Under-determined Training and Estimation for Distributed Transmit Beamforming Systems

Jian A. Zhang, Senior Member, IEEE, Tao (Tom) Yang, Member, IEEE, and Zhuo Chen, Member, IEEE

Abstract—Distributed transmit beamforming (DTB) can significantly boost the signal-to-noise ratio (SNR) of a wireless communication system. To realize the benefits of DTB, generating and feeding back beamforming vector are very challenging tasks. Existing schemes have either enormous overhead or weak robustness in noisy channels. In this paper, we investigate the design of training sequences and beamforming vector estimators in DTB systems. We consider an under-determined case, where the length of training sequence $N$ sent from each node is smaller than the number of source nodes $M$. We derive the optimal estimation of the beamforming vector that maximizes the beamforming gain and show that it can be well approximated as the linear minimum mean square error (LMMSE) estimator. Based on the LMMSE estimator, we investigate the optimal design of training sequences and propose efficient DTB schemes. We analytically show that these schemes can achieve approximately $N$ times increased SNR in uncorrelated channels, and even higher gain in correlated ones. We also propose a concatenated training scheme which optimally combines the training signals over multiple frames to obtain the beamforming vector. Simulation results demonstrate that the proposed DTB schemes can yield significant gains even at very low SNRs, with total feedback bits much less than those required in the existing schemes.

I. INTRODUCTION

Distributed transmit beamforming (DTB), where multiple geometrically separated source nodes cooperatively transmit common messages with aligned phases, can tremendously boost the signal-to-noise ratio (SNR) at the destination [1], [2]. To achieve such a beamforming gain, a beamforming vector is required by the source nodes to adjust the phases, as well as magnitudes, of transmitted signals so that they can constructively combine at the intended destination. Generating such a beamforming vector is a very challenging task. This is because not only fixed phase offsets between source nodes need to be estimated and compensated, but the estimation also needs to be updated frequently due to time varying phase shift caused by both residual carrier frequency offset (CFO) and channel variation. Minimizing the cost associated with training and feedback to achieve the beamforming gain has attracted intensive research efforts [3]–[6].

Focusing on minimizing the feedback requirement, some iterative training based schemes have been reported. A one-bit feedback iterative training scheme is proposed in [3]. Its convergence in noiseless channels is proven in [3], [7]. Its convergence speed, which is linearly proportional to the number of source nodes, is very slow due to the randomness of phase adjustment. In addition, when the initial SNR is low, decision errors with regard to the variation of the SNR at the destination node can happen frequently and impede the convergence significantly [4]. Some improved schemes are proposed in [4]–[6], [8], [9]. These improvements mainly focus on improving the convergence speed by exploiting the received information. The problem associated with the decision error is not yet solved. Therefore, the convergence speed is still more than 10 times to the number of source nodes in noiseless channels [6], [8], [9] or in the case of high SNRs, e.g., 30 dB [4]. The inefficiency of the aforementioned schemes are largely due to the negligence of the training sequence design, while focusing on reducing feedback bits.

Training sequence design for channel estimation has been well studied for conventional multiple-input multiple output (MIMO) systems [10]–[15]. Optimization of training sequences is generally formulated as minimizing the mean square error (MSE) of the estimate [12]–[15] or maximizing the capacity [10], [11]. In most existing work, the length of the training sequence sent from each transmit antenna, denoted by $N$, is assumed to be greater than or equal to the number of transmit antennas $M$. However, in DTB systems, the number of source nodes $M$ could be very large, for example, 20 sensor nodes are required to form a DTB in an application scenario of using unmanned aerial vehicle for sensor data collection [16]. Frequently sending $N \geq M$ training symbols from each node leads to enormous overhead in typically low data-rate and power-consumption sensor networks. To reduce the training overhead, this paper distinctively considers the case of $N < M$, which leads to an under-determined problem. The under-determined problem has been studied for detection and demodulation of signals with finite states, for example, in [17], [18]. It receives little attention in channel estimation, as solving the under-determined problem w.r.t. the case of $N < M$ can cause large estimation error and severe inter-symbol interference [19]. However, the DTB system can tolerate relatively large estimation errors [2] which only affects the achievable SNR. As to be seen in this paper, beamforming gain proportional to the number of training symbols $N$ can be achieved in under-determined cases.

In this paper, we investigate the training sequence and estimation design for $N < M$. We propose several DTB schemes which achieve good balance between complexity and beamforming gain. Our main contributions are summarized below:

- We derive the optimal beamforming vector which maxi-
mizes the beamforming gain for given training sequences. We also show that the optimal beamforming vector can be well-approximated by the normalized linear Minimum MSE (LMMSE) estimate;

- Based on the LMMSE estimator, we derive optimal training sequences for both spatially uncorrelated and correlated channels, for $N < M$, under the total training power constraint. The optimization function and methodology in our under-determined scenario are different to those well-known in the determined channel estimation problems. The solution is found to be a function of the eigenvector matrix of the channel correlation matrix, which is similar to those obtained in determined channel estimation problems. However, we provide a more general solution, which enables better design under per-node power constraint and covers known solutions as special examples.

- We propose four DTB schemes using different training sequences and/or estimators. These schemes have different levels of tradeoff among implementation complexity, feedback requirement, and beamforming gain. Then we analytically show the beamforming gains achieved by these schemes, which are approximately proportional to $N$ for spatially uncorrelated channels, and even higher for spatially correlated channels;

- We also propose a concatenated training scheme which can efficiently combine training sequences transmitted during different frames to generate a better beamforming vector. In slow time varying channels, as few as 1 training bit and 1 feedback symbol are required.

We note that although we specifically consider the case of $N < M$, the results in this paper can be directly applied to the case of $N = M$, should a comparison with the performance in such a case be required.

The rest of the paper is organized as follows. In Section II, the system model of the DTB is presented. In Section III, we derive the optimal beamforming vector which maximizes the beamforming gain and show that it can be well-approximated by the LMMSE estimator for typical scenarios. In Section IV, we design training sequences for spatially uncorrelated and correlated channels, aiming at maximizing the beamforming gain. In Section V, we present four training and estimation schemes, and derive closed-form expressions for their beamforming gains. In Section VI, we extend our results to the concatenated training case. Information feedback approaches for generating beamforming vectors are also briefly discussed. Section VII presents simulation results and Section VIII concludes the paper.

**Notation:** Lower-case bold face variables ($\mathbf{x}$) indicate vectors, and upper-case bold face variables ($\mathbf{X}$) indicate matrices. For a diagonal matrix $\mathbf{X}$, $\mathbf{X}^a$ denotes the power $a$ operation to each diagonal element. $\mathbf{I}$ denotes the identity matrix, $E(\cdot)$ denotes expectation, and $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^{-1}$ and $(\cdot)^*$ denote transposition, conjugate transposition, inverse and pseudo-inverse, respectively.

### II. System Model

We consider a DTB system where $M$ source nodes collaboratively send common messages to a destination node and each node has a single antenna. The protocol under consideration consists of training phases and data transmission phases. In the training phases, all source nodes transmit their training symbols simultaneously to the destination. Then, the destination feedback either the estimated beamforming vector or received signal to the source nodes. We will mainly consider the former approach and discuss the latter in Section VI-C. In the data transmission phases, the source nodes utilize the estimated beamforming vector to adjust their phases and amplitudes. We assume that timing and carrier frequency synchronization have been achieved using approaches in, e.g., [16], [20]. Based on the structure of training signals and the ways of exploiting them, we will first consider block training and then extend it to concatenated training in Section VI.

In the block training scheme, each of $M$ source nodes simultaneously and continuously transmits pre-designed training sequence of length $N$ to the destination node. These training symbols from all source nodes are represented by an $N \times M$ matrix $\mathbf{P}$, and each column of $\mathbf{P}$ is called a training sequence of length $N$. The $m^{th}$ source node only needs to know the $m^{th}$ column of $\mathbf{P}$, and the destination node knows the entire $\mathbf{P}$. The total energy of the training signals is given by $\text{Tr}(\mathbf{P}^H \mathbf{P})$. In this paper, we focus on the total training energy constraint $\text{Tr}(\mathbf{P}^H \mathbf{P}) \leq \varepsilon$, and will discuss the suitability of the derived results under the per-node power constraint $\sum_{n=1}^{N} |p_{nm}|^2/N \leq \varepsilon/(NM)$, $m = 1, \ldots, M$, where $p_{nm}$ is the $(n,m)^{th}$ element of $\mathbf{P}$.

Since source nodes are generally close to each other and far away from the destination node, we assume that the path losses between these source nodes and the destination node are the same and have value 1. Let an $M \times 1$ vector $\mathbf{h} = [h_1, h_2, \cdots, h_M]^T$ denote the flat fading channel coefficients from the $M$ source nodes to the destination node, where $h_m, m = 1, \ldots, M$, are circularly symmetric complex Normal ($CN$) random variables with zero mean. The covariance matrix of $\mathbf{h}$ is denoted by $E(\mathbf{h} \mathbf{h}^H) = \Phi$. Note that $\Phi$ is a positive definite matrix. Without loss of generality, we assume that the diagonal elements of $\Phi$ are 1s. The channels are assumed to be static during the training and signal transmission period. During the training period, the received signal at the destination node can be represented as

$$\mathbf{y} = \mathbf{Ph} + \mathbf{z},$$  \hspace{1cm} (1)

where $\mathbf{z}$ is an $N \times 1$ noise vector whose elements are i.i.d. $CN$ random variables with zero mean and variance $\sigma_z^2$.

From $\mathbf{y}$, an $M \times 1$ beamforming vector $\mathbf{w}$, $\mathbf{w}^H \mathbf{w} = 1$, will be determined. The achieved beamforming gain is defined as $\rho = |\mathbf{w}^H \mathbf{h}|^2$. In addition, we define the normalized beamforming gain as the ratio between $\rho$ and the ideal gain $\rho_{\text{ideal}}$ of the maximum ratio transmission with perfect knowledge of $\mathbf{h}$, i.e.,

$$\rho_{\text{norm}} = \frac{\rho}{\rho_{\text{ideal}}} = \frac{|\mathbf{w}^H \mathbf{h}|^2}{|\mathbf{h}^H \mathbf{h}/|\mathbf{h}^H \mathbf{h}||} = \frac{|\mathbf{w}^H \mathbf{h}|^2}{|\mathbf{h}^H \mathbf{h}|} \leq 1.$$

$$\rho_{\text{norm}} = \frac{\rho}{\rho_{\text{ideal}}} = \frac{|\mathbf{w}^H \mathbf{h}|^2}{|\mathbf{h}^H \mathbf{h}/|\mathbf{h}^H \mathbf{h}||} = \frac{|\mathbf{w}^H \mathbf{h}|^2}{|\mathbf{h}^H \mathbf{h}|} \leq 1.$$

$$\rho_{\text{norm}} = \frac{\rho}{\rho_{\text{ideal}}} = \frac{|\mathbf{w}^H \mathbf{h}|^2}{|\mathbf{h}^H \mathbf{h}/|\mathbf{h}^H \mathbf{h}||} = \frac{|\mathbf{w}^H \mathbf{h}|^2}{|\mathbf{h}^H \mathbf{h}|} \leq 1.$$
III. ESTIMATION OF BEAMFORMING VECTOR

In this section, we derive the beamforming vector $w$, $w^Hw = 1$, that maximizes the beamforming gain $\rho$ for a given training matrix $P$. Given the observed signal $y$, this problem can be formulated as finding

$$w = \arg \max_{g \in \mathbb{C}^M} E\left(\left|g^H h\right|^2 | y\right)$$

\[
\cong \arg \max_{g \in \mathbb{C}^M} g^H E\left(\left|hh^H y\right|\right) g,\]  

where $g$ is any $M \times 1$ vector with $g^H g = 1$, the expectation is over $h$ and $z$, and (a) follows from the fact that $g$ is deterministic for a given $y$. Since the constraint is given by $g^H g = 1$, the optimal beamforming vector is just the eigen-vector that corresponds to the maximum eigenvalue of $E(\tilde{h}h^H | y)$. Therefore, the remaining task is to find

$$E(\tilde{h}h^H | y).$$

The a-posteriori mean of $h$ is denoted as $E(h | y)$. As is well-known [21], it is also the minimum mean square error (MMSE) estimate of $h$ from $y$. When $h$ and $z$ are both Gaussian random vectors, $E(h | y)$ equals to the linear MMSE (LMMSE) solution, given by

$$\hat{h}_{LM} = E(h | y) = W_{LM} y,$$

where

$$W_{LM} = \Phi^H (P \Phi P^H + \sigma^2 I_N)^{-1}.$$

The error covariance of the MMSE estimate is

$$C = E\left((h - \hat{h}_{LM})(h - \hat{h}_{LM})^H\right) = E\left(hh^H | y\right) - \hat{h}_{LM} E\left(h | y\right) \hat{h}_{LM}^H + \hat{h}_{LM} \tilde{h}_{LM}^H$$

\[
= \Phi - \Phi \Phi^H W_{LM}. \]  

(6)

where (a) follows from the fact that $E(y(h - \hat{h}_{LM})^H) = 0$.

It is also known that the error covariance $C$ is independent of the observation signal $y$, that is,

$$C = E\left((h - \hat{h}_{LM})(h - \hat{h}_{LM})^H\right)$$

$$= E\left(hh^H | y\right) - \hat{h}_{LM} E\left(h | y\right) \hat{h}_{LM}^H + \hat{h}_{LM} \tilde{h}_{LM}^H$$

$$= E\left(hh^H | y\right) - \hat{h}_{LM} \tilde{h}_{LM}^H.$$  

(7)

Then, we have

$$E(\tilde{h}h^H | y) = E(hh^H | y) - E(h \hat{h}_{LM}^H) + \hat{h}_{LM}\tilde{h}_{LM}^H$$

$$= \Phi - \Phi \Phi^H W_{LM} + W_{LM} y (W_{LM} y)^H.$$  

(8)

When SNR and $N/M$ are not too small, for example, $\varepsilon/(N/M\sigma^2) > -10$ dB and $N/M \geq 0.3$, which is true in most practical scenarios, the estimation error term $C = \Phi - \Phi \Phi^H W_{LM}$ is small and has much smaller contribution to $E(\tilde{h}h^H | y)$ than $\hat{h}_{LM}\tilde{h}_{LM}^H$. Hence for typical applications, $E(\tilde{h}h^H | y)$ can be well-approximated as $\hat{h}_{LM}\tilde{h}_{LM}^H$, and the beamforming vector becomes

$$w \approx \hat{h}_{LM}/\sqrt{\tilde{h}_{LM}^H \hat{h}_{LM}}.$$  

(9)

This suggests that the beamforming gain achieved by using (9) will be very close to that with the optimal estimate in (3). Hereafter we focus on the LMMSE estimator thanks to its better tractability compared to the optimal solution.

IV. DESIGN OF TRAINING SEQUENCES $P$

In this section, we investigate the design of training matrix $P$ for both spatially uncorrelated and correlated channels. With the approximation given in (9) and $E(\tilde{h}h^H | y) \approx \hat{h}_{LM} \tilde{h}_{LM}^H$, the conditional mean of the beamforming gain as originally defined in (3) can be written as

$$\eta \triangleq w^H E(\tilde{h}h^H | y) w \approx \hat{h}_{LM}^H \tilde{h}_{LM}.$$  

Taking the expectation of $\eta$ with respect to $y$, we get

$$E_y(\eta) \approx \text{Tr}(\Phi \Phi^H (P \Phi P^H + \sigma^2 I_N)^{-1} P \Phi).$$  

(11)

On the other hand, for the LMMSE estimator $W_{LM}$, the MMSE is given by

$$e = \text{Tr}(C) = \text{Tr}(\Phi - \Phi \Phi^H (P \Phi P^H + \sigma^2 I_N)^{-1} P \Phi).$$  

(12)

Comparing (12) with (11), we see that statistically, maximizing the expectation of $\eta$ with respect to $P$ is approximately equivalent to minimizing the MMSE. Based on this relationship, the goal of this section is to design a training matrix $P$ that minimizes the MMSE in (12). The optimality of $P$ in this section is also in this sense.

A. A General Solution

Let the SVD of $\Phi$ be $\Phi = U_{\Phi} D_{\Phi} U_{\Phi}^H$, where $U_{\Phi} = [u_{\Phi1}, \ldots, u_{\Phi M}]$ is a unitary matrix and $D_{\Phi} = \text{diag}(d_{\Phi1}, \ldots, d_{\Phi M})$ is a diagonal matrix with non-negative elements. Let $\Phi^{1/2} = U_{\Phi} D_{\Phi}^{1/2}$ and $B = P \Phi^{1/2}$. Rewrite the MMSE in (12) as

$$e = \text{Tr}(\Phi) - \text{Tr}(D_{\Phi} B^H (BB^H + \sigma^2 I_N)^{-1} B).$$  

(13)

Let the compacted SVD of $B$ be $B = U_B D_B V_B^H$, where $U_B, D_B = \text{diag}(d_{B1}, \ldots, d_{BN})$, and $V_B$ are $N \times N, N \times N$ and $M \times N$ matrices, respectively. Then, (13) becomes

$$e = \text{Tr}(\Phi) - \text{Tr}(D_{\Phi} V_B B^H (U_B (D_B D_B^H + \sigma^2 I_N) U_B^H)^{-1} U_B D_B V_B^H)$$

$$= \text{Tr}(\Phi) - \text{Tr}(D_{\Phi} V_B (I_N + \sigma^2 (D_B D_B^H)^{-1})^{-1} V_B^H).$$  

(14)

To find the $P$ that minimizes $e$ in (14), we need the following lemma, with proof provided in Appendix A.

1This is verified by the simulation results to be presented in Section VII. In particular, for spatially uncorrelated channels, it can be shown that $w$ in (9) is the solution to (3) when the optimal $P$ derived in Section IV-C is adopted.
Lemma 1: For any given $M \times M$ positive definite matrix $A$ with rank$(A) \geq N$, $M \geq N$, the $N \times M$ matrix $X$ that maximizes $\text{Tr}(XAX^H)$ under the constraint of $\text{Tr}(XX^H) \leq c$ is given by $X = DXU_A^H$, where $D = \text{diag}(d_{X1}, \ldots, d_{XN})$ is a diagonal matrix with $\sum_{n=1}^N |d_{Xn}|^2 = c$ and $|d_{X1}| \geq |d_{X2}| \geq \cdots \geq |d_{XN}|$. $U_A = [u_{A1}, u_{A2}, \ldots, u_{AX}]$ and $u_{An}$, $n \in [1, N]$, is the eigenvector corresponding to the $n^{th}$ largest eigenvalue of $A$.

Note that if we let $(I_N + \sigma_z^2(D_BD_B^H)^{-1})^{-1} V_H^B$ be $X$ and $D_\Phi$ be $A$ in Lemma 1, we can see that the MMSE in (14) is minimized when $V_H^B$ is the eigenvector matrix corresponding to the $N$ largest eigenvalues of $D_\Phi$.

For the diagonal matrix $D_\Phi$, with elements $d_{\phi1} \geq d_{\phi2} \geq \cdots \geq d_{\phiM} > 0$, its typical singular vector matrix is known as an $M \times M$ identity matrix. However, when $\Phi$ has multiple identical singular values, there are actually some unnoticed eigenvector matrices other than the well-known identity matrix, which are rarely discussed in the literature. These extra eigenvector matrices lead to better solutions under per-node power constraint than the identity matrix as to be seen next.

For an identity matrix, we can easily see that any orthonormal matrix can be its eigenvector matrix. For a general diagonal matrix $D_\Phi$, let us first consider the case that $q$ consecutive diagonal elements in $D_\Phi$ are identical, that is, $d_{\phi n} = d$, $n \in [n_1, n_1 + q - 1]$. Let $U_D$ be an $M \times M$ block diagonal matrix of the form

$$U_D = \begin{bmatrix} I_{M-n_1-1} & 0 & 0 \\ 0 & U_{D1} & 0 \\ 0 & 0 & I_{M-n_1-q+1} \end{bmatrix},$$

where $U_{D1}$ is any $q \times q$ orthonormal matrix. We have

$$U_D D_\Phi U_D^H = D_\Phi,$$

which implies that $U_D$ is also one of the eigenvector matrices of $D_\Phi$.

Therefore $V_H^B$ equals the submatrix of $U_D$, $U_{DU}$ which contains its first $N$ rows, and we have

$$V_H^B = U_{DU}.$$

We can then rewrite $B$ as

$$B = U_B D_B U_{DU} = P U_\Phi D_\Phi^{1/2},$$

which leads to

$$P = U_B D_B U_{DU} D_\Phi^{-1/2} U_H^B.$$

Note that $D_B$ is the only variable matrix in (19). Using (17), (14) can be represented as

$$e = \text{Tr}(\Phi) - \text{Tr}\left(D_\Phi (I_N + \sigma_z^2(D_B D_B^H)^{-1})^{-1}\right)$$

$$= \sum_{n=1}^N d_{\phi n} \sum_{n=1}^N \frac{d_{\phi n}}{1 + |d_{Bn}|^2}$$

$$= \sum_{n=N+1}^M d_{\phi n} + \sum_{n=1}^N \frac{d_{\phi n}^2}{|d_{Bn}|^2 + \sigma_z^2},$$

(20)

where $D_{\Phi L} = \text{diag}(d_{\phi1}, d_{\phi2}, \ldots, d_{\phiN})$.

According to (19), minimization of $e$ in (20) with respect to $P$ is equivalent to with respect to $D_B$. The optimal solution is presented in the following theorem, with proof provided in Appendix B.

Theorem 1: Under the total energy constraint $\text{Tr}(P^H P) \leq \varepsilon$, a general form of the optimal training sequences $P$ is given by

$$P = U_B D_B U_{DU} D_\Phi^{-1/2} U_H^B,$$

where both $U_{D1}$, the submatrix of $U_D$ in (15), and $U_B$ can be chosen as any orthonormal matrices, the power of the $n^{th}$ diagonal element in $D_B$, $|d_{Bn}|^2$, is determined by

$$|d_{Bn}|^2 = \varepsilon d_{\phi n}/N + \sigma_z^2 \left(\frac{d_{\phi n}}{\sum_{m=1}^{N'} |d_{Bm}|} - 1\right).$$

(22)

In (22), $N'$ is the maximum integral value of $q$, satisfying

$$\frac{\varepsilon}{\sigma_z^2} > \frac{q}{d_{\phi n}} - \sum_{m=1}^q \frac{1}{|d_{Bm}|}, q \in [1, N].$$

(23)

Theorem 1 indicates that given $\varepsilon$, $\sigma_z^2$ and $\Phi$, there exists an optimal value for the length of training sequences. This is consistent with the results in the power allocation literature, where water filling algorithms are applied to determine the power allocation.

When $D_\Phi$ has multiple segment of identical diagonal elements, $U_D$ in (15) can be modified to accommodate multiple ortho-normal matrices, and the same optimization process applies.

Remark 1: Theorem 1 enables the design of optimal training matrices which can distribute training energy evenly over $M$ nodes for channels with any correlation matrix $\Phi$. For DTB systems with per-node power constraint, the matrix $U_{D1}$ and $U_B$ in Theorem 1 should be chosen as those having least zero elements, for example, Hadamard matrices. It is noted that, however, the resulted $P$ does not guarantee equal power allocation, hence may be not optimal under the per-node power constraint.

B. A Typical Solution

Let $U_{D1}$ in (15) be an identity matrix, we get

$$P = U_B D_B U_{DU} D_\Phi^{-1/2} U_H^B = U_B D_P U_{DU} U_H^B,$$

(24)

where $D_P = \text{diag}(d_{P1}, \cdots, d_{PN}) = D_B D_{\Phi L}^{-1/2}$, and $U_{\Phi L} = [u_{\Phi1}, \cdots, u_{\PhiN}]$ is an $M \times N$ submatrix of $U_\Phi$. Note that $U_B$ can be any $N \times N$ orthonormal matrix.

From Theorem 1, we can derive the following corollary:

Corollary 1: For a given $\Phi$, one optimal $P$ under the total energy constraint is given by (24), where

$$|d_{Pn}|^2 = \varepsilon/N' + \sigma_z^2 \left(\frac{1}{N'} \sum_{m=1}^{N'} \frac{1}{|d_{Bm}|} - 1\right).$$

(25)

with $N'$ determined similarly as in Theorem 1.
design of training sequences under per-node power constraint when $\Phi$ has several identical eigenvalues. For example, when only the $(1,2)^{th}$ and $(2,1)^{th}$ non-diagonal elements of $\Phi$ are non-zero, the optimal $P$ determined according to Corollary 1 will only have at most two non-zero values in each row. This implies that only two nodes transmit non-zero signals regardless of the total training power.

When $\varepsilon/N \gg \sigma_N^2 (\sum_{m=1}^{N} (N_d q_m)^{-1} - d_m^{-1})$, that is, when the SNR is large and the ratio $d_{q1}/d_{qN}$ is not too large, $|d_{Pm}|^2$ in (25) can be well-approximated by $\varepsilon/N$. Thus we get a simplified solution without requiring water filling optimization.

C. Solution for Spatially Uncorrelated Channels

For spatially uncorrelated channels, $d_{qn} = 1, n \in [1, M]$. From (19), $P$ is given by the product of several ortho-normal matrices. Hence we get the following corollary, which is also directly proven from (12) in Appendix C:

**Corollary 2:** For $\Phi = I_M$, the optimal training sequence for $N < M$ is given by any $P$ satisfying $PP^H = \varepsilon/N I_N$ under either total energy or per-node power constraints. The optimal receiver is as simple as a Matched Filter (MF). The optimal beamforming vector is given by $P^H y / \sqrt{y^H P y}$.

V. AVERAGE BEAMFORMING GAIN ANALYSIS

In this section, we analyze the average beamforming gain achieved by the DTB system. Depending on the knowledge on channel correlation matrix $\Phi$ and affordable complexity, we investigate four schemes where different training matrices and beamforming vector estimators are chosen as follows:

S1. Any pre-designed training matrix $P$ satisfying $PP^H = \varepsilon/N I_N$ and the MF estimator $W = P^H$;

S2. Any pre-designed $P$ satisfying $PP^H = \varepsilon/N I_N$ and the MMSE estimator $W = W_{\text{MMSE}}$;

S3. Channel-dependent training matrix $P = \sqrt{\varepsilon/N} U_B U_{\Phi L}^H$ for $\Phi \neq I_M$ and the MF estimator $W = P^H$; and

S4. Channel-dependent optimal training matrix $P = U_B D_P U_{\Phi L}^H$ and the MMSE estimator $W = W_{\text{MMSE}}$, where $D_P$ is given in Section IV-B.

S1 has the smallest implementation complexity and feedback requirement, and is optimal when $\Phi = I_M$. S2 has the same feedback requirement as in S1, and has an improved beamforming gain over S1 when the channel coefficients become spatially correlated. Yet it has a higher computational complexity in the destination node. Both S3 and S4 have much higher complexity, requiring eigenvalue decomposition and feeding back $P$ from the destination node to all source nodes. S3 has lower computational complexity than S4, without requiring matrix inversion and water-filling power allocation, at the cost of a small loss in beamforming gain, as we will see later.

We next analyze the average beamforming gains for Schemes 1, 3 and 4. A closed-form performance expression for Scheme 2 was not obtained, and its performance will be demonstrated numerically. For simplicity, we will use the results in Section IV-B for $\Phi \neq I_M$. Note that the averaged ideal beamforming gain $\rho_{\text{ideal}}$ as defined in (2) is $M$.

A. Scheme 1

When $\Phi = I_M$, the beamforming gain achieved by S1 is given by

$$\rho_1 = \frac{|h^H P H y|^2}{y^H P P^H y}$$

$$= \frac{N}{\varepsilon} \left( \frac{|h^H P H z|^2}{h^H P H z^H P H z} + 2 \text{Re}(h^H P H z^H z) \right)$$

$$\approx \frac{N}{\varepsilon} \left( \frac{|h^H P H z|^2}{h^H P H z^H P H z + z^H z} \right),$$

where the approximation is based on that the cross-correlation term $|h^H P H z|^2$ is much smaller than the product of power term $(h^H P H z)^2$ and the sum of power terms $(h^H P H z + z^H z)$. Averaging $\rho_1$ with respect to $h$ and $z$, we can have the following approximation in the case of $\varepsilon/N \gg \sigma^2_z$:

$$E(\rho_1) \approx \frac{N}{\varepsilon} (\text{Tr}(P \Phi P^H) - N \sigma^2_z)$$

(when $\Phi = I_M$) $N(1 - N \sigma^2_z/\varepsilon)$. (26)

**Remark 2:** The result in (26) indicates that for $\Phi = I_M$ and a fixed $\varepsilon$, the average beamforming gain is independent of the number of total nodes $M$. This implies that using $M$ nodes and solving the under-determined equations in (1) can achieve the same beamforming gain as in the case where $N$ nodes use an $N \times N$ training matrix. Using $M$ nodes, however, achieves other benefits in practical implementations. For example, the average energy consumption of each node is reduced. In addition, the concatenated training scheme as to be presented in Section VI can be implemented to consistently improve beamforming gain.

**Remark 3:** From (27), we can also see that the average beamforming gain is approximately proportional to $N$ at a high SNR. Indeed, differentiating $N(1 - N \sigma^2_z/\varepsilon)$ w.r.t $N$, we get $1 - 2N \sigma^2_z/\varepsilon$. When $\varepsilon/(N \sigma^2_z) > 2$, which is true in most of practical systems, $N(1 - N \sigma^2_z/\varepsilon)$ is a monotonically increasing function of $N$.

**Remark 4:** Based on (26), we can deduce another interesting result. For S1, whether the beamforming gain increases when channels become correlated depends on the actual correlation matrix $\Phi$ and the training matrix $P$.

To see this, referring to (26), we compare the difference of $\text{Tr}(P \Phi P^H)$ between a general $\Phi$ and a $\Phi = I_M$:

$$\text{Tr}(P \Phi P^H) - \text{Tr}(P P^H) = 2 \sum_{m=1}^{M} \sum_{n=1}^{m-1} \phi_{m,n} \text{Re}(p_n H p_m),$$

(28)

where $p_m$ and $p_n$ are the $m^{th}$ and $n^{th}$ columns in $P$, respectively. Since the sign of $\text{Re}(p_n H p_m)$ depends on $P$, (28) shows the dependence of the beamforming gain on $\Phi$ and $P$. For a given $\Phi \neq I_M$, we may refer to (28) to choose a suitable $P$, to ensure beamforming gain increases with channel correlation coefficients.
B. Scheme 3

For S3, where \( \mathbf{P} = \sqrt{\frac{\varepsilon}{N}} \mathbf{U}_B \mathbf{U}_{\phi_1}^H \) and \( \mathbf{\Phi} \neq \mathbf{I}_M \), the beamforming gain can be similarly derived as in S1, and we can obtain (26) in the case of \( \varepsilon/N \gg \sigma_z^2 \). Therefore we get the high-SNR approximation

\[
E(\rho_3) \approx \sum_{n=1}^{N} d_{\phi n} - N\sigma_z^2 / \varepsilon. \tag{29}
\]

Comparing (29) with (27), we can see that \( \sum_{n=1}^{N} d_{\phi n} \geq N \) and thus \( E(\rho_3) \geq E(\rho_1) \). This means that taking into account the channel correlation in designing the training matrix can always improve the beamforming gain, even when a suboptimal MF is used to estimate the beamforming vector. Note that (29) and (27) become the same when \( N = M \).

C. Scheme 4

In this case, we have \( \mathbf{P} = \mathbf{D}_p \mathbf{U}_{\phi L}^H \) and \( \mathbf{W}_{LM} = \mathbf{U}_{\phi L} \mathbf{D}_p^{-1} \mathbf{\Lambda} \), where \( \mathbf{\Lambda} \triangleq \left( \mathbf{I}_N + \frac{\varepsilon}{N} |\mathbf{D}_p|^{-2} \mathbf{D}_{\phi L}^2 \right)^{-1} \) is an \( N \times N \) diagonal matrix. The beamforming gain can be represented as in (30) at the bottom of this page, where the approximation is obtained by ignoring the cross-correlation terms between \( \mathbf{h} \) and \( \mathbf{z} \).

Since each diagonal element in \( \mathbf{\Lambda} \) is positive and is smaller than 1, we have

\[
\mathbf{h}^H \mathbf{U}_{\phi L} \mathbf{\Lambda}^2 \mathbf{U}_{\phi L}^H \mathbf{h} < \mathbf{h}^H \mathbf{U}_{\phi L} \mathbf{\Lambda} \mathbf{U}_{\phi L}^H \mathbf{h}. \tag{31}
\]

Then we get a lower bound for \( \rho_4 \) as

\[
\rho_4 > \frac{(\mathbf{h}^H \mathbf{U}_{\phi L} \mathbf{\Lambda} \mathbf{U}_{\phi L}^H \mathbf{h})^2}{\mathbf{h}^H \mathbf{U}_{\phi L} \mathbf{\Lambda} \mathbf{U}_{\phi L}^H \mathbf{h} + \mathbf{z}^H |\mathbf{D}_p|^{-2} \mathbf{\Lambda}^2 \mathbf{z}}
= \frac{(\mathbf{h}^H \mathbf{U}_{\phi L} \mathbf{\Lambda} \mathbf{U}_{\phi L}^H \mathbf{h})(\mathbf{h}^H |\mathbf{D}_p|^{-2} \mathbf{\Lambda}^2 \mathbf{z})}{\mathbf{h}^H \mathbf{U}_{\phi L} \mathbf{\Lambda} \mathbf{U}_{\phi L}^H \mathbf{h} + \mathbf{z}^H |\mathbf{D}_p|^{-2} \mathbf{\Lambda}^2 \mathbf{z}}.
\]

For \( \varepsilon/N \gg \sigma_z^2 \), a lower bound for the statistical average of \( \rho_4 \) can be found as

\[
E(\rho_4) \geq \sum_{n=1}^{N'} \frac{d_{\phi n}}{1 + \sigma_z^2 |d_{\phi n}|^2}. \tag{32}
\]

VI. EXTENSION TO CONCATENATED TRAINING

From the analytical results in the preceding section, we see that in the medium to high SNR region, beamforming gain increases rapidly with \( N \), until \( N \) approaches \( M \). This motivates us to improve the beamforming gain by combining multiple training matrices received in different frames in slow time varying channels. Such a mechanism is referred to as concatenated training.

The concatenated training scheme is illustrated in Fig. 1. Each frame consists of a data block and a training block. Let \( \mathbf{P}_k \), \( k = 1, 2, \cdots \) of size \( N_s \times M \) denote the training matrix in the \( (k-1) \)th frame. The training matrix \( \mathbf{P}_k \) is attached to the end of the \( (k-1) \)th data block, in order to minimize the mismatch between the estimated and the actual channels due to time lapse. By combining several latest received training blocks, the beamforming vector \( \mathbf{w}_k \) is determined and then used for data transmission in the \( k \)th frame.

We assume channels change smoothly over time, and within \( K \) frames, the channel variation is reasonably small so that the training matrices can be combined to improve beamforming gain. Here, we focus on designing the concatenated training scheme without exploiting the channel temporal correlation\(^2\).

Consider the combination of received training signals from the \((k-K) \)th to \((k-1) \)th frames. The beamforming vector in the \((k-1) \)th frame is obtained by estimating one single vector \( \mathbf{h}_k \) which approximates the channels for the \( K \) frames. Small differences between these channels and \( \mathbf{h}_k \) is absorbed into the noise terms. The received training signals in the \( K \) frames can be represented as

\[
\mathbf{y}_{k'} = \mathbf{P}_k \mathbf{h}_k + \mathbf{z}_{k'}, \tag{33}
\]

where \( k' = k - K + 1, k - K + 2, \cdots, k \), the noise term \( \mathbf{z}_{k'} \) absorbs the channel difference and the elements in \( \mathbf{z}_{k'} \) are assumed to be i.i.d. \( CN \) random variables. Let \( N_s = \sum_{k'=k-K}^{K} N_{k'} \), and \( N_s < M \). Notice that the concatenated training scheme is also applicable to the case of \( N_s \geq M \). Since the beamforming gain increases slowly after \( N_s \geq M \), we limit our discussions to the case of \( N_s < M \), to be consistent with the preceding scenario. In this section, we study the design of the concatenated training scheme in spatially uncorrelated channels. Our results can be extended to spatially correlated channels.

A. Training Sequence Design

Referring to (33), spreading training sequences of length \( N_s \) to \( K \) frames in concatenated training is equivalent to a block training using sequences of length \( N_s \). Hence results for block training, such as the principle of training sequences design, can be applied to the concatenated training case. For spatially uncorrelated channels, the \( N_s \times M \) training matrix \( \mathbf{P}_k \) shall be a submatrix of an \( M \times M \) orthogonal matrix \( \mathbf{T} \).

To enable simple and iterative beamforming vector update, we propose to generate \( \mathbf{P}_k \) from \( \mathbf{T} \) cyclically. Supposing the last row in \( \mathbf{P}_{k-1} \) is the \( q \)th row of \( \mathbf{T} \), \( \mathbf{P}_k \) will have the rows of \( \mathbf{T} \) with the indexes \( \{q + 1, q + 2, \cdots, q + N_k\} \), for \( q \leq M - N_k \) or \( \{q+1, \cdots, M, 1, \cdots, N_k + q - M\} \), for \( q > M - N_k \). Mathematically, the \( p \)th row of \( \mathbf{P}_k \), \( p \in [1, N_k] \) is the \((\text{mod}(q + p, M))\)th row of \( \mathbf{T} \), where \( \text{mod}(x, y) \) denotes

\(^2\)The optimal design of concatenated training scheme, which takes into account the temporal correlation, will be investigated in our future work.
the operation of $x$ modulo $y$. In this way, orthogonal training sequences of length $N_s$ are always available at the time of updating beamforming vector in each frame. The $N_s$ latest received training symbols can be combined to achieve beamforming gain equivalent to that obtained by transmitting block training sequences of length $N_s$.

B. Update beamforming vector

With the training blocks proposed in Section VI-A, the beamforming vector can be computed iteratively. Combining $y_{k'}, k' = k - K + 1, k - K + 2, \ldots, k$ in (33) and using the matched filter estimator, the estimate for $h_k$ is obtained as

$$\hat{h}_k = \sum_{k'=k-K+1}^{k} P^H_{k'} y_{k'}.$$  \hfill (34)

Recall that only $K$ training blocks are combined. When a new training block $P_{k_0 + K + 1}$ is received, the channel estimate can be recursively updated as

$$\hat{h}_{k+1} = \sum_{k'=k-K+2}^{k+1} P^H_{k'} y_{k'} = \hat{h}_k + P^H_{k+1} y_{k+1} - P^H_{k-K+1} y_{k-K+1}.$$  \hfill (35)

The achieved average beamforming gain is thus proportional to $N_s$ and remains stable if new training matrices are received periodically and channel temporal correlation is unchanged.

C. Feedback Requirement

For S1 and S2, the destination node only needs to feedback either the estimated $M \times 1$ beamforming vector or the received signal. The choice depends on the available feedback bandwidth and affordable computational complexity. Feeding back beamforming vector can avoid the computation in the source nodes, particularly when MMSE estimator is implemented. As each source node can be identified and numbered via the training matrices, the $m^{th}$ element in the beamforming vector is for the $m^{th}$ source node. Thus no extra information other than the beamforming vector needs to be sent. Alternatively, the destination node may feedback $N_k$ symbols by sending the $N_k \times 1$ vector $y_k/\sqrt{y_k^H y_k}$ in the block training case or $y_k$ in the concatenated training case. Consider that the MF is used to determine the beamforming vector. For block training, the $m^{th}$ source node can obtain its beamforming weight by

$$w_m = P^H_m (y_k/\sqrt{y_k^H y_k}),$$

where $p_m$ is the $m^{th}$ column vector in $P$. For concatenated training, the source nodes can compute its beamforming weight as

$$\hat{h}_{m,k} = \hat{h}_{m,k-1} + p_m y_k - p_{m,k-K} y_{k-K},$$

$$w_{m,k} = \frac{\hat{h}_{m,k}}{\sqrt{\hat{h}_{m,k}^H \hat{h}_{m,k}}}.$$  

Setting $N = 1$ in our concatenated training scheme, it becomes similar to the one-bit feedback scheme [3], [5], [6] in terms of the frame structure. However, our scheme shows three advantages: faster training, higher robustness and smaller total feedback requirement. The first two advantages are to be detailed in Section VII. For the feedback requirement, it is shown numerically in [3], [5], [6] that the iterations needed for convergence are generally more than 10 times of the number of the source nodes in noiseless channels. Therefore, the one-bit feedback scheme requires in total more than $10M$ feedback bits. Referring to the signal feedback approach in our scheme as discussed above, 10 bits are more than enough for quantizing one received symbol $y_k$ in each frame. Therefore, although our scheme requires more bits for quantized $y_k$ in one feedback, it needs much less total bits for achieving any given beamforming gain.

VII. Numerical Results

In this section, we present simulation results for the proposed DTB schemes in Rayleigh fading channels, with and without spatial correlations. In all simulations, $M = 16$ source nodes are used. Presented results will be concentrated on total energy constraint, with $\varepsilon/\sigma_z^2$ denoting the level of constraints. It can be translated to the per-node SNR $\varepsilon/(MN\sigma_z^2)$, and the results for the per-node power constraint in spatially uncorrelated channels can be inferred accordingly.

We first present numerical results for block training cases where training matrix $P$ is of size $N \times M$. These results are directly applicable to concatenated training should channels remain almost unchanged during the combination period.

In Fig. 2, we plot the normalized beamforming gains achieved by various DTB schemes for $\varepsilon/\sigma_z^2 = 19$ dB. The training matrices under consideration include a randomly generated matrix with i.i.d Gaussian coefficients, a Hadamard matrix and a discrete Fourier transform (DFT) matrix. We see that for the two types of orthogonal training matrices, almost identical beamforming gains are achieved. In addition, for orthogonal training matrices, the MF (MMSE) and optimal estimators in (8) almost achieve the same normalized gain,
which is approximately proportional to \( N/M \). These observations are consistent with the analytical results in Section V-A. When random training matrix is used, the beamforming gain is significantly degraded. The Hadamard matrix will be used as the training sequences for Scheme 1 and 2 hereafter.

Fig. 3 shows the normalized beamforming gain versus per-node SNR for \( N = 3 \) and \( N = 10 \). The beamforming gain increases rapidly in the low SNR region, for example, from \(-20 \) dB to \(-5 \) dB, and are flatten out when the SNR becomes large. Our derived analytical beamforming gains in (27) are also plotted, and match the numerical results well for SNRs higher than \( 0 \) dB. At lower SNRs, they become inaccurate due to the high-SNR assumption used in the derivation. For comparison, the beamforming gain achieved by the one-bit feedback scheme [3], with a perturbation value \( \pi/20 \), is also presented. The curve shows the normalized beamforming gain at the \( 320 \)-th iteration, which is equivalent to \( N = 320 \) in our proposed scheme. Unlike our proposed scheme, which achieves stable beamforming gain even at very low SNRs, the one-bit feedback scheme has very low efficiency in the simulated SNR range. Our proposed scheme achieves much higher gain with much less training, for example, more than 6 times of beamforming gain with 30 times less training period at per-node SNR of \( 5 \) dB.

Figs. 4 to 6 present results for spatially correlated channels, where the correlation matrix \( \Phi \) has all 1 diagonal elements and each of non-diagonal elements is uniformly distributed over \([0, \delta]\). Fig. 4 compares the normalized beamforming gain achieved by the four schemes studied in Section V. It is observed that beamforming gain is significantly improved when channel correlation is exploited, e.g., in S2, S3 and S4. The closed-form analytical results in (32) and (29) agree with the numerical results. The normalized beamforming gain versus per-node SNR is shown in Fig. 5. Similar to what we have observed from Fig. 3, beamforming gain increases more rapidly at lower SNR than at higher SNR. S4 clearly outperforms S3 in low SNR region. The lower bound in (32) is tight and the analytical gain in (29) is in good match with the numerical one when the per-node SNR is larger than \(-10 \) dB. The analytical gain becomes inaccurate at very low SNRs due to the same approximation error as the theoretical gain for Scheme 1. In Fig. 6, the normalized beamforming gain versus the correlation parameter \( \delta \) is presented. The figure indicates that beamforming gain increases rapidly as the channel becomes more correlated.

From Figs. 2 to 6, we see that the gain achieved by the MMSE solution closely approaches the optimal gain obtained by directly solving (8). This consolidates the effectiveness of the results derived based on the LMMSE estimator.

The concatenated training scheme is also simulated in time-varying Rayleigh fading channels, which is generated by the Jake’s model [22]. The temporal autocorrelation function of each channel is characterized by the zeroth-order Bessel function of the first kind \( R(\tau) = J_0(2\pi f_d \tau) \), where \( f_d \) is the
maximum Doppler shift. $N = 2$ sequences are transmitted in each training. The discrete channels for training blocks are sampled at the interval of $\tau = 224 \mu s$, to emulate a system with a sampling rate of 2 $\mu$s, and $T_d = 200 \mu s$, $T_p = 24 \mu s$ in Fig. 1. The correlation coefficients corresponding to the sampling setup is shown in Fig. 7. Fig. 8 shows how the normalized beamforming gain varies with the value of $K$ and the Doppler shift which can be translated into the correlation coefficients in Fig. 7. By combining received training signals from multiple frames, we see the concatenated scheme efficiently reduces the training cost, while achieving good beamforming gain over a large range of Doppler shifts.

VIII. CONCLUSIONS

In this paper, we studied the design of DTB systems for the under-determined case, where the length of the training sequences is smaller than the number of source nodes, i.e., $N < M$. We investigated the design of optimal training sequences and estimators. Four exemplified schemes, which achieve different tradeoffs between implementation complexity, feedback requirement and beamforming gain, have been proposed for spatially uncorrelated and correlated channels. We analytically showed that at a sufficiently high SNR, the proposed scheme can achieve a beamforming gain of about $N$ for spatially uncorrelated channels. We also showed that the channel correlation can benefit to the achievable beamforming gain, by employing our designed training matrices. Our derivation also shows that, by allowing all $M$ source nodes to participate in the training and beamforming vector estimation, the power consumption of each source node can be made very small, which is attractive for real implementations such as a low power wireless sensor network. We also proposed a concatenated training scheme, where carefully designed training sequences are distributed over multiple frames cyclically and combined in the receiver. The scheme is shown to improve beamforming gain significantly in slow time varying channels...
even when only one training bit is sent from each node in a frame. Based on this work, training symbols can be flexibly allocated as pilots distributed in the data signal, and different number of pilots, determined according to the channel correlation time, can be efficiently combined in the receiver using the schemes proposed here. The work in the paper can be enriched from various aspects. For example, joint design of estimation and quantization for the beamforming vector, and improvement of concatenated training scheme in time-varying channels are worthy of further investigations.

**APPENDIX A**

**Proof of Lemma 1**

**Proof:** For any complex matrix X in \( \mathbb{C}^{N \times M} \), \( M \geq N \), we can always represent it as the weighted summation of \( N \) out of \( M \) basis vectors in the space of \( \mathbb{C}^{M \times M} \). That is, \( X = D_X U_X \), where \( U_X = (u_{X1}, \ldots, u_{XN}) \), \( u_{Xn} \) is the \( n^{th} \) \( M \times 1 \) orthonormal basis vector, and \( D_X = \text{diag}(d_{X1}, \ldots, d_{XN}) \) is a \( N \times N \) diagonal matrix. Without loss of generality, we assume \( |d_{X1}| \geq |d_{X2}| \geq \cdots \geq |d_{XN}| \). The optimization problem can then be formulated as

\[
\max_{D_X, U_X} \text{Tr}(D_X U_X^H B U_X D_X^H) \\
\text{subject to } \text{Tr}(D_X D_X^H) = c.
\]

Rewrite

\[
\text{Tr}(D_X U_X^H A U_X D_X^H) = \sum_{n=1}^{N} |d_{Xn}|^2 u_{Xn}^H A u_{Xn}.
\]

Since \( A \) is a positive definite matrix, \( |d_{Xn}|^2 u_{Xn}^H A u_{Xn} \geq 0 \). Since \( |d_{Xn}|^2 \) decreases with \( n \) increasing, it is well known that the term \( \sum_{n=1}^{N} |d_{Xn}|^2 u_{Xn}^H A u_{Xn} \) is maximized when \( u_{Xn} \) is the eigenvector of \( A \), corresponding to the \( n^{th} \) largest eigenvalue.

The value of \( d_{Xn} \) can then be determined following the condition \( \sum_{n=1}^{N} |d_{Xn}|^2 = c \) and any other constraints on \( X \).

**APPENDIX B**

**Proof of Theorem 1**

**Proof:** From (19), the total power constraint condition can be rewritten as

\[
\text{Tr}(PP^H) = \text{Tr} \left( U_D U_D^H \Phi^{-1} U_{DU}^H B^H B \right) = \text{Tr} \left( D_{\Phi}^{-1} B^H B \right) = \sum_{n=1}^{N} \frac{|d_{Bn}|^2}{d_{\Phi n}} \leq \varepsilon,
\]

(36)

where the second equality follows from the inversion of \( D_{\Phi} \) in (16).

According to (20) and (36), we can formulate the following optimization problem with respect to \( |d_{Bn}|^2 \)

\[
\min_{|d_{Bn}|^2} \sum_{n=1}^{N} \frac{d_{\Phi n}}{|d_{Bn}|^2 + \sigma_z^2} \\
\text{subject to } \sum_{n=1}^{N} \frac{|d_{Bn}|^2}{d_{\Phi n}} \leq \varepsilon.
\]

Using the Lagrange multiplier function, the optimal solution for (37) is found to satisfy

\[
\frac{|d_{Bn}|^2 + \sigma_z^2}{d_{\Phi n}} = c, \ \forall n \in [1, N], \ \text{and} \ \sum_{n=1}^{N} \frac{|d_{Bn}|^2}{d_{\Phi n}} = \varepsilon, \quad (38)
\]

where \( c \) is a constant. This leads to the expression of \( |d_{Bn}|^2 \) in (22). Since \( |d_{Bn}|^2 > 0 \), the length of training sequences \( N' \), \( N' \leq N \), is determined as the maximal integer \( q \) that satisfies the condition (23).

**REFERENCES**


Jian (Andrew) Zhang (M’04, SM’11) received the B.S. degree from Xi’an JiaoTong University, China, in 1996, the M.Sc. degree from Nanjing University of Posts and Telecommunications, China, in 1999, and the Ph.D. degree from the Australian National University, in 2004.

Currently, Dr. Zhang is a senior research scientist in the ICT Centre, CSIRO, Sydney, Australia. From 1999 to 2001, he worked as a system and hardware engineer in ZTE Corp, Nanjing, China. From 2004 to 2010, he was a researcher in the Networked Systems, NICTA, Canberra, Australia. He is an adjunct honorary fellow in the Macquarie University and University of South Australia. Dr. Zhang’s research interests are in the area of signal processing for wireless communications, with focus on MIMO, multicarrier, ultra wideband and sensor networks. He has published more than 60 papers on leading international Journals and conferences. He is a recipient of CSIRO Chairman’s Medal and the Australian Engineering Innovation Award in 2012 for exceptional research achievements in multi-gigabit wireless communications.

Tao Yang (S’07, M’10) received B.Sc. degree in electronic engineering in 2003 from Beijing University of Aeronautics and Astronautics (Beihang University), Beijing, China. He received Master by research and Ph.D degrees in electrical engineering from the University of New South Wales, Sydney, Australia, in 2006 and 2010, respectively. He is currently a Research Fellow in the Wireless and Networking Technologies Laboratory (WNTL) at Commonwealth Scientific and Industrial Research Organization (CSIRO), Sydney, Australia. His research expertise and interests include multi-user and MIMO communications, error-control coding, iterative signal processing and decoding, physical-layer network coding and network information theory. He was the recipient of Australian Postgraduate Award (APA), NICTA research project award (NRPA) and Supplementary Engineering Award (SEA) from the University of New South Wales.

Zhuo Chen (S’01-M’05) received the B.S. degree in electrical engineering from Shanghai Jiao Tong University, Shanghai, China, in 1997 and the M.S. and Ph.D. degrees from the School of Electrical and Information Engineering, University of Sydney, Sydney, Australia, in 2001 and 2004, respectively.

He is currently a Senior Research Scientist with the Wireless and Networking Technologies Laboratory, Commonwealth Scientific and Industrial Research Organization (CSIRO), Sydney. He is the Team Leader for the Digital Communications Group, and a Project Leader for several industrial projects. His research interests include wireless communications, multiple-input-multiple-output systems, error control coding techniques, and wireless sensor networks.