# Expectation-Maximization Belief Propagation for Sparse Recovery I: Algorithm Construction

Main ref.: F. Krzakala *et al.*,"Statistical physics-based re construction in compressed sensing," *Phys. Rev X*, May, 2012 [1]

Presenter: Jaewook Kang
In Proceeding of CS Journal Club in GIST

Date: Nov. 2012, updated at Jan, 2013

## I. Introduction

In this report, the author introduces a expectation maximization (EM) based belief propagation algorithm (BP) for sparse recovery, named EM-BP. The algorithm have been mainly devised by Krzakala *et al.* from ParisTech in France [1]. The properties of EM-BP are as given below:

1) It is A low-computation approach to sparse recovery,
2) It works well without the prior knowledge of the signal,
3) It overcomes the $l_1$ phase transition given by Donoho and Tanner [11] under the noiseless setup,
4) It is further improved in conjunction with seeding matrices (or spatial coupling matrices).

The main purpose of this report regenerates a precise description of EM-BP algorithm construction from the reference paper [1]. It might be very helpful for understanding of EM-BP algorithm, and an answer for such a question: How and why does the algorithm work ? Therefore, we will focus on the explanation of 1) and 2) in the properties, and just show the result of the paper with respect to that of 3) and 4).

In addition to EM-BP, the belief propagation approach to the sparse recovery problem has been widely investigated in [2],[3],[4],[5],[6],[7].

## II. Problem setup

In the sparse recovery problem, the aim is to recovery a sparse signal $\mathbf{X} \in \mathbb{R}^N$ whose elements have nonzero value independently each other, with a probability rate $q$ called sparsity rate. Therefore, the $q$ determines the density of signal $\mathbf{X}$. Then, the algorithm performs the recovery from the measurements $\mathbf{Y} \in \mathbb{R}^M$, given as

$$\mathbf{Y} = \mathbf{\Phi X} + \mathbf{N}, \tag{1}$$

where $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ is a fat measurement matrix with $M < N$, and $\mathbf{N} \in \mathbb{R}^{M}$ denotes a additive Gaussian noise vector following $\mathcal{N}(0, \mathbf{I}\sigma_N^2)$.

## III. Algorithm Construction of EM-BP

Krzakala *et al.* has taken a probabilistic approach to devise EM-BP. From the Bayesian point of view, the posterior density of the signal $\mathbf{X}$ is represented in the form of $\text{Posterior} = \text{Prior} \times \frac{\text{Likelihood}}{\text{Evidence}}$ as

$$f_{\mathbf{X}}(\mathbf{x}|\mathbf{y}, \mathbf{\Phi}) = f_{\mathbf{X}}(\mathbf{x}|\mathbf{\Phi}) \times \frac{f_{\mathbf{Y}}(\mathbf{y}|\mathbf{\Phi}, \mathbf{X})}{f_{\mathbf{Y}}(\mathbf{y}|\mathbf{\Phi})}. \tag{2}$$

Then, using the knowledge of $\mathbf{z}$ and $\mathbf{\Phi}$, the signal posterior is given as

$$f_{\mathbf{X}}(\mathbf{x}|\mathbf{y}, \mathbf{\Phi}) = \frac{1}{C} f_{\mathbf{X}}(\mathbf{x}) \times \prod_{j=1}^{M} \frac{1}{\sqrt{2\pi\sigma_N^2}} \exp\left[-\frac{1}{2\sigma_{N_j}^2}(y_j - \textstyle\sum_{i=1}^{N} \phi_{ji} x_i)^2\right], \tag{3}$$

where $C$ is a normalization constant for $\int f_{\mathbf{X}}(\mathbf{x}|\mathbf{y}, \mathbf{\Phi})\, d\mathbf{x} = 1$. In addition, we consider a mixture type prior density function represented as

$$f_{\mathbf{X}}(\mathbf{x}) := \prod_{i=1}^{N} [(1-q)\delta_0 + q\theta(x_i)], \tag{4}$$

where $\theta(x_i)$ is a Gaussian PDF with mean $\bar{x}$ and variance $\sigma_X^2$.

Exact finding of the signal posterior is computationally infeasible. Therefore, researchers have employed BP as a standard approach to approximate the signal posterior where BP finds marginal posterior density of each signal element $X_i$. In addition, Guo *et al.* showed that the marginal posterior finding is exact if the matrix $\mathbf{\Phi}$ is a sparse matrix and $N \to \infty$ [8],[9],[10]. BP seeks the signal posterior by iteratively exchanging probabilistic messages over the signal elements, where the messages are classically described as

Measurement to signal (MtS) message :

$$m_{j \to i}(x_i) := \frac{1}{C_{j \to i}} \int\limits_{\{x_k\}_{k \neq i}} \prod_{k \neq i} m_{k \to j}(x_i) \times \exp\left[-\frac{1}{2\sigma_N^2}(\sum_{k \neq i} \phi_{jk} x_k + \phi_{ji} x_i - y_j)^2\right] \left(\prod_{k \neq i} dx_i\right), \tag{5}$$

Signal to measurement (StM) message :

$$m_{i \to j}(x_i) := \frac{1}{Z_{i \to j}} [(1-q)\delta_0 + q\theta(x_i)] \times \prod_{k \neq j} m_{k \to i}(x_i), \tag{6}$$

where $C_{j \to i}$ and $Z_{i \to j}$ are normalization constants to make the messages as PDFs. Then, the marginal posterior approximately is obtained as

$$f_{X_i}(x|\, \mathbf{y}, \mathbf{\Phi}) \overset{\text{BP}}{=} \frac{1}{C_i} [(1-q)\delta_{x_i} + q\theta(x_i)] \times \prod_{k} m_{k \to i}(x_i). \tag{7}$$

However, the message update rule in (5) and (6) is practically intractable because each BP message is probability density function (PDF). Therefore, we need to convert the density-passing procedure to a parameter-passing procedure using some relaxation techniques.

Using Hubbard-Stratonovich transformation (HST) from spin glass theory which is

$$\exp\left(-\frac{w^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \int \exp\left(-\frac{\lambda^2}{2\sigma^2} + \frac{iw\lambda}{\sigma^2}\right) d\lambda, \tag{8}$$

the exponent in (5) can be rewritten as

$$\exp\left[-\frac{1}{2\sigma^2_{N_j}}(y_j - \textstyle\sum_{i=1}^N \phi_{ji}x_i)^2\right] = \exp\left[\underbrace{-\frac{\left(\sum\limits_{k\neq i} \phi_{jk}x_k\right)^2}{2\sigma^2_{N_j}}}_{\text{Here, HST applied}} - \frac{\sum\limits_{k\neq i} \phi_{jk}x_k(\phi_{ji}x_i - y_j)}{\sigma^2_{N_j}} - \frac{(\phi_{ji}x_i - y_j)^2}{2\sigma^2_{N_j}}\right]$$

$$= \frac{1}{\sqrt{2\pi\sigma^2_{N_j}}} \int\limits_{\lambda} \exp\left(-\frac{\lambda^2}{2\sigma^2_{N_j}} + \frac{\sum\limits_{k\neq i} \phi_{jk}x_k(\phi_{ji}x_i - y_j + i\lambda)}{\sigma^2_{N_j}} - \frac{(\phi_{ji}x_i - y_j)^2}{2\sigma^2_{N_j}}\right) d\lambda. \tag{9}$$

By applying (9) to (5), we have

$$\begin{aligned} m_{j\to i}(x_i) = &\frac{\exp\left(-\frac{(\phi_{ji}x_i - y_j)^2}{2\sigma^2_{N_j}}\right)}{C_{j\to i}\sqrt{2\pi\sigma^2_{N_j}}} \int\limits_{\lambda} \exp\left(-\frac{\lambda^2}{2\sigma^2_{N_j}}\right) \\ &\times \left\{ \int\limits_{\{x_k\}_{k\neq i}} \prod_{k\neq i} m_{k\to j}(x_i) \exp\left(\frac{\sum\limits_{k\neq i}\phi_{jk}x_k(\phi_{ji}x_i - y_j + i\lambda)}{\sigma^2_{N_j}}\right) \prod_{k\neq i} dx_i \right\} d\lambda \end{aligned} \tag{10}$$

In (10), we observe that the integration over $\{x_k\}_{k\neq i}$ can be decomposed into integration over each scalar $x_k$. In addition, the integration over scalar $x_k$ takes the form of the moment generating function. Therefore,

$$\begin{aligned} m_{j\to i}(x_i) &= \frac{\exp\left(-\frac{(\phi_{ji}x_i - y_j)^2}{2\sigma^2_{N_j}}\right)}{C_{j\to i}\sqrt{2\pi\sigma^2_{N_j}}} \int\limits_{\lambda} \exp\left(-\frac{\lambda^2}{2\sigma^2_{N_j}}\right) \times \prod_{k\neq i} \left\{ \int\limits_{\{x_k\}_{k\neq i}} m_{k\to j}(x_i) \exp\left(\frac{x_k\phi_{jk}(\phi_{ji}x_i - y_j + i\lambda)}{\sigma^2_{N_j}}\right) dx_i \right\} d\lambda \\ &= \frac{\exp\left(-\frac{(\phi_{ji}x_i - y_j)^2}{2\sigma^2_{N_j}}\right)}{C_{j\to i}\sqrt{2\pi\sigma^2_{N_j}}} \int\limits_{\lambda} \exp\left(-\frac{\lambda^2}{2\sigma^2_{N_j}}\right) \times \prod_{k\neq i} \mathbf{E}_{X_k}\left[\exp\left(\frac{x_k\phi_{jk}(\phi_{ji}x_i - y_j + i\lambda)}{\sigma^2_{N_j}}\right)\right] d\lambda \end{aligned} \tag{11}$$

By assuming that each scalar $X_k$ is Gaussian distributed during the BP-iteration with mean $\mu_{i\to j}$ and variance $\sigma^2_{i\to j}$, we can approximate the MtS message expression as

$$m_{j\to i}(x_i) \approx \frac{\exp\left(-\frac{(\phi_{ji}x_i-y_j)^2}{2\sigma^2_{N_j}}\right)}{C_{j\to i}\sqrt{2\pi\sigma^2_{N_j}}}\times$$
$$\times\int_\lambda \exp\left(-\frac{\lambda^2}{2\sigma^2_{N_j}}\right)\prod_{k\neq i}\exp\left(\frac{\mu_{k\to j}\phi_{jk}(\phi_{ji}x_i-y_j+i\lambda)}{\sigma^2_{N_j}}+\frac{\sigma^2_{k\to j}}{2}\left(\frac{\phi_{jk}(\phi_{ji}x_i-y_j+i\lambda)}{\sigma^2_{N_j}}\right)^2\right)d\lambda. \tag{12}$$

By evaluating the Gaussian integration over $\lambda$, the expression in (12) becomes

$$m_{j\to i}(x_i) \simeq \frac{\sqrt{A_{j\to i}/2\pi}}{\phi_{ji}C_{j\to i}}\times\exp\left(-\frac{x_i^2}{2}A_{j\to i}+x_iB_{j\to i}+\frac{B^2_{j\to i}}{2A_{j\to i}}\right). \tag{13}$$

where

$$A_{j\to i} := \frac{\phi^2_{ji}}{\sigma^2_{N_j}+\sum\limits_{k\neq j}\sigma^2_{k\to j}\phi^2_{jk}}, \tag{14}$$

$$B_{j\to i} := \frac{\phi_{ji}(y_j-\sum\limits_{k\neq j}\mu_{k\to j}\phi_{jk})}{\sigma^2_{N_j}+\sum\limits_{k\neq j}\sigma^2_{k\to j}\phi^2_{jk}}. \tag{15}$$

Then, the expression of the StM message is rewritten as

$$m_{i\to j}(x_i) := \frac{1}{\widetilde{Z}_{i\to j}}[(1-q)\delta_0+q\theta(x_i)]\times\exp\left(-\frac{x_i^2}{2}\sum_{k\neq j}A_{k\to i}+x_i\sum_{k\neq j}B_{k\to i}+\frac{1}{2}\frac{\left(\sum\limits_{k\neq j}B_{k\to j}\right)^2}{\sum\limits_{k\neq j}A_{k\to j}}\right), \tag{16}$$

where we use an approximation $\sum\limits_{k\neq j}B^2_{k\to j}\approx\left(\sum\limits_{k\neq j}B_{k\to j}\right)^2$. The exponent can be rewritten as

$$-\frac{x_i^2}{2}\sum_{k\neq j}A_{k\to i}+x_i\sum_{k\neq j}B_{k\to i}+\frac{1}{2}\frac{\left(\sum\limits_{k\neq j}B_{k\to j}\right)^2}{\sum\limits_{k\neq j}A_{k\to j}}$$
$$=-\frac{1}{2\frac{1}{\sum\limits_{k\neq j}A_{k\to i}}}\left(x_i^2-2\frac{\sum\limits_{k\neq j}B_{k\to i}}{\sum\limits_{k\neq j}A_{k\to i}}+\left(\frac{\sum\limits_{k\neq j}B_{k\to i}}{\sum\limits_{k\neq j}A_{k\to i}}\right)^2\right)=-\frac{\left(x_i-\frac{\sum_{k\neq j}B_{k\to i}}{\sum_{k\neq j}A_{k\to i}}\right)^2}{2\frac{1}{\sum_{k\neq j}A_{k\to i}}} \tag{17}$$

Hence, equations (14) and (15) together with (19) fully describe the iterative BP-process. We define two variable given as

$$\Sigma_i^2 := \frac{1}{\sum\limits_{k\neq j} A_{k\to i}},\ R_i := \frac{\sum\limits_{k\neq j} B_{k\to i}}{\sum\limits_{k\neq j} A_{k\to i}}, \tag{18}$$

Using the notations, we rewrite the expression of the StM message given as Then, the expression of the StM message is rewritten as

$$m_{i\to j}(x_i) := \frac{1}{\widetilde{Z}_{i\to j}}[(1-q)\delta_0 + q\theta(x_i)] \times \exp\left(-\frac{(x_i - R_i)^2}{2\Sigma_i^2}\right), \tag{19}$$

Then, the mean $\mu_{k\to j}$ and variance $\sigma_{k\to j}^2$ of the StM message are calculated as

$$\begin{aligned}
\mu_{i\to j} &:= \int\limits_{X_i} x_i m_{i\to j}(x_i) dx_i \\
&= \frac{q}{Z(\Sigma_i^2, R_i)} \int\limits_{X_i} x_i \theta(x_i) \exp\left(-\frac{(x_i - R_i)^2}{2\Sigma_i^2}\right) dx_i \\
&= \frac{q}{Z(\Sigma_i^2, R_i)} \times \frac{\Sigma_i(\overline{x}\Sigma_i^2 + R\sigma_X^2)}{(\Sigma_i^2 + \sigma_X^2)^{3/2}} \exp\left(-\frac{(R-\overline{x})^2}{2(\Sigma_i^2 + \sigma_X^2)}\right),
\end{aligned} \tag{20}$$

and

$$\begin{aligned}
\sigma_{i\to j}^2 &:= \int\limits_{X_i} x_i^2 m_{i\to j}(x_i) dx_i - \mu_{i\to j}^2 \\
&= \frac{q}{Z(\Sigma_i^2, R_i)} \int\limits_{X_i} x_i^2 \theta(x_i) \exp\left(-\frac{(x_i - R_i)^2}{2\Sigma_i^2}\right) dx_i - \mu_{i\to j}^2 \\
&= \frac{q(1-q)\exp\left(-\frac{R_i^2}{2\Sigma_i^2} - \frac{(R-\overline{x})^2}{2(\Sigma_i^2+\sigma_X^2)}\right) \frac{\Sigma_i}{(\Sigma_i^2+\sigma_X^2)^{5/2}} \left(\sigma_X^2\Sigma_i^2(\Sigma_i^2+\sigma_X^2) + (\overline{x}\Sigma_i^2 + R\sigma_X^2)^2\right)}{Z(\Sigma_i^2, R_i)^2} \\
&+ \frac{q^2 \exp\left(-\frac{(R-\overline{x})^2}{2(\Sigma_i^2+\sigma_X^2)}\right) \frac{\sigma_X^2\Sigma_i^4}{(\Sigma_i^2+\sigma_X^2)^2}}{Z(\Sigma_i^2, R_i)^2},
\end{aligned} \tag{21}$$

where the normalization constant is

$$\begin{aligned}
Z(\Sigma_i^2, R_i) :&= (1-q)\int\limits_{X_i} \delta_0 \exp\left(-\frac{(x_i - R_i)^2}{2\Sigma_i^2}\right) dx_i + q\int\limits_{X_i} \theta(x_i) \exp\left(-\frac{(x_i - R_i)^2}{2\Sigma_i^2}\right) dx_i \\
&= (1-q)\exp\left(-\frac{R_i^2}{2\Sigma_i^2}\right) + q\frac{\Sigma_i}{\sqrt{\Sigma_i^2 + \sigma_X^2}} \exp\left(-\frac{(R-\overline{x})^2}{2(\Sigma_i^2 + \sigma_X^2)}\right).
\end{aligned} \tag{22}$$

The authors stated that the parameters $\bar{x}$, $\sigma_X^2$, and $q$ of the prior density $f_{\mathbf{X}}(\mathbf{X})$ can be learned and updated at every iteration. A statistical approach for the parameter learning is the use of EM. For the object function in EM, they used Bethe free-entropy. It is known that BP algorithm is constructed by applying Lagrange multipliers to Bethe entropy [12]. Therefore, the fixed point of BP-iteration corresponds to the stationary points of the Bethe free-entropy minimization, in the signal posterior finding problems. For details about the relationship between Bethe free-entropy and BP, please see Yedidia's paper.

The Bethe entropy is defined as

$$H_{\text{Bethe}} := -\sum_{i}^{N} H(Z_{x_i}) - \sum_{j}^{M} H(Z_{y_j}) + \sum_{j}^{M} \sum_{i \in N(j)} H(Z_{x_i}), \tag{23}$$

where the concept of free-entropy, defined as $H(Z) := \log Z$, is used and $Z_{x_i}$ and $Z_{y_j}$ are an approximated marginal partition function of $x$, that is,

$$Z_{x_i} = \int [(1-q)\delta_{x_i} + q\theta(x_i)] \times \prod_j m_{j\to i}(x_i) dx_i, \tag{24}$$

$$Z_{y_j} = \int \prod_i m_{i\to j}(x_i) \times \exp\left[-\frac{1}{2\sigma_N^2}(\sum_i \phi_{ji}x_i - y_j)^2\right] \prod_i (dx_i)\,. \tag{25}$$

Thus, the parameters $(\bar{x}, \sigma_X, q)$ are learned by seeking the stationary point of the Bethe free-entropy function given in (23). We update the parameter for the prior knowledge from

$$\overline{x} = \frac{\sum\limits_i \mu_i}{Nq} \tag{26}$$

$$\sigma_X^2 = \frac{\sum\limits_i (\sigma_i^2 + \mu_i^2)}{Nq} - \overline{x}^2 \tag{27}$$

$$q = \frac{\sum_i \frac{1/\sigma_X^2 + \sum_j A_{j\to i}}{\sum_j B_{j\to i} + \overline{x}/\sigma_X^2}\mu_i}{\sum_i \left(1 - q + \frac{q}{\sigma_X\sqrt{1/\sigma_X^2 + \sum_j A_{j\to i}}} \exp\left(\frac{(\sum_j B_{j\to i} + \overline{x}/\sigma_X^2)^2}{2(1/\sigma_X^2 + \sum_j A_{j\to i})} - \frac{\overline{x}^2}{2\sigma_X^2}\right)\right)^{-1}}. \tag{28}$$

I implemented the EM-BP algorithm using the equations of (14), (15), (20), (21), (22), (26), (27) in C language. I did not update the sparsity rate $q$ in the BP-iteration. The performance is not working well as shown in Fig.1. I need to check my implementation by translating the code to MATLAB. I think the EM update not much improve the performance. So, we need to modify the update rule to elementwise update rule like SuPrEM Algorithm.

## REFERENCES


[1] F. Krzakala, M. Mezard, F. Sausset, Y. Sun, L. Zdeborova, "Statistical physics-based reconstruction in compressed sensing," *Phys. Rev X*, No. 021005, May, 2012

[2] J. Kang, H.-N. Lee, and K. Kim, "Detection-directed sparse estimation using bayesian hypothesis test and belief propagation," submitted to *IEEE Trans. Signal Process.*, 2012

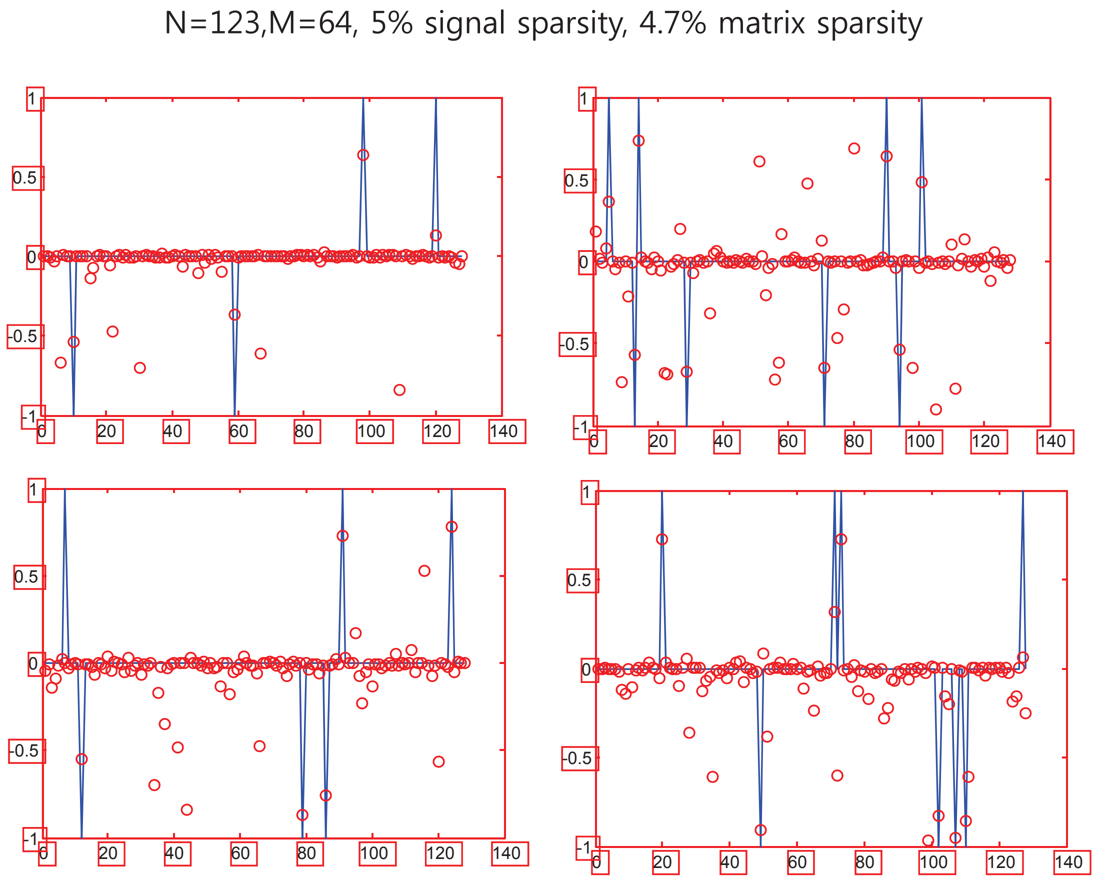


Fig. 1. Some simulation results of EM-BP when $N = 128, M = 64, q = 0.05$, $\sigma_X = 1$, and $L = 3$.


[3] D. Baron, S. Sarvotham, and R. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 269-280, Jan. 2010.

[4] X. Tan and J. Li, "Computationally efficient sparse Bayesian learning via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2010-2021, Apr. 2010.

[5] M. Akcakaya, J. Park, and V. Tarokh, "A coding theory approach to noisy compressive sensing using low density frame," *IEEE Trans. Signal Process.,* vol. 59, no. 12, pp. 5369-5379, Nov. 2011.

[6] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing,", *Proc. Nat. Acad. Sci.*, vol. 106, pp. 18914-18919, Nov. 2009.

[7] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," *Proc. in IEEE Int. Symp. Inform. Theory (ISIT)*, pp. 2168-2172, Aug. 2011.

[8] D. Guo and S. Verdu, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 1983-2010, Jun. 2005.

[9] D. Guo and C.-C. Wang, "Asymptotic mean-square optimality of belief propagation for sparse linear systems," *Proc. in*

*IEEE Inform. Theory Workshop*, pp. 194-198, Chengdu, China, Oct. 2006.

[10] D. Guo and C.-C. Wang, "Random sparse linear systems observed via arbitrary channels: a decoupling principle,"*Proc. in IEEE Int. Symp. Inform. Theory (ISIT)*, pp. 946-950, Nice, France, June. 2007.

[11] D.L. Donoho and J. Tanner, "Precise undersampling theorems," *Proceeding of the IEEE*, vol. 98, issue 6, pp. 913-924, June, 2010.

[12] J.S. Yedidia, W.T. Freeman, and Y. Weiss, "Construcing Free-Energy Approximations and Generalized Belief Propagation Algorithm," *IEEE Trans. Inform. Theory*, vol. 51, no. 7, pp. 2282-2312, July. 2005.