Transfer-matrix methodology with stability-control techniques

A. MAYER & J.-P. VIGNERON
Laboratoire de Physique du Solide
Facultés Universitaires Notre-Dame de la Paix
Rue de Bruxelles, 61 ; B-5000 Namur
BELGIUM

Abstract: The transfer-matrix methodology is frequently used to deal with elastic scattering problems that require a solution of the Schrödinger or homogeneous Maxwell equations in the continuous part of their spectra. In its basic formulation, this technique however reveals limited by numerical instabilities, which can be drastically reduced by application of the layer addition algorithm. By providing each transfer matrix with an accuracy estimator, it is possible to predict and monitor the precision of the computation as a function of the number of layers, thus enabling a quantitative control of the stability. Beside this control of stability, other recent progress were achieved in the transfer-matrix methodology, i.e. improvements due to group theory, consideration of non-square matrices, combination with the Green-function formalism. This paper is essentially an overview of these various extensions of the method, which are illustrated by simulations of electronic scattering by a C$_{60}$ molecule in a projection configuration.

Key-Words: electronic scattering, non-square transfer matrix, numerical stability, Green functions, group theory

1 Introduction
Linear systems of differential equations are frequently encountered in theoretical physics. Such equations indeed appear when dealing with the Schrödinger equation in quantum mechanics or with the homogeneous Maxwell equations in electromagnetism. Solving a scattering problem, whose relevant propagation equations are linear, can be achieved within the transfer-matrix methodology [1-2] when the physical system considered is located between two separate boundaries.

In its basic formulation, this technique however reveals limited by numerical instabilities, whose origin can be traced to the exponential behavior of solutions in regions where tunneling is encountered. The layer addition algorithm, introduced by Pendry [3] in dynamic low-energy electron diffraction (LEED) computations, improves drastically the numerical stability of the transfer-matrix algorithm by subdividing the diffusive part of the system into several adjacent layers.

Recent developments [4] enable to control quantitatively the stabilizing effect of this algorithm. By providing each transfer-matrix with an accuracy estimator, it is indeed possible to monitor the precision of all quantities during the computation and thus provide final results with an estimation of accuracy. The dependence of this accuracy on the number of layers is explained by a physical model which predicts the number of layers to consider in order to achieve a given precision. Applying the layer addition algorithm with these accuracy-control techniques is of high interest, since the performance and range of the technique are widely increased.

This paper is essentially an overview of recent progress achieved in the transfer-matrix methodology, i.e. improvements due to group theory, consideration of non-square matrices, stability control, combination with the Green-function formalism. The methodology is presented in Sec. 2. The layer addition algorithm and the theoretical material needed to achieve a control of accuracy are given in Sec 3. The combination of the transfer-matrix method with the Green-function formalism is the subject of Sec. 4. These various extensions of the methodology are illustrated in Sec. 5 by simulations of electronic field-emission [5] and diffraction by a C$_{60}$ molecule in a projection configuration.
2 The transfer-matrix methodology
2.1 Objective
Let us consider the general problem of scattering from Region I (0 ≤ D) through Region II (0 ≤ z ≤ D) to Region III (z ≥ D). Let \{\Psi_j\} be a set of basic states used for the expansion of the wave function. If the physical system is characterized by a given set of basic states \{\Psi_j\} into independent sets \{\Psi_{jk}\} corresponding to the various inequivalent irreducible representations of that symmetry \[6\] :

\[
\Psi_{jk} = \frac{1}{g} \sum_{R} \chi^*_k(R) \tilde{R} \Psi_j
\]  

where g is the order of the symmetry group and \(\chi_k(R)\) the character associated to the operator \(\tilde{R}\) in the 4th representation.

By writing these basic states with « + », « - », « I » or « III » superscripts according to whether they describe propagation towards positive (+) or negative (-) z values in Region I or III, the solutions corresponding to scattering boundary conditions take the form:

\[
\Psi_{jk} = \Psi_{j(k)}^{I+} + \sum_i t_{i,k}^{I+} \Psi_{i(k)}^{I-} = \sum_i t_{i,k}^{I+} \Psi_{jk}^{III+} \]

(2)

\[
\Psi_{jk} = \Psi_{j(k)}^{I+} + \sum_i t_{i,k}^{I-} \Psi_{i(k)}^{I-} = \sum_i t_{i,k}^{I-} \Psi_{jk}^{III-} \]

(3)

where only basic states belonging to the same set \(k\) are involved.

2.2 Derivation by numerical propagation
The usual way to derive the solutions given in Eqs. (2) and (3) consists in deriving intermediate solutions corresponding to a single outgoing basic state:

\[
\Psi_{j(k)}^{I+} = \sum_i A_{i,j}^{I+} \Psi_{i(k)}^{I+} + \sum_i B_{i,j}^{I-} \Psi_{i(k)}^{I-} = \Psi_{j(k)}^{III+} \]

(4)

\[
\Psi_{j(k)}^{I-} = \sum_i A_{i,j}^{I-} \Psi_{i(k)}^{I-} + \sum_i B_{i,j}^{I-} \Psi_{i(k)}^{I-} = \Psi_{j(k)}^{III-} \]

(5)

When the number of basic states in Region I is the same as in Region III, the four transfer matrices associated to a given set \(k\) of basic states and defining the solutions in Eqs. (2) and (3) are derived from \(t_{+} = A^{-1}, t_{-} = B^{+} A^{-1}, t_{+} = A^{-1}\) and \(t_{-} = B^{+} A^{-1}\).

When the number of basic states in Region I is not the same as in Region III, the previous relations imply the inversion of non-square matrices and require therefore a generalization of the technique. This extension is developed in Ref. [7] and enable more freedom in choosing the representation of the wave function at the two sides of the diffusive part of the system.

2.3 Derivation by conjugate sets of solutions
For two sets of basic states to be conjugate, the corresponding values in the table of characters associated to the symmetry of the problem must be complex conjugate of each other. When the potential energy takes strict real values, it is possible to bypass the numerical propagation step of Sec. 2.2 for the sets of basic states that are conjugate to another set for which the transfer matrices are already computed. Applying this technique results essentially in a reduction of the total computation time. See Ref. [6] for details.

3 Stability control
One of the most significant progress achieved recently in the development of the transfer-matrix methodology is the quantitative control of accuracy that results from the application of the layer addition algorithm under the control of accuracy estimators. This section aims at presenting the layer addition algorithm, the techniques enabling a control of accuracy and practical results.

3.1 The layer addition algorithm
The numerical instabilities of the transfer-matrix method come basically from the fact that the matrices \(A^{+}\) and \(A^{-}\) to invert contain the amplitudes of solutions that grow exponentially (with various characteristic lengths) when tunneling is encountered. The condition number of these matrices is therefore an exponential function of the distance \(D\). The idea of the layer addition algorithm, introduced by Pendry [3] for dynamic low-energy electron diffraction (LEED) simulations, is to split the distance \(D\) into \(n\) adjacent layers and compute the transfer matrices associated to each slab separately. Since each slab is associated only to a fraction \(D/n\) of the length of the diffusive part of the system, the corresponding transfer matrices will achieve a better accuracy. In a final step, the transfer matrices associated to the \(n\) layers are combined to derive those corresponding to the whole distance \(D\).
The peculiar combination formula can be found in Ref. [3].

3.2 Mathematical representation of accuracy
To control the accuracy of a transfer-matrix computation, we relate the true (but unknown) components of a matrix \( A \) to those of its known representation \( \overline{A} \) by

\[
\overline{A}_{i,j} = (1 + \delta_{A,i,j}) A_{i,j}
\]

and define the corresponding accuracy estimator \( \varepsilon_A \) by a weighted average of the relative errors \( \delta_{A,i,j} \):

\[
\varepsilon_A = \frac{\sum_{i,j} |\delta_{A,i,j} A_{i,j}|}{\sum_{i,j} |A_{i,j}|}
\]

where the values \( \delta_{A,i,j} \) are weighted by the associated matrix element.

For a digital representation of numbers with \( n \) bit binary digits, the best precision a matrix \( A \) can achieve is that obtained when all these binary digits have a meaningful value, i.e. \( \varepsilon_{\text{comp}} = 2^{-n_{\text{bit}}} \).

When performing matrix operations (multiplication, addition, inversion), the accuracy estimators are updated according to [4]:

\[
\varepsilon_{AB} = \varepsilon_A + \varepsilon_B
\]

\[
\varepsilon_{A+B} = \frac{\sum_{i,j} \left| \frac{\varepsilon_A \overline{A}_{i,j} + \varepsilon_B \overline{B}_{i,j}}{\overline{A}_{i,j} + \overline{B}_{i,j}} \right| (\overline{A} + \overline{B})_{i,j}}{\sum_{i,j} |\overline{A} + \overline{B}|_{i,j}}
\]

\[
\varepsilon_{A^{-1}} = \varepsilon_A \text{cond}(A)
\]

where \( \text{cond}(A) \) is the condition number of the matrix \( A \).

3.3 Accuracy of the layer addition algorithm
The definitions and rules of Sec. 3.2 can be used to monitor the accuracy of all matrices involved by the application of the layer addition algorithm and thus provide a final result with a value for its average relative error.

By analyzing the physical origin of numerical instabilities, it is possible to predict the precision a transfer-matrix computation will achieve [4]. Indeed, the condition number of the matrices \( A^+ \) and \( A^- \) associated to a slab of length \( D/n \) can be estimated by the ratio between the minimal and maximal transmission through that slab, i.e. \( e^{\frac{K_{\text{max}} D}{n}} \) where \( K_{\text{max}} \) is the maximal wave vector encountered in the case of tunneling \( (K_{\text{max}} = \sqrt{\frac{2m}{\hbar^2}V} \) for a potential barrier with height \( V \)). By modeling the whole process [4], the relative error on the transfer matrices obtained by dividing the distance \( D \) into \( n \) adjacent layers is found to be:

\[
\varepsilon_n = 2^{-n_{\text{bit}}} e^{\frac{K_{\text{max}} D}{n}} \times \{\text{cond}(P)n^2 + [1 + \text{cond}(P)]n - 2\text{cond}(P)\}
\]

(10)

where \( \text{cond}(P) \) is the condition number of matrices describing multiple reflection between adjacent layers. It takes typical values around 5.

This result is important, since it predicts whether a transfer-matrix computation will be successful or not. It demonstrates the possibility to deal with large distances \( D \) by simply increasing the number \( n \) of subdivisions. For practical purposes, the minimal number of subdivisions to consider in order for a transfer-matrix computation to be successful is given by:

\[
n_{\text{min}} = \frac{K_{\text{max}} D}{n_{\text{bit}} \ln 2}
\]

(11)

We recommend to consider \( n=4n_{\text{min}} \) subdivisions.

4 Combination with the Green-function formalism
For problems requiring the propagation of the scattered wave function from the diffusive part of the system to large distances, the Green-function formalism [8] is well suited once the problem is solved in the diffuser. However this part of the procedure, when treated by this formalism, requires prohibitively large data storing when the diffuser is not localized.

The transfer-matrix methodology has significantly less storage requirements, since it just « connects » the basic states at the two sides of the diffuser, without providing and depending on the wave function in the diffuser. This technique is therefore better suited for solving the scattering problem in the diffuser. Once this first part is completed, the propagation from the diffuser can be achieved within the Green-function formalism by just using the values of the wave function and its derivatives on the external surface of the diffuser [9].
In fact, according to this formalism, the wave function at any point \( \vec{r} \) of space can be computed by the expression:

\[
\Psi(\vec{r}) = \frac{-\hbar^2}{2m} \int_S \mathbf{n} \cdot [G(\vec{r}, \vec{r}'), E] \nabla_{\vec{r}'} \Psi(\vec{r}') - \Psi(\vec{r}') \nabla_{\vec{r}'} G(\vec{r}, \vec{r}', E)] dS'
\]

where \( \mathbf{n} \) is a unit vector normal to the surface element \( dS' \) of a closed surface \( S \) surrounding the point \( \vec{r} \). This surface lies on the external surface \( z=D \) of the diffuser and is closed at infinite distance where the contributions of the wave function and its derivatives cancel. Within the Kirchhoff approximation \([10]\), we can use in the surface \( z=D \) the values of \( \Psi(\vec{r}') \) and \( \nabla_{\vec{r}'} \Psi(\vec{r}') \) provided by the transfer-matrix computation, so one finds:

\[
\Psi_{j(k)}^{\alpha}(\vec{r}) = \sum_i I_{i,j}^{\alpha} \sigma(\vec{r}, i[k], E)
\]

with

\[
\sigma(\vec{r}, i[k], E) = -\frac{\hbar^2}{2m} \int_S \mathbf{n} \cdot [G(\vec{r}, \vec{r}'), E] \nabla_{\vec{r}'} \Psi_{i(k)}^{\alpha}(\vec{r}') - \Psi_{i(k)}^{\alpha}(\vec{r}') \nabla_{\vec{r}'} G(\vec{r}, \vec{r}', E)] dS'
\]

5 Application: observation of a \( C_{60} \) molecule by projection microscopy

The concepts presented in Sec. 2, 3 and 4 have been developed in previous papers \([4-7,9,11-12]\) in order to simulate electronic field emission and the observation of carbon fibers by projection microscopy. In order to take advantage of the symmetry of the field-emission tip and the sample, the problem was treated in cylindrical coordinates \((\rho, \varphi, z)\) by assuming the central \(z\)-axis to describe a \( C_n \) symmetry. We will present briefly the model corresponding to this situation and the simulated observations of a \( C_{60} \) molecule by projection microscopy.

5.1 Main assumptions

The virtual projection microscope we consider consists of a field-emission tip (standing on a metallic holder), a conducting grid (that supports the sample) and a 10 cm distant screen. The metallic tip holder is limited by the plane \( z=0 \) and is referred to as « Region I ». The conducting grid is described by the specific form:

\[
\Psi_{(m,j)}^{\pm}(\vec{r}) = e^{\pm i \beta} \frac{J_m(k_{m,j}\rho)}{\sqrt{2\pi} \int_0^\rho \rho |J_m(k_{m,j}\rho)|^2 d\rho} e^{im\varphi}
\]

\[
\Psi_{(m,j)}^{\pm}(\vec{r}) = e^{\pm i \beta} \frac{J_m(k_{m,j}\rho)}{\sqrt{2\pi} \int_0^R \rho |J_m(k_{m,j}\rho)|^2 d\rho} e^{im\varphi}
\]

where the wave vectors \( k_{m,j} \) are solution of \( J'_m(k_{m,j}R) = 0 \). Since the wave function is not confined in Region III, we restrict expression (16) to \( z=D \).

5.2 Wave function expansion

To enforce a quantization of the basic states in Region I and II, the wave function is assumed to remain localized in a cylinder with radius \( R \) in these two regions. In this context, the basic states \( \{\psi\} \) used in Sec. 2 to expand the wave function take the specific form:

5.3 Simulation technique

The techniques presented in Sec. 2, 3 and 4 have been developed with the previously given wave function representation in Ref. \([4-7,9]\). The propagation equations are derived in Ref. \([11-12]\). In the case where the \(z\)-axis describes a \( C_n \) symmetry, the propagation equations reveal the existence of \( n \) sets of independent basic states, as predicted by group theory \([6]\). Each set is characterized by \( m \) subscripts that are separated by integer multiples of the symmetry axis order \( n \). The peculiar relations enabling the propagation of the wave function to the imaging screen by the Green-function formalism are given in Ref. \([9]\).

If we describe the tip by a cone and assume the \( C_{60} \) molecule to lay on a pentagon and be centered on \( z=D \). The region 0 \( \leq z \leq D \) contains the field-emission tip and the sample and is referred to as « Region II ». The free-propagation region \( z \geq D \) that extends to the screen is referred to as « Region III ».

Let us assume Region I to be a Sommerfeld metal characterized with empirical values for the Fermi energy \( E_F \) and work function \( W \). If \( V \) is the electric bias established between the tip holder and the conducting grid and if the potential energy in Region III is set conventionally to the constant value 0, the potential energy in Region I is \( V_{\text{met}} = eV - W - E_F \).
the tip axis, we can take advantage of the $C_3$ symmetry. As predicted by group theory, there are five sets $k$ of independent basic states that contain the following $m$ subscripts:

\begin{align*}
  k=0 & \quad m = ..., -10, 0, 5, 10, ...
  \\
  k=1 & \quad m = ..., -9, -4, 1, 6, ...
  \\
  k=2 & \quad m = ..., -8, -3, 2, 7, ...
  \\
  k=3 & \quad m = ..., -7, -2, 3, 8, ...
  \\
  k=4 & \quad m = ..., -6, -1, 4, 9, ...
\end{align*}

(17)

There are two pairs of conjugate sets: the couple of sets $k=1$ and $k=4$ and the couple of sets $k=2$ and $k=3$. The corresponding lines in the table of characters are complex conjugate of each other. The $m$ subscripts in two conjugate sets differ only by the sign. As stated in Sec. 2.3, the transfer matrices associated to a given set of basic states can be derived directly from those associated to its conjugate set. Since there are two pairs of conjugate sets among the five independent sets and the numerical propagation step is the most time-consuming part of the process, exploiting these pairs of conjugate sets enable the problem to be solved within only 3/5 of the time that would be required by carrying out a numerical propagation for each set of basic states.

5.4 Characteristics of the system

To characterize the metallic support of the tip, we considered a Fermi energy $E_F$ of 19.1 eV and a work function $W$ of 4.5 eV (values for tungsten). The conical tip has a height and basis diameter of 1 nm. The distance $D$ from the tip holder to the conducting grid is 4 nm and the extraction bias $V$ is 40 V.

The atoms of the $C_{60}$ molecule are modeled by gaussian electronic distributions that move rigidly around the positive nuclei as a result of the local electric field. For more details on the computation of the potential energy in Region II, see Ref. [13].

5.5 Results

The potential-energy distribution in Region II is illustrated in Fig. 1 by a vertical section in the XZ plane. The figure reveals a potential-energy distribution that presents strong variations at the atomic positions.

The total current density on the conducting grid $z=D$, which results from the transfer-matrix computation, is presented in Fig. 2. This figure shows the electronic beam that results from the field-emission process. The variations in the central part of the figure are related to the atomic structure of the $C_{60}$ molecule. The 5-fold symmetry is clearly visible.

![Figure 1: Potential-energy distribution in the XZ plane. A $C_{60}$ molecule is observed in a virtual projection microscope. The extraction bias is 40 V.](image1)

![Figure 2: Current density (z-component) on the conducting grid. A $C_{60}$ molecule is observed in a virtual projection microscope. The extraction bias is 40 V.](image2)

The total current density on the 10 cm distant imaging screen is presented in Fig. 3. The figure reveals only the global spherical shape of the molecule, despite the fact that the atomic structure appears in the potential-energy distribution and in the current density on the conducting grid.

The lack of information on the atomic structure of the molecule in the projected image is however expected since the resolution limit due to diffraction, which is given by $\Delta_d = \frac{1}{2} \sqrt{\lambda d}$ where $\lambda$ is the
electronic wavelength in the sample and \( d \) the distance between the sample and the virtual point source, takes here the value \( \Delta d = 0.37 \text{ nm} \). This value is not small enough to enable the observation of the atomic structure of the C\(_{60}\) molecule.

Figure 3: Current density (\( r \)-component) on the 10 cm distant imaging screen. A C\(_{60}\) molecule is observed in a virtual projection microscope.

The extraction bias is 40 V.

6 Conclusion

The transfer-matrix methodology and recent extensions of this technique were presented. These extensions include the construction of independent sets of basic states by group theory, the generalization of the formalism to non-square matrices, the concept of conjugate sets, the control of accuracy and the combination with the Green-function formalism.

These various concepts were applied to simulate the observation of a C\(_{60}\) molecule by projection microscopy. In the situation considered, the atomic structure of the molecule, although present in the local current density, does not induce visible effects in the projected image.

Acknowledgments: A.M. was supported by the Belgian National Fund for Scientific Research (FNRS). The authors acknowledge the national program on the Interuniversity Research Project (PAI) and the use of the Namur Scientific Computing Facility, a common project between the FNRS, IBM-Belgium and the FUNDP.

References: