SUPERIMPOSED OR TIME-MULTIPLEXED TRAINING: A PERFORMANCE COMPARISON

Abla Kammoun[1] *Karim Abed-Meraim*[1,2] *and Sofiène Affes*[3]

Télécom ParisTech 46, rue Barrault, 75634 Paris Cedex 13, France[1]
ECE Department, College of Engineering, University of Sharjah, P.O. Box 27272, UAE[2]
INRS-EMT, 800, de la Gauchetière, Bureau 6900, Montreal, QC, H5A 1K6 Canada[3]
E-mails: kammoun, abed@tsi.enst.fr, affes@emt.inrs.ca

ABSTRACT

In this paper, we derive a performance comparison between the conventional time-multiplexing training based scheme and the most-recently proposed data-dependent superimposed training scheme, when using linear receivers. For both schemes, we derive an approximative closed-form expression for the Bit Error Rate (BER) and determine the optimal power allocation between pilot and data that minimizes the BER. Simulations are conducted to assess the accuracy of the provided expressions and to determine contexts for which it is more interesting to opt for the data-dependent superimposed training scheme.

Index Terms— Bit error rate, data-dependent superimposed training, time multiplexing, optimal power allocation.

1. INTRODUCTION

Time-Division Multiplexed Training (TDMT) is the most commonly used technique for channel estimation [1]. Because of its simplicity, it has been used in many practical communication systems, e.g. in Global System for Mobile Communications (GSM) [2]. Although accurate channel estimation can be obtained with low-computational-complexity receivers, this technique results in a low bandwidth efficiency, especially when a large number of pilots is required.

Recently, Superimposed Training (ST) has evolved as a new promising alternative to TDMT schemes due to its high bandwidth efficiency. It consists in transmitting pilot and data symbols simultaneously at the same time and on the same frequency domains. Channel estimation is then performed by treating data as an additive source of noise. The first proposed (ST) schemes use induced cyclostationarity (eg. periodic pilot sequence) to *asymptotically* mitigate the cross-correlation between training and data symbols, [3, 4, 5]. The major problem in this case is the data-pilot non-zero cross-correlation for small or finite sample sizes, thus limiting its potential over TDMT schemes to the cases of multi-carrier based systems [6] or short channel coherence times [7].

In order to enhance the channel estimation quality in superimposed training based systems, Ghogho *et al.* proposed in [8, 9] to introduce linear distortion to the data prior to insertion of the pilot symbols so as to guarantee the orthogonality between pilot and data sequences for finite length data frames. The proposed method was referred to as the Data-Dependent Superimposed Training (DDST) scheme and was shown in [10] to outperform the conventional superimposed training technique.

In this paper, we derive a performance comparison between the TDMT and DDST schemes. More particularly, we provide approximative theoretical closed-form expressions for the BER and also determine for both schemes the optimal power allocation between data and pilot[1]. For applications in which the BER should be less than a given threshold, we show that the DDST scheme can be interesting only for large frames.

Notations: Subscriptis $^{\text{H}}$ and $^{\#}$ denote, respectively, hermitian and pseudo-inverse operators. The $(K \times K)$ identity matrix is denoted by $\mathbf{I}_K$. The (i, j)-th entry of a matrix $\mathbf{A}$ is denoted by $[\mathbf{A}]_{i,j}$ or $\mathbf{A}(i, j)$. $\|.\|$ denotes the Euclidean norm of a vector. Moreover, $\mathbf{1}_K$ denotes the $(K \times K)$ matrix of all ones. $\Re$ denotes the real part of a complex entity.

2. SYSTEM MODEL

2.1 Time-Division Multiplexing

We consider a $M \times K$ Multiple-Input Multiple Output (MIMO) system operating over a flat fading channel. Two phases are considered:

First Phase: In the first phase, each transmitting antenna sends N_1 pilot symbols. The received symbol $\mathbf{Y}_1$ writes as:

$$\mathbf{Y}_1 = \mathbf{H}\mathbf{P}_t + \mathbf{V}_1$$

where $\mathbf{H}$ is the $M \times K$ channel matrix with independent and identically distributed (iid) Gaussian variables with zero mean and variance $\frac{1}{K}$, $\mathbf{V}_1$ is the $M \times N_1$ matrix whose entries are iid zero mean with variance σ_v^2 and $\mathbf{P}_t$ is the $K \times N_1$ pilot matrix. It is well known that the Mean Square Error of the channel estimation is minimized subject to a fixed training power σ_P^2, when the pilot matrix satisfies [11]:

$$\mathbf{P}_t \mathbf{P}_t^{\text{H}} = N_1 \sigma_P^2 \mathbf{I}_K.$$

Second Phase: In the second phase, N_2 data symbols with power σ_w^2 are sent by each antenna so that the received signal $\mathbf{Y}_2$ writes as:

$$\mathbf{Y}_2 = \mathbf{H}\mathbf{W}_t + \mathbf{V}_2,$$

where $\mathbf{W}_t$ is the $K \times N_2$ data matrix with iid data symbols of power σ_w^2 and $\mathbf{V}_2$ is the $M \times N_2$ additive noise matrix.

2.2 Data-Dependent-Based Scheme

One of the major shortcomings of Conventional Superimposed Training (CST) schemes is the low quality of the channel estimation caused by the embedded unknown data which

[1]Note that this optimal power allocation and the one in [3] are distinct as the considered pilot design schemes are different.

acts as an additive source of noise. In order to improve the channel estimation quality, Ghogho *et al.* [8] proposed to distort the data so that it becomes orthogonal to the training sequence. The proposed distortion matrix $\mathbf{D}$ is defined by:

$$\mathbf{D} = \mathbf{I}_N - \mathbf{J}$$

where $\mathbf{J} = \frac{K}{N}\mathbf{1}_{\frac{N}{K}} \otimes \mathbf{I}_K$ (we assume that $\frac{N}{K}$ is integer valued, N being the sample size). This distortion matrix was shown to be optimal in the sense that it minimizes the averaged Euclidian distance between the distorted and non-distorted data, [12]. The received signal at each block is therefore given by:

$$\mathbf{Y} = \mathbf{H}\mathbf{W}_d(\mathbf{I}_N - \mathbf{J}) + \mathbf{H}\mathbf{P}_d + \mathbf{V}$$

where $\mathbf{W}_d$ is the data matrix with iid data symbols of power $\sigma_{w'}^2$, $\mathbf{P}_d$ is the $K \times N$ training matrix and $\mathbf{V}$ is the $M \times N$ matrix whose entries are i.i.d zero mean with variance σ_v^2. The chosen pilot matrix $\mathbf{P}_d$ should fulfill two requirements. It should be orthogonal to the distorsion matrix $\mathbf{D}$, thus satisfying $\mathbf{D}\mathbf{P}_d^{\mathrm{H}} = \mathbf{0}$, and also verify the orthogonality relation $\mathbf{P}_d\mathbf{P}_d^{\mathrm{H}} = N\sigma_{p'}^2\mathbf{I}_K$ in order to minimize the channel estimation error subject to fixed training power. A possible pilot matrix that fulfills these requirement is [8]: $\mathbf{P}_d(k,n) = \sqrt{\sigma_{p'}^2}\exp(j2\pi kn/K)$, with $k = 0,\cdots,K-1$ and $n = 0,\cdots,N-1$.

3. CHANNEL ESTIMATION AND DATA DETECTION

3.1 Time-Division Multiplexing

In the first phase, we assume that the receiver estimates the channel in the least square sense. Hence, the channel estimate is given by:

$$\begin{aligned}\widehat{\mathbf{H}}_t &= \mathbf{Y}_1\mathbf{P}_t^{\mathrm{H}}(\mathbf{P}_t\mathbf{P}_t^{\mathrm{H}})^{-1} \\ &= \mathbf{H} + \mathbf{V}_1\mathbf{P}_t^{\mathrm{H}}(\mathbf{P}_t\mathbf{P}_t^{\mathrm{H}})^{-1} \\ &= \mathbf{H} + \Delta\mathbf{H}_t\end{aligned}$$

where $\Delta\mathbf{H}_t = \mathbf{V}_1\mathbf{P}_t^{\mathrm{H}}(\mathbf{P}_t\mathbf{P}_t^{\mathrm{H}})^{-1}$.

In the data transmission phase, the linear receiver uses the channel estimate in order to retrieve the transmitted data, thus yielding the estimated data matrix given by:

$$\widehat{\mathbf{W}}_t = \left(\widehat{\mathbf{H}}_t\right)^{\#}\mathbf{Y}_2. \tag{1}$$

Assuming that the channel estimation error is small, the pseudo-inverse of the estimated matrix can be approximated by the linear part of the Taylor expansion as:

$$\left(\widehat{\mathbf{H}}_t\right)^{\#} = \mathbf{H}^{\#} - \mathbf{H}^{\#}(\Delta\mathbf{H}_t)\mathbf{H}^{\#}.$$

Hence, the zero-forcing estimate of the transmitted matrix can be expressed as:

$$\widehat{\mathbf{W}}_t = \mathbf{W}_t - \mathbf{H}^{\#}\Delta\mathbf{H}_t\mathbf{W}_t + \left(\mathbf{H}^{\#} - \mathbf{H}^{\#}\Delta\mathbf{H}_t\mathbf{H}^{\#}\right)\mathbf{V}_2.$$

Consequently, the effective post-processing noise $\Delta\mathbf{W}_t = \widehat{\mathbf{W}}_t - \mathbf{W}_t$ could be written as:

$$\Delta\mathbf{W}_t = -\mathbf{H}^{\#}\Delta\mathbf{H}_t\mathbf{W} + \left(\mathbf{H}^{\#} - \mathbf{H}^{\#}\Delta\mathbf{H}_t\mathbf{H}^{\#}\right)\mathbf{V}_2.$$

3.2 Data-Dependent Superimposed Training

For the data-dependent superimposed training scheme, the channel estimate is given by:

$$\begin{aligned}\widehat{\mathbf{H}}_d &= \mathbf{Y}\mathbf{P}_d^{\mathrm{H}}(\mathbf{P}_d\mathbf{P}_d^{\mathrm{H}})^{-1} \\ &= \mathbf{H} + \Delta\mathbf{H}_d\end{aligned}$$

where $\Delta\mathbf{H}_d = \mathbf{V}\mathbf{P}_d^{\mathrm{H}}\left(\mathbf{P}_d\mathbf{P}_d^{\mathrm{H}}\right)^{-1}$. The zero-forcing estimate is given by:

$$\begin{aligned}\widehat{\mathbf{W}}_d &= \left(\widehat{\mathbf{H}}_d\right)^{\#}\mathbf{Y}(\mathbf{I}_N - \mathbf{J}) \\ &= \mathbf{W}_d + \Delta\mathbf{W}_d\end{aligned}$$

where using the Taylor expansion, we have:

$$\begin{aligned}\Delta\mathbf{W}_d &= -\mathbf{W}\mathbf{J} - \mathbf{H}^{\#}\Delta\mathbf{H}_d\mathbf{W}(\mathbf{I}_N - \mathbf{J}) \\ &\quad + \left(\mathbf{H}^{\#} - \mathbf{H}^{\#}\Delta\mathbf{H}_d\mathbf{H}^{\#}\right)\mathbf{V}(\mathbf{I}_N - \mathbf{J}).\end{aligned}$$

4. PERFORMANCE ANALYSIS

For finite system dimensions, the performance analysis of the TDMT and DDST schemes is difficult. Instead, we will work under the asymptotic regime when N, K and M grow to infinity with a constant rate, $\frac{K}{N} \to c_1$, with $0 < c_1 < 1$ and $\frac{M}{K} \to c_2 > 1$. For the TDMT scheme, we assume also that N_1 and N_2 go to infinity such that $\frac{N_2}{N_1} \to r$. Moreover, the notation $N \to \infty$ will refer to this asymptotic regime. Note that all proofs are omitted due to space limitation but are available on line at [13].

4.1 Asymptotic Post-Processing Noise Distribution

Under the asymptotic regime, it is possible to prove the asymptotic convergence of the post-processing noise by using the 'characteristic function' approach:

4.1.1 Time-division multiplexing

Theorem 1 *Under the asymptotic regime, and conditioned on the channel, the post-processing noise experienced by the i-th antenna at each time k for the TDMT scheme behaves asymptotically as a Gaussian random variable:*

$$\mathbb{E}\left[e^{j\Re(s^*\Delta\mathbf{W}_t(i,k))}\right] - e^{-\frac{\sigma_w^2\delta_t\left[\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\right]_{i,i}|s|^2}{4}} \xrightarrow[N\to\infty]{} 0$$

where

$$\delta_t = c_1(1+r)\frac{\sigma_v^2}{\sigma_P^2} + \frac{\sigma_v^2}{\sigma_w^2} + \frac{c_1(1+r)\sigma_v^4}{\sigma_w^2\sigma_P^2(c_2-1)}.$$

4.1.2 Data-dependent superimposed training

Theorem 2 *Under the asymptotic regime, and conditioned on the channel, the post-processing noise experienced by the i-th antenna at each time k behaves asymptotically as a Gaussian mixture random variable, i.e:*

$$\mathbb{E}\left[e^{j\Re s^*\Delta\mathbf{W}_d(i,k)}\right] - \sum_{i=1}^{\mathscr{Q}} p_i e^{(j\Re(s^*\alpha_i))} e^{-\frac{|s|^2\delta_d\sigma_w^2\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i}}{4}} \xrightarrow[N\to\infty]{} 0$$

where $\mathscr{Q}$ is the cardinal of the set of all possible values of $\overline{\mathbf{W}}(i,k) = c_1 \sum_{k=1}^{1/c_1} \mathbf{W}(i,k)$ and p_i is the probability that $\overline{\mathbf{W}}(i,k)$ takes the value α_i. Moreover, δ_d is given by:

$$\delta_d = (1-c_1)\left(\frac{c_1\sigma_v^2}{\sigma_{P'}^2} + \frac{\sigma_v^2}{\sigma_{w'}^2} + \frac{c_1\sigma_v^4}{(c_2-1)\sigma_{P'}^2\sigma_{w'}^2}\right).$$

4.2 Bit Error Rate Expression

In the sequel, we derive the closed-form expression for the BER under QPSK constellation and Gray encoding.

4.2.1 Time-division multiplexing

Following the same lines as in [14], it can be proved that for the TDMT based scheme, the Bit Error Rate (BER) is given by:

$$\mathrm{BER} = J(M-K+1, K\delta_t, 1), \tag{2}$$

where:

$$J(m,a,b) = \frac{a^m}{\Gamma(m)}\int_0^\infty e^{-ax}x^{m-1}Q(\sqrt{bx})dx. \tag{3}$$

Let $c = \frac{b}{2a}$. Then, it can be easily seen that if $c = 0$, $J(m,a,b) = 0.5$. If c is strictly positive, then this integral has the following closed-form expression [15]:

$$J(m,a,b) = \frac{\sqrt{c/\pi}\,\Gamma(m+\frac{1}{2})}{2(1+c)^{m+\frac{1}{2}}\Gamma(m+1)}\,{}_2F_1(1, m+\frac{1}{2}, m+1, \frac{1}{1+c}), \tag{4}$$

where ${}_2F_1(p,q,n,z)$ is the hypergeometric function [16], and Γ is the complete gamma function defined as:

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt.$$

4.2.2 Data-dependent superimposed training

Let $\mathbf{N}(i,j)$ denote the post-processing noise such that we have:

$$\widehat{\mathbf{W}}_d(i,j) = \mathbf{W}_d(i,j) + \mathbf{N}(i,j).$$

As it has been shown, $\mathbf{N}_{i,j}$ behaves as a mixture of Gaussian random variables. Using the symmetry of noise and data, the BER expression under QPSK constellation is given by:

$$\begin{aligned}
\mathrm{BER} &= P\left[\Re\widehat{\mathbf{W}}_d(i,j) > 0 \,|\, \Re\mathbf{W}_d(i,j) = -\sqrt{\frac{\sigma_w}{2}}\right] \\
&= P\left[\Re(\mathbf{N}(i,j)) > \sqrt{\frac{\sigma_w}{2}}\right] \\
&= \sum_{k=1}^{\mathscr{Q}} p_k \mathbb{E} Q\left(\sqrt{X}\left(1-\sqrt{\frac{\Re\alpha_k}{\sigma_w}}\right)\right)
\end{aligned}$$

where the expectation is taken over the distribution of $X = \frac{1}{\delta_d\left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\right]_{i,i}}$ given by:

$$f_X(x) = \frac{(K\delta_d)^{M-K+1}x^{M-K}}{(M-K)!}e^{-K\delta_d x}.$$

Hence, the BER expression is given by:

$$\mathrm{BER} = \frac{1}{2^{\frac{1}{c_1}-1}}\mathbb{E}\sum_{k=1}^{\frac{1}{c_1}-1}\binom{k}{\frac{1}{c_1}-1}Q\left(\left(1+c_1\left(2k-\frac{1}{c_1}\right)\right)\sqrt{X}\right).$$

Using (3), the BER can be expressed as:

$$\mathrm{BER} = \frac{1}{2^{\frac{1}{c_1}-1}}\sum_{k=1}^{\frac{1}{c_1}-1} J(M-K+1, K\delta_d, (1+c_1(2k-\frac{1}{c_1}))^2) \tag{5}$$

where $\frac{1}{c_1} = \frac{N}{K}$, is assumed to be integer.

5. OPTIMAL POWER ALLOCATION

Referring to (2) and (5), we can easily see that the optimal amount of power allocated to data and pilot for the TDMT scheme (resp. for the DDST scheme) is the one that minimizes δ_t (resp. δ_d).

5.1 Time-Division Multiplexing Scheme

Let $r = N_2/N_1$ and $\tilde{c}_1 = (1+r)c_1$. Minimizing δ_t with respect to σ_w^2 and σ_P^2 under the constraint that $N_1\sigma_P^2 + N_2\sigma_w^2 = (N_1+N_2)\sigma_T^2$ (this is a simple second-order polynomial optimization), we get:

$$\sigma_w^2 = \frac{(1+r)\sigma_T^2\sqrt{r\left((1+r)\sigma_T^2+\frac{\tilde{c}_1\sigma_v^2}{c_2-1}\right)}}{r\left(\sqrt{r\left((1+r)\sigma_T^2+\frac{\tilde{c}_1\sigma_v^2}{c_2-1}\right)}+\sqrt{\tilde{c}_1(\left((1+r)\sigma_T^2+\frac{r\sigma_v^2}{c_2-1}\right)}\right)}, \tag{6}$$

$$\sigma_P^2 = \frac{r(1+r)\sigma_T^2\sqrt{\tilde{c}_1\left((1+r)\sigma_T^2+\frac{r\sigma_v^2}{c_2-1}\right)}}{r\left(\sqrt{r\left((1+r)\sigma_T^2+\frac{\tilde{c}_1\sigma_v^2}{c_2-1}\right)}+\sqrt{\tilde{c}_1\left((1+r)\sigma_T^2+\frac{r\sigma_v^2}{c_2-1}\right)}\right)}. \tag{7}$$

5.2 Data-Dependent Superimposed Training Scheme

The optimal data and pilot powers that minimize δ_d subject to $\sigma_{w'}^2 + \sigma_{P'}^2 = \sigma_T^2$ are given by:

$$\sigma_{w'}^2 = \frac{\sqrt{(1-c_1)\left(\sigma_T^2+\frac{c_1\sigma_v^2}{c_2-1}\right)}\sigma_T^2}{(1-c_1)\left(\sqrt{(1-c_1)\left(\sigma_T^2+\frac{c_1\sigma_v^2}{c_2-1}\right)}+\sqrt{c_1\sigma_T^2+\frac{c_1(1-c_1)\sigma_v^2}{c_2-1}}\right)}, \tag{8}$$

$$\sigma_{P'}^2 = \frac{\sqrt{c_1\sigma_T^2+\frac{c_1(1-c_1)\sigma_v^2}{c_2-1}}\sigma_T^2}{\sqrt{(1-c_1)\left(\sigma_T^2+\frac{c_1\sigma_v^2}{c_2-1}\right)}+\sqrt{c_1\sigma_T^2+\frac{c_1(1-c_1)\sigma_v^2}{c_2-1}}}. \tag{9}$$

6. SIMULATIONS

6.1 Accuracy of the Asymptotic Results

Despite being valid only for the asymptotic regime, our results are found to yield good accuracy even for very small system dimensions. Fig. 1 plots the empirical and theoretical BER using QPSK modulation for $N = 32$, $K = 2$, and $M = 4$ for the TDMT and DDST based schemes. All comparisons are conducted in the context when both schemes (TDMT and DDST) have the same total energy. The number of pilots is set to $N_1 = 2$ ($N2 = 30$) for the TDMT scheme. For low SNR values (SNR until 6 dB), both schemes achieve approximatively the same BER performance, and therefore the DDST scheme outperforms its TDMT counterpart in terms of data rate since it has a better bandwidth efficiency. For high SNR values, the noise caused by the data distorsion is higher than the Gaussian additive noise, thus affecting the performance of the DDST scheme.

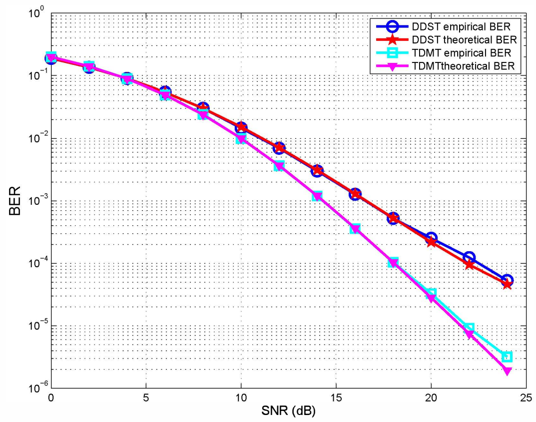

Figure 1: Theoretical and empirical BER for the TDMT and DDST based schemes.

6.2 Application 1

We consider an application in which the BER should be below a certain threshold, say 10^{-2}. This may be the case for instance of circuit-switched voice applications. We set the SNR $\triangleq \frac{\sigma_T^2}{\sigma_v^2}$ to 15 dB and the number of transmitting and receiving antennas to 2 and 4, respectively (i.e, $K = 2$ and $M = 4$). We then vary the ratio c_1 from 0.01 to 0.5. For each value of c_1, we compute the BER by using our results as illustrated in Fig. 2. We note that the data-dependent superimposed training may be interesting for low values of c_1 (say below 0.125), i.e., for long enough frames. For small frames (high distorsion ratio c_1), the distortion of the data becomes too high thus reducing the interest of the DDST scheme.

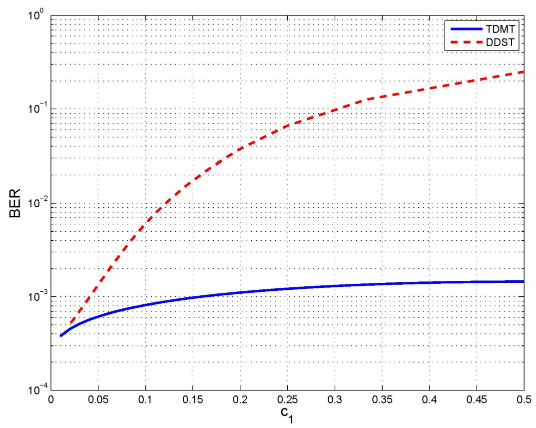

Figure 2: BER with respect to c_1 when $K = 2$, $M = 4$ and SNR=15 dB

6.3 Application 2

In this experiment, we also consider a scenario where the BER should be below 10^{-2}. We set the packet length N to 32 and the number of transmitting and receiving antennas to 2 and 4. Using (6), (7) and (2), we determine the minimum number of required pilot symbols to meet the BER lower bound requirement. We note that if the SNR is below 2 dB, the BER requirement could not be achieved. This is to be compared with the DDST scheme where the SNR should be set at least to 10.5 dB so as to meet the BER lower bound requirement as it can be shown in fig. 3. Moreover, for a BER more than 8.5 dB, the minimum number of pilot symbols for channel identification (equal to K) is sufficient to meet the BER requirement.

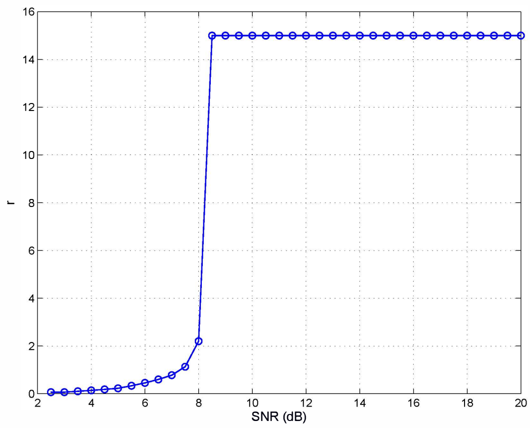

Figure 3: Required r versus SNR for BER $\leq 10^{-2}$.

7. CONCLUSION

Based on an asymptotic analysis, we have derived in this paper closed-form expressions for the BER for the TDMT and DDST schemes, when using zero-forcing detection. Based on these expressions, we have calculated the optimal pilot and data power expressions that minimize the BER. We have shown that in applications in which the BER should be less than a given threshold, the DDST scheme with linear receiver might be interesting for large enough frames. For the same kind of application, we have determined for the TDMT scheme the minimum required number of pilot symbols that could meet the BER lower bound requirement. However, overall, and for the considered context, the TDMT scheme seems to be more interesting (performant) than the DDST scheme when using zero-forcing receivers. Future extension of this work would be to analyse the performance of the DDST scheme when using non-linear receivers as the one considered in [8].

REFERENCES

[1] J. K. Cavers, "An Analysis of Pilot Symbol Assisted Modulation for Rayleigh Fading Channels," *IEEE Transactions Vehicular Technology*, vol. 40, pp. 686–693, November 1991.

[2] S. Redl, M. K. Weber, and M. Oliphant, *An Introduction to GSM*, Boston: Artech House, first edition, 1995.

[3] J. K. Tugnait and X. Meng, "On Superimposed Training for Channel Estimation: Performance Analysis, Training Power Allocation, and Frame Synchronization," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 752–765, February 2006.

[4] S. H. Jitendra and K. Tugnait, "On Superimposed Training for MIMO Channel Estimation and Symbol Detection," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 3007–3021, June 2007.

[5] A. G. Orozco-Lugo, M. Mauricio, and D. C. Mclernon, "Channel Estimation Using Implicit Training," *IEEE Transactions on Signal Processing*, vol. 52, no. 1, pp. 240–254, January 2004.

[6] M. A. Khalighi and S. Bourennane, "Semi-Blind Channel Estimation Based on Superimposed Pilots for Single-Carrier MIMO Systems," *IEEE Vehicular Technology Conference*, pp. 1480–1484, April 2007.

[7] P. Bohlin and M. Tapio, "Performance Evaluation of MIMO Communication Systems Based on Superimposed Pilots," *Interational Conference on Acoustics, Speech, and Signal Processing ICASSP*, vol. IV, pp. 425–428, May 2004.

[8] M. Ghogho and A. Swami, "Channel Estimation for MIMO Systems Using Data-Dependent Superimposed Training," *Allerton Conference*, 2004.

[9] M. Ghogho, D. McLernon, E. A. Hernandez, and A. Swami, "Channel estimation and symbol detection for block transmission using data-dependent superimposed training," *IEEE Signal Processing Letters*, vol. 12, no. 3, pp. 226–229, March 2005.

[10] A. Kammoun, K. Abed-Meraim, and S. Affes, "Performance of Linear Receivers Based on Superimposed Training," *IEEE SPAWC*, June 2007.

[11] B. Hassibi and B. M. Hochwald, "How Much Training is Needed in Multiple-Antenna Wireless Links?," *IEEE Trans. Information Theory*, vol. 49, no. 4, pp. 951–963, April 2003.

[12] M. Ghogho and A. Swami, "Optimal Training for Affine-Precoded and Cyclic-Prefixed Block Transmissions," *IEEE Workshop on Statistical Signal Processing*, July 2005.

[13] A. Kammoun, K. Abed-Meraim, and S. Affes, "A Performance Comparison Between Superimposed and Time-Multiplexed Training - Mathematical Derivation Details," Report#EMT-049-0519 available at http://www.wirelesslab.ca/en_1026_index.php, May 2009.

[14] R. Xu and F. C. M. Lau, "Performance Analysis for MIMO Systems Using Zero Forcing Detector over Rice Fading Channel," *IEEE International Symposium on Circuits and Systems*, vol. 5, pp. 4955– 4958, May 2005.

[15] T. Eng and L.B. Milstein, "Coherent DS-CDMA Performance in Nakagami Multipath Fading," *IEEE Transactions on Communications*, vol. 43, pp. 1134–1143, Feb./Mar./Apr. 1995.

[16] I. S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series and Products*, CA:Academic Press, Sixth edition, 2000.