Phys. Status Solidi B **250**, No. 11, 2481–2494 (2013) / **DOI** 10.1002/pssb.201349247

Part of Special Issue on
Quantum Transport at the Molecular Scale

physica status solidi b
www.pss-b.com
basic solid state physics

Time-dependent quantum transport theory and its applications to graphene nanoribbons

Hang Xie*,[1], Yanho Kwok[1], Yu Zhang[1], Feng Jiang[2], Xiao Zheng[3], YiJing Yan[2,3], and GuanHua Chen*,[1]

[1] Department of Chemistry, The University of Hong Kong, Pokfulam, Hong Kong
[2] Department of Chemistry, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
[3] Hefei National Laboratory for Physical Sciences at the Microscale, University of Science and Technology of China, Hefei 230026, China

Received 29 May 2013, revised 2 October 2013, accepted 4 October 2013
Published online 6 November 2013

Keywords density functional theory, electronic transport, equation of motion, graphene, nanoribbons, nonequilibrium Green's function, time-dependent DFT

* Corresponding author: e-mail xiehanggm@gmail.com, Phone: +852 28592164; Fax: +852 2915 5176
** e-mail: ghc.hku@gmail.com, Phone: +852 28592164; Fax: +852 2915 5176

Time-dependent quantum transport parameters for graphene nanoribbons (GNR) are calculated by the hierarchical equation of motion (HEOM) method based on the nonequilibrium Green's function (NEGF) theory [Xie et al., J. Chem. Phys. **137**, 044113 (2012)]. In this paper, a new initial-state calculation technique is introduced and accelerated by the contour integration for large systems. Some Lorentzian fitting schemes for the self-energy matrices are developed to effectively reduce the number of Lorentzians and maintain good fitting results. With these two developments in HEOM, we have calculated the transient quantum transport parameters in GNR. We find a new type of surface state with delta-function-like density of states in many semi-infinite armchair-type GNR. For zigzag-type GNR, a large overshooting current and slowly decaying transient charge are observed, which is due to the sharp lead spectra and the "even–odd" effect.

1 Introduction Quantum transport is an important field of research, due to the rapid improvement in nanotechnology and the semiconductor industry [1–3]. First-principles nonequilibrium Green's function (NEGF) theory has been widely employed to simulate the steady-state currents through electronic devices [4–6]. Besides the steady-state calculations, the time-dependent quantum transport theory has been developed within the framework of NEGF [7, 8] and time-dependent density functional theory (TDDFT) [9–13]. To deal with the transient current in open systems, a major approach is the embedding scheme (calculate a finite region in an extended environment with the help of self-energy). For example, the open boundary condition derived from partitioned propagators is employed for the equation of motions (EOM) of the density matrix [14]; and the transparent boundary condition is employed for the wavefunction propagations [12, 15]. Another approach is to use finite leads with large lengths to mimic the open terminals and calculate the currents in a limited duration, as in Ref. [16].

In time-dependent quantum transport calculations, the wide band limit (WBL) approximation is often used [7, 13]. Recently, we have developed a first-principles hierarchical equation of motion (HEOM) for the reduced single-electron density matrix (RSDM; denoted as RSDM-HEOM), which goes beyond the WBL approximation [17–21]. Multi-Lorentzian expansion is employed to approximate the spectra of the leads [20]. As RSDM is much simpler than the many-electron density matrix, the RSDM-HEOM method is much more efficient so that it is employed to model the realistic systems.

However, in the HEOM calculation for large systems, it is difficult to calculate the initial values, as the large number of auxiliary density matrices (ADM) leads to a very huge matrix equation [20]. Also, the computational load of HEOM is heavily determined by the number of Lorentzians. For large systems, it is crucial to find a good set of Lorentzians, which mimic the lead spectra profiles with the minimal number of Lorentzians. In this paper, we derive a new integral formula to solve the initial HEOM values of

large systems more quickly and accurately. We also develop a set of Lorentzian fitting algorithms, which can automatically fit all the self-energy curves with a small number of Lorentzians. With these two developments, our HEOM can be extended to much larger systems.

With this method, we calculate the time-dependent transport parameters of some graphene nanoribbons (GNR). Graphene has attracted a lot of research interest in recent years, mainly due to its novel band structure with linear dispersion on the Dirac points [22]. But most of the works study its static properties. Some time ago, Perfetto et al. [16] calculated the transient current of very large GNR. They found for large enough zigzag GNR, the current versus time curve shows two temporal plateaus: the first one results from the relativistic spectrum of bulk graphene; and the second one is related to the steady-state current. In this paper we focus on the small-sized GNR, and some new properties such as the establishment of "even–odd" effect are observed.

This paper is organized as follows. The methodology is described in Section 2, including the HEOM introduction, the initial-value calculation method, and three Lorentzian fitting algorithms. In Section 3, numerical results for the GNR are given. Two types of nanoribbons (armchair and zigzag) are calculated and their transport properties are investigated. A summary is given in Section 4.

2 Methodology The HEOM method is based on the NEGF theory. Instead of calculating the time evolution of the Green's functions, this method defines some time integrals of the Green's functions and self-energies as the ADM. Then, with the equations of motion for the Green's functions and self-energies, a set of differential equations for the density matrices and ADM are derived. The following part gives the details of this method.

2.1 Introduction to HEOM In a lead–device–lead system, the EOM for the device's RSDM is given as below [13]:

$$i\dot{\boldsymbol{\sigma}}_{\mathrm{D}}(t)=[\boldsymbol{h}_{\mathrm{D}}(t),\boldsymbol{\sigma}_{\mathrm{D}}(t)]+\sum_{\alpha=1}^{N_\alpha}[\boldsymbol{h}_{\mathrm{D}\alpha}\boldsymbol{\sigma}_{\alpha\mathrm{D}}-\boldsymbol{\sigma}_{\mathrm{D}\alpha}\boldsymbol{h}_{\alpha\mathrm{D}}].$$

With some derivation, this EOM can be written as

$$\begin{aligned}i\dot{\boldsymbol{\sigma}}_{\mathrm{D}}(t)=[\boldsymbol{h}_{\mathrm{D}}(t),\boldsymbol{\sigma}_{\mathrm{D}}(t)]+i\sum_{\alpha=1}^{N_\alpha}\int_{-\infty}^{t}\mathrm{d}\tau[\boldsymbol{G}_{\mathrm{D}}^{<}(t,\tau)\cdot\boldsymbol{\Sigma}_{\alpha}^{>}(\tau,t)\\-\boldsymbol{G}_{\mathrm{D}}^{>}(t,\tau)\cdot\boldsymbol{\Sigma}_{\alpha}^{<}(\tau,t)+\mathrm{H.C.}],\end{aligned}\tag{1}$$

where $\boldsymbol{\sigma}_{\mathrm{D}}$ and $\boldsymbol{h}_{\mathrm{D}}$ are the RSDM and Hamiltonian of the device, respectively. $\boldsymbol{\Sigma}_{\alpha}^{x}(t,\tau)$ are the lesser ($x=<$) or greater ($x=>$) self-energy for the lead α; $\boldsymbol{G}_{\mathrm{D}}^{x}(t,\tau)$ are the lesser or greater Green's function of the device. H.C. means the Hermitian conjugate. The second term above accounts for the dissipation effect due to the leads. In Appendix A, we show that this term is also related to the transient-current expression.

According to Ref. [19], if we introduce the energy-resolved self-energies $\boldsymbol{\Sigma}_{\alpha}^{<,>}(\varepsilon,\tau,t)$ (with $\boldsymbol{\Sigma}_{\alpha}^{<,>}(\tau,t)=\int\mathrm{d}\varepsilon\cdot\boldsymbol{\Sigma}_{\alpha}^{<,>}(\varepsilon,\tau,t)$), and the following 1st- and 2nd-tier ADM,

$$\begin{aligned}\boldsymbol{\varphi}_{\alpha}(\varepsilon,t)=i\int_{-\infty}^{t}\mathrm{d}\tau[\boldsymbol{G}_{\mathrm{D}}^{<}(t,\tau)\cdot\boldsymbol{\Sigma}_{\alpha}^{>}(\varepsilon,\tau,t)\\-\boldsymbol{G}_{\mathrm{D}}^{>}(t,\tau)\cdot\boldsymbol{\Sigma}_{\alpha}^{<}(\varepsilon,\tau,t)],\end{aligned}\tag{2}$$

$$\begin{aligned}\boldsymbol{\varphi}_{\alpha\alpha'}(\varepsilon,\varepsilon',t)=i\int_{-\infty}^{t}\mathrm{d}t_1\int_{-\infty}^{t}\mathrm{d}t_2\{[\boldsymbol{\Sigma}_{\alpha'}^{<}(\varepsilon',t,t_1)\cdot\boldsymbol{G}_{\mathrm{D}}^{a}(t_1,t_2)\\+\boldsymbol{\Sigma}_{\alpha'}^{r}(\varepsilon',t,t_1)\cdot\boldsymbol{G}_{\mathrm{D}}^{<}(t_1,t_2)]\boldsymbol{\Sigma}_{\alpha}^{>}(\varepsilon,t_2,t)\\-[\boldsymbol{\Sigma}_{\alpha'}^{>}(\varepsilon',t,t_1)\cdot\boldsymbol{G}_{\mathrm{D}}^{a}(t_1,t_2)\\+\boldsymbol{\Sigma}_{\alpha'}^{r}(\varepsilon',t,t_1)\cdot\boldsymbol{G}_{\mathrm{D}}^{>}(t_1,t_2)]\boldsymbol{\Sigma}_{\alpha}^{<}(\varepsilon,t_2,t)\},\end{aligned}\tag{3}$$

we may derive a set of equations of motion:

$$\begin{aligned}i\dot{\boldsymbol{\sigma}}(t)=[\boldsymbol{h}_{\mathrm{D}}(t),\boldsymbol{\sigma}_{\mathrm{D}}(t)]\\-\sum_{\alpha}\int\mathrm{d}\varepsilon\cdot[\boldsymbol{\varphi}_{\alpha}(\varepsilon,t)-\boldsymbol{\varphi}^{\dagger}(\varepsilon,t)],\end{aligned}\tag{4}$$

$$\begin{aligned}i\dot{\boldsymbol{\varphi}}_{\alpha}(\varepsilon,t)=[\boldsymbol{h}_{\mathrm{D}}(t)-\varepsilon-\Delta_{\alpha}(t)]\cdot\boldsymbol{\varphi}_{\alpha}(\varepsilon,t)+[f_{\alpha}(\varepsilon)\\-\boldsymbol{\sigma}_{\mathrm{D}}(t)]\boldsymbol{\Lambda}_{\alpha}(\varepsilon)+\sum_{\alpha'}^{N_\alpha}\int\mathrm{d}\varepsilon'\boldsymbol{\varphi}_{\alpha\alpha'}(\varepsilon,\varepsilon',t),\end{aligned}\tag{5}$$

$$\begin{aligned}i\dot{\boldsymbol{\varphi}}_{\alpha,\alpha'}(\varepsilon,\varepsilon',t)=-[\varepsilon+\Delta_{\alpha}(t)-\varepsilon'-\Delta_{\alpha'}(t)]\cdot\boldsymbol{\varphi}_{\alpha,\alpha'}(\varepsilon,\varepsilon',t)\\+\boldsymbol{\Lambda}_{\alpha'}(\varepsilon')\cdot\boldsymbol{\varphi}_{\alpha}(\varepsilon,t)-\boldsymbol{\varphi}_{\alpha'}^{\dagger}(\varepsilon',t)\cdot\boldsymbol{\Lambda}_{\alpha}(\varepsilon),\end{aligned}\tag{6}$$

where $\boldsymbol{\Lambda}_{\alpha}(\varepsilon)$ is the linewidth function, $f_{\alpha}(\varepsilon)$ is the Fermi–Dirac function for lead α, $f_{\alpha}(\varepsilon)=1/(1+\exp[\beta(\varepsilon-\mu_{\alpha})])$, $\beta=1/k_{\mathrm{B}}T$, is the reciprocal temperature, and μ_{α} is the chemical potential for lead α. $\Delta_{\alpha}(t)$ is the time-dependent bias potential in lead α. We note that Eq. (5) (and Eq. 6 as well) is derived as follows: the time derivatives act on three parts in the RHS of Eq. (2): $\boldsymbol{G}_{\mathrm{D}}^{<,>}(t,\tau)$, $\boldsymbol{\Sigma}_{\alpha}^{<,>}(\varepsilon,\tau,t)$ and the upper limit of integration. The last part contributes to term 2 in Eq. (5); and the first two parts contribute to term 1 and term 3 due to the EOM of $\boldsymbol{G}_{\mathrm{D}}^{<,>}(t,\tau)$ and $\boldsymbol{\Sigma}_{\alpha}^{<,>}(\varepsilon,\tau,t)$. The derivation details for Eqs. (4)–(6) are given in Refs. [19, 20]. This is the RSDM-HEOM.

In Eqs. (4) and (5), there are integrals of $\boldsymbol{\varphi}_{\alpha}(\varepsilon,t)$ and $\boldsymbol{\varphi}_{\alpha\alpha'}(\varepsilon,\varepsilon',t)$ over energy ε or ε', which are very computationally expensive. In numerical calculations, Cauchy's residue theorem is used to transform these integrals into summations [20]. Here we show the brief steps.

Step (1): The integral for the steady-state self-energy $\overline{\boldsymbol{\Sigma}}_{\alpha}^{<,>}(\tau,t)$ is transferred into summation by the residue

theorem. For example,

$$\overline{\boldsymbol{\Sigma}}_{\alpha}^{<}(\tau-t)=\frac{i}{2\pi}\int_{-\infty}^{+\infty}f_{\alpha}(\varepsilon)\boldsymbol{\Lambda}_{\alpha}(\varepsilon)\,\mathrm{e}^{-i\varepsilon(\tau-t)}\,\mathrm{d}\varepsilon =\frac{i}{2\pi}\oint f_{\alpha}(z)\boldsymbol{\Lambda}_{\alpha}(z)\,\mathrm{e}^{-iz(\tau-t)}\,\mathrm{d}z. \tag{7}$$

This transformation holds only if the contribution from the contour integrals in the upper (or lower) complex plane tends to zero [20]. Using the multi-Lorentzian expansion to approximate the linewidth function ($\Lambda_{\alpha}(\varepsilon)\approx\sum_{d=1}^{N_d}(\eta_d/((\varepsilon-\Omega_d)^2+W_d^2))\overline{\Lambda}_{\alpha d}$), and the Padé spectrum decomposition to expand the Fermi–Dirac function [23] ($f_{\alpha}(z)=(1/(1+\exp(z)))\approx(1/2)+\sum_{p=1}^{N_p}(R_p/(z-z_p^{+}))+(R_p/(z-z_p^{-}))$), we may find the poles of the integrand in the contours and write down the residue summation form as:

$$\overline{\boldsymbol{\Sigma}}_{\alpha}^{<,>}(\tau-t)=\sum_{k}^{N_k}\boldsymbol{A}_{\alpha k}^{<,>\pm}\,\mathrm{e}^{\mp\gamma_{\alpha k}^{\pm}(t-\tau)},$$

where "+" and "−" correspond to different contours, due to the sign of $\tau-t$. The expressions for $\boldsymbol{A}_{\alpha k}^{<,>\pm}$ and $\gamma_{\alpha k}^{\pm}$ are given in Ref. [20].

Step (2): Adding a phase factor to Eq. (7), $\boldsymbol{\Sigma}_{\alpha}^{<,>}(\tau,t)$ can also be written as a residue summation:

$$\boldsymbol{\Sigma}_{\alpha}^{<,>}(t,\tau)=\mathrm{e}^{-i\int_{\tau}^{t}\Delta_{\alpha}(\xi)\,\mathrm{d}\xi}\cdot\overline{\boldsymbol{\Sigma}}_{\alpha}^{<,>}(t-\tau) =\sum_{k=1}^{N_k}\boldsymbol{\Sigma}_{\alpha k}^{<,>}(t,\tau). \tag{8}$$

Then, the integral of $\boldsymbol{\varphi}_{\alpha}(\varepsilon,t)$ is transformed into the summation form,

$$\begin{aligned}\int\mathrm{d}\varepsilon\cdot\boldsymbol{\varphi}_{\alpha}(\varepsilon,t)&=\int\mathrm{d}\varepsilon\cdot i\int_{-\infty}^{t}\mathrm{d}\tau[\boldsymbol{G}_{\mathrm{D}}^{<}(t,\tau)\cdot\boldsymbol{\Sigma}_{\alpha}^{>}(\varepsilon,\tau,t)\\&\quad-\boldsymbol{G}_{\mathrm{D}}^{>}(t,\tau)\cdot\boldsymbol{\Sigma}_{\alpha}^{<}(\varepsilon,\tau,t)]\\&=\sum_{k=1}^{N_k}\boldsymbol{\varphi}_{\alpha k}(t),\end{aligned} \tag{9a}$$

where

$$\begin{aligned}\boldsymbol{\varphi}_{\alpha k}(t)=i\int_{-\infty}^{t}\mathrm{d}\tau[&\boldsymbol{G}_{\mathrm{D}}^{<}(t,\tau)\cdot\boldsymbol{\Sigma}_{\alpha k}^{>}(\tau,t)\\&-\boldsymbol{G}_{\mathrm{D}}^{>}(t,\tau)\cdot\boldsymbol{\Sigma}_{\alpha k}^{<}(\tau,t)]\end{aligned} \tag{9b}$$

are called the discretized 1st-tier ADM.

Step (3): Similarly, the integral of $\boldsymbol{\varphi}_{\alpha\alpha'}(\varepsilon,\varepsilon',t)$ can be expressed as the summation form

$$\iint\mathrm{d}\varepsilon\,\mathrm{d}\varepsilon'\cdot\boldsymbol{\varphi}_{\alpha\alpha'}(\varepsilon,\varepsilon',t)=\sum_{k,k'}^{N_k}\boldsymbol{\varphi}_{\alpha k,\alpha' k'}(t) \tag{10a}$$

with each discretized term defined as

$$\begin{aligned}\boldsymbol{\varphi}_{\alpha k,\alpha' k'}(t)=i\int_{-\infty}^{t}\mathrm{d}t_1\int_{-\infty}^{t}\mathrm{d}t_2\{&[\boldsymbol{\Sigma}_{\alpha' k'}^{<}(t,t_1)\cdot\boldsymbol{G}_{\mathrm{D}}^{a}(t_1,t_2)\\&+\boldsymbol{\Sigma}_{\alpha' k'}^{r}(t,t_1)\cdot\boldsymbol{G}_{\mathrm{D}}^{<}(t_1,t_2)]\boldsymbol{\Sigma}_{\alpha k}^{>}(t_2,t)\\&-[\boldsymbol{\Sigma}_{\alpha' k'}^{>}(t,t_1)\cdot\boldsymbol{G}_{\mathrm{D}}^{a}(t_1,t_2)\\&+\boldsymbol{\Sigma}_{\alpha' k'}^{r}(t,t_1)\cdot\boldsymbol{G}_{\mathrm{D}}^{>}(t_1,t_2)]\boldsymbol{\Sigma}_{\alpha k}^{<}(t_2,t).\end{aligned} \tag{10b}$$

From these definitions, and with the similar time derivative approach for Eqs. (4) and (5), the following HEOM can be derived [20]

$$i\dot{\boldsymbol{\sigma}}(t)=[\boldsymbol{h}_{\mathrm{D}}(t),\boldsymbol{\sigma}(t)]-\sum_{\alpha}^{N_\alpha}\sum_{k=1}^{N_k}(\boldsymbol{\varphi}_{\alpha k}(t)-\boldsymbol{\varphi}_{\alpha k}^{\dagger}(t)), \tag{11}$$

$$\begin{aligned}i\dot{\boldsymbol{\varphi}}_{\alpha k}(t)=&[\boldsymbol{h}_{\mathrm{D}}(t)-i\gamma_{\alpha k}^{+}-\boldsymbol{\Delta}_{\alpha}(t)]\boldsymbol{\varphi}_{\alpha k}(t)\\&-i[\boldsymbol{\sigma}(t)\boldsymbol{A}_{\alpha k}^{>+}+\overline{\boldsymbol{\sigma}}(t)\boldsymbol{A}_{\alpha k}^{<+}]+\sum_{\alpha'}^{N_\alpha}\sum_{k'=1}^{N_k}\boldsymbol{\varphi}_{\alpha k,\alpha' k'}(t),\end{aligned} \tag{12}$$

$$\begin{aligned}i\dot{\boldsymbol{\varphi}}_{\alpha k,\alpha' k'}(t)=&-[i\gamma_{\alpha k}^{+}+\boldsymbol{\Delta}_{\alpha}(t)+i\gamma_{\alpha' k'}^{-}-\boldsymbol{\Delta}_{\alpha'}(t)]\cdot\boldsymbol{\varphi}_{\alpha k,\alpha' k'}(t)\\&+i(\boldsymbol{A}_{\alpha' k'}^{>-}-\boldsymbol{A}_{\alpha' k'}^{<-})\boldsymbol{\varphi}_{\alpha k}(t)-i\boldsymbol{\varphi}_{\alpha' k'}^{\dagger}(t)(\boldsymbol{A}_{\alpha k}^{>+}-\boldsymbol{A}_{\alpha k}^{<+}),\end{aligned} \tag{13}$$

where $\overline{\boldsymbol{\sigma}}(t)=\boldsymbol{I}-\boldsymbol{\sigma}(t)$ and $N_k=N_d+N_p$ is the total number of the Lorentzian and Padé poles in the contour. We term Eqs. (11)–(13) and their solutions as the Lorentzian–Padé decomposition scheme.

The accuracy of this scheme has been tested with a one-level system. The transient current agrees with that in reference [24] very well and thus the accuracy of this HEOM method is validated.

2.2 Initial-state calculation In Ref. [20], we set the time derivatives of all the density (and auxiliary density) matrices to be zero and solve the matrix equations for the initial values or the static solutions. In the case of a large system with a large number of orbitals in the device region or a large number of expansion terms in the spectra of the leads, it is very time consuming to solve such a huge system of linear equations, even with the sparse matrix technique. In this paper, we develop a new method for the initial-state calculation.

First, we calculate density matrix $\boldsymbol{\sigma}_{\mathrm{D}}$ by integrating the lesser-Green's function in energy domain

$$\boldsymbol{\sigma}_{\mathrm{D}}=\frac{-1}{\pi}\int_{-\infty}^{+\infty}f_{\mathrm{P}}(E)\mathrm{Im}[\boldsymbol{G}_{\mathrm{D}}^{r}(E)]\,\mathrm{d}E, \tag{14}$$

where $f_{\mathrm{P}}(E)$ is the Padé-expansion approximant for the Fermi–Dirac function,

$$f_{\mathrm{P}}(E) = \frac{1}{2} + \frac{1}{\beta}\sum_{p=1}^{N_p}\left(\frac{R_p}{E-\mu-z_p^+/\beta} + \frac{R_p}{E-\mu-z_p^-/\beta}\right). \tag{15}$$

The retarded Green's function is calculated by

$$\boldsymbol{G}_{\mathrm{D}}^r(E) = [E\cdot\boldsymbol{I} - \boldsymbol{h}_{\mathrm{D}} - \boldsymbol{\Sigma}_{\mathrm{Lrz}}^r(E)]^{-1}, \tag{16}$$

where $\boldsymbol{\Sigma}_{\mathrm{Lrz}}^r(E)$ is the self-energy matrix fitted by the multi-Lorentzian expansion (see details in Section 2.3),

$$\begin{aligned}\boldsymbol{\Sigma}_{\mathrm{Lrz}}^r(E) &= \boldsymbol{\Sigma}_{\mathrm{Lrz}}^R(E) + i\cdot\boldsymbol{\Sigma}_{\mathrm{Lrz}}^I(E)\\ &= \sum_{d=1}^{N_d}\left[\frac{-\boldsymbol{A}_d/W_d\cdot(E-\Omega_d)}{(E-\Omega_d)^2+W_d^2}\right.\\ &\quad\left. + i\cdot\frac{\boldsymbol{A}_d}{(E-\Omega_d)^2+W_d^2}\right].\end{aligned} \tag{17}$$

As $\boldsymbol{G}_{\mathrm{D}}^r(E)$ has many singularities near the real axis, which makes a lot of narrow peaks for $\boldsymbol{G}_{\mathrm{D}}^r(E)$ on the real axis, so the accurate numerical integration in Eq. (14) needs very fine energy grids. However, if the integrand in Eq. (14) has analytic continuation into the upper complex plane, $\boldsymbol{G}_{\mathrm{D}}^r(E)$ can behave very smoothly and the integration on the upper-complex-plane contour is much easier [25]. Then we can do the integral with the help of the residue theorem [26]. To construct an analytic integrand, we rewrite Eq. (14) as

$$\begin{aligned}\boldsymbol{\sigma}_{\mathrm{D}} &= \frac{-1}{\pi}\,\mathrm{Im}\left\{\left[\int_{-\infty}^{+\infty} f_{\mathrm{P}}(E)\cdot\boldsymbol{G}_{\mathrm{D}}^r(E)\,\mathrm{d}E\right]\right\}\\ &= \frac{-1}{\pi}\,\mathrm{Im}[\boldsymbol{I}_\sigma],\end{aligned} \tag{18}$$

where $\boldsymbol{I}_\sigma$ is analytic and can be calculated by the contour integral: we extend the integral to the complex plane and use the Cauchy's residue theorem,

$$\begin{aligned}\boldsymbol{I}_\sigma &= \int_{-\infty}^{+\infty} f_{\mathrm{P}}(E)\boldsymbol{G}_{\mathrm{D}}^r(E)\,\mathrm{d}E\\ &= -\int_{C_R} f_{\mathrm{P}}(z)\boldsymbol{G}_{\mathrm{D}}^r(z)\,\mathrm{d}z + 2\pi i\cdot\sum_k \mathrm{Residue}(k),\end{aligned} \tag{19}$$

where $C_R = C_{R1} + C_{R2} + C_{R3}$, are the contours in the upper complex plane, as shown in Fig. 1. Residue(k) is the residue of the kth pole in the contour. As all the poles of $\boldsymbol{G}^r(E)$ are in the lower complex plane [25], only the Padé poles of $f_{\mathrm{P}}(E)$ are accounted for here.

Secondly, we calculate the 1st-tier ADM $\boldsymbol{\varphi}_{\alpha k}(t)$ from its definition (Eq. 9b). In the equilibrium state ($\Delta_\alpha(t) = 0$), we have

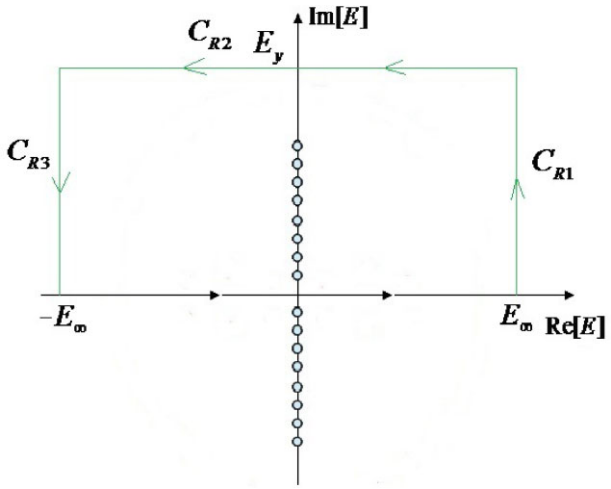

Figure 1 The contour path and poles of the contour integral in Eq. (19). The filled dots on the y-axis represent the Padé poles.

$$\boldsymbol{G}_{\mathrm{D}}^x(t-\tau) = \frac{-is_x}{\pi}\int_{-\infty}^{+\infty} f_x(E)\,\mathrm{Im}[\boldsymbol{G}_{\mathrm{D}}^r(E)]\,\mathrm{e}^{-iE(t-\tau)}\,\mathrm{d}E,$$

$$\boldsymbol{\Sigma}_{\alpha k}^x(\tau-t) = \boldsymbol{A}_{\alpha k}^{x,+}\,\mathrm{e}^{-\gamma_{\alpha k}^+(t-\tau)}, \quad \tau - t < 0,$$

where $x = <$ or $>$, $s_< = 1$, $s_> = -1$, $f_<(E) = f_{\mathrm{P}}(E)$, $f_>(E) = 1 - f_{\mathrm{P}}(E)$. Substituting them into the $\boldsymbol{\varphi}_{\alpha k}(t)$ definition for an equilibrium state as follows (in which case all the two-time quantities are reduced to the one-time quantities)

$$\begin{aligned}\boldsymbol{\varphi}_{\alpha k}(t) = i\int_{-\infty}^{t}\mathrm{d}\tau[&\boldsymbol{G}_{\mathrm{D}}^<(t-\tau)\cdot\boldsymbol{\Sigma}_{\alpha k}^>(\tau-t)\\ &- \boldsymbol{G}_{\mathrm{D}}^>(t-\tau)\cdot\boldsymbol{\Sigma}_{\alpha k}^<(\tau-t)]\end{aligned}$$

and doing the time integration, we achieve the following result after some derivations

$$\begin{aligned}&\boldsymbol{\varphi}_{\alpha k}(t=0)\\ &= \frac{1}{\pi}\int_{-\infty}^{+\infty}\mathrm{d}E\,\frac{\mathrm{Im}[\boldsymbol{G}_{\mathrm{D}}^r(E)]\cdot[\boldsymbol{A}_{\alpha k}^{>,+}f_{\mathrm{P}}(E) + \boldsymbol{A}_{\alpha k}^{<,+}(1-f_{\mathrm{P}}(E))]}{\gamma_{\alpha k}^+ + iE}.\end{aligned} \tag{20}$$

Similarly, this integral can be calculated by the residue theorem with much high accuracy and small computation load. Appendix B gives the details of this calculation.

Thirdly, after solving out $\boldsymbol{\sigma}_{\mathrm{D}}$ and $\boldsymbol{\varphi}_{\alpha k}$, the 2nd-tier ADM ($\boldsymbol{\varphi}_{\alpha k,\alpha' k'}$) in the equilibrium state is obtained directly from Eq. (13),

$$\boldsymbol{\varphi}_{\alpha k,\alpha' k'} = \frac{1}{\gamma_{\alpha k}^+ + \gamma_{\alpha k}^-}\left[(\boldsymbol{A}_{\alpha' k'}^{>,-} - \boldsymbol{A}_{\alpha' k'}^{<,-})\boldsymbol{\varphi}_{\alpha k} - \boldsymbol{\varphi}^\dagger{}_{\alpha' k'}(\boldsymbol{A}_{\alpha k}^{>,+} - \boldsymbol{A}_{\alpha k}^{<,+})\right]. \tag{21}$$

2.3 Automatic Lorentzian fitting In our previous HEOM calculations we did the Lorentzian fitting for the surface Green's function of the leads [20]. However, we find that for a real system or a large model system, the number of nonzero self-energy functions is much less than that of the surface Green's functions. So in this paper we directly fit the imaginary part of the self-energy matrix functions

$$\mathrm{Im}\left[\boldsymbol{\Sigma}_{\alpha}^{r}(E)\right]=\sum_{d=1}^{N_d}\frac{1}{(E-\Omega_d)^2+W_d^2}\cdot\boldsymbol{A}_{\alpha,d}. \tag{22}$$

The real part of the self-energy is obtained from its imaginary part by the Kramers–Kronig relation [20], as shown in Eq. (17), and the linewidth function is related to the self-energy by $\boldsymbol{\Lambda}_{\alpha}(E)=-2\,\mathrm{Im}[\boldsymbol{\Sigma}_{\alpha}^{r}(E)]$.

We use a least-square (LS) solver MINPACK for the Lorentzian fitting. MINPACK is a software package for solving the nonlinear equations and nonlinear LS problems within Powell's hybrid algorithm and the Levenberg–Marquardt algorithm, respectively [27]. In our problem, we use the subroutine "lmdif1" in MINPACK to find all the values of $x_j\,(j\in[1,n])$ for m "deviation functions" f_i (where $i\in[1,m]$, $m>n$) by minimizing the following formula

$$\min\left\{\sum_{i=1}^{m}f_i(\boldsymbol{x})^2:\boldsymbol{x}\in R^n\right\}. \tag{23}$$

These deviation-functions origin from the fitting object and the functions with fitting parameters. Iteration is employed to find the optimized fitting parameters from some initial guess.

In our fitting process, a series of discrete self-energy ($S_i=\mathrm{Im}[\boldsymbol{\Sigma}_{\alpha}^{r}(E_i)]$) values is calculated by the principle-layer method [28] for the fitting object. The deviation function is obtained after setting the initial fitting parameters $\{x_j\}$, $f_i(x_1,x_2,\ldots,x_n)=S_i-L(E_i,x_1,x_2,\ldots,x_n)$, where L is the multi-Lorentzian function,

$$\begin{aligned}L(E,x_1,x_2,\ldots,x_n)&=L(E,\Omega_1,W_1,\eta_1,\Omega_2,W_2,\eta_2,\ldots,\eta_{N_d})\\&=\sum_{d}^{N_d}\frac{\eta_d}{(E-\Omega_d)^2+W_d^2}\end{aligned}$$

and $\{x_j\}$ correspond to the Lorentzian parameters: $\{\Omega_d,W_d,\eta_d\}$, so $n=3N_d$.

Since n is very large, $\sum_{i=1}^{m}f_i(x)^2$ has a lot of local minima in such a high-dimensional solution space ($\{x_j,x\in R^n\}$). Different initial guess leads to different optimal solution for fitting parameters. A proper setting of the initial Lorentzian parameters is very crucial for a good fitting result. Some programs are employed to automatically find the initial Lorentzian parameters.

After fitting all the self-energy (imaginary part) curves (with the total number N_z), we have

$$\begin{aligned}\mathrm{Im}\left[\Sigma^{(1)}(E)\right]&=\sum_{d=1}^{N^{(1)}}\frac{\eta_d^{(1)}}{(E-\Omega_d^{(1)})^2+W_d^{(1)2}};\\\mathrm{Im}\left[\Sigma^{(2)}(E)\right]&=\sum_{d=1}^{N^{(2)}}\frac{\eta_d^{(2)}}{(E-\Omega_d^{(2)})^2+W_d^{(2)2}};\ldots,\end{aligned}$$

where $N^{(k)}$ is the number of Lorentzians used in the kth curve; $\Omega_d^{(k)}$, $W_d^{(k)}$, and $\eta_d^{(k)}$ are the Lorentzian center, width and amplitude of the dth Lorentzian in the kth curve respectively. For small systems, we directly combine all these Lorentzians into Eq. (22) by defining a global Lorentzian index, such as that in Ref. [17]. The total number of Lorentzians is $N_{d0}=\sum_{k=1}^{N_z}N^{(k)}$. We term this as the direct (element by element) fitting method.

For large systems, the number of Lorentzians has to be minimized to reduce the computational cost. Since the shape of a Lorentzian function is mainly determined by its center (Ω) and width (W), we represent each Lorentzian function as one point on a $W-\Omega$ diagram. If some points are closely located, we combine them into a new point (see Appendix B for the combination algorithm). After this combination, the number of Lorentzians is reduced from N_{d0} to N_{d1}. Then we refit all these curves with these combined Lorentzians for optimization.

One refitting way is to fix the combined Lorentzian parameters: $\{W_d\}$ and $\{\Omega_d\}$, and only to fit their amplitudes $\{\eta_d^{(k)}\}$, where $\eta_d^{(k)}$ is for the dth Lorentzian (with the total number N_{d1}) in the kth curve (with the total number N_z). The number of fitting parameters is N_zN_{d1}. It is noted that in each curve fitting, $\{\Omega_d\}$ and $\{W_d\}$ come from the combined Lorentzian points, which are much more than those in the original fitting. So the fitting results are much better. We term this process the combined fitting method.

Another refitting way is to treat the combined Lorentzian parameters $\{W_d\}$ and $\{\Omega_d\}$ as the initial fitting values. They are to be optimized together with the amplitudes. To do this, we combine all the curves into one large fitting object S ($S=\{\mathrm{Im}[\Sigma_1(E_i)],Im[\Sigma_2(E_i)],\ldots,\mathrm{Im}[\Sigma_{N_z}(E_i)]\}$) and we can simultaneously fit all these curves with a large set of Lorentzian parameters $\{\Omega_d,W_d,\eta_d^{(k)}\}$. In other words, we minimize the following sum of squared residues (SSR) in the LS process:

$$\begin{aligned}\mathrm{SSR}=&\sum_{i=1}^{N}\left|S_1(i)-\sum_{d=1}^{N_{d1}}\frac{\eta_d^{(1)}}{(E_i-\Omega_d)^2+W_d^2}\right|^2\\&+\sum_{i=1}^{N}\left|S_2(i)-\sum_{d=1}^{N_{d1}}\frac{\eta_d^{(2)}}{(E_i-\Omega_d)^2+W_d^2}\right|^2+\cdots\\&+\sum_{i=1}^{N}\left|S_{N_z}(i)-\sum_{d=1}^{N_{d1}}\frac{\eta_d^{(N_z)}}{(E_i-\Omega_d)^2+W_d^2}\right|^2.\end{aligned}$$

In this refitting process not only the N_zN_{d1} amplitudes $\{\eta_d^{(k)}\}$, but also the centers $\{\Omega_d\}$ and widths $\{W_d\}$ are to be adjusted simultaneously. The number of fitting parameters is

$2N_{d1} + N_z N_{d1}$. We term this type of fitting the global fitting method.

In most cases the result from global fitting is best with the same Lorentzians, among these three fitting schemes. But since the fitting object is very large (N curves with $N \cdot N_z$ data points or object functions for each), the computational time for the fitting is usually much longer than the other two schemes.

3 Numerical implementation GNR are used in the HEOM calculations. The nearest-neighbor tight-binding model is employed with one p_z orbital for each carbon atom. The calculations for the armchair GNR (AGNR) and zigzag GNR (ZGNR) are treated separately in the following section.

3.1 Armchair GNR calculation In this section, AGNR ($M = 16$) is used in the HEOM calculation, where M means the number of atoms in one GNR unit cell.

Figure 2a shows the atomic structure and the labels of this AGNR. With the nearest TB model, there are 20 nonzero self-energy curves. To reduce the number of Lorentzians, we only fit for 10 different curves because there are some degenerate curves due to the geometry symmetry. Figure 2b shows 4 of them (the imaginary part), which are of complicated shapes. A lot of Lorentzians are needed for the fitting and the accuracy of HEOM calculation depends on the fitting results. Two fitting schemes are used, as shown below.

3.1.1 The delta-function-like peak in the self-energy curve Before fitting, we notice that near the energy range $E = 0$ eV, the real part of some self-energy is very large (which can be infinite) while the imaginary part is of finite values (as shown in Fig. 2c). The relation of these real and imaginary parts of self-energy differs from the Kramers–Kronig relation that they should obey [29].

To find out the discrimination reason in the self-energy curve, we use a very fine sampling interval that is comparable to or smaller than the small imaginary number ($i^*\eta$) in the surface Green's function calculation [4]. We find in each imaginary-part curve there exist a "delta-function-like" peak in the range of $E = 0$ eV (inset of Fig. 2c). This is reasonable because if the $\mathrm{Im}[\Sigma(E)]$ curve behaves as a delta function near $E = 0$ eV, the $\mathrm{Re}[\Sigma(E)]$ curve has the fractional-function-like peak ($1/(E - E_0)$). They are related by the Kramers–Kronig relation,

$$\mathrm{Im}[\Sigma(E)] = -A_{xx}\pi \cdot \delta(E - E_0),$$
$$\mathrm{Re}[\Sigma(E)] = \frac{1}{\pi}\int \frac{\mathrm{Im}[\Sigma(E')]}{E' - E}\,\mathrm{d}E' = \frac{A_{xx}}{E - E_0}.$$

So in Fig. 2c the $\mathrm{Re}[\Sigma(E)]$ has a fractional-function-like singularity near $E = E_0 = 0$ eV. This explains the large values of the real-part curves near $E = 0$ eV.

We find this type of delta-function-like peaks exist in the spectra of many types of AGNR leads. They also exist in the real system such as the carbon atom chain lead in the density-functional tight-binding (DFTB) model. To further investigate this state, we plot the local density of states (LDOS) distribution of this semi-infinite AGNR at $E = 0$ eV, as shown in Fig. 2d. We see that the LDOS decays from the terminal side to the inner region (the LDOS on some atom sites does not decay to zero, but remains constant, which corresponds to the normal state near $E = 0$ eV). So we can see that this delta-function-like state is a surface state with a discrete energy level [30]. This state is similar to the famous "edge state" in ZGNR, which has a very sharp density-of-state peak at $E = 0$ eV [31]. In fact, the atomic structure on the terminal side of AGNR is also zigzag. But the surface state in our AGNR has a discrete energy, not the continuous spectrum for the edge state. From our knowledge, this surface state in the semi-finite AGNR is reported for the first time.

Now we return to the fitting calculation. A narrow Lorentzian (i.e., the width value W in the Lorentzian expression is very small, such as 0.00001 eV) has to be used to stand for such delta-function-like peaks. To evaluate the amplitude (A) of this narrow Lorentzian, the following steps are used: (i) Obtain A_{xx} from the real-part curve ($A_{xx}/(E - E_0)$) numerically. (ii) Assume the energy integrals for the delta-function-like peak and the narrow Lorentzian are equal,

$$\frac{-A\pi}{W} = \int_{-\infty}^{+\infty} \frac{-A}{(E - E_0)^2 + W^2}\,\mathrm{d}E = \int_{-\infty}^{+\infty} -A_{xx}\pi \cdot \delta(E - E_0)\,\mathrm{d}E = -A_{xx}\pi.$$

Thus, the Lorentzian amplitude is obtained: $A = A_{xx}W$.

3.1.2 Combined fitting and the HEOM calculation We use the combined fitting method to fit all the self-energy curves of this AGNR. After the initial fitting of all the 10 self-energy curves, totally 130 Lorentzians are obtained. Then, we combine the closed Lorentzian points to reduce the number of Lorentzians. Figure 3a shows a magnified part of this diagram, where the black dots represent the original Lorentzian points and the red dots represent the combined points. The combination algorithm can be seen in Appendix C.

With the combination parameter $r = 0.4$, the number of Lorentzian points are reduced to 55. With these 55 pairs of $\{\Omega_d\}$ and $\{W_d\}$, we optimize the amplitudes $\{\eta_d^{(k)}\}$ to refit all the curves. Figure 3b shows one self-energy curve of the refitting results. We see both the imaginary and the real parts of $\Sigma_{1,1}$ are fitted very well.

After fitting all the 10 self-energy curves, we use the NEGF formula [4] to calculate the transmission spectrum, which agrees well with the accurate calculation, as shown in Fig. 3c.

The contour integral method in Section 2.2 is used to evaluate the initial state. By propagating the HEOM starting from initial state with the 4th-order Runge–

Kutta algorithm, the dynamic currents of this system are calculated, as shown in Fig. 3d. A bias voltage of 2.0 V is applied across the device region symmetrically and the onsite energy changes linearly between the two leads. The bias voltage varies as a step function at $t = 0$ in time domain. We see after some oscillations in the early time, the currents tend to a steady value at about 4 fs.

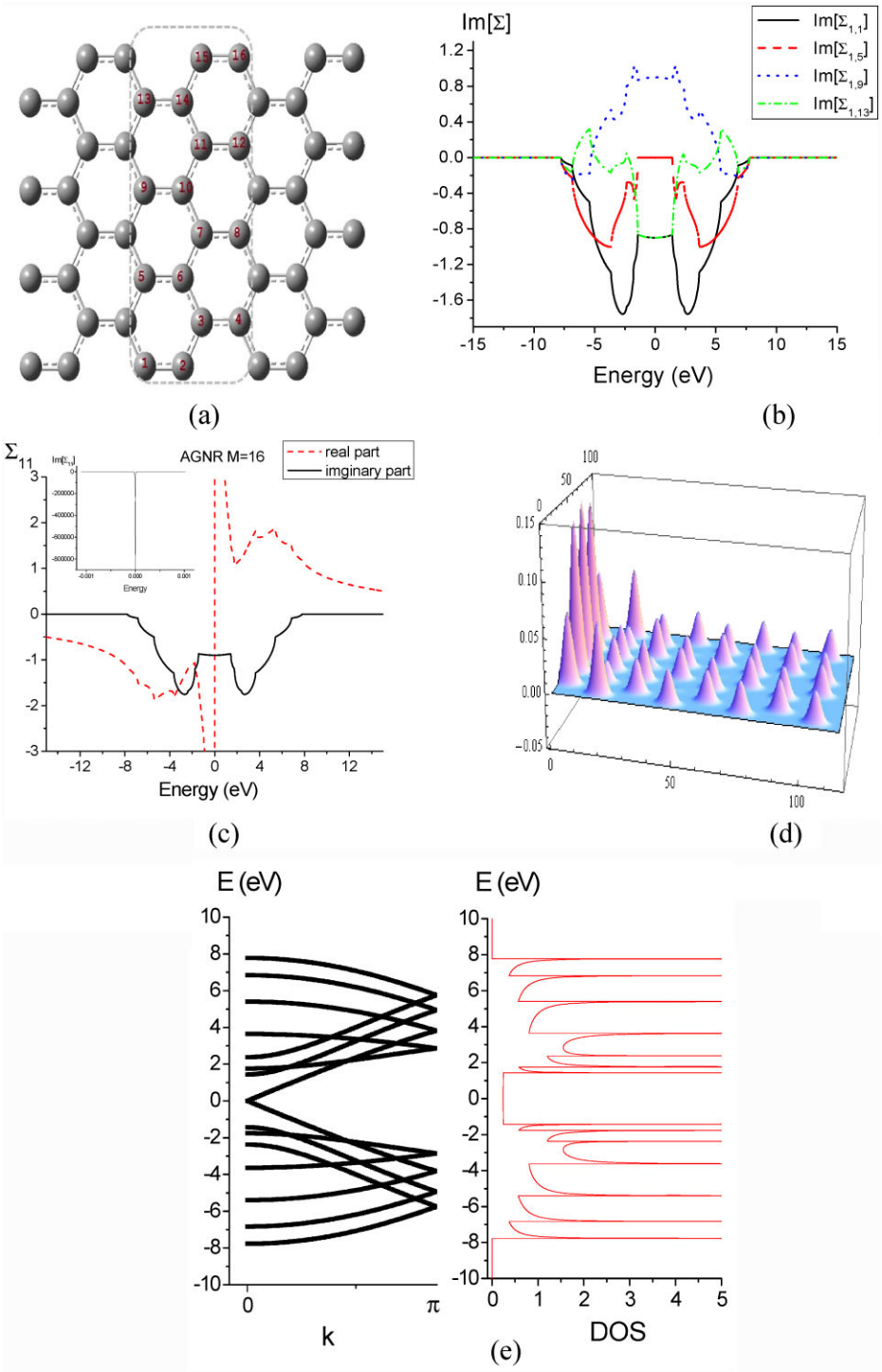

Figure 2 (a) Atomic structure and labels of AGNR ($M = 16$). The rectangular region denotes the device. (b) Four examples of self-energy curves (imaginary part) of this AGNR. (c) The imaginary and real part of self-energy ($\Sigma_{1,1}$). The inset shows the magnified delta-function-like peak near $E = 0$ eV. (d) The local density of states of the semi-infinite AGNR lead (left is the terminal side). (e) The band structure (left) and the density of states (right) of the AGNR ($M = 16$).

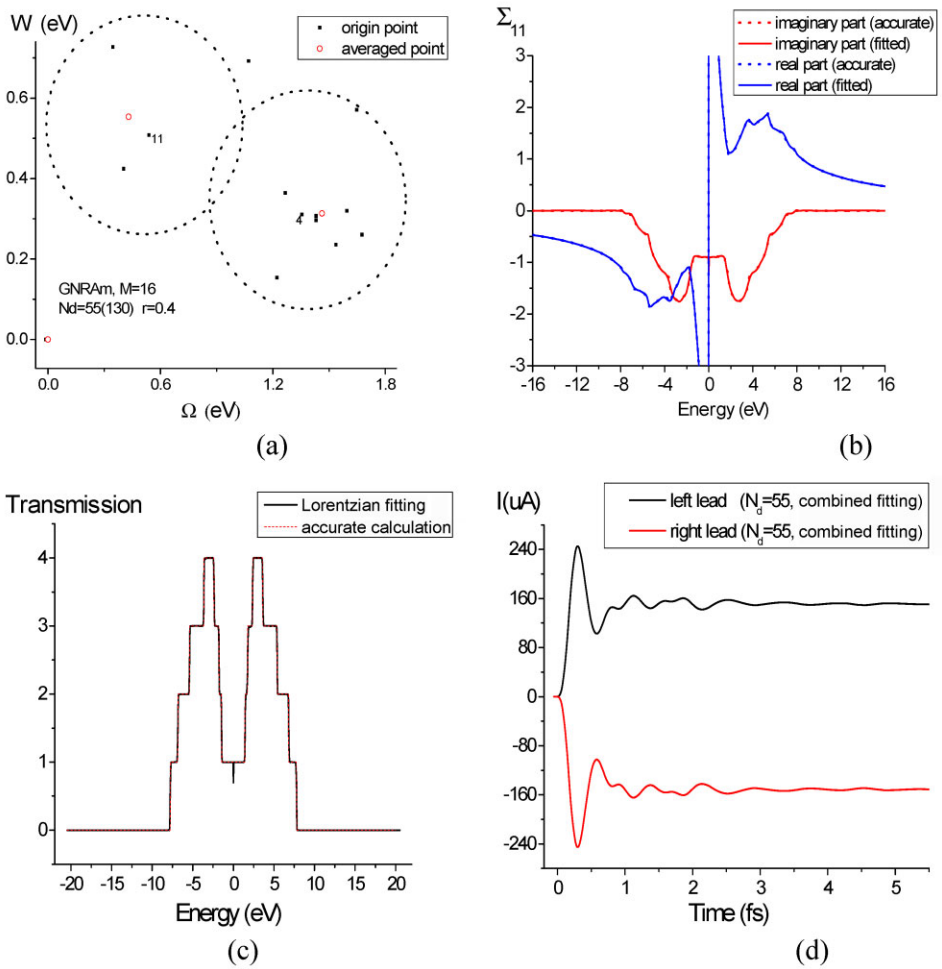

Figure 3 The fitting and HEOM calculation results for AGNR ($M = 16$). (a) Combination process for the Lorentzian points in $W - \Omega$ diagram. The originally fitted Lorentzians are represented as the black square dots and combined points are represented as the red round dots. The two dotted circles mean the combination region for each center points (labeled as 4 and 11). (b) The fitted (solid) and accurate (dotted) self-energy curve $\Sigma_{1,1}$ with the real (blue) and imaginary (red) part. (c) The transmission spectrum with accurate calculation (dotted line) and the combined fitting method (solid line). (d) The transient currents of left lead (black solid line) and right lead (red solid line) under a bias voltage of 2.0 V. The dotted line is the left current calculated from the global fitting method, as discussed in Section 2.3.

3.2 Zigzag-type GNR calculation Now, we turn to the zigzag-type GNR. For ZGNR, there are several new properties in the Lorentzian fitting and HEOM calculations, as will be detailed below.

3.2.1 Global fitting and bandstructure calculation Figure 4a shows the atomic structure of ZGNR ($M = 8$). Figure 4b shows the band structure (left) and some LDOS spectra (right) of this ZGNR. We see in the zero-energy range there are two flat bands near the Brillouin zone boundary, which corresponds to the edge state of ZGNR [31]. Due to this flat band, the group velocity at zero energy vanishes, as a result of strong Bloch wave scattering. The LDOS on the edge atoms (atom 1 or 8) also has a narrow and largest peak in this zero-energy range. This narrow peak results in the large overshooting current, which will be discussed later.

Here we show the fitting details of this ZGNR ($M = 8$). Twelve different self-energy curves are obtained from 20 nonzero self-energy elements. In each curve there exist several very narrow peaks. These peaks can only be fitted by very narrow Lorentzians. Some fitting skill is used for fitting these narrow peaks (see details in Appendix C).

With the global fitting method stated in Section 2.3, all these 12 different self-energy (imaginary part) curves are fitted by 127 Lorentzians. Figure 4c shows the first five curves for both the imaginary part (upper panel) and the real part (lower panel) with the shifted energy.[1] We see the fitted curves agree well with the accurate ones. Using these fitting

[1] In order to show all these curves in one plot, the shifted energy is used: $E_{\text{shift}} = E(k) + (k-1) \cdot E_{\text{width}}$, where $E(k)$ means the energy for the kth curve ($k = 1, \ldots, 12$); $E_{\text{width}} = E_{\max} - E_{\min}$ is the energy range of each self-energy curve.

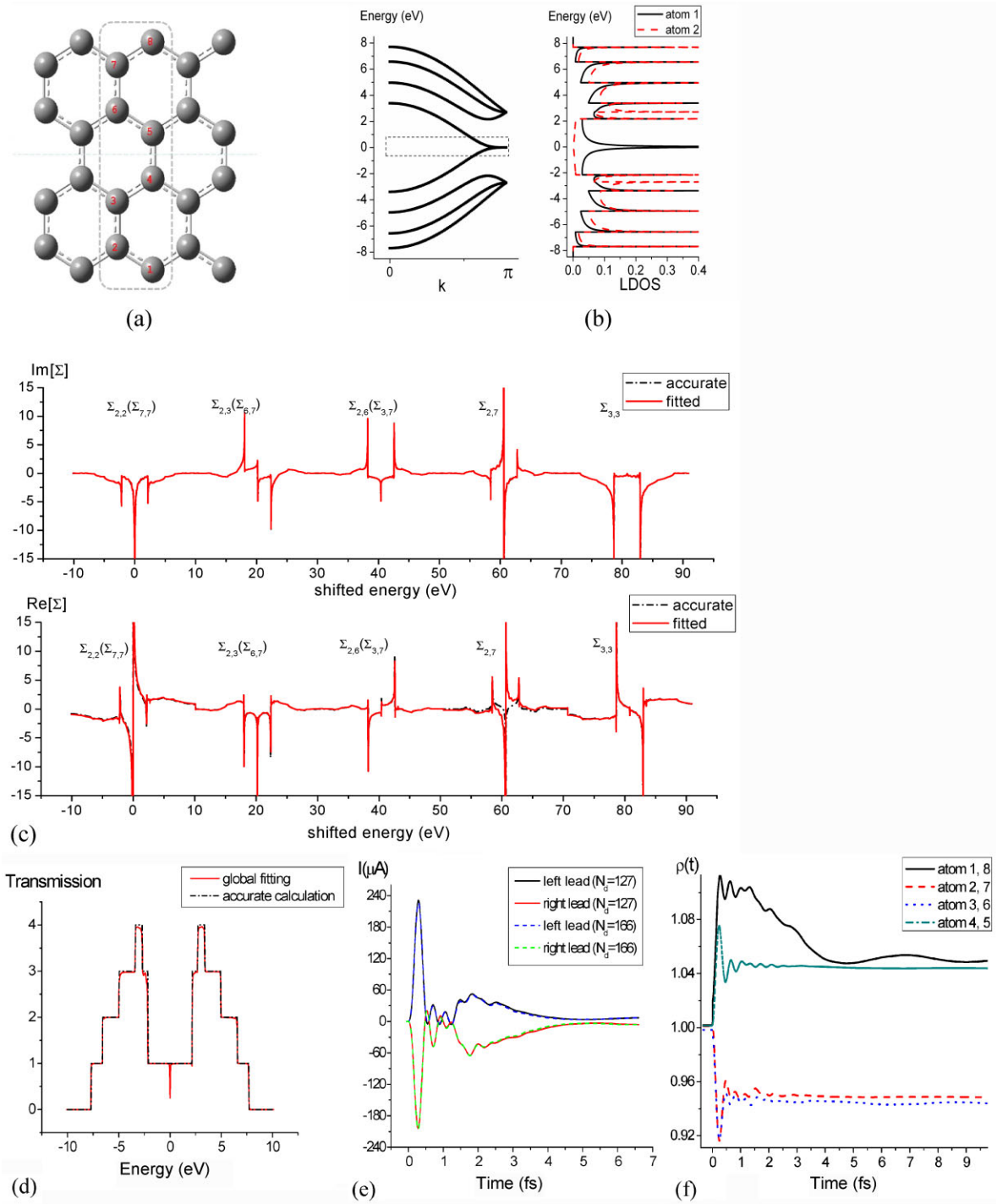

Figure 4 (a) Atomic structure and labels of ZGNR ($M = 8$). The rectangular box denotes the device region. The dotted line in the middle is the mirror symmetric line (see the even–odd effect in the next part). (b) Band structure (left) and LDOS (right) of this ZGNR ($M = 8$). The dashed rectangular box denotes the energy window for the subbands to join the transport under a small bias voltage. (c) First 5 different self-energy curves of ZGNR ($M = 8$) from the accurate calculation (dotted line) and the global Lorentzian fitting method (solid line). The upper panel is for the imaginary part and the lower panel is for the real part. (d) The transmission spectrum of ZGNR ($M = 8$) from the global Lorentzian fitting (solid line) and accurate calculation (dotted line). (e) The transient current of ZGNR ($M = 8$) under a voltage of 1.0 V. The solid and dotted lines are for two sets of Lorentzian parameters. (f) The charges vs. time curves on different atoms for this ZGNR ($M = 8$) system with a bias voltage of 1.0 V. Here are only 4 different curves due to the symmetry of this system. In (e) and (f), the bias voltage is step-wise turned on at time $t = 0$ and symmetrically applied on two leads. The onsite energies in the device region are linearly changed between the two leads.

parameters, the transmission spectrum is calculated, which also agrees well with the accurate one, as shown in Fig. 4d.

3.2.2 Dynamic currents and charges in ZGNR After the initial-state calculation, the dynamic current (Fig. 4e) and charges on different atoms (Fig. 4f) under a step-wise bias voltage (1.0 V) are calculated. In Fig. 4e, we also use another set of fitting parameters (166 Lorentzians with combination parameter $r = 0.1$). We see these two sets of parameters generate very close current curves, which ensures the convergence of our Lorentzian fitting scheme.

For the current of this ZGNR, there exist very large oscillations at the initial stage. Then, the currents slowly decay and reach to very small (close to zero) values after about 5 fs. For the charge curves, we see that in the initial time interval the charges on the right-side atoms (atoms 1, 4, 5, and 8) rapidly jump to high values and then decay to lower values with several oscillations (very slowly for atoms 1

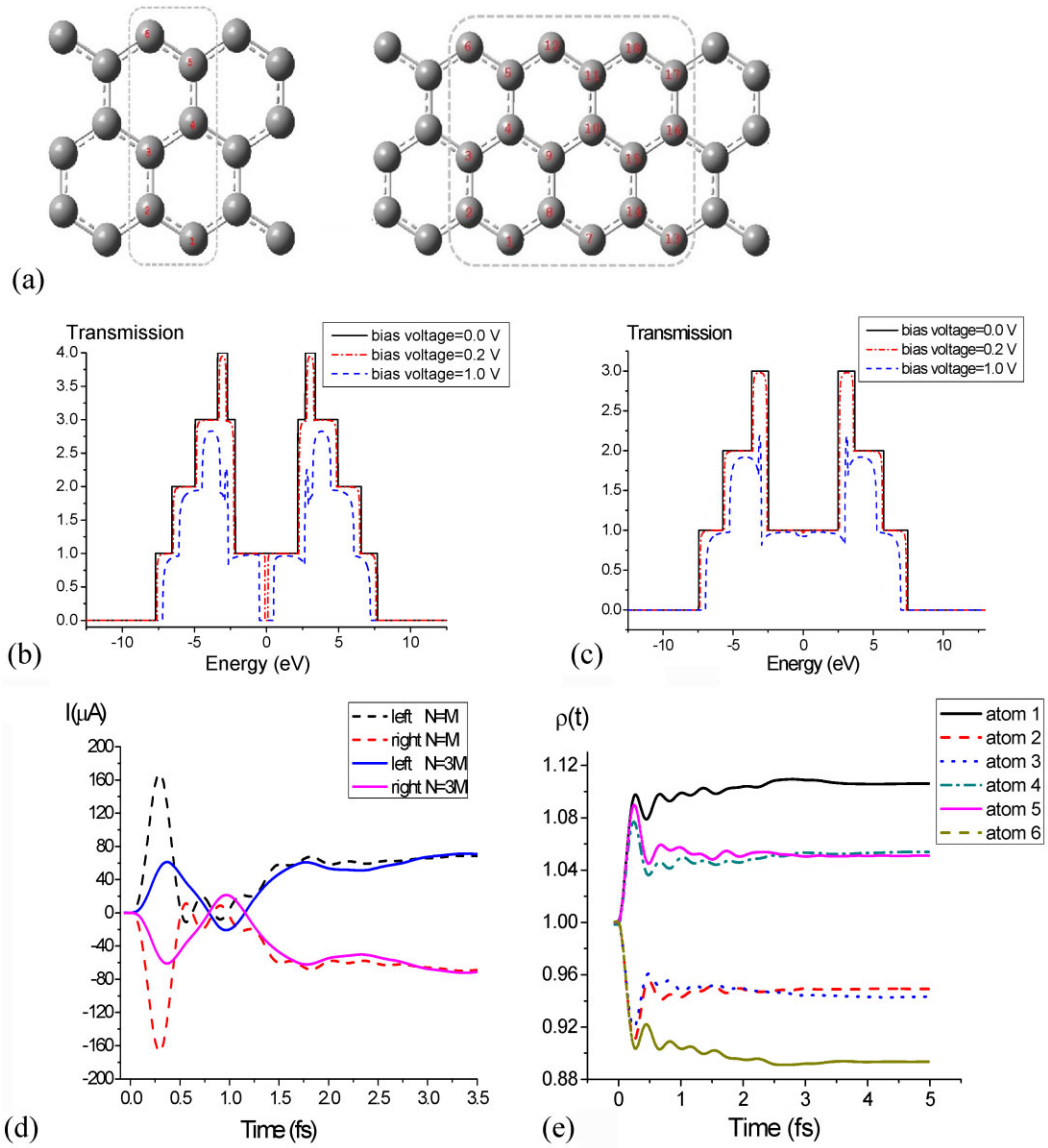

Figure 5 (a) Atomic structure and labels of ZGNR ($M = 6$). The left (right) part shows the narrow (wide) device case. The rectangular box denotes the device region. (b) The transmission spectra of ZGNR ($M = 8$) with the bias voltages of 0 V (solid line), 0.2 V (dotted line) and 1.0 V (dashed line). (c) The transmission spectra of ZGNR ($M = 6$) with the bias voltages of 0 V (solid line), 0.2 V (dotted line) and 1.0 V (dashed line). (d) The transient current of ZGNR ($M = 6$) under a bias voltage of 1.0 V. The dashed lines are for the narrow device ($N = 6$) and the solid lines are for the wide device ($N = 18$). (e) The charges vs. time curves on different atoms for the ZGNR ($M = 6$) system (narrow device) with a bias voltage of 1.0 V. In (d) and (e), the bias voltage is step-wise turned on at time $t = 0$ and symmetrically applied on two leads. The onsite energies in the device region are linearly changed between two leads.

and 8). On the left-side atoms (atoms 2, 3, 6, and 7), the charge decreases under the bias at first and then return to higher values.

Here we give some analysis of these curves. During the initial time interval (about 0.25 fs), the rapid and large current oscillation is from the large overshooting phenomenon. This overshooting current is much stronger compared with the current of AGNR (Fig. 3d). This is because near the Fermi level ($E = 0$ eV), the edge-state band in ZGNR makes a very narrow DOS spectrum both in device (Fig. 4b) and lead region (Fig. 4c, $\mathrm{Im}[\Sigma_{11}]$), compared to the very flat spectrum of AGNR (see Fig. 2b and e). This narrow spectrum peak makes the incident electron wave very difficult to dissipate to the right lead. A similar case is also shown in our previous paper for the 1D model system with poor contact [20]. (Here it is the GNR configuration or the edge state that makes such "poor contact" effect.) So when the incident current reaches the device–lead boundary most of them will be reflected and oscillate for several times until it dissipates into the right lead to form a steady state. The charge curves also show such large overshooting process. We note that during the overshooting, the charges on atoms 1 and 8 reach a much high value and decay very slowly. This is because under a small bias only the edge states go into the device (see the dashed box in the band structure of Fig. 4b) and the charges on the edge atoms increase mostly (see the large LDOS of atoms 1 and 8 in Fig. 4b). The low steady current is due to the zero transmission (see the following), which makes it very difficult for the additional charges to dissipate into the right lead.

For this ZGNR ($M = 8$). It seems strange that since near the zero energy the transmission is 1.0 (Fig. 4d), the steady current is about zero, which is "inconsistent" with the Landauer formula [4]. The reason for this is that the transmission in Fig. 4d is under a zero bias. With a finite bias, the transmission spectrum will have a sharp dip at around $E = 0$ eV with the width equal to the bias value (see Fig. 5b). This new transmission is consistent with the steady current. This phenomenon was first proposed by Li et al. and is explained as the parity mismatch of bonding (with odd parity) and antibonding (with even parity) subbands [32]. They find only the symmetric ZGNR have such an even–odd effect. Our HEOM calculations here have unveiled the dynamic process of this effect and show how the transient currents and charges evolve to the steady state.

For the asymmetric ZGNR, such as $M = 6$ case (Fig. 5a), there is no subband parity and the transmission does not drop to zero under bias voltages (Fig. 5c). For this ZGNR, we choose one (narrow) and three (wide) periods of units as the device part. Our global fitting approach has been used for the HEOM calculation. Ninety-nine Lorentzians are used for fitting the lead spectrum matrices. Figure 5d shows the transient currents for a narrow device (dashed line) and a wide device (solid line). Figure 5e shows the transient charges on different atoms of the narrow device. The bias voltage of 1.0 V is applied symmetrically on two leads and the onsite energies in the device are linearly changed.

As expected, the steady currents have finite values, which is consistent with the transmission spectrum. Now we discuss the transient case. First, the overshooting currents are also very large, which is due to the narrow lead spectra as in the symmetric ZGNR. At the beginning the charges on right-side atoms increase but on the left-side atoms they decrease. This is because the tilt potential energy profile (higher in the left) in the device makes the left atoms lose some charge and push them to the right atoms. This is a common phenomenon on a homogeneous chain with a linear change of the onsite energies between two leads.

Finally, we compare the transient currents in wide and narrow devices. There is some delay for the wide-device current to reach the first peak, compared with the narrow-device one. This is reasonable since in wide devices the incident electron has to spend more time to reach the right boundary and then gets reflected back. In fact, our previous work shows that in much larger systems such as a 1000-atom chain, the transient current increases linearly and then bends to the steady current [21]. This rise time is equal to L/v_F, where L is the length of device and v_F is the Fermi velocity. In our case the device size is not very large, so we only observe some qualitative phenomenon. In Perfetto's paper the first current plateau also lasts up to $\tau = L/v_F$ [16]. In that case the electrons propagate in a very wide ZGNR and behave as in 2D graphene sheet. After that time the electrons exploring the reservoirs and then smoothly drop to the steady current. This is very different from our narrow ZGNR structure.

4 Conclusions Two types of HEOM developments been proposed for time-dependent quantum transport in this paper. One is the energy-integration method for the initial-state calculation, including the contour integral technique for accurate computation. Another concerns systematic schemes of Lorentzian fittings. With these schemes, the number of Lorentzians can be effectively reduced by the Lorentzian combination algorithm; on the other hand, the fitting accuracy is still maintained. Armchair and zigzag GNR are used for transient calculations by this method. We find some delta-function-like spectrum peaks in many semi-infinite AGNR, which are the surface states in the AGNR lead. Our HEOM calculation also shows the current and charge evolutions in ZGNR with the "even–odd" effect for the first time. The slowly decaying charges on the edge atoms are due to the zero transmission for the edge states.

Acknowledgements We would like to thank Dr. Chi Yung Yam and Dr. Hui Cao for some useful discussions on quantum transport and Dr. Jian Sun for the help with the computer. Support from the Hong Kong Research Grant Council (HKU 700912P, HKU 700711P, HKU700909P, HKUST9/CRF/11G) and AoE (AOE/P-04/08). The support from the NSF of China (Nos. 21103157 and 21233007), the Fundamental Research Funds for Central Universities (Nos. 2340000034 and 2340000025), and the Strategic Priority Research Program (B) of the CAS (XDB01020000) are gratefully acknowledged.

Appendix A: Transient-current expression

The transient current through the lead (α)/device interface can be evaluated via

$$J_\alpha(t) = -e\int_\alpha \mathrm{d}\boldsymbol{r}\frac{\partial}{\partial t}\sigma_\alpha(t) = e\ \mathrm{Tr}([\boldsymbol{H},\boldsymbol{G}^<]_{\alpha\alpha}).$$

From the device–lead partition scheme,

$$(\boldsymbol{H}\boldsymbol{G}^<) = \begin{pmatrix} \boldsymbol{H}_{\alpha\alpha} & \boldsymbol{H}_{\alpha\mathrm{D}} \\ \boldsymbol{H}_{\mathrm{D}\alpha} & \boldsymbol{H}_{\mathrm{DD}} \end{pmatrix}\begin{pmatrix} \boldsymbol{G}^<_{\alpha\alpha} & \boldsymbol{G}^<_{\alpha\mathrm{D}} \\ \boldsymbol{G}^<_{\mathrm{D}\alpha} & \boldsymbol{G}^<_{\mathrm{DD}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{H}_{\alpha\alpha}\boldsymbol{G}^<_{\alpha\alpha} + \boldsymbol{H}_{\alpha\mathrm{D}}\boldsymbol{G}^<_{\mathrm{D}\alpha} & \boldsymbol{H}_{\alpha\alpha}\boldsymbol{G}^<_{\alpha\mathrm{D}} + \boldsymbol{H}_{\alpha\mathrm{D}}\boldsymbol{G}^<_{\mathrm{DD}} \\ \boldsymbol{H}_{\mathrm{D}\alpha}\boldsymbol{G}^<_{\alpha\alpha} + \boldsymbol{H}_{\mathrm{DD}}\boldsymbol{G}^<_{\mathrm{D}\alpha} & \boldsymbol{H}_{\mathrm{D}\alpha}\boldsymbol{G}^<_{\alpha\mathrm{D}} + \boldsymbol{H}_{\mathrm{DD}}\boldsymbol{G}^<_{\mathrm{DD}} \end{pmatrix},$$

we have $(\boldsymbol{H}\boldsymbol{G}^<)_{\alpha\alpha} = \boldsymbol{H}_{\alpha\alpha}\boldsymbol{G}^<_{\alpha\alpha} + \boldsymbol{H}_{\alpha\mathrm{D}}\boldsymbol{G}^<_{\mathrm{D}\alpha}$. Note that $\boldsymbol{H}_{\alpha\alpha}$ commutes with $\boldsymbol{G}^<_{\alpha\alpha}$ [7],

$$J_\alpha(t) = e\ \mathrm{Tr}([\boldsymbol{H}_{\alpha\mathrm{D}},\boldsymbol{G}^<_{\mathrm{D}\alpha}]) = e\ \mathrm{Tr}([\boldsymbol{H}_{\mathrm{D}\alpha},\boldsymbol{G}^<_{\alpha\mathrm{D}}]).$$

Then, with the similar derivation process as in Eq. (1) (applying Langreth's rule [13]), the transient-current expression is given below

$$J_\alpha(t) = -2e\ \mathrm{Tr}\Big\{\mathrm{Re}\Big(\int_{-\infty}^{t}\mathrm{d}\tau[\boldsymbol{G}^<_{\mathrm{D}}(t,\tau)\cdot\boldsymbol{\Sigma}^>_\alpha(\tau,t) - \boldsymbol{G}^>_{\mathrm{D}}(t,\tau)\cdot\boldsymbol{\Sigma}^<_\alpha(\tau,t)]\Big)\Big\}.$$

Appendix B: Initial-state calculation of the 1st-tier ADM by contour integral

Similar to the density matrix calculation in Section 2.2, we use the residue theorem to the integral calculation of the 1st-tier ADM. We rewrite Eq. (20) as:

$$\boldsymbol{\varphi}_{\alpha k} = \frac{1}{\pi i}\int_{-\infty}^{+\infty}\mathrm{d}E\frac{\mathrm{Im}[\boldsymbol{G}^r_{\mathrm{D}}(E)]f_{\mathrm{P}}(E)\cdot[\boldsymbol{A}^{>+}_{\alpha k} - \boldsymbol{A}^{<+}_{\alpha k}] + \mathrm{Im}[\boldsymbol{G}^r_{\mathrm{D}}(E)]\cdot\boldsymbol{A}^{<+}_{\alpha k}}{E - i\gamma^+_{\alpha k}} = \frac{1}{\pi i}(\boldsymbol{I}_a + \boldsymbol{I}_b),$$

where

$$\boldsymbol{I}_a = \textstyle\int_{-\infty}^{+\infty}\mathrm{d}E((\mathrm{Im}[\boldsymbol{G}^r_{\mathrm{D}}(E)]f_{\mathrm{P}}(E)\cdot[\boldsymbol{A}^{>+}_{\alpha k} - \boldsymbol{A}^{<+}_{\alpha k}])/(E - i\gamma^+_{\alpha k})),$$

$$\boldsymbol{I}_b = \textstyle\int_{-\infty}^{+\infty}\mathrm{d}E((\mathrm{Im}[\boldsymbol{G}^r_{D}(E)]\cdot\boldsymbol{A}^{<+}_{\alpha k})/(E - i\gamma^+_{\alpha k})).$$

Since only the analytic function can be used in the residue theorem, we have to change the integrand above (such as $(\mathrm{Im}[\boldsymbol{G}^r_{\mathrm{D}}(E)]\cdot\boldsymbol{A}^{<+}_{\alpha k})/(E - i\gamma^+_{\alpha k})$) into the real and imaginary part form with respect to the variable E, with the help of the formula below:

$$\frac{1}{E - i\gamma^+_{\alpha k}} = \frac{1}{E - i(\gamma_1 + i\gamma_2)} = \frac{1}{E + \gamma_2 - i\gamma_1} = \frac{(E+\gamma_2) + i\gamma_1}{(E+\gamma_2)^2 + \gamma_1^2},$$

where γ_1 and γ_2 are the real and imaginary parts of $\gamma^+_{\alpha k}$. If g_1 and g_2 stand for the real functions of E in the integrands except $\mathrm{Im}[G^r_{\mathrm{D}}]$, the two integrands above can be written as the following form

$$\mathrm{Im}[G^r_{\mathrm{D}}](g_1 + ig_2) = \mathrm{Im}[G^r_{\mathrm{D}}]g_1 + i\,\mathrm{Im}[G^r_{\mathrm{D}}]g_2 = \mathrm{Im}[G^r_{\mathrm{D}}g_1] + i\,\mathrm{Im}[G^r_{\mathrm{D}}g_2].$$

Then, the two integrals above can be rewritten as the form below

$$\int \mathrm{Im}[G^r_{\mathrm{D}}](g_1 + ig_2)\,\mathrm{d}E = \int \mathrm{Im}[G^r_{\mathrm{D}}]g_1\,\mathrm{d}E + i\int \mathrm{Im}[G^r_{\mathrm{D}}]g_2\,\mathrm{d}E = \mathrm{Im}\Big[\int G^r_{\mathrm{D}}g_1\,\mathrm{d}E\Big] + i\,\mathrm{Im}\Big[\int G^r_{\mathrm{D}}g_2\,\mathrm{d}E\Big].$$

The integrals in imaginary part above are analytic, which can be solved by the residue theorem.

For example, $\boldsymbol{I}_a$ is written as

$$\boldsymbol{I}_a = \mathrm{Im}\left[\int_{-\infty}^{+\infty}\mathrm{d}E\frac{\boldsymbol{G}^r_{\mathrm{D}}(E)f_{\mathrm{P}}(E)\cdot(E+\gamma_2)\cdot[\boldsymbol{A}^{>+}_{\alpha k} - \boldsymbol{A}^{<+}_{\alpha k}]}{(E+\gamma_2)^2 + \gamma_1^2}\right] + i\,\mathrm{Im}\left[\int_{-\infty}^{+\infty}\mathrm{d}E\frac{\boldsymbol{G}^r_{\mathrm{D}}(E)f_{\mathrm{P}}(E)\cdot\gamma_1\cdot[\boldsymbol{A}^{>+}_{\alpha k} - \boldsymbol{A}^{<+}_{\alpha k}]}{(E+\gamma_2)^2 + \gamma_1^2}\right] = (\mathrm{Im}[\boldsymbol{I}_{a1}] + \mathrm{Im}[\boldsymbol{I}_{a2}])\cdot[\boldsymbol{A}^{>+}_{\alpha k} - \boldsymbol{A}^{<+}_{\alpha k}].$$

The two integrals $\boldsymbol{I}_{a1}$ and $\boldsymbol{I}_{a2}$ can be calculated by the contour integral

$$\boldsymbol{I}_{a1} = \int_{-\infty}^{+\infty}\mathrm{d}E\frac{\boldsymbol{G}^r_{\mathrm{D}}(E)f_{\mathrm{P}}(E)\cdot(E+\gamma_2)}{(E+\gamma_2)^2 + \gamma_1^2} = \oint\mathrm{d}E\frac{\boldsymbol{G}^r_{\mathrm{D}}(E)f_{\mathrm{P}}(E)\cdot(E+\gamma_2)}{(E+\gamma_2)^2 + \gamma_1^2} = 2\pi i\sum_k^{N_k}\mathrm{Residue}[k],$$

$$\boldsymbol{I}_{a2} = \int_{-\infty}^{+\infty}\mathrm{d}E\frac{\boldsymbol{G}^r_{\mathrm{D}}(E)f_{\mathrm{P}}(E)\cdot\gamma_1}{(E+\gamma_2)^2 + \gamma_1^2} = \oint\mathrm{d}E\frac{\boldsymbol{G}^r_{\mathrm{D}}(E)f_{\mathrm{P}}(E)\cdot\gamma_1}{(E+\gamma_2)^2 + \gamma_1^2} = 2\pi i\sum_k^{N_k}\mathrm{Residue}[k],$$

where Residue(k) stands for the residues in the integral contours, which include the Padé poles from $f_P(E)$ and the pole from the fractional function $1/((E+\gamma_2)^2+\gamma_1^2)$. It is easy to prove that the integral on the upper-complex-plane contours (like $C_R = C_{R1} + C_{R2} + C_{R3}$ in Fig. 1) tends to zero when R becomes infinity, so the integral on the real axis can be transformed to the contour integral. Another integral $\boldsymbol{I}_b$ can also be calculated in the same way.

Appendix C: Details of the Lorentzian fitting

C1: Lorentzian points combination In Section 2.3, an algorithm is used to combine the neighboring Lorentzian points into one point on the Omega–Width diagram. Here, we show the details of this algorithm. There are several steps to search all the Lorentzian points and combine (group) all the neighboring points (the total number is N).

(1) Search for each point with index i ($i = 1, \ldots, N$) except the grouped points. The searched point is selected as a center point.
(2) For each center point i, search for the remaining points with index j ($j = i+1, \ldots, N$), except for the grouped points.
(3) Calculate the distance D_{ij} between point i and j. If D_{ij} is less than the critical radius r, then group the point j as the combined point with respect to the center point i. (Array party(i, k) $= j$ and array If group(j) $= 1$ is used for index recording.)
(4) Do the search for points with the index j (inner loop) and i (outer loop) again, until all the points are grouped.
(5) Use the recording information to calculate the average Ω and W values for each group of points. Finally, combine all the points of each group (the points in the circles, see Fig. 3a) into the new point with the averaged Ω and W values.

C2: Fitting the sharp peaks In some systems like the zigzag GNR, there are some very narrow peaks in the lead spectrum curves. To fit these curves, very narrow Lorentzians has to be used. However, from the Kramers–Kronig relation

$$\mathrm{Im}[\Sigma^r] \approx \sum_d \frac{A_d}{(E-\Omega_d)^2 + W_d^2},$$

$$\mathrm{Re}[\Sigma^r] \approx \sum_d \frac{-A_d/W_d \cdot (E-\Omega_d)}{(E-\Omega_d)^2 + W_d^2}.$$

We see a small W_d leads to a large $\mathrm{Re}[\Sigma^r]$. Thus, a tiny fitting error in the imaginary part (for width W_d) will result in a huge error in the real part. So, when fitting the imaginary-part curves, very narrow Lorentzians (with widths like 0.000001 eV) must be avoided to ensure good behaviors of the corresponding real-part ones.

Here, a strategy is employed to effectively fit these narrow-peak curves. We assign some fixed small-width Lorentzians. These widths will no longer be involved in the fitting. For example, for a curve with 3 narrow peaks, we employ 6 narrow Lorentzians with a fixed width (W_s) of 0.01 eV to fit these narrow peaks (in each peak 2 Lorentzians are used). Only their centers and amplitudes ($\{A_d, \Omega_d\}, d \in [1,6]$) will be adjusted in fitting. For other normal peaks, the Lorentzian centers, widths and amplitudes are the fitting parameters as well. The fitting parameters from each Lorentzian are listed below.

Lorentzian 1 (narrow peak 1) : A_1, Ω_1 ($W_s = 0.01$ eV);
Lorentzian 2 (narrow peak 1) : A_2, Ω_2 ($W_s = 0.01$ eV);
Lorentzian 3 (narrow peak 2) : A_3, Ω_3 ($W_s = 0.01$ eV);
Lorentzian 4 (narrow peak 2) : A_4, Ω_4 ($W_s = 0.01$ eV);
Lorentzian 5 (narrow peak 3) : A_5, Ω_5 ($W_s = 0.01$ eV);
Lorentzian 6 (narrow peak 3) : A_6, Ω_6 ($W_s = 0.01$ eV);
Lorentzian 7 (normal peak) : A_7, Ω_7, W_7.

With these parameters ($A_1, \Omega_1, A_2, \Omega_2, \ldots, A_6, \Omega_6$; $A_7, \Omega_7, W_7, \ldots, A_{N_d}, \Omega_{N_d}, W_{N_d}$), the fitting results will not generate very narrow Lorentzians.

References

[1] S. Datta, Quantum Transport, Atom to Transistor (Cambridge University Press, Cambridge, 2005).
[2] M. Auf der Maur, M. Povolotskyi, F. Sacconi, A. Pecchia, G. Romano, G. Penazzi, and A. Di Carlo, Opt. Quantum Electron. **40**, 1077 (2008).
[3] D. K. Ferry and S. M. Goodnick, Transport in Nanostructures (Cambridge University Press, Cambridge, 1997).
[4] J. C. Cuevas and E. Scheer, Molecular Electronics: An Introduction to Theory and Experiment (World Scientific, Singapore, 2010), chap. 4–6;
see also: S. Datta, Electronic Transport in Mesoscopic Systems (Cambridge University Press, Cambridge, 1995), chap. 2.
[5] Y. Xue, S. Datta, and M. A. Ratner, Chem. Phys. **281**, 151 (2002).
[6] J. Taylor, H. Guo, and J. Wang, Phys. Rev. B **63**, 245407 (2001).
[7] A. P. Jauho, N. S. Wingreen, and Y. Meir, Phys. Rev. B **50**, 5528 (1994).
[8] Y. Zhu, J. Maciejko, T. Ji, H. Guo, and J. Wang, Phys. Rev. B **71**, 075317 (2005).
[9] C. Y. Yam, X. Zheng, G. H. Chen, Y. Wang, T. Frauenheim, and T. A. Niehaus, Phys. Rev. B **83**, 245448 (2011).
[10] S. H. Ke, R. Liu, W. Yang, and H. U. Baranger, J. Chem. Phys. **132**, 234105 (2010).
[11] S. Z. Wen, S. K. Koo, C. Y. Yam, X. Zheng, Y. J. Yan, Z. M. Su, K. N. Fan, L. Cao, W. P. Wang, and G. H. Chen, J. Phys. Chem. B **115**, 5519 (2011).
[12] S. Kurth, G. Stefanucci, C.-O. Almbladh, A. Rubio, and E. K. U. Gross, Phys. Rev. B **72**, 035308 (2005).
[13] X. Zheng, F. Wang, C. Y. Yam, Y. Mo, and G. H. Chen, Phys. Rev. B **75**, 195127 (2007).
[14] E. J. McEniry, D. R. Bowler, D. Dundas, et al., J. Phys.: Condens. Matter **19**, 196201 (2007).

[15] C. Moyer, Am. J. Phys. **72**, 351 (2004).
[16] E. Perfetto, G. Stefanucci, and M. Cini, Phys. Rev. B **82**, 035446 (2010).
[17] Y. Tanimura and R. Kubo, J. Phys. Soc. Jpn. **58**, 101 (1989).
[18] J. S. Jin, X. Zheng, and Y. J. Yan, J. Chem. Phys. **128**, 234703 (2008).
[19] X. Zheng, G. H. Chen, Y. Mo, S. K. Koo, H. Tian, C. Y. Yam, and Y. J. Yan, J. Chem. Phys. **133**, 114101 (2010).
[20] H. Xie, F. Jiang, H. Tian, X. Zheng, Y. H. Kwok, S. G. Chen, C. Y. Yam, Y. J. Yan, and G. H. Chen, J. Chem. Phys. **137**, 044113 (2012).
[21] S. G. Chen, H. Xie, Y. Zhang, X. D. Cui, and G. H. Chen, Nanoscale **5**, 169 (2013).
[22] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, Rev. Mod. Phys. **81**, 109 (2009).
[23] J. Hu, M. Luo, F. Jiang, R. X. Xu, and Y. J. Yan, J. Chem. Phys. **134**, 244106 (2011).
[24] J. Maciejko, J. Wang, and H. Guo, Phys. Rev. B **74**, 085324 (2006).
[25] M. Brandbyge, J.-L. Mozos, P. Ordejón, J. Taylor, and K. Stokbro, Phys. Rev. B **65**, 165401 (2002).
[26] J. Mathews and R. L. Walker, Mathematical Methods of Physics (W.A. Benjamin, New York, 1970).
[27] MINPACK is a subprogram package for the numerical solution of nonlinear equations and nonlinear least square problems. It was written by J. More, B. Garbow, and K. Hillstrom. It is free to download from the website http://www.netlib.org/minpack/. About least-squares, see Ake Bjorck, Numerical Methods for Least Squares Problems (SIAM, Philadelphia, 1996).
[28] M. P. L. Sancho, J. M. L. Sacho, and J. Rubio, J. Phys. F, Met. Phys. **15**, 851 (1985).
[29] J. D. Jackson, Classical Electrodynamics (Wiley, New York, 1999), pp. 332–333.
[30] I. Tamm, Phys. Z. Soviet Union **1**, 733 (1932).
[31] K. Nakada, M. Fujita, G. Dresselhaus, and M. S. Dresselhaus, Phys. Rev. B **54**, 17954 (1996).
[32] Z. Y. Li, H. Y. Qian, J. Wu, B. L. Gu, and W. H. Duan, Phys. Rev. Lett. **100**, 206802 (2008).

