Hierarchical Poisson Disk Sampling Distributions

Michael McCool
Eugene Fiume

Dynamic Graphics Project
Computer Systems Research Institute
University of Toronto
6 King's College Road
Toronto, Ontario, Canada M5S 1A1
{mccool,elf}@dgp.utoronto.ca

Abstract

A relaxation method is presented that generates a series of spatially distributed stochastic samples following the Poisson disk distribution over a range of scales. The series is structured so that a prefix subsequence forms a lower frequency sampling pattern that is also Poisson disk distributed. Unlike strict dart-throwing, the relaxation procedure is guaranteed to terminate. The hierarchical structure allows straightforward adaptive sampling, using techniques analogous to ordered dithering. Also presented is an efficient method to optimize the location of samples under a minimum root mean square quantization error criterion.

Résumé

Une méthode de relaxation est présentée qui génére une série d'échantillons distribuée de façon aléatoire selon une distribution de disque Poisson pour une gamme d'échelles. La série est construite de façon à ce que les premiers échantillons correspondent aussi à une distribution de disque Poisson, mais d'une fréquence plus basse. Contrairement à la méthode du lancé de darts, la methode presentée termine tourjours. La structure hierarchique permet un échantillonage adaptatif simplement en se servant d'une technique semblable à celle connu sous le nom de ordered dithering. Une methode efficace pour optimizer le placement des échantillons dans le sens du valeur quadratique moyenne est également presentée.

Keywords: ray tracing, stochastic sampling, Poisson disk, adaptive sampling, dithering, optimization, antialiasing.

1 Introduction

Stochastic sampling has arisen as an important technique for the estimation of multidimensional integrals in computer graphics [Coo86b, Mit91, LRU85, Mit91].

Area integrals are implicit in the solution of such problems as antialiasing, soft shadows, and the simulation of finite camera apertures (causing depth-of-field effects). Integration over time is needed for motion blur. All these problems can be treated together, resulting in a large multidimensional domain of integration. Ray tracing in particular has benefited from this approach because of its inherent point sampling limitations.

Error bounds can be derived from the variability of the samples: variance, contrast, etc. An approximate error bound on the results of stochastic sampling using variance is given by

$$\epsilon \propto V \sqrt{\frac{s^2}{N}},$$

where N is the number of samples, s^2 is the variance of the sum, and V is the volume of the domain of integration [LRU85, PFTV89]. This error relationship is typically used in numerical applications of the Monte Carlo (stochastic sampling) method. Note, however, that it only decreases as a function of $\sqrt{N}$. An error bound based on contrast is given in [Mit87]; this should more closely match the psychophysical characteristics of the human visual system. Contrast is defined as

$$C = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}.$$

These expressions are particularly useful in adaptive sampling techniques, where only enough samples are taken to satisfy an error bound in a local region. For the contrast measure, supersampling can be initiated if the contrast is above a certain bound, indicating high variability.

Research has focused on the problem of choosing sample positions so as to minimize the impact of this inevitable error while minimizing the number of samples. Both can be accomplished by carefully choosing the position of the random samples.

Instead of structured artifacts such as moiré patterns or staircasing ("jaggies"), aliasing error due to undersampling appears as noise. The distribution of random samples shapes the frequency spectrum of this noise.

Visually, low-frequency noise is particularly objectionable, appearing as "structure" in the image. Also, a low-pass reconstruction filter will typically be applied after the sampling, filtering out high-frequency noise; often this filter will just be a summation of the sampled values (a box filter). For both these reasons, the sampling strategy should try to concentrate the noise in high spatial frequencies.

Mitchell [Mit91] has recently studied the problem of choosing samples across all dimensions of the rendering problem to optimize the spectral characteristics of the noise on the resulting image. This study has some results in common with his; in particular, he gives an algorithm which will generate hierarchical distributions very close to the ones we give in Section 4, and can be used as an alternative to the algorithm given in that section. However, this paper also presents a new and very efficient approach to high-quality adaptive sampling using these distributions, and presents an efficient optimization algorithm that can be applied to other problems such as the quantization of normals [McC91] or generation of random meshes [Tur91].

We will focus on the sampling problem in two spatial dimensions, which applies to the subproblems of area light sources, antialiasing, and camera aperture simulation.

The problem of selecting a spatial sampling distribution has been studied in graphics [CPC84, DW85, Coo86b, Coo86a, CPC84, Mit87], dithering [Uli87], and coded aperture reconstruction [BS81]. The analysis in this paper will follow that in [Uli87]; we will analyse distributions empirically using frequency-domain techniques.

Totally random selection of points is not a good solution, as will be shown in Section 3.1. Other approaches to this problem have included jittering samples from a periodic grid, n-rooks, and the Poisson disk. Because of space restrictions, not all of these sampling approaches have been analyzed in this paper. The following sections will focus on optimizing the Poisson disk approach after first introducing some analysis tools.

There are indications that the Poisson disk distribution is one of the best from a spectral point of view. In comparison with jittering it is difficult and expensive to generate and use: the following analysis assumes a table-based implementation.

2   Analysis Tools

We need a set of tools to analyse the frequency domain characteristics of different sampling patterns. These tools will reduce the response to two one-dimensional graphs, radial power and radial anisotropy. Radial power will measure power at each spatial frequency, regardless of orientation; anisotropy will measure the variability of this measure about all orientations.

An empirical analysis will be used. It is often simpler to see how a sampling pattern actually behaves than to attempt to derive its response, since any derivation will require some simplifying assumptions that may

limit the applicability of the solution. For example, the sequential generation of samples using dart throwing (Section 3.2) is usually not modelled.

We assume that radially symmetric responses are appropriate. Alternatives are possible, i.e. highpass two-dimensional filters with a diamond-shaped stopband. These should be simple extensions of the approach outlined here.

A distribution of samples can be represented by a set of impulses in the plane. To analyse the frequency domain response, the *periodogram* of the distribution is evaluated and measures of power and power variance in a set of circles about the centre are computed.

The periodogram is the Fourier transform of the autocorrelation function. The *autocorrelation* r_f of a real signal f is defined as

$$r_f(\vec{y}) = f(\vec{x}) \star f(-\vec{x}) = \int f(\vec{x}) f(\vec{x} - \vec{y}) \, d\vec{x}$$

For real signals the Fourier transform of the autocorrelation function is equal to the square of the magnitude of the Fourier transform of $f(\vec{x})$:

$$R_f(\vec{\omega}) = \mathcal{F}[r_f(\vec{y})] = |\mathcal{F}[f(\vec{x})]|^2,$$

according to the complex identity $zz^* = |z|^2$, the convolution theorem, and the time-reversal theorem. Using the periodogram will allow the power and anisotropy at different frequencies to be evaluated, since we can compute the anisotropy at any radius. At points closer to the origin, energy in the periodogram corresponds to low-frequency energy in the autocorrelation, and therefore long-distance spatial correlation. Since the periodogram is also the spectrum squared, low frequency power also corresponds to low frequency, or large-scale, structure in the function $f(\vec{x})$.

We want to avoid low-frequency structure in our sampling patterns because this structure could be erroneously perceived as structure arising from the source image. The validity of this quality measure depends on the quality of the reconstruction filter, which should reject high frequencies. It should be noted that a box filter has a $\text{sinc}(\omega_x)\text{sinc}(\omega_y)$ frequency response, which allows some leakage at high frequencies and is consequently not the best choice.

To evaluate the periodogram of a distribution, it is necessary to find the mean periodogram of samples drawn from the distribution. In the following analysis, average periodograms are computed by summing together 100 periodograms of sample images drawn from each distribution. Each image has a resolution of 32×32 and contains 64 impulses, a density of $1/16$. Images are defined over the unit square. An impulse is represented by a value of 1 at the appropriate location in the image. Note that the following approximations and assumptions are made in this analysis:

1. The impulses are finite-width.

2. The location of the impulses are rounded to the nearest grid location.

3. A finite-resolution grid is used.

4. The distribution on the unit square is assumed to tile the plane periodically.

5. A finite number of periodograms are used to estimate the mean periodogram.

Most of these approximations are made so that the discrete Fourier transform, in its efficient FFT form, can be used to evaluate the periodograms.

Note that the assumption that the distribution tiles the plane periodically is exactly the situation in many uses of the distribution, such as stochastic oversampling in ray tracing. The period of the pattern should be much larger than the pixel size.

Once the average periodogram R_f has been evaluated, we can reduce the information to two graphs by computing radial statistics. We define a set of annulli as in Figure 1. Within each annullus, the mean radial power P_i is computed as well as the variance s_i^2. This is the variance *within* an annullus, not between the samples used to create the average periodogram:

$$P_i = \frac{1}{A_i} \int_0^{2\pi} \int_{f_i}^{f_{i+1}} R_f(f\cos\theta, f\sin\theta) f \, df \, d\theta$$

$$s_i^2 = \frac{1}{A_i} \int_0^{2\pi} \int_{f_i}^{f_{i+1}} (R_f(f\cos\theta, f\sin\theta) - P_i)^2 f \, df \, d\theta$$

where $A_i = \pi(f_{i+1}^2 - f_i^2)$ is the area of annullus i defined by radii f_i and f_{i+1}. In practice, the "area" is given by the number of samples on the discrete 32×32 grid that fall within the bounds of a given annullus. The sample size increases approximately linearly up to the edge of the square, then decreases for the annulli that have samples only in the corners of the periodogram. For this analysis, 20 equally-spaced annulli from the centre (0 frequency, or DC) to the corner (highest possible frequency) were used. Figure 2 shows the sample size at each frequency. A dotted line has been placed where the edge of the square is first encountered, and will be present on all diagrams in which it is relevant.
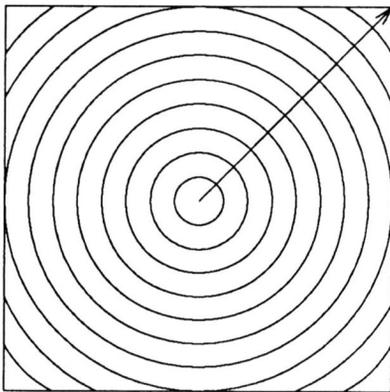

Figure 1: Annulli defined to average the periodogram radially. Twenty annulli were used for this analysis.
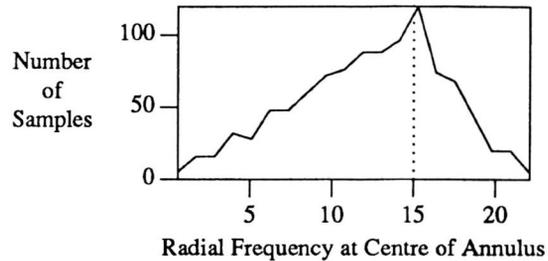

Figure 2: Sample size as a function of radial frequency for estimation of radial statistics in each annullus.

Two graphs will be shown for each distribution analysed: the mean radial power $P_r(f)$ and the anisotropy $A_r(f)$ defined as

$$A_r(f) = \frac{s^2(f)}{P_r^2(f)}.$$

Larger values of anisotropy indicate a greater uneveness in the radial distribution of power. Power and anisotropy vary over a wide range. For this reason they will be plotted in decibels; a value x is expressed in decibels according to the relation $x_{dB} = -20\log_{10}(x)$. A doubling of power or anisotropy approximately corresponds to a 6dB change.

The periodograms will also be displayed in three-dimensional graphs. In these graphs, the DC peak has been reduced by a factor of 4 and actual power is shown, not decibels. These plots are for qualitative comparison of sampling distributions.

3 Random Distributions

Several random sampling patterns have been studied. The one most used in practice is jitter sampling, which moves samples off a uniform grid by a random amount. This technique is simple to implement, but is not spectrally optimal.

This section analyses only Poisson disk sampling and derivatives, which have a better frequency response than jitter (less low frequency noise). They also have another advantage over jitter sampling: an arbitrary number of samples may be used.

The following subsection first analyses strict random sampling for comparison purposes.

3.1 Poisson

The simplest random distribution is the so-called "Poisson" distribution. The coordinates of a point are selected at random from a uniform distribution. Every point is completely independent of the others, making the generation of points a Poisson process. The number of points in an area A is Poisson-distributed with mean equal to the area times the density. For this reason, the distribution is often called a Poisson distribution. This

unfortunate terminology should not be confused with the univariate discrete Poisson distribution.

A sample two-dimensional Poisson distribution is shown in Figure 3. The tile outlined in the centre of the diagram tiles the plane periodically in a square array; parts of adjacent tiles are shown. All distributions in this study will be shown this way.

The centre tile shown here would extend over several pixels if this sampling pattern were being used for antialiasing.

The periodogram of this distribution, as well as an average of 100 other instances, is shown in Figure 4. Radial averages for these distribution are shown in Figure 5. Anisotropy is shown in Figure 6.

Figure 4: Sample and Average ($n = 100$) periodograms for the Poisson distribution. The DC peak has been reduced by a factor of 4.

Figure 3: Uniform distribution of impulses on the unit square.

Figure 5: Radially averaged power for the spatial Poisson distribution. Top: single sample distribution; bottom: average ($n = 100$) radial power.

Note how clumpy the spatial distribution in Figure 3 appears; this is due to the image having significant power at all frequencies, as we can see in looking at the periodogram. The clumps arise because the low frequency power in the spectrum of the distribution allows a variation in the local average of intensity to occur. However, we see that the periodogram is a reasonably good approximation to an impulse function.

The distribution is seen to have an even amount of anisotropy over all frequencies, which is desirable for our application; we do not want any preferred direction in the noise. The clumpiness, however, will undoubtably give poor performance. To eliminate the clumpiness we will have to eliminate the low frequencies in the spectral response.
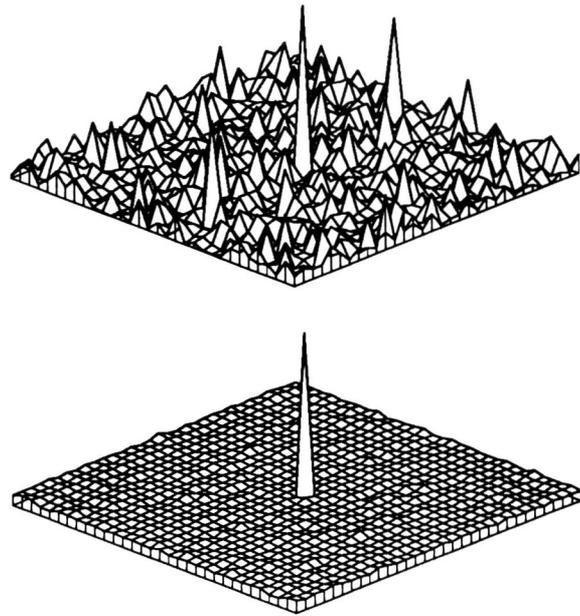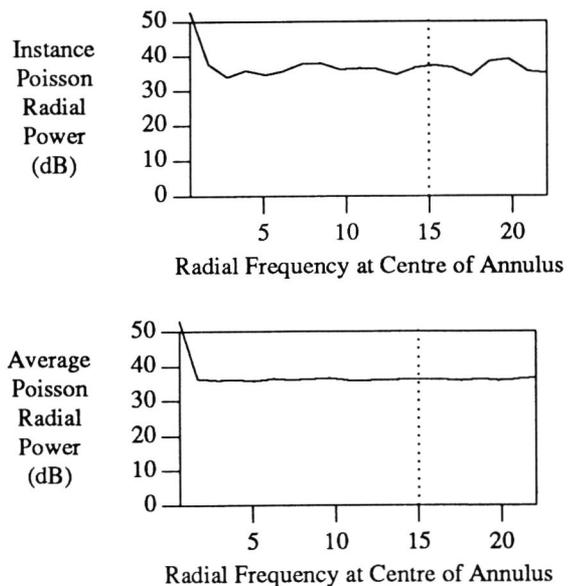
Instance Poisson Radial Anisotropy (dB)

Radial Frequency at Centre of Annulus

Average Poisson Radial Anisotropy (dB)

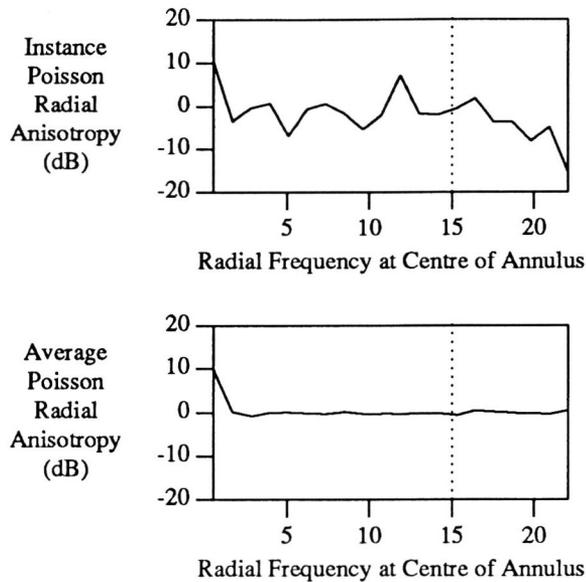Radial Frequency at Centre of Annulus

Figure 6: Radially averaged anisotropy for the spatial Poisson distribution. Top: single sample distribution; bottom: average ($n = 100$) anisotropy.

Figure 7: Poisson disk distribution in the plane. Center tile drawn with disks having the same diameter as the minimum intersample spacing.

3.2 Poisson Disk

The Poisson disk distribution is a Poisson distribution in which no two points are closer than a minimum distance[1].

In theory, to get a sample of a Poisson disk distribution one must generate Poisson distributions and evaluate each one according to the minimum distance requirement. This is not a very practical algorithm, and effective approximations have been developed.

The *dart-throwing* algorithm [Coo86b] places points sequentially. Each new point is compared to points already placed; if it is too close it is discarded. The algorithm terminates after a specific number of points have been placed, or it has proven impossible to place new points after a large number of attempts. A Poisson disk distribution in the plane, generated using dart-throwing, is shown in Figure 7. Circles are drawn about each point in the centre tile at half the minimum intersample spacing. A separation (disk diameter) of 0.1 that would result in a reasonable run-time was chosen experimentally. Intertile interference was also checked so that the periodic tiling of the plane would also satisfy the Poisson disk criterion.

Another class of algorithms that generate similar distributions are the *error diffusion* algorithms [FS75, Uli87]. These have been used in ray tracers [Mit87], but

are limited to a discrete raster. It would also be possible to generate an adaptive sampling grid using an error diffusion algorithm on the grey-scale error image after the first pass of a ray tracer. However, error-diffusion dithering still has some artifacts, particularly in constant regions [Uli87]. We will not use an error-diffusion algorithm, but will use another technique adapted from *ordered* dithering in Section 5 when we discuss adaptive sampling.

It has been found that the Poisson disk distribution has desirable spectral properties, namely low energy for low frequencies. Sample and average periodograms for this distribution are shown in Figure 8. The radially-averaged statistics for this distribution are shown in Figure 9 and Figure 10.

4 Relaxation

Following a suggestion by Robert Lansdale (documented in [Lan91], and also referenced obliquely in [DW85]), the minimum radius for the distribution required by dart-throwing does not have to be fixed.

Points are placed starting with a large radius initially. Once no more space has been found at this radius for a certain (large) number of attempts, the radius is reduced by some fraction (which is just less than 1). The final distribution will still have a minimum separation given by the last radius used. A magnification fraction is also specified to increase the number of tests as the number of samples already placed increases. To save tests, the best-fitting sample during each iteration can be saved. Eventually, the minimum separation of this

---

[1] the Poisson disk distribution is characteristic of the receptors in the retina outside of the fovea [YJ83, WC83], where it prevents aliasing; the retinal image is subsampled in those regions. Within the fovea, the receptors are hexagonally packed at a density twice the highest frequency of the highest spatial frequency passed by the cornea/lens/iris system, so the Nyquist criterion is satisfied.
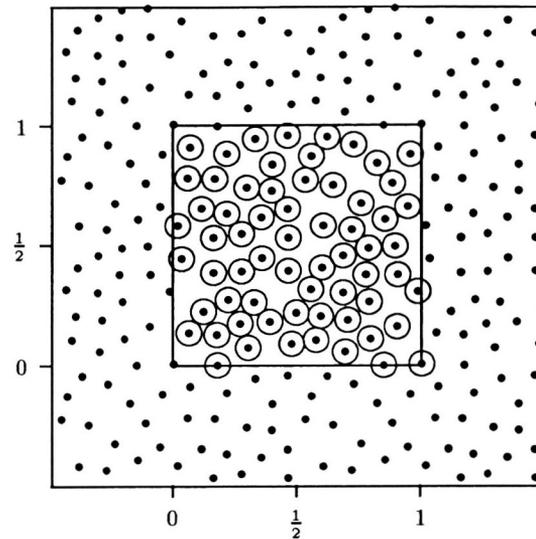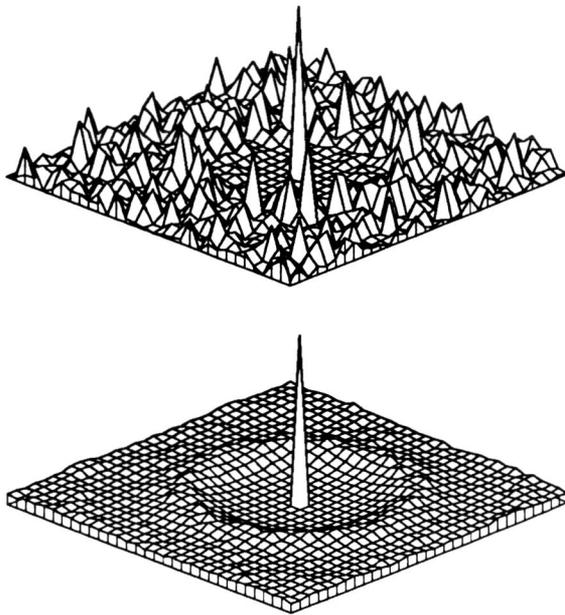
Figure 8: Sample and average ($n = 100$) periodograms for dart-throwing Poisson disk distribution. The central DC peak has been reduced by a factor of 4.
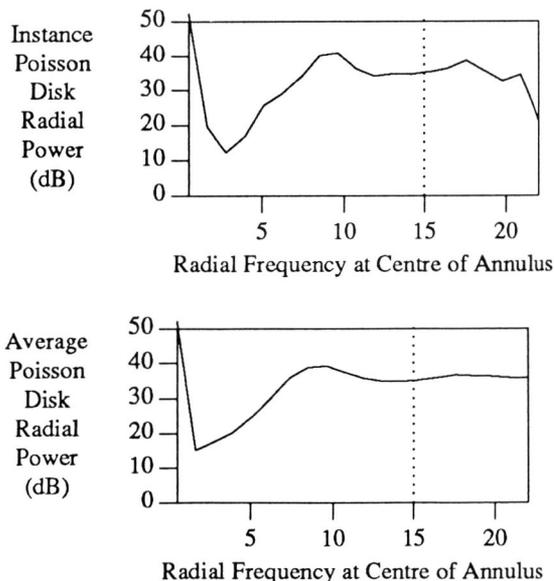
Figure 9: Radial average power for the Poisson disk distribution. Top: single sample distribution; bottom: average ($n = 100$) distribution.
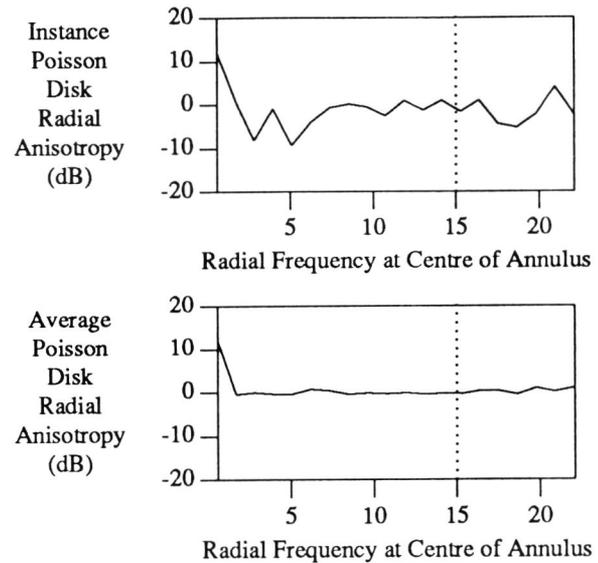
Figure 10: Anisotropy for the Poisson disk distribution. Top: single sample distribution; bottom: average ($n = 100$) distribution.

sample and the decaying radius will meet. In this way work from previous tests will not be wasted.

Such a distribution is shown in Figure 11. The circles surrounding each point in the centre tile indicate the radius in use when that point was placed. Visually, the distribution generated is similar to that generated by a fixed radius.

A similar approach was taken by [Mit91], but was not exploited for its adaptive sampling potential. Our algorithm relaxes the radius condition only when necessary, obtaining the best possible fit for the number of samples considered. Unfortunately, this may result in a longer running time. We will show empirically that the spectrum retains the desirable properties of the original Poisson disk distribution, at least over a limited range of scales. The algorithm differs from Mitchell's in two ways: a hard radius constraint is used; and the radii are guaranteed to be strictly decreasing (the last point could be addressed by sorting the samples generated by Mitchell's algorithm).

The new algorithm has many advantages. First, it will always terminate with a position for any desired number of samples, since the radius should eventually get small enough to allow every one to be placed. It is very possible that a dart-throwing algorithm with a fixed radius will not allow all samples to be placed, and will therefore fail to produce a distribution.

Second, this algorithm is less sensitive to its initial choice of parameters than strict dart-throwing. A larger threshold of trials or a larger decay fraction will increase running time and (hopefully) increase the "tightness" of the distribution, but will not cause the algorithm to fail totally. There might, however, be some regions which are more densely packed than others. This can
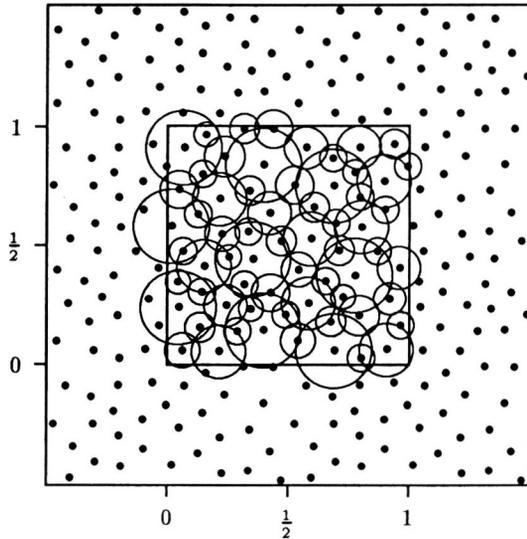
Figure 11: Approximation to a Poisson disk distribution generated by relaxation dart-throwing. The circles about points on the left indicate half the minimum intersample separation in use at the time the point was placed.

happen if the radius decay is too steep or the number of tests per iteration too few. Since the algorithm is adaptive, the same set of parameters can be used to generate distributions for a wide range of circumstances.

Finally, and probably most importantly, the resulting list of samples has a pyramidal property. Only one high-resolution sample distribution has to be generated for a wide range of sampling densities; if a lower sampling density is required, then only the initial points from the sample list are selected. These points will satisfy a larger minimum distance criterion than the full set.

In ray tracing, a common problem when using stochastic sampling is how to adaptively improve the sampling density. In the past, jitter sampling has been used rather than Poisson disk sampling because of the expense of generating each distribution. In [Mit91], Mitchell suggests scaling down the sample grid and replicating it; this, however, will introduce some unwanted periodicity which may compromise the effectiveness of the random grid and reintroduce coherent noise. This will only be a problem if a short period is used. With this algorithm, a single long list can be used and the required number of samples chosen, resulting in a consistent long period.

We need to analyze this distribution quantitatively. It will be shown that, empirically, its spectral properties closely match those of the fixed-radius dart-throwing algorithm. The distribution in Figure 11 was generated with an initial minimum spacing of 0.3 (disk radius of

0.15). The radius was multiplied by 0.99 after 1000 samples were tried without any success in placing a new point. Its periodogram is shown in Figure 12, along with an average periodogram of 100 other distributions generated with the same parameters. The radially-averaged statistics for this distribution are shown in Figure 13 and Figure 14.
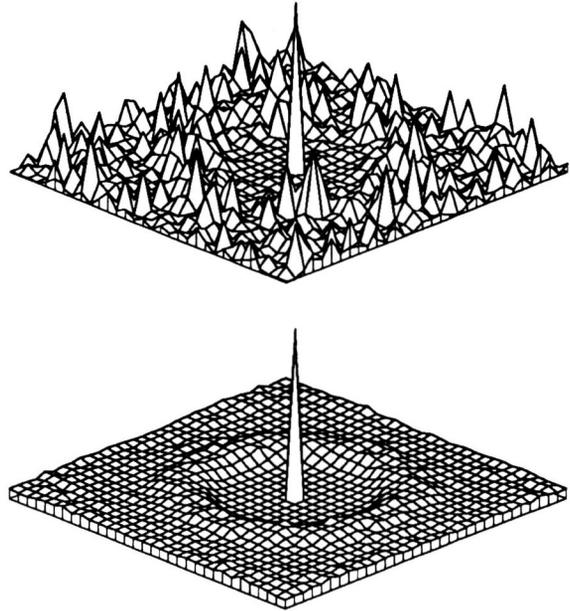

Figure 12: Periodograms for Poisson disk distributions generated by relaxation. Top: single sample distribution; bottom: average ($n = 100$) periodogram. The central DC peak has been reduced by a factor of 4.

The sequence of radii for the samples is given in Figure 15, as well as the average over the 100 tested distributions. Note that only the final samples approach the radius limit of 0.05 used for the basic dart-throwing in Section 3.2.

5   Adaptive Sampling

5.1   Continuously Variable Sampling Density

In many cases, it is desirable to locally modify the sampling density. For example, in the Introduction a formula was given for estimating the error, which depended on the variance of the samples collected. Selective supersampling corresponds to discrete levels of sample density, and can be triggered by a set of contrast thresholds. Using the pyramidal property of the relaxation-generated sequence, as many samples can be taken as needed, until an error criterion is satisfied. Every new sample will fit into a Poisson disk distribution, at ever-decreasing minimum spacing.

Instance
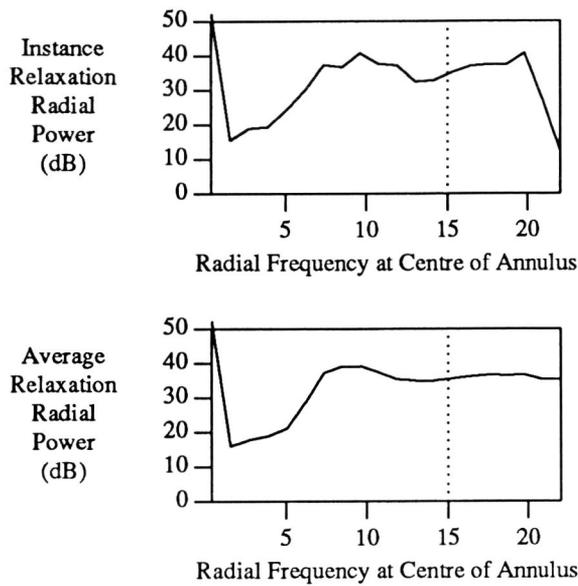Relaxation
Radial
Power
(dB)

Instance
Radius

Figure 13: Radial average power for the Poisson disk distribution generated by relaxation. Top: single sample distribution; bottom: average ($n = 100$) distribution.

Average
Relaxation
Radial
Power
(dB)

Average
Radius

Figure 15: Sample Radii for the Poisson disk distribution generated by relaxation. Top: single sample distribution; bottom: average ($n = 100$) distribution.

A set of subsequences can be extracted and stored; each subsequence has samples that fall into predetermined partitions. The refinement can then be localized to only the partitions whose error estimates exceed a bound. The ultimate extension of this approach is importance sampling.

5.2 Importance Sampling

If the approximate shape or some factor of the function to be integrated is known, then the number of samples should be increased in areas in which the function has a high value. The sampling density then replaces a weighting factor. Many examples of this occur in computer graphics.

For example, consider the ray tracing simulation of a glossy surface. The energy reflected towards the eye is actually an integration over the hemisphere of the incoming energy, multiplied by the reflectance function. The reflectance function is known; the incoming energy is not. The integral can be performed using stratified Monte Carlo sampling. Random samples are shot in a distribution of orientations, but more samples should be shot in directions from which incoming energy will be strongly reflected.

For another example, it may be possible to replace a complex reconstruction filter with a simple box filter (mean of samples in a pixel) but to use importance sample according to the weight given by a higher-quality filter. This would have the effect of concentrating more samples in the center of the pixel.

Importance sampling can be trivially solved using an approach from ordered binary dithering. In ordered

Instance
Relaxation
Radial
Anisotropy
(dB)
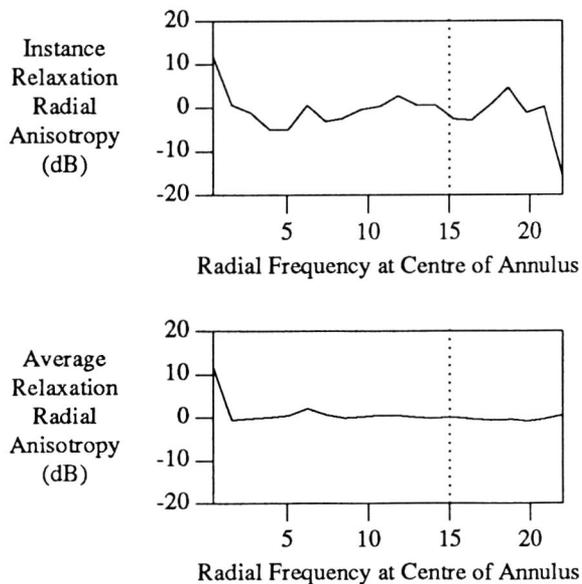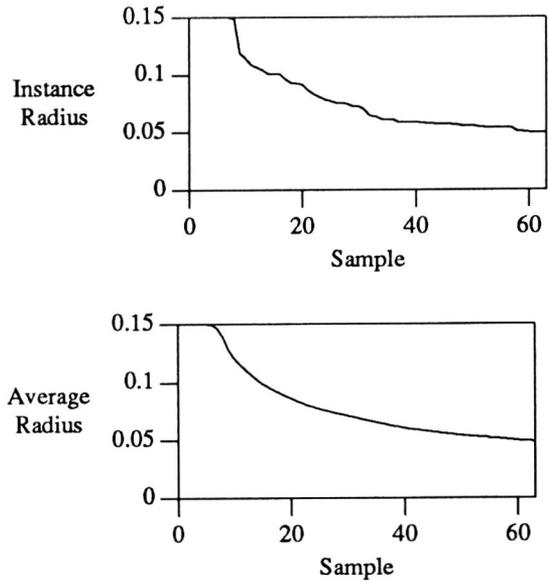
Average
Relaxation
Radial
Anisotropy
(dB)

Figure 14: Anisotropy for the Poisson disk distribution generated by relaxation. Top: single sample distribution; bottom: average ($n = 100$) distribution.
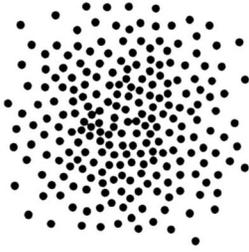
Figure 16: Adaptive Poisson disk sampling; density weighted by $\exp(-5(x^2 + y^2))$.

dithering, a mask is used which is an array of thresholds, one per pixel. A full mask is usually generated by repeating some basic pattern. At any point, if the pixel in the source image is greater than the threshold at that point in the mask, then the pixel is turned on.

A similar idea can be used if we have previously generated a large Poisson disk distribution with uniform density using relaxation. We know that the more samples chosen from the list and displayed, the denser the samples will be, although they will all follow a Poisson disk distribution at some scale. Therefore, select the samples according to the following algorithm:

1. For each sample $\vec{s_i}$, compute its rank as $r(\vec{s_i}) = i/n$, $1 \leq i \leq n$.

2. Let
$$x(\vec{s_i}) = \frac{f(\vec{s_i}) - \min(f)}{\max(f) - \min(f)},$$
where f is the known importance function. In the first example above, this would be the value of the reflectance function; in the second, it would be the weight of the reconstruction filter.

3. If $r(\vec{s_i}) \leq x(\vec{s_i})$, choose $\vec{s_i}$ as an evaluation sample point; otherwise, discard it. If a sample is chosen, it is part of the pattern needed to represent the density x.

This is very efficient, since the same table of samples can be used over and over for a variety of importance functions, as long as care is taken to not allow overly periodic use of the same sampling pattern.

An example of a set of samples whose density is weighted by $\exp(-5(x^2 + y^2))$ over $[-0.5, 0.5]^2$ is shown in Figure 16. A subsequence of 227 samples were selected out of 512.

Once a subsequence has been generated, it can *still* be used in the manner described in Section 5.1, since it will still satisfy the increasing density criteria. Only as many sample points are taken as are needed.

6 Optimization

Although a Poisson disk distribution has some nice qualities, it is possible to generate a better one from a quantization viewpoint.

We would like the distance from every point on the plane to the nearest sample point to be as small as possible. If the function to be estimated varies smoothly, then this strategy will result in the minimum root mean square (RMS) error. It will also mean that no large "gaps" will be left in the sampling structure.

We know the optimal solution to this problem, at least on the plane: a periodic distribution of samples in a hexagonal grid. However, we want to avoid periodicity. This should not be a problem, since optimization procedures that can find global optimums are rare. In this case, we actually *want* an optimization procedure that will return a somewhat suboptimal solution close to our initial configuration. We can then initialize with a Poisson disk distribution and hopefully derive a result which still retains most of its useful properties.

The Generalized Lloyd Algorithm (GLA) described further in [Llo82, JS86, For88] is based on a minimization of the (root) mean square error in quantizing each position in a multidimensional metric space to the nearest sample point. It is a strictly descending method and as such is unlikely to find global optimums in complex situations.

Given $\vec{x}$ distributed according to probability density function $p(\vec{x})$ over a space with distance metric $d(\vec{x}_1, \vec{x}_2)$, the mean square error is:

$$x_{\mathrm{MSE}} = \sum_{i=1}^{N} \int_{\mathcal{D}_i} p(\vec{x}) d^2(\vec{x}, \vec{x}_i) \, d\vec{x},$$

where $\mathcal{D}_i$ is the domain of $\vec{x}_i$: the set of all points that map to $\vec{x}_i$. Note that both probability and error are accounted for in this metric. This metric more heavily weights extremes of error, while not neglecting the mean error rate. The mean square metric is convenient analytically because many useful results can be proven using it.

Two results that are useful in the design of quantizers give necessary (but not sufficient) conditions for an optimal MSE quantizer. In an optimal MSE quantizer,

1. Borders between two domains will be equidistant from the respective quanta under the distance metric d.

2. Quanta will be placed at the mass centres of their domains, where the mass centre of a domain is defined as
$$\dot{x}_i = \frac{\int_{\mathcal{D}_i} \vec{x} p(\vec{x}) \, d\vec{x}}{\int_{\mathcal{D}_i} p(\vec{x}) \, d\vec{x}}$$

In other words, the quanta should be the average value of their domain.

These conditions are easily proven.

Condition 1: Every point x with a non-zero probability should be mapped to the quantum which it is closest to under the distance measure d. Obviously, if it was not, the error would be higher and the mean square error metric larger. The boundaries are therefore equidistant under d, since points on the boundary

can choose either quanta with no penalty in the error metric. If this was not true of the boundary, then the points on this boundary could be moved to the lower error domain and the metric reduced.

Condition 2: The value $\int_{\mathcal{D}_i} p(x)d^2(x, x_i)\,dx$ is the *moment of inertia* of $p(\mathcal{D}_i)$ around the point x_i; it obtains its minimum when x_i is $\hat{x}_i$, the mass centre of $p(\mathcal{D}_i)$.

The same results hold if the distance measure is angle on the surface of a sphere, with appropriate modifications to definitions (the edges of the domains will be great circles, for example).

The GLA alternates attempting to satisfy the above conditions. First, a set of initial quanta are chosen. Boundaries between domains are set using condition 1, and then the mass centres of each domain are computed. The quanta are moved to the mass centres to satisfy condition 2, which of course changes the boundaries and invalidates condition 1. The process is repeated until no more changes are necessary, assuming the $\vec{x}_i$ converge.

The optimization procedure can be used to generate a non-uniform distribution; simply vary the $p(\vec{x})$ term when computing the centroid. The probability distribution $p(\vec{x})$ should be smooth or the optimization may get wedged into a local minimum.

This technique will converge faster if the initial distribution is nonuniform. Such a distribution can be generated by the algorithm in Section 5. Alternatively, relaxation can be used directly, but the radius modified by the weighting function.

The positions of the quanta and the domain boundaries may not converge and may in fact go to infinity, depending on the shape of the probability distribution. This problem is avoided in the current problem by optimizing on a finite domain; the sides of the rectangle wrap around and so the optimization effectively takes place on the surface of a toroid.

Each iteration can be implemented by constructing a Voronoi diagram and finding the centre of mass of each cell. Construction of each diagram takes $O(n \log n)$ time, although in a sophisticated optimization procedure the diagram could be built incrementally; only small changes are required for each new iteration.

To generate the results shown here, the Monte Carlo method was used to estimate the centroid of each cell. This is an inefficient method compared to the Voronoi approach, but is simple to implement.

In other uses of this optimization scheme, a globally optimal solution may be important. Even if the positions of the quanta do converge, there is no guarantee that the quantizer thus derived will be globally optimal. In fact it is unlikely in complex situations, since the method is descending. If an optimal solution is desired, decaying noise can be added to simulate annealing. This will not be too expensive because the underlying optimization procedure is accurate.

A distribution optimized using a uniform probability distribution is shown in Figure 17. Ten iterations were used. As with the random distributions, 100 of these distributions were analysed and the results appear in Figures 18, 19 and 20. The average spectra shown in Figure 18 displays peaks corresponding to the horizon-

tal and vertical directions. The danger of using this technique is that the optimization may actually find a global optimum and regain periodicity. An optimized distribution might be combined with jitter to avoid this problem.
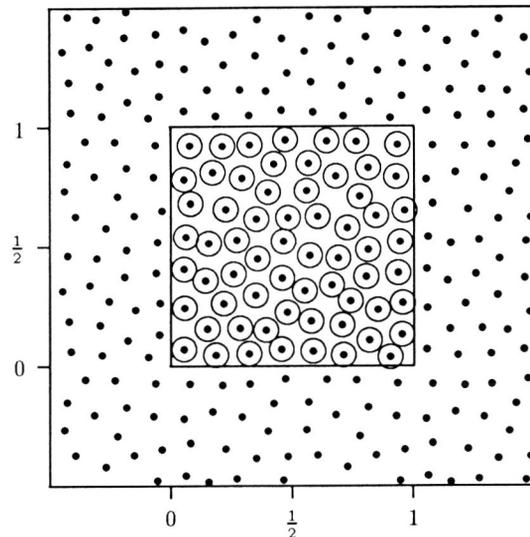

Figure 17: Optimization of a uniform distribution generated by dart-throwing. The circles about points on the central tile indicate half the minimum intersample separation in use at the time the point was placed.

7 Conclusions

A number of refinements in the generation and use of the Poisson disk stochastic sampling strategy have been presented and analysed empirically. It has been shown that by using relaxation, a sequence of samples can be generated such that prefix subsequences are also Poisson disk distributed. Sequences with this property can be used in both adaptive and importance sampling in a very efficient manner.

Unfortunately, to use these distributions in practice still requires large lookup tables, since the distributions were generated using a refinement of dart-throwing; the generation process is still relatively expensive and should not be included within the rendering loop.

An optimization technique has been presented which has uses in sampling and quantization. It will also be useful in other areas where points have to be placed uniformly within a finite domain, such as mesh generation over arbitrary smooth shapes. It can optimize sample positions to fit an arbitrary density distribution, although a smooth distribution has more chance of success.
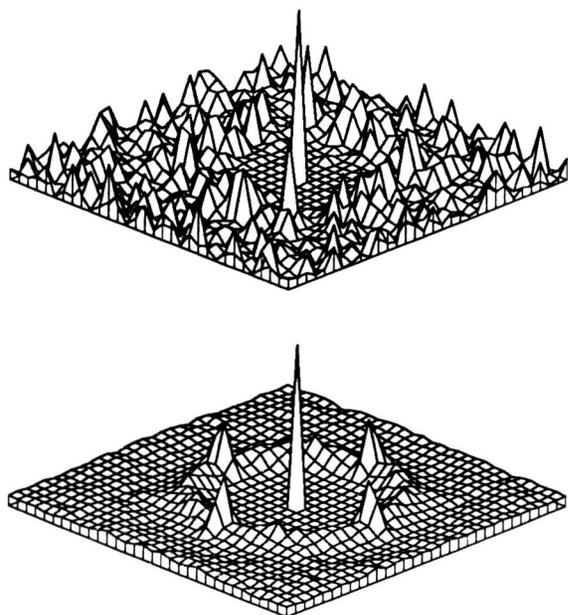
Figure 18: Periodograms for optimized distributions. Top: single sample distribution; bottom: average ($n = 100$) periodogram. The central DC peak has been reduced by a factor of 4.
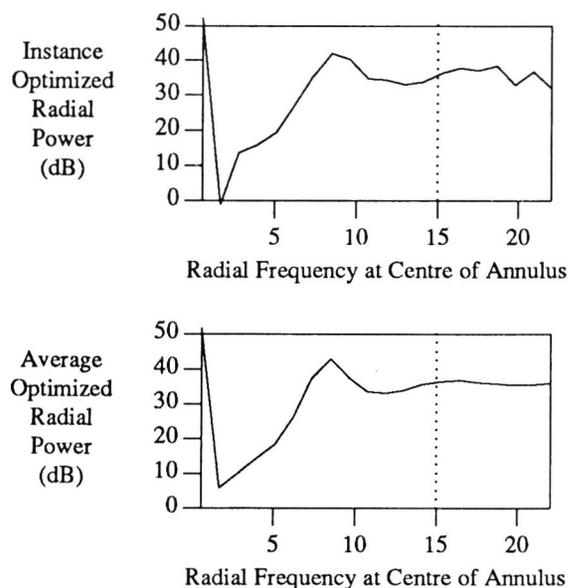

Figure 19: Radial average power for the optimized distributions. Top: single sample distribution; bottom: average ($n = 100$) distribution.
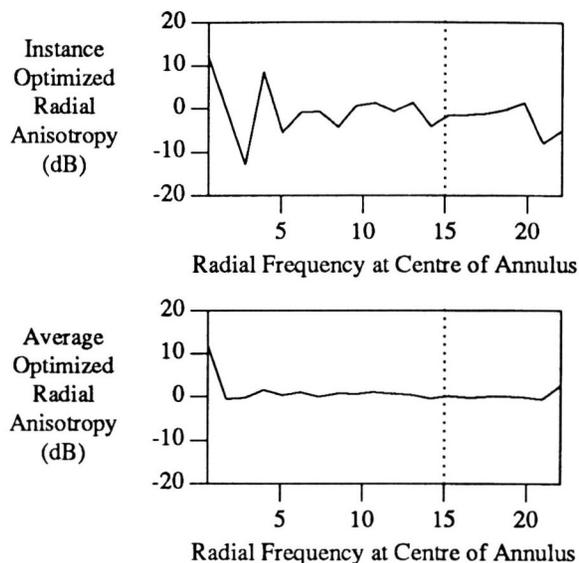

Figure 20: Anisotropy for the optimized distributions. Top: single sample distribution; bottom: average ($n = 100$) distribution.

8 Acknowledgements

References

[BS81]   Harrison H. Barrett and William Swindell. *Radiological Imaging*, volume 2. Academic Press, 1981.

[Coo86a]   Robert L. Cook. Practical aspects of distributed ray tracing. In *ACM SIGGRAPH '86 Developments in Ray Tracing seminar notes*, August 1986.

[Coo86b]   Robert L. Cook. Stochastic sampling in computer graphics. *ACM Transactions on Graphics*, 5(1):51–72, January 1986.

[CPC84]   Robert L. Cook, Thomas Porter, and Loren Carpenter. Distributed ray tracing. *ACM Computer Graphics (ACM SIGGRAPH '84 Proceedings)*, 18(3):137–145, July 1984.

[DW85]   Mark A. Z. Dippé and Erling Henry Wold. Antialiasing through stochastic sampling. *ACM Computer Graphics (ACM SIG-*

GRAPH '85 Proceedings), 19(3):69–78, July 1985.

[For88]   Bruno Forte. Topics in information theory: Lecture notes, April 1988. School for Advanced Studies in Industrial and Applied Mathematics, Department of Applied Mathematics, University of Waterloo.

[FS75]   R. Floyd and L. Steinberg. An adaptive algorithm for spatial grey scale. *Society for Information Display (SID) International Symposium—Digest of Technical Papers*, pages 36–37, 1975.

[JS86]   Neil Judell and Louis Scharf. A simple derivation of Lloyd's classical result for the optimum scalar quantizer. *IEEE Transactions on Information Theory*, IT-32(2), March 1986.

[Lan91]   Robert Lansdale. Applications of 2D and 3D texture mapping in computer graphics. Master's thesis, University of Toronto, Department of Electrical Engineering, 1991.

[Llo82]   S. P. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, IT-28(2):129–136, March 1982. reprint of paper presented in 1957 at Inst. of Math. Stat. Mtg. in Atlantic City.

[LRU85]   M. Lee, R. A. Redner, and S. P. Uselton. Statistically optimized sampling for distributed ray tracing. *ACM Computer Graphics (ACM SIGGRAPH '85 Proceedings)*, 19(3):61–67, July 1985.

[McC91]   Michael McCool. Compact data structures for volume visualization. Master's thesis, University of Toronto, Department of Computer Science, 1991.

[Mit87]   Don P. Mitchell. Generating antialiased images at low sampling densities. *ACM Computer Graphics (ACM SIGGRAPH '87 Proceedings)*, 21(4):65–72, July 1987.

[Mit91]   Don P. Mitchell. Spectrally optimal sampling for distribution ray tracing. *ACM Computer Graphics (ACM SIGGRAPH '91 Proceedings)*, 25(4):157–164, July 1991.

[PFTV89] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes*. Cambridge University Press, 1989.

[Tur91]   Greg Turk. Generating textures on arbitrary surfaces using reaction-diffusion. *ACM Computer Graphics (ACM SIGGRAPH '91 Proceedings)*, 25(4):289–298, 1991.

[Uli87]   Robert Ulichney. *Digital Halftoning*. MIT Press, 1987.

[WC83]   D. R. Williams and R. Collier. Consequences of spatial sampling by a human photoreceptor. *Science*, 221(4608):385–387, 22 July 1983.

[YJ83]   John I. Yellot Jr. Spectral consequences of photoreceptor sampling in the rhesus retina. *Science*, 221(4608):382–385, 22 July 1983.