ABSTRACT
Sparse Bayesian Learning (SBL) provides sophisticated (state) model order selection with unknown support distribution. This allows to handle problems with big state dimensions and relatively limited data by exploiting variations in parameter importance. The techniques proposed in this paper allow to handle the extension of SBL to time-varying states, modeled as diagonal first-order auto-regressive (DAR(1)) processes with unknown parameters to be estimated also. Adding the parameters to the state leads to an augmented state and a non-linear (at least bilinear) state-space model. The proposed approach, which applies also to more general non-linear models, uses a combination of belief propagation (BP), Variational Bayes (VB) or mean field (MF) techniques, and Expectation Propagation (EP) to approximate the posterior marginal distributions of the scalar factors. We propose Fisher Information Matrix analysis to determine the variable split between the use of BP and VB allowing to stay optimal in terms of Laplace approximation.

1. INTRODUCTION
The signal model for the recovery of a time varying sparse signal can be formulated as, $y_t = A^{(t)} x_t + v_t$, where $y_t$ is the observations or data at time $t$, $A^{(t)}$ is called the measurement or the sensing matrix which is known and is of dimension $N \times M$ with $N < M$. $x_t$ contains only $K$ non-zero entries, with $K << M$ and is modeled by a diagonal AR(1) (auto-regressive) process. In the static case, in Bayesian learning, the sparse Bayesian learning (SBL) algorithm was first proposed by [1, 2], which got extended to dynamic SBL in [3]. However, in order to render low complexity or low latency solutions, online processing algorithms (which process a small set of measurement vectors at any time) will be necessary. Dynamic autoregressive SBL (DAR-SBL) considered here is a case of joint Kalman filtering (KF) with a linear time-invariant diagonal state-space model, and parameter estimation, which can be considered an instance of nonlinear filtering.

In [4], they introduce a belief propagation (BP) based SBL algorithm which is more computationally efficient than the original algorithm. The authors use BP to infer the posterior pdf of $x$ and the hyperparameters are estimated using the EM algorithm. The authors in [5] propose a message passing (MP) approach for inferring the posteriors combining BP and mean field (MF) approximations. MF is a special case of Variational Bayes (VB) in which the partitioning of variables is pushed to the scalar granularity. The advantages of the MF approach are that it always admits a convergent implementation while BP yields a good approximation of the posterior marginals if the factor graph has no cycles. The authors show that the MF fixed-point equations for a combination of BP and the MF approximation correspond to stationary points of one single constrained region-based free energy approximation and provide a clear rule stating how to couple the messages propagating in the BP and MF part. Hence, it is advantageous to apply BP and the MF approximation on the same factor graph in such a combination that their respective virtues can be exploited while overcoming their drawbacks (complexity for BP, potential suboptimality for MF). However, [6] does not treat at all the topic of how to split nodes between BP and MF. We also note that the approximate message passing algorithms [6, 7] suffer from the limitation that the large system limits assume i.i.d. Gaussian or right rotationally invariant $A^{(t)}$, and the algorithms may exhibit convergence problems.

1.1. Contributions of this paper
- We propose new low complexity SBL algorithms for the static and dynamic cases, with joint hyperparameter estimation.
- Building on the framework of [5], we combine BP and MF approximations in such a way as to optimize the message passing framework, unlike most of the existing applications of the framework, which apply BP and MF to the variable subsets with discrete and continuous distributions resp.
- Using Fisher Information Matrix (FIM) analysis, we propose an optimal partitioning of the unknown parameters in the factor graph such that we can combine BP and (EP) VB in an efficient way, with low complexity and no suboptimality in terms of Laplace approximation (FIM).
- Various new algorithms in this paper are an application of these parameter partitioning and BP/VB split guidelines. For both a static (classical) compressed sensing model or a dynamic case with autoregressive evolution of the unknown $x$ (corresponding to a classical linear state-space model apart from sparsity considerations). We furthermore show in Lemma 1, in another application of the FIM analysis, that identifiability of the hyperparameters (state space model parameters) requires smoothing (filtering is not sufficient). Although (regardless of sparsity) KF with joint parameter estimation has been the subject of many approaches over decades, this smoothing requirement has never been pointed out or certainly not been analyzed before.

2. DYNAMIC SBL SYSTEM MODEL
1 Sparse signal $x_t$ is modeled using an AR(1) process with a diagonal correlation coefficient matrix $F$, which can be written as follows, State Update: $x_t = F x_{t-1} + \omega_t$, Observation: $y_t = A^{(t)} x_t + v_t$, (1) where $x_t = [x_{1,t}, ..., x_{M,t}]^T$. Diagonal matrices $F$ and $\Gamma$ are defined with its elements, $F_{i,i} = f_i, f_i \in (-1, 1)$ and $\Gamma_{i,i} = \sigma_i^2$. Diagonal matrices $\Theta$ and $\Gamma$ are defined with its elements, $\Theta_{i,i} = \sigma_i^2, \Gamma_{i,i} = \sigma_i^2$.

Notations: The operator $(\cdot)^H$ represents the conjugate transpose or conjugate for a matrix or a scalar respectively. The operators tr$(\cdot)$ represents trace if a matrix.
In the following, the pdf of a complex Gaussian random variable $x$ with mean $\mu$ and variance $\sigma^2$ is given by $CN(x; \mu, \sigma^2)$. $K L(q||p)$ represents the Kullback-Leibler distance between the two distributions $q, p$. $A_{n,\cdot}$ or $A_{\cdot,\cdot}$ represents the $n^{th}$ row or $n^{th}$ column of $A$ respectively. $blkdiag(\cdot)$ represents blockdiagonal part of a matrix. diag$(X)$ or diag$(x)$ represents a vector obtained by the diagonal elements of the matrix $X$ or the diagonal matrix obtained with the elements of $x$ in the diagonal respectively.
\[ \Gamma = \text{diag}(\alpha), \quad \alpha = [\alpha_1, \ldots, \alpha_M]. \] Here \( \alpha_i \) represents the inverse variance of \( x_i \sim \mathcal{N}(0, \frac{1}{\alpha_i}) \). Further, \( \nu_t \sim \mathcal{N}(0, \Lambda^{-1}) \), where \( \Lambda^{-1} = \Gamma (1 - F^H F) \) is \( \mathcal{N}(\frac{1}{\alpha_1}, \ldots, \frac{1}{\alpha_M}) \) and \( \nu_t \sim \mathcal{N}(0, \frac{\gamma}{I}). \) \( \nu_t \) and \( \nu_i \) are the complex Gaussian mutually uncorrelated state innovation sequences. Hence we sparsify the prediction error variance \( \nu_t \) also, with the same support as \( \nu_0 \) and henceforth enforces the same support set for \( x_i \), \forall \ i. \nu_t \) is independent of the \( \nu_0 \) process. Although the above signal model seems simple, there are numerous applications such as 1) Bayesian adaptive filtering [8], 2) Wireless channel estimation: multipath parameter estimation as in [9]. In this case, \( x_t \) is FIR filter response, and \( \Gamma \) represents e.g. the power delay profile.

In Bayesian compressive sensing, a two-layer hierarchical prior is assumed for the \( x \) as in [1]. The hierarchical prior is chosen such that it encourages the sparsity property of \( x \) or of the innovation sequences \( \nu_t \). The state update gets represented as,

\[
p(x_t|x_{t-1}, F, \Gamma) = \prod_{i=1}^{M} \mathcal{N}(f(x_t|x_{t-1}), \frac{1}{\alpha_i}).
\]

For the convenience of analysis, we reparameterize \( \alpha_i \) in terms of \( \lambda_i \) and assume a Gamma prior for \( \Lambda \), \( p(\Lambda) = \prod_{i=1}^{M} p(\lambda_i/a, b) = \prod_{i=1}^{M} \Gamma^{-1}(a)\theta_i^{-1}e^{-b\theta_i}. \)

The inverse of noise variance \( \gamma \) is also assumed to have a Gamma prior, \( p(\gamma/c, d) = \Gamma^{-1}(c)\theta_i^{-1}e^{-d\theta_i} \), such that the marginal pdf of \( x_t \) (student-t distribution) becomes more sparsity inducing than e.g., a Laplacian prior. The advantage is that the whole machinery of linear MMSE estimation can be exploited, such as e.g., the Kalman filter. This is but embedded in other layers making things eventually non-Gaussian. Now the likelihood distribution can be written as,

\[
p(y_t|x_t, \gamma) = (2\pi)^{-N_A} e^{-\frac{1}{2\gamma}y_t^H A(\theta_x) y_t}.\]

To make this priors non-informative (Jeffrey’s prior), we choose them to be small values \( a = c = b = d = 10^{-5} \). For the AR(1) coefficients \( f_k \), we don’t assume any prior distribution. We define the unknown parameter vector \( \theta = \{x, \Lambda, \gamma, F\} \) and \( \theta_t \) represents each scalar in \( \theta \).

3. COMBINED BP/MF APPROACH

The fixed points of the standard BP algorithm are shown to be the stationary points of the Bethe free energy (BFE) [5]. However, for the MF approximation in variational Bayes, the approximate posteriors are shown to be converging to a local minimum of the MF free energy, which is an approximation of the BFE. Moreover, we observe in [10,11] that for estimation of the signals from interference corrupted observations, MF is a poor choice since it doesn’t give the accurate posterior variance (posterior variance of \( x \) is observed to be independent of the error variances of other \( x, l \neq i \)). Assume that the posterior be represented as,

\[
p(\theta) = \frac{1}{Z} \prod_{a \in A_B} f_a(\theta_a) \prod_{b \in A_M} f_b(\theta_b),
\]

where \( A_B, A_M \) represent the set of nodes belonging to the BP part and MF part respectively with \( A_B \cap A_M = \emptyset \). \( Z \) is the normalization constant. Throughout the paper, the vector \( \theta \) represents a subset of \( \theta \) and \( \theta_t \) represents a scalar parameter in \( \theta \). \( N(\nu), N(a) \) represent the number of neighbouring nodes of any variable node \( i \) or factor node \( a \). \( N_{BP}(\nu) \) represents the number of neighbouring nodes of \( i \) which belong to the BP part, similarly \( N_M(\nu) \) is defined. Also, we define \( I_{BP} = \bigcup_{\nu \in N_{BP}(\nu)} N(\nu) \), \( I_{BP} = \bigcup_{\nu \in N_{BP}(\nu)} N(\nu) \).

The resulting free energy obtained by the combination of BP and MF are written as below (Note that we use an abuse of notation and let \( q_i(\theta_t) \) represents the belief about \( \theta_t \) (the approximate posterior),

\[
F_{BP,MF} = \sum_{a \in A_B} \sum_{\theta_a} q_a(\theta_a) \ln \frac{q_a(\theta_a)}{f_a(\theta_a)} + \sum_{a \in A_M} \sum_{\theta_a} \prod_{i \in \nu \in N(a)} q_i(\theta_t) \ln f_a(\theta_a) - \sum_{i \in \nu \in N(\nu)} \ln(N_{BP}(i)) - 1 \sum_{i \in \nu \in N(\nu)} q_i(\theta_t) \ln q_i(\theta_t).
\]
where \( \sigma_t^{-2} = \lambda_t + d_t^{-1} \), \( \tilde{x}_t = \frac{r_t x_t}{1+d_t \sigma_t^{-2}} \). (6)

One remark here is that compared to our previous work using VB [10], combining BP and MF gives a more accurate approximation of the error variance as shown in (6), where \( \sigma_t^2 \) incorporates the effect of all \( \sigma_t^2, t \neq t \). Since the factor node \( f_n \in \mathbb{A}_{BP} \), the message \( n_{x_t} \rightarrow f_n \) from variable node \( x_t \) to \( f_n \) is updated by the BP rule as follows,

\[
n_{x_t} \rightarrow f_n \left( x_t \right) = \frac{m_{y_t} \circ g_{x_t}}{m_{y_t} \circ g_{x_t}} \propto CN \left( x_t; \tilde{n}_t, r_t \right), \quad \tilde{n}_t = r_t \left( \frac{\tilde{\sigma}_t^2}{\tilde{\sigma}_t^2} \right) = \frac{\tilde{\sigma}_t^2}{\tilde{\sigma}_t^2} + \frac{\tilde{n}_t}{\tilde{n}_t}. \quad (7)
\]

3.2. Dynamic BP-MF-EF based SBL

The joint distribution of all the observations and parameters can be written as, \( p(y_t, \theta_t | y_{t-1}, \theta_{t-1}) = p(y_t | \theta_t)p(\theta_t | y_{t-1}, \theta_{t-1}) \), and \( p(\theta_t | y_{t-1}, \theta_{t-1}) \) denotes the predictive distribution. Similar as in KF, first we compute the posterior distribution of \( \theta_t \) given the observations till \( (t-1) \), which is called as the prediction stage. Since the correlation coefficient matrix \( F \) is diagonal, all the \( x_t, t \) are decoupled in the state update model and we exploit this fact to predict the states and the hyperparameters in the state update model using MF.

3.2.1. Diagonal AR(1) (DAR(1)) Prediction Stage

Assuming that the belief \( \gamma_t(\cdot) \) at time \( t \), of noise precision \( \gamma_t \) is known, the message \( m_{y_t} \rightarrow n_{x_t} \) from the factor node \( f_n \) to \( n_{x_t} \) is calculated using the MF rule \( m_{y_t} \rightarrow n_{x_t} \propto CN \left( x_t; \tilde{n}_t, r_t \right) \).

\[
m_{y_t} \rightarrow n_{x_t} = \frac{m_{y_t} \circ g_{x_t}}{m_{y_t} \circ g_{x_t}} \propto CN \left( x_t; \tilde{n}_t, r_t \right), \quad \tilde{n}_t = r_t \left( \frac{\tilde{\sigma}_t^2}{\tilde{\sigma}_t^2} \right) = \frac{\tilde{\sigma}_t^2}{\tilde{\sigma}_t^2} + \frac{\tilde{n}_t}{\tilde{n}_t}. \quad (8)
\]

3.2.2. Measurement Update Stage

In the measurement update stage, the posterior for \( x_t \) is inferred using BP as in Section 3.1 and we represent the messages by \( \tilde{x}_t, \sigma_t^2 \) and the beliefs by \( \tilde{x}_t, \sigma_t^2 \). In the measurement stage, the prior for \( x_{k,t} \) gets replaced by the belief from the prediction stage and thus the term \( r_t \) need to be rewritten as, \( r_{t} = d_{t} \left( \sum_{n=1}^{N} \tilde{z}_{n,t} \right) \).

3.2.3. Lag-1 Smoothing Stage

We show in Lemma 1 that KF is not enough to adapt the hyper-parameters, instead we need at least a lag 1 smoothing (i.e. the computation of \( \tilde{x}_{k,t}, \sigma_{k,t}^2 \) through BP). All the hyperparameters \( \lambda_t, f_t, r_t, \gamma_t \) belong to \( \mathbb{A}_{MF} \). Note that the notations \( \tilde{f}_t, \tilde{r}_t, \tilde{\sigma}_t, \tilde{\gamma}_t \) refers to the mean of the posterior (which is equal to the LMMSE point estimates) for the respective hyperparameters at time \( t \) and \( \tilde{\sigma}_t^2 \) represents the posterior variance of \( f_t \) at time \( t \). For the smoothing stage, we use BP with Gaussian Markov Random Fields (GMRF) based factorization. GMRF refers to the representation of BP [14], when the underlying Gaussian distribution is expressed in terms of pairwise connections between scalar variables \( x_{t} \). Substituting the state update equation into the observation model (1), we obtain the system model for the smoothing stage as follows,

\[
y_t = A^{(t)} F x_{t-1} + \tilde{v}_t, \quad \tilde{v}_t = A^{(t)} w_t + v_t, \quad (9)
\]

where \( \tilde{v}_t \sim CN \left( 0, \tilde{R}_t \right) \) with \( \tilde{R}_t = A^{(t)} \Sigma^{-1} A^{(t)^T} + \frac{1}{\lambda_t} I \). The joint distribution can be factorized as, \( p(y_t, \theta_t | y_{t-1}, \theta_{t-1}) = p(y_t | \theta_t)p(\theta_t | y_{t-1}, \theta_{t-1}) \).

In the case of KF, the following update rule is applied to update the hyper parameters:

\[
\tilde{f}_{t} = \arg \min_{l \in \mathbb{A}_{MF}} \{ l | \tilde{f}_{t} | \tilde{r}_{t} | \tilde{\gamma}_{t} \} \quad \text{and} \quad \tilde{r}_{t} = \arg \min_{l \in \mathbb{A}_{MF}} \{ l | \tilde{r}_{t} | \tilde{\gamma}_{t} \} \quad \text{where} \quad \tilde{v}_{t} = A^{(t)} u_{t} + v_{t}.
\]

4. OPTIMAL PARTITIONING OF BP AND MF NODES

In this section, we show that the partitioning of BP and MF nodes can be characterized through the computation of FIM = \( E \left[ \ln p(y_t | \theta_t) \right] \). For our analysis, we will allude briefly to an extended concept of Cramer-Rao bound (CRB), the mismatched CRB (mCRB) [15] of VB (mCRB_{BP}), which is a version of the CRB under model misspecification, and corresponds to the Laplace approximation covariance. Let CRB corresponds to the proper Bayesian CRB and mCRB_{BP} refers to the mCRB for the BP.
Theorem 1 If the parameter partitioning in VB is such that the different parameter blocks are decoupled at the level of Fisher Information Matrix, then VB is not suboptimal in terms of (mismatched) Cramer-Rao Bound. If a finer partitioning granularity is used (such as up to scalar level as in MF), then VB becomes quasi suboptimal, which can be alleviated by using BP instead.

\[ mCRB_{BP} = \text{blkdiag}(CRB) = \text{blkdiag}(FIM^{-1}) \]
\[ mCRB_{VB} = \left( \text{blkdiag}(FIM) \right)^{-1} \]

So, \( mCRB_{BP} = mCRB_{VB} \) if \( FIM = \text{blkdiag}(FIM) \).

Proof: We briefly outline the proof here. Laplace approximation refers to the evaluation of marginal likelihood or free energy using Laplace’s method [16]. This is equivalent to a Gaussian approximation of the posterior \( q(\theta|y) \) around a maximum a posteriori (MAP) estimate \( \theta_0^{(0)} \), motivated by the fact that in the asymptotic limit (large amount of data or high SNR), the posterior approaches a Gaussian around the MAP point. Under the Laplace approximation, the belief becomes \( q(\theta) = \mathcal{N}(\theta_0^{(0)}, \Sigma_0^{(0)}) \). Further we evaluate the free energy [14] \( F \), which is the free energy and \( L = \ln p(y, \theta) \).

\[ F = L(\theta_0^{(0)}) + \frac{1}{2} \sum_{i=1}^{M} (G_i + \ln \det \Sigma_0^{(0)} + k_i \ln(2\pi e)) \]
\[ \ln q_i(\theta_i) = L(\theta_i, \theta_{-i}) + \frac{1}{2} \sum_{j=1, j \neq i}^{M} G_{ij} = \text{tr} \left[ \Sigma_0 \left( \frac{\partial L(\theta_0^{(0)})}{\partial \theta_i} \right)^H \right] \]

Here \( k_i \) refers to the number of scalars in \( \theta_i \) and \( \ln(2\pi e) \) is the entropy of a Gaussian random variable. Now by differentiating \( \ln q_i(\theta_i) \) w.r.t. the posterior covariance, we obtain the approximate covariances as,

\[ \Sigma_0^{(0)} = \left( \frac{\partial L(\theta_0^{(0)})}{\partial \theta_i} \right)^H, \]

The posterior covariance in (14) is computed by the Hessian at the variational mode or maximum a posteriori (MAP) point. This variational mode can be obtained as \( \theta_0^{(0)} = \max q_\theta \ln q(\theta) \).

In the Laplace approximation, all pdfs are Gaussian with CRB (portions) as covariances and LMMSE estimates as means. So in the too fine partitioning case, the VB partitioning is applied to the FIM, taking a too fine blockdiagonal part, and since that partitioning is finer than the blockdiagonal FIM structure, then the inverse of the too fine blockdiagonal part of the FIM does not give the correct CRB. So \( mCRB_{VB} \neq CRB \). The nodes in the factor graph are decided based on the partitioning of the blocks in the FIM block diagonal structure, such that the \( mCRB_{VB} = CRB \). Here ends the proof.

4.2. Optimal Partitioning for DAR-SBL:

In this section we formulate the optimal partitioning between VB and BP for the dynamic SBL case. Here we need to consider the FIMs recursively, i.e. FIM of the time update stage followed by the measurement stage. For the time update stage, we abbreviate \( \hat{p}(x_{k,t}, x_{k,t-1}, f_k, \lambda_k|y_{t-1}) = p \) for convenience here.

\[ \ln p = \ln \lambda_k - \lambda_k |x_{k,t} - f_k x_{k,t-1}|^2 - \sigma_{k,t-1}^{-1} |x_{k,t-1} - \bar{x}_{k,t-1}|^2 + \sum_{i=1}^{3M} \ln q_{\lambda_i}(\lambda_i) \]

The measurement FIM (15) is the prior FIM for the next time update. Thus it follows that BP is needed for the inference of \( \lambda_t \) and MF for \( \gamma \). One remark here is that the prior \( x_t \) covariance for the measurement update is the inverse FIM of the time update and is diagonal here.

Lemma 1 The \( AR(1) \) model parameters require (at least lag 1) smoothing for identifiability.

Proof: Considering, with augmented state \( \theta_t = [x_t; f_t; \text{diag}(\Delta); \gamma] \) \((3M + 1 \text{ dimension})\), we obtain the FIM, \( J_{\theta}, \theta = \text{blkdiag}(J_{x,t}, J_{\gamma,t}, J_{\alpha,t}, J_{\lambda,t}) \). In [17], Tichavský et al. derived an elegant recursive approach to calculate the FIM recursions for a general discrete-time nonlinear filtering problem. Based on a similar derivation, we arrive at the following recursions for the sequence \( J_{\theta,t} \) of posterior information submatrices for estimating \( \theta_t \).

\[
\begin{align*}
J_{\lambda,t} &= \Lambda + \gamma A^{(1)} H, \\
J_{x,t} &= J_{x,t-1} + A F A^H + J_{x,t-1}^{-1}, \\
J_{\alpha,t} &= J_{\alpha,t-1} - D^T D + J_{\lambda,t}^{-1}, \\
J_{\gamma,t} &= \Lambda - D D^T + J_{\lambda,t}^{-1} D \end{align*}
\]

Note that if \( J_{x,F,t-1} = 0 \), then \( J_{x,F,t} = 0, \forall t \geq 0 \). FIM recursions show that filtering may be enough for the estimation of \( AR(1) \) parameters. However, closely looking at the expressions for \( f_{x|t} \) derived in our work [12, eq. (24-25)] shows that \( f_{x|1} = f_k \). This implies that we need to know the true \( f_k \) to estimate it, in the joint estimation framework. Further to prove the unidentifiability, we use the concept of global identifiability provided in [18].

\[ p(f/x_t, y_t) = p(y_t/x_t) p(x_t/f) p(f)/p(y_t, x_t) \]

This above expression (17) suggests that posterior of \( f \) given \( x_t \) does not depend on \( y_t \) or in other words the observations doesn’t provide any extra information about \( f \) other than the prior \( p(f/x_t) \) and hence \( f \) is globally not identifiable. This proves the Lemma. (17) also shows that \( f, x_t \) are coupled in the estimation unlike the decoupling property shown by the FIM analysis.

Few remarks follows: Th \( mCRB \) analysis in Theorem 1 indicates that the \( x \) part needs to be treated jointly, motivating joint VB or BP. We conjecture that whatever local identifiability analysis indicates as necessitating joint treatment for optimality requires indeed joint treatment. But local analysis may not capture all dependencies. The local analysis (recursive CRB) shows that filtering would be sufficient for local identifiability of \( f \) and that the \( f_k \) and the \( x_t \) are decoupled. However, global identifiability analysis reveals that filtering is not enough for identifiability of \( f \) and that the estimation of \( x_t \) and \( f_t \) is coupled. The gap between local and global analysis may perhaps be reflected in the observation that the hyperparameters could be estimated (in what corresponds to filtering) by Type-II Maximum Likelihood (ML) [19] (ie ML for hyperparameters, with the random parameters \( x \) integrated out). Such Type-II ML approach for hyperparameter estimation in the dynamic problem considered here will be investigated further in future work.

Corollary 1.1 For the smoothing stage (9), an optimal partitioning is to apply BP for estimation of the sparse vector, \( \hat{x}_{t-1} \) and MF for the correlation coefficient \( F \).
We presented a fast SBL algorithm called BP-MF-EP DAR-SBL, which uses a combination of BP, MF and EP techniques to approximate the posteriors of the data and parameters and track a time varying sparse signal. BP-MF-EP DAR-SBL helps to circumvent the matrix inversion operation required in Bayesian SAGE algorithm. We propose for the first time in the literature an optimal way to select the partitioning of BP and MF nodes with CRB as a performance evaluation criteria. Future work include extension of the combined BP-MF framework for Kronecker structured dictionary learning [20].

## 7. REFERENCES


