Channel Estimation with Amplitude Constraint: Superimposed Training or Conventional Training ?

Gongpu Wang†, Feifei Gao*, and Chintha Tellambura†
†Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada
*Department of Automation, Tsinghua University, Beijing, China
Email: †{gongpu, chintha}@ece.ualberta.ca , *feifeigao@ieee.org

***Abstract*—This paper utilizes a general superimposed training based transmission scheme that includes superimposed training and pilot symbol assisted modulation (PSAM) as special cases. The channel estimator of the scheme is the linear minimum mean square error (LMMSE) estimator. By taking into account errors of this method, we derive the closed-form lower bound of the data throughput under the constraint of limited amplitude for each symbol. Our study shows that with the constraint of total amplitude for each symbol, the conventional PSAM performs better in the high signal-to-noise ratio (SNR) region while at low SNR, the superimposed scheme performs better.**

I. Introduction

The performance of wireless communication systems is critically related to the accuracy of channel estimation. For most cases, the channel estimates are obtained from the pilot symbols that are multiplexed with the data symbols; this scheme is called pilot symbol assisted modulation (PSAM) [1]. An alternative way is to superimpose the pilot symbols onto each data symbol in the frame, known as superimposed training [2], [3].

Superimposed training utilizes the first-order statistics to estimate the channel [3] when the data and the noise are zero-mean sequences. A frequency-domain estimation method was suggested in [2] for superimposed training in the cases when the noise cannot be considered as zero-mean sequences. Although the estimation performance of superimposed training falls short of PSAM, the data-dependent superimposed training (DDST) suggested in [4] improves both estimation and detection and thus shorten the performance gap.

The performance comparison between superimposed training and conventional training is discussed in [5] and [6] for orthogonal frequency division multiplexing (OFDM) systems. These references show that the superimposed training yields higher system capacity than PSAM. However, these works adopt the least square (LS) estimation results and use an approximate lower bound of channel capacity. Moreover, these works do not determine the optimal training for either PSAM or superimposed training scheme, which may limit the validity of their capacity comparison.

A more general model that includes superimposed training and conventional training is suggested in [9], and it is found that superimposed training can offer an increased performance over conventional training for fast fading channels. The general model in [9] considers the constraint of total transmitted energy. However, for many real applications, it is more practical to consider the total amplitude constraint instead of toal energy constraint.

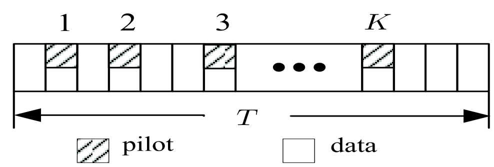

Fig. 1. The structure of one data frame.

This paper makes the following contributions. First, a unified model covering both superimposed training and PSAM over flat-fading channels is utilized. This model take into account of amplitude constraint for each transmitted symbol. Second, based on linear minimum mean square error (LMMSE) channel estimation, the closed-form lower bound of data throughput is derived as a function of the amount of training and the power allocation between pilot symbols and data symbols. Third, since there is no analytical expression for optimal allocation, we utilize the two-dimensional search and show that the conventional PSAM maximizes the lower bound at high signal-to-noise ratio (SNR) while at low SNR the conventional superimposed training yields better performance.

II. System Model

We consider single-input single-out put (SISO) systems with a flat-fading channel. Suppose one frame contains $T(\geq 1)$ symbols. In this frame, $K(\leq T)$ training symbols are superimposed on K data symbols,[1] as shown in Fig. 1. The remaining $(T-K)$ symbols contain only data information. Let $\mathcal{T}_p$ denote the index set of the superimposed training symbols while denote $\mathcal{T}_d$ as the index set of pure data symbols. The full time index set is then $\mathcal{T} = \mathcal{T}_d \bigcup \mathcal{T}_p = \{n = 1, 2, \cdots, T\}$.

Instead of the total power constraint for the whole data frame,[2] a more practical average power constraint for each symbol is adopted here, denoted as ρ. For symbols with time index $n \in \mathcal{T}_p$, the data symbol is assigned with the average power ρ_d, while the pilot symbol is assigned with the power ρ_p. There is

$$\rho = \rho_d + \rho_p. \quad (1)$$

[1]The conventional superimposed training overlay pilots on all data symbols.

[2]In some existing works, e.g., [8], the power constraint for the whole data frame is adopted which may assign a "too large power" over a certain symbol period such that the linearity of the power amplifier is violated.

The received signal can be expressed as

$$y(n) = \begin{cases} h(\sqrt{\rho_d}s(n) + \sqrt{\rho_p}p(n)) + w(n), & n \in \mathcal{T}_p \\ h\sqrt{\rho}s(n) + w(n), & n \in \mathcal{T}_d \end{cases} \quad (2)$$

where h is the Rayleigh flat-fading channel with variance σ_h^2, and $w(n)$ is the additive white Gaussian noise with variance σ_w^2. Moreover, the data and pilot symbols satisfy $E\{|s(n)|^2\} = 1$ and $|p(n)| = 1$, respectively.

Remark *1:* This model considers the amplitude constraint (1), which means the amplitude for each transmitted symbol is limited. This constraint is different from the total energy constraint in [9] where the total power for the whole frame is limited.

III. Channel estimation and data detection

This section estimates the channel by using the LMMSE method, derives the expression for channel estimation error and analyzes data detection in the presence of the channel estimation error.

A. Channel Estimation

Stack $y(n)$, $s(n)$, $p(n)$ and $w(n)$ from the set $\mathcal{T}_p$ into $K \times 1$ vectors $\mathbf{y}_p$, $\mathbf{s}_p$, $\mathbf{p}$ and $\mathbf{w}_p$. We obtain

$$\mathbf{y}_p = h\sqrt{\rho_p}\mathbf{p} + \underbrace{h\sqrt{\rho_d}\mathbf{s}_p + \mathbf{w}_p}_{\mathbf{v}}, \quad (3)$$

where $\mathbf{v}$ is defined as the corresponding item.

The way of first-order statistics is a common channel estimation method for superimposed training [3]. Define $y_0 = \sum_{n\in\mathcal{T}_p} y(n)p^*(n)/K$, where $(\cdot)^*$ denotes the conjugate operation. It can be shown that

$$y_0 = \sqrt{\rho_p}h + v_0, \quad (4)$$

where

$$v_0 = \frac{h\sqrt{\rho_d}}{K}\sum_{n\in\mathcal{T}_p} s(n)p^*(n) + \frac{1}{K}\sum_{n\in\mathcal{T}_p} w(n)p^*(n), \quad (5)$$

and the variance of v_0 is $\sigma_{v_0}^2 = (\sigma_h^2\rho_d + \sigma_w^2)/K$. The LMMSE channel estimator is selected for its orthogonality property that will be further exploited in our following throughput analysis. The estimate of h is obtained as

$$\hat{h} = \frac{\sigma_h^2\sqrt{\rho_p}y_0}{\sigma_h^2\rho_p + \sigma_{v_0}^2}. \quad (6)$$

Lemma 1: The channel estimator (6) is the same as the direct LMMSE estimate from (3)

$$\hat{h}' = \sigma_h^2\sqrt{\rho_p}\mathbf{p}^H\left(\sigma_h^2\rho_p\mathbf{p}\mathbf{p}^H + (\sigma_h^2\rho_d + \sigma_w^2)\mathbf{I}_K\right)^{-1}\mathbf{y}_p. \quad (7)$$

Proof: From Woodbury's identity [7, Equation (A.50)]

$$\left(\mathbf{A} + \mathbf{x}\mathbf{x}^H\right)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{x}\mathbf{x}^H\mathbf{A}^{-1}}{1 + \mathbf{x}^H\mathbf{A}^{-1}\mathbf{x}}, \quad (8)$$

we can rewrite

$$\begin{aligned}&\left((\sigma_h^2\rho_d + \sigma_w^2)\mathbf{I}_K + \sigma_h^2\rho_p\mathbf{p}\mathbf{p}^H\right)^{-1} \\ =&\frac{1}{\sigma_h^2\rho_d + \sigma_w^2}\left(\mathbf{I}_K - \frac{\sigma_h^2\rho_p}{\sigma_h^2\rho_d + \sigma_w^2 + K\sigma_h^2\rho_p}\mathbf{p}\mathbf{p}^H\right). \quad (9)\end{aligned}$$

Substituting (9) into (7) and using $\mathbf{p}^H\mathbf{p} = K$ will produce

$$\hat{h}' = \frac{\sigma_h^2\sqrt{\rho_p}}{\sigma_h^2\rho_d + \sigma_w^2 + K\sigma_h^2\rho_p}\mathbf{p}^H\mathbf{y}_p. \quad (10)$$

After straight calculation, it can be found that the estimate (6) is the same with (10), i.e., $\hat{h}' = \hat{h}$. ■

The channel estimation error is then $\epsilon = \hat{h} - h$ whose variance can be expressed as

$$\begin{aligned}\sigma_\epsilon^2 &= \left(\frac{1}{\sigma_h^2} + \sqrt{\rho_p}\mathbf{p}^H\left(E(\mathbf{v}\mathbf{v}^H)\right)^{-1}\sqrt{\rho_p}\mathbf{p}\right)^{-1} \\ &= \left(\frac{1}{\sigma_h^2} + \frac{\rho_p K}{\sigma_h^2\rho_d + \sigma_w^2}\right)^{-1}. \quad (11)\end{aligned}$$

Corollary 1: The closed-form expression for channel estimation error is

$$\epsilon = \frac{\sigma_h^2\sqrt{\rho_p}(h\sqrt{\rho_d}\mathbf{p}^H\mathbf{s}_p + \mathbf{p}^H\mathbf{w}_p) - (\sigma_h^2\rho_d + \sigma_w^2)h}{K\sigma_h^2\rho_p + \sigma_h^2\rho_d + \sigma_w^2}. \quad (12)$$

Proof: Proved with straight calculation from $\epsilon = \hat{h} - h$. ■

From corollary 1, it can be readily checked that $E\{\epsilon s^*(n)\} = 0$ for $n \in \mathcal{T}_p$ and also $E\{\epsilon\hat{h}^*\} = 0$. Note that this does not guarantee that the error ϵ is independent from $s(n)$ or $\hat{h}$.

B. Data Detection

This subsection rearranges received signal $y(n)$ suitable for data detection in the presence of the channel estimation error and finds the covariance between data and equivalent noise and the variance of equivalent noise.

Let us collect $y(n)$, $s(n)$, and $w(n)$ from the set $\mathcal{T}_d$ into $(T-K)\times 1$ vectors $\mathbf{y}_d$, $\mathbf{s}_d$, and $\mathbf{w}_d$ respectively. With estimated channel information $\hat{h}$, (2) can be rewritten as

$$\mathbf{y}_p = \hat{h}\sqrt{\rho_d}\mathbf{s}_p + \underbrace{\mathbf{w}_p - \epsilon(\sqrt{\rho_d}\mathbf{s}_p + \sqrt{\rho_p}\mathbf{p})}_{\mathbf{u}_p}, \quad (13a)$$

$$\mathbf{y}_d = \hat{h}\sqrt{\rho}\mathbf{s}_d + \underbrace{\mathbf{w}_d - \epsilon\sqrt{\rho}\mathbf{s}_d}_{\mathbf{u}_d}, \quad (13b)$$

where $\mathbf{u}_p$ and $\mathbf{u}_d$ represent the equivalent interference-plus-noise. From the previous discussion we know $E\{\epsilon s^*(n)\} = 0$ for any $n \in \mathcal{T}$. These following cross-correlations vanish

$$E\{\mathbf{s}_d\mathbf{u}_d^H\} = \mathbf{0}_{T-K}, \quad E\{\mathbf{s}_p\mathbf{u}_p^H\} = \mathbf{0}_K, \quad (14)$$

where $\mathbf{0}_K$ means a $K \times K$ matrix with all zero entries. For the same reason, the covariance matrix of $\mathbf{u}_d$ and that of $\mathbf{u}_p$ can be expressed as

$$\mathbf{R}_{u_p} = N_p\mathbf{I}_K + \sigma_\epsilon^2\rho_p\mathbf{p}\mathbf{p}^H, \quad \mathbf{R}_{u_d} = N_d\mathbf{I}_{T-K}, \quad (15)$$

where $N_p = \sigma_w^2 + \sigma_\epsilon^2\rho_d$ and $N_d = \sigma_w^2 + \sigma_\epsilon^2\rho$.

IV. Transmission Optimization

This section finds the expression for data throughput so that the optimal power allocation and time allocation can be found.

A. Data Throughput

Let $\mathcal{C}$, $\mathcal{I}$ and $\mathcal{H}$ denote the data throughput, mutual information and entropy respectively. The data throughput is the achievable average data rate over a communication channel, which is related to mutual information. From (13), the data throughput can be computed as

$$\mathcal{C} = \frac{1}{T}\mathcal{I}(\mathbf{y}_d, \mathbf{s}_d|\hat{h}) + \frac{1}{T}\mathcal{I}(\mathbf{y}_p, \mathbf{s}_p|\hat{h}). \quad (16)$$

Based on the definition of mutual information, the first term in (16) is rewritten as

$$\begin{aligned}\mathcal{I}(\mathbf{y}_d, \mathbf{s}_d|\hat{h}) =&E_{\hat{h}}\left\{\mathcal{H}(\mathbf{s}_d|\hat{h}) - \mathcal{H}(\mathbf{s}_d|\hat{h}, \mathbf{y}_p)\right\}\\ =&E_{\hat{h}}\left\{\log\det(\pi e\mathbf{R}_{s_d}) - \mathcal{H}(\mathbf{s}_d|\hat{h}, \mathbf{y}_d)\right\}, \quad (17)\end{aligned}$$

where the property that $\mathbf{s}_d$ is independent from $\hat{h}$ is used. Moreover,

$$\begin{aligned}\mathcal{H}(\mathbf{s}_d|\hat{h}, \mathbf{y}_d) =&\mathcal{H}(\mathbf{s}_d - f(\hat{h}, \mathbf{y}_d)|\hat{h}, \mathbf{y}_d)\\ &\leq \log\det(\pi e\mathrm{cov}(\mathbf{s}_d - f(\hat{h}, \mathbf{y}_d)), \quad (18)\end{aligned}$$

where $f(\hat{h}, \mathbf{y}_d)$ is any function, $\mathrm{cov}(\cdot)$ denotes the covariance matrix, and the inequality is obtained by replacing $\mathbf{s}_d - f(\hat{h}, \mathbf{y}_d)$ with another Gaussian random vectors of the same covariance. Therefore, the lower bound of $\mathcal{I}(\mathbf{y}_d, \mathbf{s}_d|\hat{h})$ is

$$\begin{aligned}\mathcal{I}(\mathbf{y}_d, \mathbf{s}_d|\hat{h}) \geq& E_{\hat{h}}\big\{\log\det(\pi e\mathbf{R}_{s_d})\\ &- \log\det(\pi e\mathrm{cov}(\mathbf{s}_d - f(\hat{h}, \mathbf{y}_d))\big\}, \quad (19)\end{aligned}$$

for any function $f(\hat{h}, \mathbf{y}_d)$. In order to obtain the tightest lower bound, we would like to find a function $f(\hat{h}, \mathbf{y}_d)$ such that $\det(\mathrm{cov}(\mathbf{s}_d - f(\hat{h}, \mathbf{y}_d))$ is as small as possible. Since optimal function $f(\hat{h}, \mathbf{y}_d)$ cannot be computed in closed form, we can simply choose the LMMSE estimate of $\mathbf{s}_d$ that minimizes the $tr(\mathrm{cov}(\mathbf{s}_d - f(\hat{h}, \mathbf{y}_d))$ as a good candidate, which could be specifically expressed as

$$f(\hat{h}, \mathbf{y}_d) = \hat{h}\sqrt{\rho}(|\hat{h}|^2\rho\mathbf{R}_{s_d} + \mathbf{R}_{u_d})^{-1} = \frac{\hat{h}\sqrt{\rho}}{|\hat{h}|^2\rho + N_d}\mathbf{I}, \quad (20)$$

and the corresponding covariance is

$$\begin{aligned}\mathrm{cov}(\mathbf{s}_d - f(\hat{h}, \mathbf{y}_d)) =& \left(\mathbf{R}_{s_d} + |\hat{h}|^2\rho\mathbf{R}_{u_d}^{-1}\right)^{-1}\\ =&\frac{1}{1 + |\hat{h}|^2\rho/N_d}\mathbf{I}. \quad (21)\end{aligned}$$

Note that, the property (14) is utilized when computing the LMMSE estimator (20). We then obtain

$$\mathcal{I}(\mathbf{y}_d, \mathbf{s}_d|\hat{h}) \geq (T - K)E_{\hat{h}}\left\{\log(1 + \gamma_d)\right\}, \quad (22)$$

where $\gamma_d = |\hat{h}|^2\rho/N_d$.

Following the similar derivation and using the property $E\{\hat{h}s^*(n)\} = E\{\epsilon s^*(n)\} + E\{hs^*(n)\} = 0$ for $n \in \mathcal{T}_p$, we can compute the lower bound of the second term in (16) as

$$\mathcal{I}(\mathbf{y}_p, \mathbf{s}_p|\hat{h}) \geq E_{\hat{h}}\left\{\log\det\left(\mathbf{I}_K + |\hat{h}|^2\rho_d\mathbf{R}_{u_p}^{-1}\right)\right\}. \quad (23)$$

By using Woodbury's identity [7, Equation (A.50)] for $\mathbf{R}_{u_p}^{-1}$, we can further express (23) as

$$\begin{aligned}\mathcal{I}(\mathbf{y}_p, \mathbf{s}_p|\hat{h}) \geq& E_{\hat{h}}\left\{\log\det\left(\mathbf{I}_K + \gamma_p\mathbf{I}_K - c_p\mathbf{p}\mathbf{p}^H\right)\right\}\\ =&E_{\hat{h}}\left\{\log\left((1 + \gamma_p)^K\det(1 - \frac{c_p}{1 + \gamma_p}\mathbf{p}^H\mathbf{p})\right)\right\}\\ =&E_{\hat{h}}\left\{(K - 1)\log(1 + \gamma_p) + \log(1 + \phi)\right\},\end{aligned} \quad (24)$$

where $\gamma_p = \frac{|\hat{h}|^2\rho_d}{N_p}$, $c_p = \frac{|\hat{h}|^2\rho_d\sigma_\epsilon^2\rho_p}{(N_p + K\sigma_\epsilon^2\rho_p)N_p}$, and

$$\phi = \gamma_p - c_pK = \frac{|\hat{h}|^2\rho_d}{N_p + K\sigma_\epsilon^2\rho_p}. \quad (25)$$

Finally, the lower bound of data throughput (16) can be computed as

$$\begin{aligned}\mathcal{C}_L =&\frac{1}{T}E_{\hat{h}}\big\{(T - K)\log(1 + \gamma_d)\\ &+ (K - 1)\log(1 + \gamma_p) + \log(1 + \phi)\big\}. \quad (26)\end{aligned}$$

B. Optimization over Power and Time

Define $\bar{h} = \frac{\hat{h}}{\sigma_{\hat{h}}}$, $\beta = \rho_p/\rho$ and $\rho_y = \sigma_h^2\rho + \sigma_w^2$. We can rewrite the variance of channel estimation error (11) as

$$\sigma_\epsilon^2 = \frac{(\sigma_w^2 + \sigma_h^2\rho_d)\sigma_h^2}{\sigma_w^2 + \sigma_h^2\rho_d + K\sigma_h^2\rho_p} = \frac{(\rho_y - \beta\rho\sigma_h^2)\sigma_h^2}{\rho_y + (K - 1)\sigma_h^2\beta\rho}. \quad (27)$$

By the orthogonality principle of LMMSE estimate [8], the variance of channel estimate can be written as

$$\sigma_{\hat{h}}^2 = \sigma_h^2 - \sigma_\epsilon^2 = \frac{K\sigma_h^4\beta\rho}{\rho_y + (K - 1)\sigma_h^2\beta\rho}. \quad (28)$$

We can thus obtain

$$\gamma_d = \frac{|\hat{h}|^2\rho}{\sigma_w^2 + \sigma_\epsilon^2\rho} = \frac{K\beta\sigma_h^4\rho^2|\bar{h}|^2}{\rho_y^2 + \sigma_h^2\beta\rho(K\sigma_w^2 - \rho_y)}, \quad (29)$$

$$\gamma_p = \frac{|\hat{h}|^2\rho_d}{\sigma_w^2 + \sigma_\epsilon^2\rho_d} = \frac{K\beta(1 - \beta)\sigma_h^4\rho^2|\bar{h}|^2}{\rho_y^2 + \sigma_h^2\beta\rho(K\sigma_w^2 - 2\rho_y + \beta\rho\sigma_h^2)}, \quad (30)$$

$$\begin{aligned}\phi =&\frac{|\hat{h}|^2\rho_d}{\sigma_w^2 + \sigma_\epsilon^2\rho_d + K\sigma_\epsilon^2\rho_p}\\ =&\frac{K\beta(1 - \beta)\sigma_h^4\rho^2|\bar{h}|^2}{\rho_y^2 + \sigma_h^2\beta\rho(K\sigma_w^2 + (K - 2)\rho_y - (K - 1)\beta\rho\sigma_h^2)}. \quad (31)\end{aligned}$$

The final expression for $\mathcal{C}_L$ can be found by substituting (29), (30) and (31) into (26).

Unfortunately, given the closed-form expression for $\mathcal{C}_L$, it still remains a challenge to obtain optimal value for training number K and power allocation ratio β. Therefore, we resort to the two-dimensional search to find the optimal values.[3] Note that, the parameters design is not related with instant channel knowledge so the proposed two dimensional search can be conducted in an off-line manner, whose complexity is acceptable.

[3] With average power constraint, the optimization can only be carried out with searching in other existing works, e.g., [8], too.

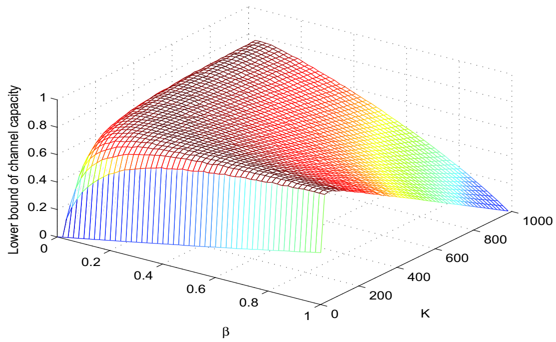

Fig. 2. Lower bound of channel capacity $\mathcal{C}_L$ at low SNR: $\rho = 1$ and $T = 1000$

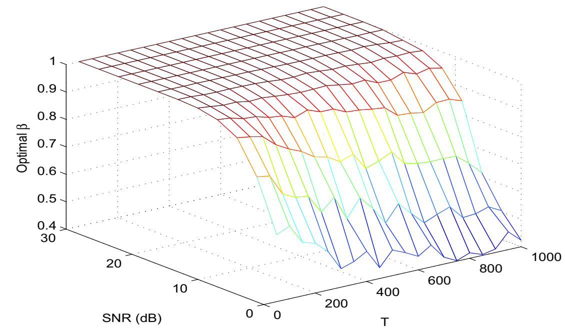

Fig. 3. Optimal value for power allocation ratio β

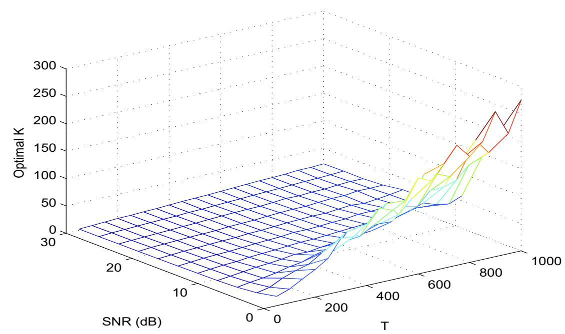

Fig. 4. Optimal value for training number K

V. Simulations

We define SNR as ρ/σ_w^2 and set $\sigma_h^2 = 1$, $\sigma_w^2 = 1$. Firstly, we fix the total number of symbols $T = 1000$ and $\rho = 1$, and change training number K from 1 to T and power allocation β from 0 to 1. For each K and β value, we generate the normalized channel estimate $\bar{h}$ and find γ_d, γ_p and ϕ according to (29), (30) and (31) respectively. Then the lower bound of data throughput $\mathcal{C}_L$ can be obtained from (26). This process is repeated for 10000 times and then the average $\mathcal{C}_L$ is found. Fig. 2 shows the average $\mathcal{C}_L$ versus K and β in the case of $\rho = 1$ and $T = 1000$.

Next we increase SNR from 0 dB to 30 dB, and total number of symbols T from 1 to 1000. For each SNR and T, we generate $\bar{h}$ for 10000 times, and in each time we search the maximum $\mathcal{C}_L$ (26) by changing K from 1 to T and β from 0 to 1. We record the maximum $\mathcal{C}_L$ and the corresponding β and K values. Thus averaging value of β and K over 10000 times can be regarded as the optimal value for optimal power allocation and number of trainings for the current SNR and the current total number of symbols T.

The optimal value for power allocation ratio β versus SNR and total number of symbols T is plotted in Fig. 3. It shows that at high SNR, the optimal value for power allocation ratio β is 1; that is, the conventional training method, PSAM, will maximize the lower bound of data throughput $\mathcal{C}_L$. However, at low SNR, the optimal value for power allocation ratio β varies. This suggests that superimposed pilots can outperform conventional trainings in low SNR.

The optimal value for training number K versus SNR and total number of symbols T is plotted in Fig. 4. It shows that at high SNR, less training is required to maximize the lower bound of data throughput $\mathcal{C}_L$, while at low SNR more training is needed. This agrees with our intuition. Furthermore, we notice that the optimal training number K is less than the current total number of symbols T, which suggests that the conventional superimposed training schemes, i.e. $K = T$, is not optimal.

VI. Conclusions

In this paper, the closed-form throughput lower bound taking into consideration the channel estimation error was derived based on a generalized superimposed training scheme. Our study shows that conventional PSAM performs better in the high SNR region while the superimposed training scheme performs better in the low SNR region.

References

[1] J. Cavers, "An analysis of training symbol assisted modulation for Rayleigh fading channels [mobile radio]," *IEEE Trans. Veh. Technol.*, vol. 40, no. 4, pp. 686–693, Nov. 1991.

[2] J. Tugnait and W. Luo, "On channel estimation using superimposed training and first-order statistics," *IEEE Commun. Lett.*, vol. 7, no. 9, pp. 413–415, Sept. 2003.

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

[4] 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 Lett.*, vol. 12, no. 3, pp. 226–229, Mar. 2005.

[5] A. Goljahani, N. Benvenuto, S. Tomasin, and L. Vangelista, "Superimposed Sequence Versus Pilot Aided Channel Estimations for Next Generation DVB-T Systems," *IEEE Trans. Broadcasting*, vol. 55, no. 1, pp. 140–144, March 2009.

[6] W. Huang, C. Li, and H. Li , "On the power allocation and system capacity of OFDM systems using superimposed training schemes," *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp.1731–1740, May 2009.

[7] H. L. Van Trees, *Optimum Array Processing Part IV: Detection and Estimation Theory*. New York: Wiley, 2002.

[8] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[9] M. Coldrey and P. Bohlin, "Training-based MIMO systems – Part I: performance comparison," *IEEE Trans. on Signal Processing*, vol. 55, no. 11, pp.5464–5476, Nov. 2006.