Ill-conditioned matrices in the scattering of waves from hard corrugated surfaces

N. Garcia,* V. Celli, and N. R. Hill
Department of Physics, University of Virginia, Charlottesville, Virginia 22901

N. Cabrera
Departamento de Fisica Fundamental, Universidad Autonoma de Madrid, Canto Blanco, (Madrid-34) Spain
(Received 25 May 1978)

In this paper we present an analysis of the conditioning character of the matrices obtained in the different methods for solving the integral equations that appear in the scattering of waves from a hard corrugated surface. It is found that the matrices derived from the Rayleigh hypothesis become ill conditioned for large corrugations when the approach is not valid at any rate. The method based on the extinction theorem developed by Masel, Merrill, and Miller is analytically correct, but it becomes ill conditioned and no practical numerical solutions can be obtained for large values of the corrugation strength, depending on the precision of the computer used. Finally the self-consistent method with the numerical procedure developed by Garcia and Cabrera is well conditioned and leads always to a good practical numerical solution for all kinds of corrugations, no matter how large.

I. INTRODUCTION

The problem of wave scattering from a hard corrugated surface (HCS) arises in optics, acoustics, wave-guide theory, and in the quantum theory of atom-surface interactions.[1-17] In this paper we discuss in detail the scattering of a scalar wave (ψ) with Dirichlet boundary conditions $\psi(x, z = D(x)) = 0$ on the surface, where x and z are coordinates parallel and perpendicular to the average direction of the surface, and $D(x)$ is the surface profile, assumed one dimensional, and with periodicity a. This model corresponds to an infinitely repulsive corrugated wall and is of current interest because it has been successfully used in the analysis of atom-surface scattering.[10,16,17] The general analysis, however, is also valid for the other problems mentioned above.

In general the scattering problem can be reduced to solving a Fredholm integral equation for the sources of the scattered field,[11,14] or an equivalent infinite set of linear equations.[10] Numerical problems arise in the inversion of the large finite matrices obtained by truncating the infinite sets. The precision used and the round-off errors play a very important role if the problem leads to ill-conditioned[18-21] matrices. In this case, small variations in the coefficients of the linear equations or in the independent term imply large variations in the unknowns. If the coefficients, for instance, are not exactly known, but are obtained after numerical manipulations, the solution obtained can be totally unreliable.

The aim of this paper is to analyze in detail and discuss the conditioning of the matrices obtained from the different methods used to solve the scattering of atoms from HCS. We delve into this problem because experience has shown that some of the methods, while analytically exact, do not lead always to convergent numerical solutions[11,12] while another exact method[14] gives always good numerical answers for any strength of the corrugation parameter $h = \zeta_0/a$ [where ζ_0 is the difference between the maximum and the minimum values of the corrugated surface $D(x)$].

In Sec. II we will discuss the different equations corresponding to the three proposed methods: (i) the Rayleigh approach[1] that is known to be exact, for small values of h for analytical $D(x)$ (Ref. 5); (ii) the exact method based on the extinction theorem developed first by Masel, Merrill, and Miller[11] (MMM), extended and applied later by Goodman[12]; and (iii) the method which has been developed by Garcia and Cabrera[14] (GC method) based on the self-consistent boundary condition obtained from Huygens' principle. The ill-conditioning criteria to be used in the computations as well as some definitions of matrix norms will be given (after Wilkinson[21]) in Sec. III. Computations using configuration and momentum-space approaches for (i)[9,10] and (iii)[14] and the proposed matrix[11,12] for (ii) are presented in Sec. IV. All the calculations are carried out for a cosinelike corrugation function:

$$D(x) = ah\cos(2\pi x/a), \tag{1}$$

which is extensively used in applications to atom-surface interactions. Finally, conclusions will be drawn in Sec. V.

II. METHODS OF SOLUTION

We describe very briefly the different methods that have been proposed to solve the scattering of

waves from a HCS. We consider an incident plane wave $e^{i\vec{k}\cdot\vec{r}}$, where $(K, -k_g)$ are the parallel and perpendicular components of the wave vector $\vec{k}$, $\vec{r} = (x, z)$ is the position vector and only two dimensions have been considered for simplicity. Because of the periodicity in the x direction, the parallel momentum for outgoing waves must be of the form $K+G$, where

$$G = 2\pi n/a \tag{2}$$

and n is an integer. Because of energy conservation, the perpendicular component of the momentum for the G beam, k_{Gg}, is given by

$$k_{Gg}^2 = k^2 - (K+G)^2 . \tag{3}$$

(i) Rayleigh method. It consists in solving for the A_G values in the following equation:

$$\sum_G A_G e^{i[k_{Gz}D(x)+Gx]} = -e^{-ik_gD(x)} . \tag{4}$$

Different numerical methods have been proposed for solving Eq. (4),[9,10,17] but the simple one that can handle any corrugation is the GR procedure,[9,10] in which the left-hand side of (4) is evaluated at a number of space points (R points) equal to the number of coefficients A_G that are kept in the expansion.

(ii) MMM solution. It consists in finding the value for the sources $f(x)$ from the integral equation

$$\delta_{G0} = \frac{1}{k_{Gg}a}\int_0^a f(x)e^{i(K+G)x}e^{ik_{Gg}D(x)}\,dx . \tag{5}$$

This is an exact solution obtained by the application of the extinction theorem.[11-14] The numerical procedures proposed determine the Fourier transform of $f(x)$, f_G, but then the method is applicable in practice only when the Fourier transform of $e^{ik_{Gg}D(x)}$ is known in closed form.

(iii) GC self-consistent solution. This is also based on finding the source function $f(x)$ from the integral equation

$$e^{i[Kx-k_gD(x)]} = \frac{1}{a}\int_0^a f(x')\sum_G \frac{e^{i(K+G)(x-x')}}{k_{Gg}} e^{ik_{Gg}|D(x)-D(x')|}\,dx' . \tag{6}$$

This is clearly an exact method, and in fact convergent solutions have been obtained without limit in the corrugation strength or in shape of the corrugation[14] by applying the RR numerical procedure consisting in discretizing the continuous variables x, x'.

The connection between the three methods has been discussed by Toigo *et al.*[13] Their complete equivalence for sufficiently small corrugations has been shown recently by expanding the exponentials in Eqs. (4)–(6) in power series.[15] It turns out that an iterative series for the unknown coefficients can be obtained in closed form, but the convergence of the series is limited to certain values of the corrugation strengths.

From previous work on this problem the following points have emerged:

(a) Equations (4) and (5) lead to some matrix elements that are very big and others that are very small because k_{Gg} becomes imaginary [see Eq. (3)] for large values of G, going to infinity when $G \to \infty$. Obviously this will create numerical troubles.

(b) In addition, for the corrugation (1) the Rayleigh method is incorrect for $h \gtrsim 0.072$.[5,22]

(c) The matrix elements obtained from the integral equation (6), however, are always well behaved and are bounded because of the modulus in the exponential. Hence the summation over G converges very rapidly.

(d) Even though formula (5) is exact, it does not lead to convergent solutions for large corrugations. The result depends drastically on the capacity of the computer and on the numerical procedure used.[12,14,17,23] Examples of this behavior are discussed in Sec. V of this paper.

(e) In a typical case, we found that by adding an independent term with $G \neq 0$ of the order of 10^{-6} in the integral equation (5), we obtain variations of 10^{-2} in the Fourier components of $f(x)$.

All these points make us suspect an ill-conditioned behavior of Eqs. (4) and (5), while integral equation (6) because of the bounded character of its kernel should be well conditioned leading always to good solutions.

III. MATRIX NORMS AND ILL-CONDITIONING CRITERIA

We give a short review of matricial norms, of the condition number and of its value for each different norm, following Chaps. 1, 2, and 4 of Wilkinson's book.[21]

The norm is a single number that gives an overall assessment of the size of a vector $\vec{X}$ or a matrix A. The norm $||\vec{X}||$ of a vector $\vec{X}$ must satisfy the following inequalities:

$$\begin{aligned} &||\vec{X}|| > 0 \quad \text{unless } \vec{X} = 0, \\ &||R\vec{X}|| = |R|\,||\vec{X}|| \quad \text{for any complex scalar } R, \\ &||\vec{X}+\vec{Y}|| \leq ||\vec{X}|| + ||\vec{Y}|| \quad \text{(triangular inequality)}. \end{aligned} \tag{7}$$

The norm $||A||$ of a matrix A must satisfy $||A|| > 0$ and

$$||AB|| \leq ||A||\,||B|| . \tag{8}$$

If the condition $||A\vec{X}|| \leq ||A||\,||\vec{X}||$ is satisfied for all A and $\vec{X}$, the two norms $||\vec{X}||$ and $||A||$ are

said to be compatible. If

$$||A|| = \max_{X\neq 0}(||A\vec{X}||/||\vec{X}||),$$

then the norm of A is subordinate to the norm of $\vec{X}$. For any subordinate norm $||I|| = 1$, where I is the unit matrix.

A useful matrix norm is the spectral norm

$$||A||_2 = (\text{max eigenvalue of } A^\dagger A)^{1/2}, \tag{9}$$

which is subordinate to the vectorial norm

$$||\vec{X}||_2 = (|X_1|^2 + |X_2|^2 + \cdots + |X_n|^2)^{1/2}. \tag{10}$$

We will find it useful to consider the Euclidean or Schur norm:

$$||A||_E = \left(\sum_{ij} |a_{ij}|^2\right)^{1/2}, \tag{11}$$

where a_{ij} is the ij element of the matrix A. This is not subordinate (because $||I||_E = n^{1/2}$) but is compatible with $||X||_2$. Since

$$||A||_E^2 = \mathrm{Tr}(A^\dagger A) = \sum_i \sigma_i^2 \tag{12}$$

(where σ_i^2 are the eigenvalues of $A^\dagger A$), we also have

$$||A||_2 \leq ||A||_E \leq n^{1/2}||A||_2. \tag{13}$$

The above norms are invariant under unitary transformations. Let us now consider the system of linear equations

$$A\vec{X} = \vec{b}, \tag{14}$$

and determine the sensitivity of the solution $\vec{X}$ to small changes in $\vec{b}$, so that (14) becomes

$$A(\vec{X} + \Delta\vec{X}) = \vec{b} + \Delta\vec{b}. \tag{15}$$

By taking norms in (15) it is easy to show that

$$||\Delta\vec{X}||/||\vec{X}|| \leq \kappa(A)(||\Delta\vec{b}||/||\vec{b}||), \tag{16}$$

where

$$\kappa(A) = ||A||\,||A^{-1}|| \tag{17}$$

is the condition number. A small value of $\kappa(A)$ assures that the system is well conditioned. For a small change in the matrix A, it turns out that again $\kappa(A)$ is the magnitude that determines the relative error $||\Delta\vec{X}||/||\vec{X}||$. It is easy to prove that for the spectral norm $||A||_2$ the condition number is

$$\kappa_2(A) = \sigma_1/\sigma_n,$$

where σ_1^2 and σ_n^2 are the largest and smallest eigenvalues of $A^\dagger A$. This is the usual way of defining the condition number but to find the eigenvalues is rather laborious. If A is Hermitian, $\kappa_2(A) = |\lambda_1|/|\lambda_n|$, where λ_1 and λ_n are the largest and smallest eigenvalues of A. From the relation (13) it follows that

$$\kappa_2(A) \leq \kappa_E(A) \leq n\kappa_2(A). \tag{18}$$

Then $\kappa_E(A)$ gives bounds for $\kappa_2(A)$ within the order of the matrix n. This factor comes because the two norms of the identity matrix are related by $||I||_E = n^{1/2}||I||_2$. The condition numbers defined above are not easy to compute. In fact, finding A^{-1} is essentially equivalent to finding the solution of (14). By working with Euclidean norms, it is possible to obtain a simpler criterion that requires only the computation of the determinant Δ of the normalized matrix

$$A_N = \begin{bmatrix} \frac{a_{11}}{\alpha_1} & \frac{a_{12}}{\alpha_1} & \cdots & \frac{a_{1n}}{\alpha_1} \\ \frac{a_{21}}{\alpha_2} & \frac{a_{22}}{\alpha_2} & \cdots & \frac{a_{2n}}{\alpha_2} \\ \vdots & \vdots & & \vdots \\ \frac{a_{n1}}{\alpha_n} & \frac{a_{n2}}{\alpha_n} & \cdots & \frac{a_{nn}}{\alpha_n} \end{bmatrix}, \tag{19}$$

where

$$\alpha_i^2 = \sum_j |a_{ij}|^2. \tag{20}$$

The normalization is chosen so that $||A_N||_E = n^{1/2}$ and it implies that $|\Delta| \leq 1$.

Writing (14) in the form $A_N\vec{X} = \vec{b}_N$, with $b_{Ni} = b_i/\alpha i$, we find, analogously to (16), the error bound

$$||\Delta\vec{X}||/||\vec{X}|| \leq \kappa(A_N)(||\Delta\vec{b}_N||/||\vec{b}_N||). \tag{21}$$

The new Euclidean condition number is

$$\kappa_E(A_N) = n^{1/2}||B||_E/|\Delta|, \tag{22}$$

where B is the matrix of the minors of A_N. Now $||B||_E^2$ is bounded by n^2M^2, where M is the modulus of the largest minor. Since all the minors of A_N have modulus no greater than one, it follows that $||B||_E^2 \leq n^2$ and

$$\kappa_E(A_N) \leq n^{3/2}/|\Delta|. \tag{23}$$

Thus one can say that $|\Delta| \ll 1$ is a necessary requirement for the system (14) to be ill conditioned. Furthermore the condition (21) indicates that an error in a component b_n is more important if the corresponding α_n is small. But Eq. (23) represents an upper bound of $\kappa_E(A_N)$. As will be shown in the calculation the bound is much larger than the actual $\kappa_E(A_N)$.

IV. COMPUTATIONAL RESULTS

We now proceed to discuss representative computations carried out on the different matrices. We have first calculated the determinant Δ of A_N given by (19) to have some indication of ill con-

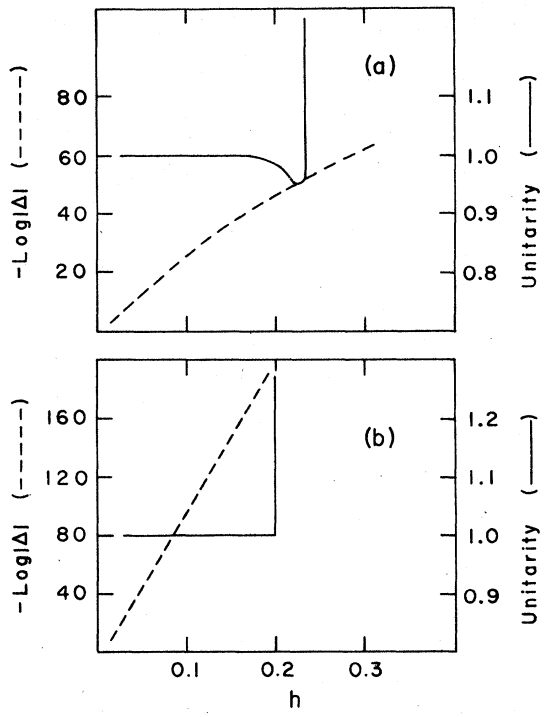

FIG. 1. Conditioning and unitarity criteria for the MMM method with $ka = 40$ applied to a sinusoidal profile for matrices of rank $N = 31$ (a) and $N = 51$ (b). The dashed line gives $-\ln|\Delta|$ and the continuous line gives the total reflected intensity as a function of h.

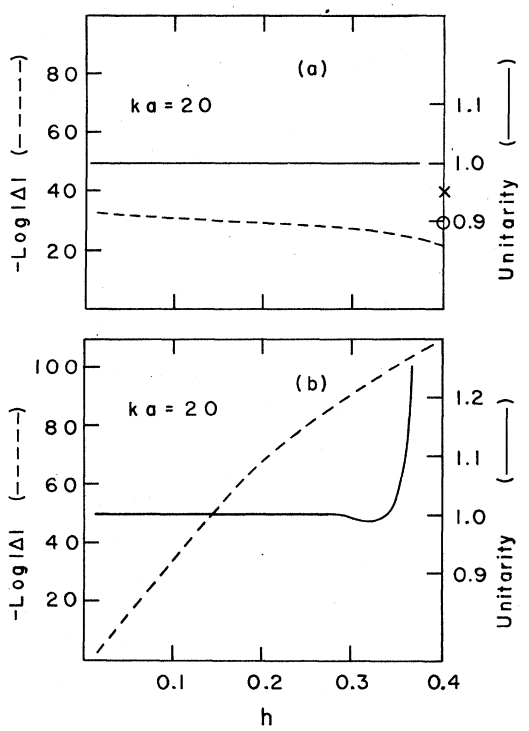

FIG. 2. Comparison of the GC method (a) with the MMM method (b). The surface profile and the symbols are the same as in Fig. 1. The GC calculation was carried out for a matrix of rank $N = 60$ using $N_G = 40$ G vectors in the evaluation of the Green's function; the results were tested at $h = 0.4$ as a function of N and N_G, as shown by × for $N = 80$, $N_G = 60$, and 0 for $N = 80$, $N_G = 80$.

ditioning because of the simplicity of this test.

Some results for the MMM method[11] are shown in Fig. 1. The dashed line in Fig. 1(a) represents $-\ln|\Delta|$ as a function of h using 31 G vectors for $ka = 40$. We observe that as h increase Δ decreases very rapidly. The continuous line is a plot of the total reflected intensity that must equal unity for the elastic scattering being considered (unitarity condition). We observe that the calculations fail for $h > 0.20$. Figure 1(b) shows that Δ becomes smaller, so that more ill conditioning appears, when the number of G vectors (i.e., the rank of the matrix) is increased to 51. This was suggested in Sec. II given the structure of the matrix elements that diverge for $G \to \infty$. The unitarity test again fails at $h = 0.20$. The value of Δ and the calculated norm (see Fig. 4 below) are only indicative of a rapidly increasing behavior, for $h > 0.20$, and should not be regarded as accurate.

Figure 2(a) gives the same plots as in Fig. 1 but for the GC self-consistent method[14] and $ka = 20$. We observe that the value of Δ keeps practically constant as h increases, and no changes are noted when increasing the dimension of the RR' matrix[14] or increasing the number of G vectors considered in summation (6). The unitarity condition (continuous curve) is always satisfied. The method seems to be well conditioned, as is to be expected from the fact, discussed in Sec. II, that the kernel of integral equation (6) is bounded. For comparison, Fig. 2(b) gives the same plots for the MMM method. Here Δ is also fastly decreasing, the unitarity improves a little with respect to Fig. 1(a) and 1(b) but it fails when $h \geq 0.30$.

This first test made us consider a further analysis of the condition number $\kappa(A)$ to infer more clearly the ill-conditioning character of the MMM method even if it is analytically exact. We calculate the Euclidean condition number $\kappa_E(A)$ but as formula (18) shows the difference with $\kappa_2(A)$ is relatively small.

We report first results for the Rayleigh method using the GR numerical procedure.[10] Figure 3 gives plots of the condition number $\kappa_E(A)$ for $ka = 20$ (a) and $ka = 40$ (b) as a function of h. Note that this increases as h does, giving ill conditioning as suggested in II. Again the matrix elements blow up or are very small as h increases and $G \to \infty$. The plots are done up to $h = 0.1$ because the method

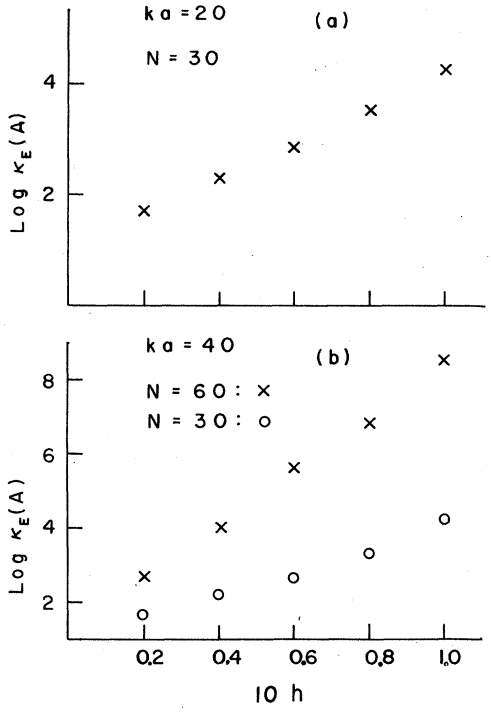

FIG. 3. Euclidean condition number $\kappa_E(A)$ as a function of h for the Rayleigh method using the GR procedure, for different values of ka and of rank N of the matrix A, as indicated.

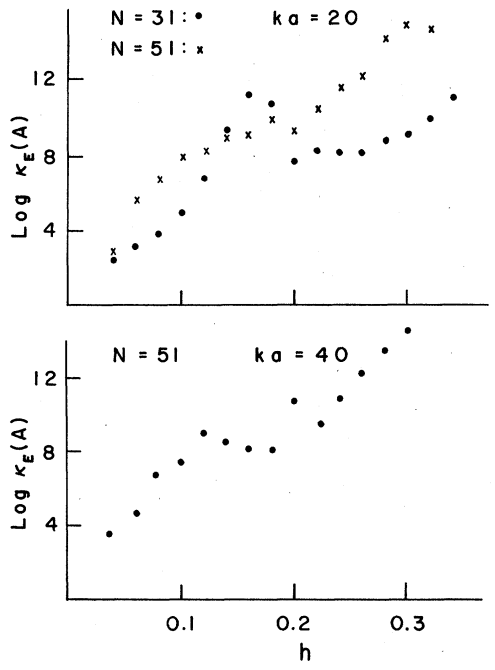

FIG. 4. As Fig. 3, but for the MMM method.

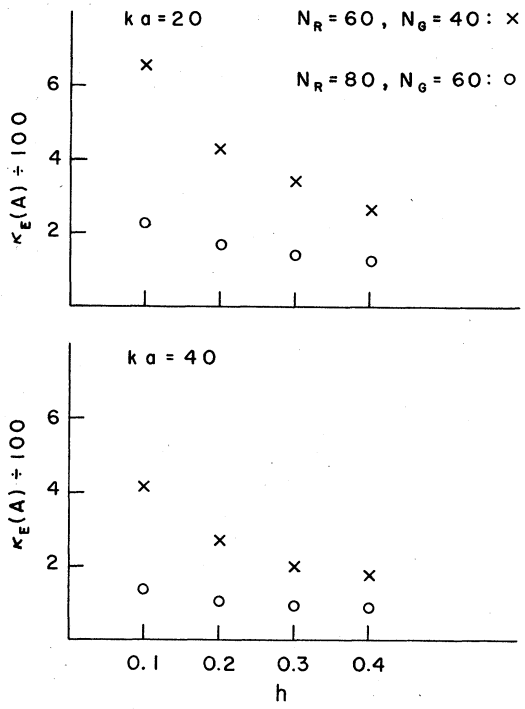

FIG. 5. As Fig. 3, but for the GC method.

is analytically incorrect when $h > 0.072$. It can be seen that the ill-conditioning character becomes worse as the rank of the GR matrix increases.

Figure 4(a) gives the condition number $\kappa_E(A)$ for the MMM method versus the value of h for two values of ka and of the number of G vectors used (which equals the rank of the A matrix). Again $\kappa_E(A)$ increases as h does, implicating ill conditioning as in Figs. 1 and 2. (The method is definitely ill conditioned.) A very interesting point emerges by relating Figs. 1, 2, and 4; unitarity fails (the method does not give good numerical results) whenever the condition number satisfies $\kappa_E(A) \gtrsim 10^{10}$–$10^{12}$. It is understood that the MMM method is valid for the range of h quoted here only when a CDC 6400 (16 digits precision) is used. In the IBM-360/65 the range of h is smaller.[24] This effect is a clear sign of ill conditioning.[18-21]

Finally in Fig. 5 we present the results for $\kappa_E(A(h))$ when using the GC self-consistent method and the numerical RR procedure. We commented before that the matrix elements arising from (6) are always bounded because of the modulus in the exponential[14] and consequently we predicted well-conditioned matrices. Figure 5 is a clear manifestation of this: the condition number decreases as h increases no matter how big is the matrix and the incident energy. In fact, Garcia and Cabrera have been able to obtain good numerical solutions

without limitations in the shape or strength of the corrugation.[14]

V. CONCLUSIONS

In the light of the above results one can draw the following conclusions: (i) The Rayleigh method is ill conditioned but its condition number is well controlled by the computer precision for $h \leq 0.10$; for larger values of h where the condition number is bigger the method is incorrect in itself. It has been proved that by using a variational approach the method can become valid for all values of h.[5,23] However, computations show that the convergence is poor in agreement with previous suggestions.[5] (ii) The MMM method based on the extinction theorem though it gives in principle the correct analytical solutions. (iii) The only well-conditioned method that works without limitations on the corrugation is the GC self-consistent method; the *RR* numerical procedure leads always to a good convergent solution and has the advantage that it is applicable to any shape of the corrugation $D(x)$, while the MMM method is only useful in practice for very simple corrugations such as a sinusoid.

ACKNOWLEDGMENTS

We thank Professor F. O. Goodman for many useful discussions. This work was supported by the NSF under grant-contract DMR.

---

*Permanent address: Departamento de Fisica Fundamental, Universidad Autonoma de Madrid, Canto Blanco, (Madrid-34) Spain.

[1]T. W. Strutt (Baron Rayleigh), Proc. R. Soc. A 79, 399 (1907); *The Theory of Sound* (Dover, New York, 1945), Vol. II.

[2]B. A. Lippmann, J. Opt. Soc. Am. 43, 408 (1953).

[3]R. Petit and M. Cadilhac, C. R. Acad. Sci. B 262, 468 (1966).

[4]T. L. Uretsky, Ann. Phys. (N.Y.) 33, 400 (1965).

[5]R. F. Millar, Radio Sci. 8, 785 (1973), and references therein.

[6]P. C. Waterman, J. Acoust. Soc. Am. 57, 792 (1975).

[7]T. A. Desanto, J. Acoust. Soc. Am. 57, 1195 (1975).

[8]U. Garibaldi, A. C. Levi, R. Spadacini, and G. E. Tommei, Surf. Sci. 48, 649 (1975).

[9]N. Garcia, T. Ibanez, T. Solana, and N. Cabrera, Surf. Sci. 60, 385 (1976).

[10]N. Garcia, Phys. Rev. Lett. 37, 912 (1976); J. Chem. Phys. 67, 897 (1977).

[11]R. J. Masel, R. P. Merrill, and W. H. Miller, Phys. Rev. B 12, 5545 (1975).

[12]F. O. Goodman, J. Chem. Phys. 66, 976 (1977).

[13]F. Toigo, A. Marvin, V. Celli, and N. R. Hill, Phys. Rev. B 15, 5618 (1977).

[14]N. Garcia and N. Cabrera, in *Proceedings of the Seventh International Vacuum Congress and Third International Conference on Solid Surfaces*, edited by R. Dobrozemski *et al.* (Berger and Söhne), Vienna, 1977), Vol. I, p. 379; Phys. Rev. B 18, 576 (1978).

[15]C. Lopez, F. T. Yndurain, and N. Garcia, Phys. Rev. B 18, 970 (1978).

[16]F. O. Goodman, Surf. Sci. 70, 578 (1978).

[17]For a general review of atom-surface scattering, see, F. O. Goodman, Crit. Rev. Solid State Mater. Sci. 7, 3 (1977).

[18]D. R. Hartree, *Numerical Analysis* (Oxford U.P., London, 1958), Chap. VIII.

[19]A. M. Turing, Q. J. Mech. Appl. Math. 1, 287 (1948).

[20]S. D. Conte, *Elementary Numerical Analysis* (McGraw-Hill, New York, 1965), pp. 163–177.

[21]J. H. Wilkinson, *The Algebraic Eigenvalue Problem* (Oxford University, London, 1965), Chaps. 1, 2, and 4.

[22]N. R. Hill and V. Celli, Phys. Rev. B 17, 2478 (1978).

[23]N. R. Hill, Ph.D. dissertation (University of Virginia at Charlottesville, 1978) (unpublished). This work gives a description also of the extension of the Rayleigh approach by a variational calculation (see Ref. 5).

[24]N. Garcia and F. O. Goodman (unpublished).