Network-Wide Distributed Carrier Frequency Offsets Estimation and Compensation via Belief Propagation

Jian Du and Yik-Chung Wu

Abstract

In this paper, we propose a fully distributed algorithm for frequency offsets estimation in decentralized systems. The idea is based on belief propagation, resulting in that each node estimates its own frequency offsets by local computations and limited exchange of information with its direct neighbors. Such algorithm does not require any centralized information processing or knowledge of global network topology, thus is scalable with network size. It is shown analytically that the proposed algorithm always converges to the optimal estimates regardless of network topology. Simulation results demonstrate the fast convergence of the algorithm and show that estimation mean-squared-error at each node approaches the centralized Cramér-Rao bound within a few iterations of message exchange.

Index Terms

Carrier frequency offsets (CFOs) estimation, heterogeneous networks, factor graph, convergence analyses.

The authors are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong (e-mail: {dujian, ycwu}@eee.hku.hk).

I. INTRODUCTION

In wireless communication systems, local oscillators are used in transceivers to generate carrier signals required for up-conversion and down-conversion. Ideally, carrier frequencies produced by oscillators of each transceiver pair should be the same. However, in practice, frequencies synthesized from independent oscillators will be different from each other due to variation of oscillator circuits. The received signal impaired by carrier frequency offsets (CFOs) between transmitter and receiver leads to a continuous rotation of symbol constellation, resulting in degradation of system capacity and bit error rate (BER) [1]–[4]. Consequently, carrier frequency synchronization has always been a momentous issue in communication systems.

As modern wireless environments become more heterogeneous and decentralized, mobile terminals in a network engage more and more in cooperative communications and distributed computations [5]. New scenarios require multiple wireless units to synchronize with each other arise. For example,

Distributed beamforming: As shown in Fig. 1(a), to improve the range of communications and save battery power during the transmission, multiple mobile terminals form a virtual antenna array and cooperatively direct a beam in the desired direction of transmission [6], [7]. Since each source node in the distributed beamformer has an independent local oscillator, common carrier frequency among all transmitters is crucial to ensure that a beam is aimed in the desired direction.

Multi-cell cooperation: In fully frequency reuse cellular systems as depicted in Fig. 1(b), despite different users interfere with each other, multiple base stations could coordinate their coding and decoding. It was shown that such joint-processing significantly outperforms a network with individual cell processing [8], [9]. Yet, multiple base-stations cooperation requires frequency synchronization so that there is no CFO between each pair of communication link [10].

Heterogenous Networks (HetNets): HetNets have attracted much attention from both industry and academia in the past few years. As shown in Fig. 1(c), in a 3-tier HetNet, a mobile may wish to be associated with different tier base stations in the uplink and downlink to obtain optimal performance [11]–[14]. However, multi-tier cooperation is possible only when different tiers of networks are frequency-synchronized to each other.

The above examples of network-wide synchronization problem can be summarized and reduced to a multi-node communication systems as shown in Fig. 2. Despite the fact that relative CFO

between each pair of nodes can be optimally estimated by existing methods [15]–[23], network-wide CFOs correction is difficult since each node needs to synchronize with multiple neighboring nodes with different relative CFOs at the same time. Making the problem more challenging is the fact that synchronization should be accomplished by local operations without knowing the global network structure since users move around and join different parts of the network randomly.

Pioneering works for network CFOs correction have been proposed in [24], [25]. By gathering all the information in a central processing unit, CFOs are estimated at the receiver and then fedback to corresponding transmitters to adjust the offsets. These methods are centralized, which are not suitable for large-scale network. On the other hand, [26]–[28] investigated methods for frequency synchronization in distributed beamforming systems. However, these methods require the formation and maintenance of special network structures (e.g., tree structure in [26], ring structure in [27] and chain structure in [28]), thus suffer from large overhead and long delay, and are not scalable with network size. Recently, consensus-type algorithms have been reported for network-wide synchronization in [29]–[31]. Distributed frequency-locked loops (D-FLL) algorithm [29] and pulse-coupled oscillator algorithm [30] have been proposed to control and synchronize the carrier frequencies of autonomous nodes based on average consensus principle. Notwithstanding the distributed carrier frequency calibration advantage, they suffer from slow convergence rate. Furthermore, the algorithms in [29], [30] are designed exclusively for single path channel. Even in a simple point-to-point case with multi-path channel, these approaches cannot be applied directly.

In this paper, we propose a network-wide fully distributed CFO estimation and compensation method which only involves local processing and information exchange between direct neighbors. The frequency offset of each oscillator is estimated and corrected locally in each node. After synchronization, the mean-square-error (MSE) for each frequency offset approaches the corresponding Cramér-Rao bound (CRB) asymptotically. The proposed algorithm is scalable with network size, and robust to topology changes. The convergence of the proposed method is also formally proved.

The rest of this paper is organized as follows. System model is presented in Section II. Fully distributed frequency offsets estimation and correction based on belief propagation (BP) is derived in Section III . The convergence property of the proposed method is analyzed in Section IV. Simulation results are given in Section V and, finally, conclusions are provided in Section

VI.

Notations: Boldface uppercase and lowercase letters will be used for matrices and vectors, respectively. $\mathbb{E}$ denotes the expectation over the random variables. Superscripts H and T denote Hermitian and transpose, respectively. The symbol $\boldsymbol{I}_N$ represents the $N \times N$ identity matrix, while $\boldsymbol{1}_K$ is an all one K dimensional vector. The symbol $\otimes$ denotes the Kronecker product and $\odot$ denotes the Hadamard product. Notation $\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{R})$ stands for the probability density function (pdf) of a Gaussian random vector $\boldsymbol{x}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{R}$. The symbol $\propto$ represents the linear scalar relationship between two real valued functions. $\mathrm{diag}\{[a_1, \ldots, a_N]\}$ corresponds to an $N \times N$ diagonal matrix with diagonal components a_1 through a_N, while $\mathrm{blkdiag}\{[\boldsymbol{A}_1, \ldots, \boldsymbol{A}_N]\}$ corresponds to a block diagonal matrix with $\boldsymbol{A}_1$ through $\boldsymbol{A}_N$ as diagonal blocks. For two matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$, $\boldsymbol{X} \succeq \boldsymbol{Y}$ means that $\boldsymbol{X} - \boldsymbol{Y}$ is a positive semi-definite matrix.

II. System Model

We consider a network consisting of K nodes distributed in a field as shown in Fig. 2. The topology of the network is described by a communication graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ of order K, where $\mathcal{V} = \{1, \ldots, K\}$ is the set of graph vertexes, and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of graph edges. In the example shown in Fig. 2, the vertices are depicted by circles and the edges by lines connecting these circles. The neighborhood of node i is the set of nodes $\mathcal{I}(i) \subset \mathcal{V}$ defined as $\mathcal{I}(i) \triangleq \{j \in \mathcal{V} | \{i, j\} \in \mathcal{E}\}$, i.e., those nodes that are connected via a direct communication link to node i. It is also assumed that any two distinct nodes can communicate with each other through finite hops, such graph is named strongly connected graph.

In general, relative CFOs exist between any pair of neighboring nodes, and can be estimated by traditional CFOs estimation methods. Let nodes i and j equipped with N_i and N_j antennas, respectively. Denote the frequency offsets [1] (with respect to a reference frequency) of the qth antenna on node i as ω_q^i, while that of kth antenna of node j as ω_k^j. Then, the relative CFO between the qth and k^{th} antenna of node i and j is $\epsilon_{q,k}^{i,j} \triangleq \omega_q^i - \omega_k^j$. Here we consider the general case where each antenna can be associated with separate oscillator circuit. Therefore, for the Multiple Input Multiple Output (MIMO) system between node i and node j, there are

---

[1] The frequency offset in this paper is the normalized CFO, defined as $2\pi\Delta f T_s$, where Δf is the CFO in Hz, while T_s is the sampling period.

$N_i N_j$ relative CFOs denoted as $\boldsymbol{\epsilon}^{i,j} \triangleq [\epsilon_{1,1}^{i,j}, \ldots, \epsilon_{N_i,1}^{i,j}, \ldots, \epsilon_{1,N_j}^{i,j}, \ldots, \epsilon_{N_i,N_j}^{i,j}]^T$. Such relative CFOs estimation in MIMO systems can be decomposed into N_j parallel Multiple Input Single Output (MISO) CFOs estimation problem [18]. For example, considering a flat-fading MISO system, for the kth receive antenna of node j, the received signal can be written as

$$y_k^{i,j}(t) = \sum_{q=1}^{N_i} h_{q,k}^{i,j} e^{j\epsilon_{q,k}^{i,j}t} z_q^i(t) + \xi_k^j(t), \quad t = 1, \ldots, N, \tag{1}$$

where $h_{q,k}^{i,j}$ is the unknown channel gain between the qth antenna of node i and k^{th} antenna of node j; $j \triangleq \sqrt{-1}$; $\{z_q^i(t)\}_{t=1}^N$ is the training sequence transmitted from the qth antennas of node i; and $\xi_k^j(t)$ is the observation noise at the kth antenna of node j. By stacking (1) with $t = 1, \ldots, N$ in vector form and omitting superscript i, j without confusion, the received vector $\boldsymbol{y}_k \triangleq [y_k(1), \ldots, y_k(N)]^T$ can be written as

$$\boldsymbol{y}_k = \boldsymbol{\Gamma}_k(\boldsymbol{\epsilon}_k) \odot \boldsymbol{Z}_k \boldsymbol{h}_k + \boldsymbol{\xi}_k \quad k = 1, \ldots, N_j, \tag{2}$$

where $\boldsymbol{\Gamma}_k(\boldsymbol{\epsilon}_k)$ is an N-by-N_i Vandermonde matrix with its tth row given by $[e^{jt\epsilon_{1,k}}, e^{jt\epsilon_{2,k}}, \cdots, e^{jt\epsilon_{N_i,k}}]$; $\boldsymbol{Z}_k$ is the N-by-N_i training sequence matrix with its tth row $[z_1(t), z_2(t), \cdots, z_{N_i}(t)]$; and $\boldsymbol{\xi}_k = [\xi_k(1), \ldots, \xi_k(N)]^T$ is the observation noise. The parameters $\boldsymbol{\epsilon}_k \triangleq [\epsilon_{1,k}, \epsilon_{2,k}, \ldots, \epsilon_{N_i,k}]^T$ and $\boldsymbol{h}_k \triangleq [h_{1,k}, \ldots, h_{N_i,k}]^T$ are the parameters need to be estimated.

If the noise is white and Gaussian, i.e., $\boldsymbol{\xi}_k \sim \mathcal{CN}(\boldsymbol{\xi}_k; \boldsymbol{0}, \sigma_k^2 \boldsymbol{I}_N)$, joint relative CFOs and channels estimation have been extensively studied in the past two decades and the optimal estimates $\hat{\boldsymbol{\epsilon}}_k$ and $\hat{\boldsymbol{h}}_k$ have been proposed in [17]–[21], with the MSEs approaching the corresponding CRBs in medium and high signal-to-noise ratio (SNR) ranges. From (2), the CRB of $\boldsymbol{\epsilon}_k$ can be shown to be [18]

$$\boldsymbol{B}_{\boldsymbol{\epsilon}_k}(\boldsymbol{\epsilon}_k, \boldsymbol{h}_k) = \frac{\sigma_k^2}{2} \left\{ \text{Re}[\boldsymbol{V}_k - \boldsymbol{T}_k^H (\boldsymbol{\Lambda}_k^H \boldsymbol{\Lambda}_k)^{-1} \boldsymbol{T}_k] \right\}^{-1}, \tag{3}$$

where $\boldsymbol{V}_k \triangleq \text{diag}\{\boldsymbol{h}_k\} \boldsymbol{\Lambda}_k^H \boldsymbol{D}^2 \boldsymbol{\Lambda}_k \text{diag}\{\boldsymbol{h}_k\}$, $\boldsymbol{T}_k \triangleq \boldsymbol{\Lambda}_k^H \boldsymbol{D} \boldsymbol{\Lambda}_k \text{diag}\{\boldsymbol{h}_k\}$, with $\boldsymbol{\Lambda}_k \triangleq \boldsymbol{\Gamma}_k(\boldsymbol{\epsilon}_k) \odot \boldsymbol{Z}_k$ and $\boldsymbol{D} \triangleq \text{diag}\{[1, 2, \ldots, N]\}$. Since there are N_j independent MISO estimation problems as in (2), the CRB for frequency estimation in MIMO system between node i and j is given by $\boldsymbol{B}_{\boldsymbol{\epsilon}}^{\{i,j\}}(\{\boldsymbol{\epsilon}_k\}_{k=1}^{N_j}, \{\boldsymbol{h}_k\}_{k=1}^{N_j}) = \text{blkdiag}\{[\boldsymbol{B}_{\boldsymbol{\epsilon}_1}(\boldsymbol{\epsilon}_1, \boldsymbol{h}_1), \ldots, \boldsymbol{B}_{\boldsymbol{\epsilon}_{N_j}}(\boldsymbol{\epsilon}_{N_j}, \boldsymbol{h}_{N_j})]\}$.

After joint estimation of relative CFOs and channels, the relative CFOs between node i and j can be obtained as

$$\boldsymbol{r}_{i,j} = \boldsymbol{A}_{i,j} \boldsymbol{\omega}_i + \boldsymbol{A}_{j,i} \boldsymbol{\omega}_j + \boldsymbol{n}_{i,j}, \tag{4}$$

where $\boldsymbol{r}_{i,j} \triangleq [\hat{\boldsymbol{\epsilon}}_1^T, \hat{\boldsymbol{\epsilon}}_2^T, \ldots, \hat{\boldsymbol{\epsilon}}_{N_j}^T]^T$ are the $N_i N_j$ relative CFOs estimates; $\boldsymbol{A}_{i,j} \triangleq \boldsymbol{I}_{N_i} \otimes \boldsymbol{1}_{N_j}$ and $\boldsymbol{A}_{j,i} \triangleq -\boldsymbol{1}_{N_i} \otimes \boldsymbol{I}_{N_j}$; and $\boldsymbol{n}_{i,j}$ is the estimation error. It is known that for the maximum likelihood (ML) estimates, $\boldsymbol{r}_{i,j}$ is asymptotically Gaussian distributed with mean $[\boldsymbol{\epsilon}_1^T, \boldsymbol{\epsilon}_2^T, \ldots, \boldsymbol{\epsilon}_{N_j}^T]^T = \boldsymbol{A}_{i,j}\boldsymbol{\omega}_i + \boldsymbol{A}_{j,i}\boldsymbol{\omega}_j$ and covariance matrix $\boldsymbol{B}_{\boldsymbol{\epsilon}}^{\{i,j\}}(\{\boldsymbol{\epsilon}_k\}_{k=1}^{N_j}, \{\boldsymbol{h}_k\}_{k=1}^{N_j})$ [32]. That is, $\boldsymbol{r}_{i,j} \sim \mathcal{N}(\boldsymbol{r}_{i,j}; \boldsymbol{\epsilon}_{i,j}, \boldsymbol{B}_{\boldsymbol{\epsilon}}^{\{i,j\}}(\{\boldsymbol{\epsilon}_k\}_{k=1}^{N_j}, \{\boldsymbol{h}_k\}_{k=1}^{N_j}))$. Notice that the CRB depends on the true value of $\{\boldsymbol{\epsilon}_k\}_{k=1}^{N_j}$ and $\{\boldsymbol{h}_k\}_{k=1}^{N_j}$, but since we have obtained the ML estimate $\{\hat{\boldsymbol{\epsilon}}\}_{k=1}^{N_j}$ and $\{\hat{\boldsymbol{h}}_k\}_{k=1}^{N_j}$, $\boldsymbol{B}_{\boldsymbol{\epsilon}}^{\{i,j\}}(\{\boldsymbol{\epsilon}_k\}_{k=1}^{N_j}, \{\boldsymbol{h}_k\}_{k=1}^{N_j})$ can be closely approximated by $\boldsymbol{R}_{i,j} = \boldsymbol{B}_{\boldsymbol{\epsilon}}^{\{i,j\}}(\{\hat{\boldsymbol{\epsilon}}_k\}_{k=1}^{N_j}, \{\hat{\boldsymbol{h}}_k\}_{k=1}^{N_j})$.

Notice that traditional CFO estimation for point-to-point link only estimates $N_i N_j$ relative CFOs given by $\boldsymbol{r}_{i,j}$ in (4). However, in order to compensate the offset of individual oscillator, we need to estimate $N_i + N_j$ absolute CFOs in $\boldsymbol{\omega}_i$ and $\boldsymbol{\omega}_j$. For simple MIMO systems, [25] provides a method to resolve $N_i + N_j$ absolute CFOs from $N_i N_j$ relative CFOs. In this paper, we take a significant step further to resolve all absolute CFOs in a distributed network. That is, to estimate and compensate $\boldsymbol{\omega}_i$ in each node based on estimation results of local relative CFOs $\boldsymbol{r}_{i,j}$.

Remark 1: The system model (2) can be extended to the cases where signals undergoing frequency selective fading channel and even doubly selective channel. Effective estimators have been extensively studied and MSE performance of these estimators were shown to approach the corresponding CRBs [33]–[35]. Thus, we can always establish the relative CFOs relationship as in (4).

Remark 2: After relative CFOs estimation, each receiver (node j in the example) obtains the estimate $\boldsymbol{r}_{i,j}$ as well as the covariance matrix $\boldsymbol{R}_{i,j}$. By feeding back this information to the corresponding transmitter, node i also obtains the relative CFOs estimates and estimation error covariance.

III. DISTRIBUTED CFOS ESTIMATION

A. *Distributed CFOs Estimation via Belief Propagation*

The optimal CFO estimator at each node is the ML estimator, which finds the maximum points of the global likelihood function:

$$[(\hat{\boldsymbol{\omega}}_2^{\mathrm{ML}})^T, \ldots, (\hat{\boldsymbol{\omega}}_K^{\mathrm{ML}})^T]^T = \arg \max_{\boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_K} p(\{\boldsymbol{r}_{i,j}\}_{\{i,j\}\in\mathcal{E}} | \boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_K). \qquad (5)$$

Here, without loss of generality, node 1 is assumed to be the reference node, so $\boldsymbol{\omega}_1$ is known. The global likelihood function is given by

$$p\big(\{\boldsymbol{r}_{i,j}\}_{\{i,j\}\in\mathcal{E}}|\boldsymbol{\omega}_1,\boldsymbol{\omega}_2,\ldots,\boldsymbol{\omega}_K\big) \propto \delta(\boldsymbol{\omega}_1) \prod_{\{i,j\}\in\mathcal{E}} p(\boldsymbol{r}_{i,j}|\boldsymbol{\omega}_i,\boldsymbol{\omega}_j), \tag{6}$$

where $p(\boldsymbol{r}_{i,j}|\boldsymbol{\omega}_i,\boldsymbol{\omega}_j) \sim \mathcal{N}(\boldsymbol{r}_{i,j}; \boldsymbol{A}_{i,j}\boldsymbol{\omega}_i + \boldsymbol{A}_{j,i}\boldsymbol{\omega}_j, \boldsymbol{R}_{i,j})$ is the local likelihood function. Notice that since the likelihood function in (6) depends on interactions among all unknown variables, the computation of $\hat{\boldsymbol{\omega}}_i^{\mathrm{ML}}$ in (5) requires gathering of all information in a central processing unit. However, such centralized processing is not favorable in large-scale networks.

In order to compute the optimal estimate (5) in a distributed way, one can exploit the conditional independence structure of the joint distribution (6), which is conveniently revealed by factor graph (FG). FG is an undirected bipartite graphical representation of a joint distribution that unifies direct and undirected graphical models. An example of FG in the context of network-wide synchronization is shown in Fig. 3. In the FG, there are two distinct kinds of nodes. One is variable nodes representing local synchronization parameters $\boldsymbol{\omega}_i$. If there is a communication link between node i and node j, the corresponding variable nodes $\boldsymbol{\omega}_i$ and $\boldsymbol{\omega}_j$ are linked by the other kind of node, factor node $f_{i,j} = p(\boldsymbol{r}_{i,j}|\boldsymbol{\omega}_i,\boldsymbol{\omega}_j)$ representing the local likelihood function [2]. On the other hand, the factor node $f_1 = \delta(\boldsymbol{\omega}_1)$ denotes value of frequency offsets of node 1, and is connected only to the variable node $\boldsymbol{\omega}_1$. Note that the FG is bipartite which means neighbors of a factor node must be variable nodes and vice versa.

From the FG, two kinds of messages are passed around: One is the message from factor node f (likelihood function $f_{i,j}$ or prior distribution f_1) to its neighboring variable node $\boldsymbol{\omega}_i$, defined as the product of the function f with messages received from all neighboring variable nodes except $\boldsymbol{\omega}_i$, and then marginalized for $\boldsymbol{\omega}_i$ [36]

$$m_{f\to i}^{(l)}(\boldsymbol{\omega}_i) = \int \cdots \int f \times \prod_{\boldsymbol{\omega}_j\in\mathcal{B}(f)\backslash\boldsymbol{\omega}_i} m_{j\to f}^{(l-1)}(\boldsymbol{\omega}_j)d\{\boldsymbol{\omega}_j\}_{\boldsymbol{\omega}_j\in\mathcal{B}(f)\backslash\boldsymbol{\omega}_i}, \tag{7}$$

where $\mathcal{B}(f)$ denotes the set of variable nodes that are direct neighbors of the factor nodes f on the FG and $\mathcal{B}(f) \backslash \boldsymbol{\omega}_i$ denotes the same set but with $\boldsymbol{\omega}_i$ removed. In (7), $m_{j\to f}^{(l-1)}(\boldsymbol{\omega}_j)$ is the other kind of message from variable node to factor node which is simply the product of the incoming

---

[2]Note that $f_{i,j} = f_{j,i}$.

messages on other links, i.e.,

$$m_{j \to f}^{(l)}(\boldsymbol{\omega}_j) = \prod_{\tilde{f} \in \mathcal{B}(\boldsymbol{\omega}_j) \backslash f} m_{\tilde{f} \to j}^{(l)}(\boldsymbol{\omega}_i), \tag{8}$$

where $\mathcal{B}(\boldsymbol{\omega}_j)$ denotes the set of factor nodes that are direct neighbors of the variable nodes $\boldsymbol{\omega}_j$ on the FG. It can be seen from (7) and (8) that for both variable nodes and factor nodes, each outgoing message is a function of all incoming messages in the last round except the incoming message from the node where the outgoing message will be directed to. This essential restriction guarantees that for cycle-free FG, incoming and outgoing messages on each edge are independent, and correct marginal distribution is obtained after convergence.

These two kinds of messages are iteratively updated at variable nodes and factor nodes, respectively. In any round of message exchange, a belief of $\boldsymbol{\omega}_i$ can be computed at variable node i as the product of all the incoming messages from neighboring factor nodes, which is given by

$$b^{(l)}(\boldsymbol{\omega}_i) = \prod_{f \in \mathcal{B}(\boldsymbol{\omega}_i)} m_{f \to i}^{(l)}(\boldsymbol{\omega}_i). \tag{9}$$

Thereupon, the estimate of $\boldsymbol{\omega}_i$ in the lth iteration is simply

$$\hat{\boldsymbol{\omega}}_i^{(l)} = \int \boldsymbol{\omega}_i b^{(l)}(\boldsymbol{\omega}_i) d\boldsymbol{\omega}_i. \tag{10}$$

Notice that after convergence, the belief $b^{(l)}(\boldsymbol{\omega}_i)$ at each variable node corresponds to the marginal distribution of that variable exactly when the underlying FG is loop free [36]. However, for the FG with loops, it is generally difficult to know if BP will converge [38]. Even if BP converges to a fixed point, there is no guarantee on the estimation accuracy. Despite the lack of general results on BP, in this paper, the convergence and optimality of BP for network-wide CFO estimation algorithm will be proved in Section IV.

B. Message Computation

In the BP framework, messages are passed and updated iteratively. In order to start the recursion, in the first round of message passing, it is reasonable to set the initial messages from factor nodes to variable nodes $m_{f_{i,j} \to i}^{(0)}(\boldsymbol{\omega}_i)$ as non-informative message $\mathcal{N}(\boldsymbol{\omega}_i; \boldsymbol{v}_{f_{i,j} \to i}^{(0)}, \boldsymbol{C}_{f_{i,j} \to i}^{(0)})$, where $\boldsymbol{v}_{f_{i,j} \to i}^{(0)}$ can be arbitrarily chosen and $[\boldsymbol{C}_{f_{i,j} \to i}^{(0)}]^{-1} = \boldsymbol{0}$. On the other hand, the message from f_1 to $\boldsymbol{\omega}_1$ is always $\delta(\boldsymbol{\omega}_1)$, which can be viewed as a Gaussian distribution with mean $\boldsymbol{\omega}_1$ and covariance $\boldsymbol{0}$. Thereupon, based on the fact that the likelihood function $f_{i,j}$ is also Gaussian,

according to (7), $m_{f_{i,j}\to i}^{(1)}(\boldsymbol{\omega}_i)$ is a Gaussian function. In addition, $m_{j\to f_{i,j}}^{(1)}(\boldsymbol{\omega}_j)$ being the product of Gaussian functions in (8) is also a Gaussian function [40]. Thus during each round of message exchange, all the messages are Gaussian functions and only the mean vectors and covariance matrices need to be exchanged between factor nodes and variable nodes.

At this point, we can compute the messages at any iteration. In general, for the lth ($l = 2, 3, \cdots$) round of message exchange, factor node $f_{i,j}$ receive messages $m_{j\to f_{i,j}}^{(l-1)}(\boldsymbol{\omega}_j)$ from its neighboring variable nodes and then compute messages using (7). After some derivations, it can be obtained that

$$
\begin{aligned}
m_{f_{i,j}\to i}^{(l)}(\boldsymbol{\omega}_i) &= \int p(\boldsymbol{A}_{i,j}, \boldsymbol{A}_{j,i}|\boldsymbol{\omega}_i, \boldsymbol{\omega}_j) m_{j\to f_{i,j}}^{(l-1)}(\boldsymbol{\omega}_j) d\boldsymbol{\omega}_j \\
&\propto \mathcal{N}(\boldsymbol{\omega}_i; \boldsymbol{v}_{f_{i,j}\to i}^{(l)}, \boldsymbol{C}_{f_{i,j}\to i}^{(l)}),
\end{aligned}
\tag{11}
$$

where the inverse of covariance matrix is

$$
\left[\boldsymbol{C}_{f_{i,j}\to i}^{(l)}\right]^{-1} = \boldsymbol{A}_{i,j}^T \left[\boldsymbol{R}_{i,j} + \boldsymbol{A}_{j,i}\boldsymbol{C}_{j\to f_{i,j}}^{(l-1)}\boldsymbol{A}_{j,i}^T\right]^{-1} \boldsymbol{A}_{i,j},
\tag{12}
$$

and the mean vector is

$$
\boldsymbol{v}_{f_{i,j}\to i}^{(l)} = \boldsymbol{C}_{f_{i,j}\to i}^{(l)}\boldsymbol{A}_{i,j}^T \left[\boldsymbol{R}_{i,j} + \boldsymbol{A}_{j,i}\boldsymbol{C}_{j\to f_{i,j}}^{(l-1)}\boldsymbol{A}_{j,i}^T\right]^{-1} (\boldsymbol{r}_{i,j} - \boldsymbol{A}_{j,i}\boldsymbol{v}_{j\to f_{i,j}}^{(l-1)}).
\tag{13}
$$

On the other hand, using (8), the messages passed from variable nodes to factor nodes can be computed as

$$
\begin{aligned}
m_{i\to f_{i,j}}^{(l)}(\boldsymbol{\omega}_i) &= \prod_{f\in\mathcal{B}(\boldsymbol{\omega}_i)\backslash f_{i,j}} m_{f\to i}^{(l)}(\boldsymbol{\omega}_i) \\
&\propto \mathcal{N}(\boldsymbol{\omega}_i; \boldsymbol{v}_{i\to f_{i,j}}^{(l)}, \boldsymbol{C}_{i\to f_{i,j}}^{(l)}),
\end{aligned}
\tag{14}
$$

where

$$
\left[\boldsymbol{C}_{i\to f_{i,j}}^{(l)}\right]^{-1} = \sum_{f\in\mathcal{B}(\boldsymbol{\omega}_i)\backslash f_{i,j}} \left[\boldsymbol{C}_{f\to i}^{(l)}\right]^{-1},
\tag{15}
$$

and

$$
\boldsymbol{v}_{i\to f_{i,j}}^{(l)} = \boldsymbol{C}_{i\to f_{i,j}}^{(l)} \sum_{f\in\mathcal{B}(\boldsymbol{\omega}_i)\backslash f_{i,j}} \left[\boldsymbol{C}_{f\to i}^{(l)}\right]^{-1} \boldsymbol{v}_{f\to i}^{(l)}.
\tag{16}
$$

Furthermore, during each round of message passing, each node can compute the belief for $\boldsymbol{\omega}_i$ using (9), which can be easily shown to be $b_i^{(l)}(\boldsymbol{\omega}_i) \sim \mathcal{N}(\boldsymbol{\omega}_i; \boldsymbol{\mu}_i^{(l)}, \boldsymbol{P}_i^{(l)})$, with the inverse of covariance matrix

$$
\left[\boldsymbol{P}_i^{(l)}\right]^{-1} = \sum_{j\in\mathcal{I}(i)} \left[\boldsymbol{C}_{f_{i,j}\to i}^{(l)}\right]^{-1},
\tag{17}
$$

and mean vector

$$\boldsymbol{\mu}_i^{(l)} = \boldsymbol{P}_i^{(l)} \sum_{j \in \mathcal{I}(i)} \left[\boldsymbol{C}_{f_{i,j} \to i}^{(l)} \right]^{-1} \boldsymbol{v}_{f_{i,j} \to i}^{(l)}. \tag{18}$$

When the algorithm converges or the maximum number of message exchange is reached, each node computes the CFOs according to (10) as

$$\hat{\boldsymbol{\omega}}_i^{(l)} = \int \boldsymbol{\omega}_i b^{(l)}(\boldsymbol{\omega}_i) d\boldsymbol{\omega}_i = \boldsymbol{\mu}_i^{(l)}. \tag{19}$$

The iterative algorithm based on BP is summarized as follows. The algorithm is started by setting the message from factor node to variable node as $m_{f_1 \to 1}^{(0)}(\boldsymbol{\omega}_1) = \delta(\boldsymbol{\omega}_1)$ and $m_{f_{i,j} \to i}^{(0)}(\boldsymbol{\omega}_i) = \mathcal{N}(\boldsymbol{\omega}_i; \boldsymbol{v}_{f_{i,j} \to i}^{(0)}, \boldsymbol{C}_{f_{i,j} \to i}^{(0)})$ with $\boldsymbol{v}_{f_{i,j} \to i}^{(0)} = \boldsymbol{0}$ and $[\boldsymbol{C}_{f_{i,j} \to i}^{(0)}]^{-1} = \boldsymbol{0}$. At each round of message exchange, every variable node computes the output messages to factor nodes according to (15) and (16). After receiving the messages from its neighboring variable nodes, each factor node computes its output messages according to (12) and (13). Such iteration is terminated when (18) converges (e.g., when $\|\boldsymbol{\mu}_i^{(l)} - \boldsymbol{\mu}_i^{(l-1)}\| < \eta$, where η is a threshold) or the maximum number of iteration is reached. Then the estimate of CFOs of each node is obtained as in (19).

Remark 3: In practical networks, there is neither factor nodes nor variable nodes. The two kinds of messages $m_{i \to f_{i,j}}^{(l)}(\boldsymbol{\omega}_i)$ and $m_{f_{i,j} \to j}^{(l)}(\boldsymbol{\omega}_j)$ are computed locally at node i, and only mean vector $\boldsymbol{v}_{f_{i,j} \to j}^{(l)}(\boldsymbol{\omega}_j)$ and covariance matrix $\boldsymbol{C}_{f_{i,j} \to j}^{(l)}(\boldsymbol{\omega}_j)$ are passed from node i to node j during each round of message exchange of BP. It can be seen the algorithm is fully distributed and each node only needs to exchange limited information with neighboring nodes.

Remark 4: Since each pair of node has knowledge of relative CFOs and channel between them, the BP message exchange can be performed as in point-to-point communications.

IV. THEORETICAL ANALYSES OF BP METHOD

It is generally known that if the FG contains cycles, such as the one shown in Fig. 3, messages can flow many times around the graph, leading to the possibility of divergence of BP algorithm [37]. A general sufficient condition for convergence of loopy FGs is given in [38]. Unfortunately, it requires the knowledge of the joint distribution of all unknown variables as shown in (6), and is difficult to verify for large-scale dynamic networks. Recently, [39] proved the convergence of BP in the context of distributed clock offset synchronization in wireless sensor network. The convergence is established for scalar variables in which sub-stochastic and irreducible properties

of BP message recursion were exploited. However, in vector variable case, the BP messages involve matrix inverses (see (12), (13), (15) and (16)), and the sub-stochastic and irreducible properties cannot be easily applied. In the following, we will prove the convergence of BP messages in vector form, and show that the BP based CFO estimates asymptotically converge to the optimal ML solution regardless of network topology.

Theorem 1. The covariance matrix $\boldsymbol{P}_i^{(l)}$ of belief $b_i^{(l)}(\boldsymbol{\omega}_i)$ at each node converges to a positive definite matrix regardless of network topology.

Proof: We begin with a few properties of positive semi-definite (p.s.d.) matrices and positive definite (p.d.) matrices. If $\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{Z}$ are N_i-by-N_i matrices and $\boldsymbol{X} \succ 0$, $\boldsymbol{Y} \succ 0$, $\boldsymbol{Z} \succeq 0$, then we have

Property i): $\boldsymbol{X}^{-1} \succ 0$.

Property ii): $\boldsymbol{X} + \boldsymbol{Y} \succ 0$.

Property iii): $\boldsymbol{X} + \boldsymbol{Z} \succ 0$.

Property iv): $\boldsymbol{X} \succeq \boldsymbol{Y}$ if and only if $\boldsymbol{Y}^{-1} \succeq \boldsymbol{X}^{-1}$ [41].

Property v): $\boldsymbol{A}_{i,j}^T \boldsymbol{X} \boldsymbol{A}_{i,j} \succ 0$ and $\boldsymbol{A}_{i,j}^T \boldsymbol{Z} \boldsymbol{A}_{i,j} \succeq 0$, where $\boldsymbol{A}_{i,j}$ is defined in (4).

Property vi): $\boldsymbol{A}_{j,i} \boldsymbol{X} \boldsymbol{A}_{j,i}^T \succeq 0$.

Properties i) to iv) are standard results in matrix analysis. Property v) is true due to the fact that $\boldsymbol{A}_{i,j}$ is of full column rank. The proof of vi) follows from the definition of $\boldsymbol{X} \succ \boldsymbol{0}$ which is $\boldsymbol{y}^T \boldsymbol{X} \boldsymbol{y} \geq 0$ for any $\boldsymbol{y}$ (including all zeros vector). The result is obtained if we let $\boldsymbol{y} = \boldsymbol{A}_{j,i}^T \boldsymbol{x}$.

Next, we investigate the updating properties of the message covariance matrix. Substituting (15) into (12), the covariance update rules from factor nodes to variable nodes are

$$\left[\boldsymbol{C}_{f_{i,j} \to i}^{(l)}\right]^{-1} = \boldsymbol{A}_{i,j}^T \left[\boldsymbol{R}_{i,j} + \boldsymbol{A}_{j,i} \left[\sum_{f \in \mathcal{B}(\boldsymbol{\omega}_j) \backslash f_{i,j}} \left[\boldsymbol{C}_{f \to j}^{(l-1)}\right]^{-1}\right]^{-1} \boldsymbol{A}_{j,i}^T\right]^{-1} \boldsymbol{A}_{i,j}. \tag{20}$$

From (20), we can deduce two consequences. First, if all message covariance $\boldsymbol{C}_{f \to j}^{(l-1)}$ on the right-hand-side of (20) are non-informative, $\left[\boldsymbol{C}_{f_{i,j} \to i}^{(l)}\right]^{-1}$ cannot be updated, i.e., $\left[\boldsymbol{C}_{f_{i,j} \to i}^{(l)}\right]^{-1} = \boldsymbol{0}$. On the other hand, if some of the $[\boldsymbol{C}_{f \to j}^{(l-1)}]^{-1}$ on the right-hand-side of (20) are p.d. while the remaining are $\boldsymbol{0}$, then $\left[\sum_{f \in \mathcal{B}(\boldsymbol{\omega}_j) \backslash f_{i,j}} [\boldsymbol{C}_{f \to j}^{(l-1)}]^{-1}\right]^{-1} \succ \boldsymbol{0}$ according to property iii). Applying property vi), we have $\boldsymbol{A}_{j,i} \left[\sum_{f \in \mathcal{B}(\boldsymbol{\omega}_j) \backslash f_{i,j}} [\boldsymbol{C}_{f \to j}^{(l-1)}]^{-1}\right]^{-1} \boldsymbol{A}_{j,i}^T \succeq 0$. Furthermore, since $\boldsymbol{R}_{i,j}$ is the relative C-FO estimation covariance, we have $\boldsymbol{R}_{i,j} \succ \boldsymbol{0}$. Thus, $\boldsymbol{R}_{i,j} + \boldsymbol{A}_{j,i} \left[\sum_{f \in \mathcal{B}(\boldsymbol{\omega}_j) \backslash f_{i,j}} [\boldsymbol{C}_{f \to j}^{(l-1)}]^{-1}\right]^{-1} \boldsymbol{A}_{j,i}^T \succ$

$\mathbf{0}$. Then, based on properties i) and v), we obtain $\left[\boldsymbol{C}_{f_{i,j}\rightarrow i}^{(l)}\right]^{-1} \succ \mathbf{0}$. We summarize the above discussion as

$$\left[\boldsymbol{C}_{f_{i,j}\rightarrow i}^{(l)}\right]^{-1} \begin{cases} = \mathbf{0}, & \text{if all } [\boldsymbol{C}_{f\rightarrow j}^{(l-1)}]^{-1} = \mathbf{0}, \\ \succ \mathbf{0}, & \text{if some of } [\boldsymbol{C}_{f\rightarrow j}^{(l-1)}]^{-1} \succ \mathbf{0}, \text{ while others are } \mathbf{0}. \end{cases} \tag{21}$$

Now we prove that for any node i, if there exists a directed path from node $1\rightarrow\ldots\rightarrow j\rightarrow i$ in the network topology, there must be a finite iteration number s such that $\left[\boldsymbol{C}_{f_{i,j}\rightarrow i}^{(s+1)}\right]^{-1} \succeq$ $\left[\boldsymbol{C}_{f_{i,j}\rightarrow i}^{(s)}\right]^{-1} \succ \left[\boldsymbol{C}_{f_{i,j}\rightarrow i}^{(s-1)}\right]^{-1} = \mathbf{0}$. Initially, all $[\boldsymbol{C}_{f_{k,j}\rightarrow j}^{(0)}]^{-1}$ over the FG equal $\mathbf{0}$ except $\boldsymbol{C}_{f_1\rightarrow 1}^{-1} = \infty\boldsymbol{I}$ at the reference node. Hence, the message update starts from the reference node. More explicitly, $\forall i: \{i, 1\} \in \mathcal{E}$, the message covariance is obtained by putting $j = 1$ into (20), which is

$$\left[\boldsymbol{C}_{f_{i,1}\rightarrow i}^{(1)}\right]^{-1} = \boldsymbol{A}_{i,1}^T \boldsymbol{R}_{i,1}^{-1} \boldsymbol{A}_{i,1} \succ \mathbf{0}, \tag{22}$$

where the p.d. property is due to property v). Furthermore, since $\boldsymbol{C}_{f_1\rightarrow 1}^{-1} = \infty\boldsymbol{I}$, it will dominate the sum $\sum_{f\in\mathcal{B}(\boldsymbol{\omega}_j)\backslash f_{i,j}} \left[\boldsymbol{C}_{f\rightarrow j}^{(l-1)}\right]^{-1}$ in (20), and lead to $\left[\boldsymbol{C}_{f_{i,1}\rightarrow i}^{(l)}\right]^{-1} = \boldsymbol{A}_{i,1}^T \boldsymbol{R}_{i,1}^{-1} \boldsymbol{A}_{i,1}$. Thus, we have

$$\ldots = \left[\boldsymbol{C}_{f_{i,1}\rightarrow i}^{(l)}\right]^{-1} = \ldots = \left[\boldsymbol{C}_{f_{i,1}\rightarrow i}^{(2)}\right]^{-1} = \left[\boldsymbol{C}_{f_{i,1}\rightarrow i}^{(1)}\right]^{-1} \succ \left[\boldsymbol{C}_{f_{i,1}\rightarrow i}^{(0)}\right]^{-1} = \mathbf{0}. \tag{23}$$

Then, we consider all nodes i with a directed path node $1\rightarrow j\rightarrow i$. In the 1st iteration, node j will take the position of node i in (23) implying $\left[\boldsymbol{C}_{f_{j,1}\rightarrow j}^{(1)}\right]^{-1} \succ \mathbf{0}$, while $\left[\boldsymbol{C}_{f_{i,j}\rightarrow i}^{(1)}\right]^{-1}$ has not been updated in the first iteration, i.e., $\left[\boldsymbol{C}_{f_{i,j}\rightarrow i}^{(1)}\right]^{-1} = \mathbf{0}$. In the second iteration, from (21), we have $\left[\boldsymbol{C}_{f_{i,j}\rightarrow i}^{(2)}\right]^{-1} \succ \mathbf{0}$. Since $\left[\boldsymbol{C}_{f_{k,j}\rightarrow j}^{(1)}\right]^{-1}$ equals $\mathbf{0}$, taking inverse on $\sum_{f\in\mathcal{B}(\boldsymbol{\omega}_j)\backslash f_{i,j}} \left[\boldsymbol{C}_{f\rightarrow j}^{(1)}\right]^{-1}$ gives $\left[\sum_{f\in\mathcal{B}(\boldsymbol{\omega}_j)\backslash f_{i,j}} \left[\boldsymbol{C}_{f\rightarrow j}^{(1)}\right]^{-1}\right]^{-1} \succeq \left[\sum_{f\in\mathcal{B}(\boldsymbol{\omega}_j)\backslash f_{i,j}} \left[\boldsymbol{C}_{f\rightarrow j}^{(2)}\right]^{-1}\right]^{-1}$. Further applying properties vi), iv) and v), we obtain

$$\underbrace{\boldsymbol{A}_{i,j}^T \left[\boldsymbol{R}_{i,j} + \boldsymbol{A}_{j,i}\left[\sum_{f\in\mathcal{B}(\boldsymbol{\omega}_j)\backslash f_{i,j}} \left[\boldsymbol{C}_{f\rightarrow j}^{(2)}\right]^{-1}\right]^{-1} \boldsymbol{A}_{j,i}^T\right]^{-1} \boldsymbol{A}_{i,j}}_{=\left[\boldsymbol{C}_{f_{i,j}\rightarrow i}^{(3)}\right]^{-1}}$$

$$\succeq \underbrace{\boldsymbol{A}_{i,j}^T \left[\boldsymbol{R}_{i,j} + \boldsymbol{A}_{j,i}\left[\sum_{f\in\mathcal{B}(\boldsymbol{\omega}_j)\backslash f_{i,j}} \left[\boldsymbol{C}_{f\rightarrow j}^{(1)}\right]^{-1}\right]^{-1} \boldsymbol{A}_{j,i}^T\right]^{-1} \boldsymbol{A}_{i,j}}_{=\left[\boldsymbol{C}_{f_{i,j}\rightarrow i}^{(2)}\right]^{-1}}. \tag{24}$$

Thus $\left[\boldsymbol{C}_{f_{i,j}\rightarrow i}^{(3)}\right]^{-1} \succeq \left[\boldsymbol{C}_{f_{i,j}\rightarrow i}^{(2)}\right]^{-1} \succ \left[\boldsymbol{C}_{f_{i,j}\rightarrow i}^{(1)}\right]^{-1} = \mathbf{0}$. In general, for any node i, if there exists a directed path from node $1\rightarrow\ldots\rightarrow j\rightarrow i$ in the network topology, there must be a finite iteration number s such that

$$\left[\boldsymbol{C}_{f_{i,j}\rightarrow i}^{(s+1)}\right]^{-1} \succeq \left[\boldsymbol{C}_{f_{i,j}\rightarrow i}^{(s)}\right]^{-1} \succ \left[\boldsymbol{C}_{f_{i,j}\rightarrow i}^{(s-1)}\right]^{-1} = \mathbf{0}. \tag{25}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

13

Finally, we divide the discussion into three cases, covering all possible relationships between two neighboring node i and j:

a) there exists a path from node $1 \to \ldots \to j \to i$ and $j \neq 1$;

b) there exists a path from node $1 \to i$;

c) there is no path from node $1 \to \ldots \to j \to i$.

For the first case, suppose $\left[\boldsymbol{C}_{f_{i,j} \to i}^{(l)} \right]^{-1} \succeq \left[\boldsymbol{C}_{f_{i,j} \to i}^{(l-1)} \right]^{-1}$ holds for $l > s$. Since $j \neq 1$, there must be a node k, such that $\left[\boldsymbol{C}_{f_{k,j} \to j}^{(l)} \right]^{-1} \succeq \left[\boldsymbol{C}_{f_{k,j} \to j}^{(l-1)} \right]^{-1}$. Then, it can be easily shown that $\sum_{f_{k,j} \in \mathcal{B}(\boldsymbol{\omega}_j) \backslash f_{i,j}} \left[\boldsymbol{C}_{f_{k,j} \to j}^{(l)} \right]^{-1} \succeq \sum_{f_{k,j} \in \mathcal{B}(\boldsymbol{\omega}_j) \backslash f_{i,j}} \left[\boldsymbol{C}_{f_{k,j} \to j}^{(l-1)} \right]^{-1}$. Following the same arguments above (24), it can be obtained that $\left[\boldsymbol{C}_{f_{i,j} \to i}^{(l+1)} \right]^{-1} \succeq \left[\boldsymbol{C}_{f_{i,j} \to i}^{(l)} \right]^{-1}$. Hence, by induction we have

$$\left[\boldsymbol{C}_{f_{i,j} \to i}^{(l)} \right]^{-1} \succeq \ldots \succeq \left[\boldsymbol{C}_{f_{i,j} \to i}^{(s+1)} \right]^{-1} \succeq \left[\boldsymbol{C}_{f_{i,j} \to i}^{(s)} \right]^{-1} \succ \boldsymbol{0}, \quad \text{for } l > s. \tag{26}$$

For the second case, if there exists a path node $1 \to i$, the corresponding result is in (23). For the third case, if the path node $1 \to \ldots \to j \to i$ does not exist, $\left[\boldsymbol{C}_{f_{i,j} \to i}^{(l)} \right]^{-1}$ never gets updated, and is always equal to $\left[\boldsymbol{C}_{f_{i,j} \to i}^{(0)} \right]^{-1} = \boldsymbol{0}$.

Since strongly connected network is considered, there is at least one $\tilde{j} \in \mathcal{I}(i)$ such that the first case is true, therefore, we obtain

$$\ldots \succeq \sum_{j \in \mathcal{I}(i)} \left[\boldsymbol{C}_{f_{i,j} \to i}^{(l+1)} \right]^{-1} \succeq \sum_{j \in \mathcal{I}(i)} \left[\boldsymbol{C}_{f_{i,j} \to i}^{(l)} \right]^{-1} \succeq \ldots \sum_{j \in \mathcal{I}(i)} \left[\boldsymbol{C}_{f_{i,j} \to i}^{(s)} \right]^{-1} \succ \boldsymbol{0}, \quad \text{for } l > s. \tag{27}$$

Applying matrix inverse to (27) and using the definition of $\boldsymbol{P}_i^{(l)}$ in (17), we have

$$\boldsymbol{P}_i^{(s)} \succeq \ldots \succeq \boldsymbol{P}_i^{(l)} \succeq \boldsymbol{P}_i^{(l+1)} \succeq \ldots \succ \boldsymbol{0}, \quad \text{for } l > s, \tag{28}$$

where the p.d. property of $\boldsymbol{P}_i^{(l)}$ is due to property i). Consequently such non-increasing p.d. matrix sequence converges to a p.d. matrix [42]. ∎

The importance of Theorem 1 is that if a reference node exists, the belief covariance matrices always converge. Next, we investigate the convergence of belief mean vectors.

Theorem 2. The mean $\boldsymbol{\mu}_i^{(l)}$ of the belief $b^{(l)}(\boldsymbol{\omega}_i)$ converges to a fixed vector regardless of the network topology.

Proof: From the proof of Theorem 1, there are three cases of relationships between node i and node j (above (26)). For the first and second cases, the evolution of $\left[\boldsymbol{C}_{f_{i,j} \to i}^{(l)} \right]^{-1}$ are described by (26) and (23), respectively. Taking matrix inverse of (23) and (26), we can readily see that $\boldsymbol{C}_{f_{i,j} \to i}^{(l)}$ is a monotonically decreasing matrix sequence and bounded by $\boldsymbol{0}$. Thus, $\boldsymbol{C}_{f_{i,j} \to i}^{(l)}$ is convergent.

For the third case, $C_{f_{i,j}\to i}^{(l)}$ is never updated, and thus can also be viewed as convergent. On the other hand, computation of $C_{j\to f_{i,j}}^{(l)}$ depends on $C_{f_{i,j}\to i}^{(l)}$ as shown in (15). So, if $C_{f_{i,j}\to i}^{(l)}$ is convergent, then $C_{j\to f_{i,j}}^{(l)}$ is also convergent. In this proof, it is assumed that $C_{f_{i,j}\to i}^{(l)}$ and $C_{j\to f_{i,j}}^{(l)}$ have already converged to $C_{f_{i,j}\to i}^{(*)}$ and $C_{j\to f_{i,j}}^{(*)}$, respectively, as the convergence of message covariance matrices do not depend on the message mean vectors.

Substituting (16) into (13), we obtain the mean update rules from factor nodes to variable nodes as

$$
\boldsymbol{v}_{f_{i,j}\to i}^{(l)} = \underbrace{\boldsymbol{C}_{f_{i,j}\to i}^{(*)}\boldsymbol{A}_{i,j}^{T}\left[\boldsymbol{R}_{i,j}+\boldsymbol{A}_{j,i}\boldsymbol{C}_{j\to f_{i,j}}^{(*)}\boldsymbol{A}_{j,i}^{T}\right]^{-1}}_{\triangleq \boldsymbol{M}_{i,j}}
\tag{29}
$$
$$
\left\{\boldsymbol{r}_{i,j}-\underbrace{\boldsymbol{A}_{j,i}\boldsymbol{C}_{j\to f_{i,j}}^{(*)}}_{\triangleq \boldsymbol{F}_{i,j}}\sum_{f\in\mathcal{B}(\boldsymbol{\omega}_j)\backslash f_{j,i}}\left[\boldsymbol{C}_{f\to j}^{(*)}\right]^{-1}\boldsymbol{v}_{f\to j}^{(l-1)}\right\}.
$$

Without loss of generality, define $\boldsymbol{v}^{(l)}$ as a vector containing all $\boldsymbol{v}_{f_{i,j}\to i}^{(l)}$ with ascending index[3] first on i and then on j. We can write (29) as

$$
\boldsymbol{v}_{f_{i,j}\to i}^{(l)} = \boldsymbol{M}_{i,j}\boldsymbol{r}_{i,j}-\boldsymbol{M}_{i,j}\boldsymbol{F}_{i,j}\boldsymbol{\Gamma}_{i,j}\boldsymbol{v}^{(l-1)},
\tag{30}
$$

where $\boldsymbol{\Gamma}_{i,j}$ is a block matrix containing $\left[\boldsymbol{C}_{f_{j,k}\to j}^{(*)}\right]^{-1}$ as component blocks such that (30) is satisfied. Stacking (30) for all i and j, and writing $\boldsymbol{v}^{(l)} \triangleq \left[(\boldsymbol{v}_x^{(l)})^T,(\boldsymbol{v}_y^{(l)})^T\right]^T$, where $\boldsymbol{v}_x^{(l)}$ containing $\boldsymbol{v}_{f_{i,j}\to i}^{(l)}$ with $j=1$, while $\boldsymbol{v}_y^{(l)}$ containing the remaining part of $\boldsymbol{v}^{(l)}$, we obtain

$$
\underbrace{\begin{bmatrix}\boldsymbol{v}_x^{(l)}\\\boldsymbol{v}_y^{(l)}\end{bmatrix}}_{\triangleq \boldsymbol{v}^{(l)}}=\underbrace{\begin{bmatrix}\boldsymbol{X}&\boldsymbol{Y}\\\boldsymbol{Q}_1&\boldsymbol{Q}_2\end{bmatrix}}_{\triangleq \boldsymbol{Q}}\underbrace{\begin{bmatrix}\boldsymbol{v}_x^{(l-1)}\\\boldsymbol{v}_y^{(l-1)}\end{bmatrix}}_{\triangleq \boldsymbol{v}^{(l-1)}}+\underbrace{\begin{bmatrix}\boldsymbol{\xi}_x\\\boldsymbol{\xi}_y\end{bmatrix}}_{\triangleq \boldsymbol{\xi}}.
\tag{31}
$$

On the other hand, putting $j=1$ into (29) and notice that $\boldsymbol{C}_{f_1\to 1}=\boldsymbol{0}$ and $\boldsymbol{C}_{1\to f_{i,1}}^{(*)}=\boldsymbol{0}$, we have

$$
\boldsymbol{v}_{f_{i,1}\to i}^{(l)} = \left[\boldsymbol{A}_{i,1}^{T}\boldsymbol{R}_{i,1}^{-1}\boldsymbol{A}_{i,1}\right]^{-1}\boldsymbol{A}_{i,1}^{T}\boldsymbol{R}_{i,1}^{-1}(\boldsymbol{r}_{i,1}-\boldsymbol{A}_{1,i}\boldsymbol{\omega}_1).
\tag{32}
$$

which shows that $\boldsymbol{v}_{f_{i,1}\to i}^{(l)}$ never changes with iteration number l. Since $\boldsymbol{v}_x^{(l)}$ containing $\boldsymbol{v}_{f_{i,1}\to i}^{(l)}$ as

---

[3]The order of $\boldsymbol{v}_{f_{i,j}\to i}^{(l)}$ arranged in $\boldsymbol{v}^{(l)}$ in fact can be arbitrary as long as it does not change after the order is fixed.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

15

components, $v_x^{(l)}$ is fixed and independent of l. Hence, we can write (31) equivalently as

$$
\underbrace{\begin{bmatrix} v_{\text{fix}} \\ v_y^{(l)} \end{bmatrix}}_{v^{(l)}} = \underbrace{\begin{bmatrix} I & 0 \\ Q_1 & Q_2 \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} v_{\text{fix}} \\ v_y^{(l-1)} \end{bmatrix}}_{v^{(l-1)}} + \underbrace{\begin{bmatrix} 0 \\ \xi_y \end{bmatrix}}_{\xi}
\tag{33}
$$

where $v_{\text{fix}} = v_x^{(l)}$ represents the stacked messages for $j = 1$. Notice that $[Q_1, Q_2]$ depends on $-M_{i,j}F_{i,j}\Gamma_{i,j}$, while ξ_y depends on $M_{i,j}a_{i,j}$. It is obvious that Q_1, Q_2 and ξ_y are independent of iteration number l. Next, we will show a property of Q_2.

Since $\mathbb{E}_{\omega_i,\omega_j,n_{i,j}}\{r_{i,j}\} = \mathbb{E}\{A_{i,j}\omega_i + A_{j,i}\omega_j + n_{i,j}\} = 0$, taking expectation on both sides of (33), we have

$$
\underbrace{\begin{bmatrix} \bar{v}_{\text{fix}} \\ \bar{v}_y^{(l)} \end{bmatrix}}_{\triangleq \bar{v}^{(l)}} = \underbrace{\begin{bmatrix} I & 0 \\ Q_1 & Q_2 \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} \bar{v}_{\text{fix}} \\ \bar{v}_y^{(l-1)} \end{bmatrix}}_{\triangleq \bar{v}^{(l-1)}}
\tag{34}
$$

or equivalently

$$
\bar{v}^{(l)} = Q^l \bar{v}^{(0)},
\tag{35}
$$

where $\bar{x}$ denotes the expectation of x. Since there is always a positive value η, satisfying $\eta > \sum_{i \neq j} |[Q]_{i,j}|$ for all i, we have $\eta I + Q$ is strictly diagonally dominant and then $\eta I + Q$ is nonsingular [43]. Hence, arbitrary initial value $\bar{v}^{(0)}$ can be expressed in terms of the eigenvectors of $\eta I + Q$ as $\bar{v}^{(0)} = \sum_{d=1}^{D} c_d q_d$, where $q_1, q_2, \cdots, q_D$ are the eigenvectors of $\eta I + Q$. Since the eigenvectors of $\eta I + Q$ is the same as that of Q, and the eigenvalues of $\eta I + Q$ are $\eta + \lambda_d$ ($1 \leqslant d \leqslant D$), where λ_d is the eigenvalue of Q, we have

$$
\bar{v}^{(l)} = Q^l \bar{v}^{(0)} = \sum_{d=1}^{D} c_d \lambda_d^l q_d.
\tag{36}
$$

Without loss of generality, suppose λ_d are arranged in descending order as below

$$
|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_D|.
\tag{37}
$$

Let the eigenvalue with the largest magnitude has a multiplicity of d_0. Then $\lambda_d/\lambda_1 < 1$ for $d > d_0$ and $(\lambda_d/\lambda_1)^l = 0$ if l is large enough. We then obtain

$$
\bar{v}^{(l)} = \lambda_1^l \sum_{d=1}^{d_0} c_d q_d,
\tag{38}
$$

for large l. Taking expectation on (32), we have $\bar{v}_{f_{i,1} \to i}^{(l)} = -\left[A_{i,1}^T R_{i,1}^{-1} A_{i,1}\right]^{-1} A_{i,1}^T R_{i,1}^{-1} A_{1,i}\omega_1$. It is obvious that $\bar{v}_{f_{i,1} \to i}^{(l)}$ never change with the iteration number l. Hence, the first element of $\bar{v}^{(l)}$

is $\bar{\boldsymbol{v}}^{(l)}(1) \triangleq \gamma_c$, which is a constant, and according to (38), we have $\lambda_1^l = \frac{\gamma_c}{\sum_{d=1}^{d_0} c_d \boldsymbol{q}_d(1)}$ for l large enough. Substituting it back to (38) yields

$$\bar{\boldsymbol{v}}^{(l)} = \frac{\gamma_c \sum_{d=1}^{d_0} c_d \boldsymbol{q}_d}{\sum_{d=1}^{d_0} c_d \boldsymbol{q}_d(1)}. \tag{39}$$

It is obvious that $\bar{\boldsymbol{v}}^{(l)}$ does not change when l is large enough, and therefore, $\bar{\boldsymbol{v}}^{(l)}$ in (35) converges. Since (34) and (35) are equivalent, $\bar{\boldsymbol{v}}_y^{(l)}$ in (34) also converges. With iteration equation in (34) rewritten as

$$\bar{\boldsymbol{v}}_y^{(l)} = \boldsymbol{Q}_2 \bar{\boldsymbol{v}}_y^{(l-1)} + \boldsymbol{Q}_1 \bar{\boldsymbol{v}}_{\text{fix}}, \tag{40}$$

and since (40) converges, the spectrum radius $\rho(\boldsymbol{Q}_2) < 1$ [44].

Now rewriting (33) as

$$\boldsymbol{v}_y^{(l)} = \boldsymbol{Q}_2 \boldsymbol{v}_y^{(l-1)} + \boldsymbol{Q}_1 \boldsymbol{v}_{\text{fix}} + \boldsymbol{\xi}_y. \tag{41}$$

With $\boldsymbol{Q}_1 \boldsymbol{v}_{\text{fix}}$ being a constant vector, and $\rho(\boldsymbol{Q}_2) < 1$, we also have (41) converges. Thus, the sequence $\boldsymbol{v}_{f_{i,j} \to i}^{(l)}$ in (29) is convergent for any initial vectors $\boldsymbol{v}_{f_{i,j} \to i}^{(0)}$ [44]. Finally, with $\boldsymbol{\mu}_i^{(l)}$ defined in (18), since $\boldsymbol{P}_i^{(l)}$, $\boldsymbol{C}_{f_{i,j} \to i}^{(l)}$ and $\boldsymbol{v}_{f_{i,j} \to i}^{(l)}$ converge, we can draw the conclusion that the vector sequence $\{\boldsymbol{\mu}_i^{(1)}, \boldsymbol{\mu}_i^{(2)}, \ldots\}$ converges. ∎

Although Theorem 2 states that the proposed iterative message mean $\boldsymbol{\mu}_i^{(l)}$ converges to a fixed point, we still need to answer the important question that how accurate the converged $\boldsymbol{\mu}_i^* = \lim_{l \to +\infty} \boldsymbol{\mu}_i^{(l)}$ is?

Theorem 3. The converged BP message mean vector $\boldsymbol{\mu}_i^*$ equals the centralized ML estimate $\hat{\boldsymbol{\omega}}_i^{\text{ML}}$. Furthermore, the estimation MSE of $\boldsymbol{\mu}^* = [(\boldsymbol{\mu}_2^*)^T, \ldots, (\boldsymbol{\mu}_K^*)^T]^T$ asymptotically (in high SNR or large training length N or both) equals the centralized CRB of $\boldsymbol{\omega} = [\boldsymbol{\omega}_2^T, \ldots, \boldsymbol{\omega}_K^T]^T$:

$$\text{CRB}(\boldsymbol{\omega}) = \left(\boldsymbol{A}^T \boldsymbol{R}^{-1} \boldsymbol{A}\right)^{-1}, \tag{42}$$

where $\boldsymbol{A}$ is obtained from stacking (4) into the form of $\boldsymbol{r} = \boldsymbol{A}\boldsymbol{\omega} + \boldsymbol{n}$, with $\boldsymbol{r}$ being a vector containing $\boldsymbol{r}_{i,j}$ with ascending index first on i and then on j; $\boldsymbol{R}$ is a block diagonal matrix with $\boldsymbol{R}_{i,j}$ as block diagonal and with the same order as $\boldsymbol{r}_{i,j}$ in $\boldsymbol{r}$.

Proof: Since the likelihood function in (6) is multivariate Gaussian and it is known that if Gaussian BP converges, the mean of the beliefs computed by BP equals the centralized ML estimate [37] from

$$\boldsymbol{r} = \boldsymbol{A}\boldsymbol{\omega} + \boldsymbol{n}, \tag{43}$$

where r is a vector containing $r_{i,j}$ with ascending indexes first on i and then on j; and n containing $n_{i,j}$ with the indexes i, j ordered in the same way as in r. Since $n \sim \mathcal{N}(n; 0, R)$, where R is a block diagonal matrix with $R_{i,j}$ as block diagonal and with the same order as $r_{i,j}$ in r, and (43) is a standard linear model, the ML estimator and thus $\mu^* = \left[(\mu_2^*)^T, \ldots, (\mu_K^*)^T\right]^T$ asymptotically approach the CRB for ω given by $\text{CRB}(\omega) = \left(A^T R^{-1} A\right)^{-1}$ [32]. ∎

V. SIMULATION RESULTS

This section presents numerical results to assess the performance of the proposed algorithm. Estimation MSE are presented for CFO estimation over the whole random network, which consists of 14 nodes randomly located in a $[0, 100] \times [0, 100]$ area. The communication range for each node is 38. In each trial, the normalized CFO of each antenna on each node (except node 1 where CFO is zero) is generated independently and is uniformly distributed in the range $2\pi[-0.2, 0.2]$. Besides, the channel between each pair of nodes is Rayleigh flat-fading. The relative CFOs and channels are first estimated based on the algorithm in [19], with training length N. Then the BP algorithm is executed for network-wide CFOs estimation and compensation. 5000 simulation runs were performed to obtain the average performance for each point in the figures.

First, consider the network shown in Fig. 2 and each node equipped with two antennas. We employ training with length $N = 16$ for relative CFOs estimation. The SNR during training stage and BP message passing are the same. Fig. 4 shows the sum MSE[4] of ω_i for $i = \{6, 2\}$ as a function of BP iteration number l. These two nodes are chosen to represent nodes close to (node 2) and far away (node 6) from the reference node. It can be seen that for both SNR$= 10$dB and 30dB, the MSEs decrease quickly and approach the corresponding CRBs in only a few iterations. Furthermore, a close inspection reveals that convergence is slightly faster at SNR $= 30$dB than that at SNR $= 10$dB, but the difference is very small.

Fig. 5 shows the average sum MSE of $\{\omega_i\}_{i \in \mathcal{V}}$ versus SNRs for different training length N. The network is randomly generated within the $[0, 100] \times [0, 100]$ area in each trial, and each node is equipped with 2 antennas. As shown in the figure, the MSEs of proposed distributed algorithm achieve the best performance as the MSEs touch the corresponding CRBs. This verifies Theorem 3. Furthermore, with increasing N, the approximation of $R_{i,j}$ to $B_\epsilon^{\{i,j\}}(\{\epsilon_k\}_{k=1}^{N_j}, \{h_k\}_{k=1}^{N_j})$

---

[4]Sum MSE over the two antennas.

becomes better at lower SNR, and thus the estimation MSEs of $\boldsymbol{\omega}_i$ achieves the corresponding CRBs earlier.

Finally, we compare the performance of the proposed algorithm with that of the consensus based D-FLL approach in [29] and pulse-coupled oscillator method in [30], where CFOs are adjusted to an average common value in each iteration based on consensus principle. The estimation error of consensus method at the lth iteration is measured by the total mean-square deviation of the individual variables from their average, which is

$$\text{MSE}^{(l)}_{\text{consensus}} = \frac{1}{K} \sum_{i=1}^{K} \mathbb{E} \left\{ \left\| \boldsymbol{\mu}_i^{(l)} - \frac{1}{K} \sum_{i=1}^{K} \boldsymbol{\mu}_i^{(l)} \right\|^2 \right\}. \tag{44}$$

On the other hand, the proposed algorithm estimates the absolute CFO values, therefore, the network estimation MSE at the lth iteration is

$$\text{MSE}^{(l)}_{\text{BP}} = \frac{1}{K-1} \sum_{i=2}^{K} \mathbb{E} \left\{ \left\| \boldsymbol{\mu}_i^{(l)} - \boldsymbol{\omega}_i \right\|^2 \right\}. \tag{45}$$

We consider the 14 nodes randomly located within the $[0, 100] \times [0, 100]$ area and for fair comparison with D-FLL and pulse-coupled oscillator method, each node is equipped with a single antenna. For D-FLL method, pilots of length 16 are transmitted by each node in each iteration, while for the pulse-coupled oscillator method, in each iteration, one pilot symbol is sent from each node for consensus updating. For the proposed method, 16 pilots are used in the joint CFO and channel estimation at the initial phase. The convergence performance of the three algorithms at different SNRs are shown in Fig. 6. It is apparent that convergence speed of the consensus based algorithms decrease with SNRs, and in general takes several hundreds of iterations to converge. For example, for D-FLL method at SNR $=$ 5dB, around 800 iterations are required. While the proposed method requires only a few iterations to approach the corresponding CRBs, the fast convergence comes from the exchanges of slightly more information (two real numbers represnting the mean and variance) compared to consensus methods.

VI. CONCLUSIONS

In this paper, a fully distributed CFOs estimation algorithm for cooperative and distributed networks was proposed. The algorithm is based on BP and is easy to be implemented by exchanging limited amount of information between neighboring nodes, thus is scalable with network size. Furthermore, it was shown analytically that the proposed distributed algorithm

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

19

converges to the optimal solution with estimation MSE coinciding with the centralized CRB asymptotically regardless of network topology. Simulation results showed that the MSE of the proposed method approaches the CRB within only a few iterations.

REFERENCES

[1] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers: Synchronization, Channel Estimation, and Signal Processing.* Wiley-InterScience, John Wiley & Sons, Inc., 1997.

[2] T. Pollet, M. V. Bladel, and M. Moeneclaey, "BER sensitivity of OFDM systems to carrier frequency offset and Wiener phase noise," *IEEE Trans. Commun.*, vol. 43, pp. 191–193, Feb./Mar./Apr. 1995.

[3] H. Steendam and M. Moeneclaey, "Synchronization sensitivity of multicarrier systems," *European Commun.*, ETT special issue on multi-carrier spread spectrum, vol. 52, no. 5, pp. 834-844, May 2004.

[4] V. K. V. Gottumukkala and H. Minn, "Capacity analysis and pilotdata power allocation for MIMO-OFDM with transmitter and receiver IQ imbalances and residual carrier frequency offset," *IEEE Trans. Veh. Technol.*, vol. 61, no. 2, pp. 553-565, Feb. 2012.

[5] I.-H. Wang and D. N. C. Tse, "Interference mitigation through limited receiver cooperation, *IEEE Trans. Inf. Theory.*, vol. 57, no. 5, pp. 2913-2940, May 2011.

[6] V. Havary-Nassab, S. Shahbazpanahi, A. Grami, and Z. Q. Luo, "Distributed beamforming for relay networks based on second-order statistics of the channel state information," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4306-4316, Sep. 2008.

[7] M. Xia, W. Wen, and S.-C. Kim, "Opportunistic cophasing transmission in MISO systems," *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3764–3770, Dec. 2009.

[8] A. D. Wyner, "Shannon-theoretic approach to a Gaussian cellular multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1713-1727, Nov. 1994.

[9] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102-111, Feb. 2011.

[10] D. Gesbert, S. Hanly, H. Huang, S. Shamai, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: a new look at interference," *IEEE J. Select. Areas Commun.*, vol. 28, no. 9, pp. 1380-1408, Dec. 2010.

[11] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, "Femtocells: Past, present, and future," *IEEE J. Select. Areas Commun.*, vol. 30, no. 3, pp. 497 – 508, April 2012.

[12] J. G. Andrews "The Seven Ways HetNets are a Paradigm Shift", to appear in *IEEE Comm. Mag.*, Mar. 2013.

[13] C. Xing, Z. Fei, S. Ma, Y.-C. Wu, and H. V. Poor, "A general robust linear transceiver Design for amplify-and-forward multi-hop MIMO relaying systems," *IEEE Trans. Signal Process.*, vol. 61, no. 5, pp. 1196-1209, Mar. 2013.

[14] F. Pantisano, M. Bennis, W. Saad, and M. Debbah, "Spectrum Leasing as an Incentive towards Uplink Macrocell and Femtocell Cooperation," *IEEE J. Select. Areas Commun.*, vol. 30, no. 3, 617 – 630, April 2012

[15] M. Morelli and U. Mengali, "Carrier-frequency estimation for transmissions over selective channels," *IEEE Trans. Commun.*, vol. 48, pp. 1580–1589, Sep. 2000.

[16] L. Sanguinetti, M. Morelli, and G. Imbarlina, "An EM-based frequency offset estimator for OFDM systems with unknown interference," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4470–4475, Sep. 2009.

[17] M. Ghogho and A. Swami, "Training design for multipath channel and frequency-offset estimation in MIMO systems," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3957–3965, Oct. 2006.

[18] O. Besson and P. Stoica, "On parameter estimation of MIMO at fading channels with frequency offsets," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 602–613, Mar. 2003

[19] M. O. Pun, M. Morelli, and C.-C. J. Kuo, "Maximum-likelihood synchronization and channel estimation for OFDMA uplink transmissions," *IEEE Trans. Commun.*, vol. 54, no. 4, pp. 726–736, Apr. 2006.

[20] T. Pham, A. Nallanathan, and Y. Liang, "Joint channel and frequency offset estimation in distributed MIMO flat-fading channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 2, pp. 648–656, Feb. 2008

[21] J. Chen, Y.-C. Wu, S. Ma, and T.-S. Ng, "Joint CFO and channel estimation for multiuser MIMO-OFDM systems with optimal training sequences," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 4008-4019, Aug. 2008.

[22] H. G. Jeon, K. S. Kim, and E. Serpedin, "An efficient blind deterministic frequency offset estimator for OFDM systesm," *IEEE Trans on Communications.*, vol. 59, no. 4, Apr. 2011.

[23] F. F. Gao, T. Cui, A. Nallanathan, and C. Tellambura, "Maximum likelihood based estimation of frequency and phase offset in DCT OFDM systems under non-circular transmissions: Algorithms, analysis and comparisons," *IEEE Trans on Communications.*, vol. 56, no. 9, pp. 1425–1429, Sep. 2008.

[24] P. A. Parker, P. Mitran, D. W. Bliss, and V. Tarokh, "On bounds and algorithms for frequency synchronization for collaborative communication systems," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3742-3752, Aug. 2008.

[25] B. Zarikoff and J. Cavers, "Coordinated multi-cell systems: Carrier frequency offset estimation and correction," *IEEE Trans. Select. Areas Commun.*, vol. 28, no. 9, pp. 1490–1501, Dec. 2010

[26] R. Mudumbai, G. Barriac, and U. Madhow, "On the feasibility of distributed beamforming in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1754–1763, May 2007.

[27] D. R. Brown III and H. V. Poor. "Time-Slotted Round-Trip Carrier Synchronization for Distributed Beamforming," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5630-5643, November 2008.

[28] R. Preuss and D. R. Brown III. "Two-Way Synchronization for Coordinated Multicell Retrodirective Downlink Beamforming," *IEEE Trans. Signal Process.*,vol. 59, no. 11, 5415-5427, November 2011.

[29] N. Varanese, U. Spagnolini, Y. Bar-Ness, "Distributed frequency-looked loops for wireless networks," *IEEE Trans on Communications.*, vol.59, n.12, pp. 3440-3451, Dec. 2011

[30] M. M. Rahman, S. Dasgupta, and R. Mudumbai, "A distributed consensus approach to synchronization of rf signals," in *IEEE Statistical Signal Processing Workshop (SSP)*, pp. 281-284, Aug. 2012.

[31] A.-S. Hu and S. Servetto, "On the scalability of cooperative time synchronization in pulse connected networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2725-2748, Jun. 2006.

[32] S. M. Kay, *Fundamentals of Statistical Signal Processing Estimation Theory,* Upper Saddle River, NJ: Prentice-Hall, 1993.

[33] P. Stoica and O. Besson, "Training sequence design for frequency offset and frequency-selective channel estimation," *IEEE Trans. Commun*, vol. 51, no. 11, pp. 1910-1917, Nov. 2003.

[34] L. L. He, S. D. Ma, Y. C. Wu, and T. S. Ng, "Semiblind iterative data detection for OFDM systems with CFO and doubly selective channels," *IEEE Trans. on Commun.*, vol. 58, no. 12, pp. 3491-3499, Dec. 2010.

[35] E. Simon, L. Ros, H. Hijazi, and M. Ghogho, Joint carrier frequency offset and channel estimation for OFDM systems via the EM algorithm in the presence of very high mobility," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 754–765, Feb. 2012

[36] 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.

[37] 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.

[38] D. M. Malioutov, J. K. Johnson, and A. S. Willsky, "Walk-sums and belief propagation in Gaussian graphical models," *Journal of Machine Learning Research,* vol, 7, Oct. 2006.

[39] M. Leng and Y.-C. Wu, "Distributed clock synchronization for wireless sensor networks using belief propagation," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5404–5414, Nov. 2011.

[40] A. Papoulis and S. U. Pillai, *Random Variables and Stochastic Processes,* 4th ed. New York: McGraw-Hill, 2002.

[41] S. Puntanen, G. Styan, and J. Isotalo, *Matrix Tricks for Linear Statistical Models Our Personal Top Twenty,* Springer, 2011.

[42] A. M. Sarhan, N. M. EI-Shazly, and E. M. Shehata, On the existence of extremal positive defnite solutions of the nonlinear matrix equation $X^r + \sum_{i=1}^{m} A_i^* X^{\delta_i} A_i = I$," Mathematical and Computer Modelling, vol. 51, no. 9- 10, 2010.

[43] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[44] J. W. Demmel *Applied Numerical Linear Algebra*. SIAM, 1997.

(a) Distributed beamforming networks

(b) Multi-cell cooperative networks

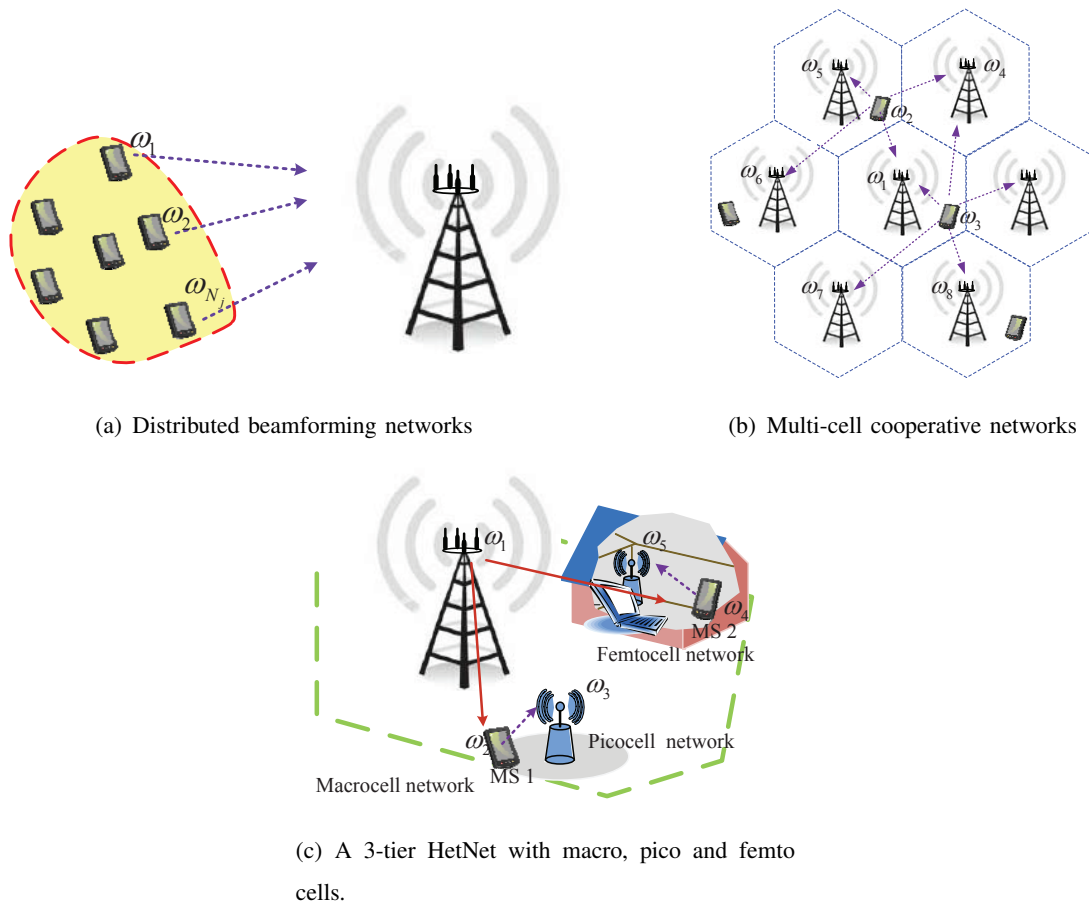(c) A 3-tier HetNet with macro, pico and femto cells.

Fig. 1. Scenarios that need network-wide frequency synchronization
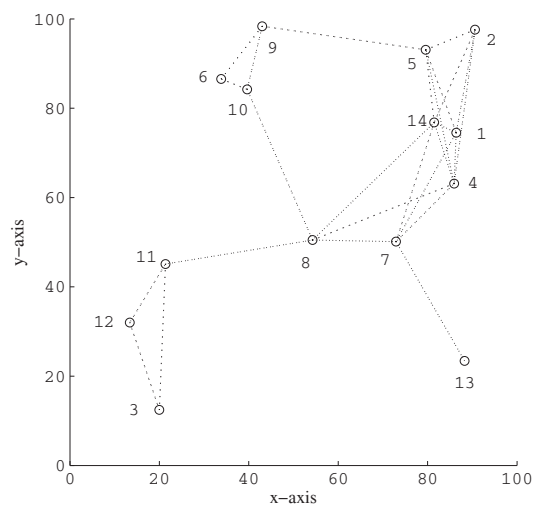

Fig. 2. An example of a network topology with 14 nodes.
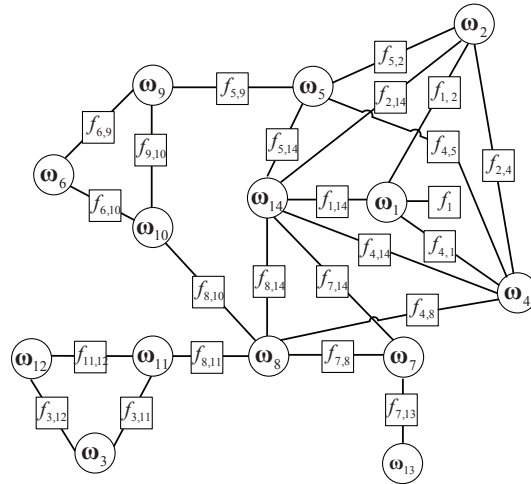
Fig. 3. The factor graph representation of the network in Fig. 2.
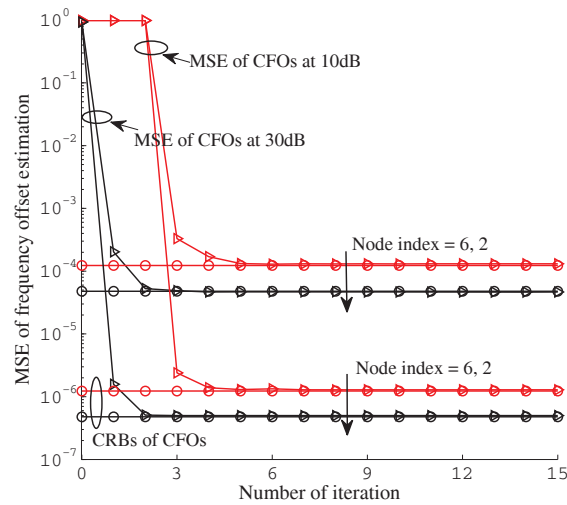

Fig. 4. Convergence performance of the proposed algorithm at Node 2 and 6.
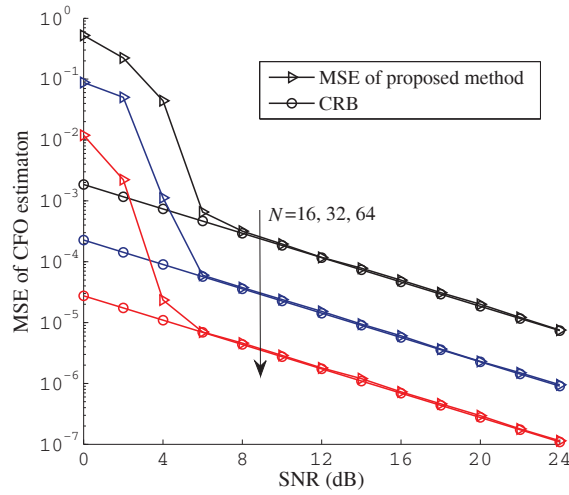
Fig. 5.   MSE of CFOs $\{\omega_i\}_{i \in \mathcal{V}}$ averaged over the whole network with respect to SNRs.
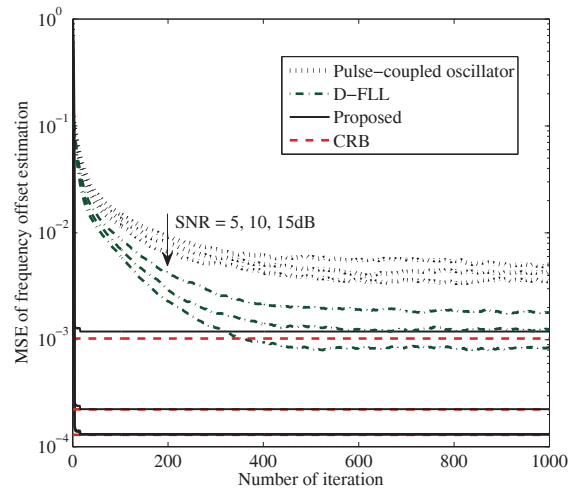

Fig. 6.   Convergence of the proposed method and consensus based methods in [29] and [30] in single antenna case.

Jian Du received the B.Eng. and M.Eng. degrees in communication and information engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2006 and 2009, respectively. He is currently working towards his Ph.D. degree with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong. His research interests are in distributed signal processing, statistical learning and inferences for wireless communication systems and smart grid.

Yik-Chung Wu received the B.Eng. (EEE) degree in 1998 and the M.Phil. degree in 2001 from the University of Hong Kong (HKU). He received the Croucher Foundation scholarship in 2002 to study Ph.D. degree at Texas A&M University, College Station, and graduated in 2005. From August 2005 to August 2006, he was with the Thomson Corporate Research, Princeton, NJ, as a Member of Technical Staff. Since September 2006, he has been with HKU, currently as an Associate Professor. He was a visiting scholar at Princeton University, in summer 2011. His research interests are in general area of signal processing and communication systems, and in particular distributed signal processing and communications; optimization theories for communication systems; estimation and detection theories in transceiver designs; and smart grid. Dr. Wu served as an Editor for IEEE Communications Letters, is currently an Editor for IEEE Transactions on Communications and Journal of Communications and Networks.