Multiple Frequency Offset Estimation for the Downlink of Coordinated MIMO Systems

Brad W. Zarikoff*, James K. Cavers†

School of Engineering Science, Faculty of Applied Sciences
Simon Fraser University
8888 University Drive
Burnaby, BC, Canada V5A 1S6
{*bwzariko@sfu.ca, †cavers@sfu.ca}

Abstract

We consider downlink MIMO beamforming from several coordinated basestations (BSs), and the associated problem of independent carrier frequency offsets (CFOs) at the BSs which cause accumulated phase errors to compromise beamforming accuracy. Correction of the CFOs requires estimation of their values, so our topic is multiple CFO estimation, a little-explored area. We present a robust and easily generalized estimator that accounts for the training sequence (TS) correlations caused by the CFOs, and show that it meets the Cramer-Rao lower bound (CRLB) at moderate signal-to-noise ratios (SNRs). The performance of the estimator is contingent upon TSs short enough to ensure convexity of the log-likelihood over the allowable CFO ranges. For combinations of TS length and CFO range that violate this constraint, we present two suboptimal estimators based on segmentation of the TS, both of which also meet the CRLB at moderate to high SNRs.

Submitted to the IEEE Journal on Selected Areas in Communications: MIMO Systems and Applications. This work was supported by a National Sciences and Engineering Research Council (NSERC) of Canada Post Graduate Scholarship. An abbreviated version has been submitted to Globecom 2007.
Multiple Frequency Offset Estimation for the Downlink of Coordinated MIMO Systems

I. INTRODUCTION

Modern wireless communications systems use the spatial domain to increase data throughput. For point-to-point links, the theory of multiple input/multiple output (MIMO) systems is already well-established. The next big step will be coordinated multicellular MIMO systems, in which multiple MIMO base stations (BSs) cooperate to form, in effect, a larger MIMO system. The desired effect is mitigation of shadowing and path loss, as well as increased trunking efficiency, multiuser diversity and a general reduction in radiated power.

We consider the downlink of a coordinated system, which is essentially a generalization on the MIMO beamformer. A coordinated downlink must share transmit data channel state information (CSI) within each group of BSs, and the CSI must be accurate in order for beamforming to be effective. Unfortunately, each BS has an independent local oscillator (LO) with its own carrier frequency offset (CFO). The CFOs can cause phase errors, which in turn degrade the performance of the beamformer over the duration of a frame [1]. A first step to the practical issue of correcting these phase errors is to estimate the values of each CFO in the multibase system, which is the topic of this paper.

While estimation at a mobile subscriber (MS) of the CFO of a single BS is a classical and well-understood problem, there have been few estimators designed to handle the CFOs of multiple BSs, and none of them is adequate for use in coordinated systems. The focus of this work is on the development of CFO estimation techniques for coordinated systems. The estimators presented in this paper make use of training sequences (TSs) transmitted from every antenna of every cooperating BS, so that a MS can estimate both the CSI and the CFOs. Compared to existing estimators, they provide more accurate estimation and/or lower peak power requirement, and they are capable of operating in realistic conditions of asynchronicity and delay spread. They also have a complexity much lower than the brute force search. Several approaches to low complexity are presented and their performances compared.

The largest hurdle to estimating multiple CFOs is the potential for the training sequences to
lose their desired orthogonality during transmission and become correlated. Existing estimators try to deal with this phenomenon by either avoiding or ignoring the correlation, and thus are limited in their applications. Ahmed et al. [2] avoids accounting for this correlation completely, and is thus limited to short, high power training sequences. Besson and Stoica [3] and Oh et al. [4] transmit the TSs at different times, avoiding the correlation problem altogether, reducing the estimation to a sequence of single CFO estimates. However, this means that the sequences must transmit with a high peak power. Moreover, the method of [3] relies on symbol-wise interleaving of sequences from different BSs to obtain orthogonality, and is therefore limited to symbol-synchronous reception in a frequency-flat channel. The newly devised techniques in this paper account for the correlation, and thus allow for simultaneous transmission of the TSs. Because of this, they can operate at a reduced peak power, yet still provide better estimation accuracy because of their greater duration.

The layout of the paper is as follows. In Section II, the system model is detailed. In Section III, a few existing estimators are reviewed. In Section IV, the proposed Newton estimator is derived. In Section V, the special case of large CFOs is solved using an approach that breaks up the TS into smaller segments. Section VI derives the Cramer-Rao bound for the coordinated system and Section VII provides a complexity comparison of the techniques. Then in Section VIII, simulations are provided to show the performance of the estimators proposed in Sections IV and V.

II. SYSTEM MODEL

The purpose of the coordinated MIMO downlink is to transmit independent data to $K$ MSs. Each MS uses an $N_R$-element antenna array. The coordinated transmitter consists of $B$ base-stations (BS), each with an $N_T$-element antenna array. Fig. 1 depicts this model, which is intentionally simplistic in order to keep the effects of CFO easily recognizable and to demonstrate the effects on coordinated systems. Our focus is on the accurate estimation of the multiple CFOs.

The radio link, or channel, from the $m^{th}$ antenna of the $b^{th}$ BS to the $q^{th}$ antenna of the $k^{th}$ MS has a random gain of $H_{kq,bm} \in \mathbb{C}$, which we consider to be static over the estimation period. We collect the gains into a matrix $H_{k,b} \in \mathbb{C}^{N_R \times N_T}$ from the $b^{th}$ BS to the $k^{th}$ MS. Although our maximum likelihood (ML) estimation does not require a statistical model for these gains, normally we will consider them to be random with second order moments that reflect path loss,
shadowing and fading, with all antennas of a given BS-MS pair having the same second moment. In our initial discussion and in most of the analysis, we assume symbol-synchronous reception with no delay spread. However, we do show a straightforward generalization that accommodates these two phenomena.

The received signal is downconverted to baseband and matched filtered. Due to the independent LOs, there exists a CFO at each BS, denoted as $c_b$ for the $b^{th}$ BS. For a CFO of $f_b$ Hz and a symbol period of $T_s$ seconds, $c_b$ is normalized to the symbol period, so $c_b = 2\pi f_b T_s$. Note that every transmit antenna at a given BS will be affected by the same CFO.

The estimation schemes covered in this paper are data-aided and require training sequences (TS). For this, we use an $N$-symbol TS for each of the $BN_T$ transmit antennas, concatenated column-wise in $X \in \mathbb{C}^{N \times BN_T}$. Since the TSs span $N$ symbols, the expression for the received samples will be vectorized appropriately. First, the rotational effect of the CFO will be imposed onto the TS matrix. The TS matrix $X$ can be expressed as the concatenation of $B$ matrices, $X = [X_1 \cdots X_B]$, one for each BS, where $X_b \in \mathbb{R}^{N \times N_T}$ denotes the sequences for the $b^{th}$ BS. The TS matrix $X$ then has each element rotated by the CFOs, or more specifically, each element is rotated by the phase shift caused by the CFOs at symbol time $n$. As a function of the CFO vector $c$, we will denote the rotated TS matrix as $V$

$$V(c) = [X_1 \odot \Omega_1(c_1) X_2 \odot \Omega_2(c_2) \cdots X_B \odot \Omega_B(c_B)]$$

$$= [V_1(c_1) V_2(c_2) \cdots V_B(c_B)]$$

$$= X \odot \Omega(c), \quad (1)$$

where the operator $\odot$ refers to the component-wise product. This expression denotes the element-by-element product of the synchronization sequences and a matrix of CFO-induced symbol rotations, where the matrix $\Omega_b(c_b) \in \mathbb{C}^{N \times N_T}$ is constructed with all $N_T$ columns equal to the same column vector $\omega_b(c_b) = [1 e^{j c_b} \cdots e^{j c_b(N-1)}]^T \in \mathbb{C}^{N \times 1}$.

The received samples for the $k^{th}$ MS can now be represented in a length-$NN_R$ vector

$$r_k = \mathbf{V}(c)h_k + n_k, \quad (2)$$

where $\mathbf{V}(c) = I_{N_R} \otimes V(c) \in \mathbb{C}^{NN_R \times BN_RN_T}$ is a Kronecker product referring to the block diagonalization of $N_R$ copies of $V(c)$, $h_k \in \mathbb{C}^{BN_RN_T \times 1}$ is the vector of channel gains created
by stacking the columns of $H_k$ where $H_k = [H_{k,1}, \cdots H_{k,B}]$, and $n_k \in \mathbb{C}^{N_R \times 1}$ is the vector obtained by stacking the $N_R$ additive white Gaussian noise sequences.

It is worth mentioning that the MSs themselves have independent oscillators. The BS CFOs estimated at a MS are therefore offsets relative to that MSs LO frequency. We will drop the subscript $k$ from here on, and will focus on the estimator performance at a single MS.

In the rest of this paper, for a matrix $A$, $A_{i,j}$ is a scalar that corresponds to the $(i^{th}, j^{th})$ entry, and $A^{<i>}$ is a vector that corresponds to the $i^{th}$ column. The matrix $I_N$ refers to an $N \times N$ identity matrix.

A. Effect of Carrier Frequency Offset

After channel estimation, a downlink beamformer is used to transmit data symbols to $K$ MSs in this coordinated system. The details of these beamformers will not be covered in this paper, but it is important to keep the target application of this system in mind. The beamformer will use the channel estimates to create a number of parallel channels to the MSs. If the channel estimates are accurate, the receive samples at each MS will be free of inter-user or inter-symbol interference.

Techniques exist that allow for accurate channel estimation in the presence of CFO [5]. However, with a large CFO, the progressive rotations can be seen as a perturbation on the CSI over the course of a data frame. These rotations degrade the orthogonality of the transmit and receive correlators [1].

While the accuracy of the LOs can be quite high, the normalized CFO can be quite large, especially in the GHz frequency bands. The accuracy of most LOs are in the range from 0.2 ppm up to 10 ppm, where ppm refers to parts per million. Assuming a 5 GHz communications system, this means that $f_b$ will range from 1 kHz to 50 kHz. With a symbol period of 1 $\mu$s, the largest value of $c_b$ is 0.1$\pi$. Compare this to the Doppler frequency shift that is experienced in mobile communications channels: for a car traveling at 100 km/h, the Doppler frequency is approximately 460 Hz. Even for high vehicular speeds, the LO CFO is significantly higher than the Doppler CFO. At 5 GHz, we take the LO CFO as the dominant offset and ignore any Doppler contributions.
B. On Training Sequences

The TS are designed for channel estimation, and ideally have a diagonal covariance matrix [6]. If the combination of signal constellation and sequence length prevents strict orthogonality, the TSs must at least be linearly independent to permit CSI estimation, and normally they are selected to be as close to orthogonal as possible. Even if they are orthogonal, so that $X$ is an orthogonal matrix, it can be seen in (1) that CFOs cause $V(c)$ not to be orthogonal, in general. Further, the segmented estimators introduced in Section V abbreviate the TSs, causing loss of orthogonality, even for zero frequency offset.

C. Maximum Likelihood CFO Estimation

The joint maximum likelihood (ML) estimation of the CSI and CFOs is based on a least squares (LS) solution for the CSI [2], [3]. It is modified here to suit the coordinated MIMO system. By substituting the LS solution for the CSI, the estimation metric is made dependent only upon the CFOs, as will be shown in the following.

The likelihood function as a function of $h$ and $c$ is

$$p(r|h, c) = \frac{1}{(2\pi)^{NN_0}} exp \left(-\frac{||r - V(c)h||^2}{2N_0}\right)$$

and can be reduced to the metric equivalent log-likelihood

$$\Sigma(c) = ||r - V(c)h||^2.$$  (4)

To perform a joint ML estimation of $h$ and $c$, we first estimate $h$ as

$$\hat{h} = (\nabla^\dagger(c)\nabla(c))^{-1}\nabla^\dagger(c)r.$$  (5)

Next, substitute into (4) and reduce, to give

$$\Lambda(c) = r^\dagger\nabla(c)(\nabla^\dagger(c)\nabla(c))^{-1}\nabla^\dagger(c)r.$$  (6)

Using this expression, the CFO estimation problem can be stated as a maximization

$$\hat{c} = \arg\max_c \Lambda(c).$$  (7)

An exhaustive search makes the complexity of the optimum solution exponential in $B$. Once the CFO $\hat{c}$ has been estimated, it can be used in (5) to calculate the CSI.
The search space for large values of $N$ can be highly irregular. As shown in Fig. 2 for a system with $B = 4$, a maximum normalized CFO of $c_{\text{max}} = 0.05$ and a low SNR of 0 dB, there are several local maxima within the search region. This can pose problems to techniques that rely upon a convex objective function, as will be explored and solved in Sections IV and V.

III. Existing Joint CSI and CFO Estimators

Some recent publications [2], [3], [4] provide methods to estimate multiple CFOs with orthogonal training sequences. In [2], [3], the model for the CFO was based on an independent Doppler shift on each multipath signal, whereas in [4], the CFO was based on LO offset.

Below, we outline the estimators of [2] and [3], which will be used for comparison to the proposed estimators in simulations (Section VIII).

A. ST-AO Estimator

In [2], Ahmed et al. exploit the fact that in certain cases, the covariance matrix off-diagonals will be very small compared to the diagonals. This approximation, that $V^\dagger V(c) \approx I_{BN_T}$, allows each CFO to be estimated independently. In practice, this approximation could be valid when the sequences are short enough to avoid accumulation of sufficient phase difference to alter the correlations between the TSs. We will refer to this technique as ST-AO for “simultaneous training - assumed orthogonalization”.

With this approximation, (6) can be rewritten as is

$$\Lambda_{\text{ST-AO}}(c_b) = r^\dagger V_b(c_b) V_b^\dagger(c_b) r$$

$$= \left( \sum_{m=1}^{N_T} \sum_{q=1}^{N_R} \left| \sum_{n=0}^{N-1} r_q(n) x_m(n) e^{j2\pi n} \right|^2 \right),$$

(8)

where $r_q(n)$ refers to the received sample at the $q^{\text{th}}$ antenna at time $n$ and $x_m(n)$ refers to the training symbol transmitted from antenna $m$ at time $n$. The second line of (8) is rewritten to demonstrate that the FFT can be used to solve for $\hat{c}_b$.

Since the performance of the ST-AO estimator is contingent on $V^\dagger V(c)$ being diagonal, it is very sensitive to accumulation of the CFO. As $N$ or the maximum CFO $c_{\text{max}}$ increase and the off-diagonals of $V^\dagger V(c)$ grow, the sequences will be long enough to allow accumulation of the CFO. The result is an error floor at high SNR, which is dependent upon the TS length $N$ and the normalized CFO $c_{\text{max}}$.
B. TO Estimator

The value of orthogonality is evident in comparison of (8) and (6). Orthogonality can be forced in spite of the CFOs by allowing only one BS to transmit in any symbol time. One such format is bunching, e.g., [4], in which each BS in turn sends its TS of length \( N/B \). This reduces computation to \( BN_T N_R \) FFTs, as in (8), followed by \( B \) independent searches. However, the method has two serious disadvantages. First, the time span, or aperture, of any TS has been reduced by a factor of \( B \), compared to the total duration of TS transmissions, with a proportional increase in the standard deviation of the frequency estimates. Second, to maintain the same energy as a TS of the original length \( N \), the peak power must be \( B \) times greater. For radio systems with several BSs, this means more expensive power amplifiers. An alternative is to keep the original length \( N \) of each TS and increase the total duration of TS transmissions by a factor \( B \), but this in turn will reduce the total throughput of the channel.

Another format that maintains orthogonality with only one BS transmitting at a time is interleaving, e.g., [3], in which BSs transmit single symbols of their length-\( N/B \) TSs in round-robin fashion. The duration of each TS remains \( N \) symbol times, so the method does not experience the loss of aperture suffered by the bunching format. It has been shown [3] to meet the Cramer-Rao lower bound (CRLB). We will refer to this method as TO, for “time orthogonalization.” Like the bunching format, TO requires \( B \) times the peak power of simultaneously transmitted TSs. It has another even more serious defect: lack of synchronization among the BSs symbols as received by the MS destroys the orthogonality. In a multi-BS, multi-MS system with a variety of propagation delays, this symbol-synchronous reception is virtually impossible to achieve. Delay spread in the has a similar effect. The format may not be appropriate for realistic channels.

IV. CFO Estimation from Simultaneous Training Sequences

In our proposed method, we employ simultaneous TSs and make no attempt to enforce orthogonality. Instead, we simply take lack of orthogonality into account in the maximization of the log-likelihood (6). This maximization is performed by Newton’s method, which converges with very few iterations in the convex region near the global maximum of \( \Lambda(c) \). The estimator is termed “simultaneous training - Newton step” (ST-N), and is described below in this section. Its performance is assessed analytically in Section VI and by simulation in Section VIII.
Compared to time orthogonal techniques, if the TS energy were kept equal, ST-N would transmit with a symbol power $B$ times lower, permitting use of less expensive power amplifiers. Orthogonality of the TSs in ST-N is degraded by the CFOs; however, we show that proper accounting of the TS correlation compensates for this effect.

For clarity of expression, the discussion below considers only symbol-synchronous arrivals and no delay spread. A straightforward extension that removes this restriction is presented in Appendix II.

The prerequisite for reliable use of the Newton technique is assurance that $\Lambda(c)$ is convex over the search domain, defined by the box constraints $|c_b| \leq c_{\text{max}}$. If the TS length $N$ is large enough, aliasing of $\Lambda(c)$ could occur, causing loss of convexity and even allowing secondary maxima in the search domain. To avoid this, a limit can be imposed on $N$ if the maximum CFO is known \textit{a priori}. For example, we can restrict the length to

$$N < \frac{\pi}{2c_{\text{max}}},$$

where the factor of $1/2$ comes from a lower approximation to the inflection point of the $\text{sinc}(x)$ windowing function in the interval $[0, 1]$. This accounts for the fact that the global maximum of $\Lambda(c)$ will not in general be centred at $c = 0$, and may be as far out as the sides of the box. Of course some situations may demand longer TSs in order to achieve a specific estimator variance, so that the bound (9) is violated. For such cases, we have proposed and detailed two segmented estimators in Section V.

In the case of a perfect quadratic function, the Newton step converges in one step [7]. Since $\Lambda(\hat{c})$ is convex and twice differentiable for $N < \frac{\pi}{2c_{\text{max}}}$, we can assume it approximates a quadratic surface near the global maximum.

Each step of the Newton method has the form [7]

$$\hat{c}_{i+1} = \hat{c}_i - \mu_{ns}(\hat{c}_i),$$

(10)

where the Newton step is

$$\mu_{ns}(\hat{c}) = [\nabla^2 \Lambda(\hat{c})]^{-1} \nabla \Lambda(\hat{c}).$$

(11)

The expressions $\nabla^2 \Lambda(c)$ and $\nabla \Lambda(c)$ refer to the Hessian and gradient of $\Lambda(c)$, respectively.

Next, expressions for the gradient and Hessian of the joint metric must be derived for a given $c$. Using (6), define $G(c) = \mathbf{V}(c)(\mathbf{V}^\dagger(c)\mathbf{V}(c))^{-1}\mathbf{V}^\dagger(c)$, so that $\Lambda(c) = r^\dagger G(c) r$. The $b^{th}$
The difference is due to the partial derivative of \( \frac{\partial}{\partial c_b} \Lambda(\hat{c}) = r^\dagger \left[ \bar{G}_b''(\hat{c}) \right] r \),

where \( \bar{G}_b''(\hat{c}) = I_{NR} \otimes G_b''(\hat{c}) \) denotes block diagonalization and

\[
G_b''(\hat{c}) = \frac{\partial}{\partial c_b} G(\hat{c})
\]

\[
= A_b(\hat{c}_b) F(\hat{c}) V^\dagger(\hat{c}) + V(\hat{c}) F(\hat{c}) A_b^\dagger(\hat{c}_b) - V(\hat{c}) F(\hat{c}) B_b(\hat{c}) F(\hat{c}) V^\dagger(\hat{c}).
\]

The matrix \( A_b(\hat{c}_b) = [0_{N \times (b-1)Nh} jZX_b \odot \Omega_b(\hat{c}_b) 0_{N \times (B-b)Nh}] \in \mathbb{C}^{N \times BN_T} \), the matrix \( 0_{N \times M} \in \mathbb{R}^{N \times M} \) is an all zero matrix, the matrix \( Z = diag(0, 1, 2 \cdots N-1) \in \mathbb{R}^{N \times N} \), the matrix \( F(\hat{c}) = (V^\dagger(\hat{c}) V(\hat{c}))^{-1} \in \mathbb{C}^{BN_T \times BN_T} \), and the matrix \( B_b(\hat{c}_b) = A_b^\dagger(\hat{c}_b) V(\hat{c}) + V^\dagger(\hat{c}) A_b(\hat{c}_b) \in \mathbb{C}^{BN_T \times BN_T} \).

The Hessian \( \nabla^2 \Lambda(\hat{c}) = \mathbb{R}^{B \times B} \) has diagonals with a slightly different form than the off-diagonals. For \( b \neq l \), the \((b^{th}, l^{th})\) entry of the \( B \times B \) matrix \( \nabla^2 \Lambda(\hat{c}) \) is

\[
\frac{\partial^2}{\partial c_b \partial c_l} \Lambda(\hat{c}) = r^\dagger \left[ \bar{G}_{b,l}''(\hat{c}) \right] r,
\]

where \( \bar{G}_{b,l}''(\hat{c}) = I_{NR} \otimes G_{b,l}''(\hat{c}) \) and

\[
G_{b,l}''(\hat{c}) = \frac{\partial^2}{\partial c_b \partial c_l} G(\hat{c})
\]

\[
= \Phi_{b,l}(\hat{c}) + \Phi_{b,l}^\dagger(\hat{c}) = G_{b,l}''(\hat{c})
\]

and

\[
\Phi_{b,l}(\hat{c}) = A_b(\hat{c}_b) F(\hat{c}) A_l^\dagger(\hat{c}_l) - V(\hat{c}) F(\hat{c}) B_b(\hat{c}) F(\hat{c}) A_l^\dagger(\hat{c}_l) - A_b(\hat{c}_b) F(\hat{c}) B_l(\hat{c}) F(\hat{c}) V^\dagger(\hat{c}) - V(\hat{c}) F(\hat{c}) (A_l^\dagger(\hat{c}_l) A_l(\hat{c}_l) - B_b(\hat{c}) F(\hat{c}) B_l(\hat{c}) F(\hat{c}) V(\hat{c}) V^\dagger(\hat{c}).
\]

For the diagonals \( b = l \), the expression for \( \Phi_{b,b}(\hat{c}) \) is

\[
\Phi_{b,b}(\hat{c}) = A_b(\hat{c}_b) F(\hat{c}) A_b^\dagger(\hat{c}_b) - V(\hat{c}) F(\hat{c}) (A_b^\dagger(\hat{c}_b) A_b(\hat{c}_b) - B_b(\hat{c}) F(\hat{c}) B_b(\hat{c}) F(\hat{c}) V^\dagger(\hat{c}) - A_b(\hat{c}_b) F(\hat{c}) B_b(\hat{c}) F(\hat{c}) V^\dagger(\hat{c}) - V(\hat{c}) F(\hat{c}) B_b(\hat{c}) F(\hat{c}) A_b(\hat{c}_b)
\]

\[
+j \left( Z A_b(\hat{c}_b) F(\hat{c}) V^\dagger(\hat{c}) + V(\hat{c}) F(\hat{c}) A_b^\dagger(\hat{c}_b) Z V(\hat{c}) F(\hat{c}) V^\dagger(\hat{c}) \right).
\]

The difference is due to the partial derivative of \( \frac{\partial}{\partial c_b} A_b(\hat{c}_b) = jZ A_b(\hat{c}_b) \), whereas \( \frac{\partial}{\partial c_l} A_b(\hat{c}_b) = 0 \).

The actual computation load is smaller than (13)-(17) suggest, since only a fraction \( 1/B \) of the columns of \( A \) are non-zero, and many partial computations can be saved and re-used.
The Newton technique presented above is reliable only for systems that satisfy the convexity constraint of (9); that is, with short enough TSs and/or small enough maximum CFO values. The next section details two extensions of the Newton technique that are able to cope with larger CFO magnitudes and longer TSs.

V. SEGMENTED ESTIMATORS

The Newton estimator may not converge if the domain of \( c \) is large enough or the TS long enough to make \( \Lambda(c) \) non-convex. To deal with this case, two extensions of the estimator in Section IV are presented here, termed the “telescoping estimator” (TE) and the “combining estimator” (CE). Since both techniques rely upon breaking the length-\( N \) TS into shorter sections, they are referred to collectively as segmented estimators (SE). The segments are chosen to be just short enough to ensure convexity. The Newton estimator of Section IV is used within each segment, although with different (shorter) sequences \( X \). We assume that the shortened sequences in \( V(c) \), although not orthogonal, still have full column rank for the CSI estimation.

The SE approaches have more than one stage. After each stage, the reduced standard deviation improves the knowledge of \( c \), and permits a reduced search space. This in turn allows longer segments in the next stage, until the segment in the final stage is the entire TS. In this way, the SEs can use long TSs, with their large aperture and energy, and still find the global maximum.

To illustrate the utility of the segmented estimators, the following example shows the surface of \( \Lambda(c) \) and how it changes with the number of training symbols. Figures 3 and 4 show \( \Lambda(c) \) for a system with \( B = 2 \), \( N = 100 \), and \( c_{\text{max}} = 0.02 \). For the sake of example, we require \( E_{TS} = 512 \) J and the energy per symbol \( E_b = 1 \) J; this requires \( N = 512 \). Applying the relation in (9), we get \( N = 512 > \pi/2(0.02) = 79 \), which means that \( \Lambda(c) \) is not convex for the full TS. This can be seen in Fig. 3 from the contour ledge in the centre-right of the graph. By only using the first 64 symbols of the TS, the problem is guaranteed to be convex. Fig. 4 shows that the contours have effectively zoomed in on the global maximum by increasing the size of the main lobe. The trade-off is a loss in TS energy and resolution in return for convexity; these losses are made up for in later stages, as the segment lengths increase.

There are two main differences between the TE and CE estimators. First, for the CE estimator, the entire TS must be received before any processing can proceed. This is not necessary for the TE estimator, which can begin processing once a certain number of symbols have been received.
Second, since the CE estimator makes use of the entire TS simultaneously for every stage, it will provide a higher quality estimate.

A. Telescoping Estimator

The TE estimator is based on breaking the TS into successive segments of increasing length, as shown in Fig. 5. For the $s^{th}$ stage, $N_{TE}(s)$ training symbols will be used by the ST-N technique. For the first stage, the number of training symbols used in the ST-N technique is based on the convexity constraint (9), so $N_{TE}(1) = \pi/2c_{max}$. The CFO estimate $\hat{c}(1)$ is made after convergence of ST-N, followed by an estimate of the standard deviation of $\hat{c}(1)$, calculated from the CRLB as in Section VI. This standard deviation, denoted $\sigma_{TE}(s)$ after stage $s$, is then used to reduce the CFO search range for the subsequent stage. To provide some room for channel and noise variations, we use $5\sigma_{TE}$ to calculate $N_{TE}$, so $N_{TE}(s+1) = \pi/10\sigma_{TE}(s)$. During the second stage, the algorithm will calculate a second estimate using $N_{TE}(2)$ symbols, and so on, until the entire TS is received and processed as a whole. To ensure the search space remains convex, the allowable range for $\hat{c}_{b,i,s}$ at the $s^{th}$ stage and $i^{th}$ Newton iteration is restricted to $\hat{c}_{b,0,s-1} \pm \sigma_{TE}(s)$. Table I details the algorithm. The process of determining the standard deviation of the estimate $\sigma_{TE}(s)$ will be explained in Section VI.

The TE Newton step is defined as

$$\mu_{TE}(c, s) = \left[\nabla^2 \Lambda_{TE}(c, s)\right]^{-1} \nabla \Lambda_{TE}(c, s).$$

(18)

The expression $\Lambda_{T}(c, s)$ refers to replacing the vector $r$ and matrix $X$ in (12)-(17) with $r_{TE}(N_{TE}(s)) = \mathcal{I}(N_{TE}(s), N_R) r$ and $X_{TE}(N_{TE}(s)) = \mathcal{I}(N_{TE}(s), 1) X$, respectively, where $\mathcal{I}(N_{TE}(s), N_R) = I_{N_R} \otimes \left[I_{N_{TE}(s)} 0_{N-N_{TE}(s)}\right]$. This simply denotes selecting the first $N_{TE}(s)$ samples from each receive antenna. In Table I, the line after the Newton update is used to ensure that $\hat{c}$ stays within the boundaries of the convex search space.

The number of stages necessary to calculate an estimate is variable because $\sigma_{TE}(s)$ is dependent on the SNR, $B$, $N_T$, and $N_R$, as well as the TS sequences themselves. However, the complexity of the TE can be calculated, since an approximate $\sigma_{TE}(s)$ is pre-calculated and used from a look-up table (Section VI).
B. Combining Estimator

The CE estimator uses the entire TS at each stage $s$ by combining its $L_{CE}$ equal-length segments incoherently, as illustrated in Fig. 6. Because of this, the entire sequence must be received before starting the estimation process. As in the TE, the CE will begin with $N_{CE}(1) = \pi/2c_{max}$, where $N_{CE}(s)$ refers to the segment length for the $s^{th}$ stage. This gives $L_{CE}(s) = \lfloor N/N_{CE}(s) \rfloor$ segments at stage $s$, where $\lfloor \cdot \rfloor$ represents the floor function. Unlike the TE, which uses the first $N_{CE}(s)$ training symbols, the CE estimator adds the log-likelihoods $\Lambda(c)$ from all segments of the TS to form an overall log-likelihood. In effect, each segment is allowed its own set of amplitudes and phases in the CSI gains $h$. The gradient and Hessian functions are the sum of the gradients and Hessians of the segments, and are used to calculate the CE Newton step. The segment sizes in $N_{CE}(s)$ are calculated in the same manner as for the TE estimator (Section VI).

The Newton step for the CE estimator is

$$\mu_{CE}(c, s) = \left[\nabla^2 \Lambda_{CE}(c, s) \right]^{-1} \nabla \Lambda_{CE}(c, s). \tag{19}$$

The gradient of the $s^{th}$ stage of the CE estimator is

$$\frac{\partial}{\partial c_b} \Lambda_{CE}(c, s) = \sum_{u=1}^{L_{CE}(s)} r_u^\dagger [\mathcal{G}'_{b}(c, u)] r_u, \tag{20}$$

where $r_u = \mathcal{I}(N_{CE}(s), 1, u) r$ is the $u^{th}$ block of received samples, where $\mathcal{I}(N_{CE}(s), N_R, u) = I_{N_R} \otimes \left[ \mathbf{0}_{(u-1)N_{CE}(s)} \mathbf{I}_{N_{CE}(s)} \mathbf{0}_{N_N-uN_{CE}(s)} \right]$. The modified $\mathcal{G}'_{b}(c, u)$ is created by replacing $X$ with $X_u = \mathcal{I}(N_{CE}(s), N_R, u) X$. The $(b^{th}, l^{th})$ Hessian is created in the same manner, with

$$\frac{\partial^2}{\partial c_b \partial c_l} \Lambda_{CE}(c, s) = \sum_{u=1}^{L_{CE}} r_u^\dagger \left[ \mathcal{G}''_{b,l}(c, u) \right] r_u, \tag{21}$$

where the modified $\mathcal{G}''_{b,l}(c, u)$ is also created by replacing the TS $X$ with $X_u$. Table II details the algorithm.

The CE is expected to have a lower estimator variance since it makes use of the entire sequence throughout the segmented process, unlike the TE estimator, which only uses information as it arrives. Also note that since we use the floor function on $L_{CE}(s)$, there are samples that are neglected. In our numerical results, we simply ignored any leftover symbols.
VI. CRAMER-RAO LOWER BOUND

To provide a measure of estimator accuracy, we derive the Cramer-Rao lower bound (CRLB) for the MS-side estimator for multiple CFOs in a coordinated MU-MIMO system. In [3], the CRLB was found for multiple CFOs resulting from independent multipath. The system model allowed for the Fisher information matrix to consist of block diagonal components, one for each receive antenna. In the coordinated system presented here, the number of CFOs is independent of the number of receive antennas, and the CRLB must be reformulated.

We use the CRLB in two ways. The first is as an analytical expression for the high-SNR performance of the ST-N method, as well as a check on the performance obtained in simulations. The second is as a way to approximate (or lower bound) the achieved standard deviation in each segment of the ST-TE-N and ST-CE-N segmented methods. As outlined in Section V, these standard deviations allow determination of the segment lengths for the subsequent stage.

The details of the derivation can be found in Appendix A. The estimate vector for the joint estimation of the channel states and the CFOs is defined as

$$\mathbf{\eta} = \left[ \mathbf{\eta}_1^T \cdots \mathbf{\eta}_{N_R}^T \mathbf{c}^T \right]^T \in \mathbb{R}^{2BN_TN_R+BN \times 1},$$

(22)

where $$\mathbf{\eta}_q = [\text{Re}(\mathbf{h}_q)^T \text{Im}(\mathbf{h}_q)^T]^T \in \mathbb{R}^{2BN_T \times 1}$$ and $$\mathbf{c} = [c_1 \cdots c_{B}]^T \in \mathbb{R}^{B \times 1}$$. The vector of transmitted symbols after distortion by the CFOs, $$\mathbf{u}$$, is formed as

$$\mathbf{u}(\mathbf{c}) = [\mathbf{u}_1^T(\mathbf{c}) \cdots \mathbf{u}_{N_R}^T(\mathbf{c})]^T \in \mathbb{C}^{NNR \times 1},$$

(23)

where $$\mathbf{u}_q(\mathbf{c}) = V(\mathbf{c})\mathbf{h}_q$$.

The Fisher information matrix is defined as

$$\mathbf{F} = \begin{bmatrix}
\mathbf{F}_H(1,1) & \cdots & 0 & \mathbf{F}_{Hp}(1) \\
\vdots & \ddots & \vdots & \vdots \\
0 & \cdots & \mathbf{F}_H(N_R,N_R) & \mathbf{F}_{Hp}(N_R) \\
\mathbf{F}_{pH}(1) & \cdots & \mathbf{F}_{pH}(N_R) & \mathbf{F}_p
\end{bmatrix},$$

(24)
where
\[
\begin{align*}
F_H(k, k) &= \frac{2}{N_0} \Re \left[ V^\dagger(c)V(c) \begin{pmatrix} jV^\dagger(c)V(c) \\ -jV^\dagger(c)V(c) \end{pmatrix} \right], \\
F_{cH}(k) &= \frac{2}{N_0} \Re \left[ -jC(k)C(k) \right], \\
F_{HC}(k) &= F_{cH}(k)^\dagger \\
F_c &= \frac{2}{N_0} \Re(P) \\
C(q) &= \begin{bmatrix} V_1^{<1T}(c_1)Z_NV_1^*(c_1)h_1^{<q>^*} & \cdots & V_B^{<NT>}(c_B)Z_NV_B^*(c_B)h_B^{<q>^*} \\
\vdots & \ddots & \vdots \\
V_1^{<1T}(c_1)Z_NV_1^*(c_1)h_1^{<q>^*} & \cdots & V_B^{<NT>}(c_B)Z_NV_B^*(c_B)h_B^{<q>^*} \\
\end{bmatrix} \\
Z &= \text{diag}(0 \cdots N-1) \\
P_{b,l} &= \sum_{n=0}^{N-1} n^2 \vartheta_b^T(n)H_H^T(b)H_l\vartheta_l(n) \in C \\
\vartheta_b(n) &= X_b(n)^T e^{j\vartheta_b(n)}, \quad (25)
\end{align*}
\]

where \(X_b(n)\) is the row corresponding to the \(n^{th}\) time slot in the \(R^{N \times NT}\) matrix \(X_b\). Using the matrix inverse lemma on (24), the CRLB for the estimation error covariance matrix is
\[
\text{CRLB}_c = \left( F_c - F_{cH}F_H^{-1}F_{HC} \right)^{-1}. \quad (26)
\]

The success of the segmented estimators in Section V-A and V-B at keeping the global maximum within the search domain is dependent upon the accuracy of the standard deviation, e.g. \(\sigma_{TE}(s)\), estimated after each stage. For this, we set the standard deviation equal to the value obtained from the CRLB; from the trace of (26), we have \(\sigma_c^2 = \text{trace}(\text{CRLB}_c)/B\). To be conservative, we then set the box constraints on the search domain as \(|c_b - \hat{c}_b(s)| \leq 5\sigma_c\). This loose constraint maintains convexity of \(\Lambda(c)\) with very high probability.

However, note that (26) depends on the channel \(h\), which in principle is unknown. We have two choices: we can substitute the ML estimate (5); or we can simply average (26) over many channel realizations in Monte Carlo simulation. For two antennas at BSs and MSs, the latter turned out to be adequate. In addition, \(\sigma_c\) will vary depending on the TS length. Therefore, a matrix of values is pre-calculated based on SNR and TS length, as \(\sigma_c(s, \Gamma)\), where \(\Gamma\) refers to the SNR. Table IV shows a typical example of how they are related to segment lengths in the ST-TE-N estimator.
VII. Complexity

The complexity of the techniques is estimated as the number of complex multiply and add (CMA) operations necessary for a single estimate of the $B$ CFOs. The results are asymptotic in $N$, since the TS length is the limiting variable in complexity. Table III compares the CMA’s for ST-AO, TO, ST-N, ST-TE-N, and ST-CE-N.

The variable $S$ used for ST-TE-N and ST-CE-N in Table III refers to the number of stages necessary for the algorithm to converge to $N_{TE}(s) = N$ and $N_{CE}(s) = N$. As explained in the Section VIII and Table IV, this is dependent on the SNR; medium SNR will require $S > 2$ while high SNR will require only $S = 2$.

The ST-N, ST-TE-N and ST-CE-N algorithms have a significant contribution from the calculation of the Hessian, which consists of the last two terms for each technique in Table III. However, since the Hessian does not vary much in the convex space that is guaranteed in these estimators, this can be pre-calculated for a range of CFO values to reduce computations.

VIII. Simulations

The model used for simulations is the same as outlined in Section II. The channel is modeled as a $N_R \times B N_T$ matrix of i.i.d. complex Gaussian random variables, and is static over the entire TS. The ST-AO and TO techniques rely upon a search over a range of CFO values, consisting of $N$ frequency bins, suitable for the FFT. Monte Carlo simulations of 500 runs were used with a range of TS energy levels. The CFO is random and uniformly distributed, with the CFO set to $c_b \in [-c_{\text{max}}, +c_{\text{max}}]$ $\forall b$. The number of iterations used for the Newton method was set to $N_{\text{ITER}} = 3$.

For TO, the TS length was reduced by $B$ to keep the total length equal between the estimators. To keep the comparison between the techniques fair, the horizontal axis was chosen to be the energy of each TS over the noise, $E_{TS}/N_0$. The total transmit energy allocated to each BS per symbol is $E_b$, so that the total transmit energy per symbol per antenna is $E_b/N_T$. With this normalization, any additional transmit antennas at a given BS will degrade the estimate quality. The TS SNR is then defined as $E_{TS}/N_0 = N E_b/N_T N_0$. Since the TO sequence lengths are reduced by $B$, the energy per symbol must be increased by a factor of $B$ to keep the same $E_{TS}/N_0$ value. As noted earlier, while this is a fair comparison of transmitted energy levels, the
transmit amplifiers used for the TO technique are required to handle $B$ times more power than in either the ST-AO or ST-N techniques.

In Fig. 7, the mean square error (MSE) curves for a system where the $\Lambda(c)$ is guaranteed to be convex for the entire sequence is shown. The system parameters are $B = 4$, $N_T = 2$, and $N_R = 2$. For $c_{\text{max}} = 10^{-2}$, the maximum TS length is $N = 157$. A more appropriate length of $N = 64$ was used. Note that the estimates make use of the fact that $\hat{c}_b \in [-c_{\text{max}}, +c_{\text{max}}]$, which is the reason why they outperform the CRLB at very low SNRs. The effect of the TS correlation is demonstrated in the curve for ST-AO, which has an error floor at an MSE of approximately $10^{-5}$ radians$^2$. The ST-N technique from Section IV is seen to achieve the CRLB at moderate to high SNR; the effects at low $E_{TS}$ are a result of the a priori knowledge of $c_{\text{max}}$. The error floor of the TO technique is a result of the FFT resolution being limited to $N$. By increasing the resolution, thus increasing the complexity, this floor can be reduced.

In Fig. 8, MSE curves are presented for a system with a high enough CFO to violate the convexity constraints. The system parameters are $B = 4$, $N_T = 2$, and $N_R = 2$. For $c_{\text{max}} = 5 \times 10^{-2}$, the maximum TS length is $N = 31$. To operate the ST-N with such a short sequence would require very high peak transmit power in order to obtain enough TS energy, so a longer sequence is preferable; a TS length of $N = 128$ was used. The ST-TE-N and ST-CE-N techniques can accommodate this by segmentation of the TS. TO is operable in this system, but it requires 4 times the peak power of ST-TE-N and ST-CE-N. The error floor for TO is again the result of the FFT resolution. The performance of both the TO and the segmented simultaneous training techniques converges to the CRLB around an MSE of $10^{-6}$, with the TO technique having the better performance at very low SNR. One reason this occurs may be due to time orthogonality of the TS, and at low SNR, the simultaneous transmission techniques are incapable of resolving the sequences. Between the two SE techniques, the ST-CE-N has the better low-SNR performance. Since the ST-CE-N uses more training symbols at each stage in the estimation process, it has better initial CFO estimates, and thus outperforms the ST-TE-N.

Table IV shows the look-up data for $N_{TE}(s)$. An interesting observation is that the SNR where the SE techniques converge with the CRLB corresponds to the point where the total number of stages jumps from 1 to 4. At a per symbol per antenna SNR of 5 dB, the TS energy is high enough to warrant decreasing the search range of CFOs; this results in an increase in the number of training symbols used in each stage, until the maximum is reached of $N = 128$. At
higher SNR, the SE techniques both converge to the CRLB using only 2 stages, since the initial estimate is of high enough quality to permit the use of the entire sequence for the second stage. This corresponds to an SNR of 15 dB and higher.

Considering the complexity of calculating the Hessian for each stage and iteration in both the ST-TE-N and ST-CE-N techniques, the idea of a look-up table solution was suggested in Section VII. With a discrete set of 100 equi-spaced values of $c_b \in (-c_{max}, +c_{max})$, simulations verified that the performance difference between the pre-calculated Hessian and the actual Hessian were insignificant.

The number of receive antennas also is a factor in the MSE of the CFO estimates. Simulations were done to demonstrate the effect of increasing $N_R$ on the CRLB. By increasing $N_R$, which effectively increases the receive diversity of the system, the received power used for CFO estimation is increased. For every doubling of $N_R$, we see a 3dB decrease in the CFO MSE.

IX. Summary

In the emerging research on coordinated MIMO systems, the problem of independent oscillators at different BSs has so far been overlooked. The effect of different CFOs is degraded beamforming and zero forcing as phase errors accumulate over time. As a first step in countering the problem, we have provided methods for simultaneous estimation of multiple CFOs (itself a relatively unexamined problem), ones that work at both large and small CFO levels. Corresponding methods for correction of the CFOs will be reported in another paper.

We have demonstrated that, unlike previously proposed methods, the BS training sequences do not have to be sent in disjoint time intervals. They can be simultaneous, with the advantages of lower peak power, greater accuracy for the same training sequence length, and robustness against asynchronous and frequency-selective reception. The core of our approach, termed ST-N, is to use Newton’s method to search for the CFO vector that maximizes the log-likelihood of the received signal. The log-likelihood is convex only near its global maximum, so we have also proposed variations that segment the training sequences to assure convexity for use of Newton’s method.

Simulations show that our proposed estimators, both the basic estimator and the segmented versions, meet the Cramer-Rao lower bound at moderate SNR levels and give good accuracy.
Our techniques do require more computation than previously proposed methods, but their advantages in performance, reduced power amplifier requirements and applicability to realistic channels make them a good choice for coordinated MIMO systems.

APPENDIX I

DERIVATION OF THE CRAMER-RAO LOWER BOUND

This bound was derived with the aid of the Slepian-Bangs formula [8]. Using the definitions of $\eta \in \mathbb{R}^{2BN_T N_R + B \times 1}$ and $u \in \mathbb{R}^{N N_R \times 1}$ from (22) and (23), the Fisher information matrix (FIM) is represented as

$$
F = \frac{2}{N_0} \text{Re} \left( \frac{\partial u^\dagger}{\partial \eta} \frac{\partial u}{\partial \eta^T} \right) \tag{27}
$$

$$
= \begin{bmatrix}
F_H(1, 1) & \cdots & F_H(1, N_R) & F_{He}(1) \\
\vdots & \ddots & \vdots & \vdots \\
F_H(N_R, 1) & \cdots & F_H(N_R, N_R) & F_{He}(N_R) \\
F_{eh}(1) & \cdots & F_{eh}(N_R) & F_c
\end{bmatrix}.
$$

The Fisher information for the channel is:

$$
F_{H}(k, q) = \frac{2}{N_0} \text{Re} \left( \frac{\partial u^\dagger}{\partial \eta_k} \frac{\partial u}{\partial \eta_q^T} \right) = \frac{2}{N_0} \text{Re} \left( \begin{bmatrix}
V^\dagger(c) V(c) & j V^\dagger(c) V(c) \\
- j V^\dagger(c) V(c) & V^\dagger(c) V(c)
\end{bmatrix} \right) \in \mathbb{R}^{2BN_T \times 2BN_T},
$$

(28)

where $V(c)$ is from (1).

The joint Fisher information for the channel and CFO is

$$
F_{eh}(q) = \frac{2}{N_0} \text{Re} \left( \frac{\partial u^\dagger}{\partial c} \frac{\partial u}{\partial \eta_q^T} \right) = \frac{2}{N_0} \text{Re} \left( \begin{bmatrix}
P \end{bmatrix} \right) \in \mathbb{R}^{B \times 2BN_T},
$$

(29)

and

$$
F_{He}(q) = \frac{2}{N_0} \text{Re} \left( \frac{\partial u^\dagger}{\partial \eta_q} \frac{\partial u}{\partial c^T} \right) = F_{eh}(q)^\dagger \in \mathbb{R}^{2BN_T \times B}
$$

(30)

where

$$
C(q) = \begin{bmatrix}
V_1^{<1>^T(c_1) Z_{ND} V_1^*(c_1) h_{1,q}^*} & \cdots & V_B^{<N_T>^T(c_B) Z_N V_1^*(c_1) h_{1,q}^*} \\
\vdots & \ddots & \vdots \\
V_1^{<1>^T(c_1) Z_N V_B^*(c_B) h_{B,q}^*} & \cdots & V_B^{<N_T>^T(c_B) Z_N V_B^*(c_B) h_{B,q}^*}
\end{bmatrix}.
$$

(31)

The Fisher information for the CFO alone is

$$
F_c(k, q) = \frac{2}{N_0} \text{Re} \left( \frac{\partial u^\dagger}{\partial c} \frac{\partial u}{\partial c^T} \right) = \frac{2}{N_0} \text{Re} \left( \begin{bmatrix} P \end{bmatrix} \right) \in \mathbb{R}^{B \times B},
$$

(32)
where
\[
\mathbf{P}_{b,l} = \sum_{n=0}^{N-1} n^2 \vartheta_l^T(n) \mathbf{H}_b^T \mathbf{H}_l \vartheta_l(n) \in \mathbb{C} \quad (33)
\]
\[
\vartheta_l(n) = \mathbf{X}_b(n)^T e^{j c_{bn}} \in \mathbb{C}^{N_T \times 1}, \quad (34)
\]
where \(\mathbf{X}_b(n)\) is the row corresponding to the \(n^{th}\) time slot in the \(R^{N_T \times N_R}\) matrix \(\mathbf{X}_b\).

Note that for \(k \neq q\), \(\mathbf{F}_H(k,q)\) collapses to a zero matrix. Substituting this into (28) gives (24).

For the CRLB, it is necessary to take the inverse of the FIM (28). Since we are primarily interested in the diagonals of the CRLB, the use of the matrix inverse lemma [9] would be efficient. First, define \(\mathbf{F}_c \in \mathbb{R}^{2BN_T \times BN_T} \) as containing the upper-left components of \(\mathbf{F}\) that pertain to the channel alone, \(\mathbf{F}_{cH} \in \mathbb{R}^{BN_T \times BN_T} \) and \(\mathbf{F}_{hc} \in \mathbb{R}^{BN_T \times BN_T} \) as containing the components of \(\mathbf{F}\) that pertain to the joint channel and CFO information, and \(\mathbf{F}_c \in \mathbb{R}^{B \times B}\).

The matrix inverse lemma then states that the inverse of the submatrix \(\mathbf{F}_c\) in
\[
\mathbf{F} = \begin{bmatrix} \mathbf{F}_H & \mathbf{F}_{hc} \\ \mathbf{F}_{ch} & \mathbf{F}_c \end{bmatrix}
\]
is
\[
\text{CRLB}_c = \left[ \mathbf{F}^{-1} \right]_c = \left( \mathbf{F}_c - \mathbf{F}_{ch} \mathbf{F}_H^{-1} \mathbf{F}_{hc} \right)^{-1}. \quad (36)
\]
The expression \(\left[ \mathbf{F}^{-1} \right]_c\) refers to the lower-right \(B \times B\) block of \(\mathbf{F}^{-1}\). For the channel state, the CRLB is
\[
\text{CRLB}_H \left[ \mathbf{F}^{-1} \right]_H = \left( \mathbf{F}_H - \mathbf{F}_{hc} \mathbf{F}_c^{-1} \mathbf{F}_{ch} \right)^{-1}. \quad (37)
\]
The trace of \(\text{CRLB}_c\) and \(\text{CRLB}_H\) correspond to the Cramer-Rao lower bound for the CFO and CSI, respectively.

**APPENDIX II**

**ASYNCHRONOUS AND FREQUENCY SELECTIVE CHANNELS**

It is straightforward to extend the ST-N method to asynchronous arrivals and delay spread (frequency selectivity). Start by reinterpreting (2) as having \(N_{ss}\) samples per symbol (sample spacing \(T_s/N_{ss}\)) in the arrays \(\mathbf{r}, \mathbf{V}(c)\) and \(\mathbf{n}\), which now have \(NN_{ss}\) rows. With \(N_{ss} > 1\), we are guaranteed Nyquist-rate sampling. Next, define the arriving signals, delayed by \(l\) samples and zero-padded, as
\[
\mathbf{\overline{V}}^{(l)}(c) = \begin{bmatrix} \mathbf{0}_{BN_T \times l} & \mathbf{V}^T(c) \mathbf{0}_{BN_T \times (L-l)} \end{bmatrix}^T \in \mathbb{C}^{(NN_{ss}+L) \times BN_T}, \quad (38)
\]
where $L$ is the length of the impulse response (to be defined). Collect these arrays into a single array of received signals

$$\tilde{V}(c) = \left[ V^{(0)}(c) \ V^{(1)}(c) \cdots V^{(L)}(c) \right].$$  \hfill (39)

Interpolation, to account for asynchronicity, has the same form as delay spread, with a vector impulse response $h^{(l)}$, $l = 0, \ldots, L$ spanning $LT_s/N_{ss}$ seconds. Form $\tilde{h} = \left[ h^{(0)T} h^{(1)T} \cdots h^{(L)T} \right]^T$. Then the expression for received samples

$$r = \tilde{V}(c)\tilde{h} + n$$ \hfill (40)

has the same form as (2) and the CSI and CFO estimates can be performed in the same way as for synchronous, unspread signals. Note that any phase shifts incurred in the delayed $V^{(l)}(c)$ will be absorbed into the channel estimates of $\tilde{h}$.

REFERENCES


CSI/CFO Estimate Feedback Channel

Coordinated BSs

Fig. 1. Coordinated MU-MIMO downlink model

Fig. 2. $A(c)$ contours for $N = 200$, SNR=0dB
Fig. 3. $A_{ST-TE-N}$ contours for $N = 512$, $L_E = 1$, $c_1 = 0.0131$, $c_2 = -0.0040$, SNR=30dB

Fig. 4. $A_{ST-TE-N}$ contours for first stage of $N = 512$, $L_E = 8$, $c_1 = 0.0131$, $c_2 = -0.0040$, SNR=30dB
Fig. 5. Telescoping Estimator, NE = Newton estimator, 4 stages

Fig. 6. Combining Estimator, NE = Newton estimator, 3 stages
Fig. 7. CFO MSE vs. TS energy, $B = 4$, $N_T = 2$, $N_R = 2$, $c_{\text{max}} = 10^{-2}$, $N = 64$ BPSK Hadamard TS

Fig. 8. CFO MSE vs. TS energy, $B = 4$, $N_T = 2$, $N_R = 2$, $c_{\text{max}} = 5 \times 10^{-2}$, $N = 128$ BPSK Hadamard TS
TABLE I
TELESCOPING ESTIMATOR

\( \hat{c}_{0,1} = 0 \)

\( s = 1 \)

\( \sigma_{TE}(s) = c_{max} \)

\( N_{TE}(s) = \frac{n}{\sigma_{TE}(s)} \)

while \( N_{TE}(s) \leq N \)

for \( i = 1 \) to \( N_{ITER} \)

\( \hat{c}_{i,s} = \hat{c}_{i-1,s} - \mu_{TE}(\hat{c}_{i-1,s}) \)

if \( |\hat{c}_{b,i,s} - \hat{c}_{b,0,s}| > \sigma_{TE}(s), \hat{c}_{b,i,s} = \hat{c}_{b,i-1,s} \pm \sigma_{TE}(s) \) \( \forall b \)

end

\( s = s + 1 \)

\( \hat{c}_{0,s} = \hat{c}_{N_{ITER},s-1} \)

\( \sigma_{TE}(s) = 5\sigma_e(s, \Gamma) \)

\( N_{TE}(s) = \frac{s}{2\sigma_{TE}(s)} \)

end
TABLE II
COMBINING ESTIMATOR

\[ \hat{c}_{0,1} = 0 \]

\[ s = 1 \]

\[ \sigma_{CE}(s) = c_{max} \]

\[ N_{CE}(s) = \frac{s}{2\sigma_{CE}(s)} \]

while \( N_{CE}(s) \leq N \)

for \( i = 1 \) to \( N_{ITER} \)

\[ \hat{c}_{i,s} = \hat{c}_{i-1,s} - \mu_{CE}(\hat{c}_{i-1,s}) \]

if \( |\hat{c}_{b,i,s} - \hat{c}_{b,0,s}| > \sigma_{CE}(s) \)

\[ \hat{c}_{b,i,s} = \hat{c}_{b,i-1,s} \pm \sigma_{CE}(s), \forall b \]

end

\[ s = s + 1 \]

\[ \hat{c}_{0,s} = \hat{c}_{N_{ITER},s-1} \]

\[ \sigma_{CE}(s) = 5\sigma_c(s, \Gamma) \]

\[ N_{CE}(s) = \frac{s}{2\sigma_{CE}(s)} \]

end

TABLE III
COMPLEXITY COMPARISON

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Approx. CMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST-AO</td>
<td>( BN_R N_T (2N + N\log_2(N)) + BN )</td>
</tr>
<tr>
<td>TO</td>
<td>( N_R N_T \left( 2N + N\log_2 \left( \frac{N}{B} \right) \right) + BN )</td>
</tr>
<tr>
<td>ST-N</td>
<td>( N_{ITER} \left[ B^3 + B^2 + BN^2 \left( N^2_R + 2BN_T \right) + \frac{1}{2}B^2 N^2 N^2_R + 5B^3 N^2 N_T + 4B^2 N^2 N_T \right] )</td>
</tr>
<tr>
<td>ST-TE-N</td>
<td>( \sum_{s=1}^{S} N_{ITER} \left[ B^3 + B^2 + BN_T^2(s) \left( N^2_R + 2BN_T \right) + \frac{1}{2}B^2 N^2_T(s) N^2_R + 5B^3 N^2_T(s) N_T + 4B^2 N^2_T(s) N_T \right] )</td>
</tr>
<tr>
<td>ST-CE-N</td>
<td>( \sum_{s=1}^{S} N_{ITER} \left[ B^3 + B^2 + L_{CE}(s) BN_R^2(s) \left( N^2_R + 2BN_T \right) + \frac{L_{CE}(s)}{2} B^2 N^2_H(s) N^2_R + L_{CE}(s) \left( 5B^3 N^2_R(s) N_T + 4B^2 N^2_R(s) N_T \right) \right] )</td>
</tr>
<tr>
<td>Brute Force</td>
<td>( N^B \left( B^3 N^3 R^3 T^3 \right) + N^B )</td>
</tr>
</tbody>
</table>
**TABLE IV**

Look-up Data for $N_{TE}(s)$, $B = 4$, $N = 128$ BPSK Hadamard TS, $c_{max} = 0.05$

<table>
<thead>
<tr>
<th>$\frac{E_b}{N_0}$</th>
<th>$s = 1$</th>
<th>$s = 2$</th>
<th>$s = 3$</th>
<th>$s = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-20dB$</td>
<td>31</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$-15dB$</td>
<td>31</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$-10dB$</td>
<td>31</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$-5dB$</td>
<td>31</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$0dB$</td>
<td>31</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$5dB$</td>
<td>31</td>
<td>46</td>
<td>86</td>
<td>128</td>
</tr>
<tr>
<td>$10dB$</td>
<td>31</td>
<td>83</td>
<td>128</td>
<td>-</td>
</tr>
<tr>
<td>$15dB$</td>
<td>31</td>
<td>128</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$20dB$</td>
<td>31</td>
<td>128</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$25dB$</td>
<td>31</td>
<td>128</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$30dB$</td>
<td>31</td>
<td>128</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$35dB$</td>
<td>31</td>
<td>128</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$40dB$</td>
<td>31</td>
<td>128</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>