ON CHANNEL ESTIMATION USING SUPERIMPOSED TRAINING AND FIRST-ORDER STATISTICS

Jitendra K. Tugnait *Weilin Luo*

Department of Electrical & Computer Engineering
Auburn University, Auburn, Alabama 36849, USA

ABSTRACT

Channel estimation for single-input multiple-output (SIMO), possibly time-varying, channels is considered using only the first-order statistics of the data. The time-varying channel is assumed to be described by a complex exponential basis expansion model (CE-BEM). A periodic (non-random) training sequence is arithmetically added (superimposed) at a low power to the information sequence at the transmitter before modulation and transmission. Recently superimposed training has been used for time-invariant channel estimation assuming no mean-value uncertainty at the receiver. We propose a different method that explicitly exploits the underlying cyclostationary nature of the periodic training sequences. It is applicable to both time-invariant and time-varying systems. Unlike existing approaches we allow mean-value uncertainty at the receiver. Illustrative computer simulation examples are presented.

1. INTRODUCTION

Consider a time-varying SIMO (single-input multiple-output) FIR (finite impulse response) linear channel with N outputs. Let $\{s(n)\}$ denote a scalar sequence which is input to the SIMO time-varying channel with discrete-time impulse response $\{\mathbf{h}(n;l)\}$ (N-vector channel response at time n to a unit input at time $n-l$). The vector channel may be the result of multiple receive antennas and/or oversampling at the receiver. Then the symbol-rate, channel output vector is given by

$$\mathbf{x}(n) := \sum_{l=0}^{L} \mathbf{h}(n;l)s(n-l). \tag{1}$$

In a complex exponential basis expansion representation [6] it is assumed that

$$\mathbf{h}(n;l) = \sum_{q=1}^{Q} \mathbf{h}_q(l)e^{j\omega_q n} \tag{2}$$

where N-column vectors $\mathbf{h}_q(l)$ (for $q = 1, 2, \cdots, Q$) are time-invariant. Eqn. (2) is a basis expansion of $\mathbf{h}(n;l)$ in the time variable n onto complex exponentials with frequencies $\{\omega_q\}$. The noisy measurements of $\mathbf{x}(n)$ are given by

$$\mathbf{y}(n) = \mathbf{x}(n) + \mathbf{v}(n) \tag{3}$$

A main objective in communications is to recover $s(n)$ given noisy $\{\mathbf{x}(n)\}$. This requires knowledge of the channel impulse response. In training-based approach, $s(n) = c(n)$ = training sequence (known to the receiver) for (say) $n = 1, 2, \cdots, M$ and $s(n)$ for $n > M$ is the information sequence (unknown apriori to the receiver) [5]. Therefore, given $c(n)$ and corresponding noisy $\mathbf{x}(n)$, one estimates the channel via least-squares and related approaches. For time-varying channels, one has to send training signal frequently and periodically to keep up with the changing channel. This wastes resources. An alternative is to estimate the channel based solely on noisy $\mathbf{x}(n)$ exploiting statistical and other properties of $\{s(n)\}$ [5]. This is the blind channel estimation approach. In semi-blind approaches, there is a training sequence but one uses the non-training based data also to improve the training-based results: it uses a combination of training and blind cost functions. This allows one to shorten the training period. More recently [1]-[3] have explored a superimposed training based approach for time-invariant systems where one takes $s(n) = c(n) + b(n)$, $\{b(n)\}$ is the information sequence and $\{c(n)\}$ is a non-random periodic training (pilot) sequence. Exploitation of the periodicity of $\{c(n)\}$ allows identification of the channel without allocating any explicit time slots for training, unlike traditional training methods. There is no loss in information rate. *In this paper we consider both time-invariant and time-varying systems: if $Q = 1$ and $\omega_1 = 0$, we have a time-invariant system.* [1]-[3] dealt with only time-invariant systems.

This material is based upon work supported in part by the US Army Research Office under Grant DAAD19-01-1-0539 and in part by the US National Science Foundation under Grant ECS-9912523.

Superimposed Training Let

$$s(n) = b(n) + c(n) \tag{4}$$

in (1) where $\{b(n)\}$ is the information sequence and $c(n) = c(n + mP)\ \forall m, n$ is a non-random periodic sequence (superimposed training) with period P.

Assume the following:

- **(H1)** The time-varying channel $\{\mathbf{h}(n;l)\}$ satisfies (2) where the frequencies ω_q ($q = 1, 2, \cdots, Q$) are distinct and known with $\omega_q \in [0, 2\pi)$.
- **(H2)** $N \geq 1$.
- **(H3)** $\{b(n)\}$ is zero-mean, white with $E\{|b(n)|^2\} = 1$.
- **(H4)** $\{\mathbf{v}(n)\}$ is **nonzero-mean** ($E\{\mathbf{v}(n)\} = \mathbf{m}$), white, uncorrelated with $\{b(n)\}$, with $E\{[\mathbf{v}(n+\tau) - \mathbf{m}][\mathbf{v}(n) - \mathbf{m}]^H\} = \sigma_v^2 I_N \delta(\tau)$. The mean vector $\mathbf{m}$ is unknown.
- **(H5)** $c(n) = c(n + mP)\ \forall m, n$ is a non-random periodic sequence with period P.

As in [1]-[3] we will exploit the first-order statistics of the received signal. The corresponding time-invariant model in [1]-[3] does not include an unknown constant term (d.c. offset) in the measurement equation ($\mathbf{m}$ in (**H4**)) – it should if we exploit $E\{\mathbf{y}(n)\}$ to estimate the channel. In practice, linear systems arise because of linearization about some operating (set) point – "bias" in BJT/FET amplifiers, e.g. These set points are typically unknown (at least not known precisely) a priori, and one does not normally worry about them since unknown means are estimated and removed before processing (blocked by capacitor-coupling etc.) and they are not needed in any processing. However, if (time-varying) mean $E\{\mathbf{y}(n)\}$ is what we wish to use, then we must include a term such as nonzero $\mathbf{m}$.

As noted earlier [1]-[3] deal with time-invariant systems with zero $\mathbf{m}$ in their model. [3] proposes the choice $c(n) = \sum_k a\delta(n - kP)$ where $\delta(n)$ is the Kronecker delta function. As noted in [2], the choice of [3] leads to a poor peak-to-average power ratio of the transmitted signal which is highly undesirable if the transmit power amplifier has some nonlinearity. In [2] a more general approach is provided where general periodic superimposed training sequences are considered. A method to synthesize "optimal" channel-independent training sequences is provided in [2]; these training sequences of [2] yield the same channel estimator performance independent of the underlying unknown channel. The training sequence can be selected to yield a peak-to-average power ratio much better than that of [3].

In this paper we follow the basic ideas of [1]-[3] but propose a different method (it explicitly exploits the underlying cyclostationary nature of the periodic training sequences) which works for nonzero $\mathbf{m}$ in (**H4**) as well as for time-varying systems described by a CE-BEM (complex exponential basis expansion model) (2).

2. SUPERIMPOSED TRAINING-BASED SOLUTION

By (1)-(3) and (**H5**), we have

$$E\{\mathbf{y}(n)\} = E\{\mathbf{x}(n)\} + \mathbf{m}$$

$$= \sum_{q=1}^{Q}\sum_{l=0}^{L} \mathbf{h}_q(l) e^{j\omega_q n} c(n-l) + \mathbf{m}. \tag{5}$$

Since $\{c(n)\}$ is periodic, we have

$$c(n) = \sum_{m=0}^{P-1} c_m e^{j\alpha_m n} \;\; \forall n, \;\; \alpha_m := 2\pi m/P, \tag{6}$$

where

$$c_m = \frac{1}{P}\sum_{n=0}^{P-1} c(n) e^{-j\alpha_m n}. \tag{7}$$

The coefficients c_ms are known at the receiver since $\{c(n)\}$ is known. Assume that $c_m \neq 0 \; \forall m$. Therefore, we have

$$E\{\mathbf{y}(n)\} = \sum_{q=1}^{Q}\sum_{m=0}^{P-1} \underbrace{\left[\sum_{l=0}^{L} c_m \mathbf{h}_q(l) e^{-j\alpha_m l}\right]}_{=:\mathbf{d}_{mq}} e^{j(\omega_q+\alpha_m)n} + \mathbf{m}. \tag{8}$$

Suppose that we pick P to be such that $(\omega_q + \alpha_m)$s are all distinct for any choice of m and q. Then $E\{\mathbf{y}(n)\}$ is (almost) periodic [7] with cycle frequencies $(\omega_q + \alpha_m)$, $1 \leq q \leq Q$, $0 \leq m \leq P-1$. A consistent (mean-square (m.s.) sense and in probability (i.p.)) estimate $\hat{\mathbf{d}}_{mq}$ of $\mathbf{d}_{mq}$, for $\omega_q + \alpha_m \neq 0$, follows as [7]

$$\hat{\mathbf{d}}_{mq} = \frac{1}{T}\sum_{n=1}^{T} \mathbf{y}(n) e^{-j(\omega_q+\alpha_m)n}. \tag{9}$$

As $T \to \infty$, $\hat{\mathbf{d}}_{mq} \to \mathbf{d}_{mq}$ m.s. and i.p. if $\omega_q + \alpha_m \neq 0$ and $\hat{\mathbf{d}}_{0q} \to \mathbf{d}_{0q} + \mathbf{m}$ m.s. and i.p. if $\omega_q + \alpha_m = 0$. In the time-invariant system case $\omega_q + \alpha_m = 0$ is true iff $m = 0$.

We now establish that given $\mathbf{d}_{mq}$ for $1 \leq q \leq Q$ and $1 \leq m \leq P-1$, we can (uniquely) estimate $\mathbf{h}_q(l)$s if $P \geq L+2$ and $\omega_q + \alpha_m \neq 0$. Since $\mathbf{m}$ is unknown, we will omit the term $m = 0$ for further discussion. For $1 \leq m \leq P-1$ define (an NQ-column vector)

$$\mathbf{D}_m := [\mathbf{d}_{m1}^T, \; \mathbf{d}_{m2}^T, \; \cdots, \; \mathbf{d}_{mQ}^T]^T \tag{10}$$

and for $0 \leq l \leq L$, define (an NQ-column vector)

$$\mathbf{H}_l := [\mathbf{h}_1^T(l), \; \mathbf{h}_2^T(l), \; \cdots, \; \mathbf{h}_Q^T(l)]^T. \tag{11}$$

Then we have

$$\mathbf{D}_m = \sum_{l=0}^{L} c_m e^{-j\alpha_m l}\mathbf{H}_l, \;\; 0 \leq m \leq P-1. \tag{12}$$

Omitting the term $m = 0$, (12) leads to

$$\underbrace{\begin{bmatrix} c_1 I_{NQ} & c_1 I_{NQ} e^{-j\alpha_1} & \cdots & c_1 I_{NQ} e^{-j\alpha_1 L} \\ c_2 I_{NQ} & c_2 I_{NQ} e^{-j\alpha_2} & \cdots & c_2 I_{NQ} e^{-j\alpha_2 L} \\ \vdots & \vdots & \vdots & \vdots \\ c_{P-1} I_{NQ} & c_{P-1} I_{NQ} e^{-j\alpha_{P-1}} & \cdots & c_{P-1} I_{NQ} e^{-j\alpha_{P-1} L} \end{bmatrix}}_{=:\mathcal{C}}$$

$$\times \underbrace{\begin{bmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_L \end{bmatrix}}_{=:\mathcal{H}} = \underbrace{\begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \\ \vdots \\ \mathbf{D}_{P-1} \end{bmatrix}}_{=:\mathcal{D}}. \tag{13}$$

Since α_ms are distinct and $c_m \neq 0 \; \forall m$, rank$(\mathcal{C}) = NQ(L+1)$ if $P \geq L+2$; hence, we can determine $\mathbf{h}_q(l)$s uniquely. Define

$$\hat{\mathbf{D}}_m := [\hat{\mathbf{d}}_{m1}^T, \; \hat{\mathbf{d}}_{m2}^T, \; \cdots, \; \hat{\mathbf{d}}_{mQ}^T]^T \tag{14}$$

and define $\hat{\mathcal{D}}$ as in (12) with $\mathbf{d}_{mq}$s replaced with $\hat{\mathbf{d}}_{mq}$s. Then we have the channel estimate

$$\hat{\mathcal{H}} = (\mathcal{C}^H\mathcal{C})^{-1}\mathcal{C}^H\hat{\mathcal{D}}. \tag{15}$$

We summarize our method in the following Lemma:

Lemma. Under (**H1**)-(**H5**), the channel estimator (15) is consistent in probability if the periodic training sequence is such that $c_m \neq 0 \; \forall m$, $P \geq L+2$ and P is such that $\omega_q + \alpha_m \neq 0 \; \forall q$ and $m \neq 0$. •

Remark 1. For time-invariant channels ($Q = 1$ and $\omega_1 = 0$), any P satisfies $\omega_q + \alpha_m \neq 0 \; \forall q$ and $m \neq 0$. □

Remark 2. Precise knowledge of the channel length L is not required; an upperbound L_u suffices. Then we estimate $\mathbf{H}_i$ for $0 \leq i \leq L_u$ with $\hat{\mathbf{H}}_i \to 0$ i.p. for $i \geq L+1$ (=true channel length) as record length $T \to \infty$. □

Remark 3. We do not need $c_m \neq 0$ for every m. We need at least $L+2$ nonzero c_ms. This can be accomplished by picking a "large" P and a suitable $\{c(n)\}$ (picked to satisfy a peak-to-average power constraint, e.g.). □

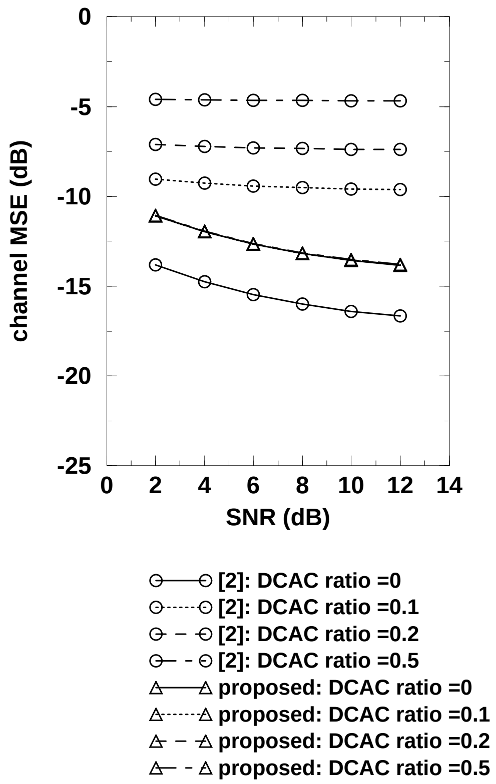

Fig. 1. Example 1: Normalized channel MSE (19) based on T =144 symbols per run, 100 Monte Carlo runs, $P = 6$. DCAC ratio $= \frac{[E\{v(n)\}]^2}{E\{|y(n)-v(n)|^2\}}$. The curves for the proposed method for different DCAC ratios are overlaid (very close).

3. SIMULATION EXAMPLES

3.1. Example 1: Time-Invariant SISO Channel

Take $N = 1$, $Q = 1$ and $\omega_1 = 0$ in (2) to get

$$y(n) = \sum_{l=0}^{L} h(l)[b(n-l) + c(n-l)] + v(n). \quad (16)$$

Let L_u be the upper bound on channel length L. We take $L_u = 4$. We consider a randomly generated channel in each Monte Carlo run with random channel length and random channel coefficients. The input information sequence $\{b(n)\}$ is i.i.d. equiprobable 4-QAM (quadrature amplitude modulation) taking values $(\pm 1 \pm j)/\sqrt{2}$. The channel length was picked randomly as $L \in \{0, 1, 2\}$ with equal probability. For a given channel length, the elements of $h(l)$, $0 \leq l \leq L$, were taken to be mutually independent complex random variables with independent real and imaginary parts, each uniformly distributed over the interval $[-1, 1]$.

The training sequence was chosen to have $P = 6$ with $\{c(n)\}$ as in [2], namely

$$\{c(n)\}_{n=0}^{5} = \{2.01, -0.87, 0.35, -0.14, 0.08, 1.02\}. \quad (17)$$

Let σ_b^2 and σ_c^2 denote the average power in the information sequence $\{b(n)\}$ and training sequence $\{c(n)\}$, respectively. As in [2] define a power loss factor

$$\alpha = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_c^2} \quad (18)$$

and power loss -10log(α) dB, as a measure of the information data power loss due to the inclusion of the training sequence. Here

$$\sigma_b^2 := E\{|b(n)|^2\}, \quad \sigma_c^2 := \frac{1}{P}\sum_{n=0}^{P-1} |c(n)|^2.$$

The training sequence was scaled to achieve a desired power loss. Complex white zero-mean Gaussian noise was added to the received signal and scaled to achieve a desired signal-to-noise (SNR) ratio at the receiver (relative to the contribution of $\{s(n)\}$). A mean-value m was added to the noisy received signal to achieve a specified DCAC power ratio $\frac{m^2}{E\{|y(n)-v(n)|^2\}}$.

Normalized mean-square error in estimating the channel impulse response averaged over 100 Monte Carlo runs, was taken as the performance measure for channel estimation. It is defined as (before Monte Carlo averaging)

$$\text{NCMSE} := \frac{\sum_{l=0}^{2} \|h(l) - \hat{h}(l)\|^2}{\sum_{l=0}^{2} \|h(l)\|^2} \quad (19)$$

The results of averaging over 100 Monte Carlo runs are shown in Fig. 1 for various SNRs and DCAC power ratios for a record length of T=144 symbols and a power loss of 2dB. Our proposed method (using $L = L_u = 4$ in (15)) and that of [2] were simulated. The method of [3] does not apply to this model. It is seen that the proposed method is insensitive to the presence of the unknown mean m whereas the method of [2] is very sensitive. For $m = 0$, the performance of our method is slightly inferior to that of [2].

3.2. Example 2: Time-Invariant SISO Channel

This example is exactly as Example 1 except for the training sequence which was taken to an m-sequence (maximal length pseudo-random binary sequence) of length 7 (=P)

$$\{c(n)\}_{n=0}^{6} = \{1, -1, -1, 1, 1, 1, -1\}. \quad (20)$$

The peak-to-average power ratio for this sequence is one (the best possible). The results of averaging over 100 Monte Carlo runs are shown in Fig. 2 for various SNRs and DCAC power ratios for a record length of T=140 symbols and a power loss of 2dB. Our proposed method and that of [2] were simulated. The method of [3] does not apply to this model. It is seen that as for Example 1, the proposed method is insensitive to the presence of the unknown mean m whereas the method of [2] is very sensitive. Unlike Example 1, for $m = 0$, the performance of our method is now slightly superior to that of [2].

3.3. Example 3: Time-Varying SISO Channel

In (2) take $N = 1$, $Q = 2$ and

$$\omega_1 = 0, \quad \omega_2 = 2\pi/50. \quad (21)$$

We consider a randomly generated channel in each Monte Carlo run with random channel length $L \in \{0, 1, 2\}$ picked with equal probabilities and random channel coefficients $h_q(l)$, $0 \leq l \leq L$, taken to be mutually independent complex random variables with independent real and imaginary parts, each uniformly distributed over the interval $[-1, 1]$. Normalized mean-square error (MSE) in estimating the channel coefficients $h_q(l)$, averaged over 100 Monte

Carlo runs, was taken as the performance measure for channel identification. It is defined as (before Monte Carlo averaging)

$$\mathrm{NCMSE}_{tv} := \frac{\left\{\sum_{q=1}^{Q}\sum_{m=0}^{2}\|h_q(m)-\widehat{h}_q(m)\|^2\right\}}{\sum_{q=1}^{Q}\sum_{m=0}^{2}\|h_q(m)\|^2} \qquad (22)$$

The training sequence was taken to be an m-sequence of length 7 specified by (20). The input information sequence $\{b(n)\}$ is i.i.d. equiprobable 4-QAM.

Our proposed method using $L = L_u = 4$ in (15) was applied for varying power losses due to the training sequence. The power loss is defined in Example 1 as -10log(α) dB where $\alpha = \frac{\sigma_b^2}{\sigma_b^2+\sigma_c^2}$. Fig. 3 shows the simulation results.

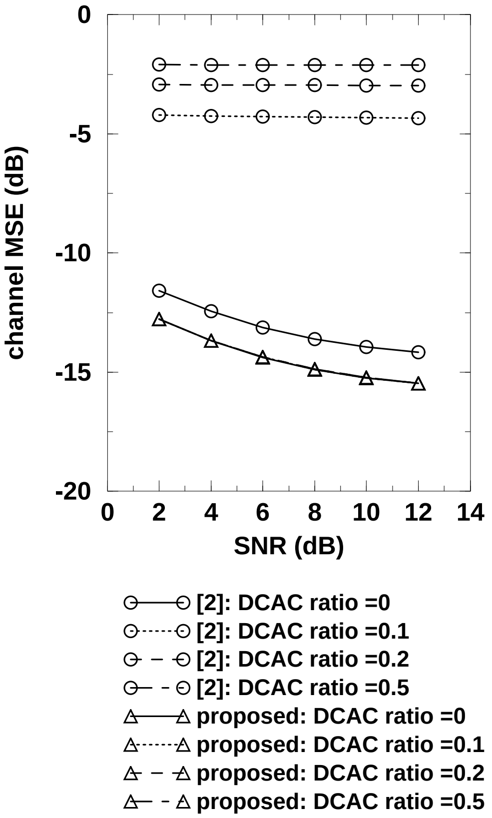

Fig. 2. Example 2: Normalized channel MSE (19) based on T =140 symbols per run, 100 Monte Carlo runs, $P = 7$. DCAC ratio $= \frac{[E\{v(n)\}]^2}{E\{|y(n)-v(n)|^2\}}$. The curves for the proposed method for different DCAC ratios are overlaid (very close).

4. REFERENCES

[1] F. Mazzenga, "Channel estimation and equalization for M-QAM transmission with a hidden pilot sequence," *IEEE Trans. Broadcasting*, vol. 46, pp. 170-176, June 2000.

[2] A.G. Orozco-Lugo, M.M. Lara and D.C. McLernon, "Blind channel estimation using implicit training," *IEEE Trans. Signal Proc.*, in review (preprint).

[3] G.T. Zhou, M. Viberg and T. McKelvey, "Superimposed periodic pilots for blind channel estimation," in *Proc. 35th Annual Asilomar Conf. Signals Systems Computers*, pp. 653-657, Pacific Grove, CA, Nov. 5-7 , 2001.

[4] T. McKelvey, "Periodic excitation for identification of dynamic errors-in-variables systems operating in closed loop," in *Proc. 13th World Congress of IFAC*, vol. J, pp. 155-160, 1996.

[5] J.K. Tugnait, L. Tong and Z. Ding, "Single-user channel estimation and equalization," *IEEE Signal Proc. Mag.*, vol. 17, pp. 16-28, May 2000.

[6] G.B. Giannakis and C. Tepedelenlioglu, "Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels," *Proc. IEEE* , vol. 86, pp. 1969-1986, Oct. 1998.

[7] A.V. Dandawate and G.B. Giannakis, "Asymptotic theory of mixed time average and kth-order cyclic-moment and cumulant statistics," *IEEE Trans. Inform. Theory*, vol. IT-41, pp. 216-232, Jan. 1995.

[8] J.K. Tugnait and W. Luo, "On blind identification of SIMO time-varying channels using second-order statistics," in *Proc. 35th Annual Asilomar Conf. Signals Systems Computers*, Pacific Grove, CA, pp. 747-752, Nov. 5-7, 2001.

[9] J.K. Tugnait and W. Luo, "Linear prediction error method for blind identification of periodically time-varying channels," *IEEE Trans. Signal Processing*, vol. SP-50, pp. 3070-3082, Dec. 2002.

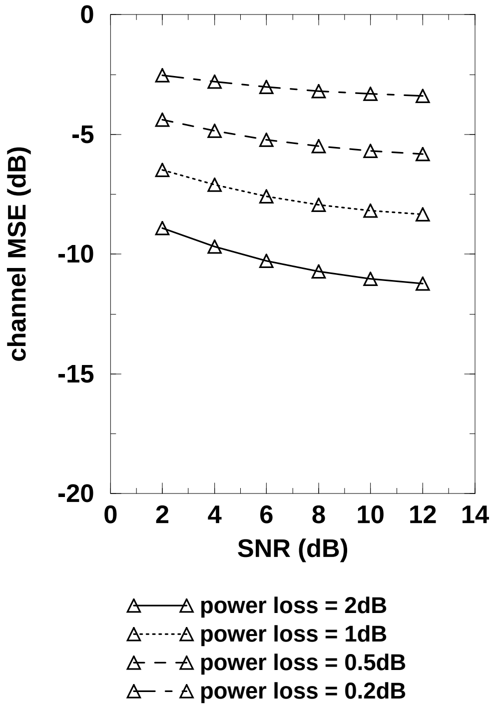

Fig. 3. Example 3: Normalized channel MSE (22) based on T =140 symbols per run, 100 Monte Carlo runs, $P = 7$. Power loss = -10log(α) dB where α is as in (18).