$$p_r(k) = \frac{a_r \hat{p}_r(k) + \sum_{s \in N_r \cap L} \sum_{k'=1}^{M} p_s(k') b(r, s, |k - k'|)}{a_r + \sum_{s \in N_r \cap L} \sum_{k'=1}^{M} b(r, s, |k - k'|)} \quad (16)$$

will converge to its unique global minimum, regardless of the initial value [12]. Suppose one chooses an initial value for p that satisfies (13), for example, by setting $p_r(k) = \hat{p}_r(k)$ (with the appropriate normalization) for all r and k. Since the weights a and b and the likelihood $\hat{p}_r(k)$ are nonnegative, one can see that each $p_r(k)$ will also remain nonnegative after each update. Summing both sides of (16) with respect to k, and using the fact that $\sum_{k=0}^{n-1} \hat{p}_r(k) = 1$, and that, from the symmetry and nonnegativity properties of the weights b,

$$\sum_{k=1}^{M} b(r, s, |k - k'|) = \sum_{m=0}^{M-1} b(r, s, m),$$

one gets that the condition $\sum_{k=1}^{M} p_r(k) = 1$ will also remain satisfied after the update. Hence, the fixed point of (16) (the optimal p) will be a valid probability distribution.

In the case of the Ising model, because of the nature of the potentials given by (12), the number of times that the field at location r takes the value q_k on a given number of independent samples of the field can depend (probabilistically) on only the number of times that the field at neighboring sites takes either the same value q_k or any other one that is different, no matter which one. In other words, the term $(p_r(k) - p_s(k'))^2$ cannot have a nonzero coefficient in (14), with $s \in N_r$, unless $k = k'$, because the conditional probability $\Pr[p_r(k) | p_s(k'), k \neq k']$ can depend only on the sum $\sum_{k \neq k'} p_s(k') = 1 - p_s(k)$. This means that, for homogeneous Ising fields, the weights b should satisfy:

$$b(r, s, m) = \begin{cases} \lambda, & \text{if } m = 0 \\ 0, & \text{otherwise,} \end{cases}$$

where λ is a positive parameter. In this case, the field p behaves like a set of decoupled membrane models, one for each layer $p(k) = \{p_r(k), r \in L\}$, and its Gibbsian energy corresponds to a generalized membrane model:

$$U(p) = \sum_{r \in L} |p_r - \hat{p}_r|^2 + \lambda \sum_{<r,s>} |p_r - p_s|^2, \quad (17)$$

where $|p_r|^2 = \sum_{k=1}^{M} (p_r(k))^2$. In intuitive terms, this model states that the posterior empirical marginal distributions should be similar to the likelihood of the observations (first term) and should vary smoothly across the lattice (second term), with the parameter λ controlling the relative weight of this smoothness constraint. The optimal estimator for the p field is obtained by minimizing this energy function and the optimal estimator for each f_r corresponds to the mode of each distribution p_r (see (24)). Since the M layers $\{p(k), k = 1, \ldots M\}$ are decoupled, the minimization of U may be done very efficiently: It is equivalent to the decoupled minimization of the M functions $U_k, k = 1, \ldots, M$, which are given by:

$$U_k(p) = \sum_{r \in L} (p_r(k) - \hat{p}_r(k))^2 + \lambda \sum_{<r,s>} (p_r(k) - p_s(k))^2. \quad (18)$$

This minimization may be performed by a number of methods. A powerful one is to use the Gauss-Seidel method combined with a multigrid approach [5] to accelerate the convergence; this method can be extended to produce high quality solutions, even for irregularly shaped regions [4].

If the region L is a rectangle of $n \times m$, where n and m are powers of 2, it is possible to obtain faster approximate solutions using transform-based methods. These methods are applicable because the linear equations that are obtained by setting $\partial U_k / \partial p_r(k) = 0$ for an interior node $r = (x, y)$ take the form of two-dimensional difference equations:

$$p_{x,y}(k) - \widehat{p}_{x,y}(k) + \lambda \left[4p_{x,y}(k) - p_{x-1,y}(k) - p_{x+1,y}(k) - p_{x,y-1}(k) - p_{x,y+1}(k) \right] = 0. \quad (19)$$

Note, however, that these equations are different for nodes at the borders of L (i.e., for $x = 0$, $y = 0$, $x = n - 1$, and $y = m - 1$), whereas transform methods require that (19) is valid for every node $(x, y) \in L$. This means that it is necessary to extend the field p so that it is defined for $x = -1, \ldots, n$ and $y = -1, \ldots, m$, introducing artificial boundary conditions. For the Fourier transform, these boundary conditions are periodic, which may introduce severe distortions in the reconstructed field. So, it is better to use the discrete cosine transform (DCT, type II) [7], which imposes the discrete equivalent of Neumann boundary conditions (see (22) below), which introduce only slight distortions; this transform (on an $n \times m$ lattice) is defined by:

$$DCT[f](u, v) = F(u, v) = \sum_{x=0}^{n-1} \sum_{y=0}^{m-1} 4 f(x, y) \cos\left[\frac{\pi u}{2n}(2x + 1)\right] \cos\left[\frac{\pi v}{2m}(2y + 1)\right], \quad (20)$$

for $(u, v) \in L$. The inverse transform is given by the expression:

$$DCT^{-1}[F](x, y) = \frac{1}{nm} \sum_{u=0}^{n-1} \sum_{v=0}^{m-1} w(u) w(v) F(u, v) \cos\left[\frac{\pi u}{2n}(2x + 1)\right] \cos\left[\frac{\pi v}{2m}(2y + 1)\right], \quad (21)$$

where $w(x)$ equals $1/2$ for $x = 0$ and equals 1 otherwise. Equation (21) is valid for all (x, y), but it imposes the conditions:

$$f_{-1,y} = f_{0,y}; f_{n,y} = f_{n-1,y}; f_{x,-1} = f_{x,0}; f_{x,m} = f_{x,m-1} \quad (22)$$

(which may be verified by direct substitution in (21)) for all $(x, y) \in L$. By using (21), one may express (19) in terms of $P_k(u, v)$ and $\widehat{P}_k(u, v)$ (the DCT of $p(k)$ and $\widehat{p}_k$, respectively) to obtain: