

RADIATIVE HEAT TRANSFER WITH QUASI-MONTE CARLO METHODS*

A. Kersch¹ W. Morokoff²
A. Schuster¹

¹Siemens AG, Corporate Research and Development
ZFE BT ACM 31, 8000 Munich 83, Germany

²Institute for Mathematics and its Applications
University of Minnesota
Minneapolis, MN 55455

Abstract

Monte Carlo simulation is often used to solve radiative transfer problems where complex physical phenomena and geometries must be handled. Slow convergence is a well known disadvantage of such methods. In this paper we demonstrate that a significant improvement in computation time can be achieved by using Quasi-Monte Carlo methods to simulate Rapid Thermal Processing, which is an important technique for the production of semiconductor wafers, as well as many other industrial processes. Several factors are considered including surface absorptivity, position of wafer surface to heat source, and choice of quasi-random sequence. A comparison of the fractional and discrete absorption methods is also made. Results show accelerated convergence and improved accuracy of QMC over the standard Monte Carlo approach, and indicate when the fractional and discrete absorption methods should be used to obtain optimal results.

*This research was supported in part by the Institute for Mathematics and its Applications with funds provided by the National Science Foundation.

1 Introduction

The Quasi-Monte Carlo method, described in detail in section 1.3, involves using deterministic, quasi-random sequences in place of random numbers in a Monte Carlo calculation. This method has been applied to the radiative transfer problem of radiation passing through an absorbing slab in [8] and [9]. In this paper we consider a related problem where the radiation is scattered and absorbed only on a surface, the interior of a reactor. In section 1.1 such a reactor is briefly described. Then a mathematical formulation of the problem as an integral equation with a series solution is given. The second half of the paper is devoted to the results of numerical experiments on this problem. First the experimental technique is described, then the results are analyzed. Finally we summarize the results and present conclusions on the optimal application of Quasi-Monte Carlo to this problem.

1.1 Radiative Heat Transfer Reactors

In the manufacturing of semiconductor products, Rapid Thermal Processing is of increasing importance. A typical application is to use thermal radiation to heat a silicon wafer placed inside a reactor. The control of the temperature distribution of the wafer in such processing equipment is of prime importance. Since the in situ control of the temperature is a partially unsolved problem, the simulation of the physical mechanisms determining the temperature cycles can provide important information for the optimization of such processes. One of the problems which can be solved by such a simulation is high accuracy modeling of the radiative heat transfer from the heater to the wafer. Figure 1 shows the draft of the cylinder symmetric projection of a typical single wafer reactor.

This reactor was the basis for the calculations described in section 2. Because of the cylinder symmetry, the surface elements shown in the figure actually represent annuli. The heater, labeled *surface 0*, had an inner radius of 5.5 cm and an outer radius of 6.0 cm. When a point source heating element was used, it was placed at a radius of 5.75 cm. The annulus of the wafer labeled *surface 1* was placed directly above the heater and had an area of 18.06 cm². *Surface 2*, on the back side of the wafer, had an inner radius of 2.5 cm and an outer radius of 3.0 cm, corresponding to an area of 8.64 cm².

The radiative heat exchange in such a reactor is a function of the geometry of the problem, the spectral absorptivity, reflectivity and transmissivity of the surfaces, the reflection law of the radiation from the surfaces, and finally on the temperature distribution on all surfaces. A mathematical formulation of this problem is given next, including an indication of when Monte Carlo methods are the appropriate means of solution.

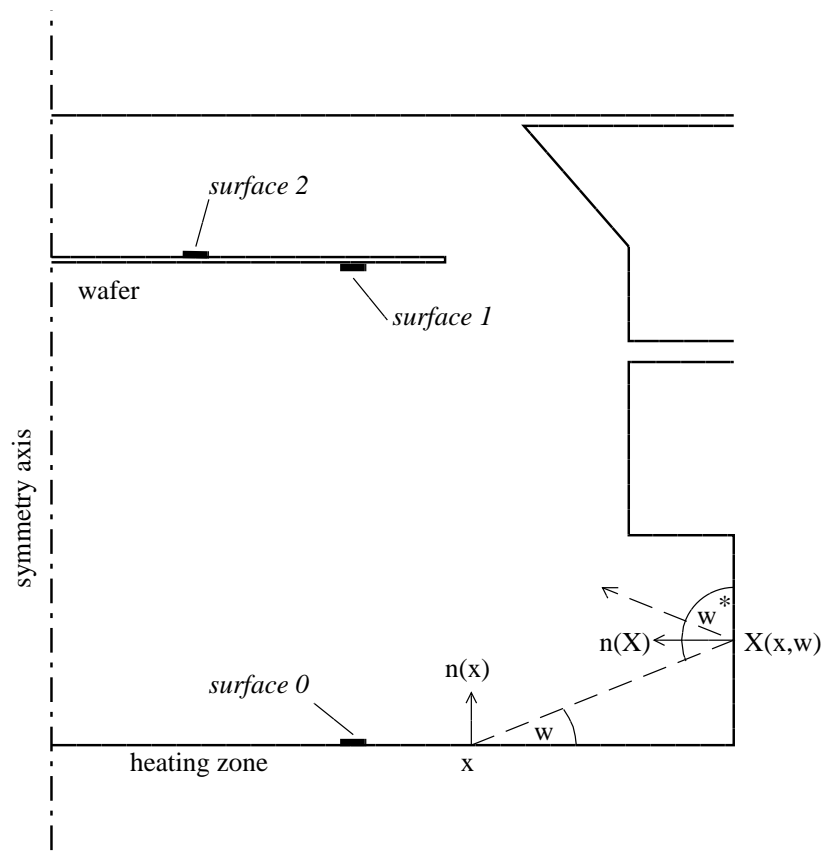


Figure 1: Geometry of the reactor and position of specific surface elements.

1.2 Mathematical Description of Problem

The radiative energy transfer inside the reactor can be described mathematically by an integral equation for the intensity of the radiation. Let D be the interior surface of the reactor, x be a point on D , and $R(x)$ be the reflectivity of the surface. Assume that there is an interior normal $n(x)$ at every point on D . Define S^+ to be the unit upper half sphere. Further define the function

$$X(x, \omega) : D \times S^+ \mapsto D$$

to be the point of intersection on D of a ray leaving the point x with direction ω relative to $n(x)$, and the function

$$\omega^*(x, \omega) : D \times S^+ \mapsto S^+$$

to be the direction, relative to $n(X)$, of the ray connecting points $X(x, \omega)$ and x . These quantities are illustrated in Figure 1. Also define the function $P_x(\omega, \omega^*)$ to be the probability that a ray arriving at point x with angle ω^* scatters with angle ω .

Let $Q(x, \omega)$ be the intensity of the source radiation emitted at point x in direction ω and $I(x, \omega)$ be the total intensity of radiation emitted (a combination of reflected and source radiation). Then I satisfies the equation

$$I(x, \omega) = Q(x, \omega) + R(x) \int_D \int_{S^+} P_x(\omega, \omega^*(y, \omega')) \delta(x - X(y, \omega')) I(y, \omega') d\omega' dy .$$

$\delta(x)$ is the standard Kronecker delta function. It is used here to select the angle which connects the points y and x .

The solution to this integral equation can be obtained by iteration. If the series of functions I_k are defined by

$$\begin{aligned} I_0(x, \omega) &= Q(x, \omega) \\ I_{k+1}(x, \omega) &= R(x) \int_D \int_{S^+} P_x(\omega, \omega^*(y, \omega')) \delta(x - X(y, \omega')) I_k(y, \omega') d\omega' dy \end{aligned}$$

then it can be shown that

$$I(x, \omega) = \sum_{k=0}^{\infty} I_k(x, \omega)$$

is formally a solution of the equation. By repeated substitution, I_k can be expressed as

$$\begin{aligned} I_{k+1}(x, \omega) &= \int_{[D]^{k+1}} \int_{[S^+]^{k+1}} R(x) R(y_k) \cdots R(y_1) \\ &\quad \cdots P_x(\omega, \omega_k^*) \cdots P_{y_1}(\omega_1, \omega_0^*) \\ &\quad \cdots \delta(x - X_k) \cdots \delta(y_1 - X_0) Q(y_0, \omega_0) (d\omega)^{k+1} (dy)^{k+1} \end{aligned}$$

Here $\omega_k^* = \omega^*(y_k, \omega_k)$ and $X_k = X(y_k, \omega_k)$. I_k can be interpreted as the contribution to I from rays which have undergone k reflections.

Generally, the quantity of interest is $E^A(x)$, the total radiative energy absorbed at a point on the surface. Up to this point we have been working with the emitted radiation. The two quantities are related as follows. It is assumed that all radiation arriving at a point is either absorbed with absorptivity $A(x)$ or reflected with reflectivity $R(x)$. This requires that $A(x) = 1 - R(x)$. The total energy reflected, E^R , can be expressed as

$$E^R(x) = \int_{S^+} (I(x, \omega) - Q(x, \omega)) d\omega .$$

This can also be expressed as

$$E^R(x) = R(x) E^T(x)$$

where E^T is the total energy arriving at x . Similarly the total absorbed energy can be written

$$E^A(x) = A(x) E^T(x) .$$

Then by solving and substituting in for E^T , we have that

$$E^A = \frac{A(x)}{R(x)} \int_{S^+} (I(x, \omega) - Q(x, \omega)) d\omega .$$

From the series representation of I it follows that

$$E^A(x) = \sum_{k=1}^{\infty} E_k(x)$$

where

$$\begin{aligned} E_k(x) &= \int_{[D]^k} \int_{[S^+]^{k+1}} A(x) R(y_{k-1}) \cdots R(y_1) \\ &\cdots P_x(\omega, \omega_{k-1}^*) \cdots P_{y_1}(\omega_1, \omega_0^*) \\ &\cdots \delta(x - X_{k-1}) \cdots \delta(y_1 - X_0) Q(y_0, \omega_0) (d\omega)^{k+1} (dy)^k \end{aligned}$$

There are two special cases of surface reflection which are of particular interest, diffuse and specular. In the diffuse case, the probability of scattering into a given angle is independent of the incoming angle. The function P is $P_x(\omega, \omega^*) = n(x) \cdot \omega$. The delta functions select out those k -step paths which can connect the starting point y_0 and the end point x . If D does not bound a convex region, not every set of k points on D can be connected by a path that lies inside D . This is related to the visibility of various parts of the reactor, a quantity which depends only on geometry and which can be calculated separately from the energy transfer computation. This is the basis for the view factor method, an effective, non-Monte Carlo method for computing radiative transfer in the diffuse case.

The second case, specular reflection, is described by setting P equal to a delta function

$$P_x(\omega, \omega^*) = \delta[\omega - \omega^* + 2(n(x) \cdot \omega^*) n(x)] .$$

Under the assumption of specular reflection, the above formula for E_k can be considerably simplified by noting that the ray paths are completely determined by the starting point y_0, ω_0 . If the function $\hat{X}_k(y_0, \omega_0)$ is defined as the position of a ray starting at (y_0, ω_0) at the k^{th} reflection, then E_k can be written

$$E_k(x) = \int_D \int_{S^+} A(x) \hat{R}_{k-1}(y_0, \omega_0) \delta(x - \hat{X}_k(y_0, \omega_0)) Q(y_0, \omega_0) d\omega_0 dy_0 .$$

Here \hat{R}_{k-1} is the product of the $k - 1$ reflectivities encountered along the path of the ray. Although this is now a low dimensional integral, because of the discontinuity of the integrand, only in the simplest case when the integral reduces to one dimension will a grid based quadrature method outperform the standard Monte Carlo approach. Thus Monte Carlo is generally used to compute the specular reflection case, which we study in this paper.

At this point it should be mentioned that the radiation intensities Q, I and E , as well as the reflectivity R and absorptivity A are functions of the frequency ν of the radiation. Typically the direction and frequency dependencies will be separated by writing

$$Q(x, \omega, \nu) = \theta(x, \omega) Q^0(x, \nu)$$

where θ is a directional distribution (e.g. Lambert's Law $\theta(x, \omega) = n(x) \cdot \omega$), and Q^0 is the frequency distribution. Often Q^0 is taken to be Plank's black body distribution at temperature $T(x)$

$$Q^0(x, \nu) = \frac{c_1 \nu^3}{e^{c_2 \nu / T(x)} - 1} .$$

Here c_1 and c_2 are constants. The inclusion of this frequency dependence simply involves adding an additional integration to the formula for E_k over the frequency range $[0, \infty]$.

The problem described above can be considered a fractional absorptivity problem. That is, every ray loses a certain percent of its energy (the absorptivity) at each reflection. This requires that the ray be reflected an infinite number of times. Of course in a computation involving ray tracing, it is necessary to terminate the path of a ray once it has lost a given percent of its energy. This corresponds to use only a finite number of terms in the above infinite series. When this method is used, it is only necessary to use Monte Carlo to choose the initial position, energy and direction of the ray. The path of the ray is fixed according to the laws of specular reflection, and at each reflection, the amount of energy absorbed is simply the absorptivity times the amount of energy remaining. One potential disadvantage to this method is that each ray must be traced through a full set of reflections, which may prove computationally expensive.

There is another formulation of this problem of radiative heat transfer in which many rays undergo fewer reflections than in the fractional absorption method, so that the average computation time per ray is smaller. In this alternative approach, the entire energy of the ray is absorbed with probability equal to the absorptivity and the ray tracing is terminated. We call this method the discrete absorptivity problem. A mathematical formulation can be obtained by employing the identity

$$R = \int_0^1 \Theta(R - z) dz$$

where $\Theta(z)$ is the Heavyside function:

$$\Theta(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0 \end{cases} .$$

This identity can be used in the above formula for E_k to obtain

$$E_k(x) = \int_0^\infty \int_D \int_{S^+} \int_{I^k} \Theta(A(x, \nu) - z_k) \prod_{i=1}^{k-1} \Theta(R^i - z_i) \delta(x - \hat{X}_k(y_0, \omega_0)) Q(y_0, \omega_0, \nu) (dz)^k d\omega_0 dy_0 d\nu$$

Here R^i is the reflectivity encountered in the i^{th} reflection. The dimension and variance (and therefore Monte Carlo error) of this integrand are considerably larger than the corresponding quantities for the fractional absorption method. Thus, a greater number of rays will be required to achieve the same accuracy. This is weighted by the fact that rays may be computed faster using the discrete method. The two approaches have been compared experimentally for a variety of cases, and the results are given below.

Finally, it should be noted that for a typical reactor, the source term has support only on a portion of D , such as the heating element labeled *surface 0* in Figure 1. Also, while we are interested in the absorbed energy distribution at every point on the wafer, it is only practical to divide the wafer into small elements, such as those shown in Figure 1, and compute the energy absorbed by each one. In general, if D_i is a sub-region of D , the question at hand is how much energy is transferred from the source to D_i . This value can be expressed as

$$\int_{D_i} E^A(x) dx .$$

In the experiments of section 2, we calculate this for the surface elements described in Figure 1. When formulated in this way, the both the fractional and discrete problems lend themselves to Monte Carlo simulations fairly easily. The inclusion of the effects described above, as well as other possible factors is relatively straight forward. A precise description of how these methods were implemented is given in section 2.

1.3 Quasi-Monte Carlo Methods

To evaluate a d -dimensional integral, the standard Monte Carlo method uses independent, uniformly distributed random numbers on the d -dimensional unit cube I^d as the source of integration nodes. In the simplest case, if $\{\mathbf{z}_i\}$ is such a sequence of random points in I^d , then the integral of the function $f(x_1, \dots, x_d)$ over I^d is approximated by the average of f evaluated at the points \mathbf{z}_i . The error,

$$\epsilon = \int_{I^d} f(\mathbf{x}) d\mathbf{x} - \frac{1}{N} \sum_{i=1}^N f(\mathbf{z}_i),$$

satisfies the relationship involving the expectation $E(\cdot)$ of a random variable

$$E(\epsilon^2) = \frac{\sigma^2(f)}{N}$$

where $\sigma^2(f)$ is the variance of f defined by

$$\sigma^2(f) = \int_{I^d} f^2(\mathbf{x}) d\mathbf{x} - \left(\int_{I^d} f(\mathbf{x}) d\mathbf{x} \right)^2.$$

This shows the familiar convergence rate of $N^{-1/2}$ associated with random methods.

The key property of the random sequence which is used here is its uniformity, so that any contiguous subsequence is well spread throughout the cube. This idea has led to the suggestion that using other sequences which are more uniformly distributed than a random sequence may produce better results. Such sequences are called quasi-random or low discrepancy sequences.

Initially it may appear that a grid would provide optimal uniformity. However, grids suffer from several difficulties. First, the number of points required to create even a coarse mesh is exponentially large in dimension. Moreover, not all dimensions are of equal importance in particle simulation problems. Generally the majority of the energy transfer will occur in the first collisions. Thus in a 6 dimensional problem, even if a million rays are used, the first collision will be described by only 10 different values of the variable. Another difficulty is that grids have rather high discrepancy, a quantity which measures the uniformity of a set of points. This is defined and discussed below. Finally, the standard method for increasing accuracy of a grid is to halve the mesh size, which requires adding 2^d times the current number of points. It is desirable to be able to increase the number of points used without adding such an extremely large number. It is unclear how to place additional points on a grid to maintain uniformity unless the mesh is halved, though.

The solution to this problem is to use infinite sequences of points such that for every N , the first N terms of the sequence are in some sense optimally distributed throughout the cube. In order to quantify this, the discrepancy of a set of N points is defined as

follows. If Q is a rectangle contained in I^d , and $m(Q)$ is its volume, then the discrepancy D_N of the sequence $\{\mathbf{x}_i\}$ of N points is

$$D_N = \sup_{Q \in I^d} \left| \frac{\# \text{ of points in } Q}{N} - m(Q) \right|.$$

A uniformly distributed sequence may then be defined as one for which D_N goes to zero as N goes to infinity. By the law of iterated logarithms, the discrepancy of a random sequence is bounded by $(\log \log N)N^{-1/2}$. There are many quasi-random sequences known for which the discrepancy is bounded by a constant times $(\log N)^d/N$, which suggests greater uniformity than a random sequence. Two such sequences, the Halton and Sobol' sequences, were chosen for comparison in this work, and are briefly described now. Further discussion of these sequences can be found in [3]

The Halton sequence [2] in one dimension is generated by choosing a prime p and expanding the sequence of integers $0, 1, 2, \dots, N$ into base p notation. The n^{th} term of the sequence is given by

$$z_n = \frac{a_0}{p} + \frac{a_1}{p^2} + \frac{a_2}{p^3} + \dots + \frac{a_m}{p^{m+1}},$$

where the a_i 's are integers taken from the base p expansion of $n - 1$

$$[n - 1]_p = a_m a_{m-1} \dots a_2 a_1 a_0,$$

with $0 \leq a_i < p$. For example, if $p = 3$, the first terms of the sequence are

$$\left\{ 0, \frac{1}{3}, \frac{2}{3}, \frac{1}{9}, \frac{4}{9}, \frac{7}{9}, \frac{2}{9}, \frac{5}{9}, \frac{8}{9}, \frac{1}{27}, \frac{10}{27}, \frac{19}{27}, \dots \right\}.$$

Note that the numbers lie in cycles of p increasing terms, and that within the cycle, the terms are separated by $1/p$. The effect is that once a grid of refinement $1/p^m$ is filled in by repeated sweeps of these cycles, the next cycle starts filling in the grid at level $1/p^{m+1}$. The d -dimensional Halton sequence is generated by pairing d one-dimensional sequences based on d different primes; usually the first d primes are chosen. One difficulty with this sequence is that in high dimensions, the base p must be large, so that the cycle of increasing terms is rather long. When paired against another large prime based sequence, the result is that the points lie on parallel lines which slowly sweep through the unit square. Thus the distribution of points is not very uniform.

As an alternative the theory of (t,s)-nets has been developed by Sobol' [10], Faure [1] and Niederreiter [7]. The Sobol' sequence solves the problem of large primes by only using $p = 2$. The sequence is generated such that the first 2^m terms of each dimension for $m = 0, 1, 2, \dots$ are a permutation of the corresponding terms of the Halton sequence with prime base 2 (also known as the van der Corput sequence). If the proper choice of

permutations is used, the resulting d -dimensional sequence can be shown to have good uniformity properties. However, as dimension increases, more permutations must be used, and the possibility increases that a bad pairing may exist between two dimensions leading to a highly non-uniform distribution in that plane.

In his book [6], Niederreiter summarizes the properties and theory of these sequences. The key point is that their discrepancy satisfies the relationship

$$D_N \leq c_d \frac{(\log N)^d}{N} + \mathcal{O}\left(\frac{(\log N)^{d-1}}{N}\right),$$

where c_d , different for each sequence, is constant in N , but depends on d . This is an optimal upper bound in the sense that for any infinite sequence, there exist an infinite number of N such that

$$D_N \geq c \frac{(\log N)^d}{N}$$

for some constant c . These bounds suggest that at least for large N , the quasi-random sequences described above will be considerably more uniform than a random sequence.

As discussed in [3] and [4], there is a rather large gap between the theoretical bounds on discrepancy and integration error and what can be observed in practical computations. There is no theory which accurately predicts integration error for a given integrand. However, computational experiments have established various trends in behavior. For many problems, quasi-Monte Carlo methods significantly outperform random sequences in the range of practical N . The convergence rate is generally between $N^{-.5}$ and N^{-1} . There is a tendency, however, for the improvement over random to diminish as the dimension of the problem increases. Computations of discrepancy indicate that there is a transition point in N which grows exponentially with dimension such that before this point, the discrepancy of quasi-random sequences is the same as for random sequences. Also, integrands which have step function like behavior, such as characteristic functions of sets, tend to show less improvement with quasi-random sequences than continuous integrands. The best way to predict performance for a specific type of problem is to analyze the results for a test problem, as is done below.

2 Results

While the idea of quasi-Monte Carlo is almost as old as the Monte Carlo approach itself, these methods have rarely been used in real world engineering problems. This may be attributed to the fact that many Monte Carlo problems are high dimensional, and quasi-random sequences tend to lose their advantage as dimension increases [4]. Another difficulty is that for many simulations (particularly non-linear ones) great care must be taken to avoid problems arising from the fact that the quasi-random numbers are

not independent [5]. Due to their relative simplicity and robustness, standard random Monte Carlo methods have remained dominant.

As the following results show, however, a considerable advantage can be won by applying quasi-random sequences to an appropriate problem. In this case, computational experiments were performed on the radiative heat transfer problem for the cylinder symmetric reactor described above. Several parameters were tested to determine which simulations might benefit the most from the quasi-random approach. Comparisons were made between the discrete and fractional absorption methods, among several absorptivities, among different surface elements on the wafer, between point and surface sources, and among the Halton, Sobol' and pseudo-random sequences. To limit the number of parameters studied, the absorptivity was assumed independent of frequency.

2.1 Experimental Technique

The typical experiment consisted of emitting N rays from a surface or point source located at *surface 0*. The point source is actually a circle under cylinder symmetry. For the surface source, the rays were distributed sequentially by increasing radius on a one dimensional grid so the distribution on the corresponding annulus would be uniform. The initial direction was sampled using Lambert's Law, which means that the zenith angle was sampled from a cosine distribution, while the azimuthal angle was sampled from a uniform distribution on $[0, 2\pi]$. The sampling was done using one point from a multi-dimensional quasi- or pseudo-random sequence, such that two angles were assigned separate dimensions. A third dimension was used to sample the initial energy of the ray from a modified version of Planck's black body distribution. The fractional method then consisted of tracing the ray path through the reactor according to the laws of specular reflection, such that at each reflection a certain percentage of the ray's energy, given by the absorptivity at that point, was transferred to the surface element in question. The tracing continued until the ray had lost a fixed percentage of its initial energy. The discrete method varied from the fractional method only in that at each reflection, either no energy was absorbed and the ray continued on, or the entire initial energy was absorbed and the ray tracing was terminated. In both cases, the final answer was computed as the percentage of the energy which left the source that was absorbed by the target element.

The ultimate goal of the experiments was to determine the accuracy as a function of computation time. This is discussed in more detail below. The first step towards this goal was to compute error size as a function of N , the number of rays emitted from the source. The approach taken here was to compute the "expected" convergence rate and error size by performing the calculation 30 times for each value of N using different, "independent" subsequences of the sequences in question. The error was then averaged to give a better estimate of what the expected error using N rays would be. This calculation was then repeated using a larger value of N , again using different

subsequences. In this way, errors at different values of N are also independent, which allows for an unbiased description of the convergence rate. The values of N tested were chosen to be evenly spaced on a logarithmic scale, so that convergence of the type N^β could more easily be studied.

The terms expectation and independence are only strictly defined for a sequence of random variables. Performing the experiments described above with such a random sequence would show converge at a rate of $N^{-.5}$ and an error size determined by the variance of the integrand. The results for the pseudo-random sequence do indeed show a convergence rate of approximately $1/\sqrt{N}$. The same method of computing expected error is then extended to the quasi-random sequences as a means of comparison. It should be noted that this kind of convergence study using averaging and independent N s requires considerably more computation time than a straight forward simulation using a fixed number of rays.

An additional difficulty in this convergence study was that the exact solution for the energy transfer was not known. Thus it was necessary to estimate the exact answer from the Monte Carlo calculation and use this to measure the error at each value of N . The estimate for the exact answer was obtained by first averaging the 30 results obtained for each of the last four values of N , then taking a weighted average of these four values. Under the assumption of a Gaussian error distribution, it can be shown that this gives an optimal estimate to the true answer.

As stated above, the real quantity of interest is computation time to reach a specified error tolerance. This is of particular importance in comparing the fractional and discrete methods. The discrete method requires fewer reflections, but at the cost of being a higher dimensional integral. Thus N might be larger in the discrete case, but the time to compute each ray path would be smaller than in the fractional case. The experiments just described address only the question of convergence with N . The problem is that computation time depends on the coding of the programs and sequences and is very machine dependent. To avoid issues of optimal coding, which would in any case only be accurate for a specified machine, we devised the following means of comparison for the fractional and discrete methods.

The computation consists roughly in computing an average of n_r random numbers in a time $n_r t_r$ and performing an average of m ray tracing steps in a time $m t_t$. Thus the total computing time is $T = N(n_r t_r + m t_t)$. We observed that most of the time is spent on ray tracing. Because our geometry is not very complex, the most realistic case will be a situation very close to the limit, in which all computation time is spent on tracing. Then a comparison T^f/T^d of the computation time with fractional and discrete method will only depend on the ratio m^f/m^d of the number m^f of ray traces with the fractional method to the average number m^d of ray traces with the discrete method. It is simple to realize that $m^d = 1/A$, but m^f depends on the desired accuracy of the calculation. If we require that a ray be traced using the fractional method until the ratio

of its final energy to its initial energy is less than α (in our case $\alpha = 10^{-4}$), we have $m^f = \lceil \log(\alpha) / \log(1 - A) \rceil + 1$. With our energy limit and $A = 0.4$ we get $T^f / T^d = 7.6^1$. This gives the scale to allow the comparison of the computation time of the fractional and discrete method.

Three absorptivities were tested, $A = 0.1, 0.4$ and 0.7 . The associated number of reflections m^f necessary using the fractional method to achieve the accuracy $\alpha = 10^{-4}$ were 88, 19 and 8 respectively. In order to compare the fractional method across absorptivity, a reference time T^* was chosen as the time required to compute 10^5 rays with $A = 0.1$ using the fractional method. Under the assumption that ray tracing dominates the computation time, $88/19 = 4.63$ times as many rays could be calculated for $A = 0.4$ in the same T^* , and $88/8 = 11$ times as many rays for $A = 0.7$. A similar scaling is used to determine the number of rays which can be computed in T^* for the different absorptivities using the discrete method.

2.2 Experimental Results

For the sake of brevity, results are presented here for only one quasi-random sequence, Halton. The Sobol' sequence was also tested, but the results were generally the same as for Halton; on the occasions where there was a difference, Halton appeared slightly better. Also, all calculations described here were performed using a point source for the radiation. Experiments were also done using a surface source; however, as there is no significant difference in the results of the two approaches, only the point source results are given here.

The first set of calculations is illustrated in Figure 2 and Table 1. This figure shows the expected relative error $\epsilon(N)$ as a function of the number of emitted rays N for the Halton sequence and a pseudo-random sequence using the fractional absorption method with a constant surface absorptivity of 0.4. Results are given for *surface 1* on the wafer (see Figure 1), which is directly visible to the source, and for *surface 2*, on the back side of the wafer. The plotted points are the calculated errors for various N (averaged over 30 runs), while the lines are a least squares fit of the data to the functional form

$$\epsilon(N) = c N^\beta .$$

This is the correct form for the expected error using a random sequence, in which case c is the standard deviation of the integrand and β is $-1/2$. As plotted on a log scale, the error appears as a line with slope β .

Figure 2 illustrates a clear advantage of using a quasi-random sequence over a pseudo-random sequence in the calculation, both in error size and in convergence rate (i.e., β). The error in calculating the energy transfer to *surface 1* with $N = 100000$ is over a factor

¹In our actual calculation this limit was between 6.23 and 6.81 for the different number generators. Hence we were pretty close to the assumed limit.

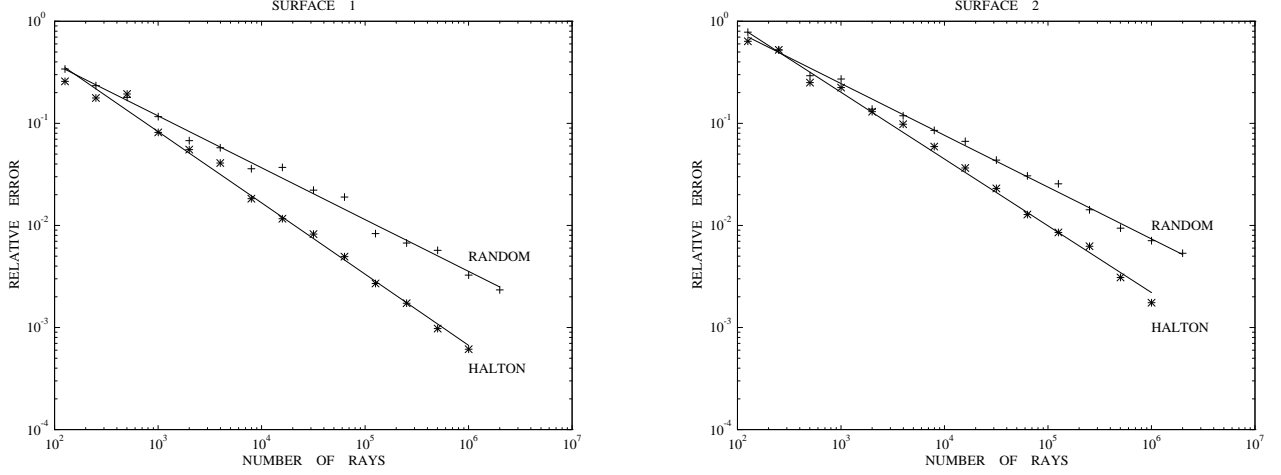


Figure 2: Comparison of random and quasi-random sequences using the fractional absorption method with absorptivity = 0.4.

Surface	Abs.	Convergence $\beta : N^\beta$	Error at $N = 10^5$	Improvement Factor	Error at Time T^*	Improvement Factor
1	0.1	-.64	.0027	2.3	.0027	2.3
	0.4	-.70	.0033	3.4	.0012	4.6
	0.7	-.70	.0042	3.7	.0008	6.0
	mix	-.66	.0036	2.8	.0017	3.3
2	0.1	-.64	.0049	2.7	.0049	2.7
	0.4	-.65	.0099	2.4	.0036	3.0
	0.7	-.65	.0152	2.4	.0032	3.4
	mix	-.65	.0097	2.3	.0047	2.7

Table 1: Convergence results for fractional absorptivity method using the Halton sequence.

of three smaller if the Halton sequence is used than if a random sequence is used. This figure also shows that the accuracy of the calculation depends on the region of the wafer in question. The error in computing the energy transfer to *surface 2* is considerably higher than for *surface 1*. This is related to the fact that a ray must undergo at least one reflection before reaching *surface 2*, as opposed to the possibility of direct energy transfer from the source to *surface 1*. This leads to a higher variance in the rays hitting *surface 2*, and thus larger error. A second consequence is that the improvement gained by using a quasi-random sequence is somewhat less pronounced. Table 1 summarizes the results of the calculations using the fractional absorptivity method for several values of absorptivity, as well as for a mixed absorptivity case in which the walls are highly reflective ($A = 0.1$), while the wafer is highly absorbing ($A = 0.7$). The mixed case is more typical of an actual reactor. The table gives the convergence rate in terms of β , the expected error for the Halton sequence for $N = 100000$, and the ratio of the expected random error to the Halton error. The expected error after computing for time T^* (defined above) is also given as well as the improvement factor of Halton over random for this case. This table shows that a significant advantage is gained by using the Halton sequence in all cases. Again, the advantage is more prominent for the directly visible *surface 1*.

Figure 3 and Table 2 provide a similar comparison of the Halton and pseudo-random sequences for the discrete absorption method. As mentioned above, the integrand being evaluated in this method has a larger variance than the integrand associated with the fractional absorption method, and therefore considerably more rays are needed to obtain the same degree of accuracy. However, the computation time required per ray in the discrete case is on average much less. In Figure 3 the values of N were chosen so as to have the same computation time as in Figure 2, based on the scaling argument given in section 2.1. The dimension of the integral is also much larger, as each reflection corresponds to a separate dimension. Depending on absorptivity, as many as 40 or more dimensions could be required, compared with the three dimensional integrand of the fractional method (in which the ray path is completely determined by its initial direction). As discussed previously, quasi-random sequences tend to lose their advantage over random sequences as the dimension of the integral increases. This is tempered by the fact that the higher dimensions play a less significant role in the calculation. For example, with an absorptivity of 0.4, 99 per cent of all rays have been absorbed after nine reflections. Nevertheless, this dimensional effect on the quasi-random approach can be seen in Figure 3 and Table 2. The factor of improvement for Halton over random is noticeably smaller than for the fractional case, although significant gains are still made by using Halton. Again, there is a considerable difference in the accuracy for *surfaces 1* and *2*, most notably in the case of high absorptivity. In this case most rays do not survive the first reflection, so that many fewer rays eventually make it to *surface 2* than *surface 1*.

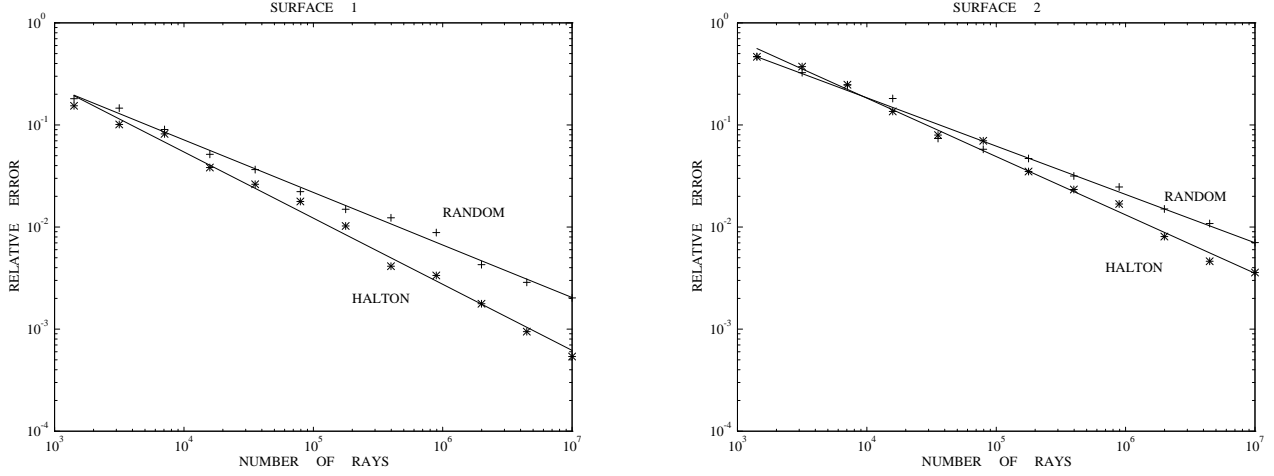


Figure 3: Comparison of random and quasi-random sequences using the discrete absorption method with absorptivity = 0.4.

Surface	Abs.	Convergence $\beta : N^\beta$	Error at $N = 8.8 \times 10^5$	Improvement Factor	Error at Time T^*	Improvement Factor
1	0.1	-.54	.0061	1.2	.0061	1.2
	0.4	-.65	.0030	2.4	.0012	2.9
	0.7	-.66	.0020	3.6	.0006	5.1
	mix	-.59	.0020	2.2	.0011	2.5
2	0.1	-.54	.0141	1.2	.0141	1.2
	0.4	-.57	.0141	1.6	.0064	1.8
	0.7	-.57	.0213	1.3	.0071	1.5
	mix	-.57	.0063	2.0	.0036	2.2

Table 2: Convergence results for discrete absorptivity method using the Halton sequence.

Surface	Abs.	Random Error	QMC Error
		Discrete/Fract.	Discrete/Fract.
1	0.1	1.70	2.23
	0.4	0.75	1.06
	0.7	0.45	0.71
	mix	0.49	0.65
2	0.1	2.17	2.85
	0.4	0.64	1.76
	0.7	0.96	2.20
	mix	0.49	0.78

Table 3: Comparison of error for discrete absorptivity method to error for fractional absorptivity method for random and Halton sequences at T^* .

Table 3 summarizes a comparison of the fractional and discrete methods for both a random sequence and the Halton sequence. For the various surfaces and absorptivities the ratio of the expected error using the discrete method to the expected error using the fractional method is shown. In the random case, both methods have the same convergence rate, so this ratio of errors depends only on the variance of the two integrands (computed from the least squares fit of the experimental data) and the computational time scaling factor described above. Thus this value is independent of the computation time, or equivalently, of N . However, since the two quasi-Monte Carlo cases do not converge at the same rate, the associated error ratio depends on the computation time. The results are given for time T^* .

At low absorptivities, the fractional method is clearly superior for both random and quasi-random, as well as for both surfaces. As the numbers for the random case indicate, the variance of the discrete method integrand is considerably larger than that of the fractional method. There is a further penalty for the discrete method in the quasi-random case coming from the use of a higher dimensional sequence. As the absorptivity increase, the variance difference becomes small enough to favor the discrete method in the random case. This trend can also be seen for quasi-random for *surface 1*, although the low dimensional advantage of the fractional method moderates the gains seen for the discrete method using the random sequence. *Surface 2*, however, shows rather different behavior, strongly favoring the fractional method for all but the mixed case. This can be explained by noting that when the absorptivity is large, many fewer particles ever make it to the back of the wafer in the discrete case. The “effective” N is then much smaller, so that the quasi-random advantage is considerably less. Figure 4 illustrates

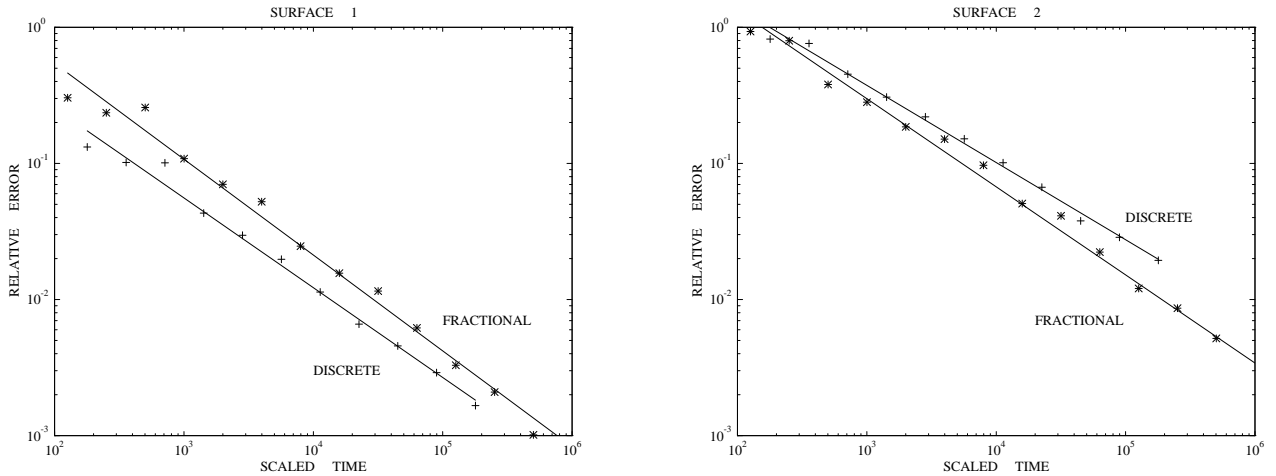


Figure 4: Comparison of fractional and discrete absorption methods using the Halton sequence with absorptivity = 0.7.

this by plotting the relative error for the two methods using the Halton sequence for absorptivity 0.7. Results for both *surface 1* and *surface 2* are shown. The abscissa is computation time in arbitrary units, which allows a comparison of the two methods. The scaling was chosen so that one time unit corresponds to the average time need to compute the path of one ray using the fractional method. On this scale $T^* = 10^5$. In the range of a practical computation, for *surface 1* the advantage of the discrete method over the fractional method starts at about a factor of two, but steadily decreases as the accuracy increases. On the other hand, for *surface 2* the fractional method has an increasing advantage over the discrete approach as the computation time increases.

The results for the mixed absorptivity case are shown in Figure 5. The success of the discrete method for both surfaces may be explained as follows. The wafer acts as a sink for the rays, absorbing 70 per cent of the rays which strike it. The walls simply reflect most rays. Thus a large percentage of the rays emitted contribute to the statistics of the wafer, both front and back, so that both methods show a comparable improvement over the random case. The relatively low variance of the discrete method in this situation then tilts the balance in favor of discrete over fractional. However, as the graphs show, the advantage of the discrete method again decreases with increasing accuracy. This is a result of the superior convergence rate of the fractional method associated with it being a low dimensional integral. In general, the choice of method may then depend on geometry, the regions of the reactor which are of most interest, and the minimum accuracy required.

To round out the computations, we will briefly describe several other results obtained. A third surface element taken at the edge of the wafer was also observed. The size of

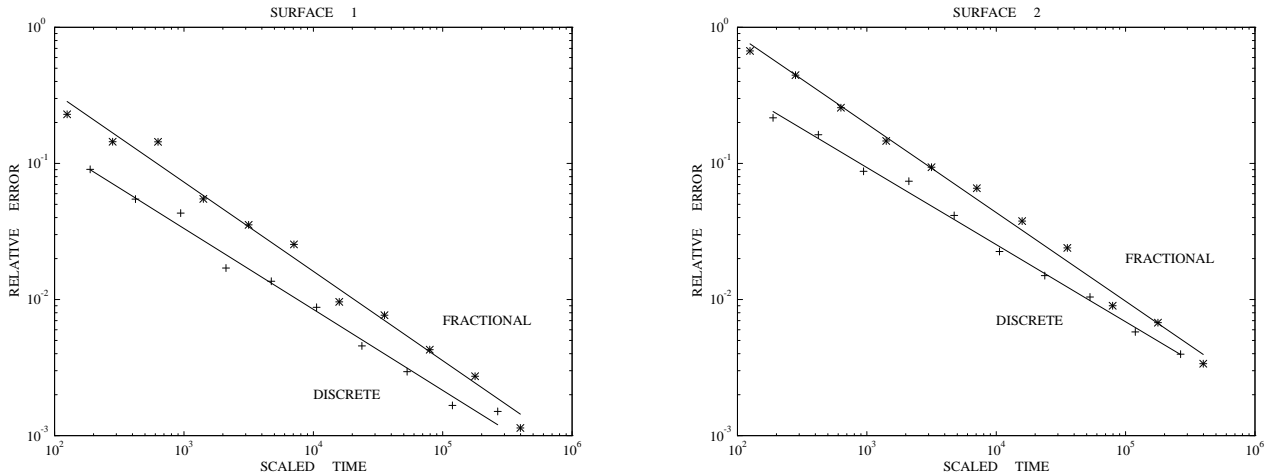


Figure 5: Comparison of fractional and discrete absorption methods using the Halton sequence with absorptivity = 0.1 on reactor walls and absorptivity = 0.7 on wafer.

this element was 2.36 cm^2 , which was much smaller than the other surface elements. Although this element had visibility relative to the heat source similar to that of *surface 1*, the error was almost an order of magnitude larger, and the performance of the Halton sequence was identical or only slightly better than the pseudo-random sequence. This can be attributed to the difference in size of the elements. Many fewer rays strike the smaller surface, leading to an effectively smaller N . Thus for the smaller surface, over the range of N considered, the calculation was still in the range where random and quasi-random are close, similar to what appears in the above graphs at small N . Moreover, the variance of the integrand being evaluated by the simulation is related to the surface size such that the smaller the surface, the higher the variance, and therefore error. This is not surprising, as the use of small surface elements leads to a greater amount of information about the temperature distribution on wafer. To obtain this more refined temperature distribution to the same degree of accuracy of course requires more work (larger N).

Another case considered was that of diffuse reflection at the walls instead of specular. Frequently such cases are often handled by non-Monte Carlo methods. However, as diffuse reflection is easily implemented in a Monte Carlo code, and the results are potentially of interest, a calculation with diffuse reflection for absorptivity 0.4 was carried out. The results using fractional absorption show a significant improvement for Halton over pseudo-random for *surface 1* similar to the results of Figure 2. However, for *surface 2*, there results are the same for Halton and random. The difference between the performance for *surface 1* and *2* can again be explained in terms of visibility. Much of the energy transferred to *surface 1* comes from rays which have not undergone a

reflection. In contrast, to reach *surface 2* requires at least one reflection. Each reflection increases the dimension of the problem by two (two numbers are needed to sample the new direction). Thus the integrand for *surface 2* has a much stronger dependence on the higher dimensions, which leads to a weaker performance by the quasi-random sequence. It should be noted, however, that in no case tested did a pseudo-random sequence outperform the quasi-random sequence.

3 Conclusions

For the simulation of heat transfer reactors and other related radiative transfer problems, the results of this paper indicate that quasi-Monte Carlo methods can be effectively implemented. The error associated with using a quasi-random sequence is always at least as small as in the standard random Monte Carlo approach, and frequently a significant reduction in error can be obtained, up to a factor of 6 in some cases within the range of realistic computations. Because these calculations tend to be dominated by time spent in the ray tracing algorithm, there is no additional computational cost associated with generation of the quasi-random sequence over the pseudo-random sequence. Moreover, the quasi-random sequence may simply be substituted in for the random sequence in existing codes, so that virtually no extra work is required to obtain the quasi-Monte Carlo advantage.

The advantage of quasi-random over random appears to increase as the absorptivity of the reactor increases, for fixed computation time. At lower absorptivities the fractional absorption method is superior to the discrete method. However, as absorptivity increases the discrete method gives better accuracy for surfaces directly visible to the heat source. For the case of highly reflective walls and a highly absorbing wafer, the discrete method was superior for all surfaces, but the advantage over the fractional method decreased with increasing accuracy due to the faster convergence of the fractional method.

The choice of quasi-random sequence (in this case Halton or Sobol') appeared to have little impact on the results. There was also little difference between using a point source or a surface source for the radiation. This additional element of robustness makes quasi-Monte Carlo methods attractive for use in a wide range of radiative transport problems.

References

- [1] Henri Faure, Discrépance de suites associées à un système de numération (en dimension s), *Acta Arithmetica* 41, 337-351, 1982.
- [2] J.H. Halton, On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals, *Numer. Math.* 2, 84-90, 1960.
- [3] W.J. Morokoff and R.E. Caflisch, Quasi-random sequences and discrepancy, *SIAM Sci. Comp.* (to appear).
- [4] W.J. Morokoff and R.E. Caflisch, Quasi-Monte Carlo Integration, submitted to *J. Comp. Phys.* (1991).
- [5] W.J. Morokoff and R.E. Caflisch, A quasi-Monte Carlo approach to particle simulation of the heat equation, *SIAM Num. Anal.* (to appear).
- [6] H. Niederreiter, *Random number generation and quasi-Monte Carlo methods*, SIAM Regional Conference Series in Applied Mathematics, CBMS-NSF 63, 1992.
- [7] H. Niederreiter, Quasi-Monte Carlo methods for multidimensional numerical integration, *Numerical Integration III*, International Series of Numerical Math. 85, H. Brass and G. Hämmerlin eds., Birkhäuser Verlag, Basel, 1988.
- [8] D.M. O'Brien, Accelerated quasi-Monte Carlo integration of the radiative transfer equation, *J. Quant. Spectrosc. Radiat. Transfer* 48, no. 1, 41 - 59, 1992.
- [9] P.K. Sarkar and M.A. Prasad, A comparative study of pseudo and quasi random sequences for the solution of integral equations, *J. Comp. Phys.* 68, 66 - 88, 1987.
- [10] I.M. Sobol', The distribution of points in a cube and the approximate evaluation of integrals, *U.S.S.R. Computational Math. and Math. Phys.* 7, no. 4, 86-112, 1967.