Belief Propagation with Gaussian Approximation for Joint Channel Estimation and Decoding

Yang Liu*†, Loïc Brunel†, and Joseph J. Boutros‡
*ENST, 46 rue Barrault, 75013 Paris, France - yliu@enst.fr
†Mitsubishi Electric ITE-TCL, 1 allée de Beaulieu, 35700 Rennes, France - brunel@tcl.ite.mee.com
‡Texas A&M University at Qatar, PO Box 23874, Doha, Qatar - boutros@tamu.edu

***Abstract*—In order to increase the performance of joint channel estimation and decoding through belief propagation on factor graphs, we approximate the distribution of channel estimate in the factor graph as a mixture of Gaussian distributions. The result is a continuous downward and upward message propagation in the factor graph instead of discrete probability distributions. Using continuous downward messages, the computation complexity of belief propagation is reduced without performance degradation. With both continuous upward and downward messages, belief propagation almost achieves the same performance as expectation-maximization under good initialization and outperforms it under bad initialization.**

I. Introduction

Factor graphs [1] are utilized to solve a large variety of problems in decoding, channel estimation and detection [2] by propagating messages in the graph according to the belief propagation algorithm (BP), also called sum-product algorithm [3], and can lead to a clean derivation of iterative algorithms in a systematic way. In [4], the unified design of iterative receivers is presented, based on factor graphs and using canonical distributions (i.e., a quantization method) for handling continuous variables. However, rough quantization degrades estimation accuracy and fine quantization results in large complexity. A better performance-complexity trade-off is achieved through the expectation-maximization algorithm (EM) [5] [6]. Indeed, EM directly handles continuous variables by using a posteriori probabilities (APPs) of transmitted symbols.

In order to improve the accuracy of channel estimation and synchronization through BP, [7] and [8] introduce adaptive quantization methods, called *particle filtering*, in each iteration. However, the computation complexity is still much higher than EM complexity. Here, we propose a Gaussian approximation method which increases the precision of BP channel estimation and reduces the computation complexity simultaneously.

The paper is structured as follows. Section II describes the transmission system model and explains how BP is applied on the related factor graph. Section III shows why the distribution of channel estimates in BP can be approximated by a mixture of Gaussian distributions. In Section IV, APPs are calculated with the approximated distribution; the approximation validity is verified by numerical simulations. Channel estimation with continuous upward messages in factor graph is presented in Section V together with simulation results. Section VI draws some conclusions.

II. System Model

We consider a coded system with transmission over a quasi-static single-input single-output (SISO) channel as shown in Fig.1. An information binary sequence b_i is encoded and modulated into N BPSK symbols x_k. After multiplication by a single complex Gaussian channel coefficient $H_{\text{true}} \sim \mathcal{CN}(0,1)$ and addition of a complex Gaussian noise $n_k \sim \mathcal{CN}\left(0, 2\sigma_n^2\right)$, the channel outputs y_k are processed by a receiver performing joint channel estimation and decoding. Finally, the receiver outputs the estimated information sequence $\hat{b}_i$. The system model is described by

$$y_k = H_{\text{true}}\, x_k + n_k, \qquad 0 \le k \le N-1. \tag{1}$$

According to [4], the corresponding factor graph is shown

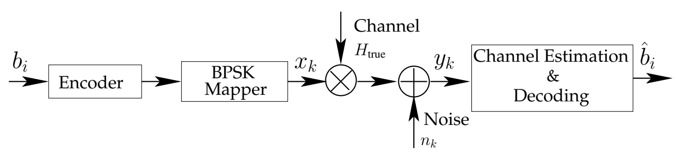

Fig. 1. System model.

together with upward messages in Fig. 2 and downward messages in Fig. 3. In node CODE, extrinsic information ξ_k is computed for each transmitted symbol (or equivalently coded bit) x_k using a forward-backward algorithm and propagated to node f_k. In each node f_k, a discrete distribution $\mu_{f_k \to p(H)}$ of channel estimate H is computed based on a marginalization of the likelihood $p\left(y_k|x_k, H\right)$ with respect to the transmitted symbol x_k. A common discrete distribution $p\left(H\right)$ is obtained from the product of all $\mu_{f_k \to p(H)}$ and propagated down to all nodes f_k. Finally, the APP of each transmitted symbol x_k is computed based on this discrete distribution, marginalizing the likelihood $p\left(y_k|x_k, H\right)$ with respect to H. The whole process of propagating upward and downward messages is then iterated. Initial estimate is obtained from pilot symbols.

III. Distribution of Channel Estimate

As pilots are known at the receiver whereas data symbols are not, the distribution of channel estimate will differ depending on if it is based on pilots or on data symbols.

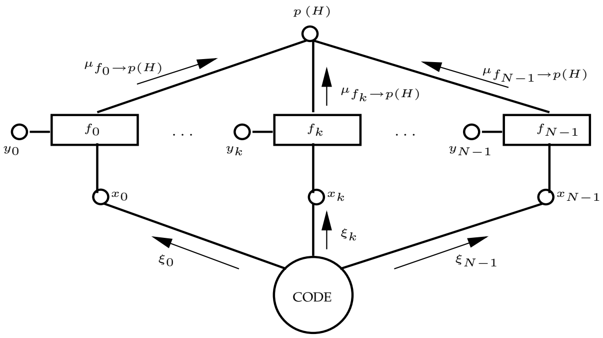

Fig. 2. Factor graph for upward message.

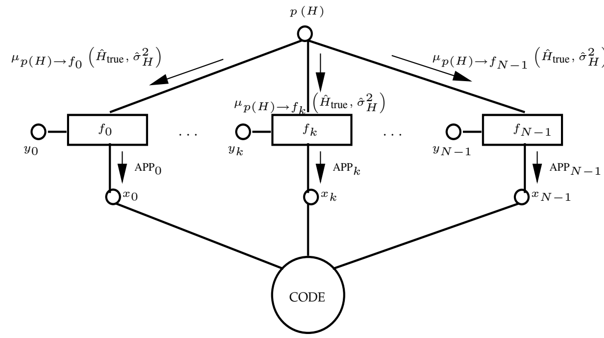

Fig. 3. Factor graph for downward message.

A. Estimation based on pilots

L_p pilot symbols $x_{\text{pilot}} = +1$ are included in the transmitted sequence. From the L_p messages provided by $f_{\text{pilot},k}$ nodes corresponding to pilots, we get the discrete distribution of H [3]:

$$p(H) = \prod_{k=0}^{L_p-1} \mu_{f_{\text{pilot},k} \to p(H)} \propto \sum_{i=0}^{q-1} \delta(H - H_i) \exp\left(-\frac{1}{2\sigma_n^2} \sum_{k=0}^{L_p-1} |y_{\text{pilot},k} - H_i|^2\right), \tag{2}$$

where

$$y_{\text{pilot},k} = H_{\text{true}}\, x_{\text{pilot},k} + n_{\text{pilot},k}, \tag{3}$$

$\{H_i\}$ is a quantization codebook of size q for channel estimate's probability density function (pdf) and $\delta(\cdot)$ denotes the Dirac delta function. For $L_p \gg 1$, we have

$$\sum_{k=0}^{L_p-1} |y_{\text{pilot},k} - H_i|^2 \approx L_p |H_i - H_{\text{true}}|^2 + 2L_p\sigma_n^2. \tag{4}$$

Surprisingly, this rough approximation for small L_p also yields a good performance. Substituting (4) into (2), we obtain

$$p(H) \propto \sum_{i=0}^{q-1} \exp\left(-\frac{L_p}{2\sigma_n^2} |H_i - H_{\text{true}}|^2\right) \delta(H - H_i). \tag{5}$$

Hence, $p(H)$ can be approximated as a Gaussian distribution $\mathcal{CN}(H_{\text{true}}, \frac{2\sigma_n^2}{L_p})$.

B. Estimation based on data

Let ξ_k be the extrinsic information for $x_k = +1$. The message from f_k to $p(H)$ can be expressed as [3]:

$$\mu_{f_k \to p(H)} \propto \sum_{i=0}^{q-1} \left\{ \exp\left(-\frac{|y_k - H_i|^2}{2\sigma_n^2}\right) \xi_k + \exp\left(-\frac{|y_k + H_i|^2}{2\sigma_n^2}\right) (1 - \xi_k) \right\} \delta(H - H_i). \tag{6}$$

Using (1), we have

$$\mu_{f_k \to p(H)} \propto \sum_{i=0}^{q-1} \left\{ \exp\left(-\frac{|H_{\text{true}} - H_i \pm n_k|^2}{2\sigma_n^2}\right) \alpha_k + \exp\left(-\frac{|H_{\text{true}} + H_i \pm n_k|^2}{2\sigma_n^2}\right) (1 - \alpha_k) \right\} \delta(H - H_i), \tag{7}$$

where

$$\alpha_k = \begin{cases} \xi_k & \text{for} \quad x_k = +1, \\ 1 - \xi_k & \text{for} \quad x_k = -1. \end{cases} \tag{8}$$

Hence, the discrete distribution of H is

$$p(H) = \prod_{k=0}^{N-1} \mu_{f_k \to p(H)} \tag{9}$$

$$\propto \sum_{i=0}^{q-1} \delta(H - H_i) \sum_{j=1}^{2^N} \left\{ \exp\left(-\frac{1}{2\sigma_n^2} \sum_{u=1}^{U_j} |H_{\text{true}} - H_i \pm n_u|^2 - \frac{1}{2\sigma_n^2} \sum_{v=1}^{V_j} |H_{\text{true}} + H_i \pm n_v|^2\right) \prod_{u=1}^{U_j} \alpha_u \prod_{v=1}^{V_j} (1 - \alpha_v) \right\} \tag{10}$$

where U_j (resp. V_j) is the number of items with α_k (resp. $(1 - \alpha_k)$) in sequence j. In (10), using mean and variance properties of n_k, we get

$$\sum_{u=1}^{U_j} |H_{\text{true}} - H_i \pm n_u|^2 + \sum_{v=1}^{V_j} |H_{\text{true}} + H_i \pm n_v|^2 \approx 2N\sigma_n^2 + N\left|H_i + \frac{V_j - U_j}{N} H_{\text{true}}\right|^2 + \left(N - \frac{(V_j - U_j)^2}{N}\right) |H_{\text{true}}|^2. \tag{11}$$

Substituting (11) into (10), we obtain

$$p(H) \propto \sum_{i=0}^{q-1} \delta(H - H_i) \sum_{j=1}^{2^N} \exp\left\{-\frac{1}{2\sigma_n^2} N \left|H_i + \frac{V_j - U_j}{N} H_{\text{true}}\right|^2\right\} \times \exp\left\{-\frac{|H_{\text{true}}|^2}{2\sigma_n^2} \left(N - \frac{(V_j - U_j)^2}{N}\right)\right\} \prod_{u=1}^{U_j} \alpha_u \prod_{v=1}^{V_j} (1 - \alpha_v). \tag{12}$$

1) At low SNR, $\xi_k \to 0.5$, i.e., $\alpha_k \to 0.5$. Hence, (12) can be approximated as

$$p(H) \propto \sum_{i=0}^{q-1} \delta(H - H_i) \left\{ \exp\left\{ -\frac{N}{2\sigma_n^2} |H_i - H_{\text{true}}|^2 \right\} \prod_{u=1}^{N} \alpha_u + \exp\left\{ -\frac{N}{2\sigma_n^2} |H_i + H_{\text{true}}|^2 \right\} \prod_{v=1}^{N} (1 - \alpha_v) \right\}. \quad (13)$$

2) At high SNR, extrinsic information from the decoder is almost perfect and (12) can be simplified into

$$p(H) \propto \sum_{i=0}^{q-1} \delta(H - H_i) \exp\left\{ -\frac{N}{2\sigma_n^2} |H_i - H_{\text{true}}|^2 \right\} \prod_{u=1}^{N} \alpha_u. \quad (14)$$

Thus, $p(H)$ can be approximated as a mixture of two Gaussian distributions:

$$p(H) \propto \sum_{i=0}^{q-1} \delta(H - H_i) \left\{ \exp\left\{ -\frac{N}{2\sigma_n^2} |H_i - H_{\text{true}}|^2 \right\} \beta + \exp\left\{ -\frac{N}{2\sigma_n^2} |H_i + H_{\text{true}}|^2 \right\} (1 - \beta) \right\}, \quad (15)$$

where β represents the product of α_k and shows how close to a single Gaussian distribution $p(H)$ is.

IV. Continuous Downward Messages

With the conclusions in Section III, the discrete distribution message $\mu_{p(H) \to f_k}$ can be reduced to one pair $(\hat{H}_{\text{true}}, \hat{\sigma}_H^2)$ characterizing $p(H)$. $\hat{H}_{\text{true}}$ is computed as the mean value of the discrete distribution in (2) (resp. (9)) and $\hat{\sigma}_H^2$ is equal to σ_n^2 / L_p (resp. σ_n^2 / N). So, we can calculate each downward message APP_k in a continuous way instead of computing it for each codebook value H_i and then marginalizing with respect to H. It reduces the computation complexity.

A. APP from pilot-based estimation

From Section III-A, for the estimation based on pilots, the channel estimate's distribution can be approximated as one Gaussian distribution. Thus, the APP for transmitted symbol $x_k = s$ ($s = +1$ or $s = -1$) can be expressed as

$$\text{APP}_{x_k = s} \propto \int_H p(y_k | x_k = s; H) \, p(H) \, dH. \quad (16)$$

After some calculations, we get

$$\text{APP}_{x_k = s} \propto \frac{1}{8\pi(\hat{\sigma}_H^2 + \sigma_n^2)} \exp\left\{ -\frac{|y_k - s\,\hat{H}_{\text{true}}|^2}{2(\hat{\sigma}_H^2 + \sigma_n^2)} \right\}. \quad (17)$$

Normalizing $\text{APP}_{x_k=+1} + \text{APP}_{x_k=-1}$ to 1, we obtain

$$\text{APP}_{x_k=+1} = \left\{ 1 + \exp\left(-\frac{-2\Re e \left\{ y_k \hat{H}_{\text{true}}^* \right\}}{\hat{\sigma}_H^2 + \sigma_n^2} \right) \right\}^{-1}. \quad (18)$$

B. APP from data-based estimation

From section III-B, we know that, for the estimation based on data, $p(H)$ can be approximated by a mixture of two Gaussian distributions. Thus,

$$\text{APP}_{x_k = s} \propto \beta \exp\left\{ -\frac{|y_k - s\,\hat{H}_{\text{true}}|^2}{2(\hat{\sigma}_H^2 + \sigma_n^2)} \right\} + (1 - \beta) \exp\left\{ -\frac{|y_k + s\,\hat{H}_{\text{true}}|^2}{2(\hat{\sigma}_H^2 + \sigma_n^2)} \right\}. \quad (19)$$

Normalizing $\text{APP}_{x_k=+1} + \text{APP}_{x_k=-1}$ to 1, we get

$$\text{APP}_{x_k=+1} = \frac{\beta + (1 - \beta) \exp\left\{ -\frac{-2\Re e \left\{ y_k \hat{H}_{\text{true}}^* \right\}}{\hat{\sigma}_H^2 + \sigma_n^2} \right\}}{1 + \exp\left\{ -\frac{-2\Re e \left\{ y_k \hat{H}_{\text{true}}^* \right\}}{\hat{\sigma}_H^2 + \sigma_n^2} \right\}}. \quad (20)$$

C. Numerical results

We compare BP using APP computation based on the Gaussian approximation (BP-QT-DGA) with BP using APP computation based on the discrete distribution (BP-QT). We also compare these two approaches using an adaptive quantization (BP-ADQT-DGA vs BP-ADQT). We use the adaptation method presented in [7]: the range of quantization codebook is shrinked in order to filter out the probability values lower than a threshold. We choose 0.2 as threshold in our simulations. For the Gaussian approximation, we always set $\beta = 1$. Indeed, our simulations have shown that this choice does not degrade performance compared to accurate evaluation of β and is less complex.

From Fig. 4 and Fig. 5, we observe that BP-QT-DGA and BP-ADQT-DGA perform as well as BP-QT and BP-ADQT respectively. Nevertheless, thanks to the computation in (18) and (20), a single APP computation instead of q computations is performed for each symbol x_k with the Gaussian approximation. Since the derivation of each message $\mu_{f_k \to p(H)}$ still involves q computations, one per codebook value, the global complexity reduction brought by the Gaussian approximation in the downward messages is approximately 50% for large q.

However, all BP approaches simulated in this section do not outperform or even approach the performances of EM.

V. Continuous Upward Messages

In order to improve the performance of the Gaussian approximation, we propose to increase the accuracy of $\hat{H}_{\text{true}}$ using a continuous upward message.

A. Upward messages from pilots

By replacing the quantization codebook $\{H_i\}$ in (2) with a continuous value H and considering normalization, we get

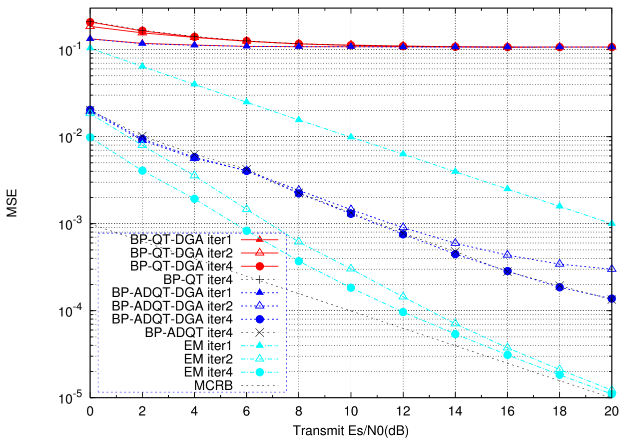

Fig. 4. Mean square error performance comparison of BP and EM. Rate 1/2 convolutional code (23, 35) with block length 1000. Number of pilots is 10. The initial quantization interval is [-10,+10] with a step equal to 0.8.

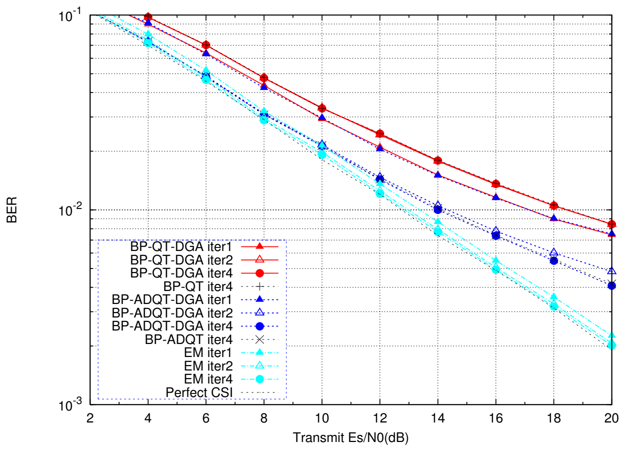

Fig. 5. Bit error rate performance comparison of BP and EM. Rate 1/2 convolutional code (23, 35) with block length 1000. Number of pilots is 10. The initial quantization interval is [-10,+10] with a step equal to 0.8.

the following estimated mean value of H:

$$\hat{H}_{\text{true}} = \frac{\int_H H \frac{1}{(2\pi\sigma_n^2)^{L_p}} \exp\left\{-\frac{1}{2\sigma_n^2}\sum_{k=0}^{L_p-1}|y_k - H|^2\right\} dH}{\int_H \frac{1}{(2\pi\sigma_n^2)^{L_p}} \exp\left\{-\frac{1}{2\sigma_n^2}\sum_{k=0}^{L_p-1}|y_k - H|^2\right\} dH}. \tag{21}$$

The integration of numerator and denominator in (21) results in the following simple formula:

$$\hat{H}_{\text{true}} = \frac{1}{L_p}\sum_{k=0}^{L_p-1} y_k. \tag{22}$$

B. Upward messages from data

Replacing the discrete computation in (6) by an integral, we get the mean value of H in (23), in which

$$\prod_{k=0}^{N-1}\left[\exp\left(-\frac{|y_k - H|^2}{2\sigma_n^2}\right)\xi_k + \exp\left(-\frac{|y_k + H|^2}{2\sigma_n^2}\right)(1-\xi_k)\right]$$
$$= \sum_{j=0}^{2^N-1}\exp\left(-\frac{1}{2\sigma_n^2}\sum_{k=0}^{N-1}|y_k - s_{k,j}H|^2\right)\Delta_j, \tag{24}$$

where $s_{k,j}$ represents the value (-1 or +1) of the kth symbol in sequence j and

$$\Delta_j = \prod_{s_{k,j}=+1}\xi_k \prod_{s_{k,j}=-1}(1-\xi_k). \tag{25}$$

After some calculations, we obtain (26). In order to reduce the computation in (26), we only consider the item with the largest Δ_j in both numerator and denominator:

$$\hat{H}_{\text{true}} \approx \frac{1}{N}\sum_{k=0}^{N-1} s_{k,j_{\max}} y_k, \tag{27}$$

where $j_{\max} = \arg\max_j \Delta_j$.

$$\hat{H}_{\text{true}} = \frac{\int_H H \prod_{k=0}^{N-1}\left[\exp\left(-\frac{|y_k - H|^2}{2\sigma_n^2}\right)\xi_k + \exp\left(-\frac{|y_k + H|^2}{2\sigma_n^2}\right)(1-\xi_k)\right] d\,H}{\int_H \prod_{k=0}^{N-1}\left[\exp\left(-\frac{|y_k - H|^2}{2\sigma_n^2}\right)\xi_k + \exp\left(-\frac{|y_k + H|^2}{2\sigma_n^2}\right)(1-\xi_k)\right] d\,H}. \tag{23}$$

$$\hat{H}_{\text{true}} = \frac{\sum_{j=0}^{2^N-1}\Delta_j \frac{1}{N}\left(\sum_{k=0}^{N-1} s_{k,j}y_k\right)\exp\left\{\frac{1}{2N\sigma_n^2}\left|\sum_{k=0}^{N-1} s_{k,j}y_k\right|^2\right\}}{\sum_{j=0}^{2^N-1}\Delta_j \exp\left\{\frac{1}{2N\sigma_n^2}\left|\sum_{k=0}^{N-1} s_{k,j}y_k\right|^2\right\}}. \tag{26}$$

C. Numerical results

We now compare BP with continuous downward and upward messages (BP-DUGA) with EM. From Fig. 6, we observe that, with 10 pilots, i.e., good initialization, the proposed BP-DUGA achieves bit error rate (BER) performance close to EM.

Furthermore, from Fig. 7 and Fig. 8, with 1 pilot for initial estimate, we observe that the proposed BP-DUGA performs better than EM: BP-DUGA's mean square error (MSE) performance in Fig. 7 is closer to modified Cramér-Rao bound (MCRB) [9]; for BER performance in Fig. 8, we obtain a gain of about 1 dB with BP-DUGA compared to EM and almost achieve the performance with perfect channel state information. Using continuous upward messages brings a complexity reduction compared to BP-QT-DGA and BP-ADQT-DGA. As a result, BP-DUGA complexity becomes equivalent to EM complexity.

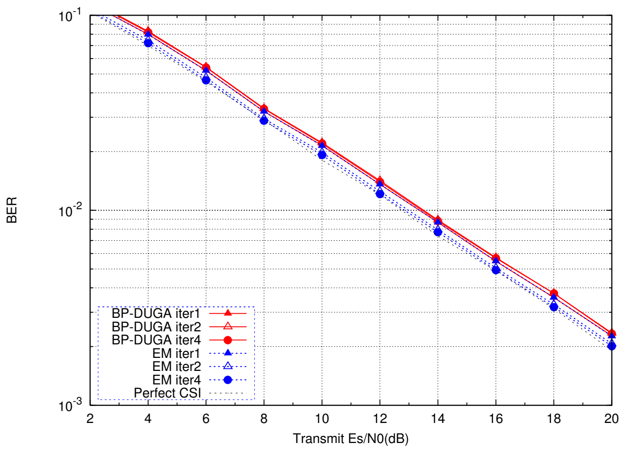

Fig. 6. Bit error rate performance comparison for channel estimation: BP-DUGA vs EM. Rate 1/2 convolutional code (23, 35) with block length 1000. Number of pilots is 10.

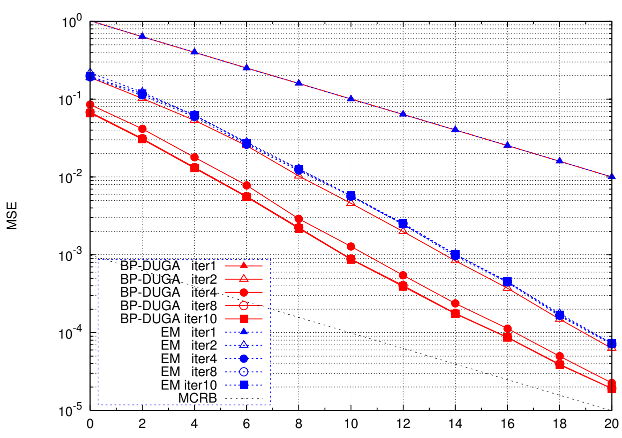

Fig. 7. Mean square error performance comparison for channel estimation: BP-DUGA vs EM. Rate 1/2 convolutional code (23, 35) with block length 1000. Number of pilots is 1.

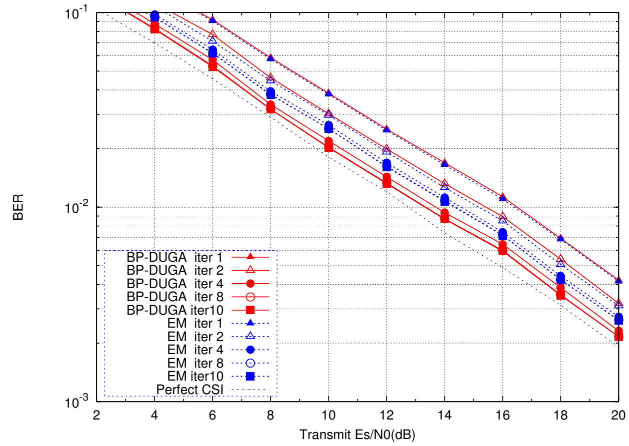

Fig. 8. Bit error rate performance comparison for channel estimation: BP-DUGA vs EM. Rate 1/2 convolutional code (23, 35) with block length 1000. Number of pilots is 1.

VI. Conclusion

We validated that the distribution of channel estimate in a factor graph can be approximated as a mixture of Gaussian distributions. With this approximation, we derived continuous downward and upward messages to be propagated in the factor graph by BP. Through performance improvement brought by a continuous upward message, BP-DUGA almost achieves EM performance under a good initialization and outperforms it under a bad initialization. Furthermore, thanks to both continuous downward and upward messages, BP-DUGA computation complexity is equivalent to EM. This paper is focusing on a single-path channel. Nevertheless, the extension of the Gaussian approximation principle to a multipath channels is natural.

References

[1] H.-A. Loeliger, "An introduction to factor graphs," *IEEE Signal Process. Mag.*, vol. 21, pp. 28–41, Jan. 2004.

[2] H. Wymeersch, H. Steendam, H. Bruneel, and M. Moeneclaey, "Code-aided frame synchronization and phase ambiguity resolution," *IEEE Trans. Signal Process.*, vol. 54, no. 7, pp. 2747–2757, Jul. 2006.

[3] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[4] A. P. Worthen and W. E. Stark, "Unified design of iterative receivers using factor graphs," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 843–849, Feb. 2001.

[5] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society (B)*, vol. 13, no. 6, pp. 1–38, 1977.

[6] A. S. Gallo and G. M. Vitetta, "Soft-in soft-output detection in the presence of parametric uncertainty via the bayesian em algorithm," *EURASIP JWCN*, vol. 2005, no. 2, pp. 100–116, 2005.

[7] F. Z. Merli and G. M. Vitetta, "A factor graph approach to the iterative detection of ofdm signals in the presence of carrier frequency offset and phase noise," in *Proc. IEEE ICC '07*, Jun. 2007, pp. 2865–2870.

[8] J. Dauwels, S. Korl, and H.-A. Loeliger, "Particle methods as message passing," in *Proc. IEEE ISIT '06*, Jul. 2007, pp. 2052–2056.

[9] A. N. D'Andrea, U. Mengali, and R. Reggiannini, "The modified cramer-rao bound and its application to synchronization problems," *IEEE Trans. Commun.*, vol. 42, no. 234, Part 2, pp. 1391–1399, Feb./Apr. 1994.