# Hierarchical MCMC Sampling

Paul Fieguth

Department of Systems Design Engineering
University of Waterloo
Waterloo, Ontario, Canada
pfieguth@uwaterloo.ca

**Abstract.** We maintain that the analysis and synthesis of random fields is much faster in a hierarchical setting. In particular, complicated long-range interactions at a fine scale become progressively more local (and therefore more efficient) at coarser levels. The key to effective coarse-scale activity is the proper model definition at those scales. This can be difficult for locally-coupled models such as Ising, but is inherent and easy for those models, commonly used in porous media, which express constraints in terms of lengths and areas.
Whereas past methods, using hierarchical random fields for image estimation and segmentation, saw only limited improvements, we find reductions in computational complexity of two or more orders of magnitude, enabling the investigation of models at much greater sizes and resolutions.

**Keywords:** Posterior sampling, MCMC methods, Hierarchical sampling, Porous media, Ising, Random Fields, Energy minimization

## 1 Introduction

The cure to arthritis and collapsing buildings lies in the fast random sampling of large images!

A trifle optimistic, to be sure, however drug delivery in cartilage and the cracking of concrete both rely on a detailed understanding of *porous media* [8, 10], and thus a corresponding need to model, generate, and manipulate large stochastic 2D images and 3D volumes.

As motivated by Figure 1, we seek hierarchical approaches to modelling and computation, specifically for two reasons: first, for those media which are inherently multi-scale (concrete, for example, has pore sizes ranging from sub-micron to millimetre) and, secondly, to more effectively model those non-local relationships on the finest scale, but which become progressively more local (and simple) on coarser levels.

To be sure, a variety of hierarchical [1,2,5,6,7] and region-based [9] methods exist, however they differ from our current context in a few ways:

- Most methods, certainly among those in the image processing literature, are designed for estimation [1,2,6,7] (thus guided/driven by measurements), not random sampling (purely model based).

A. Campilho, M. Kamel (Eds.): ICIAR 2004, LNCS 3211, pp. 114–121, 2004.

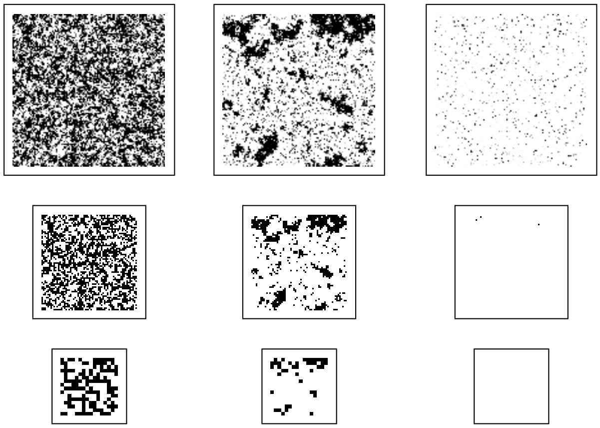


**Fig. 1.** The analysis and synthesis of random fields is easier and faster in a hierarchical setting. For Ising models, weakly and strongly coupled models become progressively random and uniform, respectively, at coarser scales [11]. Although critically-coupled structures do not simplify with scale, there are computational benefits: the long-range interactions implied by a large cluster at a fine scale become progressively more local (and therefore more efficient).

– In most cases there is a certain ambiguity or arbitrariness in the selection of coarse-scale models.

In many cases, the coarse scales served only as a weak regularizer for a densely-measured, well-conditioned fine-scale estimation problem, and led to only marginal improvements. We are interested in problems involving sparse or no measurements, in which case the finest scale is very poorly conditioned, and where the coarse scales, if appropriately defined, have a great deal to offer. In the following three sections will examine three different models which, although greatly simplified, are representative of models used in studies of porous media. A hierarchical approach will, in each case, lead to computational improvements at or exceeding two orders of magnitude.

## 2 Ising Model

We first look at the well-known Ising model [3] — very widely studied — in which the elements of a binary field $x$ interact with their immediate neighbours:

$$E = \beta \sum_{ij} x_{i,j} x_{i+1,j} + \beta \sum_{ij} x_{i,j} x_{i,j+1}. \tag{1}$$

The coupling $\beta$ controls the degree to which adjacent $x_{ij}$ should be the same, and as $\beta$ increases so does the inter-pixel correlation (Figure 3).

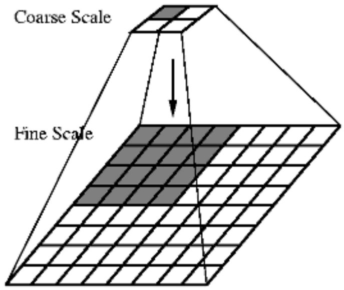


**Fig. 2.** A standard model of hierarchical random fields: to evaluate the energy of a coarse-scale field, we project it to the finest scale and evaluate the energy there.

For small $\beta$ the sampled field is essentially random, and all samplers, whether flat or hierarchical, converge quickly. However as $\beta$ increases the structures present in $x$ grow larger, and the longer-range relationships become increasingly difficult to deduce from a local model, thus hierarchical methods begin to outperform (Figure 4). The hierarchical samplers proceed from coarse to fine, in a single pass, sampling for some iterations at each scale.

The key challenge relates to the definition of coarse-scale models. The coarsification of (1) is not obvious [5], and is most often circumvented by defining coarse-scale models implicitly in terms of the finest-scale by projection [6] (Figure 2). In the Ising model this implies the widely-used model $\beta_s = 2^s\beta$ at $s$ scales above the finest.

The problem is that this is wrong. Figure 3 makes it clear that for small $\beta$ the coupling should *decrease* with scale. Using $\beta_s = 2^s\beta$ leads to stiff, large-scale structures created on coarse scales which then need to be undone at finer scales. If a properly-renormalized model (here derived experimentally) is used, with the correct value of $\beta$ at each scale, then the sampler at each scale needs only to insert the details unresolvable at the coarser scale, a *much* easier task than undoing incorrect structure, and thus converging *far* faster, as seen in Figure 4.

The key is the production of properly renormalized coarse-scale models.

## 3 Correlation Model

Although effective as an illustration, the familiar Ising model is arguably not a good representation of cartilage, concrete, or other porous media, nor was the Ising coarse-scale model easily found.

The key difficulty was to infer, from a local, interpixel model such as Ising, the interpixel relationships on coarser scales. Instead, common stochastic porous media models [10] often involve concepts of correlation, chord-length, or area — all of which are *nonlocal* constraints on the finest scale, and thus rescale almost trivially, since it is relatively easy to express a correlation, for example, between coarse pixels, on the basis of a stipulated correlation model.

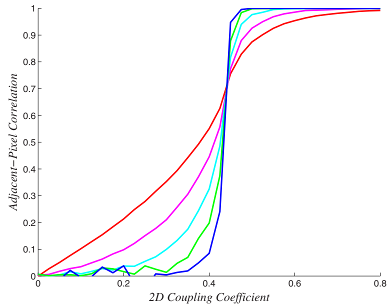


**Fig. 3.** The correlation between adjacent pixels as a function of the coupling $\beta$ and scale. Coarser scales (blue) are progressively random or uniform, depending on the value of $\beta$ relative to criticality (at $\beta \approx 0.44$).

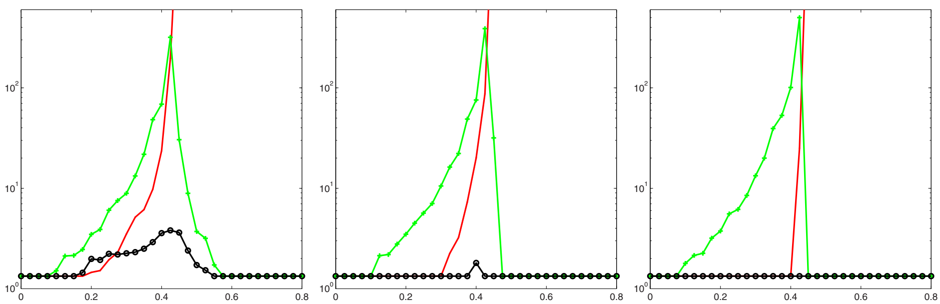


**Fig. 4.** The convergence of a flat (red), standard hierarchy (green), and renormalized hierarchical model (black) as a function of coupling, for three different scales. Around criticality, where structures are most complex, the properly renormalized model converges more than 100 times faster.

Consider, for example, the correlation structure of Figure 5: we seek regions on the order of 10 to 20 pixels in size. We can sample from such a model by computing the empirical correlation of a sample $x$, and accepting/rejecting a pixel change on the basis of the degree of improvement in the empirical correlation toward the ideal. The model rescales inherently, since a coarsification of $x$ is simply equivalent to rescaling the horizontal axis of the correlation plot in Figure 5.

Since the chosen correlation of Figure 5 is arbitrary, the pictures which result are not of any particular importance, however the convergence results of Figure 6 are significant, where the degree of fit between a desired correlation $c(d)$ and an average, empirical correlation $\hat{c}(d)$ is chosen to be

$$\|c - \hat{c}\| = \sum_{d} \frac{|c(d) - \hat{c}(d)|}{d} \qquad (2)$$

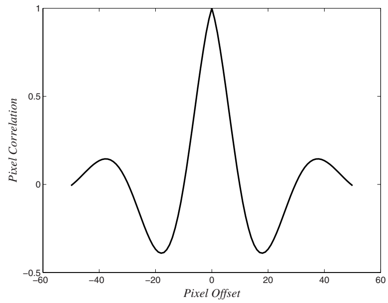


**Fig. 5.** The asserted correlation function: we seek random fields which are correlated out to a typical size of 10 pixels, and negatively correlated beyond that, out to roughly 30 pixels. The model rescales trivially, since this involves only an axis rescaling.

where $d$ is the spatial offset, measured in pixels, and where division by $d$ de-emphasizes correlation at long ranges.

When the coupling between sample field $x$ and the desired correlation is weak, then $x$ is essentially random and local, devoid of large-scale structure, and is easily sampled by a flat model. However as the coupling increases, large-scale structures begin to appear, and the flat-model convergence time increases rapidly. This slowing down is well understood; in the Ising case this is a critical phenomenon, related to the random-walk nature of information propagation in the pixellated lattice, and where the number of iterations to produce large regions grows very rapidly with region size. In this correlation model, the walk is not quite random, however the problem is analogous: the criterion at the pixel level does not provide strong guidance in the formation of large-scale regions, and so a random sampler wastes a great deal of time on trial-and-error before finding improved solutions.

With strong coupling, since the sample fields consist of relatively large regions, much of the desired structure can be realized at coarse scales where there are fewer pixels, where the regions are smaller, and where iterations proceed much more rapidly. As before, only the details of the region boundaries remain to be refined at finer scales, rather than the induction of entire regions at the finest scale of a flat model, leading to one or two orders of magnitude improvement.

## 4 Multi-scale Porous Media

The previous section introduced a correlation model, which scales inherently, however most of the structure lives on a single scale (roughly 10-20 pixels in size). A persuasive example of a porous medium needs to be truly multi-scale in nature, containing both large and small structure.

A final example is based on the criteria of Figure 7: we permit tiny regions (0% – 1% of domain), embedded in larger regions (5% – 20% of domain), in

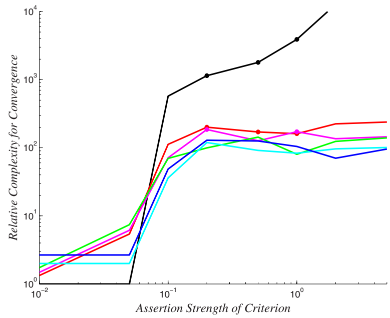


**Fig. 6.** Convergence time as a function of coupling strength. For weakly-coupled (nearly random) models, left, a flat sampler is ideal, however synthesizing larger structures becomes much more difficult. Asterisks (*) denote extrapolated convergence time for cases which failed to converge. Colour indicates change in iterations with scale (Red: fewer iterations, Blue: more iterations, Black: flat model, no iterations at coarse scales).

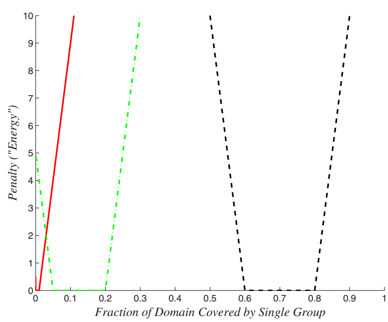


**Fig. 7.** Criteria for a true, multi-scale porous-media model: the lines plot penalty as a function of region size. The model consists of small pores (red) embedded in large regions (green) embedded in a background medium.

turn embedded in a background. To prevent the growth of long, tendril-like regions, there is a constraint limiting region shape (by penalizing the ratio of region perimeter to square-root of area). The model is trivially renormalized across scale, since all parameters are expressed in terms of lengths and areas. The model is also binary: the tiny pores and the background are of the same material.

The random sampling results, shown in Figure 8, follow a pattern by now familiar. For weakly-constrained problems (Figure 8 left), where the sample is mostly random, a flat sampler performs well. However as the constraints increase (right), and larger regions begin to appear the flat sampler fails completely, whereas the hierarchical sampler begins to produce very interesting, very credible samples, involving a wide range of structural scales.

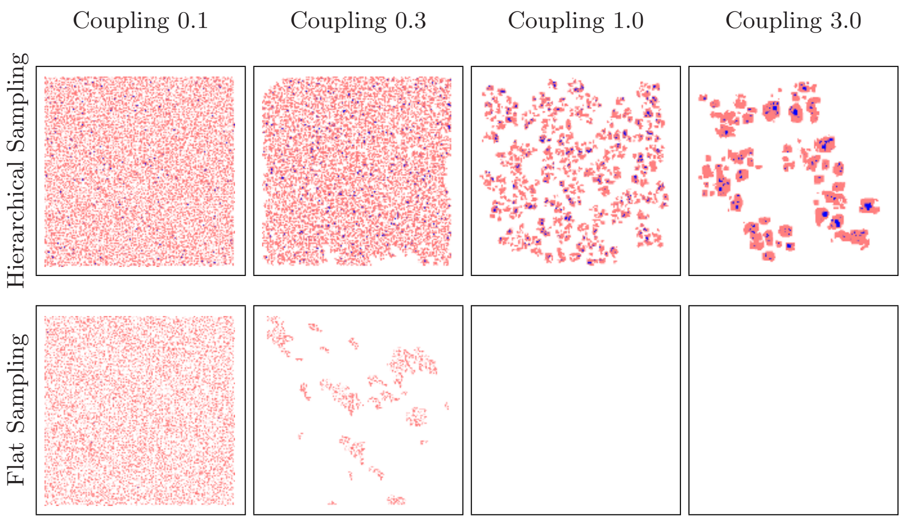


**Fig. 8.** Random samples from hierarchical and flat approaches, for four coupling strengths. Consistent with other experiments, the differences are minimal for weakly-coupled models (left), but are striking in strongly-coupled cases (right): the flat model fails to be able to synthesize large-scale structure, because of an inhibition on tiny regions. All cases were initialized with a flat, white field. Initializing with a random field had no effect on the hierarchical case, and produced very different, although no better, results for flat sampling.

In this case the computational benefit of the hierarchical approach is unmeasurable (that is, almost infinite), since the time constant to convergence for the flat sampler diverges, and may be infinite. The failure of flat sampling is easy to diagnose: there are strong barriers inhibiting the creation of small foreground regions within a uniform background. Initializing with a white field does not allow the production of local regions; initializing with random pixels does not allow a background to form. Only by initializing with a credible solution (small regions within larger regions on a background) can a flat sampler converge. A critic could charge that a redesign of the criteria in Figure 7 could solve this problem, however this leaves us with a flat sampler, sensitive to minor perturbations in the energy function, and still converging orders of magnitude slower than a hierarchical sampler.

## 5 Conclusions

It may be argued that flat samplers are not meant to perform well on strongly-coupled fields, that such problems are meant to be solved by annealing[4]. Indeed, our research in hierarchical sampling is driven by a long term interest in hierarchical annealing. However, we maintain that the best annealer is the one built around the best, fastest, and most robust sampler.

Although hierarchical annealing and sampling is not new, flat sampling and annealing methods are still widely practiced, certainly for porous media. In this paper we have clearly shown the degree of improvement available with hierarchical approaches, with only very modest algorithmic changes, and the importance of properly renormalizable or rescalable models. The improvement in computational complexity is very clearly seen to be the synthesis of large-scale structures at coarse scales, with only local details remaining to be refined at finer scales.

Motivated by the clear successes in this paper, our research foci are twofold: one, the development of less arbitrary or contrived models, more physically meaningful for a particular porous-media context, and secondly the development of annealing techniques and appropriate temperature schedules built around hierarchical samplers.

## References


1. 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.
2. C. Bouman and M. Shapiro, *A multiscale random field model for Bayesian image segmentation*, IEEE Image Processing **3** (1994), no. 2, 162–177.
3. D. Chandler, *Introduction to Modern Statistical Mechanics*, Oxford University Press, 1987.
4. 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.
5. Basilis Gidas, *A renormalization group approach to image processing problems*, IEEE Trans. PAMI **11** (1989), no. 2, 164–180.
6. Z. Kato, M. Berthod, and J. Zerubia, *A hierarchical Markov random field model* . . . , GMIP **58** (1996)
7. Jan Puzicha and Joachim M. Buhmann, *Multiscale annealing for grouping and unsupervised texture segmentation*, CVIU **76** (1999), no. 3, 213–230.
8. Dietrich Stoyan and Helga Stoyan, *Fractals, random shapes and point fields*, J. Wiley, 1994.
9. R. H. Swendson and J. S. Wang, *Nonuniversal critical dynamics in Monte Carlo simulations*, Physical Review Letters **58** (1987), 86–88.
10. M. S. Talukdar, O. Torsaeter, and M. A. Ionnidis, *Stochastic reconstruction of particulate media from two-dimensional images*, Journal of Colloid and Interface Science **248** (2002), 419–428.
11. K. Wilson and J. Kogut, *The renormalization group and the* $\epsilon$*-expansion*, Phys. Rep. **C12** (1974), 75–200.