# Data-Dependent Channel Estimation and Superimposed Training Design in Amplify and Forward Relay Networks

Gongpu Wang†, Feifei Gao*, Guoqing Li‡, and Chintha Tellambura†
†Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada
*Department of Automation, Tsinghua University, Beijing, China
‡China Telecommunication Technology Labs, Beijing, China
Email: †{gongpu, chintha}@ece.ualberta.ca

***Abstract*—In this paper, we apply the data-dependent superimposed training (DDST) in *amplify-and-forward* (AF) relay networks with cyclic-prefix single carrier (CPSC) modulation. We consider various issues such as channel estimation, training design and data detection. A sub-optimal training sequence that can minimize the upper bound of the mean square error of the estimator is derived. Since the DDST estimator can only find the overall channel information, we further propose a doubly cooperative estimator (DCE) to track the individual channel knowledge at the cost of some performance loss. Simulations are then provided to corroborate the proposed studies.**

## I. Introduction

Accurate estimation of channel state information (CSI) is often required for successful data transmission for any communication systems. Generally, CSI is obtained via separately sending the pilot symbols to the receivers, which is also known as pilot symbol assisted modulation (PSAM) [1].

An alternative approach is to superimpose pilot symbols onto data symbols such that the bandwidth efficiency is improved. The idea of superimposed pilots first appeared in [2] for analog communication systems, and were further exploited recently for both synchronization and channel estimation in digital communication systems [3]–[6].

In [3], the data and the noise were assumed as zero-mean sequences, and the receiver used the first-order statistics for channel estimation. A frequency-domain estimation method in the cases when the noise cannot be assumed as zero-mean was suggested in [4]. However, the performance of channel estimation can not compete with the performance by traditional PASM [1] since the first-order statistics cannot fully cancel the influence of unknown data and noise on channel estimation. To improve both estimation and detection, the data-dependent superimposed training (DDST) was proposed in [5]. In DDST the transmitter sends data together with an auxiliary sequence such that the DFT of the combined data and the sequence will be zeros at certain frequencies. By further superimposing training during the transmission, both the channel estimation and data detection can be improved.

In this paper, we apply DDST for *amplify-and-forward* (AF) relay networks. Cyclic-prefix single carrier (CPSC) modulation is adopted to cope with the frequency-selective environment. Based on minimizing the upper bound of channel estimation MSE, the optimal training design is derived. To obtain the separate channel knowledge from source to relay and from relay to destination, we further propose a doubly cooperative estimator (DCE). Finally, simulations are provided to corroborate the proposed studies.

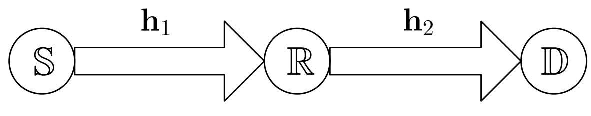


Fig. 1. System configuration for a three-node relay network.

## II. System Model

Consider a relay system with one source node $\mathbb{S}$, one relay node $\mathbb{R}$, and one destination node $\mathbb{D}$. Let $\mathbf{h}_1 = [h_{1,0}, h_{1,1}, \cdots, h_{1,L-1}]^T$ denote the channel between $\mathbb{S}$ and $\mathbb{R}$, and $\mathbf{h}_2 = [h_{2,0}, h_{2,1}, \cdots, h_{2,L-1}]^T$ the channel between $\mathbb{R}$ and $\mathbb{D}$, as shown in Fig. 1. Suppose $\mathbf{s}$ is the $(N\times 1)$ data vector to be transmitted, and $\mathbf{p}$ is the $(N\times 1)$ superimposed training vector with period $P \geq 2L-1$. Without loss of generality, we assume $N = PQ$ for some integer $Q$. Note that the energy of the periodic sequence $\mathbf{p}$, is only concentrated at the $P$ equi-spaced frequency bins $[0, Q, \cdots, (P-1)Q]$, which are referred to as pilot frequencies by [5].

DDST suggests that the transmitter distorts the data vector $\mathbf{s}$ by another vector $\mathbf{e}$ so that DFT of $(\mathbf{s}-\mathbf{e})$ at these pilot frequencies are zero. The vector $\mathbf{e}$ is set as

$$e(i+P) = \frac{1}{Q}\sum_{m=0}^{L-1} s(i+mP), \quad i = 0, 1, \cdots, P-1. \quad (1)$$

The data block at source node is then

$$\mathbf{x} = \mathbf{s} - \mathbf{e} = \mathbf{s} - \mathbf{J}\mathbf{s}, \quad (2)$$

where $\mathbf{J} = (1/Q)\mathbf{1}_Q \otimes \mathbf{I}_P$. Here $\otimes$ is the Kronecker product, $\mathbf{1}_Q$ is a $Q\times Q$ matrix with all 1 entries, and $\mathbf{I}_P$ is a $P\times P$ identity matrix. The source node further superimpose $\mathbf{p}$ to the data block $\mathbf{x}$ and then transmit the combined signals.

In order to avoid inter-block interference (IBI) between source and relay node, a CP of length $L-1$ will be inserted in the front of each block. The signal received at the relay node, after removing the CP, can be expressed as

$$\mathbf{r} = \mathbf{H}_1(\mathbf{x}+\mathbf{p}) + \mathbf{w}_1, \tag{3}$$

where $\mathbf{H}_1$ is the $(N \times N)$ circulant matrix with first column $\bar{\mathbf{h}}_1 = [h_{1,0}, h_{1,1}, \cdots, h_{1,L-1}, 0, \cdots, 0]^T$, and $\mathbf{w}_1$ is the additive complex Gaussian noise with covariance $\sigma_{w_1}^2\mathbf{I}_N$. The relay then scales the received signal by a factor of $\alpha$ and inserts a new CP to $\mathbf{r}$ before sending it to the destination node. The amplifier factor $\alpha$ is a constant that is related with the channel statistics and powers whose explicit expression is omitted for brevity. After removing the CP, the signal received at the destination is

$$\mathbf{d} = \alpha\mathbf{H}_2\mathbf{r} + \mathbf{w}_2, \tag{4}$$

where $\mathbf{H}_2$ is the $(N \times N)$ circulant matrix with first column $\bar{\mathbf{h}}_2 = [h_{2,0}, h_{2,1}, \cdots, h_{2,L-1}, 0, \cdots, 0]^T$, and $\mathbf{w}_2$ is the additive complex Gaussian noise with covariance $\sigma_{w_2}^2\mathbf{I}_N$.

## III. DDST ESTIMATOR

### A. Estimation algorithm

Combining (2), (3) and (4), we can obtain

$$\mathbf{d} = \alpha\underbrace{\mathbf{H}_2\mathbf{H}_1}_{\mathbf{H}}(\mathbf{s}-\mathbf{e}) + \alpha\mathbf{H}_2\mathbf{H}_1\mathbf{p} + \underbrace{\alpha\mathbf{H}_2\mathbf{w}_1 + \mathbf{w}_2}_{\mathbf{w}}, \tag{5}$$

where $\mathbf{H}$ and $\mathbf{w}$ defines the effective channel matrix and the effective noise, respectively. The covariance matrix of $\mathbf{w}$ can be computed as

$$\mathbf{R}_w = E(\mathbf{w}\mathbf{w}^H) = \sigma_{w_2}^2\mathbf{I}_N + \alpha^2\sigma_{w_1}^2\mathbf{H}_2\mathbf{H}_2^H. \tag{6}$$

Since $\mathbf{H}_i$, $i = 1, 2$ is a circulant matrix, it can be rewritten as

$$\mathbf{H}_i = \mathbf{F}^H\mathbf{\Xi}_i\mathbf{F}, \tag{7}$$

where $\mathbf{F}$ is the $N \times N$ normalized discrete Fourier transform (DFT) matrix with $(m, n)$ entry $e^{-j2\pi mn/N}/\sqrt{N}$, $\mathbf{\Xi}_i = \mathrm{diag}(\boldsymbol{\xi}_i)$ is a diagonal matrix, and $\boldsymbol{\xi}_i = \sqrt{N}\mathbf{F}\bar{\mathbf{h}}_i$ for $i = 1, 2$.

Denote $\tilde{\mathbf{a}}$ as the DFT transform of any $(N \times 1)$ vector $\mathbf{a}$, i.e., $\tilde{\mathbf{a}} = \sqrt{N}\mathbf{F}\mathbf{a}$, and denote $\tilde{a}(i)$ as the $i$-th element of the vector $\tilde{\mathbf{a}}$. It can be shown that

$$\tilde{x}(kQ) = 0, \quad k = 0, 1, \cdots, P-1 \tag{8}$$

$$\tilde{d}(kQ) = \alpha\xi(kQ)\tilde{p}(kQ) + \tilde{w}(kQ), \tag{9}$$

where $\xi(kQ) = \xi_2(kQ)\xi_1(kQ)$.

Then the frequency response of the channel at the pilot frequencies can be estimated as

$$\hat{\xi}(kQ) = \tilde{d}(kQ)/(\alpha\tilde{p}(kQ)). \tag{10}$$

Define $\mathcal{P} = \{0, Q, \cdots, (P-1)Q\}$ as the index set for pilot frequency and define $\mathcal{L} = \{0, 1, \cdots, 2L-2\}$. We then denote $\mathbf{F}_{\mathcal{P},\mathcal{L}}$ as a $P \times (2L-1)$ submatrix of the DFT matrix $\mathbf{F}$, selected from $\mathbf{F}$ with the row indices in $\mathcal{P}$ and the column indices in $\mathcal{L}$. It can be readily checked that

$$\mathbf{\Xi}_2\mathbf{\Xi}_1 = N\mathrm{diag}(\mathbf{F}\bar{\mathbf{h}}_2)\mathrm{diag}(\mathbf{F}\bar{\mathbf{h}}_1) = \sqrt{N}\mathrm{diag}(\mathbf{F}\bar{\mathbf{h}}), \tag{11}$$

where $\bar{\mathbf{h}} = [\mathbf{h}^T, 0, \cdots, 0]^T$, $\mathbf{h} = \mathbf{h}_2 \star \mathbf{h}_1$, and $\star$ represents the linear convolution between $\mathbf{h}_2$ and $\mathbf{h}_1$.

The estimate of the combined channel $\mathbf{h}$ can be estimated from

$$\begin{aligned}\hat{\mathbf{h}} &= \frac{1}{\sqrt{N}}(\mathbf{F}_{\mathcal{P},\mathcal{L}}^H\mathbf{F}_{\mathcal{P},\mathcal{L}})^{-1}\mathbf{F}_{\mathcal{P},\mathcal{L}}^H\hat{\boldsymbol{\xi}}_{\mathcal{P}} \\ &= \frac{Q}{\sqrt{N}}\mathbf{F}_{\mathcal{P},\mathcal{L}}^H\hat{\boldsymbol{\xi}}_{\mathcal{P}}\end{aligned} \tag{12}$$

where $\hat{\boldsymbol{\xi}}_{\mathcal{P}} = [\hat{\xi}(0), \hat{\xi}(Q), \cdots, \hat{\xi}((P-1)Q)]^T$.

Note that this estimator can only estimate combined channel $\mathbf{h}$. In the following parts we refer to this estimator as DDST estimator and we will suggest a new estimator that can find individual channel information about $\mathbf{h}_1$ and $\mathbf{h}_2$.

### B. Training design

The estimation error of $\boldsymbol{\xi}$ is

$$\mathbf{\Delta}_\xi = \boldsymbol{\xi} - \hat{\boldsymbol{\xi}}. \tag{13}$$

It can be shown that

$$\Delta_\xi(kQ) = -\frac{\tilde{w}(kQ)}{\alpha\tilde{p}(kQ)} \tag{14}$$

Clearly, $E(\mathbf{\Delta}_\xi) = 0$. Define $\mathbf{\Delta}_{\xi_\mathcal{P}}$ as the $(P \times 1)$ vector selected from $\mathbf{\Delta}_\xi$ by set $\mathcal{P}$.

Note that when $Q \geq L$, $\tilde{\mathbf{w}}(kQ)$ is uncorrelated with each other for $kQ \in \mathcal{P}$. Thus we can have

$$E(\mathbf{\Delta}_{\xi_\mathcal{P}}\mathbf{\Delta}_{\xi_\mathcal{P}}^H) = c_1\mathbf{D}^{-1}, \tag{15}$$

where

$$c_1 = \sigma_{w_2}^2/\alpha^2 + \sigma_{w_1}^2\sum_{l=0}^{L-1}|h_{2,l}|^2, \tag{16}$$

$$\mathbf{D} = \mathrm{diag}\{|\tilde{p}(0)|^2, |\tilde{p}(Q)|^2, \cdots, |\tilde{p}((P-1)Q)|^2\}. \tag{17}$$

The estimation error of the combined channel $\mathbf{h}$ is

$$\mathbf{\Delta}_h = \mathbf{h} - \hat{\mathbf{h}}. \tag{18}$$

The channel estimation in (12) is unbiased as $E(\mathbf{\Delta}_h) = 0$. According to (12), the covariance matrix of channel estimation error is

$$\begin{aligned}E(\mathbf{\Delta}_h\mathbf{\Delta}_h^H) &= \frac{Q^2}{N}\mathbf{F}_{\mathcal{P},\mathcal{L}}^H E\left(\mathbf{\Delta}_{\xi,\mathcal{P}}\mathbf{\Delta}_{\xi,\mathcal{P}}^H\right)\mathbf{F}_{\mathcal{P},\mathcal{L}}^\dagger \\ &= \frac{c_1 Q}{P}\mathbf{F}_{\mathcal{P},\mathcal{L}}^H\mathbf{D}^{-1}\mathbf{F}_{\mathcal{P},\mathcal{L}}.\end{aligned} \tag{19}$$

From (19), we know that the estimation MSE of channel $\mathbf{h}$ is

$$\begin{aligned}\mathrm{MSE} &= \frac{1}{2L-1}E\left\{\sum_{l=0}^{2L-2}|\hat{h}(l) - h(l)|^2\right\} \\ &= \frac{c_1 Q}{P(2L-1)}\mathrm{Tr}\{\mathbf{F}_{\mathcal{P},\mathcal{L}}^H\mathbf{D}^{-1}\mathbf{F}_{\mathcal{P},\mathcal{L}}\}.\end{aligned} \tag{20}$$

Note that $\mathbf{D}$ is a diagonal matrix, we can thus find the upper bound of MSE as

$$\begin{aligned}\text{MSE} =&\frac{c_1 Q}{P(2L-1)}\text{Tr}\{\mathbf{D}^{-1}(\mathbf{F}_{\mathcal{P},\mathcal{L}}\mathbf{F}_{\mathcal{P},\mathcal{L}}^H)\}\\ \leq&\frac{c_1 Q}{P(2L-1)}M_1\text{Tr}\{(\mathbf{F}_{\mathcal{P},\mathcal{L}}\mathbf{F}_{\mathcal{P},\mathcal{L}}^H)\}\\ =&\frac{c_1 Q M_1}{P(2L-1)}\times\frac{2L-1}{Q}=\frac{c_1 M_1}{P}\end{aligned} \tag{21}$$

where $M_1 = \max_{i\in\mathcal{P}}\{\frac{1}{|\tilde{p}(i)|^2}\}$. Therefore, the optimization problem can be expressed as

$$\begin{aligned}\min \quad & \frac{c_1}{P}\max_{i\in\mathcal{P}}\{\frac{1}{|\tilde{p}(i)|^2}\}\\ \text{s. t.} \quad & \sum_{i=0}^{P-1}|\tilde{p}(iQ)|^2=\Phi_p,\end{aligned} \tag{22}$$

where $\Phi_p$ is the total power allocated for DDST. The optimal solution can be obtained as all $\tilde{p}(i)$ has equal power, that is, $M_1 = 1/|\tilde{p}(i)|^2 = P/\Phi_p$. Interestingly, in this case the estimation MSE is equal to the MSE upper bound,

$$\text{MSE} =\frac{c_1}{\Phi_p}. \tag{23}$$

### C. Data detection

First note that $\mathbf{J}^2 = \mathbf{J}$ which yields $(\mathbf{I}_N - \mathbf{J})^2 = \mathbf{I}_N - \mathbf{J}$. Since $\mathbf{H}$, $\mathbf{I}_N$ and $\mathbf{J}$ are circulant matrices, we can remove the contribution of the superimposed training by simply computing

$$\begin{aligned}\mathbf{z} =&(\mathbf{I}_N-\mathbf{J})\mathbf{d}\\ =&(\mathbf{I}_N-\mathbf{J})\mathbf{H}(\mathbf{I}_N-\mathbf{J})\mathbf{s}+(\mathbf{I}_N-\mathbf{J})\mathbf{w}\\ =&\mathbf{H}(\mathbf{I}_N-\mathbf{J})\mathbf{s}+(\mathbf{I}_N-\mathbf{J})\mathbf{w},\end{aligned} \tag{24}$$

where the fact $(\mathbf{I}_N - \mathbf{J})\mathbf{p} = \mathbf{0}$ is used.

***Remark*** *1:* In time domain, it can be readily checked that $(\mathbf{I}_N-\mathbf{J})\mathbf{p} = \mathbf{0}$ since $\mathbf{p}$ is periodical. In frequency domain, multiplying $\mathbf{p}$ with $\mathbf{I}_N - \mathbf{J}$ is equivalent to setting the DFT of $\mathbf{p}$ at pilot frequencies to zero.

Since the overall channel $\mathbf{h}$ has been estimated, we can construct the $N \times 1$ vector $\hat{\hat{\mathbf{h}}} = [\hat{\mathbf{h}}^T, 0, \cdots, 0]^T$ and construct a new vector

$$\mathbf{y} =\frac{1}{\sqrt{N}}\mathbf{F}^H\left(\text{diag}(\mathbf{F}\hat{\hat{\mathbf{h}}})\right)^{-1}\mathbf{F}\mathbf{z}. \tag{25}$$

Ideally, it can be shown that $\mathbf{y} = (\mathbf{I}_N - \mathbf{J})\mathbf{s}$ if there is no noise and $\mathbf{h}$ is estimated perfectly. However, since $(\mathbf{I}_N - \mathbf{J})$ is singular, we cannot recover $\mathbf{s}$ linearly.

We will then choose the iterative symbol-by-symbol method suggested in [5] to detect $\mathbf{s}$. The initial detector of $\mathbf{s}$ is given by

$$\hat{\mathbf{s}}^{(0)} = \lceil \mathbf{y} \rfloor, \tag{26}$$

where $\lceil \mathbf{y} \rfloor$ represents the vector of constellation points that are closest to the vector $y$. And the $i$-th iterative detector of $\mathbf{s}$ is obtained from

$$\hat{\mathbf{s}}^{(i)} = \lceil \mathbf{y} + \mathbf{J}\hat{\mathbf{s}}^{(i-1)} \rfloor. \tag{27}$$

## IV. Doubly Cooperative Estimator

The proposed DDST estimator can only find the overall cascaded channels $\mathbf{h}$, while the the separate channel information of $\mathbf{h}_1$ and $\mathbf{h}_2$ still remain unknown. However, knowing $\mathbf{h}_1$ and $\mathbf{h}_2$ is also crucial to improve the system performance and achieve the optimal design, for example, the carrier permutations, and relay beamforming [7].

In order to find individual parameters of $\mathbf{h}_1$ and $\mathbf{h}_2$, we propose that not only the source node adds a superimposed training $\mathbf{q}$ to the data signals, but also relay adds another superimposed training $\mathbf{u}$ to the received signal. The two superimposed training $\mathbf{q}$ and $\mathbf{u}$ should occupy different pilot frequencies.

The training $\mathbf{q}$ and $\mathbf{u}$ can be designed in the following way. Define $\mathcal{A}_q$ and $\mathcal{A}_u$ as the index sets of pilot frequencies for $\mathbf{q}$ and $\mathbf{u}$, with cardinality $A_q$ and $A_u$ respectively. We require $\mathcal{A}_q \cap \mathcal{A}_u = \phi$, $A_q \geq L$, $A_u \geq A_u \geq L$, and $A_q + A_u = P$.

For example, define $\mathbf{u}_1 = [u_0, u_1, \cdots, u_{N-1}]^T$, $\mathbf{q}_1 = [q_0, q_1, \cdots, q_{N-1}]^T$ and the value of the elements $u_z$ and $q_z$ for $z \in [0, N-1]$ are set as

$$\begin{cases} u_z = 1, & z\in\mathcal{A}_u\\ u_z=0, & z\notin\mathcal{A}_u,\end{cases}\qquad \begin{cases} q_z = 1, & z\in\mathcal{A}_q\\ q_z=0, & z\notin\mathcal{A}_q.\end{cases}$$

The training sequences $\mathbf{q}$ and $\mathbf{u}$ can be respectively obtained as

$$\mathbf{q} = \mathbf{F}^H\mathbf{q}_1, \quad \mathbf{u} = \mathbf{F}^H\mathbf{u}_1. \tag{28}$$

***Remark*** *2:* The condition $\mathcal{A}_q \cap \mathcal{A}_u = \phi$ is not necessary when there are sufficient pilot frequencies, i.e., $P$ is large. Here, we utilize orthogonal training to demonstrate our suggested method in a clearer way.

The relay adds $\mathbf{u}$ to the received $N \times 1$ symbols, and then adds CP of length $L - 1$ for this block and transmits it to the destination. Thus the signal received at the destination node $\mathbb{D}$, after removing CP, is

$$\mathbf{d} = \alpha\mathbf{H}_2\mathbf{H}_1(\mathbf{x}+\mathbf{q}) + \alpha\mathbf{H}_2\mathbf{u} + \mathbf{w}. \tag{29}$$

### A. Estimation algorithm

The DCE includes two process: initial estimation and iterative estimation. The DCE will take the following steps to get the initial estimates of channels $\mathbf{h}_1$ and $\mathbf{h}_2$.

1) After DFT transform of $\mathbf{d}$, we obtain
$$\tilde{d}(kQ) = \alpha\xi(kQ)\tilde{q}(kQ) + \alpha\xi_2(kQ)\tilde{u}(kQ) + \tilde{w}(kQ). \tag{30}$$
2) Note that $\tilde{q}(z) = 0$ and $\tilde{u}(z) \neq 0$ for $z \in \mathcal{A}_u$. Therefore, the estimate of $\boldsymbol{\xi}_2$ at the $A_u$ pilot frequencies can be found as
$$\xi_2(z) = \tilde{d}(z)/\tilde{u}(z), \quad z \in \mathcal{A}_u. \tag{31}$$
3) Let us collect the estimated $\boldsymbol{\xi}_2(z)$ of set $\mathcal{A}_u$ into a $A_u \times 1$ vector $\boldsymbol{\xi}_{2u}$. Define $\mathcal{L}_0$ as the index set for $[0, 1, \cdots, L-1]$ and denote $\mathbf{F}_{\mathcal{A}_u,\mathcal{L}_0}$ as the $A_u \times L$ submatrix of the DFT matrix $\mathbf{F}$, selected from $\mathbf{F}$ with

the leading $L$ columns and the rows of set $\mathcal{A}_u$. Then we can estimate $\mathbf{h}_2$ as

$$\hat{\mathbf{h}}_2 = \frac{1}{\sqrt{N}}(\mathbf{F}^H_{\mathcal{A}_u,\mathcal{L}_0}\mathbf{F}_{\mathcal{A}_u,\mathcal{L}_0})^{-1}\mathbf{F}^H_{\mathcal{A}_u,\mathcal{L}_0}\hat{\boldsymbol{\xi}}_{2_u}. \quad (32)$$

Next $\hat{\boldsymbol{\xi}}_2$, the estimate of $\boldsymbol{\xi}_2$, can be obtained from

$$\hat{\boldsymbol{\xi}}_2 = \sqrt{N}\mathbf{F}\hat{\bar{\mathbf{h}}}_2, \quad (33)$$

where $\hat{\bar{\mathbf{h}}}_2 = [\hat{\mathbf{h}}_2^T, 0, \cdots, 0]^T$.

4) Similarly for $z \in \mathcal{A}_q$, we can also find the value of $\boldsymbol{\xi}_1$ at the $A_q$ pilot frequencies as

$$\hat{\xi}_1(z) = \frac{\tilde{d}(z) - \alpha\hat{\xi}_2(z)\tilde{u}(z)}{\alpha\hat{\xi}_2(z)\tilde{q}(z)}. \quad (34)$$

Let us collect the estimated $\hat{\boldsymbol{\xi}}_1(z)$ for all $z \in \mathcal{A}_q$ into a $A_q \times 1$ vector $\hat{\boldsymbol{\xi}}_{1q}$ and then we can estimate $\mathbf{h}_1$ as

$$\hat{\mathbf{h}}_1 = \frac{1}{\sqrt{N}}(\mathbf{F}^H_{\mathcal{A}_q,\mathcal{L}_0}\mathbf{F}_{\mathcal{A}_q,\mathcal{L}_0})^{-1}\mathbf{F}^H_{\mathcal{A}_q,\mathcal{L}_0}\hat{\boldsymbol{\xi}}_{1q}, \quad (35)$$

where $\mathbf{F}_{\mathcal{A}_q,\mathcal{L}_0}$ is the $A_q \times L$ submatrix of the DFT matrix $\mathbf{F}$, selected from $\mathbf{F}$ by the leading $L$ columns and by the rows of set $\mathcal{A}_q$. Similar as (33), $\hat{\boldsymbol{\xi}}_1$ can then be obtained from $\hat{\mathbf{h}}_1$ by using the DFT transform.

5) The estimate of overall channel can be found as $\hat{\mathbf{h}} = \hat{\mathbf{h}}_1 \star \hat{\mathbf{h}}_2$.

Next, with these initial estimates, iteration can be applied by DCE to further improve our estimation performance. For notation conciseness, the superscript $(k)$ is used to denote the $k$th iteration.

For all $k = 0, 1, \cdots, P-1$ we can obtain

$$\eta(kQ) = \frac{\tilde{d}(kQ)}{\alpha\hat{\Xi}_1(kQ)\tilde{q}(kQ) + \alpha\tilde{u}(kQ)}. \quad (36)$$

Define $\boldsymbol{\eta}_{\mathcal{P}} = [\eta(0), \eta(Q), \cdots, \eta((P-1)Q)]^T$. We can then update the estimate of $\mathbf{h}_2$ as

$$\hat{\mathbf{h}}_2^{(i)} = \frac{1}{\sqrt{N}}(\mathbf{F}^H_{\mathcal{P},\mathcal{L}_0}\mathbf{F}_{\mathcal{P},\mathcal{L}_0})^{-1}\mathbf{F}^H_{\mathcal{P},\mathcal{L}_0}\boldsymbol{\eta}_{\mathcal{P}}, \quad (37)$$

where $\mathbf{F}_{\mathcal{P},\mathcal{L}_0}$ is the selected rows by set $\mathcal{P}$ and the leading $L$ columns of DFT matrix $\mathbf{F}$. The iterative algorithm can benefit from the fact that more information are used than the initial estimation (32). Then update $\hat{\boldsymbol{\xi}}_2^{(i)}$ is similar to (33). Finally, $\hat{\mathbf{h}}_1^{(i)}$, the updated estimate of $\mathbf{h}_1$ in the $i$th iteration, can be found by (35).

As can be seen from the simulation part, the iteration converges after several times and the first iteration obtains the biggest gain.

***Remark** 3:* From above analysis, we know that the minimum training length required for DCE is $2L$ while it is $2L-1$ for DDST estimator.

### B. Data detection

Since $(\mathbf{I}_N - \mathbf{J})\mathbf{q} = (\mathbf{I}_N - \mathbf{J})\mathbf{u} = \mathbf{0}$, we can apply the same detection process in Section. III-C, which is then omitted for brevity.

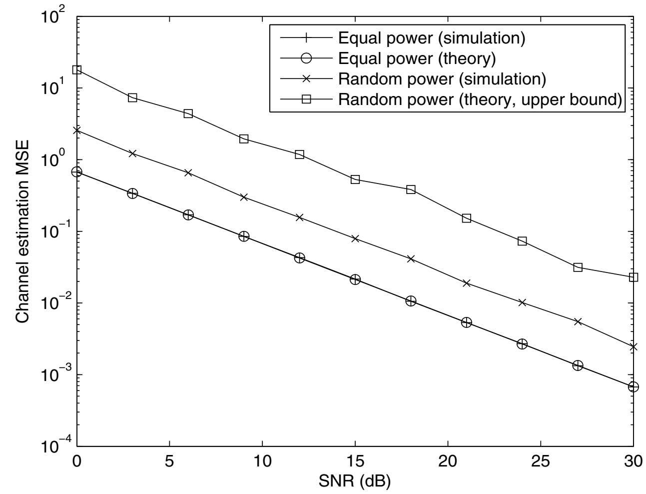


Fig. 2. MSEs of channel estimation versus SNR for DDST estimator

## V. Simulation Results

In this section, we numerically examine the the performance of our proposed channel estimators.

A four-tap channel model with the exponential delay profile $\sigma_{il} = e^{-l/10}$, $i = 1, 2$, $l = 0, 1, 2, 3$ is assumed for $\mathbf{h}_i$. The variance of noise is taken as 1. The signal-to-noise ratio (SNR) is defined as the ratio of symbol power to the noise power and the superimposed pilots take 20 percent of the whole power. Moreover, $P$ and $Q$ are set as 32 and 8, respectively so that $N = 256$. Data symbols are extracted from QPSK constellation while superimposed pilots are taken from BPSK constellation. For all examples, 1000 Monte-Carlo trials are used for the averaging. The channel estimation MSEs are chosen as the figure of merit.

### A. Performance of DDST estimator

Firstly, we investigate the performance of DDST estimator. We compare two types of DDST, the optimal equal power training and the random power training. The MSEs of the two trainings versus SNR are shown in Fig. 2. Theoretical values of upper bounds of MSEs (21) are also shown in Fig. 2. The optimal equal power training outperforms the random power training by 5 dB. Moreover, for the optimal equal power DDST, the theoretical upper bounds of MSEs are equal to the MSEs obtained from simulation, which agrees with our analysis.

### B. Performance of DCE

Next we examine the estimation accuracy of DCE. The initial and iterative estimation MSEs of separate channel $\mathbf{h}_1$ and $\mathbf{h}_2$ versus SNR are given in Fig. 3. Iteration runs for 6 times and for conciseness only the first and last iterative MSEs are shown. The estimation performance increases as SNR increases. In addition, the estimation performance of $\mathbf{h}_2$ is better than estimation performance of $\mathbf{h}_1$ because the latter depends on estimated channel of $\mathbf{h}_2$. The iterative algorithm works at low SNR and improves the estimation accuracy by

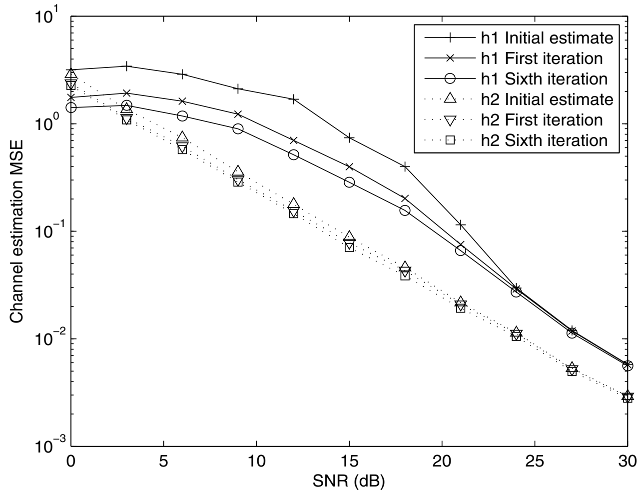


Fig. 3. MSEs of channel estimation versus SNR for DCE

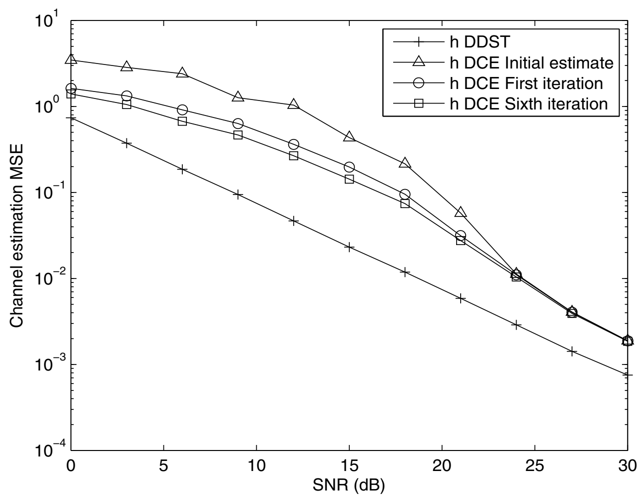


Fig. 4. MSE comparison for DDST estimator and DCE

2-4 dB. At high SNR, little performance improvement can be made by iteration since the initial estimation can be accurate enough.

### C. Comparison

Finally, we compare the performance of DDST estimator and DCE. The MSEs of overall channel $\mathbf{h}$ versus SNR obtained from DDST estimator and DCE are shown in Fig. 4. DDST estimator outperforms DCE in estimation accuracy by 3 dB, which indicates that DCE obtains individual channel information at the cost of some performance loss.

The symbol error rates (SERs) of systems that utilize DDST estimator or DCE versus SNR are demonstrated in Fig. 5. The iterative detection algorithm can improve detection accuracy and the biggest gain is obtained from the first iteration.

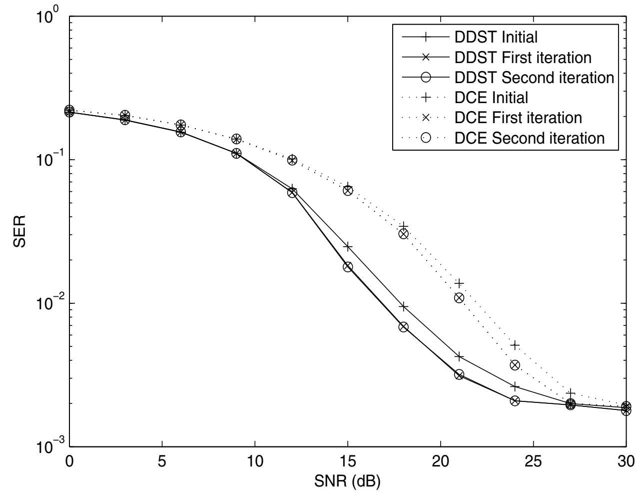


Fig. 5. SER comparison for DDST estimator and DCE.

## VI. Conclusion

In this paper, we applied DDST to AF relay networks for channel estimation. The optimal training sequence was derived based on minimizing the upper bound of mean square error of channel estimation. Since the DDST can only estimate the cascaded channel information, we propose a new DCE estimator to further extract the individual channel parameters. Our simulation results show that both DDST estimator and DCE work well, whereas DCE can obtain the separate channel knowledge at the cost of some performance loss.

## References


[1] J. Cavers, "An analysis of pilot symbol assisted modulation for rayleigh fading channels [mobile radio]," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 683-693, Nov. 1991.

[2] C. Kastenholz and W. Birkemeier, "A simultaneous information transfer and channel-sounding modulation technique for wide-band channels," *IEEE Trans. on Communications*, vol. 12, pp.162-165, June 1965.

[3] G. T. Zhou, M. Viberg, and T. McKelvey, "First-order statistical method for channel estimation," *IEEE Signal Processing Letters*, vol. 10, no. 3, pp. 57-60, Mar. 2003.

[4] J. Tugnait and W. Luo, "On channel estimation using superimposed training and first-order statistics," *IEEE Communication Letters*, vol. 7, pp. 413-415, Sept. 2003.

[5] M. Ghogho, D. McLernon, E. Alameda-Hernandez, and A. Swami, "Channel estimation and symbol detection for block transmission using data-dependent superimposed training," *IEEE Signal Processing Letters*, vol. 12, pp. 226-229, Mar. 2005.

[6] S. He and J.Tugnait, "On doubly selective channel estimation using superimposed training and discrete prolate spheroidal sequences," *IEEE Trans. on Signal Processing*, vol. 56, pp.3214-3228, July 2008.

[7] Y. Rong, "Optimal linear non-regenerative multi-hop MIMO relays with MMSE-DFE receiver at the destination," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 2268–2279, Jul. 2010.