First-principles calculation of nonlinear optical responses by Wannier interpolation

Chong Wang,[1] Xiaoyu Liu,[1] Lei Kang,[2] Bing-Lin Gu,[1,4] Yong Xu,[3,4,5,*] and Wenhui Duan[1,3,4,†]

[1]*Institute for Advanced Study, Tsinghua University, Beijing 100084, China*

[2]*Beijing Computational Science Research Center, Beijing 100193, China*

[3]*State Key Laboratory of Low Dimensional Quantum Physics and Department of Physics, Tsinghua University, Beijing, 100084, China*

[4]*Collaborative Innovation Center of Quantum Matter, Tsinghua University, Beijing 100084, China*

[5]*RIKEN Center for Emergent Matter Science (CEMS), Wako, Saitama 351-0198, Japan*

(Received 16 July 2017; published 22 September 2017)

Various nonlinear optical (NLO) responses, like shift current and second harmonic generation (SHG), are revealed to be closely related to topological quantities involving the Berry connection and Berry curvature. First-principles prediction of NLO responses is of great importance to fundamental research and device design, but efficient computational methods are still lacking. The main challenge is that the calculations require a very dense k-point sampling that is computationally expensive and a proper treatment of the gauge problem for topological quantities. Here we present a Wannier interpolation method for first-principles calculation of NLO responses, which overcomes the challenge. This method interpolates physical quantities accurately for any desired k point with little computational cost and constructs a smooth gauge by the perturbation theory. To demonstrate the method, we study shift current of monolayer GeS and WS_2 as well as SHG of bulk GaAs, getting good agreements with previous results. We show that the traditional sum rule method converges slowly with the number of bands, whereas the perturbation way does not. Moreover, our method is easily adapted to build tight-binding models for the following theoretical investigations. Last but not least, the method is compatible with most first-principles approaches, including density functional theory and beyond. With these advantages, Wannier interpolation is a promising method for first-principles studies of NLO phenomena.

DOI: 10.1103/PhysRevB.96.115147

I. INTRODUCTION

Nonlinear optical (NLO) phenomena play an important role in modern optics and condensed matter physics [1], which have found great applications in various fields, especially the laser-related science and technology [1–3]. For instance, through appropriate frequency conversion, NLO materials are used to generate new coherent light sources from ultraviolet to infrared spectral regions with tunable frequencies where ordinary lasers perform poorly [2,3]. Despite intensive efforts over half a century, the exploration of novel NLO processes with promising applications is still under active investigation [4–11]. Emerging research directions in this field are inspired by the findings of Berry phase effects and topological phenomena in solids [12–14]. Importantly, recent theoretical works revealed that various nonlinear optical responses, including shift current, second harmonic generation (SHG), and photovoltaic Hall effect, are closely related to topological quantities involving the Berry connection and Berry curvature [7,8]. Moreover, both theory and experiment demonstrated that topological materials show large NLO conductivity and strong SHG effect [6,11]. In this context, the exploration of topological effects in NLO responses becomes important, which requires more research effort.

In material science, first-principles approaches are indispensable and powerful since they are able to predict material properties with no need of empirical parameters. There have been a few attempts to develop first-principles algorithms for calculating NLO responses [4,9,10,15]. However, these approaches are typically computationally very expensive, because a very dense k-point sampling is required by the calculation. Moreover, the random phases of Bloch wave functions from first principles, if not treated properly, would cause the gauge problem, which is crucial since the calculation involves the integration of the Berry connection and its derivatives with respect to k. Further methodological developments are needed to overcome these challenges. Recently, based on the computational method of maximally localized Wannier functions (MLWFs) developed by Mazari *et al.* [16,17], the Wannier interpolation method has been developed to calculate various physical observables, including anomalous Hall conductivity and orbital magnetization [18–20], for which the first-order perturbation theory is applied to deal with the gauge problem. This method features low computational cost and can be used in conjunction with most first-principles approaches, including density functional theory (DFT) and beyond (e.g., hybrid functionals and GW). However, as far as we know, there has been no theoretical development for studying NLO responses in the framework of Wannier interpolation.

In this work, we generalized the Wannier interpolation method to explore properties of NLO responses by first-principles calculation. This method provides accurate interpolations of physical quantities for any desired k point with little computational cost and constructs a smooth gauge by the second-order perturbation theory. As example studies, we calculated shift current of monolayer GeS and WS_2 as well as SHG effect of bulk GaAs, obtaining results in good agreement with the published ones [9,21]. Also we demonstrated that the perturbation algorithm is more advantageous over the traditional sum rule method [10,15] in calculating the derivative of the Berry connection, which

*yongxu@mail.tsinghua.edu.cn

†dwh@phys.tsinghua.edu.cn

does not suffer from slow convergence with the number of bands. Moreover, we showed that Wannier interpolation is easily adapted to tight-binding models that facilitate following theoretical investigations. These advantages make Wannier interpolation a promising method for first-principles studies of NLO phenomena.

This article is organized as follows. In Sec. II, the theoretical expression of shift current and SHG effect will be presented and their symmetry properties will be analyzed. The Wannier interpolation formalism of NLO responses will be discussed in Sec. III. In Sec. IV, the method will be applied to compute NLO responses of different materials. Finally, discussions and conclusions will be provided in Sec. IV.

II. DEFINITION AND BACKGROUND

A. Shift current

When an electron is pumped by light from valence band to conduction band, it may undergo a shift in position, which results in an electric current. Under linearly polarized light with electric field $\mathbf{E}$ at frequency ω, this shift current density $\mathbf{J}$ is a second-order response [10,15,22]:

$$J^a = \sigma^{abb}(\omega)E^b(\omega)E^b(-\omega), \tag{1}$$

where a, b, c are Cartesian indices and σ^{abb} is given by

$$\sigma^{abb}(\omega) = \frac{2g_s\pi e^3}{\hbar^2}\int \frac{d^3\mathbf{k}}{(2\pi)^3}\sum_{n,m} f_{nm} I^{abb}_{nm}\delta(\omega_{nm}-\omega), \tag{2}$$

where g_s is the spin degeneracy, $\hbar\omega_{nm} = E_n - E_m$, E_n and E_m are the energy eigenvalues of bands n and m for a given wave vector $\mathbf{k}$ ($\mathbf{k}$ is omitted for simplicity), and $f_{nm} = f(E_n) - f(E_m)$ is the difference of Fermi-Dirac occupation. The integrand I^{abb}_{nm} is

$$I^{abb}_{nm} = \mathrm{Im}\left[r^b_{mn} r^b_{nm;a}\right], \tag{3}$$

where r^b_{nm} is defined to be the Berry connection $A^b_{nm} \equiv i\langle u_n|\partial_{k_b} u_m\rangle$ when $n \neq m$ or zero when $n = m$, and $r^b_{nm;a} = \partial_{k_a} r^b_{nm} - i(A^a_{nn} - A^a_{mm})r^b_{nm}$. We will omit the k in the derivative and write $|\partial_{k_b} u_m\rangle$ as $|\partial_b u_m\rangle$ hereafter.

Equation (2) can be reformulated into a more transparent form [4,15]:

$$\sigma^{abb}(\omega) = \frac{2g_s\pi e^3}{\hbar^2}\int \frac{d^3\mathbf{k}}{(2\pi)^3}\sum_{n,m} f_{mn} r^b_{nm} r^b_{mn} R^{a,b}_{mn}\delta(\omega_{nm}-\omega), \tag{4}$$

where $R^{a,b}_{mn} \equiv \partial_a \phi^b_{mn} - A^a_{mm} + A^a_{nn}$ and ϕ^b_{mn} is the phase of $r^b_{mn} = |r^b_{mn}|e^{i\phi^b_{mn}}$. $R^{a,b}_{mn}$ has the unit of length and can be physically interpreted as position change of a wave packet during its pumping from band m to band n, which is usually referred to as shift vector [23]. On the other hand, the product $r^b_{nm} r^b_{mn}\delta(\omega_{nm}-\omega) = |r^b_{nm}|^2\delta(\omega_{nm}-\omega)$ can be interpreted as transition rate from band m to band n according to the Fermi golden rule. Therefore, shift current, according to this formula, is expressed as shift vector multiplied by transition rate.

Sometimes r^b_{nm} is approximately referred to as the matrix of position operator $\hat{\mathbf{r}}$. However, one should notice that r^b_{nm} cannot be viewed as the matrix of a physical observable, since the expectation value of the position operator is not well defined for Bloch wave functions. Nevertheless, r^b_{nm} does capture some features of $\hat{\mathbf{r}}$. For example, they share the same transformation rules under time reversal $\hat{T}$ and space inversion $\hat{I}$. It is well known that $\hat{T}\hat{\mathbf{r}}\hat{T}^{-1} = \hat{\mathbf{r}}$ and $\hat{I}\hat{\mathbf{r}}\hat{I}^{-1} = -\hat{\mathbf{r}}$. Similarly, $\mathbf{r}_{nm}(\mathbf{k}) = \mathbf{r}^*_{nm}(-\mathbf{k})$ if time reversal symmetry exists, and $\mathbf{r}_{nm}(\mathbf{k}) = -\mathbf{r}_{nm}(-\mathbf{k})$ if inversion symmetry exists. Importantly, shift current vanishes in the presence of inversion symmetry, since $\hat{I}$ takes E^b to $-E^b$ and J^a to $-J^a$ in Eq. (1), which thus requires $\sigma^{abb} = -\sigma^{abb} = 0$. Therefore, shift current only exists in materials with broken inversion symmetry.

B. Second harmonic generation effect

SHG effect of an input light with electric field $\mathbf{E}$ at frequency ω is described as

$$P^c(2\omega) = \epsilon_0 \chi^{abc}(\omega)E^b(\omega)E^c(\omega),$$

where $\mathbf{P}(2\omega)$ is the induced second-harmonic dipole per unit volume, ϵ_0 is the vacuum permittivity, χ^{abc} is the second-order susceptibility, and a, b, c are Cartesian indices. Following the same symmetry analysis as for shift current, SHG effect vanishes if the system respects inversion symmetry. For a material with inversion symmetry, the bulk introduces no SHG effect, and the observed SHG signals all come from surfaces or interfaces where inversion symmetry gets broken. Therefore, SHG effect has been widely applied to characterize material surfaces and interfaces.

The expression of χ^{abc} is rather lengthy [24], which is presented in Appendix A. Importantly, χ^{abc} can also be expressed by r^b_{nm} and $r^b_{nm;a}$, similar to σ^{abb}. Thus, to calculate both σ^{abb} of shift current and χ^{abc} of SHG effect, we need to compute r^b_{nm} and $r^b_{nm;a}$, preferably by first-principles calculations with no need for empirical parameters.

III. METHODS

A. Gauge problem

Since Bloch wave functions are only determined up to phase factors, one should be careful when calculating their derivatives with respect to the wave vector $\mathbf{k}$, for instance, in the calculations of r^b_{nm} and $r^b_{nm;a}$. The finite difference method, which only applies to continuous functions, would usually fail, considering that phases of Bloch wave functions obtained from first-principles calculations are typically distributed randomly in the Brillouin zone (BZ). Phase factors of Bloch wave functions are sometimes referred to as the gauge. Thus, this problem related to phase factors is called the gauge problem, which must be treated properly in first-principles calculations of nonlinear optical responses. To the best of our knowledge, there are currently two computational methods developed to avoid the gauge problem. To calculate r^b_{nm} and $r^b_{nm;a}$, one method uses a gauge invariant discrete expression; the other uses a momentum expression, as shown in the following.

B. Gauge invariant discrete expression

Ultimately, any physical observables should never depend on the choice of phases of Bloch wave functions. This property is called gauge invariant. The first computational method of overcoming the gauge problem is to use a carefully designed