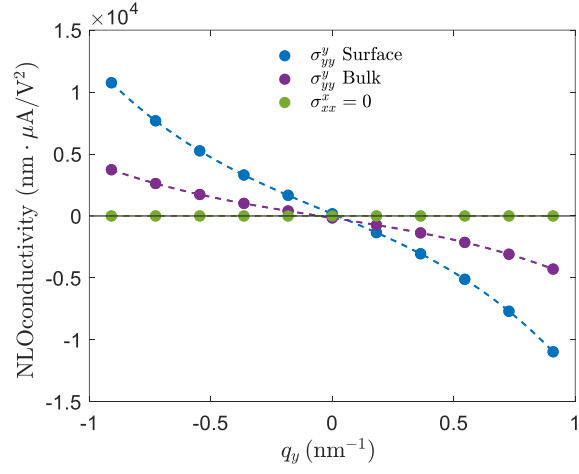
Supplementary Figure 17: The relationship between σ_{xx}^x, σ_{yy}^y of T_d-WTe$_2$ and q_y. Since q_y cannot break $\mathcal{M}_x$, σ_{xx}^x is always zero. The dashed curves are fittings of the solid dots with cubic functions.

Supplementary Figure 18: layer resolved BPV conductivity of T_d-WTe$_2$ for PLs from the surface to the bulk. The Fermi level is tuned to be 0.1 eV below the charge neutral point.
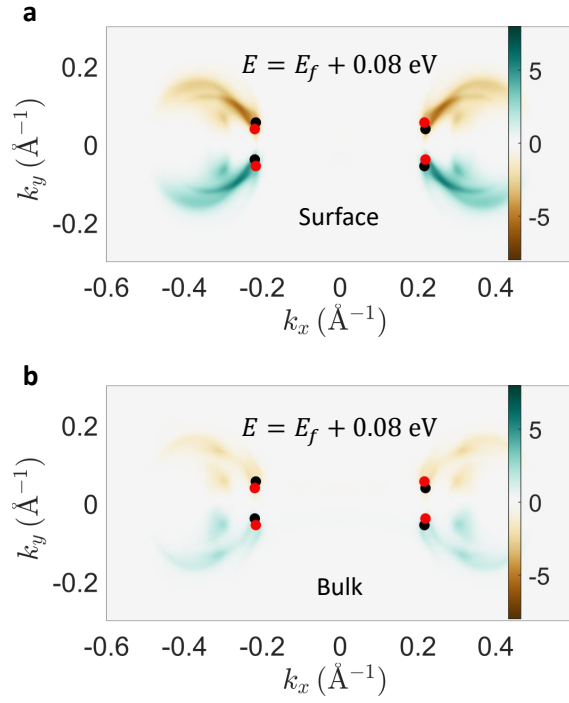
Supplementary Figure 19: Same as Figures 2(c,d) in the main text, but for $E = E_{\mathrm{F}} + 0.08$ eV

Supplementary Figure 20: Convergence test of the third-order conductivity. Two meshes with 192×96 and 128×64 k-points in the first Brillouin zone are used. The difference is negligible. Similar convergence tests have been carried out for other responses functions as well.

Supplementary References

[1] Rangel, T. *et al.* Large bulk photovoltaic effect and spontaneous polarization of single-layer monochalcogenides. *Physical review letters* **119**, 067402 (2017).

[2] Liu, J. & Dai, X. Anomalous hall effect, magneto-optical properties, and nonlinear optical properties of twisted graphene systems. *npj Computational Materials* **6**, 1–10 (2020).

[3] Georges, A., Kotliar, G., Krauth, W. & Rozenberg, M. J. Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions. *Reviews of Modern Physics* **68**, 13 (1996).

[4] Kotliar, G. *et al.* Electronic structure calculations with dynamical mean-field theory. *Reviews of Modern Physics* **78**, 865 (2006).

[5] Sancho, M. L., Sancho, J. L. & Rubio, J. Quick iterative scheme for the calculation of transfer matrices: application to mo (100). *Journal of Physics F: Metal Physics* **14**, 1205 (1984).

[6] Sancho, M. L., Sancho, J. L. & Rubio, J. Highly convergent schemes for the calculation of bulk and surface green functions. *Journal of Physics F: Metal Physics* **15**, 851 (1985).

[7] Soluyanov, A. A. *et al.* Type-ii weyl semimetals. *Nature* **527**, 495–498 (2015).

[8] Morimoto, T. & Nagaosa, N. Topological nature of nonlinear optical effects in solids. *Science advances* **2**, e1501524 (2016).

[9] de Juan, F., Grushin, A. G., Morimoto, T. & Moore, J. E. Quantized circular photogalvanic effect in weyl semimetals. *Nature communications* **8**, 1–7 (2017).

[10] Wang, H. & Qian, X. Ferroicity-driven nonlinear photocurrent switching in time-reversal invariant ferroic materials. *Science advances* **5**, eaav9743 (2019).

[11] Xu, H., Wang, H., Zhou, J. & Li, J. Pure spin photocurrent in non-centrosymmetric crystals: bulk spin photovoltaic effect. *Nature Communications* **12**, 1–9 (2021).

[12] Sodemann, I. & Fu, L. Quantum nonlinear hall effect induced by berry curvature dipole in time-reversal invariant materials. *Physical review letters* **115**, 216806 (2015).

[13] Hamamoto, K., Ezawa, M., Kim, K. W., Morimoto, T. & Nagaosa, N. Nonlinear spin current generation in noncentrosymmetric spin-orbit coupled systems. *Physical Review B* **95**, 224430 (2017).

[14] Xu, H. *et al.* Colossal switchable photocurrents in topological janus transition metal dichalco-genides. *npj Computational Materials* **7**, 1–9 (2021).