Improved Maximum Entropy Method applied to Real-time Time-Dependent Density Functional Theory

M. Toogoshi, S. S. Kano, and Y. Zempo

Computer and Information Sciences, Hosei University, 3-7-2 Kajino, Koganei 184-8584 Japan

E-mail: `16t9001@cis.k.hosei.ac.jp`

Abstract. The maximum entropy method (MEM) is a key technique for spectral analysis. Its main feature is to describe spectra at low frequency with short time-series data. We adopted MEM to analyze the spectrum from the dipole moment obtained via time-dependent density functional theory (TDDFT) calculations in real time, which are intensively studied and applied to computing optical properties. With the MEM analysis, we proposed the use of a concatenated data-set made from repeated raw data coupled with the phase shift. We applied this technique to the spectral analysis of the TDDFT dipole moment of oligo-fluorene with $n = 8$. As a result, higher resolution was obtained without any peak shift due to the phase jump. The peak position is in good agreement with that of Fourier transform with just raw data. This paper presents the efficiency and characteristic features of this technique.

1. Introduction

Time-dependent density functional theory (TDDFT) is a powerful tool to analyze the optical properties of medium-to-large sized molecules. We employ a real-time and real-space technique to solve the time-dependent Kohn-Sham equations. Within the framework of this approach, the wave functions are calculated using the finite difference method on real spatial grids [1] without using explicit bases such as plane waves or Gaussian. In our procedure to calculate the optical properties, we use the time-series data, namely the dynamic dipole moment, from which the Fourier transform (FT) optical properties are calculated using a typical technique. The spectral resolution depends on the length of the dipole moment. The computational cost to obtain good resolution is quite large.

To solve this difficulty, we use the maximum entropy method (MEM) [2], which is a key technique in spectrum analysis. The principles of MEM are based on information theory for estimating unknown probability distributions based on the information about their expected values. MEM is widely used to solve a variety of problems related to spectral estimation in fields such as earth planetary science [3] and spectroscopy [4]. Its primary feature is that it can obtain fairly high resolution and accuracy with a relatively small number of time-series data. MEM is based on the autocorrelation function, that Fourier transformation is the power spectrum. We applied this technique to the spectral analysis of the time-dependent dipole moments of molecules using real-time TDDFT. In a comparison of the MEM and FT spectra for several molecules, we observed that MEM requires less time steps than does FT to obtain an optical spectrum with the same resolution [5]. As a new improved MEM, we propose using a concatenated data-set made from repeated raw data [6] coupled with the phase shift so as to avoid the side effect of artificial periodicity. With this improvement, we successively obtained a much better spectral resolution of the target peak. Therefore, this improvement will constitute a further advantage of MEM.

We applied this technique to a spectral analysis of the TDDFT dipole moment of typical molecules such as benzene, oligo-fluorene, and other materials. In the analysis of their optical properties, we are interested in the lower energy peak, which corresponds to the band gap area. The results show higher resolution and emphasized peaks near the band gap without the effects of artificial periodicity. In this paper, we present the characteristic features of this technique compared to FT and conventional MEM.

The paper is organized as follows. In the next section, we briefly describe the procedure of our TDDFT calculation. Then, we explain MEM and our new improvement. In the following section, the results of the absorption spectra of molecules are presented, and the observed characteristics of our MEM are followed our conclusions. We adopt atomic units throughout this paper.

2. Method

2.1. Time-dependent density functional theory

In this section, we briefly describe the procedure of our TDDFT calculation [7]. The basis of the procedure is the density functional theory (DFT) [8] with the local density approximation (LDA). The total energy of the ground state can be derived from the Kohn-Sham equation (KS) [9]. DFT is much less successful in describing optical responses and absorption spectra when electronic excited states are involved. However, this difficulty is, in principle, solved by the extension of DFT to its time-dependent version, TDDFT, which was established by Runge and Gross [10]. In analogy to the time-independent case, the TDDFT equation of motion coupled with the pseudopotentials is given by

$$\Big\{ -\frac{1}{2}\nabla^2 + V_{ion}^{ps}(\mathbf{r}) + V_H(\mathbf{r},t) + V_{XC}[\rho(\mathbf{r},t)] + V_{ext}(\mathbf{r},t)\Big\}\psi_j(\mathbf{r},t) = i\frac{\partial}{\partial t}\psi_j(\mathbf{r},t)\ , \tag{1}$$

where V_{ion}^{ps} is an ionic pseudopotential, V_H is the Hartree potential, and V_{XC} is the exchange correlation potential. Since the exact time-dependent xc kernel is not known, the originally non-local time-dependent xc kernel is replaced with a time-independent local one. This is reasonable when the density varies slowly with time. This approximation allows the use of a standard local ground-state xc functional in the TDDFT frame work. The Hartree and exchange-correlation potentials can be determined from the electronic charge density,

$$\rho(\mathbf{r},t) = \textstyle\sum_j |\psi_j(\mathbf{r},t)|^2. \tag{2}$$

The summation is over all occupied states j. The Hartree potential is determined by $\nabla^2 V_H = -4\pi\rho$, and as the usual LDA is used in our study as the xc potential V_{XC}. For the ionic potential, we employed the pseudopotential V_{ion}^{ps} in its separable form so that only the valence electrons were considered [11]. Prior to the calculation of the optical responses, we first solved the usual, time-independent formulation of the pseudopotential-DFT method [12] to obtain the optimized electronic structure [13]. Then, we applied an external field V_{ext} to the system as a perturbation and followed the linear responses of the system in real time. The xc functional we adopted is originally for the electronic ground states. This method of TDDFT has been effectively used for cases where the potential was time-dependent to study the behaviors of electrons in oscillating electric and magnetic fields and therefore excited-state reactions [14]. In our calculations, the real-time and real-space technique was adopted in solving Eq. (1) by means of the finite difference approach [1]. A uniform grid was used in our study for simplicity.

The time dependent wave function is given by $\psi(t) = \exp[-iHt]\tilde{\psi}(0)$, with the initial wave function $\tilde{\psi}$ at $t = 0$,

$$\tilde{\psi}|_{t=0} = e^{ik_z z}\psi(0), \tag{3}$$

where H is the Hamiltonian of the system, and k_z is a small wave number corresponding to the external perturbation in the z direction. In the linear response, the time-dependent polarizability is proportional to the dipole matrix element $\mu_\xi(t) = \langle\psi(t)|\xi|\psi(t)\rangle$, where ξ represents x, y and z. The polarizability $\alpha_\xi(\omega)$ is numerically calculated using the Fourier transformation (FT) of $\mu_\xi(t)$ and averaged as $\alpha = (\alpha_x + \alpha_y + \alpha_z)/3$. The oscillator strength distribution $S(\omega)$ is related to the imaginary part of the polarizability,

$$S(\omega) = \frac{2\omega}{\pi}\mathrm{Im}\ \alpha(\omega). \tag{4}$$

2.2. Maximum entropy method

In this section, we briefly describe the procedure and derivation of MEM. The entropy in information theory has been recognized as a measure of the uncertainty [15, 16]. Any inferences made from incomplete information should include a probability distribution that maximizes the entropy under the constraints of the available information [17].

The present study is based on Burg's method [2]. When each random variable $\mu(t)$ obeying the Gaussian distribution,

$$h \propto \int_{-\pi/\Delta t}^{\pi/\Delta t} \log P(\omega) d\omega, \tag{5}$$

is maximized, this h is taken as the entropy. The time-series data, e.g., the dipole moment discretized at time $t = n\Delta t$ with the time step Δt and the number of steps n is newly denoted as $\mu(n)$. In Eq. (5), $P(\omega)$ is the power spectrum, which is directly comparable to $S(\omega)$.

$$P(\omega) = \Delta t \sum_{m=-\infty}^{\infty} C_m \exp(-i\omega m\Delta t) \tag{6}$$

$$C_m = \frac{1}{2\pi}\int_{-\pi/\Delta t}^{\pi/\Delta t} P(\omega)\exp(i\omega m\Delta t) d\omega, \tag{7}$$

where C_m is the calculated autocorrelation at the time lag m as mentioned below. The solution is

$$P(\omega)/\Delta t = \left| \frac{\beta}{1 + \sum_{m=1}^{M} a_m \exp(-i\omega m\Delta t)} \right|^2, \tag{8}$$

where M is the maximum length of the autocorrelation C_M. The parameters a_m and β are the Lagrange multipliers and they are the solution of the Yule-Walker equation [18],

$$\begin{pmatrix} C_0 & C_1 & \cdots & C_M \\ C_1 & C_0 & \cdots & C_{M-1} \\ \vdots & \vdots & & \vdots \\ C_M & C_{M-1} & \cdots & C_0 \end{pmatrix} \begin{pmatrix} 1 \\ a_1 \\ \vdots \\ a_M \end{pmatrix} = \begin{pmatrix} |\beta|^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{9}$$

We adopt the Levinson-Durbin algorithm to solve Eq. (9) efficiently. The autocorrelation C_m at the lag m is

$$C_m = \frac{1}{N}\sum_{n=0}^{N-m-1} \mu^*(n)\mu(n+m), \tag{10}$$

where N is the length of the time-series data. We extend the autocorrelation to the complex data. The periodicity of the signal is expressed by the lag in the autocorrelation. The value at shorter lags shows the contribution from high frequencies, and the value at longer lags shows the contribution from low frequencies. Even though the information in the low energy region is important for the optical spectrum analysis, the value of the autocorrelation at larger lags can become quite small. The effective value of M is restricted by the number of data N. Therefore, we require a sufficiently large N to obtain a reasonable spectral resolution in the low frequency region; the same as in the case of FT.

2.3. Improvements to MEM

As a new improved MEM, we propose using a concatenated data-set made from repeated raw data coupled with the phase shift so as to avoid the side effect of artificial periodicity.

We are interested in the lower energy region. To emphasize the signal in this region, we need a larger lag M, which is the maximum lag of the autocorrelation. However, increasing M sometimes causes unphysical results such as peak splits and false peaks. In addition, the number M is limited by the total number of time steps N. To solve this difficulty, we repeat the raw data many times as shown in Fig. 1(b). The resolution in the lower energy region is attributed to the maximum lag M. With the repeated signal, the value of M may be selected at sufficiently large values without additional computations. With this procedure, we can save on computational costs.

At the concatenated point, of course, there is a phase jump for each frequency component, which is recognized as noise in the total signal. To reduce side effects of the phase jump such as peak shifts, we introduced a phase shift at each concatenated data in the repeated data.

$$\tilde{\mu}(n) = \mu'(n) \exp(ik\phi), \tag{11}$$

where $\mu'(n)$ is the repeated raw data, and $\tilde{\mu}(n)$ is the k-times repeated and concatenated data $\mu'(n)$ with an appropriate phase ϕ for the target frequency ($-\pi \leq \phi \leq \pi$). This introduction of the phase to the signal does not affect to the total power spectrum due to the definition used in Eq. (8). Of course, a different target frequency has a different phase. Since the phase ϕ is a constant value, it is necessary to choose the appropriate phase for the target frequency to minimize the side effects of the phase jump due to the concatenated repeated data. Figure 1(c) shows the repeated data $N = 5000 \times 4$ with the phase $\phi = 0.25\pi$, whose data is obtained from the TDDFT signal applied to the oligomer of fluorene with $n = 8$. In this case, we can see that the raw data is in the first 5×10^3 steps and $\phi = 0.25\pi$ is introduced in the next 5×10^3 steps, $2 \times \phi$ in the following 5×10^3 steps, and so on. We can select the ϕ so as to maximize the target peak. A mismatched phase naturally makes the signal cancel, weakening the strength. Therefore, we choose this phase to match the bandgap peak appropriately in our study.

3. Results and Discussion

In this section, we show the calculated absorption spectrum of the total oscillator strength of the improved MEM. Then, we discuss the characteristics of our method compared to the results of conventional MEM and FT.

3.1. Benzene, Naphthalene, Anthracene, and Tetracene

As a simple example to confirm the efficiency of our proposed method, we applied our improved MEM to the analysis of benzene, naphthalene, anthracene, and tetracene. The molecular structure of these molecules is based on the ground state. The time evolution is carried out according to real-time TDDFT [7, 13]. The spectrum $S(E)$ is calculated from the time-series data in our MEM technique. The results are shown in Fig. 2, where the solid line is the result

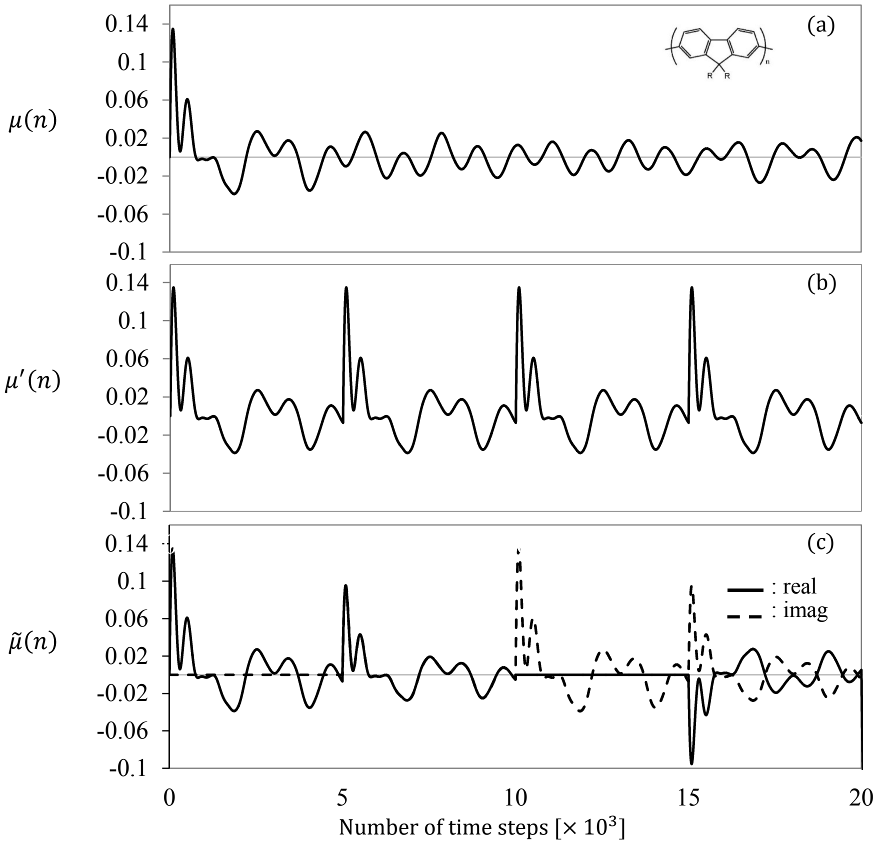

Figure 1. Comparison of the time-series data for oligo-fluorene with $n = 8$, which was made using real-time TDDFT. (a) Real time evolution up to $N = 20000$ for FT. (b) Four times repeated data of the dipole moment with $N = 5000 \times 4$. (c) The data in panel (b) together with the phase $\phi = 0.25\pi$. The solid line is the real part of the data, and the dashed line is the imaginary part of the data.

for benzene, the dashed line is that for naphthalene, the dash-dotted line is that for anthracene, and the dotted line is that for tetracene. These spectra are normalized by the intensity at the first peak. Figure 2(a) shows the results of the improved MEM. The parameters (M and ϕ) of our MEM are $M = 1800$ and $\phi = -0.25\pi$ for the analysis of benzene, $M = 2500$ and $\phi = -0.20\pi$ for that of naphthalene, $M = 2000$ and $\phi = -0.23\pi$ for that of anthracene, and $M = 3000$ and $\phi = -0.21\pi$ for that of tetracene. For both MEM calculations, the same number of the time steps, $N = 10000 \times 100$, is used. Figure 2(b) shows the results of FT with $N = 10000$. Our method provides good resolution compared with that of FT. In addition, we can see a clear spectrum in comparison to FT. This is one feature of our MEM.

3.2. Fluorene

Poly(9,9-dialkyl-fluorene) and their substituted derivatives are used as organic LED materials and are expected to be basic materials for blue emission LEDs. Their electronic structures have been extensively studied [19, 20, 21, 22, 23]. We employed the oligomer of fluorene (oligo-FL) with $n = 8$ and performed the MEM calculation. The molecule size is fairly large; therefore, then the calculation for each step is quite expensive. It takes a long time to perform the calculation to obtain the low energy part of the spectrum. If we can save on the calculation

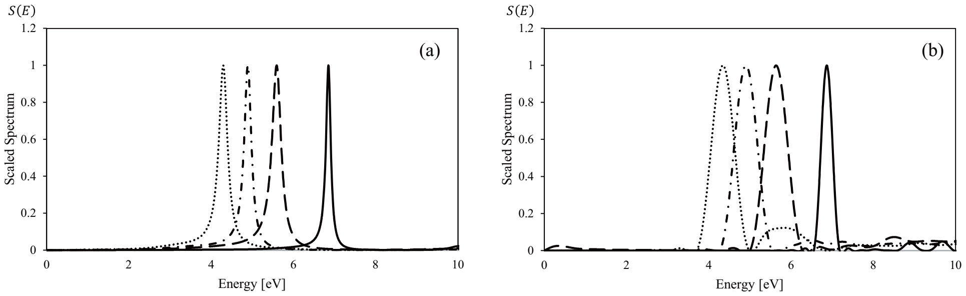

Figure 2. Comparison of the improved MEM and FT spectra for benzene (solid line), naphthalene (dashed line), anthracene (dash-dotted line) and tetracene (dotted line). (a) Results of the improved MEM using $N = 10000 \times 100$ with $M = 1800$ and $\phi = -0.25\pi$ for benzene, $M = 2500$ and $\phi = -0.20\pi$ for naphthalene, $M = 2000$ and $\phi = -0.23\pi$ for anthracene, and $M = 3000$ and $\phi = -0.21\pi$ for tetracene. (b) Results of FT with $N = 10000$. All the spectra are normalized at the first low energy peak.

cost when calculating the lower energy part of the spectrum, our method will be considered effective. We applied our improved MEM to this signal analysis. Figure 3 shows a comparison of the absorption spectra using four different methods. The solid line shows the result of the improved MEM with $N = 5000 \times 100$, $M = 2500$, and $\phi = 0.25\pi$, the dashed line is that of FT for $N = 20000$, the dash-dotted line is our conventional MEM spectrum with just the real time steps $N = 5000$ and $M = 2500$, and the dotted line is that of MEM with $N = 5000 \times 100$ and $M = 2500$. These spectra are normalized at the second low energy peak.

FT (the dashed line) and the simple MEM (the dash-dotted line) provide a broad peak at the bandgap area. With only the repeated signal (the solid and dashed lines), the strength of the first peak is emphasized. Without a phase shift, we can observe that the peak position shifts due to the side effect of the phase jump [6]. However, we can obtain a clear first peak using our new method (the solid line). In this case, we choose the matched phase corresponding to the first peak. Therefore, the position of the second peak differs from those in the other methods. (Note that the normalization of the signal was done at the second peak.) If we choose the mismatched phase for the target peak, the position of the peak shifts and the strength decreases, as we mentioned above. The bandgap peak at 2.53 eV for the improved MEM may correspond to the experimentally observed peak at 3.56 eV for poly-FL. This discrepancy is due to an inherent problem in DFT [7].

3.3. C60 fullerene

Our new MEM does not always works well for the target frequency. In this section, we show a case where our MEM does not work well, when there are several peaks close to each other. To separate each peak clearly, we need sufficient time series data, not only for MEM but also for FT.

Figure 4 shows the result for C60 fullerene. We can see that the separation is not sufficient, even if we use the concatenated repeated data with the phase shift. The solid line is the result of our MEM with the signals $N = 10000 \times 100$ together with the phase $\phi = -0.125\pi$ and $M = 6000$. The dashed line shows the result of FT with $N = 10000$. In this case, we can see a clear first peak to compare to that of MEM.

This is because, if the length of the data N is not sufficient, we obtain a broad peak, similar to the Lorentz function, in any MEM calculation. In particular, when the strength of one of the closer peaks is high, the slope of the peak affects the strength of the smaller peaks. Of course, if

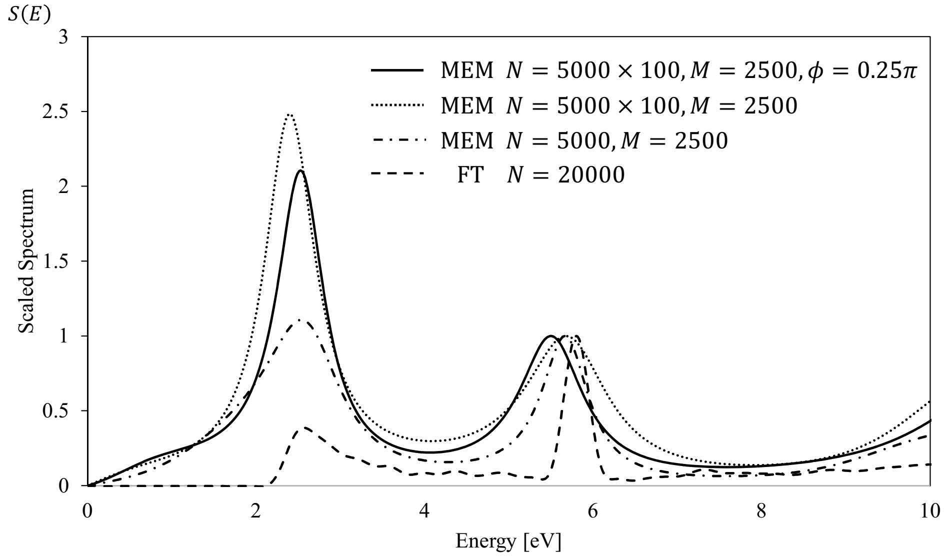

Figure 3. Comparison of the absorption spectrum for oligo-fluorene with $n = 8$. The dashed line is the result of FT with $N = 20000$, the dash-dotted line is that of MEM with $N = 5000$, the dotted line is that of MEM with repeated data $N = 5000 \times 100$, and the solid line is that of the improved MEM with $N = 5000 \times 100$, $M = 2500$, and $\phi = 0.25\pi$. The spectra are normalized at the second low energy peak.

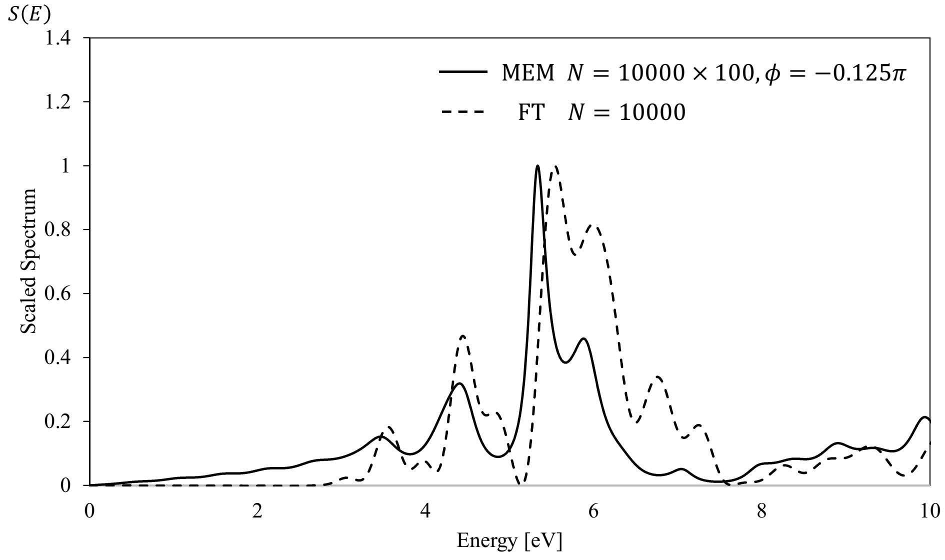

Figure 4. Absorption spectrum for C60 fullerene. The dashed line is the result of FT with $N = 10000$, and the solid line is that of the improved MEM with $N = 10000 \times 100$, $M = 6000$, and $\phi = -0.125\pi$. The spectra are normalized at the highest peak in the region of 0-10 eV. The experimental absorption peaks at 3.8, 4.8 and 5.8 eV [13].

we have a sufficiently large total time step N, a clear separation should be obtained. Therefore, we need calculate the time evolution until the information of the low energy part is obtained.

4. Concluding remarks

We employed MEM to obtain the optical spectra of molecules, and the spectral resolution was significantly improved compared to that of FT with the same number of total time steps. As an improvement, we introduced the use of repeated signals together with the phase to make seemingly long data, in which a larger time lag was included. In particular, this is an efficient technique to find the low energy peak. We obtained significantly better resolutions for the target peak without adding further calculated data. This means that we can efficiently reduce our calculation costs. However, our new MEM does not always work well for the target peak. Using this method for the spectrum analysis, we cannot obtain a low energy component that is not included in the data. We need to calculate a certain amount of time evolution before the information of the low energy part can be obtained.

Acknowledgments

The authors would like to express their gratitude to Professor H. Hanaizumi for useful discussions and encouragement. This work was partially supported by JSPS Grants-in-Aid for Scientific Research (C) Grant number 16K05047, Sumitomo Chemical, Co. Ltd. and Simulatio Corporation.

References

[1] Chelikowsky J, Troullier N, Wu K and Saad Y 1994 *Phys. Rev. B***50**, 11355

[2] Burg JP 1968 *Advanced Study Institute on Signal Processing*, 15-0, NATO, Erschede, Netherlands

[3] Sakurai T 2002 *Earth Planets Space* **54**, 153

[4] Vartiainen EM, Peiponen KE, Kishida H, and Koda T 1996 *J. Opt. Soc. Am.* B**13**, 2106

[5] Toogoshi M, Kato M, Kano SS, and Zempo Y 2014 *J. Phys. Conf. Ser.* **510**, 012027

[6] Toogoshi M, Kano SS, and Zempo Y 2015 *J. Phys. Conf. Ser.* **640**, 012069

[7] Zempo Y, Akino, Ishida M, Ishitobi M, and Kurita Y 2008 *J. Phys. Condens. Matter* **20**, 064231

[8] Hohenberg P and Kohn W 1964 *Phys. Rev.* **136**, B864

[9] Kohn W and Sham L. J 1965 *Phys. Rev.* **140**, A1133

[10] Runge E and Gross EKU 1984 *Phys. Rev. Lett.* **52**, 997

[11] Troullier N and Martins JL 1991 *Phys. Rev.* B**43**, 1993

[12] Pickett WE 1989 *Comput. Phys. Reports* **9**, 115

[13] Yabana K and Bertisch GF 1996 *Phys. Rev.* B**54**, 4484

[14] Tateyama Y, Oyama N, Ohno T, and Miyamoto Y 2006 *J. Chem. Phys.* **124**, 124507

[15] Shannon C 1948 *Bell Syst. Tech. J.*,**27**, 379 and 623

[16] Jaynes ET 1957 *Phys. Rev.* **106**, 620

[17] Jaynes ET 1982 *Proc. IEEE* **70**, 939

[18] Haykin S and Kesler S 1979 *"Non linear Methods of Spectral Analysis,"* ed Haykin S, Springer(Berlin)

[19] Hummer K, Puschnig P, Sagmeister S, and Ambrosch-Draxl C 2006 *Mod. Phys. Lett.* B**20**, 261 , and the references therein

[20] Yu WL, Cao Y, Pei J, Huang W, and Heeger AJ 1990 *Appl. Phys. Lett.* **75** 3270

[21] Wang JF, Feng JK, Ren AM, Liu XD, Ma YG, Lu P, and Zhang HX 2004 *Macromolecules* **37**, 3451

[22] Brière JF and Côté M 2004 *J. Chem. Phys.* B**108**, 3123

[23] Grice AW, Bradley DDC, Bernius MT, Inbasekaran M, Woo EP and Wu WW 1998 *Appl. Phys. Lett.* **73**, 629