JOURNAL OF COMPUTATIONAL PHYSICS 64, 177–194 (1986)

Free-Flight Time Generation in the Monte Carlo Simulation of Carrier Transport in Semiconductors

RONALD M. YORSTON*

Computer Science Department, Reading University, Reading, England

Received March 18, 1985; revised June 26, 1985

The Monte Carlo simulation of carrier transport in semiconductors requires the generation of free-flight times corresponding to known probability densities. This paper reviews the methods which have been proposed to generate such random variables. It is shown that one of the fastest available algorithms, the iterative-gamma method, does not give the correct distribution. Two new methods are proposed: the constant-time method and polynomial integral evaluation. The former is similar to iterative gamma, but modified to produce more reliable results. It is also very fast. Polynomial integral evaluation can be used where the scattering rates are stored as polynomials rather than in the more usual tabular form. This may be useful when working in restricted memory.

1. INTRODUCTION

(a) *The Monte Carlo Method*

Electron transport phenomena in solids may be studied by a Monte Carlo procedure in which the path of a single particle is followed by computer simulation. The electron is considered to move quasi-classically during periods of free flight under the influence of an applied field. The free flights are interspersed with scattering events which represent the interaction of the particle with lattice vibrations and impurities. The free-flight length and outcome of a scattering event are selected at random from known probability distributions. Transport properties of the material under investigation are found from the time average of the behaviour of the individual particle.

The state of an electron is specified by its wavevector, $\mathbf{k}$, and by its real-space position vector, $\mathbf{r}$. The wavevector is related to the momentum, $\mathbf{p}$, and the energy, ε, by

$$\mathbf{p} = \hbar\mathbf{k} \tag{1.1}$$

$$\varepsilon(1 + \alpha\varepsilon) = \hbar^2 k^2/2m \tag{1.2}$$

* Current address: ISTEL Limited, Grosvenor House, Prospect Hill, Redditch, Worcs., England.

0021-9991/86 $3.00

where α is the non-parabolicity factor, a property of the material, $\hbar$ is Planck's constant divided by 2π, and m is the effective mass of an electron in the material. The motion of the particle in an electric field, $\mathbf{E}$, is taken to be Newtonian, so that the time variation of $\mathbf{k}$ is given by

$$\mathbf{k}(t) = \mathbf{k}_0 + q\mathbf{E}(t)t/\hbar \tag{1.3}$$

where q is the charge on the particle and $\mathbf{k}_0$ is the wavevector at time $t = 0$. The electron flies freely in the field until it interacts with the lattice. The probability density for the duration of a free flight is

$$P(t) = \lambda[\mathbf{k}(t)] \exp\left[-\int_0^t \lambda[\mathbf{k}(t')]\, dt' \right] \tag{1.4}$$

where $\lambda(\mathbf{k})$ is the total scattering rate for wavevector $\mathbf{k}$

$$\lambda(\mathbf{k}) = \sum_{i=1}^{N} \lambda_i(\mathbf{k}) \tag{1.5}$$

where $\lambda_i(\mathbf{k})$ is the rate due to the ith process and N is the number of processes. The first term in (1.4), $\lambda[\mathbf{k}(t)]$, is the probability that a scatter occurs at time t. The second term is the probability that the particle survives the interval $(0, t)$ without suffering a scatter.

When an interaction takes place one of the possible scattering mechanisms is chosen to occur. The choice is made such that the probability of each mechanism being picked is proportional to its scattering rate. The wavevector is then modified according to the physical nature of the process associated with the mechanism chosen. This gives the final state of the electron.

The sequence of free flight followed by scatter is repeated many times until sufficient information has been accumulated to allow the calculation of meaningful averages. It is then possible to determine the time-average values for transport properties such as the energy, drift velocity, and diffusion coefficients. Alternatively, the simulation of a number of electrons may be carried out simultaneously and the time dependence of their behaviour derived from an ensemble average.

Price [1] gives a review of the Monte Carlo technique as applied to transport in semiconductors. Boardman [2] gives a more practical description and includes a complete program.

(b) *Free-Flight Generation*

The problem to be investigated here is that of finding a value for the random free-flight time subject to the probability density of (1.4). The starting point for this calculation will be a pseudorandom number, r, chosen from a uniform distribution in the range (0, 1). Routines for generating such numbers with suitable statistical properties are readily available for most computer systems.

The probability density of flight times, $P(t)$, may be related to that of the uniformly distributed random number by

$$P(r)\,dr = P(t)\,dt. \tag{1.6}$$

Integration of this using $P(r) = 1$ gives

$$r = \int_0^t P(t')\,dt'. \tag{1.7}$$

Substituting for $P(t)$ from (1.4) and carrying out the integration produces

$$r = 1 - \exp\left[-\int_0^t \lambda[\mathbf{k}(t')]\,dt'\right] \tag{1.8}$$

and hence

$$-\ln r = \int_0^t \lambda[\mathbf{k}(t')]\,dt'. \tag{1.9}$$

Here $(1-r)$ has been replaced by r since they are random variables with the same probability distribution.

Equation (1.9) is of fundamental importance. It expresses the relationship between the known pseudorandom number, r, and the required flight time, t. For any particular theoretical model of a semiconductor the form of the total-scattering-rate function, $\lambda(\mathbf{k})$, is known. Thus, in principle, it should be possible to carry out the integration and invert the result to find t. However, the integral cannot be evaluated analytically for most types of scattering rates which are of interest. Even in those simple cases where the integral can be found, the inversion of the resulting expression is only possible numerically. Since a direct approach is either impossible or, at best, would be cumbersome and time-consuming, other more subtle methods must be used.

One of the most fruitful concepts which has been brought to bear on the problem is that of "self-scattering," first introduced by Rees [3]. The self-scattering mechanism is a virtual process which leaves the state of an electron unchanged. Indeed, it must have this behaviour because all of the physically significant "real-scattering" processes have already been included. Since the new process can have no effect on the electron distribution its associated scattering rate may be chosen to have an arbitrary value, say, $\lambda_0(\mathbf{k})$. The new total scattering rate is then

$$\Gamma(\mathbf{k}) = \lambda_0(\mathbf{k}) + \lambda(\mathbf{k}). \tag{1.10}$$

The introduction of self-scattering thus allows the insertion of an arbitrary function into the integral of (1.9) (subject to the constraint that, since it represents a scattering rate, $\lambda_0(\mathbf{k})$ is greater than or equal to zero). Clearly, such a function may be

chosen so that substituting the new total scattering rate, $\Gamma(\mathbf{k})$, into (1.9) results in an integral which is easy to evaluate.

It is illuminating to follow Fawcett et al. [4] and consider what effect the self-scattering has on the probability density of (1.4). The probability that an electron has a free flight of length t terminated by a real scatter is

$$P(t) = \lambda(t) \exp\left[-\int_0^t \Gamma(t')\, dt' \right] \tag{1.11}$$

where, for brevity, the fact that λ and Γ are related to t via $\mathbf{k}$ has been omitted. This expression assumes that no self-scatters took place in the interval $(0, t)$. If one self-scatter had occurred in that interval the corresponding probability would be

$$P(t) = \lambda(t) \int_0^t dt' \left[\exp\left[-\int_0^{t'} \Gamma(t'')\, dt'' \right] \cdot \lambda_0(t') \times \exp\left[-\int_{t'}^t \Gamma(t'')\, dt'' \right] \right]. \tag{1.12}$$

Here the term $\lambda_0(t')$ is the probability of self-scatter at time t'. The first exponential term is the probability of an uninterrupted flight from time 0 to t' and the second is the similar probability for the interval (t', t). These three terms are integrated over all values of t' to give the probability of a flight length exceeding t which includes only one self-scatter. Expressions similar to (1.12) may be found for flights which include two or more self-scatters before the real scatter at t.

When added these expressions give the total probability of a free flight terminated by a real scatter at t. Extracting the common factors gives

$$P(t) = \lambda(t) \exp\left[-\int_0^t \Gamma(t')\, dt' \right] \times \left[1 + \int_0^t \lambda_0(t')\, dt' + \int_0^t dt'\, \lambda_0(t') \int_0^{t'} dt''\, \lambda_0(t'') + \int_0^t dt' \int_0^{t'} dt''\, \lambda_0(t'') \int_0^{t''} dt^{(3)}\, \lambda_0(t^{(3)}) + \cdots \right]. \tag{1.13}$$

Using the fact that

$$\int_0^t dt' \int_0^{t'} dt'' \cdots \int_0^{t^{(n-1)}} dt^{(n)} f(t') f(t'') \cdots f(t^{(n)}) = \frac{1}{n!} \left[\int_0^t f(t')\, dt' \right]^n \tag{1.14}$$

the sum in the last factor can be seen to reduce to the series expansion of $\exp[\int_0^t \lambda_0(t')\,dt']$. Thus we find

$$P(t) = \lambda(t) \exp\left[-\int_0^t \lambda(t')\,dt' \right] \tag{1.15}$$

which is the original probability density of Eq. (1.4).

The above analysis confirms mathematically what was physically obvious: the introduction of a virtual self-scattering process does not alter the probability distribution of free-flight times. However, the proof also highlights an important fact about the nature of the function $\lambda_0(t)$. The step in which (1.14) was applied to (1.13) relies on the fact that $\lambda_0(t)$ has the same functional form regardless of the presence of one or more self-scatters. The self-scattering-rate function, $\lambda_0(t)$, must therefore remain unchanged by the occurrence of a self-scatter. On the other hand, once a real scatter has occured the particular form of $\lambda_0(t)$ has served its purpose and some other function can be used for subsequent flight choices.

To summarise, we may introduce a fictitious scattering process which changes the total scattering rate from $\lambda(t)$ to $\Gamma(t)$, where Γ has the following properties:

(i) $\Gamma(t) \geqslant \lambda(t)$,

(ii) $\Gamma(t)$ may take any arbitrary functional form (subject to (i)), and this form may be changed after each real scatter.

2. Representation of Scattering Rates

Before describing the methods used in the solution of Eq. (1.9), it is of value to consider the ways in which the total scattering rate, $\lambda(\mathbf{k})$, can be represented. Any procedure for finding free-flight times will have to, at some stage, evaluate $\lambda(\mathbf{k})$. This is true even if a method is used which conceals $\lambda(\mathbf{k})$ in the total rate including self-scattering, $\Gamma(t)$.

When the Monte Carlo problem is first formulated the individual scattering rates which sum to give $\lambda(\mathbf{k})$ are known explicitly from the theoretical models for the mechanisms to be included. These explicit formulae are rather involved. Many of them require the extraction of a square root and would take some time to calculate. Since the scattering rate must be calculated many times it is better to find a faster method.

For the scattering mechanisms considered in this work it is found that $\lambda(\mathbf{k})$ depends only on the magnitude of the wavevector. It is therefore possible to construct a table of scattering rate as a function of k. The explicit formulae then only have to be evaluated once, during the initialisation of the table. Whenever a scattering rate is required subsequently it can be obtained by looking it up in the table—a very fast process. A reasonable compromise between an accurate representation of

$\lambda(k)$ and the minimisation of storage is achieved with a table containing 500 entries. In addition to the total scattering rate, $\lambda(k)$, it is also necessary to store the scattering rates for each of the individual processes for use in determining which of them should be chosen to terminate a free flight. Since sophisticated models for semiconductors can include ten or more different scattering mechanisms, it can be seen that the storage requirement for the tables starts to become significant—especially if the simulation is to be carried out on a small computer with a limited amount of memory.

Another possibility is to represent $\lambda(k)$ by an approximating function, such as a polynomial, which is more easily evaluated than the original formulae. It is found in practice that one polynomial will not provide a good approximation over the entire range of wavevectors which is of physical interest. This is due to the presence of a number of scattering mechanisms with discontinuities in their derivatives which cannot be followed accurately by the continuous polynomials. This problem may be dealt with by splitting the range of k at each of these discontinuities and using a different polynomial for each sub-range.

Each polynomial has the form

$$\lambda(k) = \sum_{i=1}^{n} a_i k^{i-1} + a_0/k \tag{2.1}$$

where the a_i are coefficients derived at initialisation. The term in $1/k$ is included to represent ionised-impurity scattering, which has a scattering rate of the form

$$\lambda_{ii}(k) = a_0 \left[\frac{1}{k} + \frac{2\alpha\varepsilon}{k} \right]. \tag{2.2}$$

The first part of this expression appears explicitly in Eq. (2.1) while the second term can be included in the polynomial along with the other mechanisms. The separation of the term in $1/k$ increases the accuracy of the approximation by reducing the amount of variation which has to be represented by the polynomial, particularly at low values of k. An alternative way of achieving the same result is to increase the number of polynomials used so that the change in $\lambda(k)$ over each subrange is smaller.

A quintic polynomial was used in the investigations described here because theory indicated that this would produce an error of less than one part in 10^6. This is about the same magnitude as the error due to the truncation of a variable when it is stored as a 32-bit floating-point number in a computer. Although the theory cannot be applied to the case where the derivative is discontinuous, practical experience has shown that errors of 10^{-4}% are typical over much of the range, with no error exceeding 1%.

Each of the methods described has its advantages and disadvantages. Direct evaluation is slow, but only requires one or two words of memory per scattering

mechanism. Tabulation is very fast, but consumes 500 words of memory for each process. Between these two extremes is polynomial approximation, which is moderately fast and only needs about 50 words of memory per process. The choice of which method to use depends on the computer on which the simulation is to run. In some cases a hybrid procedure could be advantageous, in which table look-up is used for finding the total rate while the rates for the individual processes are found from the explicit formulae.

3. Free-Flight Generation Algorithms

There is one special case in which Eq. (1.9) can be solved exactly, no matter what the scattering rate function— that is the case where the electric field is zero. A zero field means that the wavevector remains constant, at $\mathbf{k}(0)$, and the scattering rate is fixed at $\lambda[\mathbf{k}(0)]$. The flight time can then be obtained immediately:

$$t = \frac{-\ln r}{\lambda[\mathbf{k}(0)]}. \tag{3.1}$$

This special case is important because some of the methods to be described below will not work for a zero electric field.

(a) *Constant Gamma*

The constant-gamma method is the simplest application of the concept of self-scattering. The self-scattering rate $\lambda_0(k)$ is chosen to give a new total rate $\Gamma(k)$ which has a constant value. When this is substituted in Eq. (1.9) the free-flight time is found to be

$$t = (-\ln r)/\Gamma. \tag{3.2}$$

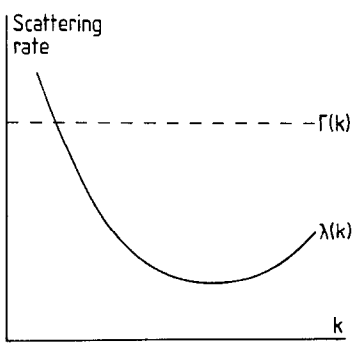

FIG. 1. Typical scattering rate curve, showing constant gamma value.

Thus, by introducing a fictitious process, the evaluation of a very complicated integral is reduced to a simple algebraic expression.

The difficulty with this procedure lies in the condition that $\Gamma \geqslant \lambda(k)$. If $\lambda(k)$ is fairly constant and bounded for all values of k there is no problem. However, Fig. 1 shows a typical scattering-rate curve for a model of silicon. Because of the $1/k$ dependence of ionised-impurity scattering it is impossible to find a value for Γ which will satisfy the condition for all k. Fortunately it is known that very large and very small values of k are unlikely to occur, so it is possible to set limits on the allowed range of k which contain all but a very few electrons. A value for Γ can then be found which satisfies the condition within the restricted interval. The chances of an electron leaving that interval and suffering an unphysical negative self-scattering rate are assumed to be sufficiently small that the effect of its behaviour on the final results can be neglected.

The disadvantage of this approach is that the required value for Γ is rather large. Over much of the possible range of wavevectors, and particularly the most occupied range, the self-scattering rate is very high compared with real scattering. Although self-scatters have no physical effect on the simulated electrons they do consume some computer time which can accumulate to become significant.

(b) *Piecewise-Constant Gamma*

As its name suggests, this method is a development of constant gamma (Borsari and Jacoboni [5]). Instead of using a single value of Γ for all wavevectors the function $\Gamma(k)$ is piecewise constant (Fig. 2). This allows a better approximation to the real scattering rate and a reduction in the number of self-scatters. The gain is slightly offset by the more complicated calculation needed to find the free-flight time.

The evaluation of Eq. (1.9) for a piecewise constant $\Gamma(k)$ must be carried out as a series of steps over each of which Γ is constant. Suppose that the first discontinuity in $\Gamma(t)$ occurs at time t_1, with $\Gamma(t) = \Gamma_1$ for $0 \leqslant t \leqslant t_1$. Integrating Γ_1 from 0 to t_1 gives the result $\Gamma_1 t_1$. Two cases now arise depending on the random value on the left-hand side of the equation.

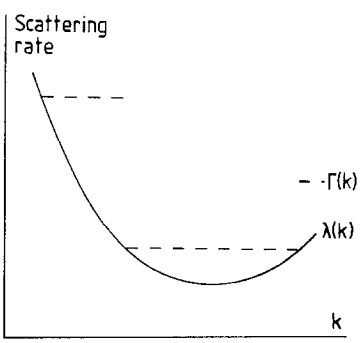

FIG. 2. Piecewise-constant gamma.

(i) $\Gamma_1 t_1 \geqslant -\ln r$. In this case putting $\lambda(k) = \Gamma(k)$ in Eq. (1.9) gives

$$-\ln r = \int_0^t \Gamma[\mathbf{k}(t')]\, dt' \tag{3.3}$$

which has a solution for some value of t less than or equal to t_1, given by

$$t = (-\ln r)/\Gamma_1. \tag{3.4}$$

(ii) $\Gamma_1 t_1 < -\ln r$. In this case the solution must be greater than t_1:

$$-\ln r = \int_0^{t_1} \Gamma_1\, dt' + \int_{t_1}^{t} \Gamma(t')\, dt' \tag{3.5}$$

or, putting $t'' = t' - t_1$,

$$(-\ln r - \Gamma_1 t_1) = \int_0^{(t-t_1)} \Gamma(t'')\, dt''. \tag{3.6}$$

This has the same form as Eq. (1.9) and may be solved for $(t - t_1)$ using the procedure described above, repeatedly if necessary. The evaluation will always terminate at some stage when the condition corresponding to case (i) is satisfied.

The above procedure requires that we know the times t_i at which Γ changes value. However, the discontinuities in Γ occur at fixed values of wavevector, not time. From (1.3) the variation of k with time is

$$k = \left[k_{0\perp}^2 + \left[k_{0\parallel} + \frac{qEt}{\hbar} \right]^2 \right]^{1/2} \tag{3.7}$$

where the subscripts $\parallel$ and $\perp$ denote, respectively, the components of wavevector parallel and perpendicular to the electric field, which has magnitude E. This function has a minimum of $k_{0\perp}$ at time

$$t_{\min} = -\hbar k_{0\parallel}/qE. \tag{3.8}$$

The time corresponding to k_i, the wavevector of the ith discontinuity in Γ during the course of the flight, is

$$t_i = \frac{\hbar}{qE} [\pm (k_i^2 - k_{0\perp}^2)^{1/2} - k_{0\parallel}]. \tag{3.9}$$

The square-root term is positive if k decreases from k_{i-1} to k_i and negative if k increases or passes through its minimum.

(c) *Iterative Gamma*

The iterative-gamma method was introduced to reduce still further the number of self-scatters (Warriner [6], Hockney and Eastwood [7]). This is achieved by

choosing a new value for Γ at each free flight. The value can be tailored to the real-scattering rate currently experienced by the particle.

The procedure is the following: for a particle with initial wavevector k_0 the first value of Γ is taken to be $\Gamma_1 = \lambda(k_0)$. Using the normal constant-gamma technique gives a corresponding flight time

$$t_1 = \frac{-\ln r}{\Gamma_1}. \tag{3.10}$$

The maximum real-scattering rate during a free flight of length t_1 is found, $\lambda_1^{\max}$. If $\Gamma_1 \geqslant \lambda_1^{\max}$ the time is accepted as being valid, but if the real-scattering rate exceeds Γ_1 the situation is unphysical. We therefore calculate a new value of constant Γ,

$$\Gamma = m\Gamma_1 \tag{3.11}$$

where m is a multiplying factor greater than unity. A typical value is $m = 1.1$. This new value of Γ is used to find a flight time, t_2. The maximum real-scattering rate over the new flight is again compared with Γ. The iterative process is repeated until a value of Γ is found which is greater than the maximum real-scattering rate. Since Γ is increased only in small steps the self-scattering rate can never form a large proportion of the total (unless $\lambda(t)$ is decreasing).

Although this procedure is superficially attractive a more detailed analysis reveals that under certain circumstances the flight times to not have the correct distribution. The graph of Fig. 3 shows the form of $\Gamma(t)$ for a particular real-scattering function. For certain time ranges Γ is a double-valued function, with the high or low branch being chosen depending on the value of the random number. To see how this comes about consider the value of $(-\ln r)$ which results in a free-flight time, t_1, just greater than t' (the time where $\lambda(t) = m\lambda(k_0) = m\lambda_0$):

$$t_1 = \frac{-\ln r}{m\lambda_0}. \tag{3.12}$$

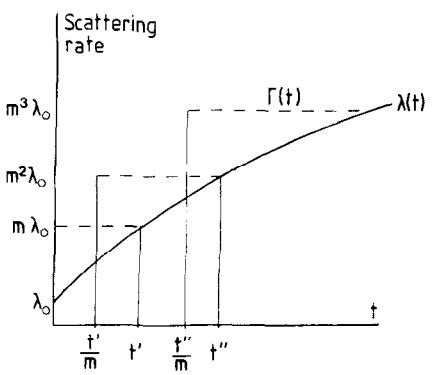

FIG. 3. Time variation of scattering rate and gamma, using the iterative-gamma method.

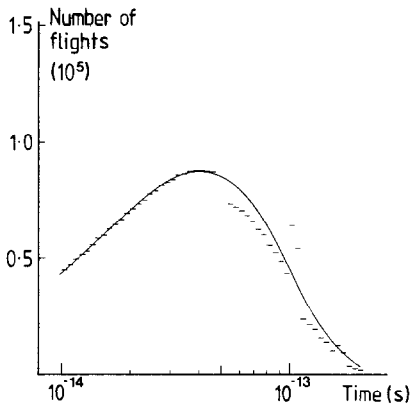

FIG. 4. Distribution of free-flight times for the scattering rate of (3.15). Continuous curve is the theoretical result; histogram depicts the Monte Carlo results using iterative gamma.

Since $\lambda(t_1) > m\lambda_0$ the iterative-gamma method requires that a new time be chosen using $\Gamma = m^2\lambda_0$:

$$t_2 = \frac{-\ln r}{m^2\lambda_0}. \tag{3.13}$$

However, this same value of time can also arise from a different random number choice, r', where $(-\ln r') = (-\ln r)/m$:

$$t = \frac{(-\ln r')}{m\lambda_0} = \frac{(-\ln r)}{m^2\lambda_0}. \tag{3.14}$$

The double-valued nature of $\Gamma(t)$ displayed here makes it difficult to apply the argument presented in Section 1(b) to show that the presence of self-scatters does not affect the distribution of free-flight times. An experimental approach has therefore been adopted.

The graph of Fig. 4 shows the result of an experiment in which iterative gamma was used to find the distribution of free-flight times for a particularly simple scattering rate:

$$\Gamma(t) = a + bt \tag{3.15}$$

where a and b are constants. The continuous curve shows the expected theoretical distribution while the histogram shows that produced by the iterative-gamma method. For short flights there is good agreement, but over the range t'/m to t', where Γ is double valued, there is a peak in the histogram as each time is generated by two possible random numbers. Similar effects are seen for other time values at which $\Gamma(t)$ is double valued. This experiment shows that the iterative-gamma method does not generate the correct distribution of free-flight times.

TABLE I

Percentage of Valid Flight Times Produced by the Iterative-Gamma Method for a Simple Model of Silicon

	Field (MV/m)				
Multiplier	0.5	1.0	2.0	4.0	6.0
1.05	98.2	95.7	93.8	91.9	90.0
1.1	99.6	98.7	96.9	95.0	94.9
1.2	99.8	99.7	98.4	96.5	97.9
1.3	99.8	99.8	98.8	96.9	97.2

Despite producing a free-flight distribution which is incorrect in places, iterative gamma is still a useful and fast method. It has been used for a number of years and has always produced results which agree well with the actual properties of semiconductors. As pointed out above, the flight distribution below t'/m is always correct. Table I shows the percentage of flights which satisfy this condition for a simple model of silicon. It is clear from this that only a small number of flights are chosen from the wrong distribution. Using larger values of the multiplier improves the situation by increasing the time t' where $\lambda(t') = m\lambda_0$. This does increase the number of self-scatters, though, and slows down the calculation.

(d) *Constant Time*

Like iterative gamma the constant-time method was designed to reduce the number of self-scatters by producing a function $\Gamma(t)$ which is a close fit to the local real-scattering rate. The procedure used is similar to that of piecewise-constant gamma, where the integration of the right-hand side of (3.3) is carried out in a series of steps. In this case, though, the times at which the discontinuities in $\Gamma(t)$ occur are always multiples of a fixed time increment, t_{inc}. Thus the first integration is from 0

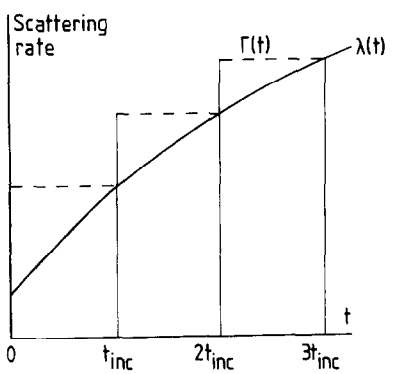

FIG. 5. Time variation of scattering rate and gamma, using the constant-time method.

to t_{inc}, the second from t_{inc} to $2t_{inc}$, etc. The constant Γ value used in each case is equal to the maximum real-scattering rate over the range of time values being considered. Figure 5 shows the form of $\Gamma(t)$ which results from this procedure.

When a free flight ends in a self-scatter special treatment is necessary. If a flight terminates with a self-scatter at time t, between $(n-1)\,t_{inc}$ and nt_{inc}, the first integration used to determine the next flight must extend from t to nt_{inc}. Also, the value of Γ used for this integration must be the same as that used for the previous one, i.e., the maximum value of the real-scattering rate over the time interval from $(n-1)\,t_{inc}$ to nt_{inc}. Subsequent integrations can proceed in steps of t_{inc} as before. These restrictions are necessary to ensure that condition (ii) of Section 1(b) is satisfied, i.e., that the functional form of $\Gamma(t)$ is not affected by the self-scatter. After a real scatter the form of $\Gamma(t)$ can change and the first integration can be over a period of length t_{inc}.

(e) *Numerical Integral Evaluation*

The methods considered so far have all avoided a direct evaluation of the integral in (1.9) by the introduction of self-scattering. Although it may not be possible to perform the required integration analytically it can certainly be done numerically. Replacing the integral with a numerical approximation using the trapezium rule gives

$$-\ln r = \sum_i \frac{\lambda(k_{i+1}) + \lambda(k_i)}{2}\,\Delta t_i. \tag{3.16}$$

The procedure works most effectively if $\lambda(k)$ is stored in the form of a table. The ordinates of the integral are then those values of time corresponding to the tabulated points. The interval between ordinates, Δt_i, is given by a formula similar to (1.9):

$$\Delta t_i = \frac{\hbar}{qE}\left[\pm(k_{i+1}^2 - k_{i\perp}^2)^{1/2} - k_{i\parallel}\right] \tag{3.17}$$

where, as before, the sign of the square root depends on the way in which k varies with time. The sum in Eq. (3.16) is evaluated term by term until it exceeds the random number on the left-hand side. The free-flight time is then the sum of the intervals Δt_i.

(f) *Polynomial Integral Evaluation*

Changing the variable of integration in (1.9) from t to k gives the following expression:

$$\frac{-eE}{\hbar}\ln r = \pm\int_{k_0}^{k} \frac{\lambda(k')\,k'}{(k'^2 - k_{0\perp}^2)^{1/2}}\,dk' = I(k_0, k) \tag{3.18}$$

TABLE II

Coefficients of the Integrated Polynomial (3.19) Related to the Coefficients of the Original Polynomial (2.1) and the Component of the Wavevector Perpendicular to the Electric Field, $k_{0\perp}$

$$c_1 = a_1 + 2a_3k_{0\perp}^2/3 + 8a_5k_{0\perp}^4/15$$
$$c_2 = a_2/2 + 3a_4k_{0\perp}^2/8 + 5a_6k_{0\perp}^4/16$$
$$c_3 = a_3/3 + 4a_5k_{0\perp}^2/15$$
$$c_4 = a_4/4 + 5a_6k_{0\perp}^2/24$$
$$c_5 = a_5/5$$
$$c_6 = a_6/6$$
$$c_0 = a_0 + a_2k_{0\perp}^2/2 + 3a_4k_{0\perp}^4/8 + 5a_6k_{0\perp}^6/16$$

where e is the magnitude of the electronic charge. The sign of the integral is the same as that of dk/dt. This integral cannot be performed if the original formulae for $\lambda(k')$ are substituted. If the polynomial approximation to $\lambda(k')$ given in (2.1) is used the integration can be carried out, with the result

$$I(k_0, k) = \pm\left[(k^2 - k_{0\perp}^2)^{1/2} \sum_{i=1}^{n} c_i k^{i-1} + c_0 \cosh^{-1}\frac{k}{k_{0\perp}} - (k_0^2 - k_{0\perp}^2)^{1/2} \sum_{i=1}^{n} c_i k_0^{i-1} + c_0 \cosh^{-1}\frac{k_0}{k_{0\perp}}\right]. \tag{3.19}$$

The polynomial coefficients of the integral, c_i, depend only on the coefficients a_i and the component of the wavevector perpendicular to the electric field, $k_{0\perp}$. Expressions for them for a quintic polynomial are given in Table II.

As was the case with piecewise-constant gamma the integrations must be carried

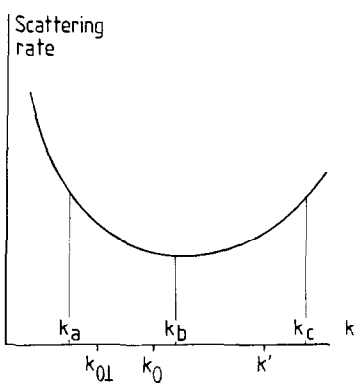

Fig. 6. Example to illustrate polynomial integral evaluation. See text for details.

out in stages. Here the limits of each step are defined by the range of the polynomial which is to be integrated. The calculation must also take into account the possibility of k decreasing to a minimum of $k_{0\perp}$ before increasing again. Consider, for example, the case illustrated in Fig. 6. The scattering rate has been represented by two polynomials—on the intervals (k_a, k_b) and (k_b, k_c). It is desired to find the integral from the initial wavevector k_0 to k'. If k decreases from k_0 the required calculation is

$$I(k_0, k') = I(k_0, k_{0\perp}) + I(k_{0\perp}, k_b) + I(k_b, k') \tag{3.20}$$

where each term is obtained from (3.19), with the first taking a negative sign because k is decreasing. The others take the positive sign.

Suppose that it is required to solve (3.18) for the situation just described. The first integral term of (3.20) is evaluated. Two possible cases may then arise depending on the random value of the left-hand side of the equation.

(i) $I(k_0, k_{0\perp}) \geqslant (-eE \ln r)/\hbar$. In this case (3.18) has a solution for some value of k between k_0 and $k_{0\perp}$. It is not possible to extract the required k from (3.19) analytically so a Newton–Raphson iterative procedure is used to find the solution numerically. Having obtained the value of k the flight time is calculated using an expression of the form (3.9)

(ii) $I(k_0, k_{0\perp}) < (-eE \ln r)/\hbar$. In this case the solution must occur at some time greater than that required to reach $k_{0\perp}$. Let the wavevector at that time be k.

$$\frac{-eE}{\hbar} \ln r = I(k_0, k_{0\perp}) + I(k_{0\perp}, k) \tag{3.21}$$

or

$$\left(\frac{-eE}{\hbar} \ln r - I(k_0, k_{0\perp})\right) = I(k_{0\perp}, k). \tag{3.22}$$

This expression has the same form as the original equation and can be solved using the same procedure with the appropriate substitution for the left-hand side. Repeated application of the same method will always result in case (i) being reached eventually, whereupon the solution is obtained.

4. Discussion

In order to test the flight-time generation methods described above a simple theoretical model for the semiconductor silicon was constructed. It is based on the parameters of Rode [8]. Despite the simplifications of this model the major features of the real-scattering-rate function, $\lambda(k)$, are well represented. Since the flight-time choice depends directly on this rather than the rates for the individual mechanisms

TABLE III

Execution Time in Microseconds for Free-Flight Generation Using Polynomial Approximation on the FPS AP-120B

Method	Field (MV/m)				
	0.5	1.0	2.0	4.0	6.0
Constant gamma	1737	2448	3220	3037	3230
Piecewise-constant gamma	159	150*	140*	139*	143*
Iterative gamma	154*	158	162	165	167
Constant time	167	171	175	178	180
Polynomial integral evaluation	311	315	318	317	316

Note. * denotes the fastest method in each case.

it is expected that the measurements made here will provide a good indication of the timings which would be obtained with more realistic models. The same conclusions should also apply for other materials with similar scattering-rate curves.

The methods of Section 3 have all been coded in FORTRAN. Each procedure has been included in a subroutine which accepts the initial state of the electron and returns a free-flight time chosen with the required probability distribution. All these flights terminate with a real scatter: self-scattering is hidden within the routine. The subroutines have been written so that they all interface with the main program in the same way and are freely interchangeable. The only differences are in the flight-time generation methods and the way in which the scattering rates are represented.

Table III shows the timings obtained for each method as the applied field was varied from 0.5 to 6.0 MV/m. In all cases the total scattering rate was represented by polynomial approximation. The timings were performed on an FPS AP-120B attached processor.

The times for *constant gamma* are considerably greater than those of any other method. This is due to the enormous number of self-scatters which result from the use of a constant Γ that is large enough to cover a wide range of wavevectors. Clearly, for materials with a wide variation in scattering rate over the physically significant range of k the constant-gamma method is quite unsuitable.

A considerable improvement in execution time is obtained by the use of a *piecewise-constant gamma.* This method is well suited to the polynomial representation of scattering rates: each polynomial may have its own associated constant Γ over its range of validity. Eight different constant values were used in this case. Further experiments have shown that even faster times may be achieved by using twelve polynomials. An increase from twelve to fourteen did not produce any marked change. Such behaviour is expected because, although increasing the number of sub-ranges will produce a better approximation to $\lambda(k)$ and reduce the number of self-scatters, there is an overhead involved every time a particle crosses from one sub-range to another so there eventually comes a time when the reduction in

TABLE IV

Execution Time in Microseconds for Free-Flight Generation Using Table Look-up on the FPS AP-120B

Method	Field (MV/m)				
	0.5	1.0	2.0	4.0	6.0
Piecewise-constant gamma	141	136	130	128	132
Iterative gamma	122	119	117	115*	115
Constant time	113*	115*	116*	116	114*
Numerical integral evaluation	116	137	189	297	—

Note. * denotes the fastest method in each case.

the number of self-scatters does not compensate for the increased overhead due to the new polynomial introduced to obtain it.

The *iterative-gamma* method was implemented here with a multiplier of 1.1. The execution times measured show that it has succeeded in its aim of reducing the number of self-scatters.

The same is true of the *constant-time* method. The time increment, t_{inc}, used in this procedure may be given any convenient value. Experiments have indicated that a value of 0.1 picoseconds results in the minimum execution time for the particular scattering rate used here. This is the case regardless of the electric field. It is possible that other scattering rates would require different time increments.

The final method, *polynomial integral evaluation*, is not particularly fast, despite parts of the routine having been coded in assembly language. The reason for this is the use of an iterative procedure to obtain the final value of wavevector at the end of a free flight. Each iteration requires the evaluation of a square root and an inverse hyperbolic cosine—both time-consuming operations.

Table IV gives the timings for four different methods using a table of scattering rates. Constant gamma was not included as it was obvious that the method would remain uncompetitive despite the faster calculation of scattering rates obtained by tabulation.

The three methods which appear in both tables, piecewise-constant gamma, iterative gamma, and constant time, all show an improvement in execution time when polynomial approximation is replaced by table look-up. This improvement is greater in the last two than in the case of piecewise-constant gamma. The explanation for this may be found in the number of times the scattering rate must be evaluated for each free flight: piecewise-constant gamma needs only one evaluation while the other methods require at least three. The latter therefore have most to gain from the change in scattering-rate representation. It may also be noted that constant time is now slightly faster than iterative gamma.

Numerical integral evaluation gives quite respectable times at low fields but becomes very slow at higher fields. This is due to the larger changes in wavevector

which occur over a free flight at high fields. These result in an increased number of terms in the numerical integration of (3.16). Since each of these terms requires the extraction of a square root the time taken to evaluate the integral increases substantially with the field.

To put the speed of calculation of the AP-120B in perspective the flight time choice program was run on two other computers. Using the constant-time method with table look-up for a field of 1 MV/m gave an execution time of 115 μsec on the AP-120B. The time for the corresponding problem on an IBM 360/195 mainframe was 54 μsec, while the Prime 750 minicomputer gave a time of 524 μsec.

5. Conclusion

Six different methods for the generation of flight times for use in the Monte Carlo simulation of semiconductors have been studied in this paper. Two of these methods, constant time and polynomial integral evaluation, are original and have arisen from a detailed consideration of the problem. It has also been shown that one of the existing methods, iterative gamma, is unreliable. The practicality of the various procedures has been demonstrated by the implementation of FORTRAN code to carry them out.

When choosing which of the methods to use it is necessary to consider the computer on which the simulation is to be run. If a large amount of memory is available the constant-time method may be used with tabulation of scattering rates. This results in the minimum execution time. A polynomial representation of the scattering rates can be used if the storage available is restricted. In this case the fastest method is piecewise-constant gamma.

References

1. P. J. Price, *Semiconductors and Semimetals* **14**, 249 (1978).
2. A. D. Boardman, "Computer Simulation of Hot Electron Behaviour in Semiconductors Using Monte Carlo Methods," *Physics Programs* (Wiley, Chichester, 1980).
3. H. D. Rees, *Phys. Lett. A* **26**, 416 (1968).
4. W. Fawcett, A. D. Boardman, and S. Swain, *J. Phys. Chem. Solids* **31**, 1963 (1970).
5. V. Borsari and C. Jacoboni, *Phys. Status Solidi B* **54**, 649 (1972).
6. R. A. Warriner, Ph.D. Thesis, Reading University, Reading, U.K., 1976 (unpublished).
7. R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles* (McGraw-Hill, Maidenhead, 1981), Chap. 10.
8. D. L. Rode, *Phys. Status Solidi B* **53**, 245 (1972).