On Bounds and Algorithms for Frequency Synchronization for Collaborative Communication Systems

Peter A. Parker, Patrick Mitran, Member, IEEE, Daniel W. Bliss, and Vahid Tarokh, Senior Member, IEEE

Abstract—Cooperative diversity systems are wireless communication systems designed to exploit cooperation among users to mitigate the effects of multipath fading. In fairly general conditions, it has been shown that these systems can achieve the diversity order of an equivalent multiple-input multiple-output (MIMO) channel and, if the node geometry permits, virtually the same outage probability can be achieved as that of the equivalent MIMO channel for a wide range of applicable signal-to-noise ratio (SNR). However, much of the prior analysis has been performed under the assumption of perfect timing and frequency offset synchronization. In this paper, we derive the estimation bounds and associated maximum a posteriori estimators for frequency offset estimation in a cooperative communication system. We show the benefit of adaptively tuning the frequency of the relay node in order to reduce estimation error at the destination. We also derive an efficient estimation algorithm, based on the correlation sequence of the data, which has mean squared error close to the Cramér-Rao Bound (CRB).

Index Terms—Cooperative diversity, Cramér–Rao bound, frequency offset estimation, maximum a posteriori (MAP) estimation.

I. INTRODUCTION

COOPERATIVE communication systems employ cooperation among nodes in a wireless network to increase data throughput and robustness to signal fading. Much of the research done in this area has concentrated on information theoretic results, protocols, and coding while assuming perfect synchronization [1]–[6]. In this paper, we explore frequency synchronization of a collaborative system and provide estimation bounds and practical algorithms having performance close to the bounds.

In a collaborative system, nodes that would have remained silent during some period of time adapt to their surroundings and collaborate with the source and destination nodes. These systems, sometimes termed cooperative diversity systems, use distributed protocols to greatly improve performance over traditional point-to-point communication systems. One improvement to system performance comes in the form of added robustness to signal fading [1], [2]. An effective way to achieve robustness is to increase spatial diversity by using multiple antennas as in a MIMO system [7], [8]. However, when considering a network of low-cost wireless devices, the size and cost of multiple antennas is prohibitive for these devices [9]. A way for low cost nodes to realize much of the benefit of a MIMO system is through collaborative (cooperative) diversity. In fact, in [10] it is shown that a collaborative system can have the same diversity order as an equivalent MIMO system. Employing a collaborative protocol in a wireless network can also increase the overall throughput of the network. The use of relaying is a special case of coding across a network and as shown in [11], the capacity of a relay (or coded) network is greater than in a traditional point-to-point network.

To design a practical collaborative communication system, one of two methods may be used. The signal coding and decoding may be designed to be naturally robust to synchronization errors [12], or alternatively, the frequency and timing offsets are estimated and subsequently compensated [13]. We explore the second option in this paper. Algorithms and bounds for standard synchronization are found in [14]–[16]. The related case of a MIMO channel with multiple frequency offsets is treated in [17] and [18]. In this paper, we provide more details and extend the results of [19]. We derive the transmission frequency the relay must use to optimally reduce the variance of the frequency estimator at the destination by minimizing the Cramér–Rao Bound (CRB) of the frequency estimators at each receive node. By using the CRB, our frequency selection algorithm is independent of algorithm choice. We also provide a pragmatic frequency estimation algorithm for the collaborative system and show that there is very little room for improvement as the performance is close to the CRB.

In [13], Shin et al. describes a specific protocol, which we use in this paper, for collaborative communication with synchronization among three nodes: a source, a relay, and a destination. The protocol is based on a two-phase transmission within each frame [1], [4], a listening phase and a cooperation phase. Within each phase there is a preamble containing synchronization signals. In the listening phase, the relay receives and decodes the source’s message. During the cooperation phase, the relay re-encodes and transmits the message cooperatively with the source. This process is illustrated in Fig. 1.

The synchronization algorithms in [13] are ad hoc and meant only to serve as a proof-of-concept that synchronization is pos-
sible with collaborative systems. In this paper, we derive the CRB for optimal frequency offset estimation for the class of systems discussed above. We show there exists an optimal (with respect to minimizing the CRB) frequency of transmission for the relay node based on: 1) the accuracy of estimation during the listening phase and 2) the signal-to-noise ratio (SNR) of all node pairs. We derive the maximum a posteriori (MAP) frequency estimators for each receive node. These estimators are asymptotically efficient, meaning they approach the CRB at high SNR and small frequency uncertainty. However, the MAP solution is computationally expensive and we therefore derive a practical correlation based estimation algorithm with performance close to the CRB. For the purposes of this paper, we assume a frequency selective channel, given by

\[ y_{a}[n] = e^{j2\pi f_{a}n} s_{a}[n] + u_{b}[n] \]  

(1)

where \( n \) is the sample index, \( f_{a} \) is the frequency offset between the two nodes of link \( a \) normalized by the sample rate (e.g., \( f_{sr} = f_{r} - f_{s} \)), \( u_{b}[n] \) is the noise generated in the electronics of receiver \( b \) \( \in \{ d, r \} \) (destination or relay node respectively), and \( s_{a}[n] \) is the combination of the known training signals \( (x_{0}, \ldots, x_{N_{t} - 1})^{T} \) and the effects of the frequency selective channel, given by

\[ s_{a}[n] = \sum_{k=0}^{P-1} h_{a}[k] x_{t}[n - k]. \]  

(2)

In this equation, \( h_{a}[n] \) are the samples of the channel response for link \( a \) and \( P \) is the duration of the channel response. We assume, for each link \( a \), the length of the channel \( P \) is the same. Note that we consider \( h_{a}[n] \) a deterministic parameter for the purposes of this paper. Writing (1) in matrix form gives

\[ y_{a} = V_{f_{a}} X_{a} h_{a} + w_{b} \in \mathbb{C}^{(N_{t}+P-1)\times 1} \]  

(3)

where \( V_{f_{a}} \) is a \( (N_{t}+P-1) \times (N_{t}+P-1) \) diagonal matrix with \( \{ V_{f_{a}} \}_{n,m} = e^{j2\pi f_{a}n} \) and \( X_{a} \) is a \( (N_{t}+P-1) \times P \) Toeplitz matrix with \( \{ X_{a} \}_{n,k} = x_{t}[n - k] \) where \( x_{t}[k] = 0 \) for \( k < 0 \) and \( k \geq N_{t} \). We set these samples in \( x_{t} \) to zero to avoid interblock-interference.

In the cooperation phase, the signal is defined as follows:

\[ y_{c} = V_{f_{sd}} X_{sd} h_{sd} + V_{f_{rd}} X_{rd} h_{rd} + w_{d} \in \mathbb{C}^{(N_{c}+P-1)\times 1} \]  

(4)

where we assume the frequency \( f_{sd} \) is constant over both phases. For each receiver \( b \) the noise is assumed to be a zero-mean circularly symmetric complex Gaussian random vector

\[ w_{b} \sim \mathcal{CN}(0,\sigma_{b}^{2}I). \]  

(5)

In the general case, the frequency offsets between nodes can take on any values within the Doppler spread of the system plus the frequency differences of the local oscillators. We assume
the maximum frequency offset is bounded and use this information to calculate the CRB and MAP frequency estimators. In the remainder of the paper, we assume the nodes are static and thus the signals have no Doppler spread. A statistical model for the frequency offset is used as prior information to aid in frequency estimation. Let the operating frequency of each node \( m \in \{r, s, d\} \) be modeled as

\[
f_m = f_0 + q_m
\]

where \( f_0 \) is the mean operating frequency and \( q_m \) is a random variable with mean zero and variance \( \sigma^2_m \). We assume the random variables \( q_m \) are independent. For this paper, we also assume \( \sigma^2_m = \sigma^2 \) for all nodes \( m \), which is an appropriate model when considering a group of identical nodes cooperating together. The frequency offsets to be estimated are the difference between two of these independent random variables and thus the frequencies, \( f_\alpha \) for \( \alpha \in \{sl, sr, rd\} \), have mean zero, variance \( 2\sigma^2 \), and are correlated.

III. LISTENING PHASE

In the listening phase, the destination and the relay receive the same signal through two different channels. We drop the subscript \( \alpha \) when considering only the single node-to-node link. To derive a good estimator for the frequency, it is useful to know the distribution of \( q_m \). However, this is not known, so it is reasonable to design an estimator based on the “worst case” distribution constrained to the known statistics, i.e., a mini-max estimator. As frequency estimation is inherently nonlinear, an asymptotic analysis is performed. Under this assumption, the variance of a MAP estimator asymptotes towards the CRB [23, Sec. 2.4.2]. In the remainder of this section, we show that a Gaussian distribution with mean zero and variance \( \sigma^2 \) for \( q_m \) maximizes the CRB of the frequency estimate over all distributions with the same mean and variance. We then derive the MAP estimator of \( f \). While both the CRB and the MAP estimators in this case are well known, we include them here for completeness and to introduce notation to be used later.

A. CRB

The unknown parameters in the single node-pair model (1) are \( f \) (which is modeled as a random variable with mean zero and variance \( 2\sigma^2_f \)) and \( h \) (which is considered deterministic).\(^1\) The CRB is defined to be the diagonal entries of the inverse Fisher Information Matrix (FIM). When one or more parameters are random variables, the FIM is expressed in the following form [23, Sec. 2.4.2]

\[
\mathbf{J}_\theta = \mathbf{E}_f(\mathbf{J}_{\theta|f}) + \mathbf{J}_f
\]

where the expectation is taken over the random variable \( f \)

\[
\mathbf{J}_{\theta|f} = -\mathbf{E}_w \left( \frac{\partial^2}{\partial \theta \partial \theta^T} L(\mathbf{y}|f) \right)
\]

where \( \mathbf{J}_f \) is the standard (nonrandom parameter) FIM, with expectation over the noise distribution, and \( L(\mathbf{y}|f) = (-1/\sigma^2) || \mathbf{y} - \mathbf{V}_f \mathbf{X} ||^2 \) is the log-likelihood of the data vector when the values of \( \mathbf{h} \) and \( f \) are held constant. The matrix \( \mathbf{J}_f \) is defined as follows:

\[
\mathbf{J}_f = -\mathbf{E}_f \left( \frac{\partial^2}{\partial \theta \partial \theta^T} L(\mathbf{y}|f) \right)
\]

where \( L(f) = \ln p(f) \) and \( p(f) \) is the distribution function of the random variable \( f \). For the parameter vector \( \theta^T = [f \mathbf{h}^T \mathbf{h}^T] \), the FIM has the following form [24]:

\[
\mathbf{J}_{\theta|f} = \begin{bmatrix} \Delta & \Lambda & \mathbf{A} \\ \Lambda^T & 0 & \Xi^T \\ \mathbf{A}^T & \Xi & 0 \end{bmatrix} \in \mathbb{C}^{(2P+1) \times (2P+1)}
\]

where \( \Delta \) is a scalar, \( \Lambda \in \mathbb{C}^{1 \times P} \), and \( \Xi \in \mathbb{C}^{P \times P} \). Let \( \mathbf{D}_f \) be a diagonal matrix with \( [\mathbf{D}_f]_{nn} = 2n - 1 - N_f \) such that

\[
\frac{\partial}{\partial f} \mathbf{V}_f = j\pi \mathbf{D}_f \mathbf{V}_f.
\]

The submatrices of (10) are computed as

\[
\Delta = \frac{2\pi^2}{\sigma^2} || \mathbf{D}_f \mathbf{X}^* \mathbf{X} ||^2, \quad \Lambda = -\frac{j\pi}{\sigma^2} \mathbf{h}^* \mathbf{X}^* \mathbf{D}_f \mathbf{X}, \quad \Xi = \frac{1}{\sigma^2} \mathbf{X}^* \mathbf{X}.
\]

None of these components depend on the random variable \( f \) and therefore the expectation in (7) goes away. The matrix \( \mathbf{J}_f \) is only nonzero in the first element and is

\[
[\mathbf{J}_f]_{11} = -\mathbf{E} \left( \frac{\partial L(f)}{\partial f^2} \right) \triangleq F_f
\]

where \( F_f \) is the Fisher information of the random variable \( f \) and \( L(f) \) is the log-likelihood of \( f \). The CRB for an estimator of \( f \) is then \( \left[ \mathbf{J}_f^{-1} \right]_{11} \), which can be calculated using the Shur complement\(^2\) to be

\[
C_f = \left( \frac{2\pi^2}{\sigma^2} || \mathbf{P}_f \mathbf{D}_f \mathbf{X}^* \mathbf{X} ||^2 + F_f \right)^{-1}
\]

where \( \mathbf{P}_f = \mathbf{I} - \mathbf{X}^* \mathbf{X} \mathbf{X}^* \mathbf{X} \) is the projection matrix onto the space orthogonal to the range of \( \mathbf{X} \). As the Fisher information is a positive number, it is clear that, to find the worst case (maximum) CRB, \( F_f \) must be minimized. We use the result from Appendix I which states that \( F_f \) is minimized when \( f \) is a Gaussian and thus (16) is maximized (over all distributions of \( f \) with variance \( 2\sigma^2_f \)) and is

\[
C_f = \left( \frac{2\pi^2}{\sigma^2} || \mathbf{P}_f \mathbf{D}_f \mathbf{X}^* \mathbf{X} ||^2 + \frac{1}{2\sigma^2_f} \right)^{-1}.
\]

\(^1\)The parameter \( \sigma_f \) is considered known as it is a property of the receiver hardware. Also, the noise variance \( \sigma^2 \) is uncoupled with the other parameters and is estimated separately with no penalty.

\(^2\)The \{1, 1\} block of a block matrix inverse is \([\mathbf{A}^{-1}]_{11} = (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} \).
B. MAP Estimator of Frequency

As a result of a Gaussian prior distribution on \( f \) maximizing (16), we use this distribution to calculate the MAP estimator. This choice of prior represents the least informative prior of all distributions with variance \( 2 \sigma_f^2 \) and mean zero. For a particular channel gain \( h \), the log-likelihood of the data is

\[
I(y, f) = \ln p(y, f) = \ln p(y|f) + \ln p(f) = -\frac{1}{\sigma_f^2} |y - X_f h|^2 + \frac{1}{4} \sigma_f^2 f^2.
\]

(18)

The apparent additional factor of two associated with \( \sigma_f^2 \) is due to the fact that \( f \) has a real Gaussian distribution as opposed to complex (as in the first term above). For any given frequency, the maximum of this expression over \( h \) is achieved when

\[
\hat{f}(y) = (X^* X)^{-1} X^* y.
\]

(19)

To find the MAP estimator of \( f \), we substitute (19) into (18) and minimize the negative

\[
\hat{f} = \arg\min_f \left\{ ||P_{X_f} y||^2 + \frac{\sigma_f^2}{4 \sigma_f^2} f^2 \right\}.
\]

(20)

We note that as \( \sigma_f \) goes to infinity (no prior information), the estimator (20) is the standard maximum likelihood frequency estimator [26].

IV. COOPERATION PHASE

In the cooperation phase, the destination node receives the superposition of signals coming from the source and relay. Each of these signals is transmitted with a slightly different frequency due to system imperfections. The purpose of this section is to derive a mini-max estimator for the two frequency offsets \( f_{sd} \) and \( f_{rd} \). The estimator is mini-max in the sense that we design the (asymptotically) minimum variance estimator given that the prior distribution on the frequencies maximizes the estimator variance. We show there exists an optimal transmit frequency for the relay, which reduces the variance of frequency estimation at the destination.

As the relay has an estimate of \( f_{sr} \) (which is correlated with \( f_{sd} \) and \( f_{rd} \)) this information is useful in reducing the variance of the estimate at the destination. We assume the frequency transmitted from the relay is adjusted according to the following rule:

\[
f_{rd} = f_{r,Tx} - f_{sr} = f_{r} - \gamma (f_{sr} + \epsilon_{sr}),
\]

(21)

where \( \gamma \) is a mixing parameter (mixing between the relay frequency and the relay’s estimate of the source frequency) and \( \epsilon_{sr} \) is the estimation error from the listening phase. We choose this rule as it is a linear function of the estimate and thus analytically tractable. When \( \gamma = 0 \), no frequency adjustment is made (e.g., when the estimate \( f_{sr} \) provides no information about the source’s frequency), and when \( \gamma = 1 \), the relay transmits its own estimate of the source’s frequency (thus trusting the estimate to provide all of the information available about the source’s frequency). We now express the frequency difference between the destination and the relay as

\[
f_{rd} = f_{rd}^{\text{opt}} + \epsilon_{sr} = f_{rd} - (1 - \gamma) f_{sr} + \gamma \epsilon_{sr}.
\]

(22)

The two frequencies to be estimated at the destination node are \( f_{sd} \) and \( f_{rd} \).

A. Covariance of Frequencies

Before calculating the MAP estimator of \( f_{sd} \) and \( f_{rd} \), we compute the least informative joint prior distribution for these random variables. First, the covariance matrix of these random variables is found by exploiting the relationship in (22) and then we show that the joint Gaussian distribution is the least informative prior.

To proceed, we calculate the covariance matrix of \( f_{sd}, f_{sr} \) and \( \epsilon_{sr} \), which is related to [see (22)] the covariance of \( f_{sd} \) and \( f_{rd} \). The mean of \( f_{sd} \) and \( f_{sr} \) are zero. To calculate the mean of \( \epsilon_{sr} \), let

\[
\xi = \arg\min_f \left\{ ||V_f P_{X_f} V_f^* y||^2 + \frac{\sigma_f^2}{4 \sigma_f^2} f^2 \right\}.
\]

(23)

Then \( \epsilon_{sr} = -f_{sr} + \xi \) and \( E(\epsilon_{sr}) = E[f_{sr} (E[\epsilon_{sr} | f_{sr}])] \). In Appendix II, it is shown for high SNR that \( E[\epsilon_{sr} | f_{sr}] = Q/(Q + K) f_{sr} \), where

\[
Q = \pi^2 ||P_{X_f} D_{r} X_f h_{sr}||^2,
\]

\[
K = \frac{\sigma_f^2}{4 \sigma_f^2}.
\]

Therefore

\[
E(\epsilon_{sr}) = E \left( \frac{Q}{Q + K} f_{sr} - f_{sr} \right) = 0.
\]

(26)

Note that \( \epsilon_{sr} \) is the estimator error for the frequency estimate at the relay in the listening phase (20). Equation (26) shows that (20) is an unbiased estimate. The covariance is now expressed as

\[
R_{f_{sd}, f_{sr}, \epsilon_{sr}} = \begin{bmatrix} f_{sd}^2 & f_{sd} f_{sr} & f_{sd} \epsilon_{sr} \\ f_{sd} f_{sr} & f_{sr}^2 & f_{sr} \epsilon_{sr} \\ f_{sd} \epsilon_{sr} & f_{sr} \epsilon_{sr} & \epsilon_{sr}^2 \end{bmatrix}
\]

(27)

where \( E(f_{sd}^2) = E(f_{sr}^2) = 2 \sigma_f^2 \) and \( E(f_{sd} f_{sr}) = \sigma_f^2 \), and using the result from Appendix II

\[
E(f_{sr} \epsilon_{sr}) = E(f_{sr} E(\epsilon_{sr} | f_{sr})) = \frac{-2K}{Q + K} \sigma_f^2
\]

(28)

and similarly \( E(\epsilon_{sr} \epsilon_{sr}) = (Q/(Q + K)) \sigma_f^2 \). To calculate \( E(\epsilon_{sr}^2) \), it can be shown that \( p(\epsilon_{sr} | y) \) approaches a Gaussian distribution for small \( \sigma_f \) and \( \sigma_r \) [27], the condition for efficiency of a MAP estimator [23, Sec. 2.4.2]. Therefore, for small frequency uncertainty and high SNR, \( E(\epsilon_{sr}^2) = (2K/(Q + K)) \sigma_f^2 \), which is the CRB (17). In summary

\[
R_{f_{sd}, f_{sr}, \epsilon_{sr}} = \sigma_f^2 \begin{bmatrix} 2 & 1 & 0-K \\ 1 & 2 & 0-K \\ 0-K & 0-K & 4K \end{bmatrix},
\]

(29)
With this covariance matrix calculated, the covariance of $f_{sd}$ and $f_{rd}$ is
\[
R_{f_{sd}f_{rd}} = \sigma_f^2 \begin{bmatrix}
1 & 1 & 0 & 0 \\
0 & \gamma & 1 & \gamma
\end{bmatrix}
\]
where the transformation matrix $T = \frac{2(1+\gamma)Q+K}{Q+K} \begin{bmatrix}
(1+\gamma)Q+K & 2(1-\gamma)Q+K \\
Q+K & 2(1-\gamma)Q+K
\end{bmatrix}$ is calculated from (22).

### B. CRB in Cooperative Phase

Recall the signal models for the cooperation phase (4) and the listening phase (3) as well as the relation between the two frequencies to be estimated $f_{rd}$ and $f_{sd}$ (22). The parameter vector for the unknown parameters is
\[
\theta = [f_{sd} f_{rd} h_{sd} h_{rd} s_{ld}]^T.
\]
For compactness, define $f = [f_{sd} f_{rd}]^T$. Recalling the notation developed in Section III, the deterministic FIM $(\mathbf{J}_{\theta f})$ is a $(2 + 6P) \times (2 + 6P)$ matrix with the structure of (10) where $\Delta \in \mathbb{C}^{2 \times 2}$, $\mathbf{A} \in \mathbb{C}^{2 \times 3P}$, and $\mathbf{E} \in \mathbb{C}^{3P \times 3P}$. Given the frequency random variables, the distributions of $y_c$ and $y_{sd}$ are independent and the joint distribution is written as
\[
p(y_c,y_{sd}|\theta,f) = p(y_c|f)p(y_{sd}|f)p(f)
\]
and the FIM is written as
\[
\mathbf{J}_{\theta} = \mathbf{J}_{\theta f}^{(c)} + \mathbf{J}_{\theta f}^{(md)} + \mathbf{J}_{f}.
\]

The blocks of the matrix $\mathbf{J}_{\theta f}^{(c)}$ are
\[
\Delta_{11,c} = \frac{2\pi^2}{\sigma_d^2} |D_c X_{sd} h_{sd}|^2
\]
(33)
\[
\Delta_{22,c} = \frac{2\pi^2}{\sigma_d^2} |D_c X_{rd} h_{rd}|^2
\]
(34)
\[
\Delta_{12,c} = \frac{2\pi^2}{\sigma_d^2} \{h_{sd}^* X_{sd} V_{f_{sd}}^* V_{f_{rd}}^* D_c^2 X_{rd} h_{rd}\}
\]
(35)
\[
\Xi_{11,c} = \frac{1}{\sigma_d^2} X_{sd}^* X_{sd}
\]
(36)
\[
\Xi_{22,c} = \frac{1}{\sigma_d^2} X_{rd}^* X_{rd}
\]
(37)
\[
\Xi_{12,c} = \Xi_{21,c} = \frac{1}{\sigma_d^2} X_{sd}^* X_{sd} V_{f_{sd}}^* V_{f_{rd}}^* D_c X_{rd}
\]
(38)
\[
\Lambda_{11,c} = \frac{-j\pi}{\sigma_d^2} h_{sd}^* X_{sd}^* D_c X_{sd}
\]
(39)
\[
\Lambda_{22,c} = \frac{-j\pi}{\sigma_d^2} h_{rd}^* X_{rd}^* D_c X_{rd}
\]
(40)
\[
\Lambda_{12,c} = \frac{-j\pi}{\sigma_d^2} h_{sd}^* X_{sd}^* V_{f_{sd}}^* V_{f_{rd}}^* D_c X_{rd}
\]
(41)
\[
\Lambda_{21,c} = \frac{-j\pi}{\sigma_d^2} h_{rd}^* X_{rd}^* V_{f_{sd}}^* V_{f_{rd}}^* D_c X_{sd}
\]
(42)

and zero for terms not listed. Note that $\Delta_{ik} \in \mathbb{C}^{1 \times 1}$, $\Xi_{ik} \in \mathbb{C}^{P \times P}$, and $\Lambda_{ik} \in \mathbb{C}^{1 \times P}$. The diagonal matrix $D_c$ is defined similar to $D_i$ in (11) with $N_c$ replacing $N_f$.

For data obtained during the listening phase, the matrix $\mathbf{J}_{\theta f}^{(md)}$ is
\[
\Xi_{11,\ell} = \frac{2\pi^2}{\sigma^2} |D_{\ell} X_{ld} h_{ld}|^2
\]
\[
\Xi_{33,\ell} = \frac{1}{\sigma^2} X_{ld}^* X_{ld}
\]
\[
\Lambda_{13,\ell} = \frac{-j\pi^2}{\sigma} h_{ld}^* X_{ld}^* D_{\ell} X_{ld}
\]
and zero for terms not listed. Note that for the submatrices (33)–(45), the structure is based on the ordering in the parameter vector $\theta = [f_{sd} f_{rd} h_{sd} h_{rd} h_{ld}]^T$.

To calculate $\mathbf{E}(\mathbf{J}_{\theta f})$, note that only the (1, 2) and (2, 1) cross terms of the submatrices above (i.e., $\Delta_{12}$, $\Xi_{33}$, $\Lambda_{13}$, . . .) are dependent on the frequencies. In each case, the dependency is of the form $\mathbf{AV}_{f_{sd}}^* \mathbf{V}_{f_{rd}} \mathbf{B}$ where $\mathbf{A}$ and $\mathbf{B}$ are deterministic matrices or vectors. Looking at the $n$th term of $\mathbf{V}_{f_{sd}}^* \mathbf{V}_{f_{rd}}$
\[
\mathbf{E}(\mathbf{V}_{f_{sd}}^* \mathbf{V}_{f_{rd}}) = \mathbf{E}(e^{j\pi d_n (f_{rd} - f_{sd})})
\]
(46)
where $d_n = 2n - 1 - N_c$. This expectation is just the characteristic function of the random variable $f_{rd} - f_{sd}$ evaluated at $\pi d_n$ (denoted as $\phi_{f_{rd} - f_{sd}}(\pi d_n)$). Let $\mathbf{M}$ be a diagonal matrix with $[\mathbf{M}]_{nn} = \phi_{f_{rd} - f_{sd}}(\pi d_n)$, then we replace $\mathbf{E}(\mathbf{V}_{f_{sd}}^* \mathbf{V}_{f_{rd}})$ with $\mathbf{M}$ in all cross terms of the FIM blocks. The FIM is then expressed as
\[
\mathbf{FIM} = \mathbf{E}(\mathbf{J}_{\theta f}) + \mathbf{J}_f
\]
(47)
where $\mathbf{J}_f$ is nonzero only in the upper left $2 \times 2$ block and this block is equal to $\mathbf{F}_f$, the Fisher information matrix of $f_{sd}$ and $f_{rd}$. Using the Shur complement of the upper left $2 \times 2$ block of (47), the CRB for the frequencies are the diagonal entries of
\[
\mathbf{C}_f = \left(\Delta - 2\Re\{\mathbf{AE}^{-1}\mathbf{A}^*\} + \mathbf{F}_f\right)^{-1}
\]
(48)

A careful examination of (48) reveals that this bound is a function of the phases of the channels. For node-pair $\alpha$, the two channels $\mathbf{h}_a$ and $e^{j\phi_a} \mathbf{h}_a$ result in different bounds (48) even though the channels only differ by a single phase. As the absolute phase of the signal at each node is hard to control and cannot be relied on to remain stable over time, we find the worst case CRB invariant to this phase. That is, for $\mathbf{h}_a = e^{j\phi} \mathbf{h}_a$, find $\phi$ maximizing the CRB (48). The resulting expression is
\[
\mathbf{C}_{f_{\text{max}}} = \left(\Delta - 2\Re\{\mathbf{AE}^{-1}\mathbf{A}^*\} + \mathbf{F}_f\right)^{-1}
\]
(49)
where $\Delta_{ii} = \Delta_{\phi}$ and
\[
\Delta_{12} = \Delta_{21} = \frac{2\pi^2}{\sigma_d^2} \Re\{h_{sd}^* X_{sd}^* D_c^2 X_{rd} h_{rd}\}
\]
(50)
As the diagonal terms of the matrices in (48) do not depend on $\phi$, the phase is chosen to maximize the magnitude of the off-diagonals of the matrix to be inverted in (48), which in turn maximizes the diagonals of the inverse (the negative signs are chosen for the off-diagonal terms because the FIM of the prior distribution, as calculated in the next section, also has negative off-diagonal terms). To see this, let $\mathbf{C}_f = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ (a Hermitian...
matrix), then \([C_F]_{11} = c/(a c - ||\hat{\theta}||^2)\). As \(a\) and \(c\) do not depend on the phase \(\phi\), the diagonals of \(C_F\) are maximized when \(||\theta||\) is maximized.

C. Distribution of Frequencies

We now desire to find the distribution of \(f_{sd}\) and \(f_{rd}\), which maximizes the CRB for a given frequency covariance \(R_F(30)\). In order to do this, we assume the training sequences are chosen to provide near optimal performance. Examining (49), an ideal set of training sequences would zero out the off-diagonal terms in \(\Delta\) and also zero out the \((\Delta \tilde{\Sigma}^{-1} \Delta^*\) term. Thus, for any constant modulus training sequences for which this is the case, the bounding CRB is

\[
C_{F,\text{opt}} = (\Delta_{\text{opt}} - F_F)^{-1}
\]  

(51)

where \(\Delta_{\text{opt}} = \text{diag}(|\Delta|)\). While it is not obvious that such training sequences can be designed, we performed exhaustive searches and show in Section VI, by simulation, sequences exist where (49) is close to (51). Under the assumption of a good set of sequences, the dependence on the distribution of \(f_{sd}\) and \(f_{rd}\) enters only through \(F_F\). We use the following lemma to find the distribution maximizing the CRB.

**Lemma 1:** For \(A, B,\) and \(C\) positive definite Hermitian matrices, if \(B > C\) (i.e., \(B - C\) is positive definite), then \((A + C)^{-1} - (A + B)^{-1}\) has positive diagonal entries.

**Proof:** By assumption, \((A + B) > (A + C)\), which implies \((A + C)^{-1} < (A + B)^{-1}\). Thus the difference of the matrices is positive definite Hermitian and therefore has positive diagonal elements.

To maximize the diagonal elements of the CRB (51), Lemma 1 implies \(F_F\) is as small as possible. Using the results of Appendix I, the Gaussian distribution satisfies this requirement and \(F_F = R_F^{-1}\). The assumptions that \(f_s, f_r, f_d\) and the estimation error from the listening phase \(e_{sr}\) are jointly Gaussian is therefore the least informative prior given the family of variances and correlations.

D. Optimal \(\gamma\)

With the aim of deriving a mini-max estimator, we desire to choose \(\gamma\) in (21) to minimize the trace of \(C_F\) (51). As this expression is not intuitive, it is helpful to consider a flat fading model. For flat fading, \(P = 1\) and the terms in the optimal CRB (51) are

\[
\Delta_{\text{opt}} = \begin{bmatrix}
\eta_c S_{sd_c} + \eta_r S_{sd}\nn_e S_{rd}
\end{bmatrix}
\]

(52)

and

\[
R_F = \begin{bmatrix}
2 & 2\eta_r(1 + \gamma) S_{sr} + \frac{\eta_c^2}{\eta_r S_{sr} + \frac{\eta_c^2}{\eta_r}} \\
2\eta_r(1 + \gamma) S_{sr} + \frac{\eta_c^2}{\eta_r} & 2\eta_r(1 - \gamma) S_{sr} + \frac{\eta_c^2}{\eta_r}
\end{bmatrix}
\]

(53)

where \(\eta_r = (2/3)\pi^2 N_r (N_F^2 - 1)\) (similarly for \(\eta_c\)) and \(S_{rd} = ||\hat{\theta}_{rd}||^2/\eta_r^2\) is the signal to noise ratio of the source-relay link (similarly for \(S_{sd_c}, S_{sd}, \) and \(S_{sr}\)).

An exact calculation of the optimal \(\gamma\) leads to a long, complicated expression that depends on the SNR of each link and the variance of the frequency oscillators. The expression is omitted here as it gives no insight into the problem. Later, we show there is minimal loss when \(\gamma\) is always set to 1. To gain some insight into the behavior of \(\gamma\), consider two limiting cases for \(\gamma_{\text{opt}}:\)

1) Large \(\sigma_f\), or no Prior Information: By taking the limit of the expression for \(\gamma_{\text{opt}}\) as \(\sigma_f \to \infty\), it is shown that \(\gamma_{\text{opt}} \to 1\). In this case, the relay transmits at a frequency equal to its estimate of the source frequency. By choosing this transmit frequency, the operating frequency of the relay \(f_r\) is removed from the estimation procedure as it contains no information about the source-destination frequency.

2) Small \(\sigma_f\): When \(\sigma_f \to 0\) (or when \(1/\sigma_f^2\) is much larger than any of the link SNRs perhaps due to poor channel SNRs), the limiting value of \(\gamma_{\text{opt}}\) is 1/2. However, in this regime, the estimate is dominated by prior information and the estimate \(f_{sr}\) is equal to its mean, which is zero. Therefore, this choice of \(\gamma\) (or any other choice) does not impact the relay transmit frequency (21) in this regime. Thus, choosing \(\gamma = 1\) does not incur any performance penalty.

As an example of the function \(\gamma_{\text{opt}}\), Fig. 2 shows plots of several curves of \(\gamma_{\text{opt}}\) versus \(\sigma_f^2\). The length of the training signal is \(N_t = N_c = 16\) and the SNR of the source-destination link is \(-3\) dB (combining the listening and cooperation phases, the effective SNR is 0 dB). The SNR from relay to destination, \(S_{rd}\), is also 0 dB and there is one curve each for \(S_{sr} \in \{-10, 0, 10\}\) dB. For each curve, the transition from \(\gamma_{\text{opt}} = 1\) to \(\gamma_{\text{opt}} = 1/2\) appears to occur roughly when \(\sigma_f^2 \approx (\eta_c S_{sd_c} + \eta_r S_{sd})/(\eta_c^2 S_{sd_c} + \eta_r^2 S_{sd})\). These values of \(\sigma_f^2\) are significant because, for example, when \(\sigma_f^2 < \eta_c S_{sd_c}\) (left half of the plot), the assumed prior knowledge of frequency has more weight than the data, whereas when \(\sigma_f^2 > \eta_c S_{sd_c}\), the information in the data is more important than the prior model.

E. MAP Estimator of \(f_{sd}\) and \(f_{rd}\)

To calculate the MAP estimate of \(f_{sd}\) and \(f_{rd}\) at the destination node during the cooperation phase, the covariance between

![Fig. 2. Plot of optimal \(\gamma\) as a function of modeled frequency variation. Three curves are shown for different values of gain between the source and relay. The SNR of the source-destination and relay-destination node pairs are held constant at 0 dB.](image-url)
these two random variables (30) is needed. Therefore, the values of $Q$ and $K$ need to be forwarded to the destination node. The log-likelihood of the data at the destination is

$$L(y_c, y_{sdl}, f) = \ln p(y_c | f) + \ln p(y_{sdl} | f) + p(f)$$

$$\propto -\frac{1}{\sigma_d^2} ||y_{sdl} - V_{f,sdl} x \hat{h}_{sdl}||^2 - \frac{1}{\sigma_d^2} ||y_c - X(f) g||^2 - \frac{1}{2} f^T R_f^{-1} f$$  (54)

where $X(f) \triangleq [V_{f,sdl} \hat{h}_{sdl} X_{f,sdl}^r]$ and $g^T = [h_{sdl}^T, h_{f,sdl}^T]$. As before, we choose estimates for $g$ and $\hat{h}_{sdl}$ to maximize the log-likelihood for any given frequency pair

$$\hat{g}(f) = (X^*_c X_c)^{-1} X^*_c y_c$$  (55)

$$\hat{h}_{sdl}(f) = (X^*_f X_f)^{-1} X^*_f V^*_f y_{sdl}.$$  (56)

Substituting these estimates into (54) and minimizing the negative log-likelihood to obtain the MAP frequency estimator

$$\hat{f} = \arg \min_f \left\{ ||P_{\hat{h}_{sdl}} V_{f,sdl} y_{sdl}||^2 + ||P_{\hat{g}} V_{f,sdl} y_{sdl}||^2 + \frac{\sigma_d^2}{2} f^T R_f^{-1} f \right\}.$$  (57)

We note the special case of $\sigma_f \to \infty$ (which implies $\gamma_{opt} = 1$). For $\gamma = 1$, the covariance (30) needed in the MAP estimator simplifies to

$$R_{f,sdl,f,sdl} = \sigma_f^2 \left[ \begin{array}{c} 2Q + K \\ 2Q + K \end{array} \right]$$  (58)

which has a finite inverse when $\sigma_f < \infty$. However, when $\sigma_f \to \infty$, we evaluate the limit of $R_f^{-1}$ resulting in

$$\lim_{\sigma_f \to \infty} R_f = \zeta^T \frac{2Q}{\sigma_f^2} = \zeta^T \frac{2\pi^2}{\sigma_f^2} ||P_{\hat{h}_{sdl}} D_{sdl} h_{sdl}||^2$$  (59)

where $\zeta = [1 - 1]$ and $C_{f,sdl}$ is the CRB of the frequency in the source-relay link (17) with $\sigma_f = \infty$. The penalty term (last term) of the MAP estimator (57) simplifies to

$$\frac{\sigma_d^2}{2} f^T R_f^{-1} f \to \frac{\sigma_d^2}{2C_{f,sdl}} (f - \hat{f})^2.$$  (60)

Thus, the penalty term is a quadratic of the frequency difference term normalized by the ratio of error variances (noise power over frequency estimation error variance). Note that this quadratic penalty term [and the more general term in (57)] is due to assuming a Gaussian prior distribution on the frequency offsets. We justify the use of the Gaussian prior because it is the least informative prior in the sense of Fisher information as this distribution maximizes the CRB when constraining the covariance.

V. SUB-OPTIMAL ALGORITHMS

The MAP frequency estimator (57) requires a two-dimensional search over the frequency range of interest. As this is a computationally expensive approach to estimation, we derive a suboptimal implementation of the MAP estimator and introduce a correlation based estimator which does not require a search over frequency.

A. One-Dimensional MAP

Through judicious choice of training sequence, estimation of the two frequencies is nearly uncoupled (an example is shown later). Therefore, performing two independent one-dimensional MAP searches for the frequencies is approximately the same as performing the full two-dimensional search as required by the MAP algorithm. Given the data vector $y_c$, the one-dimensional MAP estimates of the frequencies are

$$\tilde{f}_{sl} = \arg \min_f \left\{ ||P_{\hat{h}_{sdl}} V_{f,sdl} y_{sdl}||^2 + \frac{\sigma_d^2}{4\sigma_f^2} f^2 \right\}$$  (61)

$$\tilde{f}_{rl} = \arg \min_f \left\{ ||P_{\hat{g}} V_{f,rl} y_{rel}||^2 + ||P_{\hat{h}_{sdl}} V_{f,sdl} y_{sdl}||^2 + \frac{\sigma_d^2}{4\sigma_f^2} f^2 \right\}$$  (62)

which do not take the correlations between the frequencies into account. To improve the estimates (61) and (62), we assume the variance of each estimate meets the CRB assuming the prior information is uncorrelated for each frequency

$$\tilde{C}_f = \left( \Delta - 2R[\Lambda \Lambda^T] + \text{diag}(R_f)^{-1} \right)^{-1}$$  (63)

where $\text{diag}(R_f)$ is a diagonal matrix consisting of the diagonal entries of $R_f$ (zeroing out the other elements). This assumption is valid for high SNR. Incorporating this knowledge with the prior information, the least squares estimates of the frequencies are

$$\begin{bmatrix} \tilde{f}_{sdl,MAP} \\ \tilde{f}_{f,rl,MAP} \end{bmatrix} = R_f (R_f + \tilde{C}_f)^{-1} \begin{bmatrix} \tilde{f}_{sdl} \\ \tilde{f}_{rl} \end{bmatrix}.$$  (64)

B. Correlation Method

We first describe a standard correlation frequency estimation method as presented in [28] and then provide an extension to allow this algorithm to work in the presence of two signals with known training sequences. Assuming a single signal in the presence of flat fading

$$y[n] = e^{j2\pi f_0 n} x[n] + w[n], \quad 1 \leq n \leq N.$$  (65)

The estimated autocorrelation sequence of $y[n]$ is

$$R[k] = \frac{1}{N-k} \sum_{i=k+1}^{N} \langle y[i] \overline{y[i]} \rangle \langle y[i-k] x[i-k] \rangle,$$  (66)

The estimate of the frequency is calculated as

$$\hat{f} = \frac{1}{\pi(M+1)} \arg \left\{ \sum_{k=1}^{M} R[k] \right\}.$$  (67)
where $M$ is a design parameter and the frequency estimate is unambiguous if
\begin{equation}
|f| < \frac{1}{M+1}.
\end{equation}
(68)

Therefore, $M$ trades performance for estimation range. The performance of this algorithm (67) is shown in (28) to be close to the CRB when $M = N/2$. To ensure adequate estimation range, the maximum allowed value of $M$ is 12 (corresponding to a range of five standard deviations away from the mean of the prior). To incorporate the known prior knowledge of the frequency variance, the estimate (67) is adjusted according to the following rule:
\begin{equation}
\hat{f}_p = \frac{\sigma_f^2}{\sigma_f^2 + c_f^2} \hat{f}.
\end{equation}
(69)

where $c_f^2$ is the CRB of the frequency estimate with no prior information. Let
\begin{equation}
\hat{f} = \rho(y, x, \sigma_f)
\end{equation}
be a function that inputs the data vector $y$, training vector $x$, and prior information, and outputs the frequency estimate according to the above algorithm. This algorithm is used without modification during the listening phase to calculate the estimate $\hat{f}_{sr} = \rho(y_{sr}, x, \sigma_f)$.

For the cooperation phase, there are two signals present and the undesired signal acts as interference for the desired signal being estimated. The estimates provided by the correlation algorithm are
\begin{equation}
\hat{f}_{sd1} = \rho(y_c, x_{sd1}, \sigma_f)
\end{equation}
(71)
\begin{equation}
\hat{f}_{rd1} = \rho(y_c, x_{rd1}, \sigma_f)
\end{equation}
(72)

which exhibit a floor in mean squared error (MSE) (see Fig. 5). To improve the estimates, we project out the undesired signal in the following manner:
\begin{equation}
\tilde{y}_{cd} = \mathbf{P}_{\hat{y}_{rd}} x_{rd} y_c
\end{equation}
(73)
\begin{equation}
\tilde{y}_{cd} = \mathbf{P}_{\hat{y}_{ld}} x_{ld} y_c
\end{equation}
(74)

where the frequency estimates in (71) and (72) are used to calculate the interference signal, which is projected out. The correlation algorithm is run a second time to find
\begin{equation}
\hat{f}_{sd2} = \rho(\tilde{y}_{sd}, x_{sd2}, \sigma_f)
\end{equation}
(75)
\begin{equation}
\hat{f}_{rd2} = \rho(\tilde{y}_{rd}, x_{rd2}, \sigma_f).
\end{equation}
(76)

The final frequency estimates, with all prior information accounted for, is calculated similarly to (64)
\begin{equation}
\begin{bmatrix}
\hat{f}_{sd, corr} \\
\hat{f}_{rd, corr}
\end{bmatrix} = \mathbf{R}_f (\mathbf{R}_f + \hat{\mathbf{C}}_f)^{-1}
\begin{bmatrix}
\hat{f}_{sd2} \\
\hat{f}_{rd2}
\end{bmatrix},
\end{equation}
(77)

VI. SIMULATIONS

In Section IV, we showed the optimal $\gamma$ for extreme values of $\sigma_f$ is either 1 or 1/2 and when $\gamma_{opt}$ approaches 1/2, its effect is small because the frequency adjustment is going toward zero. In this section, we show by simulation, the penalty for choosing $\gamma = 1$ instead of $\gamma = \gamma_{opt}$ is usually limited to a few tenths of a decibel. Thus, near optimal performance is achieved without communicating any of the link SNRs back to the relay for calculation of $\gamma_{opt}$. We also show the existence of training sequences where (49) is close to (51). We next show the benefit of letting the relay set its transmit frequency based on information received during the listening phase and finally compare the performance of each estimation algorithm with the CRB.

We simulate a three node system in a frequency flat environment. In all simulations, we use the SNR of the $sd$ link (assuming $S_{sd} = S_{rd}$) as a reference value. The following configuration is considered: use $S_{rd} = S_{sd}$ and then vary the link SNR of the source-relay link relative to $S_{sd}$. Let $N_c = N_r$. The prior distribution for the operating frequency we assume is Gaussian with a standard deviation of 0.01 times the sample rate (e.g., a ±5 part-per-million standard deviation of a local oscillator at 5 GHz with 5 MHz sample rate).

For flat fading channels and constant modulus training sequences, it is sufficient to choose $x_c = 1$ (the vector of all ones) and $x_{rd} = 1$. A search is performed to find $x_{rd}$ which minimizes the CRB (49). For values of $N_r \in \{4, 8, 16\}$ an exhaustive search over all binary sequences is performed (results hold independent of choice between $\gamma = \gamma_{opt}$ or $\gamma = 1$) and for values of $N_r > 16$, a randomized search over binary sequences is performed. For each value of $N_r$ (up to 128) it is observed that the optimal sequence for $x_{rd}$ has the following structure.

A. Sequence Design

Let $a_1 = [1, -1]^T$
\begin{equation}
a_n^T = [a_{n-1}^T, -a_{n-1}^T]
\end{equation}
(78)

where $a_n$ is length $2^n$ and is the last column of a Sylvester matrix. Then the length $N_c = 2^n$ optimal sequence is
\begin{equation}
x_{rd, opt} = \begin{bmatrix}
a_{n-1} \\
J a_{n-1}
\end{bmatrix}
\end{equation}
(79)

where $J$ is the exchange matrix which reverses the order of elements in the vector it multiplies. A more general approach to training sequence design is addressed in [22].

For the configuration described above, and with $S_{sr} = 10 S_{sd}$, Fig. 3 shows the difference between the best possible CRB (51) (for any constant modulus sequence and $\gamma = \gamma_{opt}$) and the worst case CRB (49) using the binary sequence shown above and $\gamma = 1$. The 0.6 dB difference for $N_r = 4$ is primarily due to a non-optimal sequence $x_{rd}$, whereas the 0.2 dB difference for other values of $N_r$ is due to choosing $\gamma = 1$ instead of the optimal value. The loss in performance due to a non-optimal sequence decreases dramatically as $N_r$ increases. These loss values are typical of other system configurations as well. The system behavior as a function of training sequence
Fig. 3. Plot of the loss in performance caused by binary training sequence as
opposed to an arbitrary sequence, and when choosing $\gamma = 1$ versus $\gamma = \gamma_{\text{opt}}$.
Relay-destination and source-destination SNRs are the same and source-relay
SNR is 10 dB higher.

Fig. 4. Plot of the sum of Cramér-Rao Bounds for $f_{\text{sd}}$ and $f_{r\text{d}}$. Circle and
"x"-marks show bound when $\sigma_f = 0.01$ and $\gamma = \gamma_{\text{opt}}$, plus marks show
bound when $\gamma = 0$, and triangles show bound when $\gamma = 0$ and $\sigma_f = \infty$ (the
standard frequency bound assuming no prior information). All curves are for a
length 16 training sequence.

Fig. 5. Plot of mean squared error of nonadaptive (circles) and adaptive two-
step (triangles) correlation algorithms. The mean squared error is compared with
the CRB.

Fig. 6. Plot of mean squared error of full (two-dimensional search) MAP (cir-
cles), one-dimensional MAP ("x"-marks), and adaptive correlation (triangles).
The mean squared error is compared with the CRB.

illustrates the fact that the CRB is insensitive to the selection
of these sequences.

Fig. 4 shows the sum of the CRB (49) for the two frequencies
estimated at the destination node as a function of $S_{\text{sd},c}$. For this
figure, the SNRs of the source-destination link ($S_{\text{sd}}$) and relay-
destination link ($S_{r\text{d}}$) are the same. The circle and "x"-marks show the CRB when the SNR of the source-relay link $S_{\text{sr}}$ is,
respectively, the same as and 10 dB higher than $S_{\text{sd}}$. The plus
marks show the CRB when $\gamma = 0$. The difference between
the plus-marks and the circle or "x"-marks show the potential
gain in estimation performance by changing the relay's transmit
frequency (greater benefit when the SNR is large). The triangles
show the CRB when no prior information is used. This shows a
great advantage of using a prior model (curves with plus, circle,
and "x"-marks) when the SNR is low.

For the case when the source-relay link is 10 times better
than the source-destination link ($S_{\text{sr}} = 10S_{\text{sd},c}$), we illustrate
the performance of each algorithm as compared with the CRB
(49). We first demonstrate the performance of the correlation
algorithm and then compare it with the MAP algorithms. Fig. 5
shows the total MSE (summation of errors from $f_{\text{sd}}$ and $f_{r\text{d}}$) of
the correlation algorithm compared with the CRB for $N_c = 16$.
The triangle markers denote the performance of the algorithm
without any adaptation, (71) and (72), while the circle markers
denote the performance of the adaptive two-step algorithm (77).
For lower SNRs, the adaptive algorithm has about a 3 dB advan-
tage while the performance difference is much greater at higher
SNRs (above 15 dB). The performance of the adaptive algorithm
is near optimal. The slight "bump" in performance of the two al-
algorithms at $S_{\text{sd}} = -10$ dB SNR is caused by the interaction of
the threshold region (the region where the MSE performance
breaks away from the CRB) and the region dominated by prior
information (where the algorithms converge to $\sim 34$ dB MSE
relative to the sample rate).

For the same scenario, Fig. 6 compares the three estimation
algorithms: full two-dimensional search MAP (57) (circles),
one-dimensional MAP (64) ("x"-marks), and the adaptive correlation algorithm (77) (triangles). Each of these algorithms is close to the CRB for high SNR. Note that $\sigma_f$ (the frequency uncertainty) is small enough to achieve asymptotic efficiency here. The differences in behavior at lower SNRs is attributed to the different algorithms entering their threshold regions at different SNRs. A more detailed analysis of this region can be carried out using the methods of [29]. We note here that the best compromise between estimation performance and computational efficiency is provided by the two-step correlation algorithm.

VII. CONCLUSIONS

In this paper, we have detailed a practical approach to frequency offset estimation in a three-node collaborative communications system. By deriving the Cramér-Rao bounds for this system we have shown that our approach yields performance quite close to the derived bound; thus leaving no room for improvement upon these algorithms. We have shown through simulations, the performance increase obtained by allowing the relay to change its transmitting frequency. We have also shown that there exists an optimal transmit frequency for the relay node based on the other link SNRs and the assumed prior knowledge of the frequency offsets. However, there is only a small (tenths of decibels) penalty if the relay always transmits at its estimate of the source frequency. Simulation results also demonstrate the existence of binary training sequences that result in very little loss in performance. As compared with an arbitrary constant modulus sequence.

APPENDIX I

MINIMUM FISHER INFORMATION DISTRIBUTION

In this appendix, we find the distribution with mean zero and given covariance that minimizes the Fisher information. Let $p_R(z)$ represent the family of distributions with mean zero and covariance $R$. Let $z$ be a random vector distributed as $p_R(z)$ and the FIM is defined as

$$F_z \triangleq -E \left( \frac{\partial^2 \ln p_R(z)}{\partial z \partial z^T} \right).$$

(80)

We define the ordering of Hermitian matrices as follows: $A \succeq B$ implies that $(A - B)$ is positive semidefinite. We show next that the distribution (with given covariance $R$) that minimizes the FIM of $z$ is

$$p_R(z) = \mathcal{N}(0, R).$$

(81)

To show this, consider the following experiment: without any data, design an estimator $\hat{z}$ for the random variable $z$. If $z = 0$, then this estimator is unbiased and its covariance is $R$. By the Cramér-Rao Theorem

$$\text{cov}(\hat{z}) \triangleq R \geq F_z^{-1}.$$

(82)

Therefore, $F_z \geq R^{-1}$ with equality being achieved when $z \sim \mathcal{N}(0, R)$ because the FIM of a Gaussian random vector is equal to $R^{-1}$.

APPENDIX II

CONDITIONAL MEAN OF ESTIMATION ERROR

Here we derive the expression for $E_{f_{sr}}[\xi|f_{sr}]$ ($\xi$ being defined in (23)) in the asymptotic or high SNR regime. We replace $y$ with its mean and obtain

$$E(\xi|f_{sr}) \approx \arg\min_f \left\{ ||V_f P_{X_f}^\perp V_f^* X_{f_{sr}} ||^2 + \frac{\sigma_f^2}{4\eta_f} f^2 \right\}$$

(83)

where the approximation is exact in the limit $\sigma_f^2 \to 0$. We perform the change of variables $f_{sr} = 0$ and $\tilde{f} = f - f_{sr}$, therefore, $V_{f_{sr}} = I$. The first term in (83) is

$$h_{sr}^* X_{f_{sr}}^* V_f P_{X_f}^\perp V_f^* X_{f_{sr}} h_{sr}$$

(84)

which is greater than or equal to zero and only equal to zero when $f = 0$ (i.e., $f = f_{sr}$). This function is thus locally convex about the point $f = f_{sr}$ and the linear term vanishes. The second-order Taylor series approximation is then

$$\frac{\pi^2}{Q} |P_{X_f}^\perp X_{f_{sr}} h_{sr}|^2 f^2,$$

(85)

The value $Q$ can be considered the effective signal power including all systematic and estimation gains. Returning to (83)

$$E(\xi|f_{sr}) \approx \arg\min_f \left\{ Q \cdot (f - f_{sr})^2 + \frac{\sigma_f^2}{4\eta_f} f^2 \right\}$$

$$= \frac{Q}{Q + K f_{sr}}.$$
Peter A. Parker received the B.S.E.E. and M.S. degrees in electrical engineering from Brigham Young University, Provo, UT, both in 2001 and the Ph.D. degree in engineering sciences from Harvard University, Cambridge, MA, in 2008. Since 2001, he has been a staff member with Lincoln Laboratory, Massachusetts Institute of Technology (MIT), where he has worked on adaptive signal processing algorithms for multiantenna radar and communication systems. While at Harvard University (2004–2008), his research focused on information theoretic bounds and signal processing techniques for cooperative communication systems and multiantenna communication systems.

Patrick Mitran (S’01–M’07) received the B.S. and M.S. degrees in electrical engineering, in 2001 and 2002, respectively, from McGill University, Montreal, Canada, and the Ph.D. degree from the Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA, in 2006. In 2005, he interned as a research scientist for Intel Corporation in the Radio Communications Lab. During 2006–2007, he was an applied mathematics lecturer with the School of Engineering and Applied Sciences, Harvard University. Since fall 2007, he has been with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada, as an Assistant Professor. Currently he is interested in the study of cooperation and cognition in wireless networks both from an information theoretical viewpoint, as well as coding theory and signal processing perspectives.

Daniel W. Bliss received the B.S.E.E. degree in electrical engineering from Arizona State University, Tempe, in 1989. He received the M.S. and Ph.D. degrees in physics from the University of California at San Diego in 1995 and 1997, respectively. Since 1997, he has been with Lincoln Laboratory, Massachusetts Institute of Technology (MIT), where he is currently working with the Advanced Sensor Techniques group. He focuses on multiantenna adaptive signal processing and performance bounds, primarily for communication systems. He was with General Dynamics (1989–1991), where he designed avionics for the Atlas-Centaur launch vehicle and performed research and development of fault-tolerant avionics. As a member of the Superconducting Magnet Group at General Dynamics (1991–1993), he performed magnetic field calculations and optimization for high-energy particle-accelerator superconducting magnets. His doctoral degree work (1993–1997) was in the area of high-energy particle physics, searching for bound states of gluons, studying the two-photon production of hadronic final states, and investigating innovative techniques for lattice-gauge-theory calculations. His current research topics include channel phenomenology, space-time coding, information-theoretic bounds for MIMO communication systems, algorithm development for multichannel multiserter detectors (MCMUD), and multiple-input multiple-output (MIMO) radar concepts.