

Channel Estimation by Inference on Gaussian Markov Random Fields

Thomas J. Riedl, Jun Won Choi, and Andrew C. Singer
 University of Illinois at Urbana-Champaign
 Urbana, IL 61801, USA
 Email: triedl2@illinois.edu

Abstract—In this paper, we discuss a novel method for channel estimation. The approach is based on the idea of modeling the complex channel gains by a Markov random field. This graphical model is used to capture the statistical dependencies between consecutive taps in time and delay. The sum-product algorithm is finally employed to infer a MAP channel estimate from given observations.

Index Terms—Markov random field, channel estimation, channel modeling, gaussian message passing.

I. INTRODUCTION

Physical channels such as underwater acoustic channels and ionospheric radio channels which result in time-varying multipath propagation of the transmitted signal may be characterized mathematically as time-varying linear filters [1]. At the receiver, the incident signal is sampled and hence the effect of this time-varying channel on the transmitted sequence $x[i]$ can be modeled by a discrete time, time-varying filter $h[i, j]$. The sampled received signal $y[i]$ can then be represented as

$$y[i] = \sum_{j=0}^{L-1} h[i, j]x[i - j] + w[i], \quad i = 0, \dots, M - 1 \quad (1)$$

where $w[i]$ denotes additive complex white Gaussian noise with zero mean and circular symmetric variance σ_n^2 , M is the total number of observations and the delay spread of the channel is assumed to be at most L symbols.

The complex gains $h[i, j]$ are modeled as random variables, where i indicates time and different j correspond to different multipath arrivals. It is typically assumed that $h[i, j]$ is Gaussian, uncorrelated in delay j and wide sense stationary in time i . The variability of the wireless channel over time is then fully characterized by the autocorrelation of $h[i, j]$ in time defined by $R_j(m) = E[h[i + m, j]h[i, j]^*]$. Random processes of that kind can be well approximated by a Gaussian autoregressive (AR) process, i.e., a random process that is generated by an autoregressive filter when driven by white Gaussian noise. This channel model captures the dependencies of $h[i, j]$ in time i but neglects any dependence of $h[i, j]$ in delay j . Figure 1 shows a typical channel impulse response as measured from real underwater acoustic communication

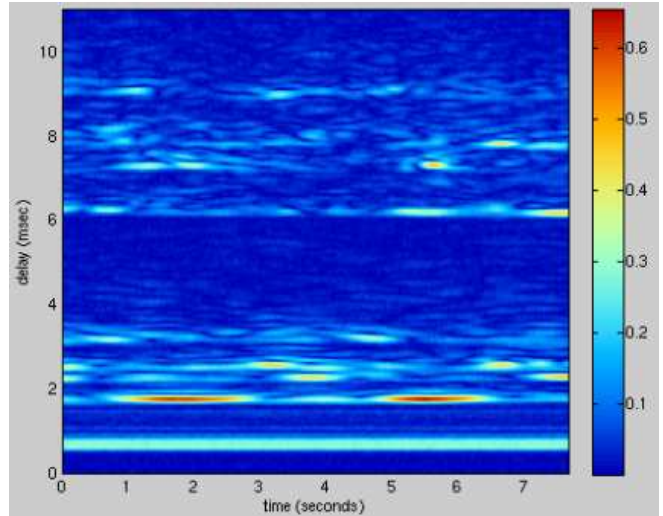


Figure 1. Typical channel realization

data. A certain continuity in delay is observed but this kind of local dependency in delay is neglected in the channel model just described. The channel estimators proposed in [2]–[5] are built on this model and hence will suffer a performance loss whenever they face a channel exhibiting strong dependencies in delay.

The underwater acoustic communications channel is rapidly time-varying and often sparse. Experimental results such as the channel plot above, however, suggest that the statistics of this channel are substantially less fluctuating and the statistical dependencies between consecutive taps in time and delay are local and relatively strong. Drawing on these observations we propose a new approach to channel modeling and channel estimation that is based on a *Markov random field* (MRF) model for the complex channel gains. This graphical model effectively captures the local statistical dependencies of the channel taps and can easily be embedded in iterative (turbo) receivers [6]. This is expected to improve the data estimation performance significantly as the MRF carries prior knowledge of the channel.

Markov random fields have been used successfully in the field of image processing and restoration for modeling spatial continuity [7]. MRFs are also used in the context of *magnetic*

This work was supported in part by the department of the Navy, Office of Naval Research, under grants ONR MURI N00014-07-1-0738 and ONR N00014-07-1-0311.

resonance (MR) imaging, where they help to unwrap MR phase images by imposing smoothness constraints on the true phase function [8].

We elaborate on the theory behind MRFs in Section III. Basically, this concept allows us to carry the conditional independence property of a *Markov Chain* over to two dimensional random processes such as $h[i, j]$ and to define the global behavior of $h[i, j]$ by local constraints on the probability distribution of $h[i, j]$. A performance evaluation of the proposed MRF channel model is presented in Section VI. It will turn out that channel estimation based on our MRF model yields competitive results in case there is no correlation in delay and it clearly outperforms common channel estimation techniques if the channel is correlated in delay.

II. NOTATION AND SYSTEM DEFINITION

For notational convenience we represent the channel taps in the matrix \mathbf{H} by identifying $[\mathbf{H}]_{(i,j)}$ with $h[i, j]$ and also introduce the vectors \mathbf{y} and \mathbf{x} with $[\mathbf{y}]_{(i)} = y[i]$ and $[\mathbf{x}]_{(i)} = x[i]$. Due to the Gaussian assumption on the channel noise, the distribution of \mathbf{y} given \mathbf{x} and \mathbf{H} then reads

$$p(\mathbf{y}|\mathbf{H}, \mathbf{x}) \sim \prod_{i=0}^{M-1} \underbrace{\exp\left(-|y[i] - \sum_{j=0}^{L-1} h[i, j]x[i-j]|^2/\sigma_n^2\right)}_{=T_i(\mathbf{H})} \quad (2)$$

where the notation \sim means that the left-hand side and the right-hand side are proportional to each other.

III. MARKOV RANDOM FIELD THEORY

We use an MRF to model the time-varying channel \mathbf{H} . By definition, an MRF is a set of random variables \mathbf{H} with an associated undirected graph. The random variables h in \mathbf{H} are assumed to satisfy the following two conditions:

$$p(h) > 0 \quad (3)$$

$$p(h|\mathbf{H}_{\setminus h}) = p(h|k \in n(h)) \quad (4)$$

where $\mathbf{H}_{\setminus h}$ is the entire set of random variables \mathbf{H} without the element h and $n(h)$ denotes the set of nodes k that neighbor h in the corresponding graph.¹ In other words, an MRF represents statistical dependencies of variables by an undirected graph. The vertices in an MRF stand for random variables and the edges impose statistical constraints on these random variables. A subset b of \mathbf{H} is called a *clique* if it is a singleton or if every pair of elements h in b are neighbors in the corresponding graph.

The lattice shaped MRF considered in this paper is depicted in Figure 2. Due to the geometric nature of lattices, we prefer to index the variables h by the two dimensional index pair $[i, j]$. So here the set \mathbf{H} represents the set of channel gains $h[i, j]$, and $\mathbf{N}[i, j]$ will denote the set $\{[l, m] : h[l, m] \text{ neighbors } h[i, j]\}$. We easily identify a

¹Note that we use the same symbol to denote a random variable and its realization, whenever the meaning is clear from the context.

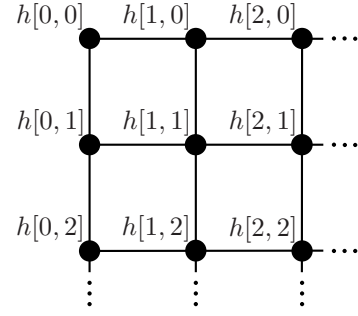


Figure 2. Lattice shaped Markov random field

singleton clique for each vertex $h[i, j]$ and a pairwise clique for each pair of adjacent vertices. These pairwise cliques have the form $\{h[i, j], h[i, j-1]\}$ or $\{h[i, j], h[i-1, j]\}$. According to the Hammersley-Clifford theorem [9], [10] \mathbf{H} and its associated graph form an MRF if and only if the distribution on the variables in \mathbf{H} is of the form

$$p(\mathbf{H}) = Z^{-1} \exp\left(-\frac{\sum_{b \in \mathcal{B}} V_b(\bar{b})}{K}\right) \quad (5)$$

This form of joint distribution is known as a Gibbs distribution. The parameter K is a temperature parameter chosen to be unity here, and Z is a normalization constant. The argument of the exponential function includes a sum of functions $V_b(\bar{b})$ which are indexed by the set \mathcal{B} of all possible cliques b . The vector \bar{b} is composed of the values of the random variables $h[i, j]$ in the clique b . The functions $V_b(\bar{b})$ are simply assumed to be nonnegative functions of their arguments and are called clique potentials.

IV. CHANNEL MODELING

We take the potential of pairwise cliques as the square of adjacent differences:

$$V_b(\bar{b}) = \alpha_{[i_b, j_b], [m_b, n_b]} |(h[i_b, j_b] - \mu[j_b]) - (h[m_b, n_b] - \mu[n_b])|^2 \quad (6)$$

where $[i_b, j_b]$ and $[m_b, n_b]$ are the coordinates of the vertices in the one pair clique b and $\alpha_{[i_b, j_b], [m_b, n_b]}$, $\mu[j_b]$ and $\mu[n_b]$ are parameters. The potentials of single cliques that are associated with channel gains $h[i, j]$ at time $i = 0$ form a Gaussian distribution

$$V_b(\bar{b}) = \alpha_{[j_b]} |h[i_b, j_b] - \mu[j_b]|^2 \quad (7)$$

with the parameters $\mu[j_b]$ and $\alpha_{[j_b]}$. The other potentials are assumed to be zero. So we have

$$p(\mathbf{H}|\boldsymbol{\theta}) = Z(\boldsymbol{\theta})^{-1} \exp\left(-\sum_{b \in \mathcal{B}_1} \alpha_{[j_b]} |h[i_b, j_b] - \mu[j_b]|^2 - \sum_{b \in \mathcal{B}_2} \alpha_{[i_b, j_b], [m_b, n_b]} |(h[i_b, j_b] - \mu[j_b]) - (h[m_b, n_b] - \mu[n_b])|^2\right)$$

where the vector $\boldsymbol{\theta}$ contains all the model parameters, i.e., the $\alpha_{[i_b, j_b], [m_b, n_b]}$, the $\alpha_{[j_b]}$ and the $\mu[j]$. The set \mathcal{B}_1 comprises all

single cliques that correspond to channel gains $h[i, j]$ at time $i = 0$ and the set B_2 contains all the pairwise cliques.

Figure 3 shows the factor graph representation of $p(\mathbf{H}|\boldsymbol{\theta})$.

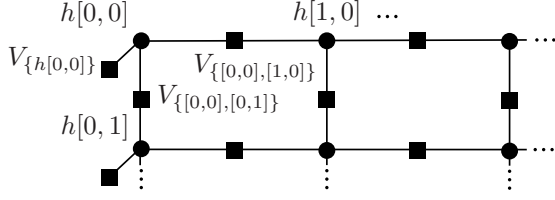


Figure 3. Factor graph based on MRF model

The black boxes and circles represent function and variable nodes, respectively. Obviously, our MRF model can be interpreted as a sequence of vertically connected Markov chains and it accounts for the correlation between consecutive taps in delay by introducing function nodes between vertically neighboring variable nodes.

The reasons for the above choice of the potentials are as follows. First it makes \mathbf{H} jointly Gaussian distributed and $h[i, j]$ has mean $\mu[j]$. As we will see later in this paper assuming a Gaussian distribution significantly simplifies the computation needed for the inference of the channel gains. Second, it imposes a certain continuity on the behavior of neighboring channel gains, as $p(h[i, j]|h[l, m] : [l, m] \in \mathbf{N}[i, j])$ for $i \neq 0$ then becomes a complex Gaussian distribution with mean

$$\mu[j] + \frac{\sum_{[l,m] \in \mathbf{N}[i,j]} \alpha_{[i,j],[l,m]} (h[l, m] - \mu[m])}{\sum_{[l,m] \in \mathbf{N}[i,j]} \alpha_{[i,j],[l,m]}} \quad (8)$$

and variance

$$\left(\sum_{[l,m] \in \mathbf{N}[i,j]} \alpha_{[i,j],[l,m]} \right)^{-1} \quad (9)$$

The mean of $h[i, j]$ is shifted by a weighted sum of the differences $h[l, m] - \mu[m]$. So if all the neighbors, for example, assumed values above their means, $h[i, j]$ is likely to assume a value that is above its mean as well. The $\alpha_{[i,j],[l,m]}$'s determine what impact each neighbor has on $h[i, j]$. For $i = 0$, $p(h[i, j]|h[l, m] : [l, m] \in \mathbf{N}[i, j])$ becomes a complex Gaussian distribution with mean

$$\mu[j] + \frac{\sum_{[l,m] \in \mathbf{N}[i,j]} \alpha_{[i,j],[l,m]} (h[l, m] - \mu[m])}{\alpha_{[j]} + \sum_{[l,m] \in \mathbf{N}[i,j]} \alpha_{[i,j],[l,m]}} \quad (10)$$

and variance

$$\left(\alpha_{[j]} + \sum_{[l,m] \in \mathbf{N}[i,j]} \alpha_{[i,j],[l,m]} \right)^{-1} \quad (11)$$

The $\alpha_{[j]}$ hence determines how likely it is that our concatenated Markov chains start with the mean $\mu[j]$.

V. MRF-BASED CHANNEL ESTIMATION

A. Problem setup

The channel estimation task is based on the following MAP estimation approach:

$$\begin{aligned} \hat{\mathbf{H}} &= \underset{\mathbf{H}}{\operatorname{argmax}} p(\mathbf{H}|\mathbf{y}, \mathbf{x}, \boldsymbol{\theta}) \\ &= \underset{\mathbf{H}}{\operatorname{argmax}} p(\mathbf{H}, \mathbf{y}, \mathbf{x}|\boldsymbol{\theta}) \\ &= \underset{\mathbf{H}}{\operatorname{argmax}} p(\mathbf{y}|\mathbf{H}, \mathbf{x})p(\mathbf{H}|\boldsymbol{\theta}) \end{aligned} \quad (12)$$

For now, we assume the values of the parameters $\boldsymbol{\theta}$ to be known.

B. Solution

The probabilities $p(\mathbf{y}|\mathbf{H}, \mathbf{x})$ and $p(\mathbf{H}|\boldsymbol{\theta})$ are both proportional to Gaussian densities in \mathbf{H} and hence $p(\mathbf{y}|\mathbf{H}, \mathbf{x})p(\mathbf{H})$ can be considered as proportional to a Gaussian density in \mathbf{H} as well. Gaussian densities assume their maximum at their mean and therefore the optimization problem above amounts to calculating the means of \mathbf{H} with respect to $p(\mathbf{y}|\mathbf{H}, \mathbf{x})p(\mathbf{H})$. As $p(\mathbf{y}|\mathbf{H}, \mathbf{x})p(\mathbf{H})$ factors nicely, the sum-product algorithm can be used for an efficient marginalization. The means of the marginals with respect to $p(\mathbf{y}|\mathbf{H}, \mathbf{x})p(\mathbf{H})$ are easily identified as the means of \mathbf{H} with respect to $p(\mathbf{y}|\mathbf{H}, \mathbf{x})p(\mathbf{H})$. Figure 4 shows the factor graph representation of $p(\mathbf{y}|\mathbf{H}, \mathbf{x})p(\mathbf{H})$.

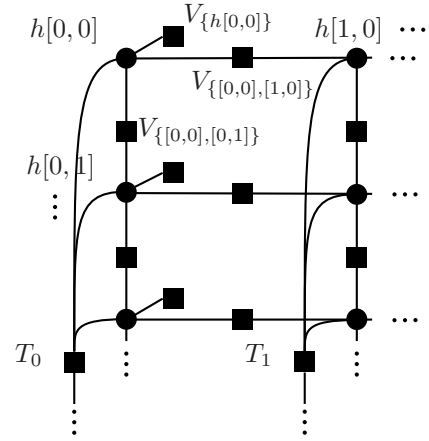


Figure 4. Factor graph for MAP channel estimation

Because of the Gaussianity of the distributions involved, the marginalized functions are completely characterized by their mean and variance. Our factor graph has many cycles and so we need to iterate the message passing until it converges [11].

Following the notation of [11] we denote messages sent from a variable node $h[i, j] \in \mathbf{H}$ to a local function node as $m_{h \rightarrow f}(h)$. Here f represents either one of the functions T_i or one of the potential functions $V_b(\bar{b})$. Furthermore we denote messages sent from a local function node to a variable node as $m_{f \rightarrow h}(h)$. Also, let $n(v)$ denote the set of nodes that neighbor a given node v in a factor graph.

Before we derive more specific message computation rules, let us focus again on the factor graph shown in Figure 4.

We see that the messages that are sent along the edges of our factor graph are of three kinds. Messages of the first kind come from a variable node, messages of the second kind come from one of the potential functions and messages of the third kind come from one of the functions T_i . These three types of messages are illustrated in Figure 5. The type of the message is superscribed in each case. As mentioned above, the messages

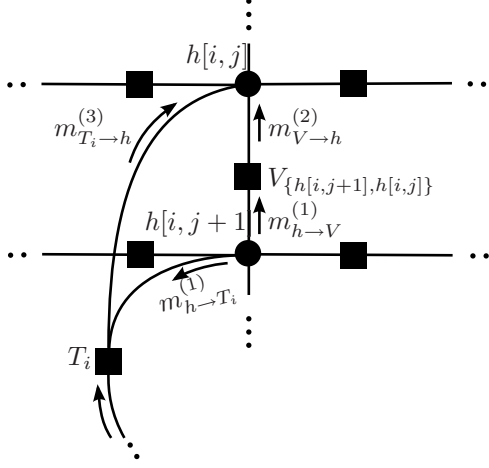


Figure 5. A closer look at the factor graph from Figure 4

are completely characterized by their mean μ_m and variance σ_m^2 . The derivation of update rules for messages of the first two kinds is straightforward; the derivation of an update rule for messages of the third kind is more involved (see Appendix A for the derivation). We just summarize the results here.

Update rule for messages of the first kind

$$\begin{aligned} \mu_{m_{h \rightarrow f}} &= \frac{\sum_{g \in n(h) \setminus \{f\}} \sigma_g^{-2} \mu_g}{\sum_{g \in n(h) \setminus \{f\}} \sigma_g^{-2}} \\ \sigma_{m_{h \rightarrow f}}^{-2} &= \sum_{g \in n(h) \setminus \{f\}} \sigma_g^{-2} \end{aligned} \quad (13)$$

where μ_g and σ_g^2 are the mean and the variance of the message $m_{g \rightarrow h}(h)$, respectively.

Update rule for messages of the second kind

$$\begin{aligned} \mu_{m_{V_b \rightarrow h}} &= \mu_u + \mu[j_h] - \mu[j_u] \\ \sigma_{m_{V_b \rightarrow h}}^{-2} &= \frac{\sigma_u^{-2} \alpha_{[i_h, j_h], [i_u, j_u]}}{\sigma_u^{-2} + \alpha_{[i_h, j_h], [i_u, j_u]}} \end{aligned} \quad (14)$$

where μ_u and σ_u^2 are the mean and the variance of the message $m_{u \rightarrow V_b}(u)$, $u \in n(V_b) \setminus \{h\}$, respectively, and $[i_h, j_h]$ and $[i_u, j_u]$ are the coordinates of the nodes h and u , respectively.

Update rule for messages of the third kind

$$\begin{aligned} \sigma_{m_{T_i \rightarrow h[i, j]}}^{-2} &= |x[i - j]|^2 \sigma_n^{-2} (1 + z)^{-1} \\ z &= \sigma_n^{-2} \sum_{l=0, l \neq j}^{L-1} |x[i - l]|^2 \sigma_{m_{h[i, l] \rightarrow T_i}}^2 \\ \mu_{m_{T_i \rightarrow h[i, j]}} &= x[i - j]^{-1} (y[i] - \sum_{l=0, l \neq j}^{L-1} x[i - l] \mu_{m_{h[i, l] \rightarrow T_i}}) \end{aligned} \quad (15)$$

The sum-product algorithm is guaranteed to converge to the correct marginals in singly connected graphs [11]. In a factor graph of arbitrary topology the marginals are in general not guaranteed to converge and even if they converge they might be incorrect. In this paper we consider the special case, where the graph features cycles and its nodes describe jointly Gaussian random variables, and in this case it can be shown that the means of the marginals are correct given that the algorithm converges [12]. As we are interested in finding the means of the marginals only, our algorithm yields a correct MAP estimate, given it converges.

VI. SIMULATION RESULTS

In order to evaluate the potential of the proposed channel model, we generate a channel (i.e., we draw the channel gains from the probability distribution that corresponds to our MRF model). In an attempt to mimic a real communications channel where the correlation in time is a lot stronger than that in delay, the parameters $\alpha_{[i, j], [l, m]}$ between vertically adjacent gains were set to 100, while the $\alpha_{[i, j], [l, m]}$ s between horizontally adjacent gains were set to 1000. The $\alpha_{[j]}$ s were all set to 100 and the means were chosen to be

$$[m[j]] = \begin{bmatrix} 0 & 0.6 & 0.4 & 1 & 0.2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .2 & .1 \end{bmatrix}$$

A typical channel realization for these parameters is depicted in Figure 6

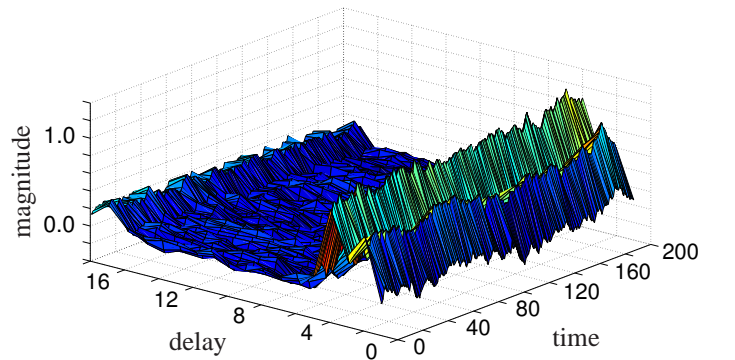


Figure 6. Channel realization used for simulations

We then applied three different channel estimators to the given observations. The first is an RLS algorithm parameterized with a forgetting factor of 0.999 to account for the strong correlation in time. The second is based on our MRF model but we set all the $\alpha_{[i, j], [l, m]}$ s between vertically adjacent gains to zero in the estimator. This way the channel

estimator neglects correlations of the channel gains in delay and channel estimation is equivalent to Kalman filtering based on a diagonal state transition matrix as proposed in [2]. The third method we evaluate is also based on our MRF model but this time the method takes correlations over delay into account. The simulation results for these three methods are shown in Figure 7. Here NMSE denotes the normalized mean

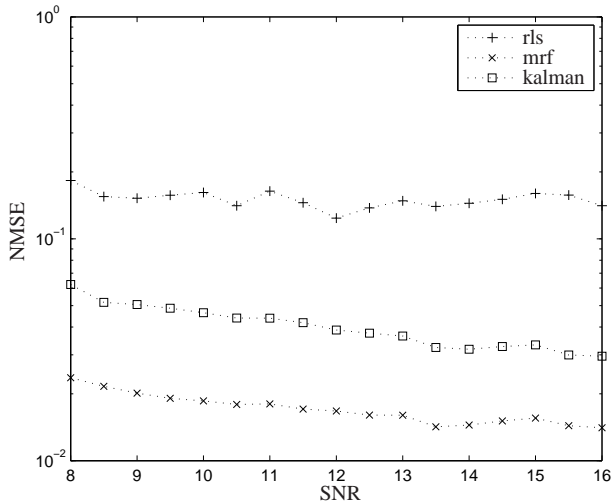


Figure 7. Channel estimation performance of our MRF model-based MAP estimation versus a conventional RLS method

squared error defined as

$$\text{NMSE} = \text{avg}_i \left[\frac{\|[\mathbf{H}]_{(i,:)} - [\hat{\mathbf{H}}]_{(i,:)}\|^2}{\|[\mathbf{H}]_{(i,:)}\|^2} \right] \quad (16)$$

where avg_i means time average. It is easily seen that taking the correlation of the channel gains in delay into account benefits the channel estimation quality significantly. In case $\alpha_{[i,j],[l,m]}$ s between vertically adjacent gains are set to smaller values for the generation of the synthetic channel, the proposed method loses part of its performance advantage, but remains superior to the conventional RLS method.

VII. CONCLUSION

So far we have assumed the parameters θ to be known. In a real communication setup this is, however, not the case. There are two ways to approach this problem. Either we guess some reasonable values a priori and accept the performance loss that comes with the introduced model mismatch, or we do actually estimate the unknown parameters from available channel observations.

In simulations we experienced that the estimation performance degraded gracefully when moving the parameters used in MRF channel estimation from their real values. Similar observations were made in [13]. So the MRF setup described in this paper achieves competitive simulation results for channel estimation as long as the MRF parameters were set in the right range.

Simulations on synthetic data, however, also showed that the estimation quality improves if the MRF model parameters match the actual channel accurately. To estimate the parameters from real channel observations, we used an *expectation maximization* (EM) like algorithm. The precise calculation of the integrals involved in the conduction of EM is not computationally feasible [14] and so we bypassed this issue by approximating the maximization step in EM by a gradient ascent step and obtained an EM-based parameter estimator.

REFERENCES

- [1] J. Proakis and M. Salehi, *Communication Systems Engineering*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, July 2002.
- [2] S. Song, A. C. Singer, and K.-M. Sung, "Soft input channel estimation for turbo equalization," *IEEE Transactions on Signal Processing*, vol. 52, no. 10, pp. 2885–2894, October 2004.
- [3] K. J. Kim and R. A. Iltis, "Joint detection and channel estimation algorithms for qs-cdma signals over time-varying channels," *IEEE Transactions on Communications*, vol. 50, no. 5, pp. 845–855, May 2002.
- [4] A. Lakhzouri, E. S. Lohan, R. Hamila, and M. Renfors, "Extended kalman filter channel estimation for line-of-sight detection in wcdma mobile positioning," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 13, pp. 1268–1278, 2003.
- [5] W. Ling and L. Ting, "Kalman filter channel estimation based on comb-type pilot in time-varying channel," *IEEE Transactions on Communications*, vol. 50, no. 5, pp. 845–855, May 2002.
- [6] A. P. Worthen and W. E. Stark, "Unified design of iterative receivers using factor graphs," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 843–849, February 2001.
- [7] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, November 1984.
- [8] L. Ying, Z.-P. Liang, D. C. Munson, R. Koetter, and B. J. Frey, "Unwrapping of mr phase images using a markov random field model," *IEEE Transactions on Medical Imaging*, vol. 25, no. 1, pp. 128–136, January 2006.
- [9] J. M. Hammersley and P. Clifford, "Markov fields on finite graphs and lattices," *Unpublished*, 1971.
- [10] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society*, vol. 36, no. 2, pp. 192–236, 1974.
- [11] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, February 2001.
- [12] Y. Weiss and W. T. Freeman, "Correctness of belief propagation in gaussian graphical models of arbitrary topology," *Neural Computation*, vol. 13, no. 10, pp. 2173–2200, 2001.
- [13] J. Zhang, "The mean field theory in em procedures for markov random fields," *IEEE Transactions on Signal Processing*, vol. 40, no. 10, pp. 2570–2583, October 1992.
- [14] D. Chandler, "Introduction to modern statistical mechanics," *Oxford University Press*, 1987.