A LETTERS JOURNAL EXPLORING THE FRONTIERS OF PHYSICS

OFFPRINT

Efficient computation of lattice Green's functions for models with nearest-neighbour hopping

M. BERCIU and A. M. COOK

EPL, **92** (2010) 40003

epl A Letters Journal Exploring the Frontiers of Physics

November 2010

EPL, **92** (2010) 40003 www.epljournal.org
doi: 10.1209/0295-5075/92/40003

Efficient computation of lattice Green's functions for models with nearest-neighbour hopping

M. Berciu and A. M. Cook

Department of Physics and Astronomy, University of British Columbia - Vancouver, BC, Canada V6T 1Z1

received 19 August 2010; accepted in final form 28 October 2010
published online 3 December 2010

PACS 02.60.Cb – Numerical simulation; solution of equations
PACS 02.70.Hm – Spectral methods
PACS 02.30.Gp – Special functions

Abstract – We show that for models with nearest-neighbour (nn) hopping, the lattice Green's functions can be calculated without the need to perform integrals. Our method applies to rectangular, triangular and honeycomb lattices in two dimensions, and to simple, face-centered and body-centered lattices in three dimensions. External magnetic fields can be dealt with trivially. As an example, we show that our method works for any ratio ϕ/ϕ_0 of the magnetic flux through the unit cell, *i.e.* irrespective of the change in the size of the magnetic unit cell. Other straightforward generalizations are to models with multiple orbitals per site, with any spin-orbit coupling, on-site disorder, and any combinations thereof. The method works equally well in the presence of surfaces. In all cases, accurate values for large distances can be obtained very efficiently and without finite-size effects. The relationship to other computational methods is also analyzed.

Lattice Green's functions are needed in a multitude of contexts in condensed matter [1], for instance in problems involving disorder [2], surfaces [3,4], interactions [5], modeling of mesoscopic devices [6] etc. The densities of states which can be obtained from these quantities are of key interest in methods such as the Dynamical Mean-Field Theories (DMFT) [7]. While in principle the Green's functions are straightforward to calculate, their evaluation by usual means turns out to be a cumbersome numerical task.

The lattice Green's function is defined as

$$G(\mathbf{R}_1,\mathbf{R}_2,\omega)=\langle 0|c_{\mathbf{R}_1}\hat{G}(\omega)c^{\dagger}_{\mathbf{R}_2}|0\rangle, \qquad (1)$$

where $|0\rangle$ is the vacuum, $c_{\mathbf{R}}$ annihilates a particle at site $\mathbf{R}$ of the lattice, and $\hat{G}(\omega)=[\omega+i\eta-\mathcal{H}]^{-1}$ is the resolvent corresponding to the Hamiltonian $\mathcal{H}$ of the system ($\hbar=1$). In the absence of disorder, in clean systems, the expression of the lattice Green's functions can be obtained if the eigenspectrum $\mathcal{H}|\phi_{n,\mathbf{k}}\rangle=E_n(\mathbf{k})|\phi_{n,\mathbf{k}}\rangle$ is known. Here, n indexes the various bands and $\mathbf{k}$ is the quasi-momentum defined in the Brillouin zone (BZ) of the specific model. In terms of these,

$$G(\mathbf{R}_1,\mathbf{R}_2,\omega)=\sum_n\int_{BZ}\mathrm{d}\mathbf{k}\frac{\phi_{n,\mathbf{k}}(\mathbf{R}_1)\phi^*_{n,\mathbf{k}}(\mathbf{R}_2)}{\omega+i\eta-E_n(\mathbf{k})}, \qquad (2)$$

where $\phi_{n,\mathbf{k}}(\mathbf{R})=\langle\mathbf{R}|\phi_{n,\mathbf{k}}\rangle=\langle 0|c_{\mathbf{R}}|\phi_{n,\mathbf{k}}\rangle$ are the Bloch wave functions. Evaluating these integrals directly is inefficient. To see why, consider the simplest possible case of nn hopping on a d-dimensional simple hypercubic lattice with no external magnetic field. There is a single band with $E(\mathbf{k})=-2t\sum_{i=1}^{d}\cos(k_ia)$, the wave functions are $\phi_{n,\mathbf{k}}(\mathbf{R})=e^{i\mathbf{k}\cdot\mathbf{R}}/\sqrt{(2\pi)^d}$ and the BZ is $|k_i|<\frac{\pi}{a},\forall i=1,d$, where a is the lattice constant. Using the invariance to translations, if $\mathbf{R}_1-\mathbf{R}_2=a(n_1,n_2,\ldots,n_d)$ then

$$G(n_1,\ldots,n_d,\omega)=\int_{BZ}\frac{\mathrm{d}k_1\ldots\mathrm{d}k_d}{(2\pi)^d}\frac{e^{i\sum_{i=1}^{d}k_in_ia}}{\omega+i\eta-\epsilon(\mathbf{k})}. \qquad (3)$$

Since $G(\ldots,n_i,\ldots,n_j,\ldots,\omega)=G(\ldots,-n_i,\ldots,n_j,\ldots,\omega)=G(\ldots,n_j,\ldots,n_i,\ldots,\omega)$ etc., we only need $G(n_1,\ldots,n_d,\omega)$ for $n_1\geqslant\cdots\geqslant n_d\geqslant 0$; all other Green's functions can be obtained from these. Moreover, since a local self-energy $\Sigma(\omega)$ may be present, *i.e.* $\omega\to\omega-\Sigma(\omega)$ in eq. (3), we may need these Green's functions for arguments with large imaginary parts. Thus, we rename $\omega+i\eta\to z$ in eqs. (1)–(3), and allow z to be any complex number.

In eq. (3), one integral can be performed analytically [2], however for $d>1$ one has to evaluate the other $d-1$ integrals numerically. This becomes difficult for larger d and for large values of $|n_i|$, where the integrand is a strongly oscillating function. Note that $d>3$ results are needed to solve many-particle problems —for instance, two-particle Green's functions for non-interacting particles

in 3D can be expressed in terms of the $d=6$ Green's functions of eq. (3) [8]. The situation gets significantly worse if a magnetic field is applied, *e.g.* transverse to a $d=2$ square lattice. If the magnetic flux per unit cell $\phi = Ba^2$ is $\phi/\phi_0 = p/q$, where $\phi_0 = h/e$ is the magnetic flux quantum and p and q are mutually prime integers, then the magnetic unit cell is q times larger, implying a BZ that is folded q times and thus hosts q distinct bands —the Hofstadter butterfly [9]. Any minute change in B, which may modify q significantly, will require a very different calculation to extract the Green's functions out of eq. (2). Similar difficulties are encountered, even at $B=0$, for other lattices like the honeycomb in 2D or FCC and BCC in 3D, which have rather oddly shaped BZ. Finally, this integration-based approach fails completely in the presence of disorder.

Multiple alternatives to the inefficient brute-force integration of eq. (2) have been proposed [1]. Some start from the equation of motion for the Green's function, which is derived from the identity $(z-\mathcal{H})\,\hat{G}(z)=1$, and which for nn hopping on the hypercubic lattice reads

$$zG(\ldots,n_i,\ldots,z)+t\sum_{i=1}^{d}\sum_{\delta=\pm1}G(\ldots,n_i+\delta,\ldots,z)=\prod_{i=1}^{d}\delta_{n_i,0}. \tag{4}$$

For example, in 2D Morita [10] found a special recurrence relation between $G(n,n,z), G(n-1,n-1,z)$ and $G(n+1,n+1,z)$. As a result, one only needs to evaluate numerically $G(0,0,z)$ and $G(1,1,z)$, and then all other $G(n,m,z)$ are obtained either from this special recurrence relation, or from eq. (4). However, this works only for isotropic 2D nn hopping and even then, there are computational issues for large n_i values [11]. Work linking these Green's functions to elliptic integrals, hypergeometric functions, wavelets and other special functions has also been pursued [1,12], however this usually either redefines the integrals which need to be performed, or expresses them as infinite power series which have their own challenges in terms of accurate and efficient evaluation.

A much more general —and probably the most used— approach is the recursion method reviewed in ref. [13], which is linked to the Lanczos method [14]. Briefly, to find $\langle u_0|\hat{G}(z)|u_0\rangle$ one uses iterations $\mathcal{H}|u_n\rangle = b_n^*|u_{n-1}\rangle + a_n|u_n\rangle + b_{n+1}|u_{n+1}\rangle$ to generate the orthonormal vectors $\{|u_n\rangle\}$ starting from $|u_0\rangle$, and to calculate the constants $\{a_n\}, \{b_n\}$. Because in this basis $\mathcal{H}$ is a tridiagonal matrix, the matrix element of its resolvent is found to be

$$\langle u_0|\hat{G}(z)|u_0\rangle = \cfrac{1}{z-a_0-\cfrac{|b_1|^2}{z-a_1-\ldots}}. \tag{5}$$

Various ways to terminate the continued fraction are reviewed in ref. [13]. To find $G(\mathbf{R},\mathbf{R},z)$ one starts with $|u_0\rangle = |\mathbf{R}\rangle$. To find $G(\mathbf{R}_1,\mathbf{R}_2,z)$ for $\mathbf{R}_1 \neq \mathbf{R}_2$, one performs two calculations for $|u_0^{(\pm)}\rangle = |\mathbf{R}_1\rangle \pm |\mathbf{R}_2\rangle$. Because our method is somewhat related to this recursion method, we compare them in more detail below, to emphasize why our approach is much more efficient for nn hopping.

For clarity, we first discuss our method for simple hypercubic lattices in d-dimensions, with isotropic hopping and no on-site disorder. We begin with $d=1$ to illustrate the approach and recover the known solution analytically. Then, we generalize to higher d and show that we can find all Green's functions with $\sum_{i=1}^{d} n_i \leqslant N_{max}$ in one run. As an example of its capabilities, we prove that addition of a homogeneous external magnetic field can be treated just as easily, and one can study everything from the formation of Landau levels in the limit of very small magnetic fields to the appearance of Hofstadter butterflies at large magnetic fields within the same framework. We also show how the method can be extended to triangular and honeycomb lattices in 2D, respectively BCC and FCC lattices in 3D, and list many other possible generalizations.

In 1D, we start with eq. (4) for $n=0$, to find

$$zG(0,z)+2tG(1,z)=1, \tag{6}$$

since $G(1,z)=G(-1,z)$. For any $n\geqslant 1$, eq. (4) reads

$$zG(n,z) = -t\left[G(n-1,z)+G(n+1,z)\right], \tag{7}$$

which is a trivial recurrence relation. We solve it using arguments which can be generalized to higher d. We expect $G(n,z)\to 0$ as $n\to\infty$ because its Fourier transform $G(n,t)$ is the amplitude of probability for the particle to move a distance na in the time t. As $n\to\infty$, this probability has to vanish for any $\mathrm{Im}\, z \neq 0$. Let N be a large enough value so that we can take $G(N+1,z)\approx 0$ (we will allow $N\to\infty$, so there is no actual truncation). From eq. (7) for $n=N$ we find that $G(N,z)\propto G(N-1,z)$. Using eq. (7) for $n=N-1$ we find that $G(N-1,z)\propto G(N-2,z)$ etc., so if $N\to\infty$ then for any $n\geqslant 1$ we have

$$G(n,z) = A(z)G(n-1,z), \tag{8}$$

from which $G(n,z)=[A(z)]^n G(0,z)$. Clearly, $|A(z)|<1$ so that $G(n,z)\to 0$ as $n\to\infty$. Substituting in eq. (7) we get $z=-t[1/A(z)+A(z)]$, thus $A(z)$ equals the root $\frac{1}{2}\left[-\frac{z}{t}\pm\sqrt{\left(\frac{z}{t}\right)^2-4}\right]$ whose modulus is less than 1. Since the product of these two roots is 1, the correct one is uniquely identified for any $\mathrm{Im}\, z\neq 0$. Finally, using $G(1,z)=A(z)G(0,z)$ in eq. (6) we find $G(0,z)=1/[z+2tA(z)]$. This completes the derivation of all 1D Green's functions using only basic arithmetics.

We now generalize this to the square 2D Green's functions $G(n,m,z)$, with $n\geqslant m\geqslant 0$. At first this seems impossible, since eq. (7) here reads $-\frac{z}{t}G(n,m,z)=G(n-1,m,z)+G(n+1,m,z)+G(n,m-1,z)+G(n,m+1,z)$ and this is not a simple recurrence relation. The solution, however, is immediate if we think in terms of the "Manhattan distance" $|n|+|m|$ and realize that the terms on the right side either increase it or decrease it by 1.

To formalize this, we introduce the vectors $V_{2n}^T = (G(2n,0,z), G(2n-1,1,z),\ldots,G(n,n,z))$ and $V_{2n+1}^T = (G(2n+1,0,z), G(2n,1,z),\ldots,G(n+1,n,z))$ that collect

all distinct Green's functions with Manhattan distances of $2n$, respectively $2n+1$. For any Manhattan distance larger than 1, eq. (4) can be written in the matrix form

$$V_n = \alpha_n(z)V_{n-1} + \beta_n(z)V_{n+1}, \tag{9}$$

where $\alpha_n(z)$ and $\beta_n(z)$ are easy to identify sparse matrices, *e.g.* $\alpha_{2n}(z)|_{i,i} = -\frac{t}{z}$, $(\forall) i = 1, n$; $\alpha_{2n}(z)|_{i,i-1} = -\frac{t}{z}$, $(\forall) i = 2, n$ while $\alpha_{2n}(z)|_{n+1,n} = -\frac{2t}{z}$, and all other matrix elements are zero. Note that none of the matrix elements depend on n, they are simple multiples of $-t/z$. However, the dimensions of these matrices depend on n.

As shown before, we must have for any $n \geqslant 1$,

$$V_n = A_n(z)V_{n-1}, \tag{10}$$

where the matrices $A_n(z)$ are found from eq. (9) to be

$$A_n(z) = [1 - \beta_n(z)A_{n+1}(z)]^{-1}\alpha_n(z). \tag{11}$$

They can be computed starting from a sufficiently large N with $A_{N+1}(z) = 0$. The multiplication by α and β matrices is very efficient; the time-consuming part is the matrix inversion. Even this is not a big problem, because the dimension of these matrices decreases fast with decreasing n. The choice of the cutoff N and a better "initial guess" than $A_{N+1}(z) = 0$ are discussed below.

Once $A_n(z)$ are known, we have $V_n = A_n(z) \dots A_1(z)V_0$, where $V_0 = G(0,0,z)$. In particular $V_1 = G(1,0,z) = A_1(z)G(0,0,z)$, which used in eq. (4) with $n = m = 0$, *i.e.* $zG(0,0,z) + 4tG(1,0,z) = 1$, gives $G(0,0,z) = 1/[z + 4tA_1(z)]$. This completes the 2D calculation using again only elementary operations and no integrals.

From $G(0,0,z)$ one should calculate other Green's functions using $V_n = A_n(z) \dots A_1(z)G(0,0,z)$, *not* eq. (9). The latter seems more efficient since it does not require storing the A_n matrices, however it leads to wrong results for large Manhattan distances. The reason is that just like eq. (7), eq. (9) has a general solution which is a linear combination of exponentially increasing and exponentially decreasing terms with the Manhattan distance. The physical solution is the exponentially decreasing one. The "initial condition" $zG(0,0,z) + 4tG(1,0,z) = 1$ is guaranteed to eliminate the non-physical contribution. However, any numerical errors in $G(0,0,z)$ or accumulated as the solution is propagated to larger Manhattan distances will mix in some of the non-physical solution, and eventually this wrong, exponentially increasing term will dominate the exponentially decreasing, physical solution.

This is avoided using $V_n = A_n(z) \dots A_1(z)G(0,0,z)$, because the $A_n(z)$ matrices are calculated coming down from asymptotically large Manhattan distances. As they are propagated towards smaller Manhattan distances, it is the physical solution that now increases exponentially and overcomes the errors made in deciding how to start the iterations at a large Manhattan distance N. This explains why even with the simple-minded initial choice of $A_{N+1}(z) = 0$ one can obtain good accuracy in rather few steps (an even better choice is presented below).

Generalization to $d > 2$ hypercubic lattices is straightforward. First, we collect all distinct Green's functions with a Manhattan distance n in a vector V_n. Then, because the Hamiltonian only involves nearest-neighbour hopping, the equation of motion (4) for $n \geqslant 1$ becomes a recurrence relation linking V_n to V_{n-1} and V_{n+1} and the solution proceeds similarly to the one described above. Of course, the size of V_n increases with d. For $d = 2$ we have (for the clean system on a square lattice) $\dim(V_n) \sim \frac{n}{2}$, for $d = 3$ we have $\dim(V_n) \sim \frac{n^2}{4}$, etc. If the system has disorder or other inhomogeneities, then the dimension of V_n will increase further as symmetries are broken. The maximum possible dimension of V_n on the hypercubic lattice is $\Delta_n^{(d)} = \sum_{n_1=-n}^{n} \dots \sum_{n_d=-n}^{n} \delta_{|n_1|+\dots+|n_d|,n}$.

Before showing results and better ways to terminate the continued fraction, let us contrast our method to the recursion method when both are terminated at the same step n by setting $A_{n+1} \equiv 0$, respectively $|u_{n+1}\rangle \equiv 0$. *Their results are identical*, however the calculation is approached very differently. The recursion method calculates a new basis $|u_n\rangle = \sum_{\mathbf{R}} \alpha_n(\mathbf{R})|\mathbf{R}\rangle$ which makes $\langle u_n|\mathcal{H}|u_m\rangle$ tridiagonal, so that the first matrix element of its inverse $\langle u_0|(z - \mathcal{H})^{-1}|u_0\rangle$ is given by eq. (5). Let us assume that $|u_0\rangle$ has the electron at the origin. Acting n times on it with $\mathcal{H}$ means that $|u_n\rangle$ is a linear superposition of states with the electron at a Manhattan distance n from the origin. This implies that to properly generate all the vectors up to $|u_n\rangle$, *one needs to keep in the basis* $|\mathbf{R}\rangle$ *all sites up to the Manhattan distance* n. In other words, one has to deal with vectors of dimension $\sum_{m=0}^{n} \Delta_m^{(d)} \sim n^d$ at each step of the iteration $\mathcal{H}|u_m\rangle = b_m^*|u_{m-1}\rangle + a_m|u_m\rangle + b_{m+1}|u_{m+1}\rangle$. Moreover, to get an off-diagonal matrix element, one needs to repeat the whole calculation twice (and to appropriately increase the dimension of the basis).

In contrast, our method deals directly with the matrix elements of the resolvent, not of $\mathcal{H}$, in the basis $\{|\mathbf{R}\rangle\}$. One immediate advantage is that we can take symmetries (if any) into account. Instead of finding a new basis in which the matrix of eqs. (4) can be inverted, we use the fact that because of nearest-neighbour hopping the equations of motion can be put into the recurrence form of eq. (9) which allows for a solution in terms of a continued fraction of matrices. The dimension of the largest matrix V_n scales like n^{d-1} since it contains only the sites on the hypersurface of Manhattan distance n, while the recurrence method groups all the sites in the volume bounded by this surface together. This is why our method involves much smaller matrices, and therefore can be used up to much higher n. Furthermore, it also gives all the Green's functions with $|\mathbf{R}_1 - \mathbf{R_2}| < n$ from the same calculation.

If $|u_0\rangle = c_0^\dagger|0\rangle$ has the electron at origin, both calculations truncated at iteration $n+1$ give the spectrum for a finite system with open boundary conditions and which includes only sites up to Manhattan distance n from the origin. The Green's function is a sum of poles describing the discrete states of this finite-size system. If

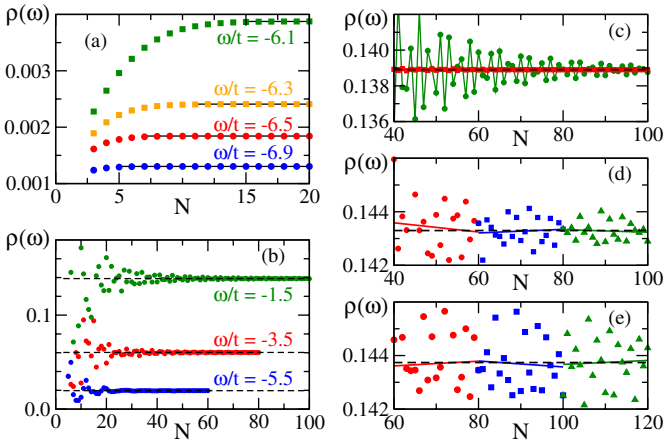

Fig. 1: (Color online) 3D DOS $\rho(\omega) = -\frac{1}{\pi}\text{Im}\ G(0,0,0,\omega)$ *vs.* the cutoff N. Panels (a), (b) and (c) are for $\eta = 0.1$, (d) is for $\eta = 0.01$ and (e) for $\eta = 0.001$. Panels (c), (d) and (e) are for $\omega/t = -1.5$. The dashed black lines in all panels show the numerical integration results. See text for more details.

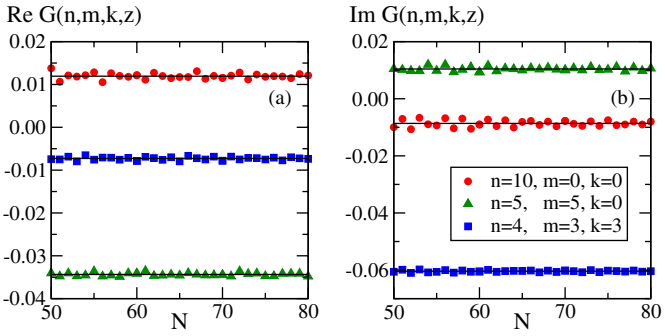

Fig. 2: (Color online) Various $G(n,m,k,z)$ for $z/t = -1.5 + 0.01i$ *vs.* N. Solid lines show numerical integration results.

n is large enough that level spacings are less than η, the Green's function converges smoothly towards that of the bulk system. Another way to see this is that $\tau \sim 1/\eta$ is the lifetime of the electron. If τ is short enough and n is large enough, the electron cannot reach the boundaries of the system, and this simple truncation procedure gives an excellent approximation. For a small η, one needs to either go to a large enough n or to find a better truncation scheme.

In figs. 1 and 2 we show results for $d = 3$, to substantiate our claims. They are compared against highly accurate (but very slow) numerical values obtained with Mathematica from eq. (3). We start with results obtained with the truncation $A_{N+1}(z) = 0$. We plot the density of states (DOS) $\rho(\omega) = -\frac{1}{\pi}\text{Im}\ G(0,0,0,\omega)$ *vs.* the cutoff value N for energies in the gap in fig. 1(a), and in the continuum in fig. 1(b). Since $\rho(\omega) = \rho(-\omega)$ we only explore the $\omega < 0$ region. As expected, for energies in the gap $|\omega| > 6t$, results converge fast to the expected values, shown by the dashed black lines —the further away from the band-edge, the faster the convergence. This is because in the gap the propagators decrease exponentially with the distance, as there are no eigenstates at these energies. The characteristic lengthscale increases with $1/(|\omega| - 6t)$ and is insensitive to η unless one is very near the band-edge. Thus, $A_{N+1}(z) = 0$ is good enough for energies in the gap, and a very small N suffices to achieve high accuracy.

Figure 1(b) shows results for $\omega/t = -5.5, -3.5, -1.5$ and a rather large $\eta/t = 0.1$. Here η controls how fast the solution convergences, since $1/\eta$ defines the "lifetime". In all cases a cutoff $N \sim 50$ suffices, although full convergence is rather slow. This is clearer in fig. 1(c), where the $\omega/t = -1.5$ data (circles) is shown on an expanded scale. It oscillates about the expected value (dashed line) with a decreasing amplitude. This is problematic, because it suggests that the appropriate cutoff value N scales like $1/\eta$, making convergence difficult for small η.

Thus, here we need a better truncation than $A_{N+1} = 0$. The red squares in fig. 1(c) show our best solution, obtained by assuming that for very large Manhattan distances $N = n + m + k \gg 1$, we have asymptotic dependence $G(n+1,m,k,z) \approx G(n,m+1,k,z) \approx G(n,m,k+1,z) = \lambda(z)G(n,m,k,z)$ with $|\lambda(z)| < 1$, in other words all propagators decrease at the same rate, as N increases. This is certainly true [11] for $d = 1, 2$ and although we have been unable to prove it in higher d, it is a reasonable guess in the continuum limit $N \gg 1$. Equation (4) then becomes $z + \text{d}t[\lambda(z) + 1/\lambda(z)] = 0$, from which the physical solution $|\lambda(z)| < 1$ is immediate. As expected, for $|\omega| < 2\text{d}t$, $1 - |\lambda(z)| \sim \eta$, *i.e.* in the continuum η indeed controls how fast the physical solution decreases with N. The new "initial guess" is extracted from eq. (4) by replacing the propagators corresponding to $N+1$ by $\lambda(z)G(n,m,k,z)$. This gives an approximate relation between V_N and V_{N-1} and therefore an initial guess for $A_N(z)$. Note that this new truncation automatically identifies the proper location of the continuum $|\omega| < 2\text{d}t$. This is to be contrasted with the various termination schemes for the recurrence method described in ref. [13].

Figures 1(d), (e) show results for $\eta/t = 10^{-2}$, 10^{-3} for this new $A_N(z)$. The spread increases with $1/\eta$, but slowly. Accurate values are obtained averaging over a range of N. The solid lines show best fits for consecutive ranges of 20. The average values for these intervals converge monotonically, *e.g.* for $\eta/t = 10^{-3}$ the average over the 60–80 range is 0.143701, while that over 100–120 is 0.143740; the integration result is 0.143739. Thus, one gets decent accuracy even with $N \leqslant 80$, which requires vectors V_n of dimension below 600. To go to $N = 80$ in 3D with the recursion method, one needs to work with vectors of dimension 695681; the mismatch is worse as N increases.

The accuracy is maintained for finite Manhattan distance, as shown in fig. 2 for $G(n,m,k,\omega)$ with $n + m + k = 10$. Averaging over the 50–60 range already gives relative errors below 0.006, and they decrease fast with increased range/N values. Of course, the average must be over values obtained from cutoffs N larger than the distances of interest, but even inside the continuum

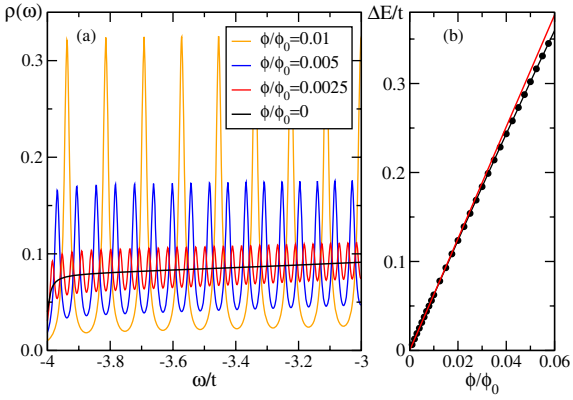

Fig. 3: (Color online) (a) DOS *vs.* ω for small magnetic fields, for a square lattice. Equally spaced Landau levels (broadened by $\eta = 0.01$) are observed; (b) energy difference ΔE between the first Landau level and the bottom of the band *vs.* ϕ/ϕ_0. Circles show the data, the line is the prediction in the continuum limit.

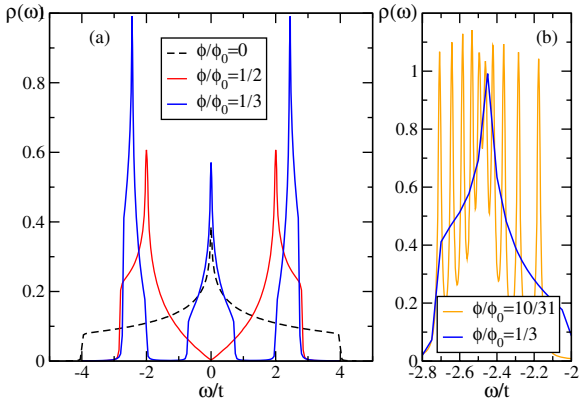

Fig. 4: (Color online) (a) DOS *vs.* ω showing two examples of the Hofstadter butterfly on the square lattice. (b) As the flux is decreased from 1/3 to 10/31, the lower band of the 1/3 Hofstadter butterfly splits into 10 distinct bands. See text for more details.

one does not need to use very much larger N to achieve convergence.

Our method generalizes trivially to hypercubic lattices with anisotropic nn hopping including phases due to an external magnetic field and/or spin-orbit coupling terms [15], to addition of on-site disorder, or to having multiple orbitals per site, since in all these cases the nn hopping links V_n to only V_{n-1} and V_{n+1} and the solution carries over. As an example, consider the 2D square lattice in a homogeneous transverse magnetic field B. If one uses the symmetric gauge $\mathbf{A}(\mathbf{r}) = \frac{B}{2}(-y, x, 0)$, then the lattice Green's functions remain highly symmetric with $G(n, m, z) = G(m, -n, z) = G(-n, -m, z) = G(-m, n, z)$ so the number of distinct Green's functions with a Manhattan distance $N = |m| + |n|$ is $\dim(V_N) = N$. The vector potential modifies the hopping integrals by adding Peierls phases $t_{ij} \to t \exp(\frac{ie}{\hbar}\int_i^j \mathbf{A}\mathrm{d}\mathbf{r})$. This changes the matrix elements of $\alpha_n(\omega)$ and $\beta_n(\omega)$, but nothing else: the same method works for any value of ϕ/ϕ_0.

To prove this, we look first at the $\phi/\phi_0 \to 0$ limit in fig. 3(a). In this case, the magnetic length is very large and the solution must agree with the one in the continuum limit. In particular, at the bottom of the band where the dispersion can be approximated in the parabolic limit, we expect to see the emergence of the Landau levels [16]. Instead of the continuous band $E(\mathbf{k}) \approx -4t + \hbar^2 k^2/(2m)$, where $m = \hbar^2/(2ta^2)$, the spectrum should consist of equidistant Landau levels $E_n = -4t + (n + \frac{1}{2})\hbar\omega_c$ with the cyclotron frequency $\hbar\omega_c = eB/m$. Indeed, as the magnetic field is turned on, the (almost) constant 2D DOS of the free electrons fractionalizes into a sum of discrete peaks (here broadened by $\eta = 0.01$). As B increases, the spacing between levels increases linearly, as does the DOS at the peak values. To check this, in fig. 3(b) we plot the distance between the location of the lowest energy peak and the bottom of the band, which should be $\Delta E = \hbar\omega_c/2 = 2\pi t\phi/\phi_0$. The thick line shows this prediction, in excellent agreement with the data (circles) for small B. As B increases, non-parabolic corrections kick in and differences become apparent (this is also seen in panel (a), where the $B = 0$ DOS is not constant, as expected in the parabolic limit). As a technical aside, this data was obtained in minutes on a regular desktop, starting from $A_N = 0$ for $N \sim 200$. As gaps open for larger ϕ/ϕ_0 values the convergence is much faster: for $\phi/\phi_0 = 0.01$ a cutoff $N \sim 50$ gives already perfectly converged data.

The same method can be used to look at the Hofstadter butterfly, for large ϕ/ϕ_0. As explained, if $\phi/\phi_0 = p/q$ we expect q distinct (possibly touching) bands. This is demonstrated in fig. 4(a) for $q = 2$ and $q = 3$. In fig. 4(b) we show the high sensitivity to the value of q: changing ϕ/ϕ_0 from 1/3 to 10/31 should result in the 3 bands splitting into 31. Indeed, the lower band (and by symmetry, the upper one) split into 10 new bands each, whereas the central band splits into 11 new bands (not shown).

While all of this is well known, it is worth emphasizing again that all these results are obtained from the same code, by simply changing the Peierls phases appropriately. The method works just as well for $\phi/\phi_0 = 1/1000$ (smallest value shown in fig. 3(b), which has 1000 bands and a magnetic unit cell of size 1000) as it does for $\phi/\phi_0 = 1/2$. In fact, one does not even need to choose $\phi/\phi_0 = p/q$ (however, in a computer any number is approximated to a rational value). Moreover, the method can deal just as easily with *inhomogeneous* magnetic fields —the LDOS will then vary in space, but it can be calculated anywhere with the same method, simply adjusting the Peierls phases.

Another advantage becomes apparent when considering that magnetic fields can just as easily be treated in 3D systems (and as shown below, on many types of lattices), for any orientation of $\mathbf{B}$ desired. Again, this simply modifies hopping integrals but nothing else. Because we obtain directly the density of states, it is then trivial to estimate the location of the Fermi energy for any

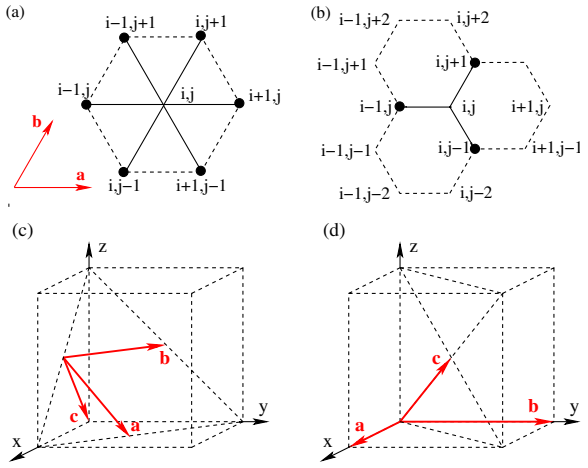

Fig. 5: (Color online) Indexing of sites on (a) 2D triangular; (b) 2D honeycomb; (c) 3D face-centered, and (d) 3D body-centered lattices, for which nn hopping links only sites with "Manhattan distances" $i+j$, respectively $i+j+k$, that vary by ± 1 or 0.

desired electron concentration, and calculate the magnetic field dependence of various thermodynamic and transport quantities. As shown in 2D, this can be done for arbitrarily small magnetic fields, and it therefore permits the easy study of quantum oscillations such as the de Haas - van Alphen phenomenology, on realistic discrete lattices, away from simple continuous models.

Indeed, the method also generalizes to other lattices, such as triangular and honeycomb in 2D and BCC and FCC in 3D. The reason is that nn hopping on these lattices also changes a certain "Manhattan distance" only by ± 1 or 0, if the lattice sites are indexed appropriately. This indexing is shown in fig. 5(a) and (b) for triangular and honeycomb lattices, and indeed links any site ij with a "Manhattan distance" $n = i+j$ to only sites with Manhattan distance $n \pm 1$ (for honeycomb) and also n, for triangular lattices. In the latter case, the recurrence relation changes to $\gamma_n(z)V_n = \alpha_n(z)V_{n-1} + \beta_n(z)V_{n+1}$, where γ_n is another known sparse matrix. The solution is still of the general form $V_n = A_n(z)V_{n-1}$, but now $A_n(z) = [\gamma_n(z) - \beta_n(z)A_{n+1}(z)]^{-1}\alpha_n(z)$. The calculation therefore proceeds in essentially the same way.

The same is true for BCC and FCC lattices, if one uses the "Manhattan distance" $n = i+j+k$ where the lattice sites $\mathbf{R}_{ijk} = i\mathbf{a} + j\mathbf{b} + k\mathbf{c}$ are defined in terms of the unit vectors shown in figs. 5(c), (d). For the FCC lattice, nn hopping links site (i,j,k) to its 6 nn in the same (111) triangular layer, namely $(i \pm 1, j, k)$; $(i, j \pm 1, k)$ and $(i \pm 1, j \mp 1, k)$; and 3 nn each in the (111) layer above, $(i,j,k+1)$, $(i-1,j,k+1)$, $(i,j-1,k+1)$, respectively below, $(i,j,k-1), (i+1,j,k-1), (i,j+1,k-1)$. All 12 nn sites have Manhattan distances of either $n-1, n$ or $n+1$. Similarly, for the BCC lattice a site (i,j,k) is linked to its 4 nn sites in the layer above, $(i,j,k+1), (i-1,j,k+1), (i,j-1,k+1)$ and $(i-1,j-1,k+1)$, respectively in the layer below, $(i,j,k-1), (i+1,j,k-1), (i,j+1,k-1)$ and $(i+1,j+1,k-1)$. All these neighbours have Manhattan distances of either n or $n \pm 1$, so the method works in the same way. Again, generalizations to anisotropic hoppings, addition of magnetic fields or spin-orbit coupling, multiple orbitals per site, disorder, etc. are all straightforward.

Finally, while we focused here on bulk calculations, the method also works near a surface, which is defined by the absence of hopping across it (other parameters near the interface can be changed as well). None of this affects the overall structure of recurrence relations. If the surface is smooth and the "in-plane" momentum $\mathbf{k}_{\|}$ is a good quantum number, then for any value of $\mathbf{k}_{\|}$ one needs to solve a 1D recurrence relation tracking the distance from the surface, which often can be solved analytically [4]. This allows one to look at the spectrum of surface states and their dispersion with the in-plane momentum for a *single surface*, since the system is semi-infinite. This is important for identifying topological insulators, and our method is much more efficient than the brute force diagonalization of finite-size systems, currently used in the literature.

In conclusion, we have presented a new efficient way to compute Green's functions for models with nn hopping, solving a long-standing cumbersome numerical challenge.

* * *

This work was supported by NSERC, CIfAR and CFI.

REFERENCES

[1] For a brief review, see Mamedov B. A. and Askerov I. M., *Int. J. Theor. Phys.*, **47** (2008) 2945.
[2] Economou E. N., *Green's Functions in Quantum Physics* (Springer-Verlag, Berlin) 1983.
[3] Einstein T. L. and Schrieffer J. R., *Phys. Rev. B*, **7** (1973) 3629; Fiete G. A. and Heller E. J., *Rev. Mod. Phys.*, **75** (2003) 933.
[4] Goodvin G. L., Covaci L. and Berciu M., *Phys. Rev. Lett.*, **103** (2009) 176402.
[5] Sawatzky G. A., *Phys. Rev. Lett.*, **39** (1977) 504; Berciu M. and Goodvin G. L., *Phys. Rev. B*, **76** (2007) 165109; Berciu M. and Fehske H., *Phys. Rev. B*, **82** (2010) 085116.
[6] Datta S., *Electronic Transport in Mesoscopic Systems* (Cambridge University Press, Cambridge, UK) 1977.
[7] Georges A. *et al.*, *Rev. Mod. Phys.*, **68** (1996) 13.
[8] This method also generalizes to interacting systems, see Berciu M., unpublished.
[9] Hofstadter D. R., *Phys. Rev. B*, **14** (1976) 2239.
[10] Morita T., *J. Math. Phys.*, **12** (1971) 1744.
[11] Berciu M., *J. Phys. A: Math. Theor.*, **42** (2009) 395207.
[12] Morita T., *J. Korean Phys. Soc.*, **6** (2002) 1015; *Interdiscip. Inf. Sci.*, **2** (1996) 63.
[13] Haydock R., Heine V. and Kelly M. J., *J. Phys. C: Solid State Phys.*, **8** (1975) 2591.
[14] Lanczos C., *J. Res. Natl. Bur. Stand.*, **45** (1950) 255.
[15] Covaci L. and Berciu M., *Phys. Rev. Lett.*, **102** (2009) 186403.
[16] Landau L. D., *Z. Phys.*, **64** (1930) 629.