Removing all periodic boundary conditions: Efficient nonequilibrium Green's function calculations

Papior, Nick Rübner; Calogero, Gaetano; Leitherer, Susanne; Brandbyge, Mads

Published in:
Physical Review B

Link to article, DOI:
10.1103/physrevb.100.195417

Publication date:
2019

Document Version
Peer reviewed version

Citation (APA):
Removing all periodic boundary conditions: Efficient non-equilibrium Green function calculations

Nick Papior,1,2 Gaetano Calogero,3,2 Susanne Leitherer,2 and Mads Brandbyge2
1 Computing Center, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark
2 Center for Nanostructured Graphene, Department of Physics, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark
3 Dipartimento di Ingegneria dell’Informazione, Università di Pisa, 56122, Pisa, Italy
(Dated: August 28, 2019)

We describe a method and its implementation for calculating electronic structure and electron transport without approximating the structure using periodic super-cells. This effectively removes spurious periodic images and interference effects. Our method is based on already established methods readily available in the non-equilibrium Green function formalism and allows for non-equilibrium transport. We present examples of a N defect in graphene, finite voltage bias transport in a point-contact to graphene, and a graphene-nanoribbon junction. This method is less costly, in terms of CPU-hours, than the super-cell approximation.

I. INTRODUCTION

Widely used, efficient computational methods have been developed for calculations of electronic structure of systems presenting perfect periodic repetition of a unit cell along one, two, or three dimensions surrounded by vacuum. These are typically based on Kohn-Sham Density Functional Theory (DFT)1 2. The infinite system is replaced by finite unit-cell with periodic boundary conditions (PBC) using Bloch’s theorem and a discrete sampling of Bloch phases/\(k\)-points. Due to the efficient implementations this method is also applied systems which lack periodicity. For example surfaces are modelled by a slab, isolated defects by periodically repeated defects surrounded by “large” regions of bulk, and isolated adsorbates on surfaces by a mix. This results in compromises due to computational feasibility with respect to slab-size and inter-defect distances, which may lead to unwanted effects related to interferences or standing-wave patterns not present in the ideal, large system. Beyond PBC methods have been around for a long time. These include matching of the wavefunctions in different regions, e.g. surface and bulk3, and Green function or embedding methods4 9 have e.g. been used to treat the isolated defect/adsorbate on a surface or electronic transport between two electrodes10. These methods are based on a screening assumption where the potential has converged to its bulk value outside the computational “active” region.

In particular, for transport calculations the treatment of systems as “open” with semi-infinite electrodes along the transport directions is essential. A number of computational implementations have been developed for more than a decade for this problem based on the non-equilibrium Green function (NEGF) method11 16. These typically represent electrodes by a unit-cell repeated as periodic layers along the semi-infinite electrode/transport direction, and use PBC and corresponding \(k\)-points in the directions transverse to this. A self-energy is then used to treat the semi-infinite direction in a numerically exact way based on a very efficient method17 which recursively removes the infinite number of degrees of freedom/states in the semi-infinite direction. This approach may also be used in “single-electrode mode” treating the surface of semi-infinite bulk with a computational load comparable to slab calculation of e.g. chemical reactions at the surface14 15 19. Indeed this avoids the periodic images and finite size effects of the slabs in the surface-normal direction, but leaves the periodicity in the surface direction.

In this paper we present a simple, efficient and precise method based on Green function theory which can be used for isolated defects as well as extended NEGF calculations using multiple probes/electrodes. Our method solves this problem by calculating the real-space self-energy which can be outlined as (details explained in the Method section)

\[
\Sigma_{00}^R = S^R z - H^R - [G_{00}^R]^{-1}. \tag{1}
\]

Equation (1) is computationally demanding since real-space quantities require a dense integration grid in reciprocal space. We emphasize that our method focuses on the efficient algorithmic implementation which has prohibited the community to extend its broader use. Secondly, our method allows non-equilibrium calculations by the regular assumption of “equilibrium” electrodes11 19. The paper is organized as follows. First we describe the theoretical and computational details of our method. Then we show DFT+NEGF results using the real-space self-energies for three illustrative cases: i) electronic structure of a nitrogen defect in a large graphene lattice, ii) non-equilibrium transport in a gold STM tip in contact with a graphene flake, and iii) a graphene/graphene nanoribbon junction.

*nicpa@dtu.dk
II. METHOD

We remark that Eq. (1) is a well known equation in the transport community and that our contribution here lies in the implementation. In the following we will describe the method for a pristine bulk system which is the basis for defected systems.

The starting point of the efficient real-space self-energy method is any system with PBC in 2 or 3 directions where one wishes to replace a predefined direction with a semi infinite direction, see Fig. 1b. Lastly our presented method replaces any number of PBC, and/or semi-infinite directions with a single self-energy, see Fig. 1c.

The Hamiltonian of an infinitely large system may be written in a block-tri-diagonal “shell” fashion:

$$H^R = \begin{bmatrix} H_{00} & H_{01} & 0 & \cdots \\ H_{10} & H_{11} & H_{12} & \cdots \\ 0 & H_{21} & H_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (2)$$

Here index 0 in $H_{00}$ is referred as the primary unit cell with only nearest neighbour couplings, $H_{01}$ is the coupling between the primary unit-cell and the first set of neighbour cells (2D: 8, 3D: 26), and $H_{i,i+1}$ is the coupling between the $i$ and $i+1$ shell. We use the superscript $R$ to indicate the real-space representation of matrices.

We want to calculate the Green function for the infinite matrix comprising the Hamiltonian $H^R$ in a subspace $H^R_{0,0,0,\cdots}$ up to some shell size $i$. The straightforward Dyson equation is sufficient for systems with short screening lengths such as metals where the convergence requires only a few shells\cite{20}. For weak screening the increasing matrix sizes with $i$ in the Dyson equation become problematic and one may replace the real-space iterations in shells with an integral over $k$-points to calculate the real-space Green function (here only shown for the primary unit cell),

$$G^R_{00}(z) = \int dk G_k(z)$$

$$= [S^R z - H^R_{00}]^{-1} = [S_{00} z - H_{00} - \Sigma^R_{00}]^{-1} \quad (4)$$

where $G_k(z)$ is the Green function for a given $k$-point, $S^R$ the overlap matrix, and energy plus imaginary part is $z = E + i\eta$. We define the subspace of interest by 0 and the coupling of this to the surrounding bulk system is described by the real-space self-energy, $\Sigma^R_{00}$. We remark that $G_k(z) = G_{-k}$ using time-reversal symmetry. This converts the inversion of infinite matrices in real-space to a problem of inverting finite-sized matrices by introducing a $k$ integral employing Bloch’s theorem. This method was employed in Ref. 9.

Two new problems arise. A sufficient accuracy in the integral is difficult because the elements of the Green function has Lorentzian peaks/step-functions (in $k$-space) for each eigenvalue (pole). In Fig. 2 we show the Green function matrix elements of the graphene tight-binding model ($t = -2.7$ eV) for $z = (0.5 + i10^{-4})$ eV ($G_k^{-1} = I_z - H_k - \Sigma^R_k - \Sigma^R_{00}$) for transverse $k$ points. Several $\delta$ peaks are seen which makes an integral extremely difficult.
some multiples of the primary unit cell such that the final matrix has dimensions \( n \prod n_i \), with \( n \) being the number of orbitals in the primary unit cell, and \( n_i \) is the number of repetitions along the \( i \)th lattice vector. For large \( n_i \), the matrix dimensions rapidly increases making a fine integral in Eq. (3) unfeasible\[21].

Our method solves this dimension problem by only performing the costly inversions on matrices of dimension \( n \), regardless of all \( n_i \). This is achieved using three well established methods and performing them in the following order, 1) a surface self-energy removes the \( k \) integral along a single semi-infinite direction, 2) recursive Green function calculation (block-tri-diagonal inversion, BTD) expands along the semi-infinite direction and 3) Bloch’s theorem efficiently expands the Green function into the remaining one (or two) dimensions. Note that the surface self-energy calculation is a particular efficient solution of the BTD algorithm for a bulk system. Hence the difference between the two methods is subtle, yet significant in their application for our method. The three steps above work for both surfaces and bulk systems, with minor variations. Note that for 3D systems, leaving a \( k \) direction out of the integral, Eq. (3), one finds the real-space Green function for a cylinder with the directions normal to the cylinder surface integrated out, retaining the \( k \) point along the cylinder. Such a particular use-case will not be covered in this study, but we remark that our code allows such calculations which may be useful for e.g. line-defects in solids.

In the following all matrices not denoted by \( R \) are implicitly \( k \) dependent.

A. Surface self-energies

The recursive surface self-energy method by Sancho et al.\[21\] calculates the left/right self-energies for a given transverse \( k \)-point. This procedure presents a \( 2^i \) convergence series such that for iteration \( i \) one has effectively decimated \( 2^i \) layers. The algorithm is given here for the sake of completeness:

\[
\begin{align*}
\Sigma^L_0 &= \Sigma^R_0 = 0, \\
L_0 &= H_{10} - S_{10}z, \\
R_0 &= H_{01} - S_{01}z
\end{align*}
\]

perform following iterative scheme until \( \Sigma^L/R_{i-1} \approx \Sigma^L/R_i. \)

\[
\begin{align*}
\tau^L &= [S_{00}z - H_{00} - \Sigma^L_{i-1} - \Sigma^R_{i-1}]^{-1}L_{i-1} \\
\tau^R &= [S_{00}z - H_{00} - \Sigma^L_{i-1} - \Sigma^R_{i-1}]^{-1}R_{i-1} \\
\Sigma^L_i &= \Sigma^L_i + L_{i-1}\tau^R \\
\Sigma^R_i &= \Sigma^R_i + R_{i-1}\tau^L \\
L_i &= L_{i-1}\tau^L \\
R_i &= R_{i-1}\tau^R.
\end{align*}
\]

The surface self-energy removes the \( k \) integral along the semi-infinite direction and immediately reduces the integral in Eq. (3) by one dimension.

B. Block-tri-diagonal inversion (BTD)

This method may be generalized to calculate the layer off-diagonals for the inverse of matrices when these can be written in block form Eq. (2)\[14\]. The pristine bulk system may be written in the following BTD form along the semi-infinite direction:

\[
H = \begin{bmatrix}
H_{00} & H_{01} & 0 \\
H_{10} & H_{00} & H_{01} \\
0 & H_{10} & \ddots
\end{bmatrix}.
\]

Calculating the Green function for an arbitrary number of blocks along the semi-infinite direction follows,

\[
\begin{align*}
\tilde{Y} &= [S_{00}z - H_{00} - \Sigma^L_0 - \Sigma^R_0]^{-1}(S_{01}z - H_{01}) \\
\tilde{X} &= [S_{00}z - H_{00} - \Sigma^L_0 - \Sigma^R_0]^{-1}(S_{10}z - H_{10}) \\
G_{mn} &= [S_{00}z - H_{00} - \Sigma^L_0 - \Sigma^R_0]^{-1} \\
G_{mn} &= -\tilde{X}G_{m-n}, \text{ for } m > n \\
G_{mn} &= -\tilde{Y}G_{m+n}, \text{ for } m < n.
\end{align*}
\]

A key-point is that the real-space Green function for a bulk system is a Toeplitz matrix, e.g. \( G_{mn} = G_{m-n} \), for \( m - n = m' - n' \). Consequently for a bulk system of \( M \) blocks one can calculate the full Green function matrix by only calculating \( G_{00} \) and \( G_{n0} \) for all \( n \) (omitting \( G_{MM} \) since it equals \( G_{00} \)). Thus only \( 2M - 2 \) matrix multiplications are required in order to calculate the full Green function once \( G_{nn}, \tilde{X} \) and \( \tilde{Y} \) are obtained. We note that if the system is not bulk (e.g. surfaces) this algorithm need only be replaced by the full BTD algorithm\[14\], which is still much faster than full matrix inversion algorithms.

C. Bloch’s theorem

We want to obtain the self-energy for the pristine system consisting of a unit cell repeated \( N \) times in the transverse direction, large enough to include the defect cf. Fig. [1]. Due to the screening approximation we assume that the potential is unperturbed at the boundary and outside this cell and thus, the self-energy can be calculated from the pristine periodic system. To this end we can apply Bloch’s theorem and express the \( N \) times bigger system transverse to the semi-infinite direction via the primary matrix for a given \( k \). In our case we are interested in the Green function for a given \( k \). The equations for expanding the Green function (or any Bloch matrix) along a single direction for a given \( K \) (defined in the large
The potential in the surrounding region is not changed. The real-space Green function for arbitrarily large pristine, periodic systems, \( \mathbf{R} \times \mathbf{M} \) may be used in any defective system as long as \( \mathbf{V} \) and the potential in the surrounding region is not changed.

\[
G_N^N = \frac{1}{N} \sum_{j=0}^{N-1} e^{ik_j} e^{i(1-N)k_j} \otimes G_1^{k_j} \equiv \Sigma_R \equiv \Sigma_{00}
\]

Here \( G_1^{k_j} \) is the primary cell Green function matrix at the primitive cell \( k \)-point \( k_j \) which is to be unfolded into the matrix \( G_N^N \) and \( \otimes \) is the tensor product. The above equation is only expressed in terms of expansion along one direction, however it is easily generalized for more than one direction.

The above three steps conclude the calculation of the real-space Green function for arbitrarily large pristine, periodic systems, \( N \times M \). The algorithm in short; the self-energies remove the integral along one \( k \) direction, the BTD algorithm expands the Green function to arbitrary length, \( M \), along the semi-infinite direction employing just matrix multiplications, and finally Bloch’s theorem expands the Green function to arbitrary width (and also depth for 3D), \( N \).

### III. RESULTS

In the following we show results on spectral and transport properties of truly single defects/junctions using the real-space self-energy. Our self-consistent DFT+NEGF is implemented in Siesta, TranSiesta and TBtrans\([14, 23]\) while the algorithms described in Sec. II are implemented in siest\([24]\).

Three systems will be shown using graphene as the real-space electrode. The different systems highlight three particular cases where the real-space self-energy is applicable. We omit the use case of cylindrical self-energies since its use is limited to 3D bulk systems with periodicity along one direction (line defect). The atomic structure of the systems is shown as insets with periodicity along one direction (line defect). The atomic structure of the systems is shown as insets with coloured atoms indicating the support of the real-space self-energy/electrode (in blue), and other electrodes (in red). A last set of atoms is high-lighted (in light green) which are used as the projection region for local density of states (LDOS) analysis.

All calculations are performed using a 300 eV mesh cut-off, single-\( \text{\&}\) polarized basis set, and PBE+GGA exchange-correlation\([25]\), and otherwise default parameters. Although LCAO calculations for graphene using simple basis sizes (DZP) misses the lowest unoccupied states\([26]\) we do not add basis orbitals to describe these. Thus our presented analysis is limited to energies below the missing unoccupied bands \((E - \mu_{\text{graphene}} < 3.35 \text{eV})\).

#### A. Validation — Graphene

To ensure a functioning method we have constructed a pristine graphene calculation (inset Fig. 4) and calculated the projected DOS on a single carbon atom (marked). The Siesta method calculates the DOS on a \( 31 \times 51 \) Monkhorst-Pack grid\([27]\) (with energy broadening \( \sigma \sqrt{2} = 0.1 \text{eV} \)), and both the Green function methods are based on 300 \( k \) points and an imaginary part of \( \eta = 0.1 \text{meV} \).

The three different methods all yield the same projected DOS and there is a negligible difference between the methods. Any small difference from the diagonalization method vs. the Green function methods lies in...
the insufficient k-point sampling and the large smearing value. The two Green function methods are almost numerically the same since the system is bulk and no mixing of k-points take effect.

B. Nitrogen defect

Single nitrogen defects in graphene intrinsically have a substantial interaction range and thus calculating defect properties at the DFT level proves difficult [28]. In this example we use the real-space self-energy and compare with a 2D periodic and 1D periodic calculation. In all 3 examples we use the same unit-cell consisting of a square graphene lattice cell replicated 8 × 9 totalling 288 atoms. In Fig. 4 we show the projected DOS on the nitrogen atom for the three cases: i) A SIESTA calculation using a 31 × 51 Monkhorst-Pack grid in agreement with other work [29], ii) a two terminal TRANSIESTA with 300 transverse k points, and finally iii) using the real-space self-energy calculated from 300 k points. We remark that 300 k points corresponds to 2400 k in the minimal square graphene unit cell (see e.g. Fig. 2).

The DOS shows distinguished differences and particularly so for energies above the graphene Fermi level. The SIESTA and two probe calculations reveal a fine structure with multiple peaks dispersed over ∼ 1 eV. A large smearing parameter (σ√2 = 0.1 eV) for the SIESTA calculation was required due to the relatively crude Monkhorst-Pack grid, which still took more than 5 hours on 20 cores. The two probe calculation shows some even more localized features which could be the same as those in the SIESTA calculation. Both look similar to prior calculations [29] where the projected DOS on the nitrogen defect (p_z) in a similar periodic simulation was dispersed across two bands with a dispersion ∼ 0.5 eV. We find the real-space method broadens the peaks to a single peak, just above the chemical potential. This result is in perfect agreement with results from a tight-binding description of the isolated N fitted to DFT [28].

Although not shown, the same localized features found for the nitrogen atom are seen for the three neighbouring carbon atoms. These carbon atoms are particularly important for STM images [30].

C. STM tip on graphene

Scanning tunnelling microscope [31] (STM) is a key experimental technique for analyzing the local electronic structure of surfaces and defects or adsorbates on surfaces. The STM technique is a single tip junction probing the spatial local DOS and yields considerable insight of surface electronic topographies. However, DFT-NEGF calculations of STM on almost isolated defects are problematic both due to periodic repetition of the surface unit-cell, including the repetition of the STM probe tips. Here a calculation of the transmission from an “STM”-like tip to graphene [32–35] is calculated via two methods. Namely, a three terminal (left/right graphene/tip) invoking transverse periodicity, and a two terminal (graphene/tip) calculation, both at an applied bias of μ_{graphene} − μ_{tip} = −0.5 eV.

Figure 5 top panel shows spectral DOS of scattering states originating from the tip electrode projected on the carbon atom in contact with tip. In the bottom panel we show the transmission from the tip into graphene.

For the occupied states there is little to no difference while we find a large difference for the unoccupied states. The spectral DOS decreases on the contact atom while the transmission increases. In both graphs we find a discontinuity at 0.45 eV for the 3-electrode simulation (non existing in the real-space method) which we attribute to periodic image interaction. This fact is supported by other work [37] as well as it matches the bias on the tip.

Such STM calculations which may be very influential when calculating inelastic contributions [38, 39] since they use the energies in the unoccupied range [39].
D. Graphene contacted to a zGNR

A typical experiment comprise large electrodes contacted through a single graphene contact with few exceptions such as e.g. self-assembled monolayers. A key issue in DFT+NEGF simulations of such devices is that, until now, the simulation had a periodic array of junctions. Such an array of junctions will have interference effects and requires extra care in convergence of the width. Using the real-space self-energy we eliminate the periodic junctions and effectively retain a single junction where interference is removed.

The example shown here is a graphene flake contacted to a zig-zag graphene nano-ribbon (zGNR). Our calculations are performed using $\mu_{\text{graphene}} - \mu_{\text{zGNR}} = -0.5 \text{ eV}$. We remark that any molecular junction (for instance Au-benzene-di-thiol-Au) could be replaced in this example since the electrodes are handled as “surfaces”. In Fig. 7 we plot the projected DOS on the first 4 atoms in the zGNR (top) and the transmission (bottom). In this example there are relatively few differences since the unit cell is already relatively wide and thus the interference is limited. There are however differences such as a larger spread on the localized states just above the graphene chemical potential. These correspond to states in the zGNR which depends on the electrode coupling and thus is sensitive to periodicities.

E. Electrostatics

For all results shown, the electrostatics has been solved using the Fourier transform. Such a method forces the Poisson solution to be periodic. This is in contrast to our real-space Green function method which inherently has open boundaries. TRANSIESTA allows external input to ensure the correct boundary conditions for the electrodes. We have ensured that adding such boundary conditions does not change the results noticeably, for further information see [14, Sec. 3.5].

F. Disk space

Currently, our method relies on storing the self-energies on-disk since the algorithm still needs to be implemented in TRANSIESTA. Each surface self-energy file has a memory footprint which can be summarized by 3 integer, $n$ (matrix dimension), $n_k$ (number of k-points) and $n_E$ (number of energy points)

$$M_{\text{GB}}(n, n_k, n_E) = 16 \cdot n_k n^2 (2 + n_E) / 1024^3, \quad (11)$$

where $2 + n_E$ stems from the Hamiltonian and overlap per $k$, and the self-energies per $E$. In Tbl. I the dimensions and maximum disk usage is shown for the largest bias used.

We find a required disk space requirement of $\sim 50 \text{ GB}$ which is large, but in no way limiting its application on common HPC systems. One generally requires many more energy points in the TBTRANS calculation, however since one can define the chemical potential for the real-space electrode to be constant for all applied bias $\mu_e N = 0$ and the other electrode(s) to be at $\mu_e = V$ one can reuse the file for all applied bias at a much reduced computational cost and with a single file.
G. Performance

We have now shown that using the real-space self-energy one can avoid using the super-cell approximation for non-periodic structures such as isolated defects or single junction conductors. In order for it to be competitive with standard methods it also needs to be competitive in terms of performance/through put. We will here show that it is in fact less demanding to do a real-space self-energy calculation when taking into account the full sequence of calculations.

An important factor in using our method is the real-space self-energy calculation. The $\Sigma^R$ method is slower compared to $\Sigma_L/\Sigma_R$ given that the self-energy is more costly to calculate because of larger $k$-point sampling and a more complex algorithm. On the other hand the SCF cycles and transport/DOS calculations are much faster since no $k$-point sampling is required. In Tbl. II we show the timings of the presented calculations divided into three segments; i) TranSiesta, ii) TBtrans, and iii) $\Sigma^R$. All timings are based on the same 20-core machine.

As can be seen, the timings for TranSiesta is more or less constant while the TBtrans calculations are much faster. Note however, that for the graphene-zGNR system the convergence for the real-space method is faster leading to decreased timings in TranSiesta. Otherwise, the clear bottleneck is the $\Sigma^R$ calculation which can easily be embarrassingly parallelized. We remark that, as noted in Fig. 3 the self-energy is generic for any defect that does not alter the coupling out to the infinite exterior. This means that a single calculation of the self-energy allows using it for more than one system. Since these are one-shot calculations there is no reason not to do an extremely fine $k$ integration when sampling the real-space self-energy. For the systems shown here it takes less than 100 s per energy point for 300 $k$ points. It should be stressed that the current implementation is done in Python/Cython and thus additional performance gains would be to port it to fortran/C code.

All-in-all we find that the proposed method is comparable to, or faster, than the existing method for equivalent $k$-point sampling.

IV. CONCLUSION

We have presented a simple, effective, and fast algorithm for constructing real-space self-energies generalized for surfaces and full 2D/3D bulk systems. The algorithm relies on already well established methods used in the community and can thus be directly integrated into existing codes without problems. The current algorithms are implemented in the TranSiesta, TBtrans and sisl toolboxes which are all open-source under GPL variant licenses.

We have applied the method in three graphene cases which are readily found in current experimental literature. A recurring difference between the analyzed DOS and transmission profiles is that the occupied energy range is largely comparable to standard DFT+NEGF methods, while the unoccupied energy range shows substantial deviations. Such differences are attributed to removed interference effects.

We have shown how the use of real-space self-energies will remove the periodic images of defects in DFT calculations. The results shown provide insights into the far-field accuracy of DFT+NEGF calculations for single defects which has been missing in the electronic structure community.
V. ACKNOWLEDGMENTS

We thank Dr. K. Kaasbjerg for useful discussions. Financial support by Villum Fonden (00013340), Danish research council (4184-0003). The Center for Nanostructured Graphene (CNG) is sponsored by the Danish Research Foundation (DNRF103).


