Hierarchical Annealing for Random Image Synthesis*

Simon K. Alexander[1], Paul Fieguth[2], and Edward R. Vrscay[1]

[1] Department of Applied Mathematics
University of Waterloo, Waterloo, Ontario, Canada, N2L-3G1
{sk2alexa,ervrscay}@uwaterloo.ca
ph: (519) 888-4567 Ext. 5455 fax: (519) 746-4319
[2] Department of Systems Design Engineering
University of Waterloo, Waterloo, Ontario, Canada, N2L-3G1
pfieguth@ocho.uwaterloo.ca

Abstract. Simulated annealing has been applied to a wide variety of problems in image processing and synthesis. However, particularly in scientific applications, the computational complexity of annealing may constrain its effectiveness, in that the demand for very high resolution samples or even three-dimensional data may result in huge configuration spaces. In this paper a method of hierarchical simulated annealing is introduced, which can lead to large gains in computational complexity for suitable models. As an example, the approach is applied to the synthesis of binary porous media images.

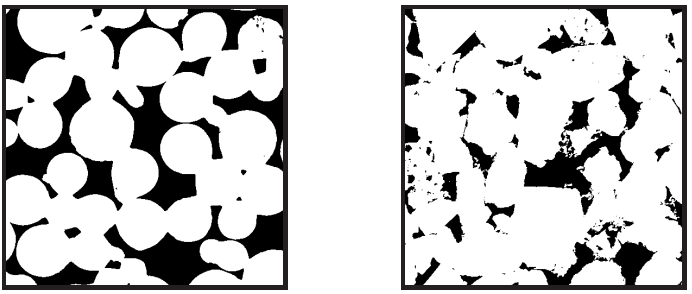
1 Introduction

In this work we are motivated by challenges in scientific imaging, in particular, by the demand for the random synthesis of very high resolution 2D images or 3D cubes. Our interest in this field is driven by studies in binary porous media, possessing solid-gas or solid-fluid distributions, two examples of which are shown in Fig. 1. Although these have been simulated extensively (stochastic geometry [19, 20], annealing methods [23, 13]) computational issues are at the heart of further research progress: Larger 2D and 3D simulations are required to validate scientific models on ever smaller scales. Thus our goal is efficient synthesis or sampling of huge random images of this type.

It is important the recognize the distinction between the superficially similar problems of random synthesis, considered here, and the much more common problem of image estimation. The problem of estimation involves finding that particular, normally unique, image which *optimizes* some criterion (with respect

* This research was supported in part by the Natural Sciences and Engineering Council of Canada (NSERC).in the form of grants (P.F. and E.R.V) and a Postgraduate Scholarship (S.K.A). We also wish to thank M. Ioannidis, Dept. of Chemical Engineering University of Waterloo for providing sample porous media images, and interesting discussions.

A. Rangarajan et al. (Eds.): EMMCVPR 2003, LNCS 2683, pp. 194–210, 2003.



(a) (b)

Fig. 1. Binary image (a) of packed glass spheres (common test data for porous media applications) and (b) a rock sample

to some model and measurements), an inherently deterministic problem. In contrast, the synthesis of appropriate images by sampling is an inherently stochastic problem, the selection of random samples from a statistical distribution. The distinction between the two is illustrated in Fig. 2. Although *many* approaches have been developed for optimization (including closed-form solvers, linear systems, and a huge variety of ad-hoc methods), the sampling problem is much more subtle.

As opposed to relatively well-conditioned problems involving densely-measured images, such as image denoising and segmentation, our interests involve the sampling of random images subject to prior constraints and sparsely sampled measurements, normally a very difficult and poorly-conditioned problem. Applying annealing approaches to such numerical stiffness typically requires huge numbers of iterations and vast computational requirements.

There is one approach, however, which blurs the distinction between sampling and estimation, and which may hold the key to an efficient sampler. The method of simulated annealing (SA) (described in §2) has been widely used for common imaging processing tasks, such estimation from noisy images [7, 8], image segmentation [7, 18, 2], and image synthesis [17, 23]. It is a promising approach in many imaging applications, in that it can accommodate a wide variety of assumptions and prior image models, even nonlinear ones. Estimation effectively becomes a limiting case of random sampling, in the limit where only the most probable samples are accepted. The well-known problem, however, is that SA is exceptionally slow.

To be sure, a wide variety of accelerated annealing methods have been proposed, many promising ones based on various hierarchical approaches [18, 14, 25, 17]. The basic premise, as with related methods such as multigrid [10, 4, 15], is that the problem is greatly accelerated by addressing long-range phenomena and dependencies on a coarse-scale, leaving only more local phenomena to be ad-

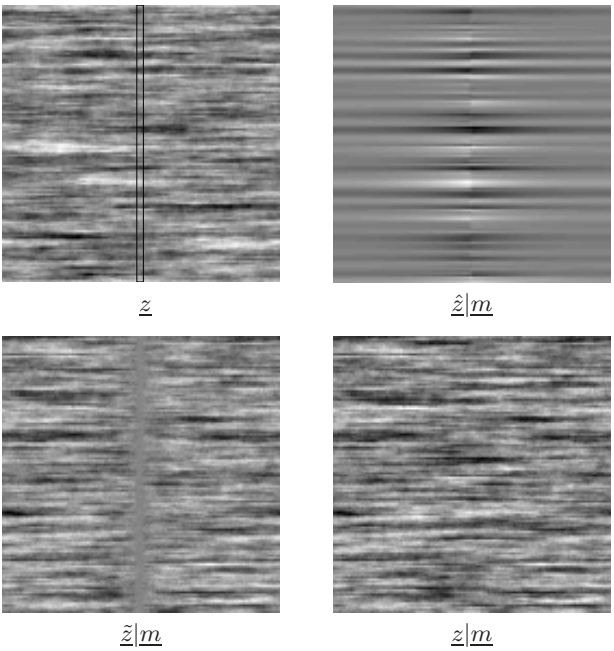


Fig. 2. A comparison of estimation and posterior sampling: the top left panel shows an anisotropic wood-grain texture and the measured central column $\underline{m}$. The estimates $\underline{\hat{z}}|\underline{m}$ clearly reveal the pattern of the measurements, and bears little resemblance to the original texture. The bottom panels show the sampled posterior estimation error $\underline{\tilde{z}}|\underline{m}$, in which a low-variance zero-mean band can be seen where the estimates are good. Finally the posterior sample $\underline{z}|\underline{m}$ shows the random texture, consistent with both the measurements and the prior statistics

dressed on finer scales. It would appear, then, that the success of these methods should lead to the desired large-scale sampling. Unfortunately, this appearance is misleading, and underlies our whole research objective.

In particular, virtually every method of accelerated annealing has been aimed at problems of estimation/optimization, such as image classification [25], segmentation [7, 3, 18], and image restoration [7, 9, 11]. In *each* of these cases the finest scale is *densely* measured (the image to be analyzed), strongly constraining the problem, thus both the passing of long-range information and the constraints imposed by a prior model are relied upon relatively weakly in arriving at a solution. That is, the hierarchical framework can be relatively heuristic and approximate and still produce excellent estimates. In sharp contrast, random sampling proceeds from a blank slate — all long-range structure has to be synthesized explicitly, based on the constraints of the prior model, which therefore needs to be implemented and followed faithfully.

Secondly, the strong conditioning of an estimation problem implies that coarse-scale convergence is fairly robust, making it possible to avoid anneal-

ing at all but the finest scale [25]; thus the coarse scales can be managed by fast, deterministic methods such as *iterated conditional mode* (ICM). Again, in distinct contrast, any hierarchical approach to random sampling will require proper sampling on each scale, which furthermore raises subtle issues of how to move from scale to scale and the range of temperatures over which to anneal.

In order to address these difficulties in a hierarchical framework we are motivated by the renormalization group theory approaches originating in statistical mechanics [24], which suggests that temperature and scale are related. Intuitively, at a particular scale and temperature, structures from coarser scales are 'slushy' (i.e., relatively frozen for some length of time) while at finer scales are a 'boiling froth' (i.e., highly variable). Hence both the coarser and finer scales can, to some extent, be ignored while the structure at the given scale is developing. To be sure, other renormalization-motivated approaches have been proposed for image modelling in applications such as restoration [9] and vision [6], however both of these are subject to the limitations of estimation-based approaches, as discussed earlier.

From the preceding discussion, we see how the application of hierarchical sampling represents a significant departure from existing methods, and yet represents a significant problem in scientific modeling and analysis. In this work, we will discuss how the reduction in the size of configuration space represented by subsampling (e.g. renormalization) can be used to reduce computation cost, concentrating on the potential gains for the computationally expensive applications of image synthesis.

This paper seeks to initiate discussion of methods to improve the computational cost of sampling by working in a multiscale/hierarchical framework. One method to do this is to use a simulated annealing approach, recast in a hierarchy of coarse-grained approximations to the configuration space. In the following sections, we discuss the algorithms involved, the construction of energy functions to describe appropriate Gibbs random fields, and the resulting image samples.

2 Simulated Annealing

2.1 Background

In this work we are interested in Markov Chain Monte-Carlo (MCMC) sampling methods [7, 5]. MCMC methods are used to draw samples from Gibbs random field — that is, a random field whose probability distribution π is written in terms of an energy function $\mathcal{E}(x)$ weighted by inverse temperature $\beta = 1/T$:

$$\pi_\beta(x) = \frac{e^{-\beta\mathcal{E}(x)}}{Z_\beta} \quad . \tag{1}$$

Note that the normalizing *partition function* Z_β, which is particularly difficult to calculate for most problems, is not needed by the sampling algorithms under consideration (e.g. Metropolis[16] and Gibbs[7] samplers), which is one of the primary benefits of MCMC methods.

Algorithm 1 Simulated Annealing

$n \Leftarrow 0$
while $\mathcal{E}(x)$ not converged **do**
  $\beta \Leftarrow 1/T_n$
  $x \Leftarrow$ sample $\pi_\beta(x)$ { draw a sample from π_β}
  $n \Leftarrow n + 1$
end while

The simulated annealing (or stochastic relaxation [7]) algorithm is as follows: Here $\{T_k\}$ describes a general *cooling schedule*, a sequence of (eventually) decreasing temperature values with

$$T_k > 0 \qquad \forall k \ , \tag{2}$$

$$\lim_{k\to\infty} T_k = 0 \ . \tag{3}$$

We note, however, that convergence to a global minimum (i.e. global optimization of the energy function) is guaranteed [7] only for impractically slow logarithmic cooling satisfying

$$T_k \geq \frac{A}{\log(k+1)} \ , \tag{4}$$

where A is some model-dependent constant. Determining the minimal value of A is not straightforward [7]. Furthermore, it can be shown that for appropriate sampling algorithms, if the energy function has multiple global minima, the algorithm will sample uniformly from these minima (independent of initial states) [7, 8]. This property will be employed in §3 to draw samples from a large class of images.

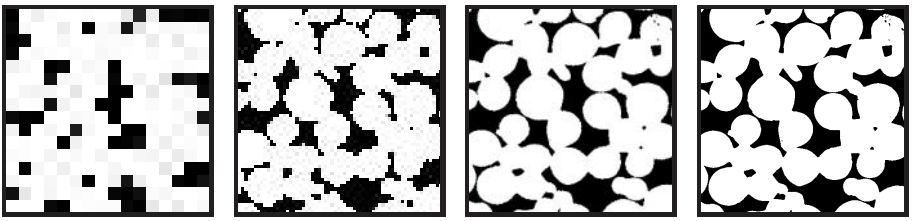
There are thus two basic considerations for the computational cost of this class of algorithms:

1. The process of drawing samples (the MCMC algorithm);
2. The cooling schedule of the annealing.

We maintain that these two issues are orthogonal; that is, that they can be separately addressed, and that gains from accelerated cooling [22] can be applied more or less equally to both regular and multiscale annealing. Additionally, extensive research into improving the performance of MCMC samplers has not resulted in general improvements, although special conditions can result in improved convergence [1, 21].

Therefore the focus of this paper is to reduce the computational cost of annealing by restructuring the process to take advantage of sampling on smaller lattices where possible. We will construct a hierarchy of coarse-grainings of the configuration space, in order to allow fast convergence of lower frequency components, as discussed in the following section.

If we consider the cost of individual site updates to be fixed (true for a large class of models), then the computational cost of drawing samples is driven by



(a) 16×16 (b) 64×64 (c) 256×256 (d) 512×512

Fig. 3. A porous medium image viewed at several resolutions: How do local and non-local features scale and map across resolutions?

the rate of convergence of the Markov chain, which increases with the size of the domain, as does the number of pixels. For large domains, the number of iterations can scale to huge proportions.

It is this last aspect, the scaling of computational cost with domain size, that we will leverage to reduce total computational cost. The key proposal is that by annealing in a multiscale hierarchy, the computational cost of annealing is greatly reduced.

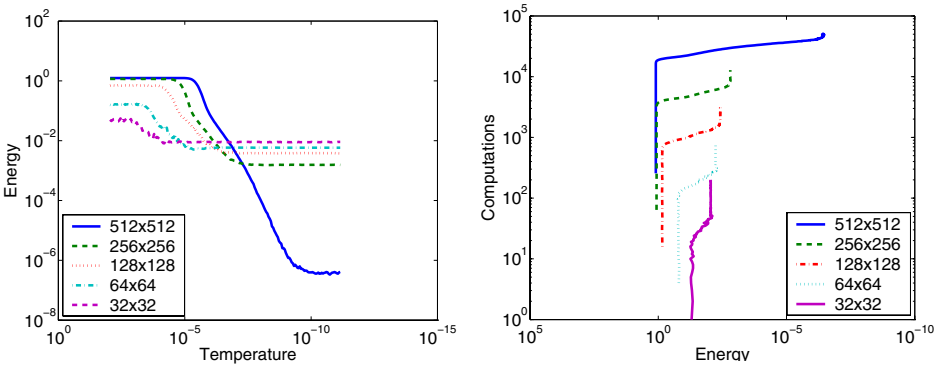
2.2 Multiscale Annealing

The purpose of this approach is to take advantage of the lower complexity at higher levels in the hierarchy, and to have these higher levels precondition, in a sense, or simplify the task at the lower levels.

Our proposal is to apply the central idea of renormalization group theory to accelerate computationally burdensome annealing in scientific applications. This approach is distinct from existing research on hierarchical structures and accelerated annealing such as "multi-temperature annealing" [12] which constructs a hierarchical Markov random field model with scale dependent cooling, but where the Markov model explicitly couples together adjacent scales. In contrast, we treat each scale as a single image, only mapping the coarse configuration onto the finer lattice, i.e. always in the direction of increasing resolution.

As we work down a hierarchy from coarsest to finest resolution, the question then is to determine what image features are preserved (or lost) across rescaling. At any particular level in this multiscale hierarchy, what image features are represented? How may we anneal in such a way that certain features of interest are represented at the current level and can be meaningfully mapped to the the next finer level? Figure 3 illustrates this view.

The key insight that we bring from renormalization methods is that the effective temperature for a given feature size is scale dependent. In particular, for some temperature at some intermediate scale, coarser scales are cold (meaning



(a) (b)

Fig. 4. Annealing to convergence at several resolutions. (a) Effective temperature shifts for convergence are evident. Note that the energy axis represents the energy as measured at finest resolution, regardless of the resolution.(b) Computations normalized to cost of 32×32 pixel flips

that the large, coarse features are 'frozen'), and finer scales are hot (meaning that the tiny features, not resolved at the intermediate scale, are still in a state of flux). What this implies is that we are able to concentrate on one intermediate scale at a time. To illustrate this approach, Figs. 4(a) and 4(b) illustrate the convergence process at several scales (empirical curves based on the model in §3), plotting the energy (that is, the convergence) versus both temperature and computational cost. We note the following:

- Each curve is characterized by a rapid drop in energy, followed by convergence.
- The computational cost of a single sweep at fine levels can be high relative to the total cost of convergence at a coarse level.
- The finer the scale, the better the ultimate convergence.
- Some features of these plots are model dependent; the large difference between minimal energy at finest resolution and all the other resolutions is an artifact of the sensitivity of the model to blocks (i.e. from supersampling coarser resolutions) in the high resolution image. This is not a measure of distance in configuration space.

From which it follows that, in general, we want to be working on the coarsest, unconverged scale.

The benefit of this approach stems from two observations. First, the size of the coarse domains is small, allowing rapid sampling. Second, and much more significant: At an intermediate scale the algorithm needs to be iterated only enough to allow relatively local structure to converge, since the larger structures

Algorithm 2 Multiscale Annealing

$k \Leftarrow 0$
for $s = S$ to 0 **do**
  while $\mathcal{E}_s(X_s)$ not converged **do**
    $\beta \Leftarrow 1/T_k$
    $X_s \Leftarrow$ sample π_β { draw a sample from π_β}
    $k \Leftarrow k + 1$
  end while
  $X_{s-1} \Leftarrow M_s(X_s)$ {map to next finer resolution}
end for

already converged at coarser scales. Our goal, then, is to work in a decimated configuration space (at less computational cost) for sufficient time and then map onto the next larger space and to continue annealing. Clearly this idea may be applied hierarchically, leading to a succession of coarse-grained configuration spaces. If we consider the Gibbs field X at the target resolution, we can denote a hierarchy $\{X_s\}_{s=0}^{S}$ of coarse-grainings of the configuration space where each increase in level represents decimation by a factor of two (and where X_0 is finest resolution). We note that this is not the only possible renormalization map [9].

Annealing with too low an initial temperature (not enough energy) approaches a greedy algorithm, prone to finding local minima. On the other hand, too much energy may destroy the larger structures passed down from coarser scales. Since the above holds true at any level in the hierarchy, clearly there is a delicate balance in achieving computational gains while retaining good optimization performance.

An analysis of interactions between the cooling schedule of the annealing, the mappings in the annealing hierarchy, and the convergence of the stochastic sampler is not straightforward. At present, we rely on heuristic rules to determine the cooling and mapping schedule. Future work will concentrate on an analysis of this process. It is worth noting that since we employ a single cooling schedule across the entire hierarchy, the 'shape' of this schedule will strongly effect overall computational costs. From a computational point of view, the more time spent in the coarser levels of the hierarchy the better. Of course this must be balanced against the effects of the cooling schedule on convergence.

The model presented in the next section uses the following heuristic. At each higher level in the hierarchy, the energy function can be viewed as an approximation for the energy function at the final resolution, denoted as $\mathcal{E}_s$. Denote the mapping operator (coarse to fine) from one level to the next as M_s : $X_s \rightarrow X_{s-1}$. Multiscale annealing is performed as shown in Algorithm 2.2.

While in no sense is this heuristic considered to be "optimal", in the next section it is applied to a simple model and shown to be effective.

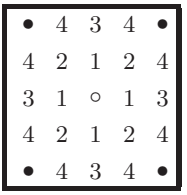


Fig. 5. Neighbourhoods up to 4^{th} order

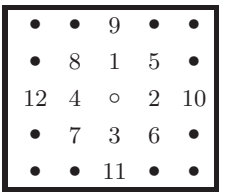


Fig. 6. Bit indices for a 3rd-order neighbourhood defining the mapping $\mathcal{F}$

3 A Simple Model for Synthesizing Porous Media

We have intentionally chosen a simplistic model in order minimize computational cost, as well as model parameters and clutter in order to interpret the results more easily. In particular we will work with binary images, which have the advantage of smaller configuration spaces (compared to non-binary images of the same size) while still being important for many application areas. Henceforth the two pixel states will be referred to as *white* (1) and *black* (0).

We consider local neighbourhoods of fairly small order. Using a common notation, neighbourhoods are determined by increasing Euclidean distance from the current pixel. The $k+1^{\text{th}}$ order neighbourhood contains all of the pixels in the k^{th} order neighbourhood, and so on. Figure 5 illustrates the first four neighbourhoods.

We denote the neighbourhood (of some fixed order) of a pixel as $\mathcal{N}$, and let b denote the size of the neighbourhood (i.e the number of pixels). Since each pixel is binary, the size of the set of all possible configurations within a neighbourhood is 2^b. For reasonably small order, it is computationally feasible to count the instances of each local configuration in an image.

For this purpose, there is a natural bijective mapping constructed by labelling each pixel in the neighbourhood uniquely from 1 to b, and treating the m^{th} pixel state as the state of the m^{th} bit in a b bit binary representation of an integer in $0 \ldots 2^b - 1$. For example with 3rd-order neighbourhoods there are 12 pixel locations. Figure 6 shows one possible indexing scheme forming a map onto $0 \ldots 4095$. Under such an indexing scheme, each location in the image has, via the neighbourhood structure, a mapping:

$$\mathcal{F}_{i,j} : X_{i,j} \to 0 \ldots 2^b - 1 \ . \tag{5}$$

Given this indexing of local configurations, one possible approach is to consider, for some class of bitmap images, the global distribution of local configurations. Designate target probability mass functions (pmfs) for the two cases of white central pixel and black central pixel as the following:

$$p_w^s[n],\ p_b^s[n], \qquad n = 0 \ldots N-1\ , \tag{6}$$

Where $N = 2^b$.

Figure 7 shows several such pmfs for the packed spheres data (Fig. 1(a)), demonstrating the difference in mass distribution at several different resolutions. For any given image (i.e configuration) the pmf may be approximated by a histogram:

$$h_x^w[n],\ h_x^b[n], \qquad n = 0 \ldots N-1\ , \tag{7}$$

with total counts C_x^w, C_x^b, respectively. Thus the sample probability of configuration k for a white central pixel is $\frac{h_x^w[k]}{C_x^w}$. These sample statistics can be efficiently maintained while performing stochastic sampling.

Given the above, one possible energy function (used in (1)) at level s is a weighted sum of errors with respect to the target pmf at level s in the hierarchy:

$$\mathcal{E}_s(x) = \sum_{n=0}^{N-1} \left[\alpha_n^s \left(p_w^s[n] - \frac{h_x^w[n]}{C_x^w} \right)^2 + \beta_n^s \left(p_b^s[n] - \frac{h_x^b[n]}{C_x^b} \right)^2 \right] \ . \tag{8}$$

At each level in the hierarchy, the target pmfs define a new energy function. This can, however, be viewed as an approximation to the energy function at the final resolution.

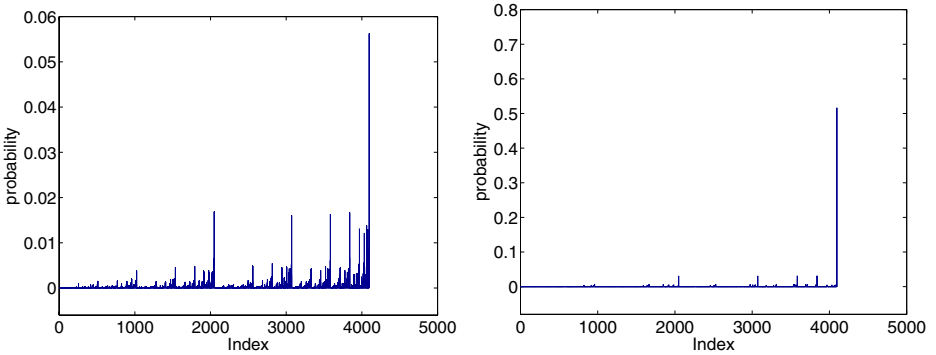
More complicated energy functions may be proposed to allow more sophisticated modelling of particular image classes. In particular, nonlocal terms may be needed to accurately reflect image morphology. These considerations, while interesting, are separate from the issues surrounding hierarchical annealing proposed here. Additionally, the best choice of distance metric (or for that matter, including a non-metric such as the Kullback-Leibler distance) is not clear for this process. No claim of optimality is made the example given in (8).

3.1 Implementation

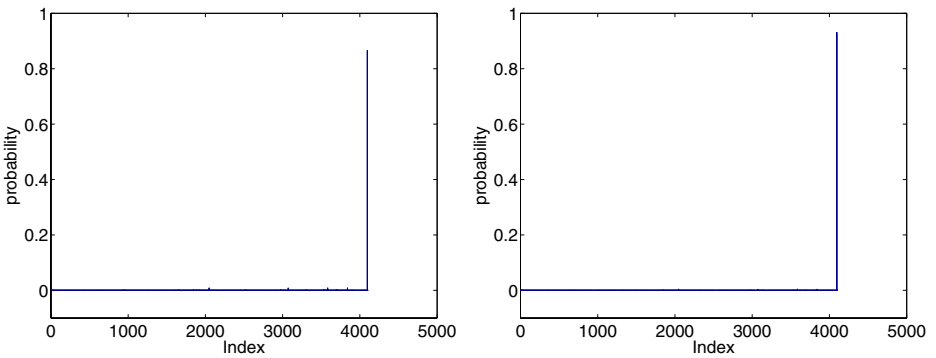
In this experiment, target pmfs were measured as the mean of sample distributions from sets (50 – 100 images) of training data. Weights in (8) are taken as 1 for this experiment, but need not be in general. The pmfs shown in Fig. 7 correspond to the resolutions shown in Fig 3, illustrating how mass distribution in the pmf varies with scale. This approach allows a very simple training of our model.

Samples were drawn from the models using a Metropolis [16] sampler with random site location. In all cases annealing was performed using a geometric cooling schedule [22]:

$$T_k = \alpha^k T_0 \qquad 0 < \alpha < 1\ . \tag{9}$$



(a) 16×16 (b) 64×64



(c) 256×256 (d) 512×512

Fig. 7. Target pmfs at several resolutions. These particular pmfs are the sample mean distributions given a white central pixel at various resolutions for a set of packed glass spheres, see Fig. 3. The mass of the distribution is mostly located at homogeneous white neighbourhoods at the finest resolution (index 4095 corresponds to all white pixels). At low resolutions, the mass distribution exhibits a more interesting structure

In the experimental data presented here, the cooling parameter was taken as $\alpha = 0.9$. This parameter could, of course, be tuned to particular applications. As previously discussed, issues of 'optimal' cooling are considered to distract from the purpose of illustrating hierarchical annealing. Improvements may be made with analysis of particular models and cooling characteristics.

The heuristic rules discussed in §2.2 were used to define progression through the hierarchy while annealing. In these experiments such an approach proved capable of reaching globally minimal energy states. It is expected that work with more complex energy functions (and therefore more expensive to compute) would reveal limitations for this simple heuristic and present directions for more sophisticated approaches.

It is worth noting that, by construction, (8) does not allow for zero energy states in general. The size of the image domain will constrain the quantization, or step size, of the histogram approximations to the target probability mass functions. Hence in the general case for each non-zero value in the target pmf we can expect an error on the order of the inverse of the number of pixels. If we denote the number of pixels in an image domain D as $|D|$, and take b as the number of bits in the mapping (5) then we may estimate the minimal energy as

$$\mathcal{E}^2_{\min} \approx 2^b \frac{1}{2} \frac{1}{|D|^2} \quad . \tag{10}$$

4 Results

The results that follow are demonstrative of typical runs of the hierarchical annealing method presented in this paper. Two training data sets were used, one of a pack of glass beads, the other, natural rock.

Figure 8 shows a plot of computational cost vs. energy. This graph is identical to the one given in Fig 4(b) with the addition of a curve for the hierarchical annealing. Here, as previously, the energy for all curves is measured in the finest resolution lattice (the convergence in higher levels is governed by the energy function at that level, of course) in order to make comparison possible. This plot clearly demonstrates the computational advantages of the approach. Additionally, it is shown that both methods converge to a minimal energy. As noted above the histogram quantization does not allow zero energy configurations.

Figure 9, similarly, contains the data of Fig. 4(a) along with the hierarchical annealing values for comparison purposes. This figure demonstrates how the annealing schedule for a hierarchical annealing compares to 'flat' annealing at the highest resolution 512×512. Initial convergence is slower (relative to temperature) as the hierarchical method roughly follows the profile of coarser resolution curves. After mapping to the highest resolution, extremely fast convergence shows the 'preconditioning' effect of the method.

Figure 10 shows examples of synthetic images resulting from this algorithm, along with representative samples of the training data used to define target pmfs for both cases.

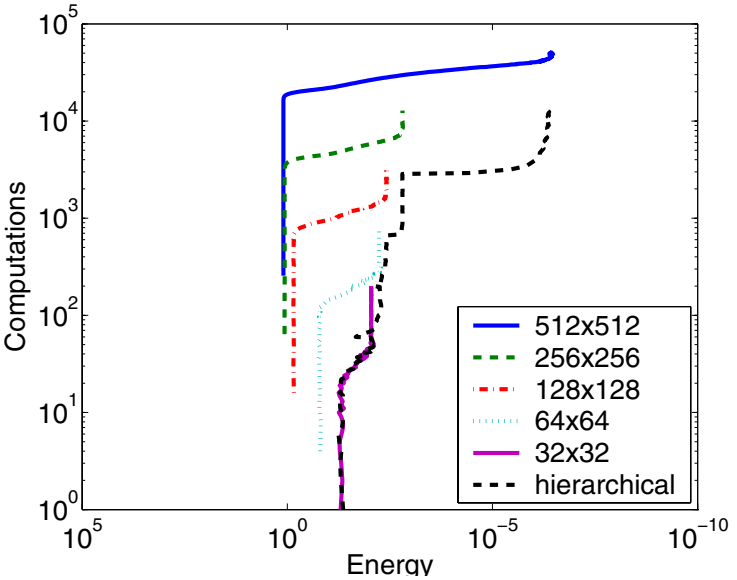


Fig. 8. Computation vs. energy for hierarchical and 'flat' annealing. Geometric cooling schedule with $\alpha = 0.9$. The target resolution is 512×512, and the hierarchy has 5 levels, with the coarsest level ($s = 4$) being 32×32. The two approaches converge to a similar minimal energy, however the hierarchical approach converges much more quickly. The repeated convergence/saturation of the energy is visible in the hierarchical curve, preceding shifts to the next scale

These images serve to demonstrate the capabilities (and limitations) of our simple model. Locally, the agreement seems qualitatively quite good; in terms of smooth edges, homogeneous regions, etc. Unsurprisingly, larger scale morphological features of the training data are *not* captured by our simple local model.

Computation times are very reasonable. The 512×512 samples shown here were generated in a few minutes on a 1Ghz PC workstation. By comparison, the synthesis of much smaller images has been reported to take on the order of 20–30 hours on an RS6000 workstation, by another stochastic relaxation approach [23].

5 Conclusions

This paper has described a method of hierarchical annealing that can result in large computational gains for appropriate models. A simple model for binary porous media images was presented to demonstrate its effectiveness. The heuristic approach to multiscale annealing given here is supported somewhat by these simple experiments, but may prove to be näive when generalizing to more complicated models. Future work must be performed with an aim to analyze

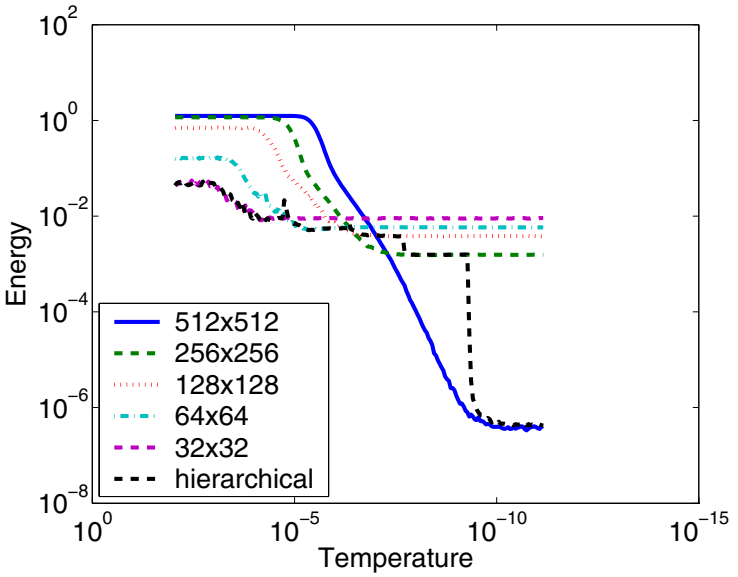


Fig. 9. Comparison of hierarchical annealing method with (flat) annealing at several resolutions. Hierarchical annealing results in highly accelerated convergence after mapping to the highest resolution

the complicated interaction between sampling, annealing, and refinement in the hierarchy.

Thus there are several directions open to future work in developing this approach. Of primary importance is analysis of the process for mapping to finer resolutions in the hierarchy. Other interesting questions abound; for example, convergence issues for the Markov chain samplers and relation to grid size, cooling schedule improvement, and introducing more complicated models as mentioned previously.

As illustrated in the previous section, the simplicity of this model restricts the class of appropriate images for direct application. It is interesting to consider the local features that are captured by this simple model, as well as the morphological differences that lie beyond the representation capability of the model.

Simple extensions to more complex models involving some non-local attributes such as chord-length [23] could prove to be quite effective, while still very computationally tractable. Extension to 3D data should be straightforward in principle, and yield even greater computational savings than seen in 2D.

In addition to refining this method, other applications are of interest. Constraining synthesis with particular data is one possibility, and could be extended to 3D reconstruction from 2D data.

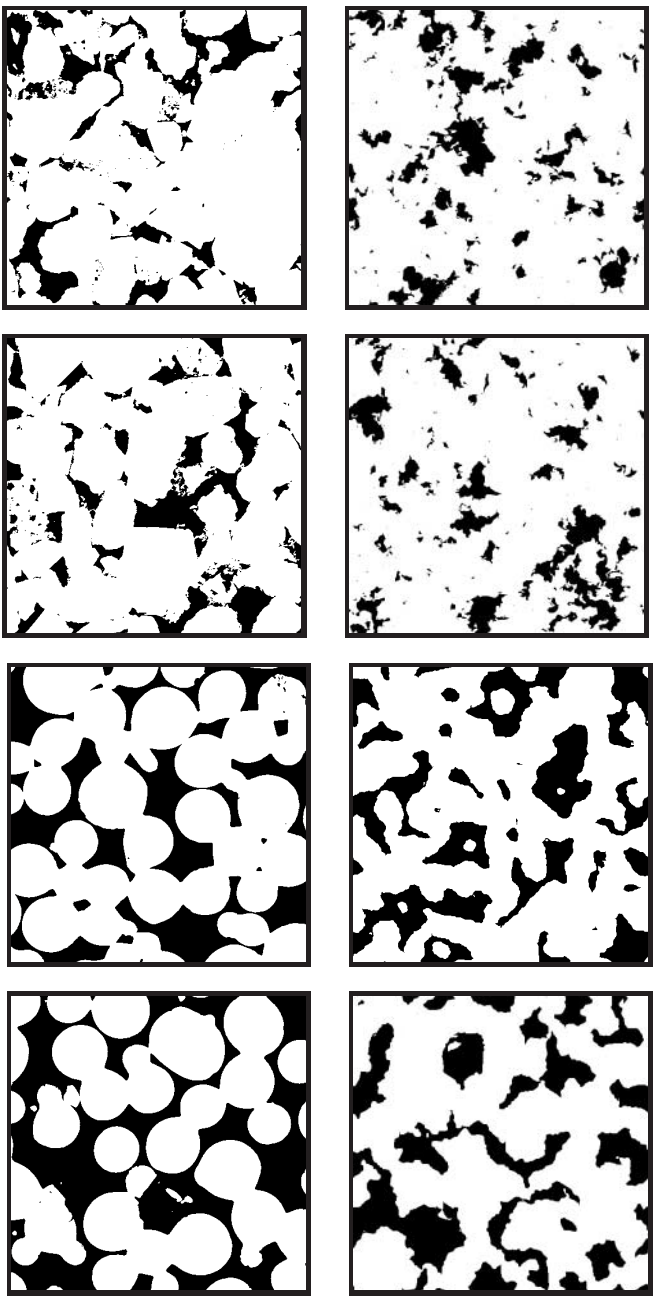


Fig. 10. Imaged (left column) and synthetic (right column) binary porous media images for two data sets; sandstone (top four) and packed glass spheres (bottom four)

References

[1] J. Besag, P. Green, D. Higdon, and K. Mengersen, *Bayesian computation and stochastic systems (with discussion)*, Statistical Science **10** (1995), 3–66. 198

[2] C. Bouman and M. Shapiro, *A multiscale random field model for Bayesian image segmentation*, IEEE Image Processing **3** (1994), no. 2, 162–177. 195

[3] Charles Bouman and Bede Liu, *Multiple resolution segmentation of textured images*, IEEE Transactions on Pattern Analysis and Machine Intelligence **13** (1991), no. 2, 99–113. 196

[4] J. Bramble, *Multigrid method*, John Wiley & Sons, 1993. 195

[5] Pierre Brémaud, *Markov chains: Gibbs fields, monte carlo simulation, and queues*, Springer, 1998. 197

[6] Davi Geiger and João E. Kogler Jr., *Scaling images and image features via the renormalization group*, Proceedings IEEE CVPR '93 (New York), June 1993. 197

[7] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Transactions on Pattern Analysis and Machine Intelligence **6** (1984), 721–741. 195, 196, 197, 198

[8] B. Gidas, *Nonstationary markov chains and convergence of the annealing algorithm*, Journal of Statistical Physics **39** (1985), 73–131. 195, 198

[9] Basilis Gidas, *A renormalization group approach to image processing problems*, IEEE Transactions on Pattern Analysis and Machine Intelligence **11** (1989), no. 2, 164–180. 196, 197, 201

[10] W. Hackbusch, *Multigrid methods and applications*, Computational Mathematics, vol. 4, Springer Verlag, 1985. 195

[11] F. Heitz, P. Perez, and P. Bouthemy, *Multiscale minimization of global energy functions in some visual recovery problems*, Computer Vision, Graphics, and Image Processing. Image Understanding **59** (1994), no. 1, 125–134. 196

[12] Zoltan Kato, Marc Berthod, and Josiane Zerubia, *A hierarchical Markov random field model and multitemperature annealing for parallel image classification*, Graphical Models and Image Processing **58** (1996), no. 1, 18–37. 199

[13] Z. Liang, M. A. Ioannidis, and I Chatzis, *Reconstruction of 3d porous media using simulated annealing*, Computational Methods in Water Resources XIII (Balkema, Rotterdam) (Bentley et al., ed.), 2000. 194

[14] M. Luettgen, W. Karl, A. Willsky, and R. Tenney, *Multiscale representations of Markov random fields*, IEEE Transactions on Signal Processing **41** (1993), no. 12, 3377–3395. 195

[15] S. McCormick, *Multigrid methods*, SIAM, Philadelphia, 1987. 195

[16] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H Teller, and E. Teller, *Equation of state calculations by fast computing machines*, Journal of Chemical Physics **21** (1953), 1087–1092. 197, 203

[17] R. Paget and I. D. Longstaff, *Texture synthesis via a noncausal nonparametric multiscale markov random field*, IEEE Transactions on Image Processing **7** (1998), no. 6, 925–931. 195

[18] Jan Puzicha and Joachim M. Buhmann, *Multiscale annealing for grouping and unsupervised texture segmentation*, Computer Vision and Image Understanding: CVIU **76** (1999), no. 3, 213–230. 195, 196

[19] Dietrich Stoyan, Wilfrid S. Kendall, and Joseph Mecke, *Stochastic geometry and its applications*, 2 ed., J. Wiley, 1996. 194

[20] Dietrich Stoyan and Helga Stoyan, *Fractals, random shapes and point fields: Methods of geometrical statistics*, J. Wiley, 1994. 194

[21] R. H. Swendson and J. S. Wang, *Nonuniversal critical dynamics in monte carlo simulations*, Physical Review Letters **58** (1987), 86–88. 198

[22] H. Szu and R. Hartley, *Fast simulated annealing*, Physics Letters A **122** (1987), 157–162. 198, 203

[23] M. S. Talukdar, O. Torsaeter, and M. A. Ionnidis, *Stochastic recontruction of particulate media from two-dimensional images*, Journal of Colloid and Interface Science **248** (2002), 419–428. 194, 195, 206, 207

[24] K. Wilson and J. Kogut, *The renormalization group and the ϵ-expansion*, Phys. Rep. **C12** (1974), 75–200. 197

[25] J. Zerubia, Z. Kato, and M. Berthod, *Multi-temperature annealing: a new approach for the energy-minimization of hierarchical markov random field models*, Proceedings of the 12th IAPR International Conference on Pattern Recognition (Jerusalem, Israel), vol. 1, 1994. 195, 196, 197