5-1-2010

# Markov random field modeling of the spatial distribution of proteins on cell membranes

Jun Zhang

## Recommended Citation

Zhang, Jun. "Markov random field modeling of the spatial distribution of proteins on cell membranes." (2010). https://digitalrepository.unm.edu/cs_etds/11

Jun Zhang
*Candidate*

Computer Science
*Department*

This dissertation is approved, and it is acceptable in quality and form for publication:

*Approved by the Dissertation Committee:*

3/5/10, Chairperson

Stanley Sternberg 3/5/10

03/05/2010

MAR 5, 2010

Bridget Wilson 3/5/2010

# Markov Random Field Modeling of the Spatial Distribution of Proteins on Cell Membranes

**by**

**JUN ZHANG**

B.S., Chemistry, Peking University, 1992

M.S., Chemistry, University of New Mexico, 1999

M.S., Computer Science, University of New Mexico, 2001

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

**Doctor of Philosophy**
**Computer Science**

The University of New Mexico
Albuquerque, New Mexico

**May 2010**

# Dedication

*To my wife, Lu Li, for her patience, support, and encouragement.*

# Acknowledgments

I would like to thank my Ph.D. advisor, Dr. Lance Williams, for his enthusiasm, inspiration, and great efforts to explain things clearly and simply. He helped to make mathematics fun for me. I would like to thank other members of my committee, Dr. Terran Lane, Dr. Shuang Luan, Dr. Stanly Steinberg, and Dr. Bridget Wilson, for their intellectual advice, insightful discussion, and enthusiastic support. I would never have been able to finish my dissertation without the guidance of my committee members.

I am grateful to Dr. Janet Oliver for her advice, encouragement, and support. I wish to thank Janet Pfeiffer for the experimental data used in the research. The financial support by a STMC/Geis Fellowship from UNM Cancer Research and Treatment Center is gratefully acknowledged.

Finally, I am forever indebted to my parents and my wife for their understanding, endless patience, and encouragement when it was most required.

# Markov Random Field Modeling of the Spatial Distribution of Proteins on Cell Membranes

**by**

**JUN ZHANG**

ABSTRACT OF DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

**Doctor of Philosophy**
**Computer Science**

The University of New Mexico
Albuquerque, New Mexico

**May 2010**

# MARKOV RANDOM FIELD MODELING OF THE SPATIAL DISTRIBUTION OF PROTEINS ON CELL MEMBRANES

**by**

**JUN ZHANG**

B.S., Chemistry, Peking University, 1992
M.S., Chemistry, University of New Mexico, 1999
M.S., Computer Science, University of New Mexico, 2001

Ph.D., Computer Science, University of New Mexico, 2010

## Abstract

Cell membranes display a range of receptors that bind ligands and activate signaling pathways. Signaling is characterized by dramatic changes in membrane molecular topography, including the co-clustering of receptors with signaling molecules and the segregation of other signaling molecules away from receptors. Electron microscopy of immunogold-labeled membranes is a critical technique to generate topographical information at the 5-10 nm resolution needed to understand how signaling complexes assemble and function. However, due to experimental limitations, only two molecular species can usually be labeled at a time. A formidable challenge is to integrate experimental data across multiple experiments where there are from 10 to 100 different proteins and lipids of interest and only the positions of two species can be observed simultaneously. As a solution, Markov random field (MRF) modeling is proposed to reconstruct the distribution of multiple cell membrane constituents from pair-wise data sets. MRFs are a powerful mathematical formalism for modeling correlations between states associated with neighboring sites in spatial lattices. The presence or absence of a protein of a specific type at a point on the cell membrane is a state. Since only two protein types can be observed, *i.e.*, those bound to particles, and the rest cannot be observed, the problem is one of deducing the conditional distribution of a MRF with unobservable (hidden) states. Here, a multiscale MRF model has been developed and mathematical programming techniques have been used to infer the conditional distribution of a MRF for proteins of three types from observations showing the spatial relationships between only two types. Application to synthesized data shows that the spatial distributions of three proteins can be reliably estimated. Application to experimental data provides the first maps of the spatial relationship between groups of three different signaling molecules. Initially, a 4-neighborhood system was used in the MRF modeling. In order to improve reconstruction quality, a larger 8-neighborhood system was subsequently used in a multiscale Gibbs random field (GRF) formulation by exploiting the Markov-Gibbs equivalence. Application of the multiscale GRF model to synthesized and experimental data shows that the quality of reconstruction is improved. This work is an important step towards a more complete understanding of membrane spatial organization and dynamics during signaling.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Cell membranes display a range of receptors that bind signaling molecules and initiate transmembrane responses. Receptors and the signaling proteins and lipids they activate are distributed non-randomly in membranes; in addition, receptor activation is accompanied by dramatic reorganization of membrane components as well as by recruitment of new signaling proteins from the cytoplasm to the membrane [42, 44]. Strict regulation of signal transduction from the outer cell surface to the cytoplasm and nucleus is crucial for cell survival, differentiation, proliferation and other activities. Unregulated signaling is an important component in the pathogenesis of many diseases, including cancer. Nevertheless, many aspects of how the cell maintains spatio-temporal control of signaling pathways remain unclear. Correlating the activities of receptors and signaling proteins and lipids with their spatial distribution and dynamics is essential to better understand the regulation of cell signaling.

To observe the topographical events associated with cell signaling, several groups have generated high resolution spatial maps of colloidal gold particles marking receptors and signaling proteins and lipids in native membranes [22, 33, 34, 40, 42, 43, 44, 45, 47]. In this technique, cell monolayers adherent to glass coverslips are inverted onto nickel electron microscopy grids that are precoated with a highly charged polymer, poly-L-lysine. Pressure is applied briefly, and then the coverslips are lifted, leaving cytoplasmic face-up sheets of native membranes (see Figure 1.1). The sheets are fixed lightly, and labeled with nanometer-sized electron-dense probes that have been functionalized, usually with antibodies, to recognize specific signaling molecules. After labeling, membranes are fixed more completely, counterstained with osmium and uranyl acetate to label features of the cell membrane like coated pits, air-dried, and imaged using transmission electron microscopy (TEM). The TEM images are digitized and probe coordinates are extracted. The spatial distributions of the probes with respect to each other and with respect to membrane features such as clathrin-coated pits and caveolae are subsequently analyzed [17, 30, 31, 49]. In general, due to limitations of applicable gold particle size, only two different protein species can be labeled with confidence in the same experiment (Figure 1.2). Experimentalists are exploring the use of new metal, semi-conductor, and ceramic electron-dense nanoprobes with different shapes to expand the number of probes than can be discriminated [1, 18]. However, even with these new tools, the limited availability of antibodies raised in different species to label signal pathway components makes it difficult to substantially expand the number of probes that can be used in a single experiment. Consequently, there is a need to integrate experimental data across multiple experiments. In this dissertation, Markov random field (MRF) modeling is used to address this problem.

MRFs provide a powerful and robust framework for modeling correlations between states associated with neighboring sites in spatial lattices. The first concept of MRF came from the physicist Ernst Ising in the 1920s [23]. Ising tried to devise a mathematical model to explain the experimental properties of ferromagnetic materials. In his model, Ising made the simplifying assumption that only interactions between neighboring bipoles need be taken into account. Since the 1980s, MRFs have been become useful in several areas of image analysis such as texture synthesis [9, 12, 32], image restoration [4, 14], image segmentation [11] and surface reconstruction [13]. The success of MRFs can be attributed

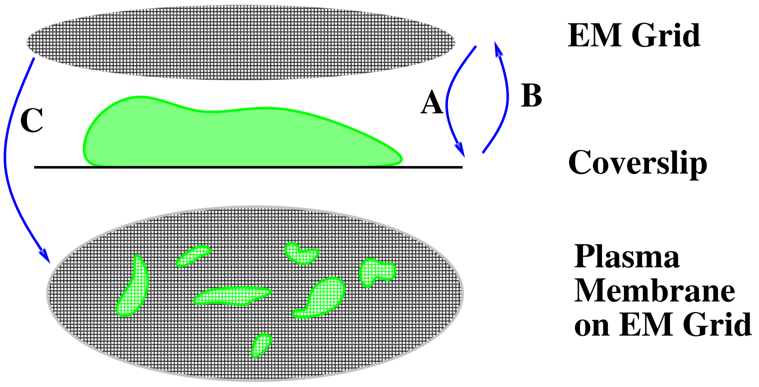

Figure 1.1: a) Lower the EM grid onto the cell on the coverslip; b) remove the EM grid; c) produce fragments of the plasma membrane on the grid.

to their versatility as stochastic image models and to the fact that the Hammersley-Clifford theorem [3] provides a simple way of specifying the joint probability distribution.

The goal of image modeling is to find a good representation of the gray level distribution of a given image. What is good often depends on the task at hand. In most of the work on MRF modeling in image processing, a key assumption is that the information contained in the local structure of images is sufficient to obtain a good image representation. The local information is captured by a conditional probability distribution because the image gray level at a particular site depends only on its neighboring pixels. There are usually two ingredients in the MRF image modeling: the *prior* and *observation* process. The prior process, $X$, is a model of the unobserved original image. $X$ is defined on the set of image attributes that are of interest. For example, in edge preserving image restoration, $X$ includes image gray levels and binary-valued edge labels. In texture segmentation, $X$ includes image gray levels and texture labels. The observation process, $Y$, is a model of the given image which is a noisy, blurred, or incomplete version of the original image. The MRF modeling fits well into a Bayesian estimation/inference paradigm, where Bayes' theorem is invoked to obtain the posterior probability distribution of the original image given the observation, and where a form of Bayesian estimation, maximum *a posteriori* estima-

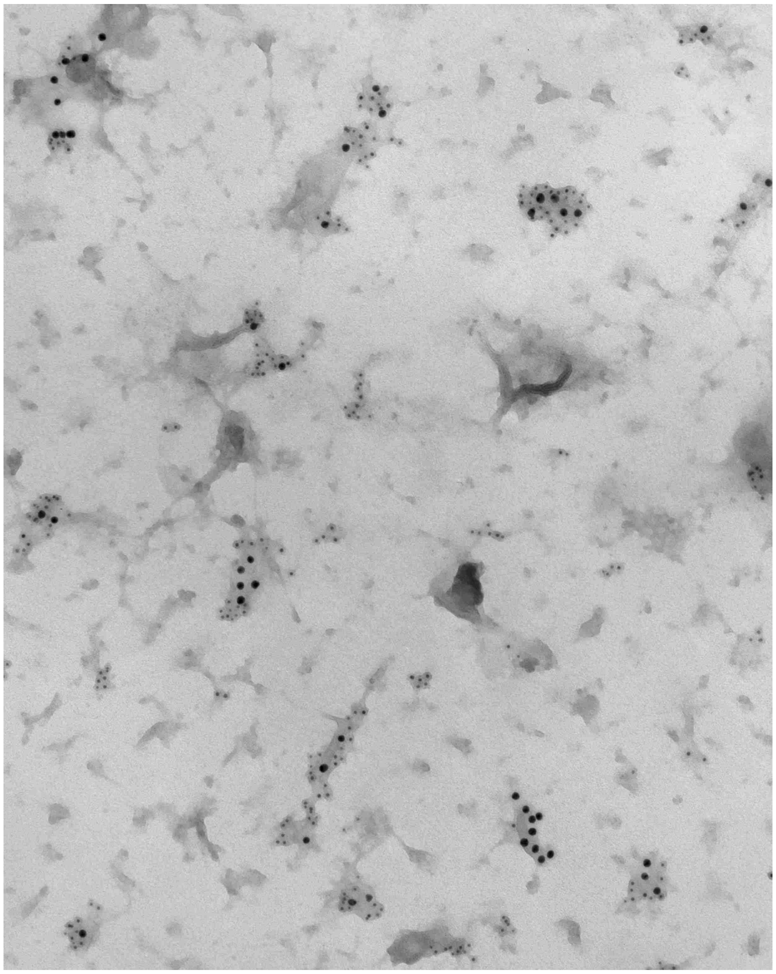

Figure 1.2: 5nm and 10nm gold probes in the image of a TEM slide.

tion, is used to restore or to segment images.

The spatial distributions of proteins on cell membranes are determined by the underlying physical and chemical interactions, including protein-protein and protein-lipid interactions. Usually, these interactions decrease quickly as distance increases. In other words, these interactions are local. Therefore, local characteristics are a good representation of protein spatial distributions which are determined by local interactions. As a result, MRFs are a good choice to characterize protein spatial distributions. Cell membranes can be modeled as a 2D lattice, and the presence or absence of a protein of a specific type at a point on the cell membrane is a state. Since only two protein types can be observed, *i.e.*, those bound to particles, and the rest cannot be observed, the problem is one of deducing the conditional probability distribution of a MRF with unobservable (hidden) states. A question of significant importance is: What fraction of the conditional probability distribution of the MRF modeling spatial relationships between proteins of all types can be inferred from particle preparations showing the spatial relationships between only two types? If the conditional probability distribution can be reliably estimated, then the Gibbs sampler [14] can be used to synthesize sample MRFs allowing the complete set of protein spatial relationships to be visualized.

The organization of this dissertation is as follows. Chapter 2 gives a brief introduction to MRF theory and Bayesian image analysis. It also reviews related work. In Chapter 3, a quadruple stochastic model for modeling the protein spatial distributions is first described, and then first-order and second-order Markov chains are used to demonstrate that the quadruple model is feasible. It then presents how the conditional probability distribution of a MRF for proteins of three types can be deduced from observations showing the spatial relationships between only two types. However, the neighborhood system of the MRF used in Chapter 3 is too small to obtain good results on experimental data. To solve this problem, a multiscale MRF model is developed in Chapter 4. Application of the multiscale MRF model to synthesized and experimental data gives satisfactory results. In Chapter 5, the Markov-Gibbs equivalence is exploited to utilize a larger neighborhood system in the modeling. Results of applying the GRF models to the synthesized and experimental data are then presented. Finally, Chapter 6 concludes and discusses future research directions.

# Chapter 2

# Background and Related Work

## 2.1 Markov Random Field

MRFs have several components: a lattice $S$ with $m$ sites $s$; a neighborhood system $\mathcal{N} = \{\mathcal{N}_s | s \in S\}$, where $\mathcal{N}_s$ is the subset of sites in $S$ which are the neighbors of $s$; a field of random variables $X = \{X_s | s \in S\}$, and a conditional probability mass function (p.m.f.), $P(X_s = x_s | X_t = x_t, t \in \mathcal{N}_s)$. Each random variable $X_s$ takes a value in a finite set $Q = \{l_1, \ldots, l_q\}$ of possible states. $X_s = x_s$ denotes the event that $X_s$ takes the value $x_s$ and $(X_1 = x_1, X_2 = x_2, \cdots, X_m = x_m)$ denotes a joint event. The joint event is abbreviated as $X = x$ in which $x = \{x_1, x_2, \cdots, x_m\}$ is a realization of $X$. Therefore, there is also a joint p.m.f., $P(X = x)$. Either the conditional p.m.f. or the joint p.m.f. can be used to specify a MRF. The Markov property means that the state at a site is dependent only on those at its neighboring sites:

$$P(X_s = x_s | X_t = x_t, t \neq s, t \in S) = P(X_s = x_s | X_t = x_t, t \in \mathcal{N}_s). \tag{2.1}$$

For example, the neighbor set of $s = (i, j)$ for a regular lattice $S$ is commonly defined as

$$\mathcal{N}_s = \{r = (k, l) \in S : 0 < (k - i)^2 + (l - j)^2 \leq o\}, \tag{2.2}$$

where $o$ is the order of the neighborhood system. Figure 2.1 a)-c) show the neighborhood systems for $o = 1$, $2$ and $8$. Alternatively, a neighborhood system can also be represented by the number of neighbors in it. For example, a first-order neighborhood can also be called a 4-neighborhood.

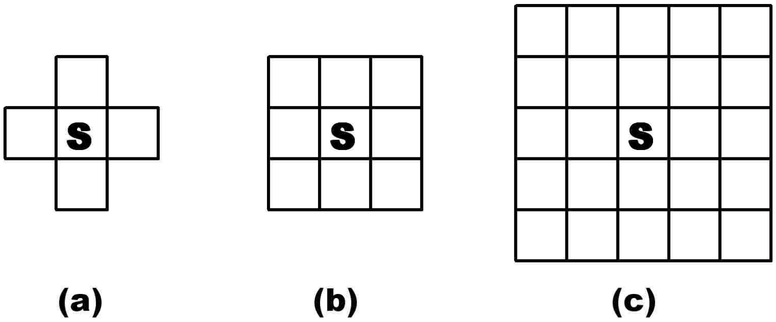

Figure 2.1: MRF neighborhood systems for site $s \in S$: a) 4-neighborhood; b) 8-neighborhood; and c) 24-neighborhood [26].

MRFs are often formulated as *Gibbs Random Fields* (GRFs), which are equivalent to MRFs by the Hammersley-Clifford theorem [3]. In Section 5.1, we give a brief introduction to Gibbs random fields.

## 2.2 Bayesian Image Analysis

When MRFs are used to model images, Bayesian methods are often used in image analysis [19, 14, 11]. The Bayesian approach has benefits in image analysis and interpretation because it permits the use of prior knowledge concerning the situation under study. Here, a brief introduction to Bayesian image analysis is given.

### 2.2.1 Bayes' Theorem

Bayesian methods are based on Bayes' theorem (also known as Bayes' rule or Bayes' law), which defines the relationship between the conditional and marginal probability distributions of two random variables. Let us consider the basic set-up in the context of image restoration, an original image described by a random process $X$, and a given image represented by another random process $Y$. The original image is not observable; it can only be observed through the given image which is a noisy or blurred version of the original. We want to restore the original image from the given image. Bayes' theorem relates their conditional and marginal probabilities as follows:

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}. \tag{2.3}$$

$P(X)$ is the *prior* probability of $X$ because it does not take into account any information about $Y$. $P(X|Y)$ is called the *posterior* probability because it depends on the specified value of $Y$. $P(Y|X)$ is called the *likelihood* of $Y$ with respect to $X$, indicating that the value $x$ for which $P(Y = y|X = x)$ is large is more "likely" to be the true value. $P(Y)$ is the *evidence* factor and can be viewed as a normalizing constant. With this terminology, the Bayes' theorem can be informally expressed as:

$$posterior = \frac{likelihood \times prior}{evidence}. \tag{2.4}$$

In words, the posterior probability is proportional to the product of the prior probability and the likelihood. Bayes's theorem provides the means that we can update or revise our knowledge of $X$ in light of new information of $Y$, both expressed in terms of a probability distribution. For more detailed materials on Bayesian theory, the reader is referred to [2].

### 2.2.2 Bayesian Analysis

The first aspect of Bayesian analysis involves the choice of the prior. In one interpretation of Bayes' theorem, the prior $P(X)$ represents general information about the process $X$, *i.e.*, the prior might be viewed as a way to restrict $X$ so that the posterior provides more information about $X$ than the likelihood. In this case, many different priors can be used in the analysis to investigate the range of possible outcomes and the proper choice for the prior depends on the problem domain. In Bayesian image analysis, MRFs are often used as prior models.

The second aspect of Bayesian analysis deals with the use of the posterior probability. When attempting to summarize the results of an analysis, it might be necessary to represent $P(X|Y)$ in more concise terms. In the process of interpretation, some information concerning $P(X|Y)$ is lost. To achieve an optimal way to interpret the posterior probability, the Bayesian approach associates the costs with making various kinds of errors in the interpretation process. The assignment of the proper cost function is usually a part of specifying the problem. When any answer other than the correct one has the same increased cost, the estimate is the value of $X$ that maximizes the posterior probability distribution:

$$\hat{X} = \mathrm{argmax}\ \ P(X|Y). \tag{2.5}$$

This is the well-known maximum *a posteriori* (MAP) estimation. Since the paper of Geman and Geman [14], many image analysis problems have been formulated in this MRF-MAP framework.

### 2.2.3 Gibbs Sampler

Gibbs sampling is a method for generating a sequence of samples from the joint probability distribution of two or more random variables. The associated algorithm, called the Gibbs sampler, was devised by Geman and Geman [14]. Gibbs sampling is applicable when the joint distribution is not known explicitly, but the conditional distribution of each variable is known. It operates by generating a realization from the distribution of each random variable in turn, conditional on the current values of neighboring random variables. Geman and Geman showed that the sequence of samples comprises a Markov chain, and the stationary distribution of the Markov chain is the designed joint distribution. Gibbs sampling is well-adapted to sampling the joint distribution of MRFs, because MRFs are typically specified as conditional probability distributions.

## 2.3 Related Work

### 2.3.1 Non-hierarchical MRF Models

#### Texture Models

MRFs have been used in computer vision and image processing for texture synthesis [9, 12, 32], image segmentation [11], and image restoration [14, 26]. In image processing, a digital image is defined as a function of a two-dimensional lattice. The value at a site in the lattice is called the gray level of the image at that point. Texture is a spatially extended, statistically homogeneous pattern of gray levels. If we assume that the gray level of a pixel only depends on those of its neighbors, textures can be modeled as MRFs. The states of the MRF texture model are all possible gray levels. A sample texture is regarded as a realization of the MRF model and is used to estimate the conditional distribution of the model through either parametric [9] or non-parametric methods [12, 32]. Texture can be synthesized by sampling from the conditional distribution. Large neighborhood systems are required to capture the visual characteristics of a natural texture, *e.g.*, 24-neighborhood and 48-neighborhood systems were used in [32] (see Figure 2.2).

#### Image Restoration Models

In addition to MRFs where all states are observable, there are also hidden Markov models which contain states that are not directly observable. These hidden models are flexible and powerful when used in applications such as image restoration and segmentation, because many kinds of prior knowledge can be modeled. For example, Besag [4] used a hidden MRF to model the local characteristics of unobserved sample images, assuming that pixels close together tend to have the same or similar gray levels. As a second example, Geman and Geman [14] used an *intensity process* $F$ to characterize the variations of gray levels, and a *line process* $L$ to characterize the presence and orientation of edges. The original image was modeled as a pair of hidden Markov processes, $X = (F, L)$. $F$ is defined over a lattice where each site corresponds to one image pixel, and $L$ is defined over a dual lattice with sites centered between each vertical and horizontal pair of pixels (Figure 2.3). The sites of $L$ represent possible edges. The given image is a realization of an observable process, $Y$, which is modeled as a blurred and noisy version of the original image:

$$Y = \Psi(\mathrm{H}(F), N)$$

where $\Psi$ is an invertible operation, $\mathrm{H}$ is a shift-invariant blurring matrix and $N$ is an independent noise field. The addition of the line process results in good performance for finding the boundaries of arbitrary shapes. The pair of intensity and line processes is flexible, for example, Tonazzini *et al.* [38] used it to solve a blind image separation problem.

### Image Segmentation Models

In image segmentation, hidden MRFs can be used to model regions where pixels have similar gray values or are statistically homogeneous. One example is Derin and Elliott's two-level hierarchical GRF, which they used to model and segment noisy and textured images [11]. Derin and Elliott's model was based on a *region process*, $X$ (the high level process characterizing formation of homogeneous regions), and a *texture process*, $T$ (the low level process modeling textures in different regions). A noisy or textured image was modeled using an observable random field, $Y$, which is defined in terms of hidden random fields, $X$ and $T$. Derin and Elliott adopted a specific class of GRF to model both the region process and the texture process. The GRF class specifies the distribution in terms of the 8-neighborhood system, and assigns one parameter to clique types with two or more sites. The clique potentials are then defined as

$$V_C(x) = \begin{cases} -\zeta & \text{if all } x_s \text{ in } C \text{ are equal} \\ \zeta & \text{otherwise} \end{cases}$$

where $\zeta$ is the parameter assigned for the clique type $C$. And the potential for the single site clique is defined as $V_C(x) = \alpha_i$ for $x_s = i$ where $i$ represents one of the $q$ different regions that $X$ can model. When parameters are properly chosen, the above GRF class is good at modeling various regions and textures. Derin and Elliott were interested in two classes of images: noisy ones and textured ones. While the region processes are identical for these two classes, the low level models are different. For noisy images, the model is $Y = \mathrm{F}(X) + N$ where $\mathrm{F}$ is simply a mapping of the region type to the matching gray level and $N$ is additive white noise. For textured images, the model is $Y = \mathrm{H}(X) = T$ where $\mathrm{H}$ is a simple mapping of the region type to the corresponding texture. Derin and Elliott developed a parameter estimation scheme for finite state space GRF based on histogramming and least squares fit. The scheme has the following steps: 1) Find the relationship between the joint probability and the parameters; 2) use histogramming to calculate these probabilities; 3) build an overdetermined system of linear equations with respect to probabilities and parameters; and 4) solve the system using the least squares fit.

### Limitations of Non-hierarchical Markov Random Field Models

MAP estimation is commonly used to restore or segment images. MAP estimation finds the $x$ which maximizes the probability distribution of $X$ conditioned on the data $y$, $P(X = x|Y = y)$. The estimation is usually implemented as maximizing

$$\log P(Y = y|X = x) + \log P(X = x) - \log P(Y = y),$$

where $P(Y = y)$ is assumed to be constant. For the non-hierarchical MRFs discussed thus far, exact MAP estimation is quite expensive. Consequently, approximation methods such as iterated conditional modes (ICM) [4], dynamic programming [11], or simulated annealing with Gibbs sampling [14] are used. Another difficulty with non-hierarchical MRFs is that the neighborhood systems need to be small because larger neighborhood systems dramatically increase the number of parameters and the complexity of the MAP algorithms. Although these models perform well for the tasks described by the authors, they have only a limited ability to characterize statistical properties at large spatial scales.

### 2.3.2 Hierarchical MRF Models

To address the problems associated with non-hierarchical models, multiscale MRF models were formulated and have been extensively discussed in the image processing literature [5, 6, 20, 21, 24, 27, 29, 46]. [1] For example, in Bouman and Shapiro's multiscale MRF (MSRF) model [6] for Bayesian image segmentation, there is a series of $L$ random fields at a range of scales or resolutions. At each scale, $i$, the segmentation is modeled as a hidden MRF, $X^{(i)}$, defined on a lattice, $S^{(i)}$. $X^{(0)}$ is the base of the pyramid, *i.e.*, it is the random field at the finest scale with a lattice corresponding to the image. Each site at the next coarser scale, $X^{(1)}$, corresponds to four sites of $X^{(0)}$, etc. Bouman and Shapiro assumed that the random field at each scale depends only on the coarser random field above it in the resolution pyramid, *i.e.*, the sequence of random fields from coarse-to-fine scales is a Markov chain. The image to be segmented was characterized by an observable MRF, $Y$. The non-iterative method used by Bouman and Shapiro to segment images is termed a *sequential MAP* (SMAP) estimator. The SMAP algorithm consists of two full passes through the pyramid. The SMAP algorithms have performance comparable to or better than that of MAP estimation by simulated annealing in a non-hierarchical

---

[1] We note that the two-level hierarchical GRF model in [11] is not multiscale because there is no correlation between the two levels.

MRF. Furthermore, SMAP algorithms require less computation than simulated annealing or ICM.

Laferté *et al.* [24] also exploited the causality between scales and developed a method called *hierarchical MAP* (HMP) based on the Viterbi algorithm, to compute an exact MAP estimate on a quadtree. Like SMAP, HMP is non-iterative and requires two passes on the quadtree. SMAP was extended by Laferté *et al.* to segmentation of multi-resolution images. Experiments demonstrated that these hierarchical approaches require less computation and result in better estimation than non-hierarchical methods. Wilson and Li [46] extended the MSRF model in [6] by adding interactions between pixels at a given scale which model regions that contains significant variation in image properties. Both 4-neighborhood and 8-neighborhood systems are used in the model. However, only pairwise interactions are included in their prior. As in [24], the observation model incorporates a multi-resolution representation of the image of interest: there is a series of observable MRFs, $Y^{(i)}, 0 \leq i \leq L$, having a pyramid structure identical to Bouman and Shapiro's [6] prior. In addition, by carefully combining the prior and observation models, the estimation algorithms can be made independent of the number of classes. As a result, the model is able to segment images containing unknown number of regions. A *sequential multi-resolution MAP* estimation process, utilizing a Gibbs sampler and simulated annealing, is used to estimate random process parameters and segment images.

In yet another example of a hierarchical MRF, Kato *et al.* [20, 21] further extended the neighborhood system used in [46], so that it contains the three scales shown in Figure 2.4. Each site has interactions with its parent as well as its neighbors at the same scale, and its children. Due to the complexity of the neighborhood system, iterative algorithms are required to compute the MAP estimate. A multi-temperature annealing scheme was developed to segment images. The approach assigns high temperatures to coarser scales so that the solutions computed will less likely correspond to local minima. At the finer scales, the annealing is performed at lower temperatures. Although this provides excellent segmentation results, it is computationally expensive.

### 2.3.3 MRF Parameter Estimation

Parameter estimation is important for Bayesian image analysis because model parameters are required for MAP estimation. Maximum likelihood (ML) estimation is often used for MRF parameter estimation [3]. When part of the data is hidden, the Expectation-Maximization (EM) algorithm [10] is commonly used for ML estimation. The EM algorithm is an iterative procedure in which each iteration has two steps: 1) the E-step which computes the conditional expectation of the likelihood of the hidden data given the observed data and current estimates; and 2) the M-step which updates the parameter estimates by maximizing the conditional expectation computed in the E-step. The procedure stops when the parameter estimates stabilize. A major difficulty in applying the EM algorithm to MRFs is in the calculation of the conditional expectation in the E-step, which is generally intractable because it requires summing over all possible configurations. Therefore, approximation techniques such as the Mean Field Approximation [7, 38, 48] and pseudo-likelihood method [8, 50] are used in the E-step. SEM, a stochastic version of the EM algorithm, has been proposed to reduce dependence on the initial estimate by adding a stochastic step after the E-step [28]. Alternatively, Monte Carlo Markov Chain sampling can be used to approximately compute the expectation [25, 35, 41]. The resulting Monte Carlo EM (MCEM) includes SEM as a special case. EM and MCEM have been extended for parameter estimation on a quadtree [6, 24].

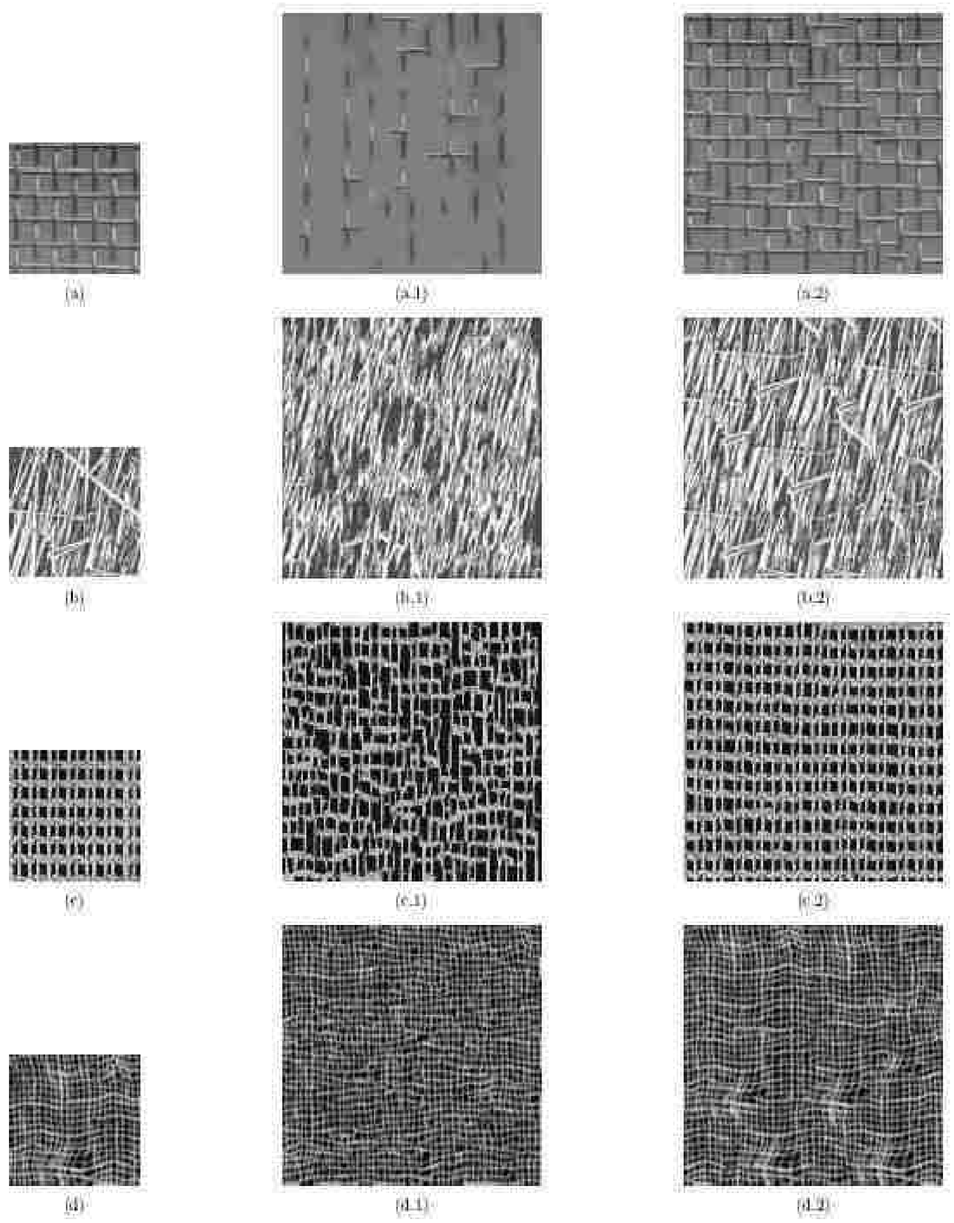

Figure 2.2: Brodatz textures: a) aluminum wire mesh; b) straw; c) magnified French canvas; d) loose burlap. The middle column and the right column are synthesized textures using $24$-neighborhood and $48$-neighborhood respectively [32].

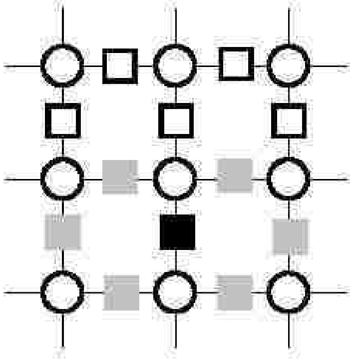

Figure 2.3: Pixel sites (circles) and edge sites (squares) in Geman and Geman's image restoration model [14]. Pixel sites have a 4-neighborhood system, and each edge site has 6 neighbors.

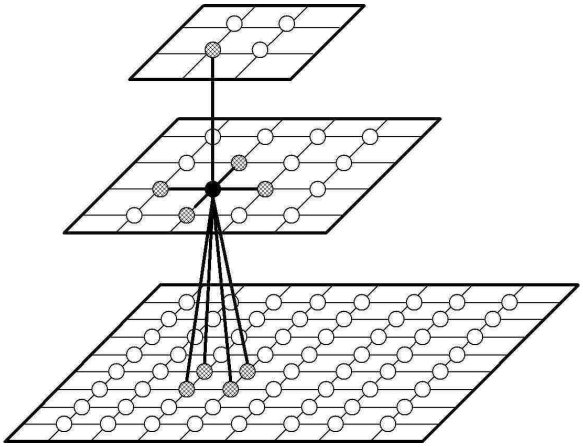

Figure 2.4: The neighborhood system in Kato's multiscale MRF model [21].

# Chapter 3

# Modeling Protein Spatial Distributions

In a specific signaling pathway, there are typically from 10 to 100 proteins of interest. Let us consider a simple, idealized case, where there are three proteins of interest which we call $\{R, G, B\}$, and where only two proteins can be observed in any single sample. For this scenario, we use a hidden process $D$ with four states $\{R, G, B, X\}$ to characterize the distribution of proteins on the cell membrane. The additional state $X$ corresponds to background. $D$ is called a *distribution process*. In addition, there are three observable processes, $O_r$, $O_g$ and $O_b$, to model observations where only two kinds of proteins can be observed at a time. These processes are called *observation processes* and have four observable states $\{R', G', B', X'\}$. As in [36], we call these observable states *observation symbols*, corresponding to the presence or absence of a gold particle bound to a protein visualized by TEM. The observation symbol probability distribution [36] is used to describe the probability that hidden state $j$ will be observed as symbol $i$.

The approach adopted here is based on the quadruple stochastic model shown in Figure 3.1. This model assumes that the three observation processes depend on the same distribution process. The behavior of the observation processes given the distribution process is defined by the observation symbol probability distribution. There are three observation symbol probability distributions, one for each observation process. It is assumed that the state in a given observation process depends only on the corresponding state in the distribution process, and that three *observation matrices*, $\mathbf{Q}_r$, $\mathbf{Q}_g$ and $\mathbf{Q}_b$, can be used to represent the conditional probability mass functions, $P((O_r)_s = i|D_s = j)$, $P((O_g)_s = i|D_s = j)$ and $P((O_b)_s = i|D_s = j)$:

$$\mathbf{Q}_r = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{Q}_g = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

$$\mathbf{Q}_b = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

where $(\mathbf{Q}_k)_{ij}$ is the probability that the hidden state $j$ will be observed as the symbol $i$ in $O_k$. The problem is to infer the conditional p.m.f. of $D$ given the observation processes $O_r$, $O_g$ and $O_b$.

In the following sections, first-order Markov chains are used for both the distribution and observation processes to demonstrate that the quadruple stochastic model is feasible. The transition probabilities of hidden first-order Markov chains can be estimated in closed form given the visible chains. It is then demonstrated that the transition probabilities of hidden second-order Markov chains can be estimated in closed form under an assumption of isotropy. Finally, we describe that the conditional p.m.f of a hidden MRF with a 4-neighborhood system can be reliably estimated given the visible MRFs.

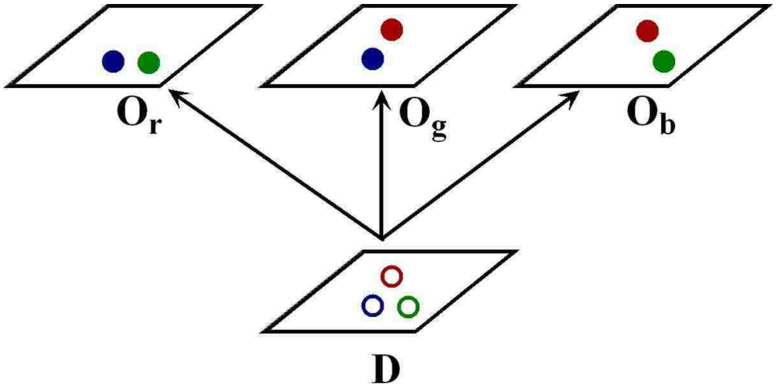

Figure 3.1: Structure of a quadruple stochastic random field used in modeling of protein spatial distributions. Red, green and blue open circles represent $R$, $G$, and $B$, and red, green and blue closed circles represent $R'$, $G'$, and $B'$.

## 3.1 A First-order Markov Chain Model

This section demonstrates that the transition probabilities of a hidden first-order chain can be accurately estimated given the visible chains.

The distribution process $D$ is a first-order Markov chain $\Gamma$ in which the Markov property implies that

$$\begin{aligned} P(\gamma_{t+1} = j|\gamma_t = i, \gamma_{t-1} = i_{t-1}, \cdots, \gamma_1 = i_1, \gamma_0 = i_0) \\ = P(\gamma_{t+1} = j|\gamma_t = i) \\ = p_{ji} \end{aligned} \tag{3.1}$$

where $i$, $j \in \{R, G, B, X\}$, and $p_{ji}$ is the probability of going from state $i$ to state $j$. Assuming that the Markov process is homogeneous, the statistics of $\Gamma$ are defined by a transition matrix $\mathbf{P}$:

$$\mathbf{P} = \begin{bmatrix} p_{rr} & p_{rg} & p_{rb} & p_{rx} \\ p_{gr} & p_{gg} & p_{gb} & p_{gx} \\ p_{br} & p_{bg} & p_{bb} & p_{bx} \\ p_{xr} & p_{xg} & p_{xb} & p_{xx} \end{bmatrix}.$$

The first step is to generate a set of visible chains. To generate visible chains, we must first generate a hidden chain $\gamma$ of length $N$. Three visible chains $\gamma_r$, $\gamma_g$, and $\gamma_b$ are then obtained by mapping the hidden states in $\gamma$ to the observable symbols using the observation matrices. For example, we replace $R, G, B, X$ with $X', G', B', X'$ respectively to generate $\gamma_r$. The visible chains, $\gamma_r$, $\gamma_g$, and $\gamma_b$, are realizations of the observation models $\Gamma_r$, $\Gamma_g$, and $\Gamma_b$. The hidden chain $\gamma$ is a realization of $\Gamma$. The transition probabilities in $\mathbf{P}$ are estimated using histogramming [16]. For example, the probability of going from $R$ to $G$ can be estimated as follows:

$$p_{gr} = P(G|R) = \frac{P(\langle G, R\rangle)}{P(\langle R\rangle)} \approx \frac{H(\langle G, R\rangle)/N}{H(\langle R\rangle)/N} = \frac{H(\langle G, R\rangle)}{H(\langle R\rangle)} \tag{3.2}$$

where $H(\langle R\rangle)$ and $H(\langle G, R\rangle)$ are the observed frequencies of the one-tuple $\langle R\rangle$ and the two-tuple $\langle G, R\rangle$ in the hidden chain. In order to accurately estimate the transition probabilities, the chain length $N$ needs to be quite large. In addition, the histograms are calculated from a sufficiently large starting point $\mathcal{S}$ so that they will be representative of the stationary distribution.

Unfortunately, the histograms for all states in the hidden chain are not directly observable. However, we can deduce them from the histograms of symbols in the visible chains. Because of the one-to-one relationships between $G$ in $\gamma$ and $G'$ in $\gamma_r$ and between $B$ in $\gamma$ and $B'$ in $\gamma_r$, we observe that

$$H(\langle G\rangle) = H_r(\langle G'\rangle), \qquad H(\langle B\rangle) = H_r(\langle B'\rangle) \tag{3.3}$$

$$H(\langle G, G\rangle) = H_r(\langle G', G'\rangle), \qquad H(\langle B, B\rangle) = H_r(\langle B', B'\rangle) \tag{3.4}$$

$$H(\langle B, G\rangle) = H_r(\langle B', G'\rangle), \qquad H(\langle G, B\rangle) = H_r(\langle G', B'\rangle) \tag{3.5}$$

where $H_r$ are histograms of one-tuples and two-tuples of symbols in $\gamma_r$. It follows that

$$p_{gg} \approx \frac{H_r(\langle G', G' \rangle)}{H_r(\langle G' \rangle)} \tag{3.6}$$

$$p_{bg} \approx \frac{H_r(\langle B', G' \rangle)}{H_r(\langle G' \rangle)} \tag{3.7}$$

$$p_{gb} \approx \frac{H_r(\langle G', B' \rangle)}{H_r(\langle B' \rangle)} \tag{3.8}$$

$$p_{bb} \approx \frac{H_r(\langle B', B' \rangle)}{H_r(\langle B' \rangle)}. \tag{3.9}$$

Using analogous methods, $p_{rr}$, $p_{rb}$, $p_{br}$, $p_{bb}$ can be estimated from $\gamma_g$, and $p_{rr}$, $p_{rg}$, $p_{gr}$, $p_{gg}$ can be estimated from $\gamma_b$. Therefore, estimates for all transition probabilities involving at most two of the non-background states, namely $p_{rr}$, $p_{rg}$, $p_{rb}$, $p_{gr}$, $p_{gg}$, $p_{gb}$, $p_{br}$, $p_{bg}$, $p_{bb}$, can be estimated directly from the visible chains. Applying the stochastic matrix property, *i.e.*, all column sums must equal one, three more entries in the transition matrix can be estimated:

$$p_{xr} = 1 - p_{rr} - p_{gr} - p_{br}$$

$$p_{xg} = 1 - p_{rg} - p_{gg} - p_{bg}$$

$$p_{xb} = 1 - p_{rb} - p_{gb} - p_{bb}.$$

Four additional transition probabilities remain to be estimated: $p_{rx}$, $p_{gx}$, $p_{bx}$, and $p_{xx}$. Since all $G$s in $\gamma$ are observed as $X'$s in $\gamma_g$, it follows that

$$H_g(\langle R', X' \rangle) = H(\langle R, X \rangle) + H(\langle R, G \rangle) \tag{3.10}$$

$$= H(\langle X \rangle) \cdot p_{rx} + H_b(\langle R', G' \rangle). \tag{3.11}$$

We now observe that all histograms of one-tuples in the hidden chain are known:

$$H(\langle R \rangle) = H_g(\langle R' \rangle) \tag{3.12}$$

$$H(\langle G \rangle) = H_r(\langle G' \rangle) \tag{3.13}$$

$$H(\langle B \rangle) = H_g(\langle B' \rangle) \tag{3.14}$$

$$H(\langle X \rangle) = N - \mathcal{S} - H(\langle R \rangle) - H(\langle G \rangle) - H(\langle B \rangle) \tag{3.15}$$

where $N$ is the chain length and $\mathcal{S}$ is the starting point for histogramming. Therefore, the only unknown variable in Equation 3.11 is $p_{rx}$. Using methods analogous to those described above for $H_g(\langle R', X'\rangle)$, eight linear equations are obtained using $H_g(\langle B', X'\rangle)$, $H_g(\langle X', X'\rangle)$, $H_r(\langle G', X'\rangle)$, $H_r(\langle B', X'\rangle)$, $H_r(\langle X', X'\rangle)$, $H_b(\langle R', X'\rangle)$, $H_b(\langle B', X'\rangle)$, and $H_b(\langle X', X'\rangle)$, where $p_{rx}$, $p_{gx}$, $p_{bx}$, and $p_{xx}$ are the unknown variables. The resulting system of nine linear equations in four unknowns can be solved in closed form using the pseudo-inverse method.

The relative error of $\hat{\mathbf{P}}$ is computed in terms of the matrix 2-norm, $\|\hat{\mathbf{P}} - \mathbf{P}\|_2 / \|\mathbf{P}\|_2$. The average error for seven experiments is approximately $10^{-4}$, *i.e.*, $\hat{\mathbf{P}}$ is accurate to 4 decimal digits. Actual data are shown in Appendix A.1.

## 3.2 A Second-order Markov Chain Model

Unfortunately, first-order Markov chains are too trivial to apply to experimental data. Consequently, we investigate whether the transition probabilities of a hidden higher-order Markov chain can be accurately estimated given the visible chains. This Section shows that under an assumption of isotropy the transition probabilities of a hidden second-order Markov chain can also be estimated in closed form. For a second-order Markov chain, the Markov property means that

$$\begin{aligned} &P(\gamma_{t+1} = k | \gamma_t = j, \gamma_{t-1} = i, \gamma_{t-2} = i_{t-2}, \cdots, \gamma_1 = i_1, \gamma_0 = i_0) \\ &\qquad = P(\gamma_{t+1} = k | \gamma_t = j, \gamma_{t-1} = i) \\ &\qquad = p_{kji} \end{aligned} \tag{3.16}$$

where $i$, $j$, $k \in \{R, G, B, X\}$, and $p_{kji}$ is the probability that the process will enter state $k$ given that the current state is $j$ and the previous state was $i$. There are $4^3 = 64$ such transition probabilities. As before, $\mathbf{P}$ is used to denote the matrix[1] of these transition

---

[1] This term is used even though $\mathbf{P}$ is three dimensional, *i.e.*, it is a *tensor*.

probabilities. Given $\mathbf{P}$, a second-order hidden chain $\gamma$ and three visible chains, $\gamma_r$, $\gamma_g$, and $\gamma_b$, are generated using an approach similar to those described in Section 3.1.

The histograms of one-tuples and two-tuples of states in $\gamma$ can be computed as in Section 3.1. However, we also need the histograms of three-tuples of states to calculate the transition probabilities for the second-order Markov chain

$$p_{nml} = \frac{H(\langle n,m,l\rangle)}{H(\langle m,l\rangle)} \tag{3.17}$$

where $l$, $m$, $n \in \{R,G,B,X\}$, and $H(\langle n,m,l\rangle)$ is the value of the histogram for the three-tuple $\langle n,m,l\rangle$ in $\gamma$. A strategy identical to that described in Section 3.1 is used to infer the transition probabilities from the visible chains. We first compute the transition probabilities involving at most two of the non-background states: 1) $p_{ggg}$, $p_{bgg}$, $p_{gbg}$, $p_{bbg}$, $p_{ggb}$, $p_{bgb}$, $p_{gbb}$ and $p_{bbb}$ from $\gamma_r$; 2) $p_{rrr}$, $p_{brr}$, $p_{rbr}$, $p_{bbr}$, $p_{rrb}$, $p_{brb}$, $p_{rbb}$ and $p_{bbb}$ from $\gamma_g$; and 3) $p_{rrr}$, $p_{grr}$, $p_{rgr}$, $p_{ggr}$, $p_{rrg}$, $p_{grg}$, $p_{rgg}$ and $p_{ggg}$ from $\gamma_b$. Three additional transition probabilities are estimated by exploiting the stochastic matrix property

$$\begin{aligned} p_{xrr} &= 1 - p_{rrr} - p_{grr} - p_{brr} \\ p_{xgg} &= 1 - p_{rgg} - p_{ggg} - p_{bgg} \\ p_{xbb} &= 1 - p_{rbb} - p_{gbb} - p_{bbb}. \end{aligned}$$

In addition, we can directly estimate six more entries in $\mathbf{P}$, namely $p_{nml}$, where $n$, $m$, $l \in \{X,R\}$, or $\{X,G\}$, or $\{X,B\}$, and $X$ appears once in the three-tuple $\langle n,m,l\rangle$. For example, because

$$H_r(\langle G',X',G'\rangle) = H(\langle G,X,G\rangle) + H(\langle G,R,G\rangle) \tag{3.18}$$

and

$$H(\langle G,R,G\rangle) = H_b(\langle G',R',G'\rangle) \tag{3.19}$$

it follows that

$$\begin{aligned} p_{gxg} &\approx \frac{H(\langle G, X, G)}{H(\langle X, G\rangle)} \\ &= \frac{H_r(\langle G', X', G') - H_b(\langle G', R', G'\rangle)}{H(\langle X, G\rangle)}. \end{aligned} \tag{3.20}$$

Using analogous methods, $p_{ggx}$, $p_{rxr}$, $p_{rrx}$, $p_{bxb}$ and $p_{bbx}$ can also be estimated. It follows that there are 34 remaining unknown variables, namely $p_{bgr}$, $p_{xgr}$, $p_{gbr}$, $p_{xbr}$, $p_{gxr}$, $p_{bxr}$, $p_{xxr}$, $p_{brg}$, $p_{xrg}$, $p_{rbg}$, $p_{xbg}$, $p_{rxg}$, $p_{bxg}$, $p_{xxg}$, $p_{grb}$, $p_{xrb}$, $p_{rgb}$, $p_{xgb}$, $p_{rxb}$, $p_{gxb}$, $p_{xxb}$, $p_{grx}$, $p_{brx}$, $p_{xrx}$, $p_{rgx}$, $p_{bgx}$, $p_{xgx}$, $p_{rbx}$, $p_{gbx}$, $p_{xbx}$, $p_{rxx}$, $p_{gxx}$, $p_{bxx}$ and $p_{xxx}$.

In a manner analogous to that described in Section 3.1, we write the observed frequencies of three-tuples of symbols in the visible chains in terms of the transition probabilities and histograms of two-tuples of states. For example,

$$\begin{aligned} H_r(\langle X', B', G'\rangle) &= H(\langle X, B, G\rangle) + H(\langle R, B, G\rangle) \\ &= H(\langle B, G\rangle) \cdot (p_{xbg} + p_{rgb}) \end{aligned}$$

where $p_{xbg}$ and $p_{rgb}$ are unknown variables. In an analogous manner, 12 additional linear equations can be derived from $\gamma_r$ using

$$\begin{array}{llll} H_r(\langle B', X', G'\rangle), & H_r(\langle X', X', G'\rangle), & H_r(\langle X', G', B'\rangle), & H_r(\langle G', X', B'\rangle), \\ H_r(\langle X', X', B'\rangle), & H_r(\langle B', G', X'\rangle), & H_r(\langle X', G', X'\rangle), & H_r(\langle G', B', X'\rangle), \\ H_r(\langle X', B', X'\rangle), & H_r(\langle G', X', X'\rangle), & H_r(\langle B', X', X'\rangle), & H_r(\langle X', X', X'\rangle). \end{array}$$

An analogous approach can be used to derive 13 linear equations from $\gamma_g$ and 13 linear equations from $\gamma_b$. Unfortunately, the resulting system of 39 linear equations in 34 unknowns is under-determined. To provide additional constraints on the matrix of transition probabilities, an assumption of symmetry under reflection (isotropy) is introduced (Figure 3.2). As a result, the number of unique transition probabilities in $\mathbf{P}$ is reduced from 64 to 40. Because the number of unknown variables is also reduced from 34 to 19, the resulting linear system can now be solved in closed form using the method of pseudo-inverse.

We compute the relative error using the matrix 2-norm as in Section 3.1. The average error of five experiments is approximately $10^{-3}$. Actual data are shown in Appendix A.2.

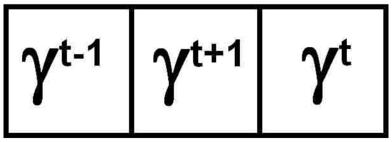

Figure 3.2: Concatenation of the future state $\gamma^{t+1}$, the current state $\gamma^{t}$, and the previous state $\gamma^{t-1}$ into a two-neighbor neighborhood system. Symmetry under isotropy means that $P(\gamma^{t+1} = n | \gamma^{t} = m, \gamma^{t-1} = l)$ equals $P(\gamma^{t+1} = n | \gamma^{t} = l, \gamma^{t-1} = m)$, *i.e.*, for all $l$, $m$, $n \in \{R, G, B, X\}$, $p_{nml}$ equals $p_{nlm}$.

## 3.3 A Markov Random Field Model

This section investigates whether the strategy used in the first-order and second-order Markov chain models works for Markov random field models. A Markov random field can be specified by a conditional probability mass function, $P(X_s = x_s | X_t = x_t, t \in \mathcal{N}_s)$. Unfortunately, the conditional p.m.f. of a hidden MRF can not be estimated in closed form using a strategy similar to the strategy described in Section 3.2. As a result, ML estimation is utilized to estimate the conditional probabilities. A method using mathematical programming techniques is developed to solve the ML estimation problem and it is shown that this approach results in good estimates for MRFs with a 4-neighborhood system.

MRFs with a 4-neighborhood system (see Figure 2.1) are used for both the distribution and observation processes. For MRFs with a 4-neighborhood system, the matrix of conditional probabilities, $\mathbf{P}$, is five dimensional and there are $4^5 = 1024$ elements in $\mathbf{P}$. Given $\mathbf{P}$, we first generate a realization of the distribution process, $d$, using the Gibbs sampler [14]:

```
Initialize $d$ randomly.
```

```
2: Choose $s \in S$ randomly and replace $D_s$ with $d_s$ drawn from $P(D_s = d_s | D_t = d_t, t \in \mathcal{N}_s)$.
3: Repeat Step 2 many times.
```

After the realization is generated, samples of three visible fields, $o_r$, $o_g$, and $o_b$, are obtained by mapping the states in $d$ to symbols of the observation process using the observation matrices. For exposition purposes, we number the sites in the 4-neighborhood system as in Figure 3.3a. A five-tuple $\langle m, l, k, j, i \rangle$ and a four-tuple $\langle l, k, j, i \rangle$ can then be used to denote the joint events $(D_4 = m, D_3 = l, D_2 = k, D_1 = j, D_0 = i)$ and $(D_3 = l, D_2 = k, D_1 = j, D_0 = i)$ respectively. The conditional probability $P(D_4 = m | D_3 = l, D_2 = k, D_1 = j, D_0 = i)$ is estimated as:

$$
\begin{aligned}
P(D_4 = m | D_3 = 1, D_2 = k, D_1 = j, D_0 = i) \\
&= p_{mlkji} \\
&\approx \frac{H(\langle m, l, k, j, i \rangle)}{H(\langle l, k, j, i \rangle)}
\end{aligned}
\tag{3.21}
$$

where $H$ are frequencies of the tuples in the samples of $D$. In order to get good estimates for the conditional probabilities, the sample size need to be quite large. The realization of the distribution process, $d$, is generated on a lattice of size $600 \times 740$. Corresponding to each $d$, there is a set of three visible samples. Our experiments showed that 100 sets of visible samples results in good estimates.

By performing a raster scan of the window shown in Figure 3.3 over the visible samples, we obtain the frequencies of four-tuples and five-tuples of observation symbols. We can directly estimate 108 of the conditional probabilities from these observed frequencies and by exploiting the stochastic matrix property. A system of linear equations can be built as in Section 3.2. Unfortunately, the resulting system of linear equations is underdetermined. To reduce the number of unknown variables, we exploit isotropy in the matrix of conditional probabilities as shown in Figure 3.4. The isotropy assumption is reasonable because the protein distributions are independent of orientation. The number of unknown variables is reduced to 181 under this assumption. Unfortunately, the system of linear equations is still under-determined. Consequently, ML estimation is adopted to infer the conditional probabilities. Although the EM algorithm is the standard method for ML estimation when part of the data is hidden, it converges very slowly. By exploiting a property of the observation matrix (*i.e.*, the probabilities in the observation matrix are either 0 or 1), we develop a non-iterative method using mathematical programming techniques to solve the ML estimation problem.

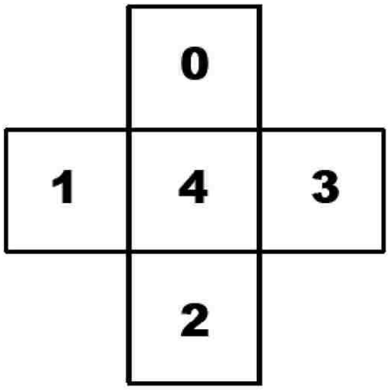

Figure 3.3: Numbering of the sites in a 4-neighborhood system.

Given a realization $d$ of the distribution process, the ML estimate maximizes the conditional probability, $P(d|\mathbf{P})$. The probability of $d$ given $\mathbf{P}$ is approximated by the pseudo-likelihood [26] which is simply a product of the conditional probabilities:

$$\mathrm{P}(d|\mathbf{P}) \approx \prod_{s \in S} \mathrm{P}(d_s|d_t, t \in \mathcal{N}_s). \tag{3.22}$$

However, the distribution process is not directly observable. Only visible samples are available. Consequently, we need to compute the probability of a visible sample $o_k$ given $\mathbf{P}$ and $\mathbf{Q}_k$. We also approximate the conditional probability of $o_k$ given $\mathbf{P}$ and $\mathbf{Q}_k$ by the pseudo-likelihood:

$$\mathrm{P}(o_k|\mathbf{P}, \mathbf{Q}_k) \approx \prod_{s \in S} \mathrm{P}((o_k)_s|(o_k)_t, t \in \mathcal{N}_s) \tag{3.23}$$

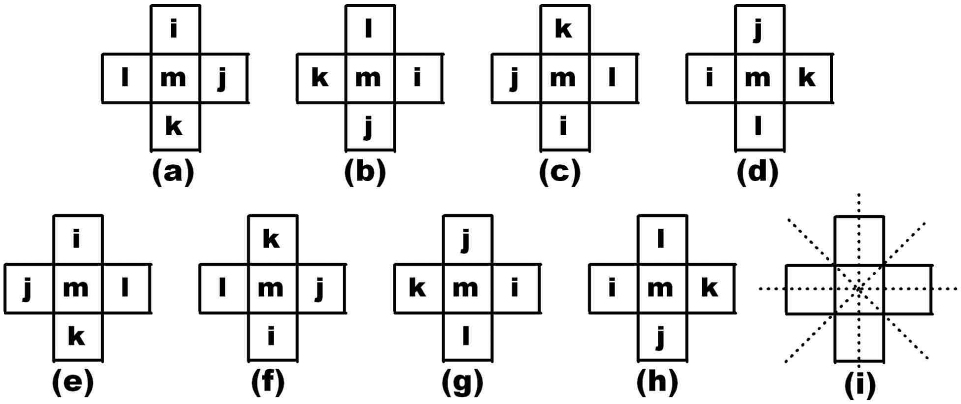

Figure 3.4: Isotropy in a 4-neighborhood system. b), c) and d) are rotations of a) by multiples of $90^\circ$. e)-h) are mirror images of a) with respect to four axes of reflective symmetry, i) axes of reflective symmetry. All eight conditional probabilities, *e.g.*, $P(m|i,l,k,j)$ and $P(m|l,k,j,i)$, are equal under the isotropy assumption.

where (by Bayes' rule):

$$\begin{aligned}
&\mathrm{P}((o_k)_s|(o_k)_t, t \in \mathcal{N}_s) \\
&= \sum_{d_t, t\in\mathcal{N}_s} \mathrm{P}((o_k)_s|d_s)\mathrm{P}(d_s|d_t, t \in \mathcal{N}_s)\mathrm{P}(d_t, t \in \mathcal{N}_s|(o_k)_t, t \in \mathcal{N}_s) \qquad (3.24)\\
&= \sum_{d_t, t\in\mathcal{N}_s} \frac{\mathrm{P}((o_k)_s|d_s)\mathrm{P}(d_s|d_t, t \in \mathcal{N}_s)\mathrm{P}((o_k)_t, t \in \mathcal{N}_s|d_t, t \in \mathcal{N}_s)\mathrm{P}(d_t, t \in \mathcal{N}_s)}{\mathrm{P}((o_k)_t, t \in \mathcal{N}_s)}. \qquad (3.25)
\end{aligned}$$

Because it is assumed that a symbol at a site in an observation process depends only on the corresponding state in the distribution process, it follows that

$$P((o_k)_t, t \in \mathcal{N}_s|d_t, t \in \mathcal{N}_s) = \prod_{t\in\mathcal{N}_s} \mathrm{P}((o_k)_t|d_t). \qquad (3.26)$$

Finally, we observe that

$$\begin{aligned}
&\mathrm{P}((o_k)_s|(o_k)_t, t \in \mathcal{N}_s) \\
&= \sum_{d_t, t\in\mathcal{N}_s} \frac{\mathrm{P}((o_k)_s|d_s)\mathrm{P}(d_s|d_t, t \in \mathcal{N}_s)\mathrm{P}(d_t, t \in \mathcal{N}_s)\prod_{t\in\mathcal{N}_s}\mathrm{P}((o_k)_t|d_t)}{\mathrm{P}((o_k)_t, t \in \mathcal{N}_s)} \qquad (3.27)
\end{aligned}$$

and where $k \in \{r, g, b\}$.

In Eq. 3.27, $P((o_k)_s|d_s)$ is known and defined in $\mathbf{Q}_k$. $P((o_k)_t, t \in \mathcal{N}_s)$ are the probabilities of four-tuples of symbols in the visible fields, and can be estimated from samples of the visible fields. $P(d_t, t \in \mathcal{N}_s)$ are the probabilities of four-tuples of states in the hidden field (determined by $\mathbf{P}$). The unknown variables are $P(d_s|d_t, t \in \mathcal{N}_s)$ and $P(d_t, t \in \mathcal{N}_s)$. The Maximum likelihood estimation of $\mathbf{P}$ is achieved by maximizing the logarithm of the probability that the process generated the visible samples as a function of $P(d_s|d_t, t \in \mathcal{N}_s)$ and $P(d_t, t \in \mathcal{N}_s)$:

$$\hat{\mathbf{P}} = \arg\max_{\mathbf{P}} \sum_{k\in\{r,g,b\}} \log \mathrm{P}(o_k|\mathbf{P}, \mathbf{Q}_k). \tag{3.28}$$

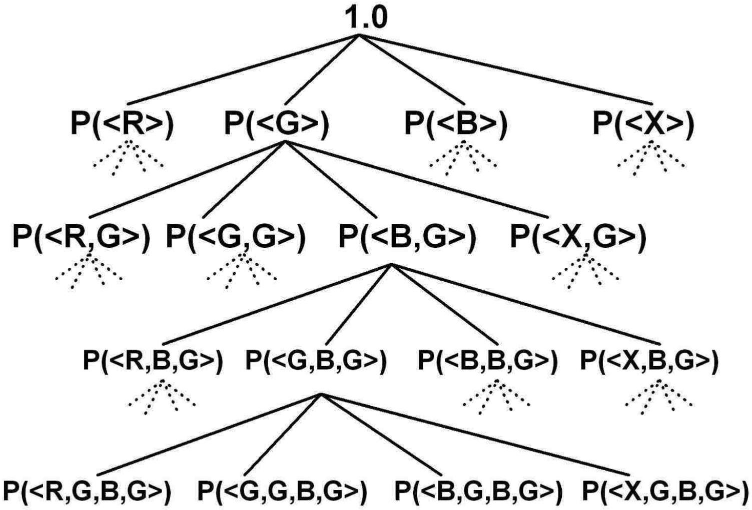

Figure 3.5: Quadtree of tuple probabilities. The probability of each vertex equals the sum of the probabilities of its children.

Under the isotropy assumption, there are 181 unknown conditional probabilities. In addition, 37 of the joint probabilities of state four-tuples are also unknown. The conditional probabilities are constrained by

$$\forall i, j, k, l \quad \sum_m p_{mlkji} = 1 \tag{3.29}$$

where $i$, $j$, $k$, $l$, $m \in \{R, G, B, X\}$. There are also constraints on the joint probabilities of state four-tuples. To illustrate these constraints, we can build a quadtree of the four-tuple probabilities (see Figure 3.5). As before, we can compute the probabilities of all one-tuples and two-tuples, and the probabilities of all tuples containing at most two non-background states. The constraint is

$$\sum_{\langle l,k,j,i\rangle \in \mathcal{C}(a)} P(\langle l, k, j, i\rangle) = P(a) \tag{3.30}$$

where $a$ is the nearest known ancestor of the four-tuple and $\mathcal{C}(a)$ is the set of leaves of the quadtree which are on the branch which starts at $a$. We know that at least one of the ancestors of the four-tuple is known because the probabilities of all two-tuples are known. Therefore, there are 218 unknown variables with constraints defined by Eq. 3.29 and Eq. 3.30. The objective function is

$$\begin{aligned}
f &= \sum_k \log \mathrm{P}(o_k|\mathbf{P}, \mathbf{Q}_k) && \text{(by Eq. 3.23)} \\
&= \sum_k \log \left[\prod_{s\in S} \mathrm{P}((o_k)_s|(o_k)_t, t \in \mathcal{N}_s)\right] \\
&= \sum_k \sum_{(o_k)_t, t\in\mathcal{N}_s} H_k((o_k)_t, t \in \mathcal{N}_s) \log \mathrm{P}((o_k)_s|(o_k)_t, t \in \mathcal{N}_s)
\end{aligned} \tag{3.31}$$

where $k \in \{r, g, b\}$ and $\mathrm{P}((o_k)_s|(o_k)_t, t \in \mathcal{N}_s)$ is defined in Eq. 3.27. $H_k((o_k)_t, t \in \mathcal{N}_s)$ is the observed frequency of the four-tuple of symbols. The problem we need to solve is

maximize: $f$

subject to:

(i) $0 \le P(d_s|d_t, t \in \mathcal{N}_s) \le 1$;

(ii) $0 \le P(d_t, t \in \mathcal{N}_s) \le 1$;

(iii) constraint defined by Eq. 3.29;

(iv) constraint defined by Eq. 3.30.

A software package, SNOPT [15], is used to solve the above problem.

## 3.4 Results for the Markov Random Field Model

To test this method for ML estimation, a conditional distribution is first specified for a MRF with a 4-neighborhood system. 100 realizations of the MRF were generated on a lattice of size $600 \times 740$ by sampling from the conditional distribution using the Gibbs sampler. For each realization, three visible samples were obtained by mapping the hidden states in the realization to the observation symbols using the observation matrices. By means of a raster scan (see Figure 3.3) over the visible samples, the frequencies of four-tuples and five-tuples of observation symbols were computed. We estimated 108 conditional probabilities directly from these frequencies. To estimate the remaining conditional probabilities, Eq. 3.31 was used to create an objective function which was maximized subject to the constraints described in Section 3.3 using SNOPT. This results in good estimates which generate realizations which are visually very similar to those generated using the specified conditional distributions (see Figure 3.6). The relative errors of estimates are shown in Table 3.1. We note that the estimates are not as good as those for the first- and second-order Markov chain models, especially for the case shown in Figure 3.6 b) and c). One factor that may contribute to the error in the estimates is the lack of sufficient data for accurate estimation of conditional probabilities. Another factor may be the fact that the visible fields may be consistent with many different hidden distributions, and the ML estimation process can yield any of these distributions [11].

|  | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $\frac{\|\hat{\mathbf{P}}-\mathbf{P}\|_2}{\|\mathbf{P}\|_2}(\times 10^2)$ | 2.98 | 14.3 | 37.4 | 6.20 |
| Figure | 3.6 a) | 3.6 b) | 3.6 c) | 3.6 d) |

Table 3.1: Relative errors of estimates for the MRF modeling.

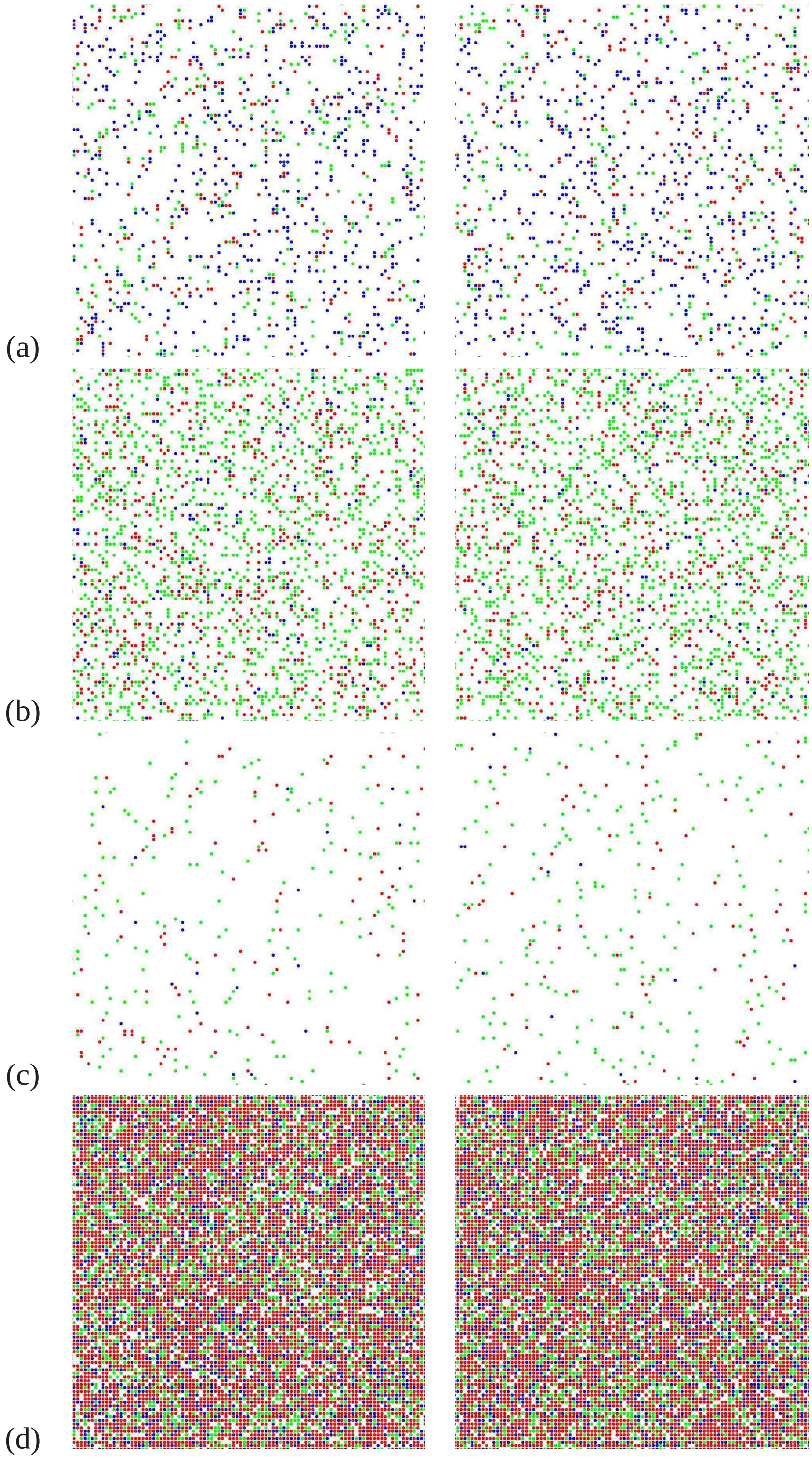

Figure 3.6: Realizations from the MRF model with specified (left) and estimated (right) conditional distributions.

# Chapter 4

# A Multiscale Modeling Approach

Unfortunately, the 4-neighborhood system is too small to achieve satisfactory results in many situations, especially when there are long range correlations between states. In addition, because 10nm particles are larger than 5nm particles, the interactions between 10nm particles are at distances longer than those between 5nm particles in the experimental data. Figure 4.1 illustrates the problem: The clustering of 10nm particles is optimally characterized at the scale of 80 pixels, but clustering of 5nm particles is optimally characterized at the scale of 40 pixels. We also note that artifacts due to the small neighborhood system, *i.e.*, particle positions aliasing with the lattice become worse with coarser grids. Although we can use larger neighborhood systems, *e.g.*, Tjelmeland and Besag [37], the number of unknown variables dramatically increases. For example, there are 19400 unknown variables for a 8-neighborhood system, which is too large for SNOPT to handle. However, a more powerful mathematical programming package was able to solve a system with such a large number of unknown variables, our datasets are not large enough to reliably estimate the conditional distributions, because only 10 to 20 images are captured in a typical experiment using nanoprobes.

## 4.1 A Multiscale MRF Model

To solve those problems, a multiscale MRF model has been developed. Like Bouman and Shapiro [6], a pyramid with three layers is built (Figure 4.2a). Each site is a parent of four sites in the finer layer beneath it. Let us denote the three layers from coarse-to-fine, $L^{(2)}$, $L^{(1)}$, and $L^{(0)}$. For each layer, there is a corresponding distribution process, $D^{(i)}$ where $i \in \{2, 1, 0\}$. Each of these is modeled as a MRF with a 4-neighborhood system. Furthermore, two additional distribution processes are used to model the correlation between the values of sites in one layer and those in the layer beneath it: $D^{(2,1)}$ between $L^{(2)}$ and $L^{(1)}$, and $D^{(1,0)}$ between $L^{(1)}$ and $L^{(0)}$. Both of these are modeled as MRFs with a (4+1)-neighborhood system, where four neighbors are in the same layer and the fifth neighbor is their parent in the coarser layer above (Figure 4.2b). There are also three corresponding observation processes for each of the distribution processes: $O_k^{(2)}$, $O_k^{(1)}$, $O_k^{(0)}$, $O_k^{(2,1)}$, and $O_k^{(1,0)}$ where $k \in \{r, g, b\}$. In the multiscale MRF model, the value of a site at a given scale depends not only on its parent in the layer above but also on its neighbors at the same scale. In this respect, the model is closely related to those presented in [20, 21, 29, 46]. However, unlike the models described by these authors, the statistical inference problem is solved by means of a sequence of related multi-resolution problems rather than as a single problem representing the entire quadtree. Multi-resolution representations of the observed data at three scales are realizations of $O^{(i)}$, $i \in \{2, 1, 0\}$, and data between two scales are realizations of $O^{(2,1)}$ and $O^{(1,0)}$. The conditional p.m.f.'s of $D^{(i)}$ can be inferred from $O_k^{(i)}$ as described in Section 3.3. Furthermore, if the conditional p.m.f.'s of $D^{(2,1)}$ can be estimated from $O_k^{(2,1)}$, and $D^{(1,0)}$ from $O_k^{(1,0)}$, then the Gibbs sampler can be used to generate samples from the conditional p.m.f.'s in a coarse-to-fine manner.

The multiscale MRF model has been tested on the data used in Figure 4.1. The data were processed at three different scales, with grid sizes of 20, 40 and 80 pixels respectively. The finest scale was chosen to be 20 pixels because that is the most frequent distance of the nearest neighbor for every particle in the data set (Figure 4.3). When there are multiple proteins present at a single site, simple majority is used to determine the state of the site. Because there are only two proteins, the frequencies of the four-tuples and five-tuples of states are computed by means of a raster scan where the 4-neighborhood is scaled to a size of $80 \times 80$ pixels (see Figure 3.3). The conditional probabilities of $D^{(2)}$ are then computed using Eq. 3.21. In an analogous way, the conditional probabilities of $D^{(1)}$ and $D^{(0)}$ are computed using the data at scales of $40$ and $20$ pixels respectively. By means of a raster scan over the data at scales of both $80$ and $40$ pixels, the frequencies of the five-tuples and six-tuples of states (see Section 4.2) can also be computed. The conditional probabilities of $D^{(2,1)}$ are then computed using Eq. 4.1. In an analogous manner, the conditional probabilities of $D^{(1,0)}$ are computed using data at scales of $40$ and $20$ pixels. The Gibbs sampler is then used to generate samples from the conditional p.m.f's:

```
1: Initialize $D^{(2)}$ randomly.
2: Sample from the conditional p.m.f. of $D^{(2)}$.
3: Initialize $D^{(1)}$ randomly.
4: While keeping $D^{(2)}$ unchanged, sample from the conditional p.m.f. of $D^{(2,1)}$.
5: Initialize $D^{(0)}$ randomly.
6: While keeping $D^{(1)}$ unchanged, sample from the conditional p.m.f. of $D^{(1,0)}$.
```

Keeping the upper layer unchanged while sampling from the MRF modeling the correlation between two layers allows long range properties to propagate from coarse scales to fine scales. The multiscale MRF model is very good at characterizing both short and long range interactions between states as shown in Figure 4.4. Compared to Figure 4.1 a), b) and c), particle distributions in Figure 4.4 c) are visually very similar to those in the experimental sample images and the artifacts due to the use of the relatively small neighborhood system are minimal. There are 10 images in our sample dataset, which is adequate to obtain good estimates of the conditional probabilities as shown in Figure 4.4.

## 4.2 Parameter Estimation: a (4+1)-Neighborhood System

In order to apply the multiscale MRF model to data where there are three proteins, the parameters of the hidden MRF with a (4+1)-neighborhood system first need to be estimated from the visible fields. A strategy similar to that described in Section 3.3 is adopted to solve the ML estimation problem. Figure 4.5 indicates how sites are numbered in a (4+1)-neighborhood system. Using this numbering scheme, a six-tuple $\langle n, m, l, k, j, i \rangle$ and a five-tuple $\langle m, l, k, j, i \rangle$ can be used to denote the joint events

$$(D_5^{(u,w)} = n, D_4^{(u,w)} = m, D_3^{(u,w)} = l, D_2^{(u,w)} = k, D_1^{(u,w)} = j, D_0^{(u,w)} = i)$$

and

$$(D_4^{(u,w)} = m, D_3^{(u,w)} = l, D_2^{(u,w)} = k, D_1^{(u,w)} = j, D_0^{(u,w)} = i)$$

respectively, where $u$ denotes an upper layer and $w$ denotes a lower layer. The conditional probability $P(D_5^{(u,w)} = n | D_4^{(u,w)} = m, D_3^{(u,w)} = l, D_2^{(u,w)} = k, D_1^{(u,w)} = j, D_0^{(u,w)} = i)$ can then be estimated:

$$\begin{aligned} P(D_5^{(u,w)} = n | D_4^{(u,w)} = m, D_3^{(u,w)} = 1, D_2^{(u,w)} = k, D_1^{(u,w)} = j, D_0^{(u,w)} = i) & \\ = & p_{nmlkji} \\ \approx & \frac{H(\langle n, m, l, k, j, i \rangle)}{H(\langle m, l, k, j, i \rangle)} \end{aligned} \tag{4.1}$$

where $H$ are histograms of the tuples. The probability of a realization of the distribution process, $d^{(u,w)}$, given $\mathbf{P}$ is also approximated by a product of the conditional probabilities:

$$\mathrm{P}(d^{(u,w)} | \mathbf{P}) \approx \prod_{s \in S^{(w)}} \mathrm{P}(d_s^{(u,w)} | d_t^{(u,w)}, t \in \mathcal{N}_s) \tag{4.2}$$

where $S^{(w)}$ is the set of sites in the lower layer. The probability of a realization of the observation process, $o_k^{(u,w)}$, given $\mathbf{P}$ and $\mathbf{Q}_k$ is approximated as:

$$\mathrm{P}(o_k^{(u,w)} | \mathbf{P}, \mathbf{Q}_k) \approx \prod_{s \in S^{(w)}} \mathrm{P}((o_k^{(u,w)})_s | (o_k^{(u,w)})_t, t \in \mathcal{N}_s) \tag{4.3}$$

where

$$\begin{aligned}&\mathrm{P}((o_k^{(u,w)})_s|(o_k^{(u,w)})_t, t \in \mathcal{N}_s)\\ &= \sum_{d_t^{(u,w)}, t\in\mathcal{N}_s} \frac{\mathrm{P}(d_s^{(u,w)}|d_t^{(u,w)}, t \in \mathcal{N}_s)\mathrm{P}(d_t^{(u,w)}, t \in \mathcal{N}_s) \prod_{t\in\{s\}\cup\mathcal{N}_s} \mathrm{P}((o_k^{(u,w)})_t|d_t^{(u,w)})}{\mathrm{P}((o_k^{(u,w)})_t, t \in \mathcal{N}_s)}\end{aligned} \tag{4.4}$$

and where $k \in \{r, g, b\}$. $P((o_k^{(u,w)})_s|d_s^{(u,w)})$ are known and defined in $\mathbf{Q}_k$. $P((o_k^{(u,w)})_t, t \in \mathcal{N}_s)$ are the probabilities of symbol five-tuples in the visible fields and can be estimated from the visible samples. $P(d_t^{(u,w)}, t \in \mathcal{N}_s)$ are the probabilities of state five-tuples in the hidden field. In Eq. 4.4, $P(d_s^{(u,w)}|d_t^{(u,w)}, t \in \mathcal{N}_s)$ and $P(d_t^{(u,w)}, t \in \mathcal{N}_s)$ are unknown variables. The maximum likelihood estimation of $\mathbf{P}$ is achieved by maximizing the logarithm of the probability that the process generated the visible samples as a function of $P(d_s^{(u,w)}|d_t^{(u,w)}, t \in \mathcal{N}_s)$ and $P(d_t^{(u,w)}, t \in \mathcal{N}_s)$:

$$\hat{\mathbf{P}} = \arg\max_{\mathbf{P}} \sum_{k\in\{r,g,b\}} \log \mathrm{P}(o_k^{(u,w)}|\mathbf{P}, \mathbf{Q}_k). \tag{4.5}$$

Isotropy is only enforced within layers because there is no corresponding symmetry between layers. The number of unknown conditional probabilities in $\mathbf{P}$ is 802. In addition, there are 181 unknown probabilities of state five-tuples. The total number of unknown variables is 983. As in Section 3.3, a quadtree of state tuples is built. The five-tuples are leaves in the quadtree. The probabilities of the five-tuples are constrained by

$$\sum_{\langle m,l,k,j,i\rangle\in\mathcal{C}(a)} P(\langle m, l, k, j, i\rangle) = P(a) \tag{4.6}$$

where $a$ is the nearest known ancestor of the five-tuples and $\mathcal{C}(a)$ is the set of leaves of the quadtree which are on the branch which starts at $a$. The problem to be solved is maximize:

$$\begin{aligned} f &= \sum_k \log \mathrm{P}(o_k^{(u,w)}|\mathbf{P},\mathbf{Q}_k) \\ &= \sum_k \sum_{(o_k^{(u,w)})_t, t\in\mathcal{N}_s} H_k((o_k^{(u,w)})_t, t\in\mathcal{N}_s)\ \log\ \mathrm{P}((o_k^{(u,w)})_s|(o_k^{(u,w)})_t, t\in\mathcal{N}_s) \end{aligned} \tag{4.7}$$

subject to:

(i) $0 \leq P(d_s^{(u,w)}|d_t^{(u,w)}, t\in\mathcal{N}_s) \leq 1;$

(ii) $0 \leq P(d_t^{(u,w)}, t\in\mathcal{N}_s) \leq 1;$

(iii) $\forall d_t^{(u,w)}, t\in\mathcal{N}_s \quad \sum_{d_s^{(u,w)}} P(d_s^{(u,w)}|d_t^{(u,w)}, t\in\mathcal{N}_s) = 1;$

(iv) constraint defined by Eq. 4.6.

The above problem is also solved using SNOPT.

Seven experiments were performed to exam the quality of estimation. A conditional distribution for a MRF with a (4+1)-neighborhood system was first specified. 100 realizations of the MRF were generated on a two layer resolution pyramid, *i.e.* a coarse lattice of size $300 \times 370$ above a fine lattice of size $600 \times 740$, by sampling from the conditional distribution using the Gibbs sampler. For each realization, three visible samples were obtained by mapping the hidden states in the realization to the observation symbols using the observation matrices. By means of a raster scan with a windows (see Figure 4.5) over the visible samples, the frequencies of the five-tuples and six-tuples of observation symbols were computed. To estimate the conditional probabilities, Eq. 4.7 was used to create an objective function and maximization subject to the constraints described above was solved by SNOPT.

Table 4.1 shows relative errors of estimates in terms of matrix 2-norm for the seven experiments. Like the MRF model with a 4-neighborhood system, there are errors in the estimates. When the number of samples are increased by a factor of ten, similar errors were also observed. It is very likely that the errors are due to the fact that the visible fields are consistent with many different hidden distributions. As demonstrated in Section 4.3 and Section 4.4, although there were errors in estimates of the conditional probabilities of the MRF with a (4+1)-neighborhood system, the multiscale MRF model worked well when applied to both synthesized and experimental data.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $\frac{\|\hat{\mathbf{P}}-\mathbf{P}\|_2}{\|\mathbf{P}\|_2}(\times 10^2)$ | 44.1 | 33.7 | 29.9 | 35.6 | 35.9 | 45.3 | 33.3 |

Table 4.1: Relative errors of estimates for MRFs with a (4+1)-neighborhood system.

## 4.3 Results on Synthesized Data

The multiscale MRF model has been applied to synthesized data where the grand truth is known. The conditional probabilities of $D^{(2)}$, $D^{(2,1)}$, and $D^{(1,0)}$ are first specified. A Gibbs sampler is then used to generate samples from the conditional distributions by means of the coarse-to-fine process described in Section 4.1. The three scales used to generate the samples are 80, 40 and 20 pixels. The coordinates of all three proteins at the finest layer in the samples are exported as data. The coordinates of $R$ are eliminated to generate a set of data where $R$ is missing. In an analogous way, a set of data where $G$ is missing and a set of data where $B$ is missing are obtained. The data sets are processed at the same three scales used to generate them. The data sets at the scale of 80 pixels consists of samples of $O_k^{(2)}$ and are used to estimate the conditional p.m.f. of $D^{(2)}$. The data sets at the scales of 80 and 40 pixels are used to estimate the conditional p.m.f. of $D^{(2,1)}$. The data sets at the scales of 40 and 20 pixels are used to estimate the conditional p.m.f. of $D^{(1,0)}$. Finally, the estimates of the conditional probabilities of $D^{(2)}$, $D^{(2,1)}$, and $D^{(1,0)}$ are used to generate samples using Gibbs sampling in the same coarse-to-fine manner described previously.

This produces satisfactory results. As shown in Figure 4.6, the samples of the finest layer generated with specified and estimated conditional probabilities are visually very similar.

## 4.4 Results on Experimental Data

As a critical test, the multiscale MRF model has been applied to two sets of experimental data obtained by immunogold labeling of membrane sheets. These were prepared from rat basophilic leukemia 2H3 (RBL-2H3) mast cells, which express the high affinity IgE receptor, Fc$\epsilon$RI. Crosslinking this receptor with multivalent ligand activates a complex, multicomponent tyrosine kinase-dependent signaling pathway leading to the release of histamine and other mediators of allergic and asthmatic responses. Early events in the Fc$\epsilon$RI signaling cascade include receptor redistribution into large clusters, the recruitment of both membrane-bound and cytoplasmic signaling proteins to receptor-rich domains and also the segregation of certain tyrosine phosphorylated scaffolding and signaling proteins away from receptor-rich domains [42, 43, 44].

The first experiment involves three signaling species in rat basophilic leukemia cell line 2H3 mast cells: high-affinity IgE receptor $\beta$ subunit ($\beta$), linker for activation of T cells (LAT), and phospholipase $C\gamma$ isoform 1 ($\mathrm{PLC}\gamma 1$). It is known that $\mathrm{PLC}\gamma 1$ colocalizes with LAT and colocalizes loosely with $\beta$, but LAT occurs in small clusters separate from $\beta$ [44]. There are three sets of double-labeled data in the first experiment: $\mathrm{PLC}\gamma 1$ labeled with 5nm particles and LAT labeled with 10nm particles in the first data set (Figure 4.7 a)), LAT labeled with 5nm particles and $\beta$ labeled with 10 nm particles in the second set (Figure 4.7 b)), and $\mathrm{PLC}\gamma 1$ labeled with 5nm particles and $\beta$ labeled with 10nm particles in the third set (Figure 4.7 c)). There are 10 images in each data set. We computed histograms of the distances of the nearest neighbor for every particle in the data sets and found that the most frequent distance is 16 pixels. Consequently, windows of size $16 \times 16$, $32 \times 32$, and $64 \times 64$ pixels are used to estimate the histograms of symbols at the three scales. LAT, $\mathrm{PLC}\gamma 1$, and $\beta$ are assigned to $R$, $G$, and $B$ respectively. In Figure 4.7 d), the sample reconstructed from the modeling demonstrates both the colocalization of PLC$\gamma$1 (green) with LAT (red) and the segregation of LAT/ PLC$\gamma$1 from $\beta$ (blue) in a single integrated image.

In the second experiment, we again chose the Fc$\epsilon$RI $\beta$ subunit but added two different signaling species, Grb2-binding protein 2 (Gab2) and the p85 subunit of phosphatidylinositol 3-kinase (p85 of PI 3-kinase). Gab2 is an adaptor protein and PI 3-kinase is an enzyme that phosphorylates phosphatidylinositol lipids in the 3 position on the inositol ring. Previous double-label studies indicated that Gab2 and p85 are both recruited to receptor-rich domains in activated RBL-2H3 cells [44]. This dataset was also composed of three double-label protocols, with 10 images each: 1) Gab2 was labeled with 5nm particles and $\beta$ was labeled with 10nm particles (Figure 4.8 a)); 2) Gab2 was labeled with 5nm particles and p85 was labeled with 10nm particles (Figure 4.8 b)); and 3) p85 was labeled with 5nm particles and $\beta$ was labeled with 10nm particles (Figure 4.8 c)). The data are processed at scales of 20, 40, and 80 pixels. In Figure 4.8 d), the sample reconstructed from the modeling confirms the colocalization of p85 (red), Gab2 (green), and $\beta$ (blue) in a single integrated image.

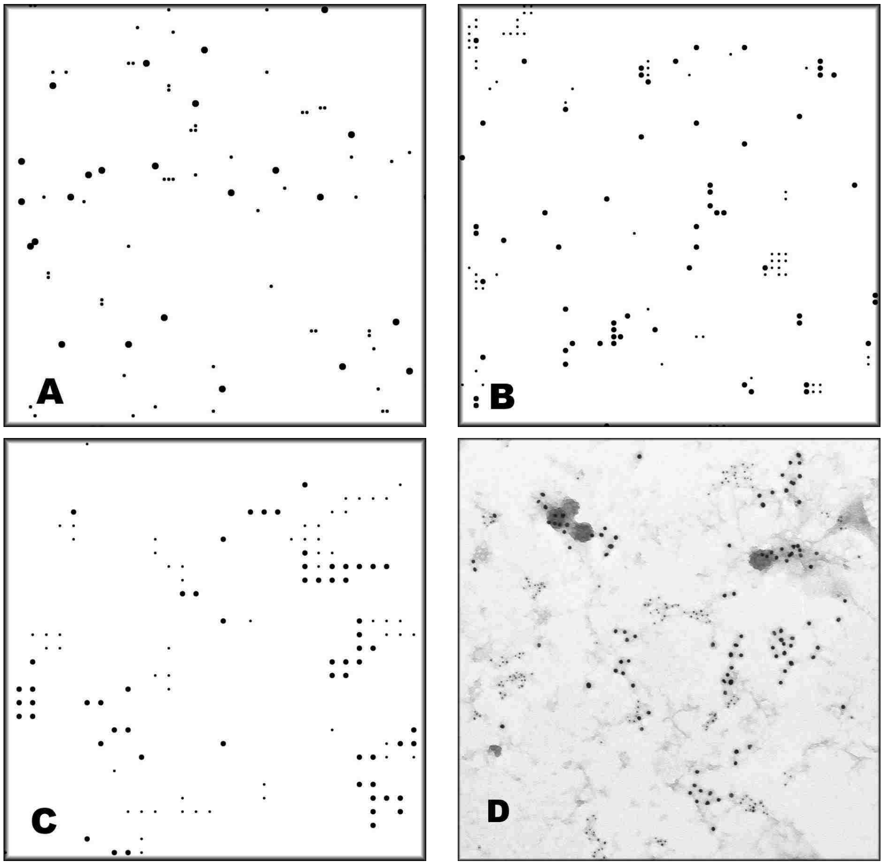

Figure 4.1: Non-hierarchical MRF modeling on experimental data with two proteins. The first protein is a glycosylphosphatidylinositol-linked protein, Thy-1 (labeled with 10nm particles), and the second is a linker for activation of T cells, LAT (labeled with 5nm particles). The MRF model has a 4-neighborhood system. Because only two proteins are involved, the frequencies of the four-tuples and five-tuples of states can be computed directly from the observed data. The conditional probabilities of the MRF are computed using Eq. 3.21. A Gibbs sampler is used to generate samples from the conditional distribution of the MRF. The reconstruction depends on the grid size used to process the data. Grid sizes of 20, 40 and 80 pixels are used to process and reconstruct samples in a), b) and c). A real TEM image is shown in d).

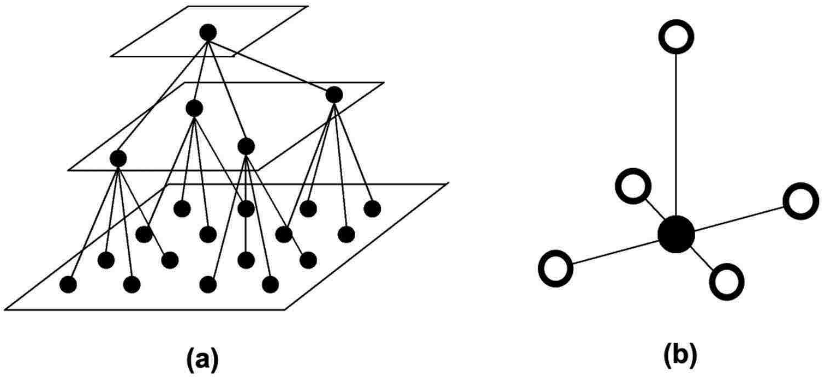

Figure 4.2: A multiscale MRF model and a (4+1)-neighborhood system.

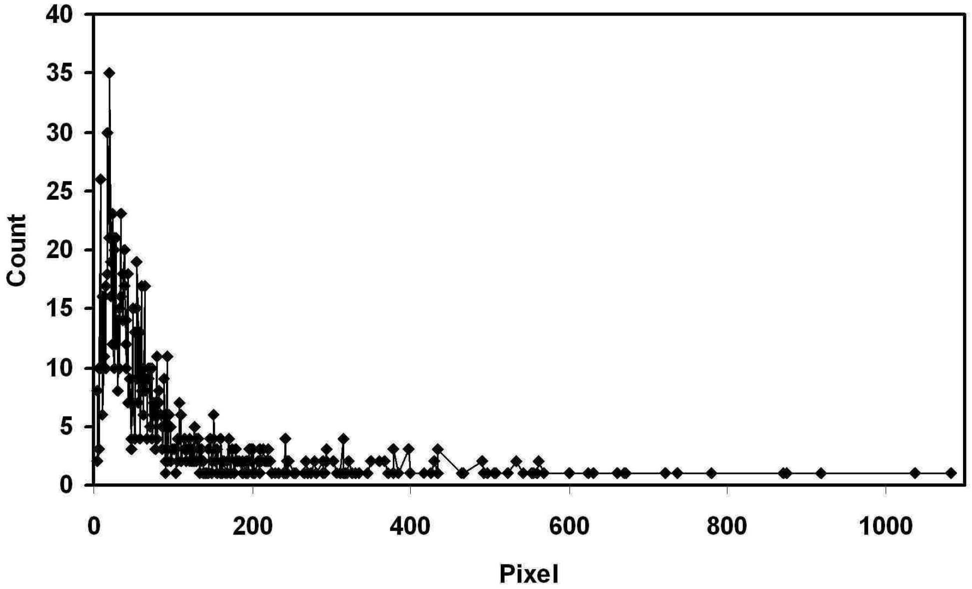

Figure 4.3: Nearest neighbor distance histogram for the images in Figure 4.4.

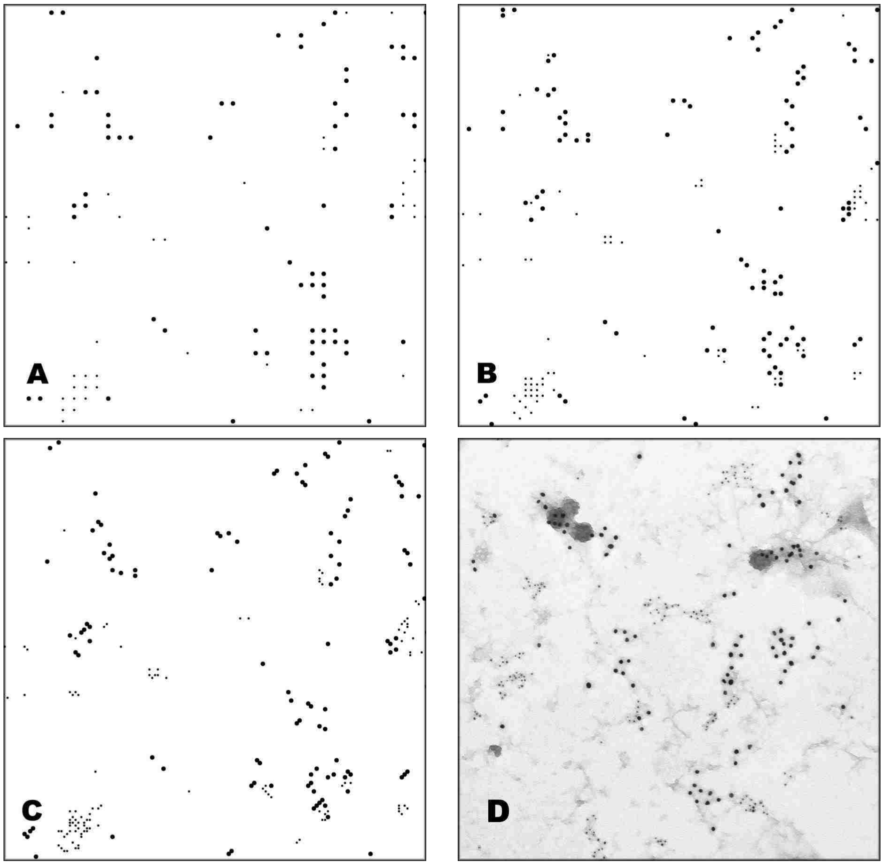

Figure 4.4: Multiscale MRF modeling on experimental data with two proteins. The first protein is a glycosylphosphatidylinositol-linked protein, Thy-1 (labeled with 10nm particles), and the second is a linker for activation of T cells, LAT (labeled with 5nm particles). a) A sample of $D^{(2)}$ generated from the conditional p.m.f. of $D^{(2)}$; b) a sample of $D^{(1)}$ after sampling from the conditional p.m.f. of $D^{(2,1)}$; c) a sample of $D^{(0)}$ after sampling from the conditional p.m.f. of $D^{(1,0)}$; and d) a real TEM image.

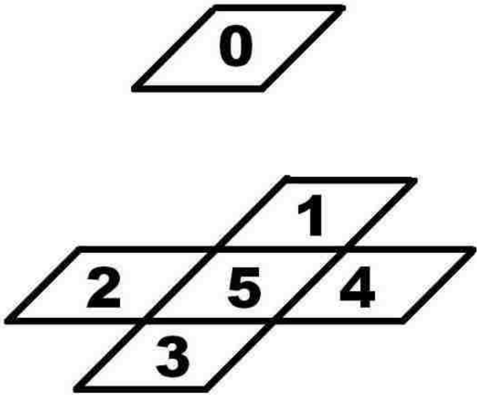

Figure 4.5: Numbering of the sites in a (4+1)-neighborhood system.

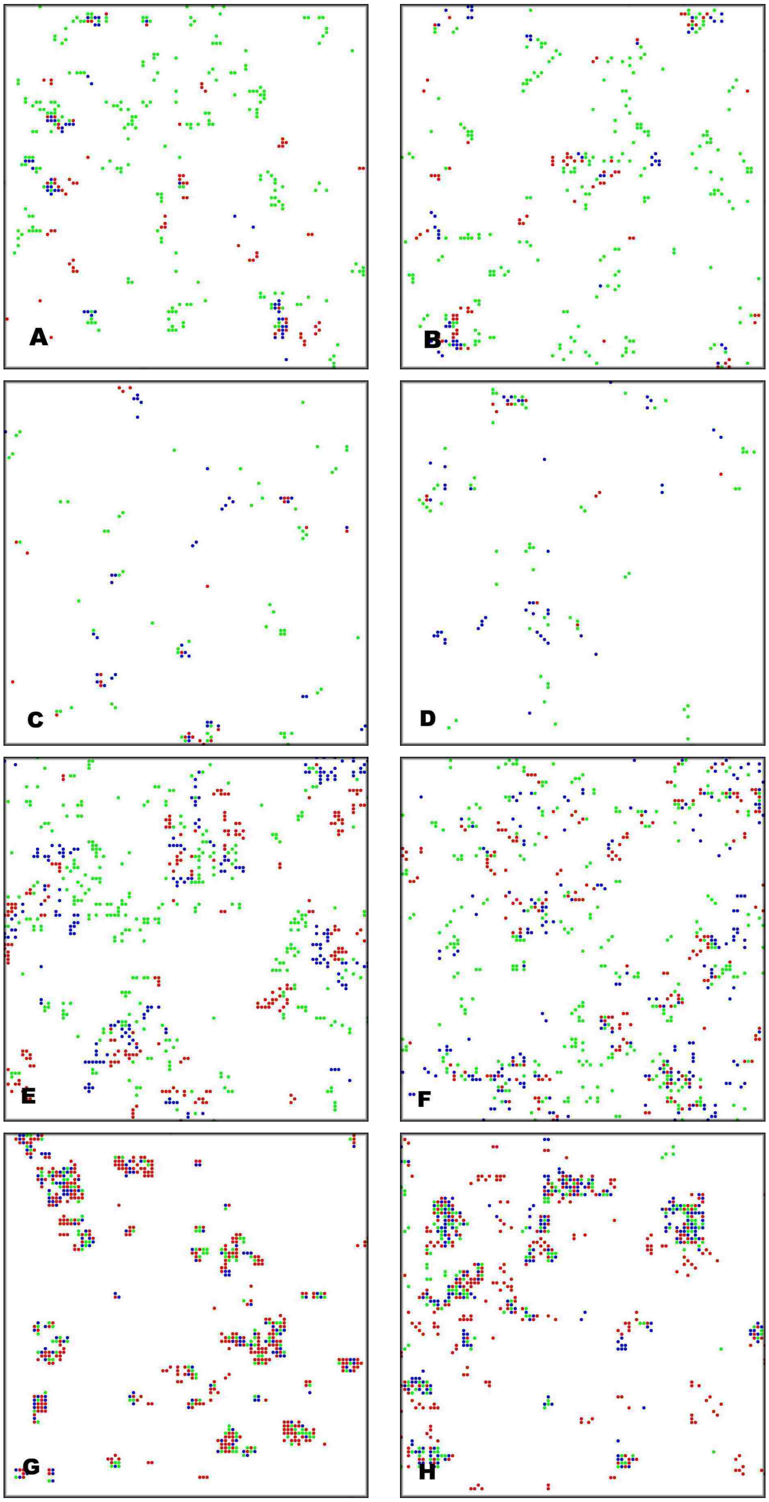

Figure 4.6: Realizations of the finest layer of the multiscale MRF model generated with specified (left) and estimated (right) conditional distributions.

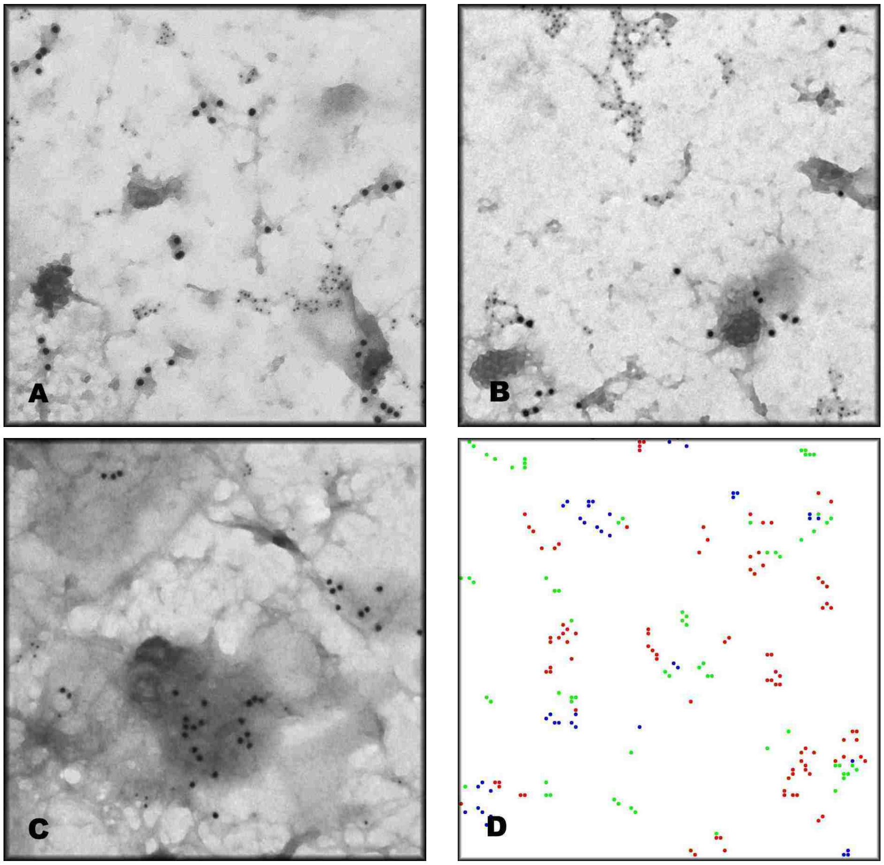

Figure 4.7: Multiscale MRF modeling for the first experiment involving three proteins where only two are observable in any single sample. a) TEM image where $\mathrm{PLC}\gamma 1$ (5nm, green) and LAT (10nm, red) are observed; b) TEM image where LAT (5nm, red) and $\beta$ (10nm, blue) are observed; c) TEM image where $\mathrm{PLC}\gamma 1$ (5nm, green) and $\beta$ (10nm, blue) are observed; and d) a reconstruction computed using Gibbs sampling that shows the distributions of all three proteins.

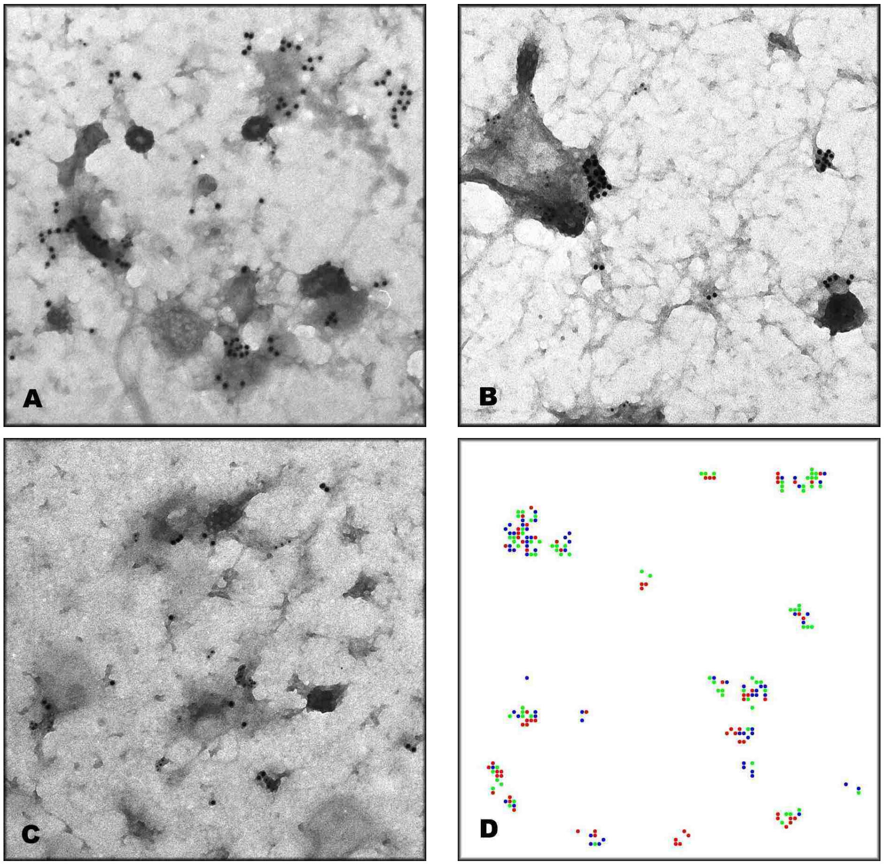

Figure 4.8: Multiscale MRF modeling for the second experiment involving three proteins where only two are observable in any single sample. a) TEM image where Gab2 (5nm, green) and $\beta$ (10nm, blue) are observed; b) TEM image where Gab2 (5nm, green) and p85 (10nm, red) are observed; c) TEM image where p85 (5nm, red) and $\beta$ (10nm, blue) are observed; and d) a reconstruction computed using Gibbs sampling that shows the distributions of all three proteins.

# Chapter 5

# Gibbs Random Field Models

In all of the experiments that have been performed so far, MRFs with a 4-neighborhood system have been used. A larger neighborhood system will improve the quality of the reconstruction computed using Gibbs sampling. However, as the size of the neighborhood system increases, the number of unknown variables becomes too large to be solved using SNOPT. In this chapter, the Markov-Gibbs equivalence is exploited to utilize a larger neighborhood system in the modeling. The Gibbs formulation will be used to compute the conditional probability for a MRF. By using the Gibbs formulation, the number of unknown variables can be reduced. Consequently, a 8-neighborhood system can be used in the modeling to improve the quality of reconstruction, as opposed to the 4-neighborhood system that has been used in the previous experiments.

## 5.1 Gibbs Random Field

As mentioned in Section 2.1, Gibbs random fields and Markov random fields are formally equivalent by the Hammersley-Clifford theorem. A GRF is defined in terms of functions termed *clique potentials*. A *clique* $C$ associated with a lattice $S$ with a neighborhood system $\mathcal{N}$ is a subset of sites in $S$ which satisfies either of the following conditions:

- $C$ consists of a single site
- Every pair of distinct sites in $C$ are neighbors, ***i.e.***, if $s, r \in C$ and $s \neq r$ then $s \in \mathcal{N}_r$ and $r \in \mathcal{N}_s$.

A random field $X$ is said to be a GRF on $S$ with respect to $\mathcal{N}$ if and only if its joint distribution is of the following form

$$P(X = x) = \frac{1}{Z}\exp(-U(x)) \tag{5.1}$$

where

$$Z = \sum_x \exp(-U(x)) \tag{5.2}$$

is a normalizing constant called the *partition function* and $U(x)$ is the *energy function*. The energy

$$U(x) = \sum_C V_C(x) \tag{5.3}$$

is a sum of clique potentials $V_C(x)$ over all possible cliques. The clique potential depends only on the values of the random variables associated with sites in the clique. Figure 5.1 a) and b) shows the cliques for the 4-neighborhood and 8-neighborhood systems.

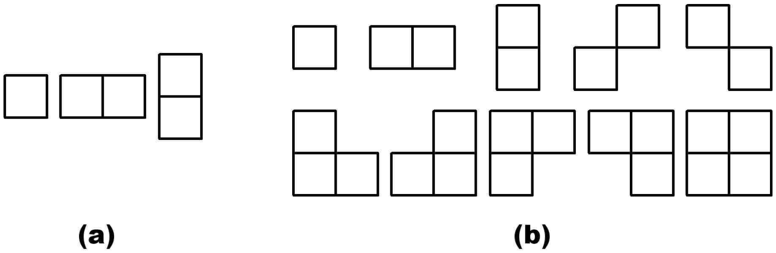

Figure 5.1: Cliques for the a) 4-neighborhood and b) 8-neighborhood systems [26].

A GRF is *homogeneous* if $V_C(x)$ doesn't depend on the relative position of the clique $C$ in $S$. It is said to be *isotropic* if $V_C$ is independent of the orientation of $C$. A homogeneous and isotropic GRF is much simpler to specify than one without such properties. Homogeneity is assumed in most MRF image modeling for the sake of mathematical and computational convenience. Isotropy is assumed in the problem treated here because the spatial distributions of proteins on cell membranes are independent of orientation.

$P(X = x)$ is the probability of a particular configuration, $x$. When the energy of the configuration, $U(x)$, is lower, it is more likely. By exploiting the Markov-Gibbs equivalence, we can write the conditional probability in terms of clique potentials:

$$P(X_s = x_s | X_t, t \in \mathcal{N}_s) = \frac{\exp[-\sum_{C \in C_s} V_C(x_s; x_t, t \neq s, t \in C)]}{\sum_{y_s \in \mathcal{V}} \exp[-\sum_{C \in C_s} V_C(y_s; x_t, t \neq s, t \in C)]} \tag{5.4}$$

where $C_s$ is the set of cliques of $S$ which contains $s$, $\mathcal{V}$ is the set of values that a random variable can take. As shown in Figure 5.1b, a 8-neighborhood system has 1-cliques, 2-cliques, 3-cliques and 4-cliques. Using the Gibbs formulation, we only need to estimate 4-dimensional joint probability mass functions rather than 9-dimensional joint probability mass functions. By exploiting isotropy, the number of unknown variables can be further reduced, so that our MRF model can be based on a 8-neighborhood system.

## 5.2 A GRF Model: 4-Neighborhood Systems

In this section, a GRF model with a 4-neighborhood system is formulated. The GRF model is defined in term of clique potentials. The mathematical programming techniques used in Chapter 3 and 4 are utilized to infer the clique potentials of a hidden random field for three proteins from visible random fields where only two proteins are observable.

### 5.2.1 Clique Potentials for a 4-Neighborhood System

GRFs with a 4-neighborhood system are used to model both the distribution and observation processes. As shown in Figure 5.1a, a 4-neighborhood system has only 1-cliques and 2-cliques. Because there are four possible states, the number of state and clique type combinations is 36. In other words, 36 clique potentials are required to define a GRF with a 4-neighborhood system. Under the assumption of isotropy, the number of unique clique potentials can be reduced to 14 as shown in Table 5.1. Compared with 218 unknowns for a MRF model with a 4-neighborhood system, the number of unknown parameters is reduced dramatically.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\mathrm{C}_R$ | $\mathrm{C}_G$ | $\mathrm{C}_B$ | $\mathrm{C}_X$ | $\mathrm{C}_{RR}$ | $\mathrm{C}_{RG}$ | $\mathrm{C}_{RB}$ | $\mathrm{C}_{RX}$ | $\mathrm{C}_{GR}$ | $\mathrm{C}_{GG}$ |
| $v_1$ | $v_2$ | $v_3$ | $v_4$ | $v_5$ | $v_6$ | $v_7$ | $v_8$ | $v_6$ | $v_9$ |
| $\mathrm{C}_{GB}$ | $\mathrm{C}_{GX}$ | $\mathrm{C}_{BR}$ | $\mathrm{C}_{BG}$ | $\mathrm{C}_{BB}$ | $\mathrm{C}_{BX}$ | $\mathrm{C}_{XR}$ | $\mathrm{C}_{XG}$ | $\mathrm{C}_{XB}$ | $\mathrm{C}_{XX}$ |
| $v_{10}$ | $v_{11}$ | $v_7$ | $v_{10}$ | $v_{12}$ | $v_{13}$ | $v_8$ | $v_{11}$ | $v_{13}$ | $v_{14}$ |

Table 5.1: Fourteen clique potentials for a GRF with a 4-neighborhood system.

Like the clique potentials of Gibbs random fields used in image segmentation and texture models [11, 14], the potentials of 1-cliques control the percentage sites in given states in the Gibbs random field, that is the marginal distribution of the random variable, while the potentials of 2-cliques, 3-cliques, and 4-cliques control the higher-order statistics.

### 5.2.2 Parameter Estimation

The distribution process is represented by a GRF with a 4-neighborhood system that is defined by the clique potentials presented in Table 5.1. However, in order to generate a realization of the distribution process using the Gibbs sampler, the conditional probabilities of the corresponding MRF are needed. The conditional probabilities are computed as follows:

$$P(D_s = d_s | D_t, t \in \mathcal{N}_s) = \frac{\exp[-V(d_s) - \sum_{t \in \mathcal{N}_s} V_2(d_s, d_t)]}{\sum_{d_s \in \{R,G,B,X\}} \exp[-V(d_s) - \sum_{t \in \mathcal{N}_s} V_2(d_s, d_t)]} \tag{5.5}$$

where $V(d_s)$ and $V_2(d_s, d_t)$ are potentials of the 1-cliques and 2-cliques. Using the conditional probabilities, we can generate samples of the distribution process and observation processes, compute the frequencies of four- and five-tuples, and estimate the conditional probabilities from the samples as described in Section 3.3.

A ML estimation process analogous to that described in Section 3.3 is used to infer the clique potentials of the distribution process. Let $o_k$ be a realization of the observation process $O_k$, and let $\mathbf{V}$ denote a vector of clique potentials, and let $\mathbf{Q}_k$ denote the observation matrices. The conditional probability $P(o_k|\mathbf{V}, \mathbf{Q_k})$ is approximated by the pseudo-likelihood:

$$P(o_k|\mathbf{V}, \mathbf{Q}_k) \approx \prod_{s \in S} P((o_k)_s|(o_k)_t, t \in \mathcal{N}_s). \tag{5.6}$$

The problem which must be addressed next is how to compute $P((o_k)_s|(o_k)_t, t \in \mathcal{N}_s)$ in terms of clique potentials. In Section 3.3, $P((o_k)_s|(o_k)_t, t \in \mathcal{N}_s)$ is computed using Eq. 3.27. We can rewrite this equation as

$$\begin{aligned} &\mathrm{P}((o_k)_s|(o_k)_t, t \in \mathcal{N}_s) \\ &\quad = \sum_{d_t, t \in \mathcal{N}_s} \frac{\mathrm{P}((o_k)_s|d_s)\mathrm{P}(d_s, d_t, t \in \mathcal{N}_s) \prod_{t \in \mathcal{N}_s} \mathrm{P}((o_k)_t|d_t)}{\mathrm{P}((o_k)_t, t \in \mathcal{N}_s)} \end{aligned} \tag{5.7}$$

where

$$\mathrm{P}(d_s, d_t, t \in \mathcal{N}_s) = \mathrm{P}(d_s|d_t, t \in \mathcal{N}_s)\mathrm{P}(d_t, t \in \mathcal{N}_s). \tag{5.8}$$

We note that $\mathrm{P}(d_s, d_t, t \in \mathcal{N}_s)$ are the joint probabilities for a Gibbs random field based on a 4-neighborhood system. As a result, we can use Eq.5.1 to compute the joint distribution for this special Gibbs random field:

$$\mathrm{P}(d_s, d_t, t \in \mathcal{N}_s) = \frac{1}{Z}\exp(-U(d_s, d_t, t \in \mathcal{N}_s)) \tag{5.9}$$

where $Z$ is the partition function and $U(d_s, d_t, t \in \mathcal{N}_s)$ is the energy function of the configuration $\{d_s, d_t, t \in \mathcal{N}_s\}$. Because the lattice of the Gibbs random field is a 4-neighborhood system, and because there are only cliques of size up to two, $U(d_s, d_t, t \in \mathcal{N}_s)$ can be calculated as follows:

$$U(d_s, d_t, t \in \mathcal{N}_s) = V_1(d_s) + \sum_{t \in \mathcal{N}_s} V_2(d_s, d_t) \tag{5.10}$$

where $V_1(d_s)$ and $V_2(d_s, d_t)$ are potentials of the 1-cliques and 2-cliques. Therefore, we can compute $P((o_k)_s|(o_k)_t, t \in \mathcal{N}_s)$ using the following equation:

$$\begin{aligned} &\mathrm{P}((o_k)_s|(o_k)_t, t \in \mathcal{N}_s) \\ &\quad = \sum_{d_t, t \in \mathcal{N}_s} \frac{\mathrm{P}((o_k)_s|d_s)\exp(-V_1(d_s) - \sum_{t \in \mathcal{N}_s} V_2(d_s, d_t)) \prod_{t \in \mathcal{N}_s} \mathrm{P}((o_k)_t|d_t)}{Z \times \mathrm{P}((o_k)_t, t \in \mathcal{N}_s)}. \end{aligned} \tag{5.11}$$

The one problem which remains is how to compute the partition function $Z$. A main difficulty in using GRF models is in the calculation of the partition functions because this requires summing over all possible configurations of random fields. Fortunately, the lattice of the Gibbs random field is small and it is possible to sum over all configurations to compute the partition function:

$$\begin{aligned} Z &= \sum_{\{d_s, d_t, t \in \mathcal{N}_s\}} \exp(-U(d_s, d_t, t \in \mathcal{N}_s)) \\ &= \sum_{\{d_s, d_t, t \in \mathcal{N}_s\}} \exp(-V_1(d_s) - \sum_{t \in \mathcal{N}_s} V_2(d_s, d_t)) \end{aligned} \tag{5.12}$$

Equation 5.11 can be rewritten as:

$$\begin{aligned} &\mathrm{P}((o_k)_s|(o_k)_t, t \in \mathcal{N}_s) \\ &\quad = \sum_{d_t, t \in \mathcal{N}_s} \frac{\mathrm{P}((o_k)_s|d_s)\exp(-V_1(d_s) - \sum_{t \in \mathcal{N}_s} V_2(d_s, d_t)) \prod_{t \in \mathcal{N}_s} \mathrm{P}((o_k)_t|d_t)}{\sum_{\{d_s, d_t, t \in \mathcal{N}_s\}} \exp(-V_1(d_s) - \sum_{t \in \mathcal{N}_s} V_2(d_s, d_t)) \times \mathrm{P}((o_k)_t, t \in \mathcal{N}_s)}. \end{aligned} \tag{5.13}$$

In Eq. 5.13, $\mathrm{P}((o_k)_s|d_s)$ and $\mathrm{P}((o_k)_t|d_t)$ are known and are defined in $\mathbf{Q}_k$. $\mathrm{P}((o_k)_t, t \in \mathcal{N}_s)$ are the probabilities of four-tuples of symbols in the observation processes, and can be estimated from samples of the observation processes. Consequently, the unknown variables are $V_1(d_s)$ and $V_2(d_s, d_t)$. The maximum likelihood estimation of $\mathbf{V}$ is achieved by maximizing the logarithm of the probability that the process generated the visible samples as a function of $V_1(d_s)$ and $V_2(d_s, d_t)$:

$$\hat{\mathbf{V}} = \arg\max_{\mathbf{V}} \sum_{k \in \{r,g,b\}} \log \mathrm{P}(o_k|\mathbf{V}, \mathbf{Q}_k). \tag{5.14}$$

There are also constraints on the clique potentials. Firstly, 1-clique potentials control the marginal distribution of the random field, and these 1-clique potentials can be defined as the logarithm of the corresponding marginal probabilities:

$$V_1(d_s) = -\ln(P(d_s)). \tag{5.15}$$

Therefore, these 1-clique potentials are constrained by

$$\sum_{d_s \in \{R,G,B,X\}} \exp(-V_1(d_s)) = 1. \tag{5.16}$$

Secondly, as described in Section 3.3, we can directly estimate the conditional probabilities that involve at most two non-background states from the samples of the observation processes. These known conditional probabilities impose constraints on the clique potentials:

$$P(d_s|d_t, t \in \mathcal{N}_s) = \frac{\exp[-V_1(d_s) - \sum_{t \in \mathcal{N}_s} V_2(d_s, d_t)]}{\sum_{d_s \in \{R,G,B,X\}} \exp[-V_1(d_s) - \sum_{t \in \mathcal{N}_s} V_2(d_s, d_t)]} \tag{5.17}$$

where $\mathrm{P}(d_s|d_t, t \in \mathcal{N}_s)$ are the conditional probabilities that involve at most two non-background states. Therefore, the fourteen unknown clique potentials are constrained by

Eq. 5.17. The objective function is

$$\begin{aligned} f &= \sum_k \log \mathrm{P}(o_k|\mathbf{V},\mathbf{Q}_k) && (5.18)\\ &= \sum_k \log \left[ \prod_{s\in S} \mathrm{P}((o_k)_s|(o_k)_t, t\in\mathcal{N}_s) \right] && (5.19)\\ &= \sum_k \sum_{(o_k)_t, t\in\mathcal{N}_s} H_k((o_k)_t, t\in\mathcal{N}_s) \log \mathrm{P}((o_k)_s|(o_k)_t, t\in\mathcal{N}_s) && (5.20) \end{aligned}$$

where $k \in \{r, g, b\}$ and $\mathrm{P}((o_k)_s|(o_k)_t, t\in\mathcal{N}_s)$ is defined in Eq. 5.13. $H_k((o_k)_t, t\in\mathcal{N}_s)$ is the observed frequency of the four-tuple of symbols. The problem we need to solve is

maximize: $f$
subject to:

(i) the constraint defined by Eq. 5.16;

(ii) the constraints defined by Eq. 5.17.

SNOPT is used to solve the above problem also.

### 5.2.3 Results for the GRF Model: 4-Neighborhood Systems

To test the ML estimation for the GRF with a 4-neighborhood system, a distribution of clique potentials is first specified. The conditional probabilities of the corresponding MRF are then computed using Eq. 5.5. One hundred realizations of the GRF were generated on a lattice of size $600 \times 740$ by sampling from the conditional distribution using the Gibbs sampler. For each realization, three visible samples were obtained by mapping the hidden states in the realization to the observation symbols using the observation matrices. By means of a raster scan (see Figure 3.3) over the visible samples, the frequencies of the four-tuples and five-tuples of observation symbols were computed. 108 conditional probabilities are estimated from these frequencies, and these were used to compute the clique potentials: Eq. 5.20 was used to create an objective function and maximization subject to the constraints defined by Eq. 5.16 and Eq. 5.17 was solved by SNOPT. This results in good estimates which generate realizations which are visually very similar to those generated using the specified clique potentials (See Figure 5.2). The relative errors of estimates are shown in Table 5.2.

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $\frac{\|\hat{\mathbf{V}}-\mathbf{V}\|_2}{\|\mathbf{V}\|_2}(\times 10^2)$ | 3.2 | 13.5 | 20.0 | 5.7 |
| Figure | 5.2 a) | 5.2 b) | 5.2 c) | 5.2 d) |

Table 5.2: Relative errors of estimates for the GRF modeling for 4-neighborhood systems.

## 5.3 A GRF Model: 8-Neighborhood Systems

This section investigates whether the strategy described in Section 5.2 also works for Gibbs random field models with a 8-neighborhood system.

As shown in Figure 5.1b, a 8-neighborhood system has cliques of size up to four. Since there are four possible states, the number of state and 1-clique type combination is 4. The numbers of state and clique type combination are 64, 256 and 256 for 2-cliques, 3-cliques and 4-cliques respectively. Therefore, 580 clique potentials are needed to define a GRF with a 8-neighborhood system. This number is already reduced dramatically, compared with 19400 unknown variables for a MRF with a 8-neighborhood system. By assuming isotropy, the number of unique clique potentials can be further reduced. For examples, the vertical, horizontal and diagonal 2-cliques are all equal under the assumption of isotropy. The number of unique clique potentials required to define a GRF with a 8-neighborhood system can be reduced to 69 under the isotropic assumption.

The conditional probabilities of the corresponding MRF with a 8-neighborhood system can be computed in terms of clique potentials as follows:

$$P(D_s = d_s | D_t = d_t, t \in \mathcal{N}_s) = \frac{\exp(-U(d_s))}{\sum_{d_s \in \{R,G,B,X\}} \exp(-U(d_s))} \tag{5.21}$$

where

$$\begin{aligned} U(d_s) = -V_1(d_s) - \sum_{t \in \mathcal{N}_s} V_2(d_s, d_t) - \sum_{t,u \in \mathcal{N}_s} V_3(d_s, d_t, d_u) \\ - \sum_{t,u,v \in \mathcal{N}_s} V_4(d_s, d_t, d_u, d_v) \end{aligned} \tag{5.22}$$

and where $V_1(d_s)$, $V_2(d_s, d_t)$, $V_3(d_s, d_t, d_u)$, and $V_4(d_s, d_t, d_u, d_v)$ are potentials of the 1-cliques, 2-cliques, 3-cliques and 4-cliques respectively. Using the conditional probabilities, we can generate samples of the distribution process and observation processes, and can compute the frequencies of eight- and nine-tuples of observable symbols, and can estimate the conditional probabilities from these samples as before.

The mathematical programming technique used to estimate the clique potentials of a GRF with a 8-neighborhood system is analogous to that used for the GRF with a 4-neighborhood system. To approximate the conditional probability $P(o_k|\mathrm{V}, \mathrm{Q_k})$ using pseudo likelihood, the joint probabilities of a Gibbs random field with a 8-neighborhood system are calculated using Eq. 5.9. The energy function is then computed using Eq. 5.22. The partition function is computed using the following equation:

$$\begin{aligned} Z = \sum \exp[-V_1(d_s) - \sum_{t \in \mathcal{N}_s} V_2(d_s, d_t) - \sum_{t,u \in \mathcal{N}_s} V_3(d_s, d_t, d_u) \\ - \sum_{t,u,v \in \mathcal{N}_s} V_4(d_s, d_t, d_u, d_v)] \end{aligned}$$

In addition, there is a constraint on 1-clique potentials

$$\sum_{d_s \in \{R,G,B,X\}} \exp(-V_1(d_s)) = 1. \tag{5.23}$$

Finally the constraints imposed by the conditional probabilities that involve at most two non-background states are defined by Eq. 5.21. The mathematical programming problem is solved by SNOPT.

Four experiments have been performed to test the ML estimation for GRFs with a 8-neighborhood system. Figure 5.3 shows the results for the four experiments. The estimates are good and result in realizations which are visually very close to those generated using the specified clique potentials. The relative errors of estimates are shown in Table 5.3.

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $\frac{\|\hat{\mathbf{V}}-\mathbf{V}\|_2}{\|\mathbf{V}\|_2}(\times 10^2)$ | 2.95 | 34.2 | 22.8 | 10.1 |
| Figure | 5.3 a) | 5.3 b) | 5.3 c) | 5.3 d) |

Table 5.3: Relative errors of estimates for the GRF modeling with a 8-neighborhood system.

## 5.4 A Multiscale GRF Model

In this section, a multiscale GRF model is built. Each layer in this model is modeled as a GRF with a 8-neighborhood system.

As was true of the multiscale MRF model described in Section 4.1, there is a pyramid with three layers in the GRF model. Each site in a coarse layer is a parent of four sites in the fine layer beneath it. The layers from coarse to fine are denoted as $L^{(2)}$, $L^{(1)}$, and $L^{(0)}$. For each layer, there is a corresponding distribution process, $D^{(i)}$ where $i \in \{2, 1, 0\}$. Each of these is modeled as a GRF with a 8-neighborhood system. Furthermore, two additional distribution processes are used to model the correlation between the values of sites in one layer and those in the layer beneath it: $D^{(2,1)}$ between $L^{(2)}$ and $L^{(1)}$, and $D^{(1,0)}$ between $L^{(1)}$ and $L^{(0)}$. Both of these are modeled as GRFs with a (8+1)-neighborhood system, where eight neighbors are in the lower layer and the ninth neighbor is their parent in the layer above (Figure 5.4a). There are also three corresponding observation processes for each of the distribution processes: $O_k^{(2)}$, $O_k^{(1)}$, $O_k^{(0)}$, $O_k^{(2,1)}$, and $O_k^{(1,0)}$ where $k \in \{r, g, b\}$. The value of a site at a given scale depends not only on its parent in the layer above but also on its neighbors at the same scale. Multi-resolution representations of the observed data at three scales are realizations of $O^{(i)}$, $i \in \{2, 1, 0\}$, and data between two scales are realizations of $O^{(2,1)}$ and $O^{(1,0)}$. We can infer the clique potentials of $D^{(i)}$ from $O_k^{(i)}$ using the process described in Section 5.3. Furthermore, if we can estimate the clique potentials of $D^{(2,1)}$ from $O_k^{(2,1)}$, and $D^{(1,0)}$ from $O_k^{(1,0)}$, the Gibbs sampler can be then used to generate sample random fields in a coarse-to-fine manner.

### 5.4.1 Parameter Estimation: a (8+1)-Neighborhood System

In order to apply the multiscale GRF model to experimental data, the clique potentials of the hidden GRF with a (8+1)-neighborhood system first need to be estimated from the visible fields. The cliques of a (8+1)-neighborhood system also have to be defined. As in the MRF multiscale model, isotropy is only enforced within layers. Figure 5.5 shows the cliques for a (8+1)-neighborhood system. The first row shows cliques with sites from only one layer, which are the same as the cliques of the 8-neighborhood system. The second row shows a clique with sites from two layers. There is only one 2-clique with sites from two layers because only site 9 in the lower layer is a neighbor of site 0 in the upper layer. Under the assumption of isotropy, there are 69 unique clique potentials for cliques in the first row of Figure 5.5. For the 2-clique in the second row of Figure 5.5, there are 16 unique clique potentials. As a result, the total number of unique clique potentials needed to define a GRF with a (8+1)-neighborhood system is 85. Using these clique potentials, the conditional probabilities of the corresponding MRF with a (8+1)-neighborhood system can be calculated as before.

A similar strategy to that described in Section 5.2.2 is adopted to solve the ML estimation problem. The sites are numbered in a (8+1)-neighborhood system in the manner shown in Figure 5.4 b). Using this numbering scheme, a ten-tuple $\langle r, q, p, o, n, m, l, k, j, i \rangle$

can be used to denote the joint event that

$$D_9^{(u,w)}, D_8^{(u,w)}, D_7^{(u,w)}, D_6^{(u,w)}, D_5^{(u,w)}, D_4^{(u,w)}, D_3^{(u,w)}, D_2^{(u,w)}, D_1^{(u,w)}, D_0^{(u,w)}$$

take the values of $r, q, p, o, m, l, k, j, i$ respectively, and a nine-tuple $\langle q, p, o, m, l, k, j, i\rangle$ can be used to denote the joint event that

$$D_8^{(u,w)}, D_7^{(u,w)}, D_6^{(u,w)}, D_5^{(u,w)}, D_4^{(u,w)}, D_3^{(u,w)}, D_2^{(u,w)}, D_1^{(u,w)}, D_0^{(u,w)}$$

take the values of $q, p, o, m, l, k, j, i$ respectively, where $u$ denotes an upper layer and $w$ denotes a lower layer. For the sake of notation, the prime symbol is used to denote $(u, w)$. The conditional probability can then be estimated as follows:

$$\begin{aligned} &P(D'_9|D'_8, D'_7, D'_6, D'_5, D'_4, D'_3, D'_2, D'_1, D'_0) \\ &\quad\approx \frac{H(\langle n, m, l, k, j, i\rangle)}{H(\langle m, l, k, j, i\rangle)} \end{aligned} \tag{5.24}$$

where $H$ are histograms of the tuples. The probability of a realization of the observation process, $o'_k$, given $\mathbf{V}$ and $\mathbf{Q}_k$ is approximated as:

$$\mathrm{P}(o'_k|\mathbf{V}, \mathbf{Q}_k) \approx \prod_{s\in S^{(w)}} \mathrm{P}((o'_k)_s|(o'_k)_t, t \in \mathcal{N}_s) \tag{5.25}$$

where

$$\begin{aligned} &\mathrm{P}((o'_k)_s|(o'_k)_t, t \in \mathcal{N}_s) \\ &= \sum_{d'_s, d'_t, t\in\mathcal{N}_s} \frac{\exp(-U(d'_s, d'_t, t \in \mathcal{N}_s)) \prod_{t\in\{s\}\cup\mathcal{N}_s} \mathrm{P}((o'_k)_t|d'_t)}{Z \times \mathrm{P}((o'_k)_t, t \in \mathcal{N}_s)} \end{aligned} \tag{5.26}$$

and where $S^{(w)}$ is the set of sites in the lower layer, and $k \in \{r, g, b\}$. In Eq. 5.26, $U(d'_s, d'_t, t \in \mathcal{N}_s)$ is the energy function of the configuration $\{d'_s, d'_t, t \in \mathcal{N}_s\}$ and $Z$ is the partition function for a Gibbs random field with is a (8+1)-neighborhood system. $U(d'_s, d'_t, t \in \mathcal{N}_s)$ can be computed as follows:

$$\begin{aligned} U(d'_s, d'_t, t \in \mathcal{N}_s) = V_1(d'_s) + \sum_{t\in\mathcal{N}_s} V_2(d'_s, d'_t) + \sum_{t,u\in\mathcal{N}_s} V_3(d'_s, d'_t, d'_u) \\ + \sum_{t,u,v\in\mathcal{N}_s} V_4(d'_s, d'_t, d'_u, d'_v) \end{aligned} \tag{5.27}$$

where $V_1(d'_s)$, $V_2(d'_s, d'_t)$, $V_3(d'_s, d'_t, d'_u)$, and $V_4(d'_s, d'_t, d'_u, d'_v)$ are potentials of the 1-, 2-, 3- and 4-cliques. And the partition function of the Gibbs random field with a (8+1)-neighborhood system can be calculated as follows:

$$\begin{aligned} Z = \sum \exp[-V_1(d'_s) - \sum_{t \in \mathcal{N}_s} V_2(d'_s, d'_t) - \sum_{t,u \in \mathcal{N}_s} V_3(d'_s, d'_t, d'_u) \\ - \sum_{t,u,v \in \mathcal{N}_s} V_4(d'_s, d'_t, d'_u, d'_v)] \end{aligned} \tag{5.28}$$

In Eq. 5.26, Eq. 5.27, and Eq. 5.28, the unknown variables are the clique potentials. The maximum likelihood estimation of $\mathbf{V}$ is achieved by maximizing the logarithm of the probability that the process generated the visible samples as a function of the clique potentials:

$$\hat{\mathbf{V}} = \arg\max_{\mathbf{V}} \sum_{k \in \{r,g,b\}} \log \mathrm{P}(o'_k | \mathbf{V}, \mathbf{Q}_k). \tag{5.29}$$

A constraint on 1-clique potentials also needs to be enforced:

$$\sum_{d'_s \in \{R,G,B,X\}} \exp(-V_1(d'_s)) = 1. \tag{5.30}$$

In addition, as described in Section 3.3, it is possible to directly estimate the conditional probabilities that involve at most two non-background states from the samples of the observation processes. These known conditional probabilities impose the following constraint on the clique potentials:

$$P(d'_s | d'_t, t \in \mathcal{N}_s) = \frac{\exp[-U(d'_s, d'_t, t \in \mathcal{N}_s)]}{\sum_{d_s \in \{R,G,B,X\}} \exp[-U(d'_s, d'_t, t \in \mathcal{N}_s)]} \tag{5.31}$$

where $\mathrm{P}(d'_s | d'_t, t \in \mathcal{N}_s)$ are the conditional probabilities involving at most two non-background states, and $U(d'_s, d'_t, t \in \mathcal{N}_s)$ are defined by Eq. 5.27. Therefore, the 85 unknown clique potentials are constrained by the constraints defined by Eq. 5.31. The objective function is

$$f = \sum_{k} \log \mathrm{P}(o_k' | \mathbf{V}, \mathbf{Q}_k) \tag{5.32}$$

$$= \sum_{k} \log \left[ \prod_{s \in S^{(w)}} \mathrm{P}((o_k')_s | (o_k')_t, t \in \mathcal{N}_s \right] \tag{5.33}$$

$$= \sum_{k} \sum_{(o_k')_t, t \in \mathcal{N}_s} H_k((o_k')_t, t \in \mathcal{N}_s) \log \mathrm{P}((o_k')_s | (o_k')_t, t \in \mathcal{N}_s) \tag{5.34}$$

where $k \in \{r, g, b\}$ and $\mathrm{P}((o_k')_s | (o_k')_t, t \in \mathcal{N}_s)$ is defined in Eq. 5.26. $H_k((o_k')_t, t \in \mathcal{N}_s)$ is the observed frequency of the nine-tuple of symbols. The problem we need to solve is

maximize: $f$
subject to:

(i) the constraint defined by Eq. 5.30;

(ii) the constraints defined by Eq. 5.31.

SNOPT is used to solve this problem also.

Five experiments have been performed to examine the quality of estimation. Clique potentials for a GRF with a (8+1)-neighborhood system were first specified, and the conditional probabilities for the corresponding MRF were computed. 100 realizations of the GRF were generated on a two layer resolution pyramid consisting of a lattice of size $300 \times 370$ above a lattice of size $600 \times 740$ by sampling from the conditional distribution using the Gibbs sampler. For each realization, three visible samples were obtained by mapping the hidden states in the realization to the observation symbols using the observation matrices. By means of a raster scan (see Figure 5.4 b)) over the visible samples, the frequencies of the nine-tuples and ten-tuples of observation symbols were computed. To estimate the clique potentials, Eq. 5.34 was used to create an objective function which was maximized subject to the constraints described above using SNOPT. Table 5.4 shows the relative errors of the estimates in terms of matrix 2-norm for the five experiments. Like the GRF model with a 8-neighborhood system, there are errors in the estimates. It is also likely that the errors are due to the fact that the visible fields are consistent with many different hidden distributions. As demonstrated in Section 5.4.2, although there were errors in the estimates of the clique potentials of the GRF with a (8+1)-neighborhood system, the multiscale GRF model worked well when applied to synthesized and experimental data.

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $\frac{\|\hat{\mathbf{V}}-\mathbf{V}\|_2}{\|\mathbf{V}\|_2}(\times 10^2)$ | 15.8 | 37.1 | 58.5 | 44.4 | 74.2 |

Table 5.4: Relative errors of estimates for GRFs with a (8+1)-neighborhood system.

### 5.4.2 Results on Synthesized Data

The multiscale GRF model is first applied to synthesized data where the ground truth is known. The clique potentials of $D^{(2)}$, $D^{(2,1)}$, and $D^{(1,0)}$ are specified. Using these clique potentials, the conditional probabilities of the corresponding MRFs are computed. A Gibbs sampler is then used to generate samples from the conditional distributions through a coarse-to-fine pass in the manner described in Section 4.1. The three scales used to generate the samples are 80, 40 and 20 pixels. The coordinates of all three proteins at the finest layer in the samples are exported as data. The coordinates of $R$ are eliminated to generate a set of data where $R$ is missing. In an analogous way, a set of data where $G$ is missing and a set of data where $B$ is missing are obtained. The data sets are processed at the same three scales used to generate them. The data sets at scale of 80 pixels are regarded as samples of $O_k^{(2)}$ and are used to estimate the clique potentials of $D^{(2)}$. The data sets at scales of 80 and 40 pixels are used to estimate the clique potentials of $D^{(2,1)}$. The data sets at scales of 40 and 20 pixels are used to estimate the clique potentials of $D^{(1,0)}$. Finally, the estimates of the clique potentials of $D^{(2)}$, $D^{(2,1)}$, and $D^{(1,0)}$ are used to generate samples by means of Gibbs sampling in a coarse-to-fine manner as before. This produces satisfactory results. As shown in Figure 5.6, the samples of the finest layer generated from GRFs with the specified and estimated clique potentials are visually very similar.

### 5.4.3 Results on Experimental Data

The multiscale GRF model has also been applied to the same two sets of experimental data that were used to test the multiscale MRF model. The first experiment involves three signaling species in rat basophilic leukemia cell line 2H3 mast cells: high-affinity IgE receptor $\beta$ subunit ($\beta$), linker for activation of T cells (LAT), and phospholipase $C\gamma$ isoform 1 ($\mathrm{PLC}\gamma 1$). In the second experiment, we are still interested in the Fc$\epsilon$RI $\beta$ subunit but two different signaling species, Grb2-binding protein 2 (Gab2) and the p85 subunit of phosphatidylinositol 3-kinase (p85 of PI 3-kinase) are added. The data are processed in an analogous way as in Section 4.4, except that the multiscale GRF model is used to reconstruct samples from the data. Figure 5.7 a) shows a sample reconstructed from the first experiment data. It also demonstrates both the colocalization of PLC$\gamma$1 (green) with LAT (red) and the segregation of LAT/ PLC$\gamma$1 from $\beta$ (blue) in a single integrated image. Figure 5.7 c) shows a sample reconstructed from the modeling on the second experiment data. It also confirms the colocalization of p85 (red), Gab2 (green), and $\beta$ (blue) in a single integrated image. Compared with samples reconstructed by MRF modeling (Figure 5.7 b) and d)), the cluster sizes in samples generated by GRF modeling data are larger and the samples from the GRF modeling are marginally better than those of the MRF modeling.

### 5.4.4 Comparison of MRF and GRF Modeling Quality

Kullback-Leibler (KL) divergence is a non-symmetric measure of the difference between two probability distributions. In this section, the quality of MRF and GRF modeling is measured in terms of KL divergence computed using samples from the modeling. Synthesized data of MRF modeling with a 4-neighborhood system and GRF modeling with a 8-neighborhood system are used in the computation. It is desirable to use experimental data. However, it is difficult to use them because, firstly, the ground truth is not known, and secondly, the available experimental data are too sparse to accurately estimate the KL divergence of the probability distributions.

Let $P(x)$ represent a true distribution of data. In the case of modeling on synthesized data, this is the distribution defined by the specified parameters. Let $Q(x)$ denote an approximation of $P(x)$. It is the distribution defined by the estimated parameters for the synthesized data. The KL divergence of $Q$ from $P$ is defined to be

$$D_{KL}(P||D) = \sum_x P(x) \log \frac{P(x)}{Q(x)}. \tag{5.35}$$

200 samples are first generated using the specified parameters and $P(x)$ are estimated using these samples. Parameters estimated from MRF/GRF modeling are then used to generated 200 samples. The $Q(x)$ are calculated using these samples. The KL divergence of $Q$ from $P$ is computed using Eq. 5.35. The KL divergences for three sets of samples from MRF and GRF modeling are shown in Table 5.5. It is clear that the larger neighborhood systems in the GRF modeling result in better reconstruction quality in terms of KL divergence.

| | 1 | 2 | 3 |
|---|---|---|---|
| MRF modeling | 0.5012 | 0.4890 | 0.2284 |
| GRF modeling | 0.0653 | 0.0572 | 0.0273 |

Table 5.5: KL divergences for samples of MRF and GRF modeling.

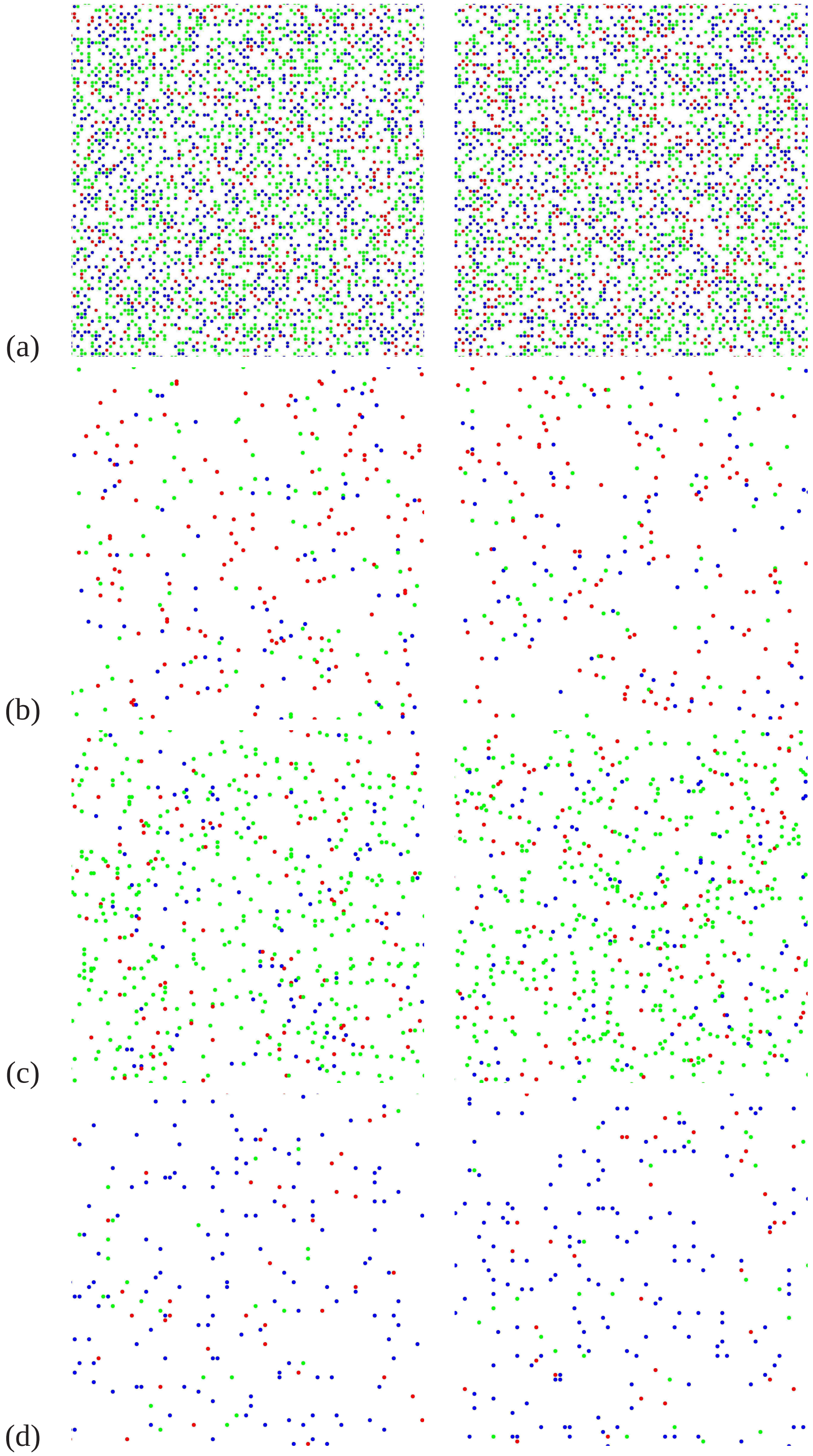

Figure 5.2: Realizations from the GRF model with specified (left) and estimated (right) clique potentials. The GRF model has a 4-neighborhood system.

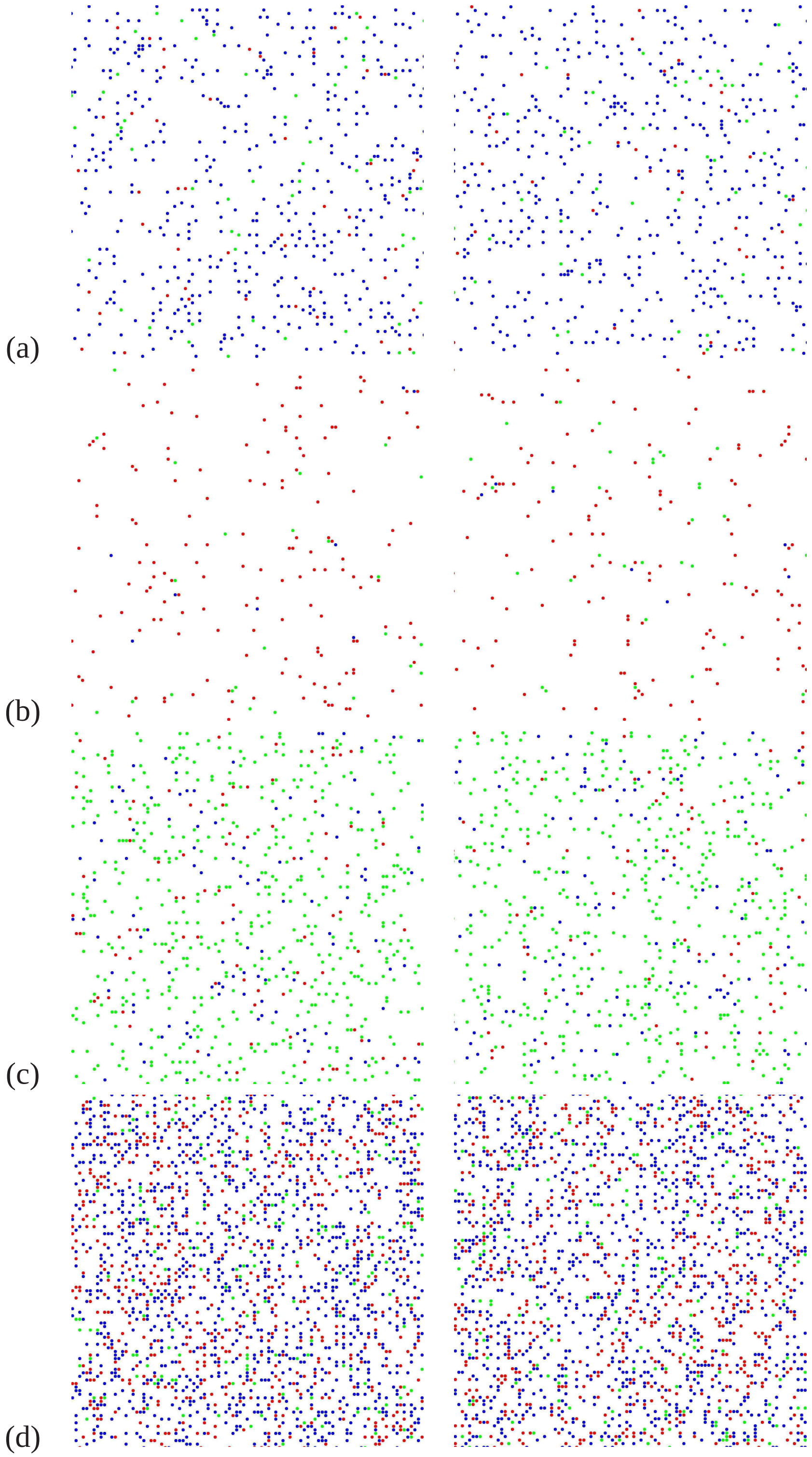

Figure 5.3: Realizations from the GRF model with specified (left) and estimated (right) clique potentials. The GRF model has a 8-neighborhood system.

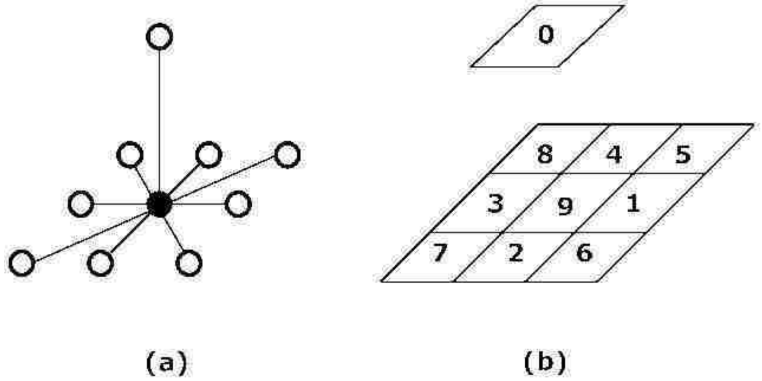

Figure 5.4: A multiscale GRF model: a) A (8+1)-neighborhood system and b) numbering of sites in the neighborhood system.

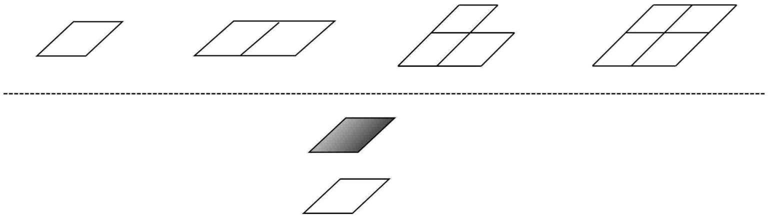

Figure 5.5: Cliques for a (8+1)-neighborhood system. The four cliques with sites from one layer are in the first row and the single clique with sites from two layers is in the second row.

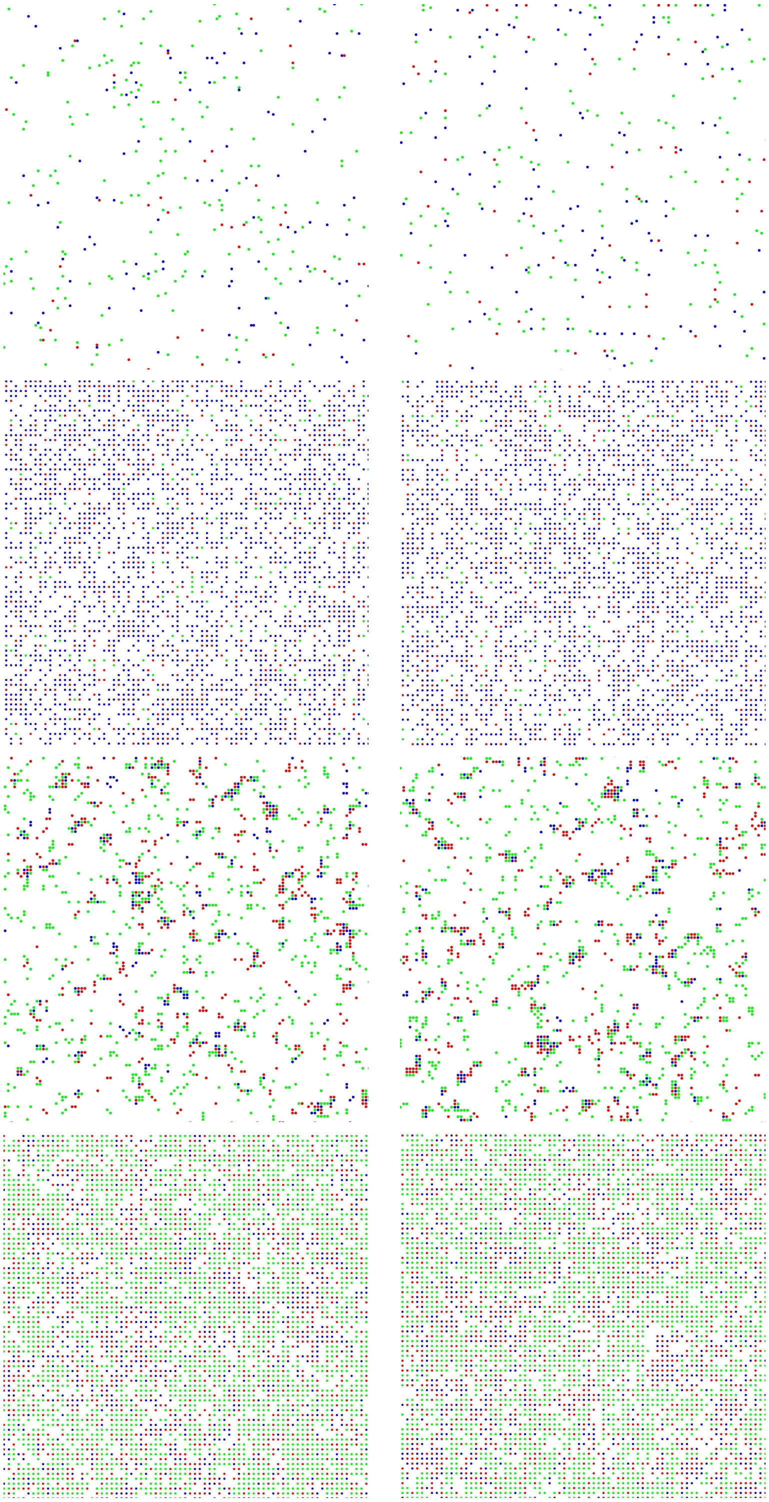

Figure 5.6: Realizations of the finest layer of the multiscale GRF model generated from GRFs with the specified (left) and estimated (right) clique potentials.

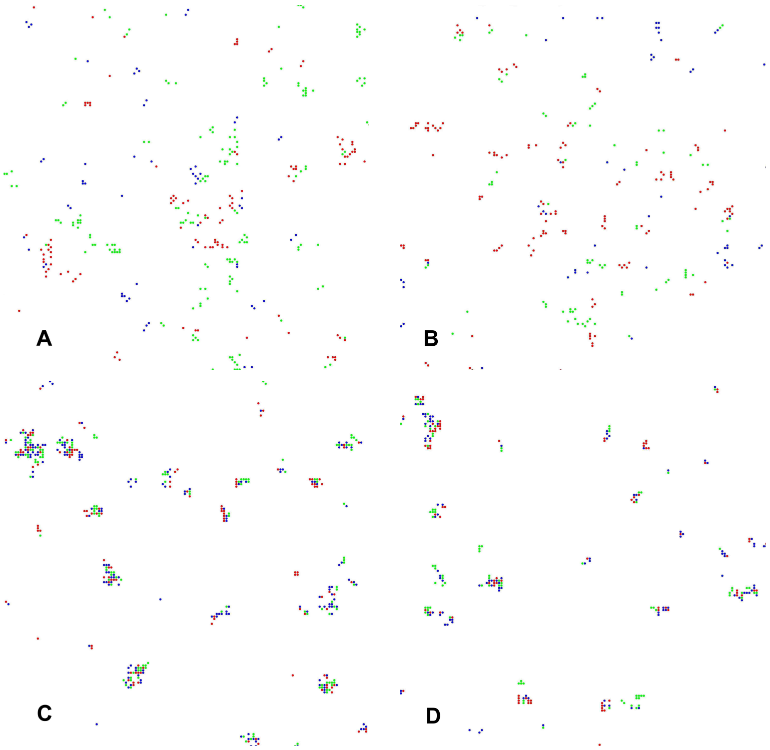

Figure 5.7: Multiscale GRF modeling of the experimental data. a) A reconstruction computed from the multiscale GRF modeling that shows the distributions of $\beta$ (blue), LAT (red), and $\mathrm{PLC}\gamma1$ (green); b) A reconstruction computed form the multiscale MRF modeling that shows the distributions of $\beta$ (blue), LAT (red), and $\mathrm{PLC}\gamma1$ (green); c) A reconstruction computed from the multiscale GRF modeling that shows the distributions of $\beta$ (blue), p85 (red), and Gab2 (green); d) A reconstruction computed form the multiscale MRF modeling that shows the distributions of $\beta$ (blue), p85 (red), and Gab2 (green).

# Chapter 6

# Conclusion

## 6.1 Contributions

A novel approach has been described for reconstructing spatial relationships between three proteins on cell membranes from samples showing relationships between only two proteins. This approach utilizes a multiscale hidden Markov random field model where mathematical programming techniques are used to deduce the conditional distributions. To our knowledge, we are the first to use Markov random fields to model the spatial distribution of proteins on cell membranes. The application of our approach to synthesized data has demonstrated that the multiscale MRF model is good at characterizing both short and long range statistical properties and that the spatial relationships among three proteins can be reliably estimated. The application to experimental data has provided the first maps of the spatial relationship between groups of three different signaling molecules. The ability to analyze the spatial organization and dynamics of multiple membrane proteins during signaling is a critical step towards a more complete understanding of the spatial and temporal regulation of signal transduction pathways.

By exploiting the Markov-Gibbs equivalence, 8-neighborhood systems have been utilized in the multiscale GRF modeling. By imposing the very reasonable constraint of statistical isotropy, the number of parameters is reduced and the consistency of the model parameters is improved. Application of the multiscale GRF model to synthesized and experimental data shows that the quality of reconstruction has been improved. In addition, it is easier to obtain biological insights when the results are interpreted in term of clique potentials. Using clique potentials, combined with MRF modeling's ability to derive confidence intervals, it is possible to provide quantitative measurements of co-clustering for molecular species.

## 6.2 Future Research Directions

As shown in the completed work, the interactions among three proteins can be reliably estimated from observations showing interactions between only two proteins using the multiscale models. However, there are still many issues to be addressed.

### 6.2.1 Labeling Efficiency

*Labeling efficiency* is used to denote the percentage of a protein labeled by gold particles in a biological experiment. In the experiments we have done so far, it is assumed that all proteins of interest on cell membranes are labeled by gold particles, *i.e.*, labeling efficiencies of gold particles are all 100% . Unfortunately, the labeling efficiencies in experiments using nanoprobes are typically less than 100%. Because of their smaller sizes, 5nm gold particles have better labeling efficiency than 10nm gold particles. For some proteins whose populations are known, we estimated labeling efficiencies using gold particle labeling data. These experiments have shown that the labeling efficiencies of 5nm gold particles are from 70% to 90% while the labeling efficiencies of 10nm gold particles are typically less than 50%. We propose to include the labeling efficiency in the observation matrices. Consequently, three observation matrices, $\mathbf{Q}_r$, $\mathbf{Q}_g$, and $\mathbf{Q}_b$, are written as:

$$\mathbf{Q}_r = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & p_r^g & 0 & 0 \\ 0 & 0 & p_r^b & 0 \\ 1 & 1 - p_r^g & 1 - p_r^b & 1 \end{bmatrix}$$

$$\mathbf{Q}_g = \begin{bmatrix} p_g^r & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & p_g^b & 0 \\ 1 - p_g^r & 0 & 1 - p_g^b & 1 \end{bmatrix}$$

$$\mathbf{Q}_b = \begin{bmatrix} p_b^r & 0 & 0 & 0 \\ 0 & p_b^g & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 - p_b^r & 1 - p_b^g & 1 & 1 \end{bmatrix}$$

where $p_i^j$ is the labeling efficiency of protein $j$ in experiments where protein $i$ is missing.

Initially, it will be assumed that the labeling efficiencies are known. Quantitative measurements of protein populations and cell surface areas can be performed by our collaborators at the UNM Cancer Research Center, and labeling efficiencies can be estimated using gold particle labeling data. These labeling efficiencies will then be put into the above observation matrices and used to model the spatial distributions of proteins. This will allow us to determine whether the mathematical programming techniques developed in this dissertation will also work for parameter estimation when the labeling efficiencies are not 100%.

Unfortunately, some protein populations are not known. In this case, labeling efficiencies will be treated as hidden data in the modeling and the EM algorithm can be used to estimate the hidden process parameters.

### 6.2.2 Using the EM Algorithm to Estimate Parameters

When labeling efficiencies are unknown, more information is needed to estimate the hidden process parameters. As before, realizations of the observation process $o_k$ are called incomplete data. And the observation matrices $\mathbf{Q}_k$ are also hidden data. It is assumed that a complete data set $(o_k, \mathbf{Q}_k)$ exists and the complete-data likelihood can be computed as

$$P(o_k, \mathbf{Q}_k|\mathbf{V}) = P(o_k|\mathbf{Q}_k, \mathbf{V})P(\mathbf{Q}_k|\mathbf{V}) \tag{6.1}$$

where $\mathbf{V}$ is a vector of clique potentials and $k \in \{r, g, b\}$. Note that this complete-data likelihood is in fact a random variable because $\mathbf{Q}_k$ is unknown and governed by an underlying distribution. The hidden process potentials can be computed by

$$\hat{\mathbf{V}} = \arg\max_{\mathbf{V}} \log[P(o_k|\mathbf{Q}_k, \mathbf{V})P(\mathbf{Q}_k|\mathbf{V})]. \tag{6.2}$$

The EM algorithm is an interactive procedure for solving Eq. 6.2. In this specific case, each iteration consists of the following two steps.

1. E-step: Calculate the expectation of $\log P(o_k|\mathbf{Q}_k, \mathbf{V})P(\mathbf{Q}_k|\mathbf{V})$ with respect to $\mathbf{Q}_k$ and conditioned on the observed data $o_k$ and the current estimate $\mathbf{V}^i$ of clique potentials. The expectation is computed as:

$$\begin{aligned} Q(\mathbf{V}|\mathbf{V}^i) &= E\left[\log P(o_k|\mathbf{Q}_k, \mathbf{V}) + \log P(\mathbf{Q}_k|\mathbf{V})|o_k, \mathbf{V}^i\right] \\ &= E\left[\log P(o_k|\mathbf{Q}_k, \mathbf{V})|o_k, \mathbf{V}^i\right] + C \end{aligned} \tag{6.3}$$

   while $C$ is an unknown constant.

2. M-step: Compute

$$\begin{aligned} \mathbf{V}^{i+1} &= \arg\max_{\mathbf{V}} Q(\mathbf{V}|\mathbf{V}^i) \\ &= \arg\max_{\mathbf{V}} E\left[\log P(o_k|\mathbf{Q}_k, \mathbf{V})|o_k, \mathbf{V}^i\right] \end{aligned} \tag{6.4}$$

The expectation in Eq.6.3 is computed as follows:

$$E\left[\log P(o_k|\mathbf{Q}_k,\mathbf{V})|o_k,\mathbf{V}^i\right] = \int_{\mathbf{Q}_k} \log P(o_k|\mathbf{Q}_k,\mathbf{V})P(\mathbf{Q}_k|o_k,\mathbf{V}^i)\,d\mathbf{Q}_k \tag{6.5}$$

where $P(o_k|\mathbf{Q}_k,\mathbf{V})$ can be calculated as described in Chapter 5 and $P(\mathbf{Q}_k|o_k,\mathbf{V}^i)$ is the posterior distribution of $\mathbf{Q}_k$ given the data and the previous estimate of $\mathbf{V}$. A major difficulty here is how to specify this posterior distribution. Using prior knowledge from biological experiments, mean values and ranges of labeling efficiencies can be guessed. It is reasonable to assume that this posterior distribution of $\mathbf{Q}_k$ is a multivariate Gaussian distribution. With this assumption, the expectation can be computed. An initial set of clique potentials can be estimated using the method described in Chapter 5 by assuming 100% labeling efficiencies. The EM algorithm can then be used to estimate clique potentials in the case that the labeling efficiencies are unknown.

### 6.2.3 Exploring More Observation Matrices

The three observation matrices used in our modeling are very constrained. The constraints on the observation matrices can be further relaxed to obtain more observation processes. For example, we could ask our collaborators to carry out experiments where half of the 5nm gold particles are used to label protein $R$, the other half of the 5nm gold particles are used to label protein $G$, and all of the 10nm gold particles are used to label protein $B$. The experiments are realizations of an observation process which is related to the hidden process through the following observation matrix:

$$\mathbf{Q}_i = \begin{bmatrix} 0 & 0 & 0 & 0 \\ .5 & .5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ .5 & .5 & 0 & 1 \end{bmatrix}$$

where labeling efficiencies are assumed to be 100%. In this way, we can obtain more observation processes and impose more constraints on the hidden process. As a result, the quality of hidden process parameter estimates can be improved. We propose to use more observation matrices to improve the modeling quality for the three protein reconstruction problem.

Because there are more constraints on the hidden process when more observation matrices are used, we can possibly extend our models to more than three proteins. We propose to investigate whether the interactions among more than three proteins can be reliably estimated when more observation processes are introduced.

An observation matrix is a channel over which information of a hidden process is transmitted, *i.e.*, the hidden random field $D$ is observed as a random field $O_i$ over the observation matrix $Q_i$. The mutual information for $Q_i$ is the mutual information of two random fields $D$ and $O_i$, which can be computed as:

$$I(D;O_i) = \sum_{d \in D} \sum_{o \in O_i} P(d,o) \log \frac{P(d,o)}{P(d)P(o)}, \tag{6.6}$$

where $P(o,d)$ is the joint probability distribution function of $D$ and $O_i$, and $P(d)$ and $P(o)$ are the marginal probability distribution functions of $D$ and $O_i$ respectively. We propose to compute the mutual information for observation matrices and design observation matrices with maximum information channel capacity by maximizing the mutual information [39] for sets of observation matrices. We will use these observation matrices to design particle labeling experiments to obtain better reconstruction of interactions among proteins.

# Appendix A

# Results of the Markov chain modeling

## A.1 The first-order Markov chain model

| | $\mathbf{P}$ | | | | $\hat{\mathbf{P}}$ | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | .40885 | .36921 | .17410 | .04320 | .40896 | .36921 | .17415 | .04332 |
| | .12930 | .52834 | .10139 | .77892 | .12922 | .52832 | .10152 | .77888 |
| | .31119 | .08227 | .24039 | .09638 | .31103 | .08230 | .24050 | .09632 |
| | .15066 | .02019 | .48412 | .08151 | .15080 | .02018 | .48383 | .08148 |
| 2 | .00391 | .09946 | .08653 | .28725 | .00392 | .09942 | .08651 | .28724 |
| | .76572 | .34447 | .52591 | .53052 | .76580 | .34451 | .52578 | .53048 |
| | .11284 | .41860 | .01115 | .13215 | .11286 | .41862 | .01119 | .13221 |
| | .11753 | .13747 | .37642 | .05008 | .11742 | .13745 | .37652 | .05007 |
| 3 | .91308 | .01150 | .05131 | .99800 | .91310 | .01149 | .05135 | .99800 |
| | .05303 | .10345 | .70939 | .00148 | .05304 | .10355 | .70931 | .00147 |
| | .02554 | .81201 | .21349 | .00031 | .02552 | .81186 | .21348 | .00031 |
| | .00835 | .07304 | .02580 | .00022 | .00834 | .07310 | .02585 | .00022 |
| 4 | .92364 | .86406 | .05311 | .32127 | .92364 | .86408 | .05300 | .32098 |
| | .02656 | .10318 | .38505 | .33645 | .02654 | .10320 | .38536 | .33643 |
| | .01700 | .02302 | .00515 | .11821 | .01700 | .02297 | .00525 | .11814 |
| | .03281 | .00974 | .55670 | .22407 | .03282 | .00975 | .55638 | .22445 |
| 5 | .28942 | .43960 | .26739 | .59277 | .28941 | .43961 | .26748 | .59285 |
| | .35661 | .04516 | .31161 | .25504 | .35664 | .04512 | .31159 | .25505 |
| | .06877 | .40161 | .14996 | .10364 | .06877 | .40150 | .15002 | .10363 |
| | .28520 | .11364 | .27104 | .04855 | .28518 | .11377 | .27091 | .04846 |
| 6 | .85000 | .01000 | .03000 | .20000 | .85001 | .00997 | .02999 | .19997 |
| | .05000 | .50000 | .02000 | .40000 | .05006 | .49973 | .01998 | .40034 |
| | .08000 | .32000 | .90000 | .32000 | .07994 | .32014 | .90006 | .31988 |
| | .02000 | .17000 | .05000 | .08000 | .01999 | .17016 | .04997 | .07981 |
| 7 | .29000 | .55000 | .03000 | .63000 | .28985 | .54995 | .03003 | .63009 |
| | .31000 | .25000 | .21000 | .05000 | .30998 | .25023 | .20985 | .05005 |
| | .30000 | .14000 | .52000 | .09000 | .30014 | .13990 | .52015 | .08995 |
| | .10000 | .06000 | .24000 | .23000 | .10003 | .05993 | .23997 | .22990 |

Table A.1: Results for seven experiments of the first-order Markov chain model. $\mathbf{P}$ is the transition matrix used to generate the hidden chains and $\hat{\mathbf{P}}$ is the estimate computed solely from the statistics of the observed chains. The chain length, $N = 7.01 \times 10^7$, and the starting point, $S = 1 \times 10^5$.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $\frac{\lVert\hat{\mathbf{P}}-\mathbf{P}\rVert_2}{\lVert\mathbf{P}\rVert_2}(\times 10^2)$ | .037 | .018 | .014 | .043 | .023 | .048 | .033 |

Table A.2: Relative error of estimates for the first-order Markov chain modeling.

## A.2 The second-order Markov chain model

| | 1 | | 2 | | 3 | | 4 | | 5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\mathbf{P}$ | $\hat{\mathbf{P}}$ | $\mathbf{P}$ | $\hat{\mathbf{P}}$ | $\mathbf{P}$ | $\hat{\mathbf{P}}$ | $\mathbf{P}$ | $\hat{\mathbf{P}}$ | $\mathbf{P}$ | $\hat{\mathbf{P}}$ |
| $p_{bgr}$ | .0342 | .0341 | .0856 | .0862 | .0007 | .0007 | .0449 | .0448 | .0056 | .0056 |
| $p_{xgr}$ | .4836 | .4838 | .1061 | .1055 | .0051 | .0051 | .1610 | .1611 | .2886 | .2887 |
| $p_{gbr}$ | .0330 | .0331 | .1857 | .1858 | .0770 | .0770 | .6297 | .6297 | .0110 | .0109 |
| $p_{xbr}$ | .0709 | .0708 | .0462 | .0461 | .3153 | .3151 | .0070 | .0070 | .6237 | .6238 |
| $p_{gxr}$ | .1771 | .1771 | .4563 | .4557 | .3132 | .3132 | .0263 | .0263 | .3544 | .3544 |
| $p_{bxr}$ | .4308 | .4311 | .1070 | .1051 | .0038 | .0038 | .0366 | .0368 | .0028 | .0028 |
| $p_{xxr}$ | .0823 | .0821 | .1012 | .1036 | .1565 | .1566 | .0468 | .0466 | .0009 | .0007 |
| $p_{rbg}$ | .8868 | .8865 | .8332 | .8337 | .6428 | .6431 | .6621 | .6622 | .7712 | .7711 |
| $p_{xbg}$ | .0234 | .0236 | .1280 | .1275 | .1991 | .1987 | .0485 | .0485 | .0717 | .0717 |
| $p_{rxg}$ | .3018 | .3019 | .0470 | .0464 | .1214 | .1214 | .2215 | .2215 | .2443 | .2442 |
| $p_{bxg}$ | .1526 | .1528 | .6111 | .6089 | .1217 | .1216 | .4911 | .4913 | .1529 | .1529 |
| $p_{xxg}$ | .0596 | .0594 | .2510 | .2538 | .2072 | .2074 | .0717 | .0715 | .0043 | .0042 |
| $p_{rxb}$ | .4922 | .4923 | .6591 | .6581 | .6789 | .6787 | .5125 | .5124 | .1198 | .1198 |
| $p_{gxb}$ | .3442 | .3443 | .2269 | .2259 | .1149 | .1149 | .4691 | .4689 | .5030 | .5033 |
| $p_{xxb}$ | .1559 | .1556 | .0585 | .0605 | .0210 | .0214 | .0019 | .0020 | .0580 | .0579 |
| $p_{rxx}$ | .5287 | .5281 | .3439 | .3467 | .1220 | .1222 | .6773 | .6777 | .1735 | .1732 |
| $p_{gxx}$ | .3714 | .3713 | .5832 | .5863 | .1648 | .1650 | .1969 | .1968 | .6322 | .6298 |
| $p_{bxx}$ | .0797 | .0777 | .0113 | .0131 | .4806 | .4804 | .0805 | .0770 | .0812 | .0805 |
| $p_{xxx}$ | .0202 | .0228 | .0616 | .0539 | .2327 | .2324 | .0454 | .0486 | .1130 | .1165 |

Table A.3: Results of five experiments for the second-order Markov chain modeling with symmetry. The chains used in the modeling are of length $N = 7.01 \times 10^8$ and the starting point is $S = 1.0 \times 10^6$.

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $\frac{\|\hat{\mathbf{P}}-\mathbf{P}\|_2}{\|\mathbf{P}\|_2}(\times 10^2)$ | .23 | .68 | .07 | .33 | .31 |

Table A.4: Relative error of estimates for the second-order Markov chain modeling.

# References

[1] N. L. Andrews, K. A. Lidke, J. R. Pfeiffer, A. R. Burns, B. S. Wilson, J. M. Oliver, and Lidke D. S. Dynamics, topography, and microdomains in Fc$\epsilon$RI signaling. *Biophys. J. Submitted*, 2007.

[2] J. M. Bernardo and A. F. M. Smith. *Bayesian Theory*. John Wiley, New York, 1994.

[3] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B*, 36(2):192–236, 1974.

[4] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society B*, 48(3):259–302, 1986.

[5] C. Bouman and B. Liu. Multiple resolution segmentation of textured images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-13(2):99–113, 1991.

[6] C. A. Bouman and M. Shapiro. A multiscale random field model for Bayesian image segmentation. *IEEE Transactions on Image Processing*, 3(2):162–177, 1994.

[7] G. Celeux, F. Forbes, and N. Peyrard. EM procedures using mean field-like approximations for Markov model-based image segmentation. *Pattern Recognition*, 36(1):131–144, 2003.

[8] B. Chalmond. An iterative Gibbsian technique for reconstruction of M-ary images. *Pattern Recogn.*, 22(6):747–762, 1989.

[9] G. R. Cross and A. K. Jain. Markov random field texture models. *IEEE Trans. Pattern Anal. Machine Intell.*, 5(1):25–39, 1983.

[10] A. P. Dempster, N. M.Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal Royal Stat. Soc., Series B*, 39(1):1–38, 1977.

[11] H. Derin and H. Elliott. Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9(1):39–55, 1987.

[12] A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *ICCV '99: Proceedings of the International Conference on Computer Vision-Volume 2*, page 1033, Washington, DC, USA, 1999. IEEE Computer Society.

[13] Davi Geiger and Federico Girosi. Parallel and deterministic algorithms from MRFs: Surface reconstruction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(5):401–412, 1991.

[14] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, 6(6):721–741, Nov. 1984.

[15] Philip E. Gill, Walter Murray, and Michael A. Saunders. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Journal on Optimization*, 12(4):979–1006, 2002.

[16] M. I. Gurelli and L. Onural. On a parameter estimation method for Gibbs-Markov random fields. *IEEE trans. on Pattern Analysis and Machine Intelligence*, 16(4):424–430, 1994.

[17] J. F. Hancock and I. A. Prior. Electron microscopic imaging of ras signaling domains. *Methods*, 37(2):165–712, 2005.

[18] B. A. Hernandez-Sanchez, T. J. Boyle, T. N Lambert, S. D. Daniel-Taylor, J. M. Oliver, B. S. Wilson, D. S. Lidke, and N. L. Andrews. Synthesizing biofunctionalized nanoparticles to image cell signaling pathways. *IEEE Trans. NanoBioscience*, 5:222–230, 2006.

[19] B.R. Hunt. Bayesian methods in nonlinear digital image restoration. *IEEE Trans. Comp.*, 26(3):219–229, March 1977.

[20] Z Kato, M. Berthod, and J. Zerubia. A hierarchical Markov random field model and multitemperature annealing for parallel image classification. *CVGIP: Graphical Model and Image Processing*, 58(1):18–37, 1996.

[21] Z. Kato, J. Zerubia, and M. Berthod. Unsupervised parallel image classification using Markovian models. *Pattern Recognition*, 32(4):591–604, 1999.

[22] J. H. Kim, L. Cramer, H. Mueller, B. S. Wilson, and B. Vilen. Independent trafficking of Ig-$\alpha$/Ig-$\beta$ and $\mu$-heavy chain is facilitated by dissociation of the B cell antigen receptor complex. *Journal of Immunology*, 175:147–154, 2005.

[23] Ross Kindermann and J. Laurie Snell. *Markov Random Fields and their Applications*, volume 1 of *Contemporary Mathematics*. American Mathematical Society, 1980.

[24] J. M. Laferté, P. Pérez, and F. Heitz. Discrete Markov image modeling and inference on the quadtree. *IEEE Transactions on Image Processing*, 9(3):390–404, 2000.

[25] S. Lakshmanan and H. Derin. Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(8):799–813, 1989.

[26] S. Z. Li. *Markov random field modeling in computer vision*. Springer-Verlag, London, UK, 1995.

[27] K. H. Liang and T. Tjahjadi. Adaptive scale fixing for multiscale texture segmentation. *IEEE Transactions on Image Processing*, 15(1):249–256, January 2006.

[28] P. Masson and W. Pieczynski. SEM algorithm and unsupervised statistical segmentation of satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, 31(3):618–633, 1993.

[29] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy. Sonar image segmentation using an unsupervised hierarchical MRF model. *IEEE Transactions on Image Processing*, 9(7):1216–1231, 2000.

[30] D. V. Nicolau, K. Burrage, R. G. Parton, and J. F. Hancock. Identifying optimal lipid raft characteristics required to promote nanoscale protein-protein interactions on the plasma membrane. *Mol Cell Biol*, 26:313–323, 2006.

[31] J. M. Oliver, J. R. Pfeiffer, Z. Surviladze, S. L. Steinberg, K. Leiderman, M. Sanders, C. Wofsy, J. Zhang, H. Fan, N. Andrews, S. Bunge, T. Boyle, P. Kotula, and B. S. Wilson. Membrane receptor mapping: the membrane topography of Fc$\epsilon$RI signaling. In P. Quinn, editor, *Membrane Dynamics and Domains*, number 37 in Subcellular Biochemistry, pages 3–34. Kluwer Academic/Plenum Publishers, Dordecht/New York, 2004.

[32] R. Paget and I. D. Longstaff. Texture synthesis via a noncausal nonparametric multiscale Markov random field. *IEEE Transactions on Image Processing*, 7(6):925–931, 1998.

[33] S. J. Plowman, C. Muncke, R. G. Parton, and Hancock J. F. H-ras, k-ras and inner plasma membrane raft proteins operate in nanoclusters with differential dependence on the actin cytoskeleton. *Proc. Nat. Acad. Sci.*, 102:15500–15505, 2005.

[34] I. A. Prior, C. Muncke, R. G. Parton, and J. F. Hancock. Direct visualization of ras proteins in spatially distinct cell surface microdomains. *The Journal of Cell Biology*, 160(2):165–170, 2003.

[35] W. Qian and D. M. Titterington. Estimation of parameters in hidden Markov models. *Philos. Trans. Roy. Soc. A*, 337(1647):407–428, 1991.

[36] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77:257–286, 1989.

[37] H. Tjelmeland and J. Besag. Markov random fields with higher order interactions. *Scandinavian Journal of Statistics*, 25:415–433, 1998.

[38] A. Tonazzini, L. Bedini, and E. Salerno. A Markov model for blind image separation by a mean-field EM algorithm. *IEEE Transactions on Image Processing*, 15(2):473–482, 2006.

[39] P. Viola and W. M. Wells III. Alignment by maximization of mutual information. *International J. Computer Vision*, 24(2):137–154, 1997.

[40] P. Volna, P. Lebduska, L. Draberova, S. Simova, P. Heneberg, M. Boubelik, V. Bugajev, B. Malissen, B. S. Wilson, V. Horejsi, M. Malissen, and P. Draber. Negative regulation of mast cell signaling and function by the adaptor lab/ntal. *Journal of Experimental Medicine*, 200(8):1001–1013, 2004.

[41] G. Wei and M. Tanner. A Monte Carlo implementation of the EM algorithm and the poor mans data augmentation algorithms. *Journal of the American Statistical Association*, 85:699–704, 1990.

[42] B. S. Wilson, J. R. Pfeiffer, and J. M. Oliver. Observing Fc$\epsilon$RI signaling from the inside of the mast cell membrane. *Journal of Cell Biology*, 149(5):1131–1142, 2000.

[43] B. S. Wilson, J. R. Pfeiffer, and J. M. Oliver. Fc$\epsilon$RI signaling observed from the inside of the mast cell membrane. *Molecular Immunology*, 38:1259–1268, 2002.

[44] B. S. Wilson, J. R. Pfeiffer, Z. Surviladze, E. A. Gaudet, and J. M. Oliver. High resolution mapping of mast cell membranes reveals primary and secondary domains of Fc$\epsilon$RI and LAT. *Journal Cell Biology*, 154(3):645–658, 2001.

[45] B. S. Wilson, S. L. Steinberg, K. Liederman, J. R. Pfeiffer, Z. Surviladze, J. Zhang, L. E. Samelson, L. h. Yang, P. G. Kotula, and J. M. Oliver. Markers for detergent-resistant lipid rafts occupy distinct and dynamic domains in native membranes. *Molecular Biology of the Cell*, 15(6):2580–2592, 2004.

[46] R. Wilson and C. T. Li. A class of discrete multiresolution random fields and its application to image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(1):42–56, 2003.

[47] M. Xue, G. Hsieh, M A Raymond-Stintz, D. Roberts, J. Zhang, Steinberg S. L., J. M. Oliver, Prossnitz E. R., and B. S. Wilson. FPR and Fc$\epsilon$RI occupy common signaling domains for localized crosstalk. *Mol. Biol. Cell.*, 18:1410–1420, 2007.

[48] J. Zhang. The mean field theory in EM procedures for Markov random fields. *IEEE Transactions on Image Processing*, 40(10):2570–2583, 1992.

[49] J. Zhang, K. Leiderman, J. R. Pfeiffer, B. S. Wilson, J. M. Oliver, and S. L. Steinberg. Characterizing the topography of membrane receptors and signaling molecules from spatial patterns obtained using nanometer-scale electron-dense probes and electron microscopy. *Micron*, 37(1):14–34, 2006.

[50] J. Zhang, J. W. Modestino, and D. A. Langan. Maximum-likelihood parameter estimation for unsupervised stochastic model-based image segmentation. *IEEE Transactions on Image Processing*, 3(4):404–420, 1994.