FLAT FADING CHANNEL ESTIMATION BASED ON DIRTY PAPER CODING

*Gabriel Domínguez-Conde#, Pedro Comesaña-Alfaro#, and Fernando Pérez-González#**

#University of Vigo, EE Telecomunicación, Campus Universitario
* Gradiant (Galician Research and Development Center in Advanced Telecommunications)
36310 Vigo-SPAIN
{gdomin|pcomesan|fperez}@gts.uvigo.es

ABSTRACT

A novel complex flat fading channel estimation scheme is proposed. Contrarily to previous schemes in the literature, this new approach is not based on introducing pilot sequences, but on reducing the interference caused by the information-bearing signal on the estimation-aiding signal by using Dirty Paper Coding. We show through simulations that our method outperforms the Partially-Data Dependent scheme, which is a state-of-the-art technique based on superimposed pilots.

Index Terms— Channel estimation, dirty paper coding, superimposed pilots.

1. INTRODUCTION

Channel estimation is a transversal problem in signal processing; it is used in a number of applications, including digital communications (e.g., estimation of the channel parameters, automatic gain control, signal-to-noise ratio estimation, etc.), multimedia forensics (e.g., estimation of the linear filter used for post-processing an image), and acoustics (e.g., estimation of the acoustic response of a room).

There are mainly two approaches to the channel estimation problem: blind estimation and pilot-based estimation. In the former, certain underlying properties of the transmitted signal are exploited to estimate the channel. Those characteristics can be statistical, such as Higher Order Statistics (c.f., [1]), or deterministic, as in Constant Modulus Algorithms (c.f., [2]). On the other hand, pilot-based systems use part of the total power budget to transmit a signal, referred to as pilot or training signal, that is known to the receiver, so it can be used to infer the channel response. In most cases, the pilot signal is transmitted in an orthogonal subspace to that of the information-bearing signal, most often through either time-domain of frequency-domain multiplexing.

Pilot-based approaches have a number of well-known drawbacks [3, 4, 5]: **1)** In fast-varying channels, the training signals must be frequently sent in order to update the channel state information, thus wasting a significant amount of resources (in terms of bandwidth increase or loss in information rate), **2)** The information-bearing signal has to be shut down, requiring the implementation of additional logic to synchronize the pilot sequence slots (in whatever domain they are allocated) at both the transmitter and the receiver, **3)** The estimate is based on particular locations of the pilot sequences (typically locations at time and/or frequency); therefore, interpolation is frequently required in order to have channel estimates at other times/frequencies. On the other hand, blind estimation approaches suffer from slow convergence (i.e., a high number of samples of the received signal are required), and possible misconvergence [6].

More recently, the so-called *superimposed training* has been proposed as an alternative to the above approaches. In superimposed training a known pilot sequence (due to the parallelism with data hiding, we will name it *watermark*) is added to the information-bearing signal (which, similarly, we will call *host*). Since both signals are simply added, explicit allocation of time/frequency slots for training purposes is not required, in contrast to traditional training methods [7]. However, assuming that the transmitter has some fixed power budget, the information-bearing signal will suffer some power loss, and will be additionally distorted by the superimposed signal.

Unfortunately, as the host and pilot sequences are not orthogonal, the former will interfere on the pilot signal. This is a well-known problem in watermarking, where it is referred to as *host interference*, and it occurs in those schemes whenever a watermark independent of the host is added to the latter (as in Additive Spread Spectrum schemes [8]). In both fields solutions have been proposed that devote some of the available power to partially cancel the host-interference in the direction of the added sequence. These schemes were independently developed by Malvar and Florencio in 2003 [9] in the watermarking field, and by He and Tugnait in 2008 [4] for channel estimation, and they were named *Improved Spread Spectrum* (ISS) and *Partially-Data-Dependent* (PDD) *superimposed training*, respectively. Interestingly, to the best of the authors' knowledge this connection between PDD and ISS has not been reported to date.

Both ISS and PDD only partially cancel the host interference, thus leaving room for improvements. In fact, full host-interference rejection has been achieved in data hiding by exploiting the *Dirty Paper Coding* (DPC) paradigm, initially proposed by Costa in [10]. Adapting Costa's code construction, Chen and Wornell [11] proposed the use of Distortion Compensated-Quantization Index Modulation (DC-QIM) which, thanks to its host-rejection feature, leads to substantial performance improvements with respect to ISS.

In this work we pursue this connection between channel estimation and data hiding further to propose a flat fading estimation technique that is based on DPC principles and significantly benefits from a larger host cancellation than in PDD. We remark that this is the first time DPC codes are used for channel estimation purposes. For the sake of simplicity, we will focus our analysis on the complex flat fading channel (which is highly relevant in communications), leaving for future work the extension to more involved cases.

The remaining of this paper is organized as follows: Sect. 2 introduces the problem formulation, while Sect. 3 describes the proposed scheme. Then, Sect. 4 presents the experimental results, and conclusions are summarized in Sect. 5

Research supported by the Spanish Government under projects DYNACS (TEC2010-21245-C02-02/TCM), and the Galician Regional Government under projects GRC2013/009.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

1.1. Notation

Random variables are denoted with capital letters (e.g., X) and their outcomes with lowercase letters (e.g., x). Similarly, L-dimensional random vectors and their outcomes are denoted by bold letters (e.g., $\mathbf{X}$ and $\mathbf{x}$, respectively), and their jth component are indicated by a subindex (e.g., X_j and x_j). The magnitude of complex number x will be denoted by $|x|$, and its phase by $\measuredangle x$; furthermore, sometimes some notation abuse will be considered by denoting the componentwise magnitude of complex vector $\mathbf{x}$ as $|\mathbf{x}|$. The probability density function (pdf) of a random variable is denoted by $f_{\cdot}(\cdot)$ and its standard deviation by $\sigma_{\cdot}$; in both cases the letter denoting the random variable is in the subindex (e.g., $f_X(x)$, σ_X). Furthermore, $\mathrm{E}\{\cdot\}$ denotes the statistical mean. The $L \times L$ identity matrix will be denoted by $I_{L\times L}$, and A^* is the Hermitian transpose of matrix A.

2. PROBLEM FORMULATION

Taking into account the drawbacks of both training sequences and superimposed pilots exposed in the Introduction, in this work we propose a novel procedure for flat fading channel estimation. We require our method to have the following properties: **1)** the channel estimation should not be based on allocating some channel slots (e.g., in time or frequency) for pilot signals. In this way the transmission of the information can be transmitted with no interruption (e.g., in time or frequency), **2)** the original signal must keep its value according to some objective measure, meaning that the estimation procedure can only slightly modify the transmitted signal. Mean Square Error (MSE) will be used for quantifying this distortion, **3)** few samples should be required in order to provide a channel accurate estimate.

Hereafter, we will denote by $\mathbf{x}$ the information signal, which will be modeled by a continuous random variable. The transmitted signal will be denoted by $\mathbf{y}$, and $\mathbf{w} \triangleq \mathbf{y} - \mathbf{x}$ will denote the estimation aiding signal. Then, the transmitted signal goes through a complex flat fading channel with complex gain t_0, which also introduces circularly-symmetric complex Additive White Gaussian Noise (AWGN), yielding the received signal $\mathbf{z} = t_0\mathbf{y} + \mathbf{n}$.

DC-QIM computes the transmitted signal as

$$\mathbf{y} = Q_v(\mathbf{x}) + (1-\alpha)[\mathbf{x} - Q_v(\mathbf{x})], \tag{1}$$

where $Q_v(\cdot)$ is a quantizer indexed by the hidden message v, and $\alpha \in (0, 1]$ is the so-called distortion compensation parameter. The second term in the right side of (1) is called self-noise, which makes the transmitted signal lie off a quantization centroid; in other words, the larger α, the closer $\mathbf{y}$ will be to a centroid of the quantizer, but the larger the distortion on the information signal introduced by the watermark. The most common implementation of DC-QIM is that where $Q_v(\mathbf{x}) = Q_\Delta(\mathbf{x} - \mathbf{d} - \mathbf{d}_v) + \mathbf{d} + \mathbf{d}_v$, where $Q_\Delta(\cdot)$ is the minimum-distance componentwise scalar quantizer with step-size Δ, $\mathbf{D} \sim U([-\Delta/2, \Delta/2]^L)$ is the so-called dither, which takes values depending on a secret key K, and $\mathbf{d}_v$ is a quantizer offset that depends on the hidden message.

Since DC-QIM is able to completely avoid the host-interference problem in the watermarking scenario, in this work we propose to use (1), setting a fixed value of v, for generating also the transmitted signal. Intuitively, in the same way that the quantization of the information signal improves the watermark decoding in watermarking (in terms, for example, of achievable rate) due to the reduction of the host interference, it makes sense to use it in the channel estimation problem. As we will see, the achieved results confirm this intuition.

Finally, let us mention that some works in the watermarking literature [12, 13] have studied the watermark decoding of DC-QIM when the received signal is a noisy scaled version of the transmitted one, i.e., $\mathbf{z} = t_0\mathbf{y} + \mathbf{n}$, where t_0 is the scaling factor. However, contrarily to the problem considered in this paper, the target of those works was not to provide an estimator of t_0, but only to decode v.

3. DESCRIPTION OF THE ALGORITHM

In this paper, we exploit the nature of the complex product (multiplicative on the magnitude, additive on the phase) by considering a codebook defined in polar coordinates. By doing so, the estimation of t_0 will be decoupled into two simpler real estimation problems: first, an estimator $|\hat{t}_0(\mathbf{z})|$ of the magnitude is obtained, and this is then used to estimate the phase $\measuredangle\hat{t}_0(\mathbf{z})$. This decoupling will introduce some loss in performance, but, on the other hand, it will allow to significantly reduce the computational cost of the estimation.

3.1. Generation of the transmitted signal

The magnitude of x_i is modified as $|y_i| = |x_i| + \alpha\left(\mathcal{Q}_\rho\left(|x_i| - \varrho_i\right) - (|x_i| - \varrho_i)\right)$, where $i = 1, \ldots, L$, $\mathcal{Q}_\rho(\cdot)$ denotes a uniform scalar quantizer with step-size ρ, and $\boldsymbol{\varrho}$ stands for a dither sequence which is uniformly distributed in $[-\rho/2, \rho/2]^L$. It is worth noting that since the real and imaginary components of X follow independent zero-mean Gaussian distributions with variance σ_X^2, then $|X|$ will be Rayleigh distributed, with scale parameter σ_X.

In order to control the distortion introduced by the estimation aiding signal, and at the same time provide a phase detection error probability similar to that of the magnitude detection, the quantization step applied to $\measuredangle x_i$, $i = 1, \ldots, L$, is chosen to yield an Euclidean distance between neighboring complex centroids sharing the same magnitude (i.e., those centroids only distinguished by their phase), nearly equal to ρ.[1] Specifically, the quantization step used for quantizing the phase coordinate of the ith sample is calculated as

$$\phi_i = \begin{cases} 2\pi\left(\left\lceil \pi\left[\cos^{-1}\left(\frac{\sqrt{\left(\mathcal{Q}_\rho(|y_i|-\varrho_i)^2-(\rho/2)^2\right)}}{\mathcal{Q}_\rho(|y_i|-\varrho_i)}\right)\right]^{-1}\right\rceil\right)^{-1} \\ \qquad \text{if } \mathcal{Q}_\rho\left(|y_i| - \varrho_i\right) \neq 0, \\ 2\pi, \qquad \text{otherwise} \end{cases}, \tag{2}$$

where the arccosine function $\cos^{-1}(\cdot)$ takes values in $[-\pi, \pi)$, $\lceil\cdot\rceil$ stands for the ceil function, and we have used the relationships between Cartesian and polar coordinates.

Consequently, the modified phase will be obtained as $\measuredangle y_i = \measuredangle x_i + \alpha\left[\mathcal{Q}_{\phi_i}\left(\measuredangle x_i - \phi_i\varphi_i\right) - \left(\measuredangle x_i - \phi_i\varphi_i\right)\right]$, where $\boldsymbol{\varphi}$ is uniformly distributed in $[-1/2, 1/2]^L$. Note that the magnitude quantization step does not depend on i, but the phase quantization step does; in fact, the larger the magnitude of the ith sample, the smaller the used phase quantization step, which makes sense in order to achieve the target of controlling the estimation aiding signal power.

3.2. Magnitude Estimation

Since, in general, *a priori* information on $|t_0|$ is not available, the Maximum Likelihood (ML) estimator will be used. In order to obtain a mathematically tractable expression of $f_{|Z|\,\big|\,|T|,K}\left(|z_i|\,\big|\,|t|, \varrho_i\right)$, the pdf of $|Z|$ given $|t|$ and ϱ_i, an approximation of that pdf is proposed based on three hypotheses (whose verification depends on

[1] In general that distance can not be exactly ρ, as the phase quantization step is required to be an integer divider of 2π, in order to verify the phase periodicity constraint.

the chosen values of of the quantization step-size ρ, and the distortion compensation parameter α): **1)** the variance of $|X|$ is much larger than the second moment of the quantization lattice (i.e., $\rho^2/12$). Therefore, the probability of the transmitted centroid given $|t_0|$ can be approximated by $\rho f_{|X|}(z/|t_0|)$, **2)** the variance of the scaled self-noise (i.e., $|t_0|^2(1-\alpha)^2\rho^2/12$) is much smaller than the variance of the channel noise σ_N^2. Therefore, the Gaussian channel noise dominates the total noise distribution, **3)** the square distance between scaled centroids (which we will quantify by using $|t_0|^2\rho^2/12$) is much larger than the variance of the total noise $(\sigma_N^2 + |t_0|^2(1-\alpha)^2\rho^2/12)$. Therefore, the noise distribution is negligible out of the quantization region of the transmitted centroid.

By jointly considering these assumptions, one can approximate

$$f_{|Z|\big||T|,K}\left(|z|\big||t|,\varrho\right) \approx \frac{|z|\rho e^{-\frac{|z|^2}{2\sigma_X^2|t|^2}}}{\sigma_X^2|t|} \frac{e^{-\frac{((|z|-\varrho|t|)\bmod \rho|t|)^2}{2\left(\sigma_N^2+\frac{(1-\alpha)^2\rho^2|t|^2}{12}\right)}}}{\sqrt{2\pi\left(\sigma_N^2+\frac{(1-\alpha)^2\rho^2|t|^2}{12}\right)}}, \tag{3}$$

where the modulo operation is defined as $A \bmod B \triangleq A - \mathcal{Q}_B(A)$. If the three hypotheses do not simultaneously hold, the validity of this pdf approximation and the accuracy of our estimator can no longer be guaranteed. Intuitively, the leftmost fraction of the previous expression approximates the probability of the centroid corresponding to $|z|$, while the rightmost fraction approximates the distribution of $|z|$ given that centroid. From (3), and given that the components of $\mathbf{z}$ are mutually independent, the ML estimation can be approximated as

$$\begin{aligned}|\hat{t}_0(\mathbf{z})| \approx \operatorname*{argmin}_{|t|\geq 0} &\left(\frac{\|\mathbf{z}\|^2}{\sigma_X^2|t|^2} + \frac{\|(|\mathbf{z}|-\boldsymbol{\varrho}|t|)\bmod\rho|t|\|^2}{\left(\sigma_N^2+\frac{(1-\alpha)^2\rho^2|t|^2}{12}\right)}\right.\\ &\left.+L\log\left(|t|^2\left(\sigma_N^2+\frac{(1-\alpha)^2\rho^2|t|^2}{12}\right)\right)\right).\end{aligned} \tag{4}$$

In order to limit the search-space of (4), a search-interval $[|t_-|, |t_+|]$ will be calculated by using the variance-based estimator of $|t_0|^2$, i.e., $|\hat{t}_0(\mathbf{z})|^2_{\text{var}} = \frac{\frac{\sum_{i=1}^L |z_i|^2}{L-1} - \left(\frac{\sum_{i=1}^L |z_i|}{L-1}\right)^2 - \sigma_N^2}{\sigma_{|X|}^2+\sigma_\omega^2}$,where $\sigma_{|X|}^2 = (4-\pi)\sigma_X^2/2$ and σ_ω^2 denotes the variance of the magnitude of the estimation aiding signal (i.e., $\sigma_\omega^2 \approx \alpha^2\rho^2/12$). It can be shown that $|\hat{t}_0(\mathbf{z})|^2_{\text{var}}$ is an unbiased estimator of $|t_0|^2$; consequently, if L is large enough to apply the Central Limit Theorem (CLT), the distribution of $|\hat{t}_0(\mathbf{z})|^2_{\text{var}}$ can be approximated by a Gaussian distribution with mean $|t_0|^2$ and variance $2(|t_0|^2(\sigma_{|X|}^2+\sigma_\omega^2)+\sigma_N^2)^2/[(L-1)(\sigma_{|X|}^2+\sigma_\omega^2)^2]$. Therefore, if $|\hat{t}_0(\mathbf{z})|^2_{\text{var}} \approx |t_0|^2$, then $|t_0|^2$ will lie with approximated probability $\operatorname{erf}(K_2/\sqrt{2})$ in the interval defined by $|t|^2_{\pm} = \max\left(\epsilon, |\hat{t}_0(\mathbf{z})|^2_{\text{var}} \pm K_2\sqrt{2\eta/(L-1)}\right)$,where $\epsilon > 0$ guarantees that $|t_+|^2$ and $|t_-|^2$ take positive values, and $\eta \triangleq \frac{(|\hat{t}_0(\mathbf{z})|^2(\sigma_{|X|}^2+\sigma_\omega^2)+\sigma_N^2)^2}{(\sigma_{|X|}^2+\sigma_\omega^2)^2}$. By applying the square root to those values, we obtain the interval we were looking for.

Notice that due to the modulo operation, (4) is a non-convex function, making it impossible to apply off-the-shelf optimization algorithms. Here we propose to sample the search interval $[|t_-|, |t_+|]$ finely enough to guarantee that two consecutive sampled points will be in the main lobe of the target function, which is indeed convex; the set of sampled points will be denoted by $\mathcal{T}$. The sampling criterion is based on setting the total noise variance to be a multiple of the square quantization step-size, iteratively assuming that the considered magnitude value $|t(l)| = |t_0|$, so

$$\begin{aligned}|t(l+1)| = \frac{|t(l)|}{\mathrm{E}\{|X|^2\}+\frac{\rho^2(1-K_1)}{12}} \times &\left(\alpha\frac{\rho^2}{12}+\mathrm{E}\{|X|^2\}\right.\\ &\left.+\frac{\rho}{\sqrt{12}}\sqrt{\frac{\rho^2}{12}\left((1-\alpha)^2+K_1(2\alpha-1)\right)+K_1\mathrm{E}\{|X|^2\}}\right),\end{aligned}$$

where $\mathrm{E}\{|X|^2\} = \sigma_{|X|}^2 + \sigma_X^2\pi/2$, $|t(1)| = |t_-|$, and the iterative sampling stops when $|t(l)| \geq |t_+|$. The parameter K_1 is introduced to control the separation between two consecutive elements of $\mathcal{T}$ and, thus, the cardinality of that set. The larger K_1, the smaller $|\mathcal{T}|$ (less computational cost), but the more likely it will be that $\mathcal{T}$ misses the main lobe of the target function, with a consequent performance reduction. The details of this sampling algorithm will be expanded in a future work.

The Matlab optimization toolbox function `fminbnd` (which implements a bounded optimization algorithm based on golden section search and parabolic interpolation) is run once for each interval defined by two consecutive elements of $\mathcal{T}$; in this way a set $\mathcal{T}^*$ (of cardinality $|\mathcal{T}|-1$) with the corresponding optimization solutions, is built. Finally, the approximated ML estimate is that point in $\mathcal{T}^*$ which minimizes the target function in (4).

3.3. Phase Estimation

Assuming that the $|\hat{t}_0(\mathbf{z})|$ obtained following the scheme described in the previous section is an accurate approximation of $|t_0|$, the normalized observation $|z_i|/(|\hat{t}_0(\mathbf{z})|)$, which is approximately equal to $|y_i|$, is used to estimate the phase quantizer step-size $\hat{\phi}_i$ as in (2).

Under the hypotheses introduced in the previous section, the distribution of Z given t_0 and the transmitted centroid, can be approximated by an i.i.d. Gaussian distribution centered at the transmitted centroid multiplied by t_0, and with variance equal to the sum of the noise channel variance and the self-noise variance scaled by $|t_0|^2$. Analogously to the magnitude estimation, the pdfs of neighboring phase centroids are approximately not overlapped. Therefore, the resulting ML estimator of $\measuredangle t_0$ can be approximated as

$$\begin{aligned}\measuredangle\hat{t}_0(\mathbf{z}) = \operatorname*{argmin}_{t\in[-\pi,\pi)} \sum_{i=1}^{L} &\left|\mathcal{Q}_\rho\left(\frac{|z_i|}{|\hat{t}_0(\mathbf{z})|}-\varrho_i\right)\right.\\ &\left.-\left(\frac{|z_i|}{|\hat{t}_0(\mathbf{z})|}-\varrho_i\right)e^{j\left((\measuredangle z_i-\hat{\phi}_i\varphi_i-t)\bmod\hat{\phi}_i\right)}\right|^2,\end{aligned}$$

where the modulo operation is used to measure the phase difference between the received samples and their closest centroids. In this work, the previous optimization is carried out by exhaustive-search.

4. EXPERIMENTAL RESULTS

In this section, we compare the MSE of the DPC-based estimator proposed in this work, with that of PDD [4]. Figs. 1-3 show the MSE as a function of $|t_0| \in [0.1, 2] \cap 0.1\mathbb{Z}$, where the results for each of those points were obtained by using 10^3 Monte Carlo runs; for each run, $\measuredangle t_0$ was independently generated according to $U(-\pi,\pi)$. The DPC-based scheme uses $K_1 = 1$, $K_2 = 10$, $\epsilon = 10^{-3}$, and the exhaustive search performed in the estimate of $\measuredangle t_0$ considers $2\cdot 10^4$ points uniformly located through $[-\pi,\pi)$.

Concerning the comparison with PDD, we consider the case where such scheme also deals with time invariant flat channels, even

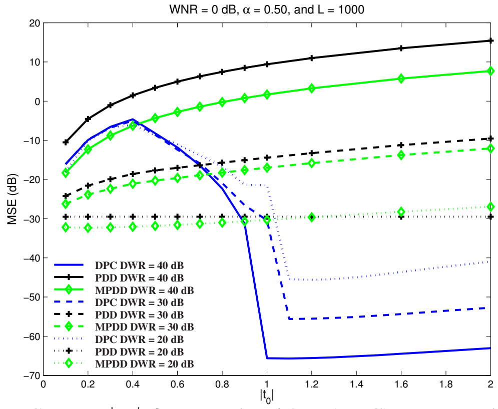

Fig. 1. MSE vs. $|t_0|$ for our algorithm (DPC), PDD and MPDD. DWR = 20, 30, 40 dB, WNR $= 0$ dB, $\alpha = 0.5$, and $L = 10^3$.

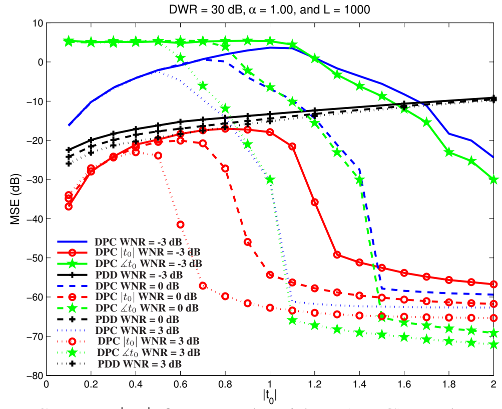

Fig. 2. MSE vs. $|t_0|$ for our algorithm (DPC), and PDD. WNR $= -3, 0, 3$ dB, DWR = 30 dB, $\alpha = 1$, and $L = 10^3$. MSEs of $|t_0|$ and $\measuredangle t_0$ are also provided.

if it can be used in more general frameworks; furthermore, the host-interference controlling parameter proposed in [4] is optimized in order to provide the best performance for that scheme. Note that the power of the distortion introduced on the host signal by PDD comprises both the power of the estimation aiding signal, and the power due to the reduction of the host interference. The channel estimator proposed in [4], once it is adapted to the complex flat fading case, is $\hat{t}_0(\mathbf{z}) = \mathbf{w}^*\mathbf{z}/\|\mathbf{w}\|^2$, i.e., it only uses the component of $\mathbf{z}$ in the direction of $\mathbf{w}$; consequently, the remaining $L-1$ components of $\mathbf{z}$ are dismissed. Since those $L-1$ components follow a $\mathcal{N}(0, |t_0|^2\sigma_X^2 + \sigma_N^2)$ distribution, they are indeed informative about $|t_0|$, and that dependence could be exploited. Therefore, we propose a suboptimal modification of PDD (denoted by MPDD), where the $L-1$ components of $\mathbf{z}$ orthogonal to $\mathbf{w}$ are fed to a variance-based estimator, and the estimate of $\measuredangle t_0$ is $\measuredangle \mathbf{w}^*\mathbf{z}$.

For the sake of comparison, we will find it useful to define two power ratios borrowed from watermarking: the Document to Watermark Ratio (DWR), which is the ratio between the power of $\mathbf{x}$ and the power of $\mathbf{w}$, and the Watermark to Noise Ratio (WNR), which is the ratio between the power of $\mathbf{w}$ and the power of $\mathbf{n}$.

Fig. 1 compares the MSE of the proposed scheme with that of PDD and MPDD as a function of $|t_0|$, for different values of DWR. One can observe that the larger the DWR (i.e., the larger the margin by which Hypothesis 1 in Sect. 3.2 is satisfied), the better the performance of the proposed scheme. Furthermore, in the proposed scheme a larger DWR helps to estimate t_0 (at the cost of increasing the estimation computational cost), contrarily to what happens with PDD and MPDD; indeed, in order for the DPC-based scheme to provide better results than PDD and MPDD, $|t_0|$ must take values larger than a DWR-dependent threshold; the larger the DWR, the smaller the $|t_0|$ value for the crossing point. Is is worth mentioning that our method generally requires more computational resources than PDD or MPDD. Related to the comparison between PDD and MPDD, the larger the DWR, the better MPDD is with respect to PDD; in that case PDD will not be able to cancel out the host interference on $\mathbf{w}$, and, as it was mentioned before, the estimator proposed in [4] does not take advantage either of the $L-1$ components of $\mathbf{z}$ orthogonal to $\mathbf{w}$ (as our proposed modification MPDD does).

Fig. 2 illustrates the contribution of $|t_0|$ and $\measuredangle t_0$ to the MSE of the estimate of t_0 for different values of WNR; again, the results for PDD are also plotted. Similarly to the discussion about Fig. 1, in this case we can check the effect of the margin by which Hypothesis 3 in Sect. 3.2 is satisfied on the performance of the estimator. Mainly, the larger the WNR, the better the provided approximation; of course, one must also take into account that a larger WNR will make easier the estimate, independently of the accuracy in the approximation of the pdf. Additionally, it must be noted that the DPC MSE curves share a similar behavior with respect to $|t_0|$: for small values of $|t_0|$, the value of the MSE increases with it; then, when the three hypotheses hold, it decreases with $|t_0|$. Furthermore, we can see that the main source of MSE seems to be the phase estimate; this is partially due to the fact that this estimator inherits the errors made by the magnitude estimator.

Finally, Fig. 3 illustrates the behavior of the MSE as a function of the distortion compensation parameter α. According to these results, the performance of our scheme shows a trade-off between the value of $|t_0|$ at the crossing point with PDD, and the value of MSE when $|t_0|$ is increased. For example, for $\alpha = 0.5$ our technique outperforms PDD for $|t_0| \geq 0.7$, and MSE ≈ -55 dB for large values of $|t_0|$; on the other hand, for $\alpha = 0.75$ the crossing point is at $|t_0| \approx 0.9$, but MSE ≈ -59 dB for large values of $|t_0|$.

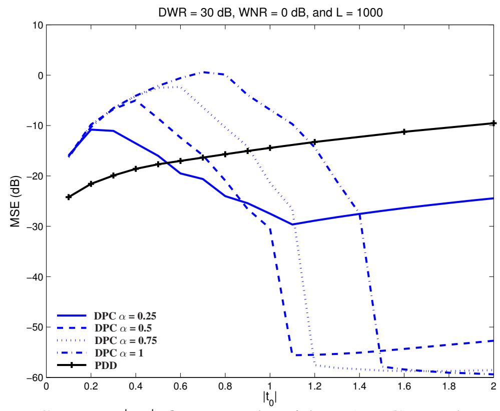

Fig. 3. MSE vs. $|t_0|$ for our algorithm (DPC), and PDD. $\alpha = 0.25, 0.5, 0.75, 1$, DWR = 30 dB, WNR $= 0$ dB, and $L = 10^3$.

5. CONCLUSIONS

In this paper, a novel flat fading channel estimator based on DPC and ML has been presented. According to the experimental results, the DPC-based estimation performance, measured by the MSE, improves with DWR and WNR. Additionally, the impact of the distortion compensation parameter α on the performance of our scheme is also analyzed. The shown results also indicate that our proposal outperforms the PDD techniques when the design hypotheses hold.

6. REFERENCES

[1] Amod V. Dandawate and Georgios B. Giannakis, “Asymptotic theory of mixed time averages and kth-order cyclic-moment and cumulant series,” *IEEE Transactions on Information Theory*, vol. 41, no. 1, pp. 216–232, January 1995.

[2] John R. Treichler and Brian G. Agee, “A new approach to multipath corretion of constant modulus signals,” *IEEE Transactions on Acoustisc, Speech, and Signal Processing*, vol. 31, no. 2, pp. 459–472, April 1983.

[3] Jitendra K. Tugnait and Xiaohong 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] Shuangchi He and Jitendra K. Tugnait, “On doubly selective channel estimation using superimposed training and discrete prolate spheroidal sequences,” *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3214–3228, July 2008.

[5] Peter Hoeher and Fredrik Tufvesson, “Channel estimation with superimposed pilot sequence,” in *Proc. of the IEEE Global Telecommunications Conference*, 1999, pp. 2162–2166.

[6] Jitendra K. Tugnait, Lang Tong, and Zhi Ding, “Single-user channel estimation and equalization,” *IEEE Signal Processing Magazine*, vol. 17, no. 3, pp. 17–28, May 2000.

[7] Jitendra K. Tugnait and Weilin Luo, “On channel estimation using superimposed training and first-order statistics,” *IEEE Communications Letters*, vol. 7, no. 9, pp. 413–415, September 2003.

[8] Ingemar J. Cox, Matthew L. Miller, and Jeffrey A. Bloom, *Digital Watermarking*, Morgan Kaufmann Publishers, 2002.

[9] Henrique S. Malvar and Dinei A. F. Florencio, “Improved spread spectrum: A new modulation technique for robust watermarking,” *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 898–905, April 2003.

[10] Max H. M. Costa, “Writing on dirty paper,” *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[11] Brian Chen and Gregory W. Wornell, “Quantization index modulation: A class of provably good methods for digital watermarking and information embedding,” *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[12] Félix Balado, Kevin M. Whelan, Guénolé C. M. Silvestre, and Neil J. Hurley, “Joint iterative decoding and estimation for side-informed data hiding,” *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 4006–4019, October 2005.

[13] Ivo D. Shterev and Reginald L. Lagendijk, “Amplitude scale estimation for quantization-based watermarking,” *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4146–4155, November 2006.