PHYSICAL REVIEW E VOLUME 52, NUMBER 6 DECEMBER 1995

Acceptance-probability-controlled simulated annealing: A method for modeling the optical constants of solids

Aleksandar D. Rakić, Jovan M. Elazar, and Aleksandra B. Djurišić
Faculty of Electrical Engineering, University of Belgrade, Bulevar Revolucije 73, P.O. Box 816, 11000 Belgrade, Yugoslavia
(Received 14 April 1995)

A simulated annealing procedure with acceptance-probability control instead of the usual temperature control is proposed. The algorithm presented here has proved to be fully insensitive to initial parameters values, free of local-minima trapping problems, and shows superior convergence compared to adaptive-step classical simulated annealing with exponential cooling schedule. Experiments on computer generated synthetic data (with noise), closely resembling the optical constants of a metal, were performed to verify the effectiveness of the algorithm. The algorithm is then applied to parameter estimation of the model of optical constants of aluminum.

PACS number(s): 02.70.−c, 78.20.Ci, 78.66.−w

I. INTRODUCTION

Fitting the optical constants of solids to a specified model is a task that is frequently performed. A model parameter estimation is performed by minimizing the merit or cost function, which is usually a χ^2 function. When the cost function has a single mininum, the conventional downhill methods can locate it. However, when the cost function has multiple minima, a nonconvex optimization technique that allows hill climbing for escaping from local minima is required. When using conventional algorithms for such problems, it is necessary to provide good initial estimates of the parameter values to locate the valley containing the global minimum. It is sometimes very difficult to supply such initial guesses.

The simulated annealing [1] (SA), having its origins in the work of Metropolis *et al.* [2], is a powerful combinatorial optimization technique that overcomes the initial state estimation problem. Starting from an arbitrary initial state, the algorithm generates a sequence of changes of model parameters or "moves" in parameter state space. The success of this approach lies in the probabilistic hill climbing capability of a SA algorithm. The search process of the simulated annealing is usually controlled by an externally specified parameter, called temperature T, with the same units as the cost function. The plan for changing the temperature with time is termed the cooling schedule. If the initial temperature $T^{\rm init}$ is high enough, according to the cooling theorem of Geman and Geman [3], the global minimum is obtained asymptotically if the cooling is not faster than $T(t) = T^{\rm init}/\ln(1+t)$. The standard, or exponential, cooling schedule by Kirkpatrick, Gelatt, and Vecchi [1] is widely accepted [4–8]. A variation of the exponential schedule is the adaptive temperature scale-factor schedule given by Catthoor *et al.* [9,10]. Szu and Hartley [11,12] have proposed the annealing schedule $T(t) = T^{\rm init}/(1+t)$, called fast simulated annealing.

All of the above algorithms have two major problems. First, these schedules imply only the *reduction* of the temperature with time. In metallurgical annealing it is not unusual to inspect the structure of the system during the cooling process in order to detect the polycrystalline state and to increase the temperature to remelt it. This concept was applied by Matsuba [13] to simulated annealing. He varied temperature according to the decrement in cost function E. Thus, when the system has to climb hills, the schedule increases the temperature, allowing occasional hill climbing. The solution suggested in Ref. [13] requires substantial analytical effort even for the simplest one-dimensional problem. It is obvious that this method could be of no use for real-size problems.

The second problem is related to the traditional choice of the temperature as a control parameter. The temperature is varied in one way or another during the simulation in order to reduce the acceptance probability (AP) and confine the system in the vicinity of the already located global minimum. However, it is evident that the really important parameter is not the temperature itself but the *acceptance probability.* Therefore, we suggest that the AP be used as a control parameter instead of temperature.

In this paper we present an efficient simulated annealing algorithm. This algorithm uses acceptance-probability control for the cooling schedule rather than temperature. The main feature of our annealing schedule is that the AP is varied in a prescribed manner in time while the temperature is adaptively changed in accordance with the average change in cost function. This schedule also performs the occasional heating of the system, solving both problems observed with temperature-controlled algorithms simultaneously.

In the following section we describe the acceptance-probability controlled simulated annealing (APCSA) technique. Section III is devoted to the evaluation of the APCSA schedule performance and the comparison with the classical SA schedules. In Sec. IV the algorithm is used to estimate the parameters of the model of the optical constants of aluminum.

II. DESCRIPTION OF THE ALGORITHM

Simulated annealing is a procedure that iteratively changes a state of the optimization problem. Moves are chosen using a state generation procedure. The decision of whether or not to actually make a move is made by an acceptance criterion. Usually, the acceptance criterion is a function of the change in cost function (ΔE) and temperature T. The most frequently used AP is the Boltzmann distribution, where the probability of accepting the uphill move producing the change in cost function ΔE at temperature T is $\pi = \exp(-\Delta E/T)$ [2].

The algorithm presented in this paper (APCSA) has two nested loops. In the outer loop, the decrease of the AP is performed directly. The inner loop keeps the temperature constant while exploring the parameter space. A more detailed description of the algorithm is given in Secs. II A and II B.

A. The generation procedure and the acceptance criterion

The domain P containing the parameter vector $\boldsymbol{p} = (p(1), p(2), \ldots, p(N))$ is a subset of N-dimensional real space. It is determined by setting the lower and upper boundaries for each parameter $p_l(k)$ and $p_u(k)$. We discretize the solution state space by dividing the interval $p_u(k) - p_l(k)$ by N_Δ in order to determine the initial value of the vector $\boldsymbol{\Delta}$ (move-step vector) containing the initial step for changing the parameter vector $\boldsymbol{p}$. Thus $\boldsymbol{\Delta}$ is given by $\Delta(k) = [p_u(k) - p_l(k)]/N_\Delta$ for $k = 1, \ldots, N$.

Reducing the number of parameters to be altered in one iteration contributes to the stability of the solution and was employed by Hsu *et al.* [8] and Wu *et al.* [14]. We use the state generation procedure of Hsu *et al.* [8] with a randomly reduced number of parameters to be changed in one move. This procedure is of a local nature, allowing only the transitions from the current state $\boldsymbol{p}_i$ to the neighboring state $\boldsymbol{p}_j$,

$$p_j(k) := p_i(k) + r\,\Delta(k), \tag{1}$$

where r is an integer chosen randomly in the set $(-1,1)$.

The impact of the step size on the quality of the solution has been addressed before [9,11,12,15]. A comparatively large step is needed in the initial stage of the procedure to provide sufficient mobility of the algorithm to cover the entire parameter space. At the same time, it is important to perform some kind of move-range reduction with time in order to reduce the fluctuations during the final stages of annealing. Here we adopt the suggestion of Catthoor *et al.* [9] and reduce the step in "nearly inverse quadratic" manner. When the ratio $\Delta(k)/p(k)$ is less than 0.005 further reduction of the move step for that parameter is stopped.

The initial temperature is determined by employing the procedure of Rees and Ball [16]. Making many random changes to the cost function E, we find the average cost function $\langle E\rangle_\infty$ and the average of the absolute change $\langle|\Delta E|\rangle_\infty$. This would be equivalent to the average change of E at an infinite temperature when all moves would be accepted. Then the initial temperature T^{init} is assigned based on the acceptance probability $\pi^{\text{init}} = 0.90$, giving

$$T^{\text{init}} = -\frac{\langle|\Delta E|\rangle_\infty}{\ln(\pi^{\text{init}})} \approx 10\,\langle|\Delta E|\rangle_\infty. \tag{2}$$

The inner-loop stopping criterion is connected with the convergence of the entity D defined by [9]

$$D = \frac{1}{n}\sum_{i=1}^{n} \exp[(\langle E\rangle - E(\boldsymbol{p}_i^{\text{acc}}))/T_M], \tag{3}$$

where the summation is performed over the states $\boldsymbol{p}_i^{\text{acc}}$ accepted at temperature T_M. $\langle E\rangle$ is the average cost function at the preceding temperature T_{M-1} defined by

$$\langle E\rangle = \frac{1}{N_{\text{acc}}}\sum_{i=1}^{N_{\text{acc}}} E(\boldsymbol{p}_i^{\text{acc}}), \tag{4}$$

where N_{acc} is the number of moves accepted at T_{M-1}, while $E(\boldsymbol{p}_i^{\text{acc}})$ is the cost function, which corresponds to the accepted states $\boldsymbol{p}_i^{\text{acc}}$. Equilibrium is considered to be achieved when the absolute value of the relative difference between the current and the preceding value of D is less than a specified criterion δ, i.e., when $|D_n - D_{n-1}|/D_{n-1} < \delta$.

For outer loop termination we accept the slightly modified *solidification condition* of Doria *et al.* [7]. At each temperature we record the lowest value of the cost function obtained. When the absolute values of the relative difference between this value and the minimal values of the cost function at three preceding consecutive temperatures were within ϵ of each other, the simulation was stopped. The value of ϵ was initially set to 10^{-4}.

B. Role of the acceptance probability

After the inner loop termination the average of the absolute change in the cost function $\langle|\Delta E|_{\text{acc}}\rangle$ is evaluated. This value provides a valid estimation for the change in the cost function at the next temperature. At this point the classical SA algorithm reduces the temperature to some extent, checks the solidification criterion, and starts the new inner loop cycle.

Here our algorithm departs from the classical SA technique. The main idea of our approach is to control the AP directly instead of controlling it via the control parameter T. The initial temperature is determined at the beginning of the annealing procedure, employing Eq. (2). In this way the initial temperature is related to the average change in the cost function and the desired AP. What really matters is not the temperature itself, nor the average change in cost function, but the AP. It is also widely accepted that at the end of the annealing process the temperature should be "sufficiently low," meaning nothing else but that the AP should be low. At all times the AP was the entity of importance for the schedule. Therefore, it is not surprising that the entire cooling process could be efficiently controlled by properly controlling the AP instead of the temperature T.

What is the crucial difference between the classical SA approach and our approach? If the annealing process flows smoothly and if the cost function has no discontinuities or deep pits, it all works well even with the usual temperature control. But in regions where the cost function has valleys with steep walls, memory of the connection between the T and ΔE established correctly at the beginning could be easily lost. It is therefore essential to update the correlation between the T and the average change in the cost function adaptively during the annealing process in the same manner we did at the initial moment while determining the initial temperature, only with the lower acceptance ratio. This leads to the concept of the acceptance-probability controlled simulated annealing.

The acceptance probability is lowered in each outer loop cycle. One of the well known annealing schedules, such as, the exponential [1,6,5], linear [17], or inverse logarithmic schedule [3], can be used to lower the AP. Here we use the Gaussian function to generate the sequence of acceptance probabilities depending on the outer loop counter M:

$$\pi_M = \pi^{\text{init}} \exp(-M^2/2\sigma^2). \tag{5}$$

The temperature T_M is then determined by employing the average of the absolute change in the cost function at the preceding temperature $\langle|\Delta E|_{\text{acc}}\rangle$ and the desired acceptance probability π_M

$$T_M = -\frac{\langle|\Delta E|_{\text{acc}}\rangle}{\ln(\pi_M)}. \tag{6}$$

The monotonically decreasing function $[-1/\ln(\pi_M)]$ provides the needed average reduction of the temperature. It was already correctly noted by Catthoor *et al.* [9] that in the regions where the cost function had sharp edges or deep valleys it was suitable to minimize the reduction of the temperature in order to enable the algorithm to cope with discontinuities. Clearly, in such regions the $\langle|\Delta E|_{\text{acc}}\rangle$ values are significant. Here we proceed one step further and allow the algorithm not only to keep the temperature constant, but to occasionally increase the temperature in the situations when the AP is not sufficient to overcome the increase in $\langle|\Delta E|_{\text{acc}}\rangle$. The ability of the APCSA schedule to raise the temperature in the instants when the system needs remelting makes it very efficient in escaping from local minima. As it will be shown in the following section, the APCSA is more efficient in escaping local minima than the usual, temperature controlled, SA algorithms. In the following sections three numerical experiments will be performed on synthetic and real data sets to evaluate the performance of the APCSA technique.

III. TEST OF THE APCSA TECHNIQUE

To investigate the performance of the APCSA algorithm we performed two experiments on synthetic data, with and without noise. We shall briefly discuss the applied model, which was employed for parametrization of the optical constants of metals [18]. It was shown [19–22] that $\epsilon_r(\omega)$ could be expressed in a form that separates explicitly the intraband effects (usually referred to as free electron effects) from the interband effects (usually associated with bound electrons). In this paper the following model is used:

$$\hat{\epsilon_r}(\omega) = \hat{\epsilon}_r^{(f)}(\omega) + \hat{\epsilon}_r^{(b)}(\omega). \tag{7}$$

The intraband part $\hat{\epsilon}_r^{(f)}(\omega)$ of the dielectric constant is a well known free electron or Drude model [23]

$$\hat{\epsilon}_r^{(f)}(\omega) = 1 - \frac{\Omega_p^2}{\omega(\omega + i\Gamma_0)} \tag{8}$$

and the interband part of the dielectric constant $\hat{\epsilon}_r^{(b)}(\omega)$ is a simple semiquantum model resembling the Lorentz result for insulators

$$\hat{\epsilon}_r^{(b)}(\omega) = -\sum_{j=1}^{k} \frac{f_j \omega_p^2}{(\omega^2 - \omega_j^2) + i\omega\Gamma_j}, \tag{9}$$

where ω_p is the plasma frequency; k is the number of interband transitions with frequency ω_j, oscillator strength f_j, and lifetime $1/\Gamma_j$, $\Omega_p = \sqrt{f_0}\omega_p$ is the plasma frequency associated with intraband transitions; f_0 is the oscillator strength for electrons contributing in intraband processes; and Γ_0 is the intraband damping constant.

We use this model to evaluate the performance of the APCSA technique. Two sets of synthetic data resembling the optical constants of a metal were generated to test the algorithm.

In the first step, we generated the values of the dielectric constants in the range from 6.3 meV to 15 eV. The initial values for the parameter vector components and the target values used to generate the $\hat{\epsilon}_r(\omega)$ are presented in Table I. In common situations, at least some of the frequencies of the interband transitions are known from the

TABLE I. Values of the semiquantum model parameters: initial values, target values employed for generation of the synthetic data sets with and without noise, and values obtained by the APCSA algorithm by fitting the model to data with noise. Γ and ω are in eV and f is dimensionless. Column σ shows the standard uncertainties of the estimated parameters.

Parameter	Initial	Target	Obtained	σ
f_0	0.500	0.700	0.696	0.008
Γ_0	0.050	0.060	0.060	0.003
f_1	0.500	0.200	0.203	0.007
Γ_1	0.500	0.300	0.309	0.009
ω_1	0.500	0.400	0.398	0.006
f_2	0.500	0.300	0.296	0.005
Γ_2	0.500	0.300	0.295	0.003
ω_2	1.600	1.500	1.499	0.001
f_3	0.500	0.200	0.199	0.009
Γ_3	0.500	1.000	0.971	0.030
ω_3	2.200	2.000	1.983	0.017
f_4	0.500	0.050	0.052	0.004
Γ_4	0.500	3.000	2.916	0.090
ω_4	4.000	4.500	4.389	0.082

band structure calculations or can be anticipated from the visible structure in optical constants. Therefore, it is possible to make valid assumptions for the initial values of ω_j. No attempt was made to confine other parameters even by an order of magnitude. The following cost function was used:

$$E(\boldsymbol{p}) = \sum_{i=1}^{i=N} \left[\left| \frac{\epsilon_{r1}(\omega_i) - \epsilon_{r1}^{\text{expt}}(\omega_i)}{\epsilon_{r1}^{\text{expt}}(\omega_i)} \right| + \left| \frac{\epsilon_{r2}(\omega_i) - \epsilon_{r2}^{\text{expt}}(\omega_i)}{\epsilon_{r2}^{\text{expt}}(\omega_i)} \right| \right]^2 . \tag{10}$$

We tried to fit the synthetic data generated from the model, employing the fitting routine based on the modified Levenberg-Marquardt algorithm with and without the strict descent option. In both cases the algorithm produced a rather low value of the cost function and "acceptable looking" plots of the optical constants, but failed to obtain the target parameter values. This behavior indicates the termination in the local minimum differing slightly in cost from the global one, but in a different location in parameter state space. This is where the parameter estimation problem appears to be more demanding compared to other optimization problems such as the circuit placement one, where the quality of the solution is measured exclusively by the final cost function value. The algorithm for the parameter estimation is required to find the global minimum, i.e., to locate correctly all the model parameters, especially if they are expected to have some physical meaning.

In the next step we used the synthetic data for $\hat{\epsilon}_r(\omega)$ with Monte Carlo generated additive Gaussian noise with frequency-dependent variance, determined in such a manner to give uncertainties of 0.5% in the reflectance calculated from the $\hat{\epsilon}_r^{\text{noise}}(\omega)$ values. As expected, it was impossible to fit the data with noise to the model and obtain the target parameter values employing the Levenberg-Marquardt algorithm.

Both problems were successfully solved with the APCSA technique. Annealing was performed starting from the same initial configuration, far from the global minimum. We examined the impact of the inner-loop stopping criterion δ on the convergence rate. Figure 1 shows the number of iterations the APCSA algorithm needed to find the solution as a function of δ. For δ above the 0.05 the convergence fails; there is not enough time spent at each temperature for the system to reach the equilibrium. For δ below 0.01, the number of iterations needed for completing the simulation is rapidly increased. Similar behavior is observed for classical simulated annealing (CSA), but with significantly longer computer CPU time. Values of δ providing the rapid convergence of the APCSA schedule lie between 0.01 and 0.05, while the values of δ needed to ensure the convergence of the CSA algorithm are significantly lower (0.001–0.0025), consequently requiring a longer time to be spent at each temperature.

The annealing parameters for both algorithms were determined to provide the fastest convergence. Temperature vs number of iterations for both APCSA and CSA schedules for synthetic data with noise, using optimal δ for both algorithms, are presented in Fig. 2. It is evident that, in spite of the occasional increase of temperature, the APCSA algorithm provides faster cooling. Figure 3 shows the cost function for the same data. The CSA cost function curve shows a typical transition from the high cost plateau to the low cost region. The adaptive change of the temperature in the APCSA results in the absence of the plateau and apparently superior convergence. Figures 4 and 5 emphasize the time evolution of the free electron parameters of the semiquantum model. Figure 4 (APCSA) shows steady convergence of the parameters towards the target values. The classical algorithm (Fig. 5) obviously had problems with the local minimum while attempting to locate the Γ_0 at 0.5 eV instead of the required 0.060 eV. Figures 6 and 7 show the evolution of the values of the third oscillator param-

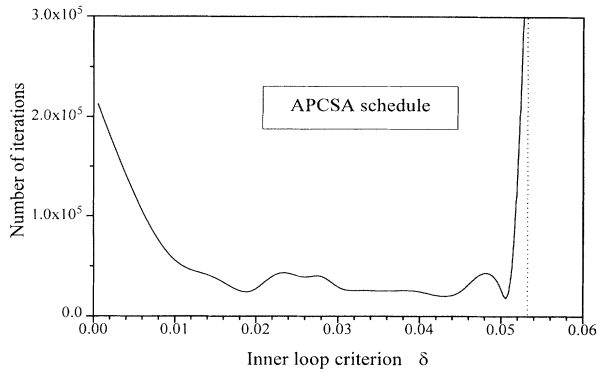

FIG. 1. Number of iterations the APCSA algorithm needs to find the solution as a function of the inner loop criterion δ (dimensionless). For δ above 0.05 the convergence fails.

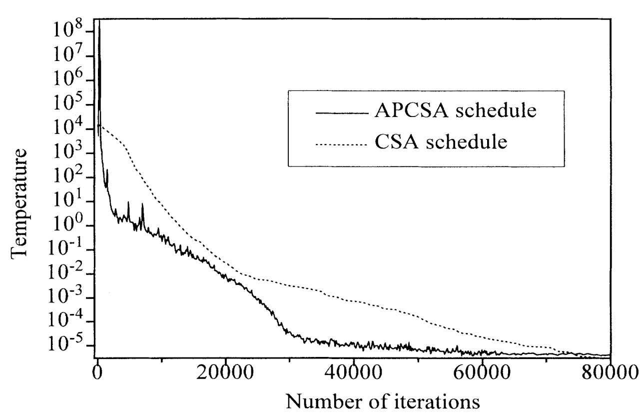

FIG. 2. Temperature (in same relative units as the cost function) vs number of iterations for APCSA (solid line) and CSA (broken line) algorithms. Results are for synthetic data with noise.

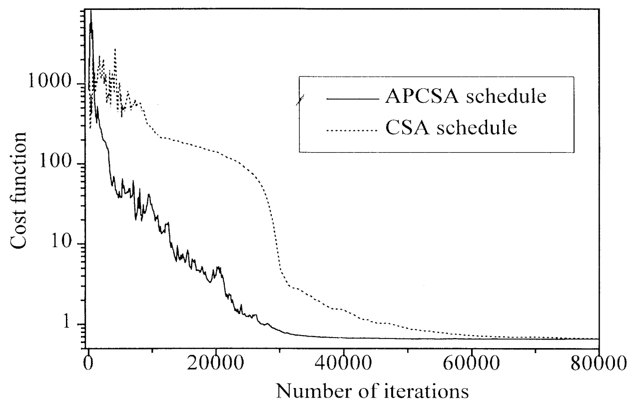

FIG. 3. Cost function (dimensionless) vs number of iterations for APCSA (solid line) and CSA (broken line) algorithms. Results are for synthetic data with noise.

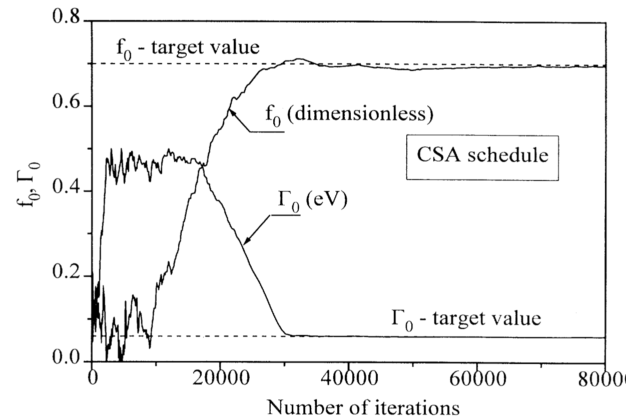

FIG. 5. Time evolution of the free electron parameters of the semiquantum model for the CSA algorithm.

eters f_3, Γ_3, and ω_3. The APCSA algorithm needed only 30 000 iterations to find the solution, while the CSA became stuck in a local minimum, with no chance of getting out, even after 80 000 iterations. The parameters obtained from the APCSA for the data set without noise have 3–5 correct digits, while the parameters obtained from data with noise are well within the uncertainty borders estimated from the covariance matrix and the initial estimate of the uncertainties in the reflectance spectrum.

IV. APPLICATION TO ALUMINUM

Finally, we applied the APCSA technique to the estimation of the parameters of the model of the optical constants of aluminum. There has been considerable theoretical and experimental interest in the optical properties of aluminum (for review see Ref. [18]).

The semiquantum model, as a simple phenomenological model, was employed for parametrization of optical constants of aluminum several times [18,24,25]. Here we intend to reperform this task, but in a different manner. Aluminum is chosen as a well-known material, ensuring that results could be anticipated in advance. Again we located intentionally the initial guess for the parameter vector in the region of the parameter space far from the expected solution. It was not possible to solve this problem by employing the Levenberg-Marquardt algorithm, while the APCSA technique finds the solution efficiently. In the final stage of the optimization procedure, when the location of the global minimum is approximately revealed, the process could be significantly accelerated by switching to the Levenberg-Marquardt algorithm to obtain more accurate parameter estimates. Final parameters values are presented in Table II.

V. CONCLUSION

Our principal aim was to determine whether the simulated annealing approach is appropriate for fitting the

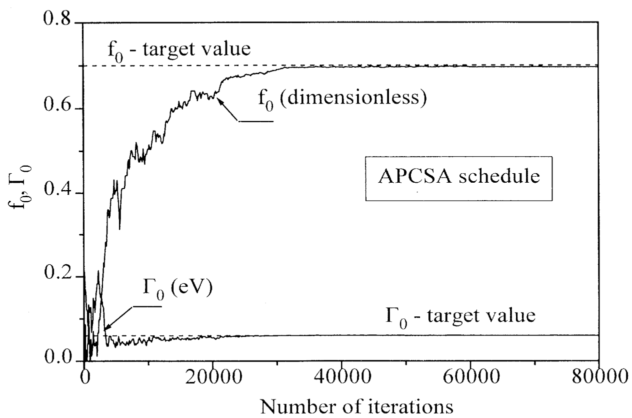

FIG. 4. Time evolution of the free electron parameters of the semiquantum model for the APCSA algorithm.

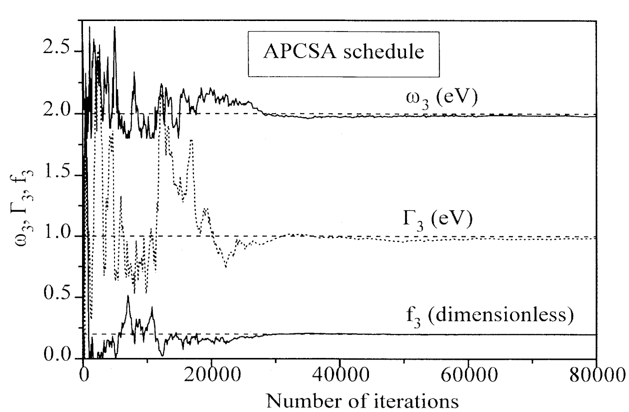

FIG. 6. Time evolution of the third oscillator parameters f_3, Γ_3, and ω_3 for the APCSA algorithm.

TABLE II. Semiquantum model (oscillator model) parameters values employed for the calculation of the optical constants of aluminum.

j	f_j	Γ_j (eV)	ω_j (eV)
0	0.610	0.057	0
1	0.142	0.436	0.306
2	0.105	0.512	1.581
3	0.130	1.917	2.092
4	0.009	4.701	5.336

optical constants of metals. We believe we have demonstrated that it is. Our second objective was to introduce the more efficient simulated annealing algorithm with acceptance-probability control instead of the usual temperature control. The algorithm presented has proved to be fully insensitive to initial parameters values, extremely efficient in escaping local minima, and shows faster convergence compared to adaptive-step classical simulated annealing with exponential cooling schedule.

Accordingly, we conclude that the acceptance-probability controlled simulated annealing algorithm is superior to classical simulated annealing in situations when the location of the global optimum cannot be anticipated even approximately. In such cases, when no valid initial guess for some of the parameters could be made, the APCSA algorithm showed its superiority exactly where it was expected — in the complete independence on the initial parameters values. This makes the APCSA a powerful tool not only for the parametrization of the optical constants of solids but also for a computer-aided design of the optical thin film coatings and filters.

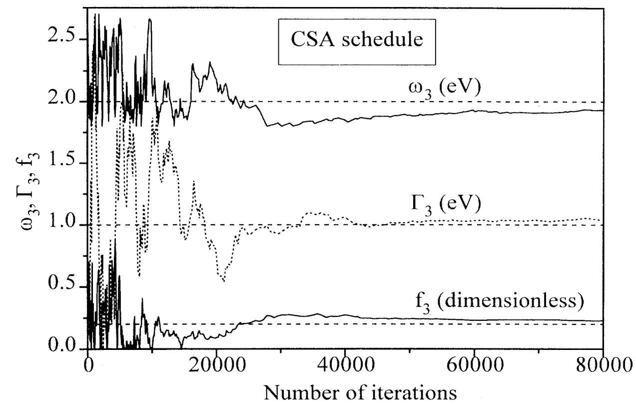

FIG. 7. Time evolution of the third oscillator parameters f_3, Γ_3, and ω_3 for the CSA algorithm.

ACKNOWLEDGMENT

The authors would like to acknowledge very useful discussions with Dr. S. Parameswaran of The University of Queensland, in Brisbane.

[1] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, Science **220**, 671 (1983).
[2] N. Metropolis *et al.*, J. Chem. Phys. **21**, 1087 (1953).
[3] S. Geman and D. Geman, IEEE Trans. Pattern Anal. Machine Intell. **PAMI-6**, 721 (1984).
[4] R. A. Rutenbar, IEEE Circuits Dev. Mag. **5**, 19 (1989).
[5] M. K. Vai, D. Ng, and S. Prasad, Electron. Lett. **26**, 892 (1990).
[6] I. Ziskind and M. Wax, IEEE Trans. Antennas Propagat. **38**, 1111 (1990).
[7] M. M. Doria, J. E. Gubernatis, and D. Rainer, Phys. Rev. B **41**, 6335 (1990).
[8] F. Hsu, P.-R. Chang, and K.-K. Chan, IEEE Trans. Antennas Propagat. **41**, 1195 (1993).
[9] F. Catthoor, H. de Man, and J. Vandewalle, INTEGRATION, **6**, 147 (1988).
[10] G. G. E. Gielen, H. C. C. Walscharts, and W. M. C. Sansen, IEEE J. Solid-State Circuits **25**, 707 (1990).
[11] H. H. Szu and R. L. Hartley, Phys. Lett. A **122**, 157 (1987).
[12] H. H. Szu and R. L. Hartley, Proc. IEEE **75**, 1538 (1987).
[13] I. Matsuba, Phys. Rev. A **39**, 2635 (1989).
[14] J. Wu, T. Kondo, and R. Ito, Jpn. J. Appl. Phys. **32**, 5576 (1993).
[15] C. P. Chang and T. H. Lee, Opt. Lett. **15**, 595 (1990).
[16] S. Rees and R. C. Ball, J. Phys. A **20**, 1239 (1987).
[17] K. H. Hoffmann and P. Salamon, J. Phys. A Math. Gen. **23**, 3511 (1990).
[18] A. D. Rakić, Appl. Opt. **34**, 4755 (1995).
[19] N. W. Ashcroft and K. Sturm, Phys. Rev. B **3**, 1898 (1971).
[20] H. Ehrenreich, H. R. Philipp, and B. Segall, Phys. Rev. **132**, 1918 (1963).
[21] H. Ehrenreich and H. R. Philipp, Phys. Rev. **128**, 1622 (1962).
[22] K. Sturm and N. W. Ashcroft, Phys. Rev. B **10**, 1343 (1974).
[23] M. I. Marković and A. D. Rakić, Appl. Opt. **29**, 3479 (1990).
[24] C. J. Powell, J. Opt. Soc. Am. **60**, 78 (1970).
[25] K. Baba *et al.*, Appl. Opt. **27**, 2554 (1988).