

Time-Multiplexed Training for Time-Selective Channels

Zijian Tang and Geert Leus

Abstract—Pilot-assisted channel estimation is considered in this letter, where the channel is assumed to be time-selective and can be accurately fit by a basis expansion model. The position and power of the pilots are crucial to the mean square error of the channel estimator. In this paper, we present nonlinear integer programming algorithms to optimize the position and power of the pilots. In comparison with the traditional equi-distant/powered pilot structure, the solution obtained from the proposed algorithms yields a better performance.

Index Terms—Basis expansion model (BEM), pilot-assisted channel estimation, time-selective channels.

I. INTRODUCTION

RECENTLY, parsimonious channel models to track the channel's time variation have drawn increasing attention. One of the approaches, known as the basis expansion model (BEM) will be treated in this letter. Examples of such BEMs are the complex exponential BEM (CE-BEM) [1]–[4], the polynomial BEM (P-BEM) [5], the discrete prolate spheroidal BEM (DPS-BEM) [6], etc. Note that the CE-BEM in [2], [4] is slightly different than that in [1], [3]: the former is sometimes referred to as the critically-sampled CE-BEM [(C)CE-BEM] because the BEM period equals the window length, whereas the latter uses a longer period and is thus also referred to as the oversampled CE-BEM [(O)CE-BEM].

Based on a certain BEM assumption, time-selective channel estimation reduces to estimating the BEM coefficients. We focus on pilot-assisted channel estimators corresponding to different BEM assumptions using time-multiplexed (TM) training, i.e., pilots are inserted in the time-domain as considered in [2], [4]–[6]. Usually, the total number of pilots (bandwidth constraint) as well as the total power (power constraint) is limited. The question arises as to what will be the optimal pilot structure (positions and powers). Note that the optimal pilot structure will depend on both the optimality criterion and the considered BEM. For instance, in terms of the maximum BEM modeling error and under the P-BEM assumption, the optimal pilot positions correspond to the roots of the Chebyshev polynomials [7]. In this letter, we will not consider the influence of the BEM modeling error, which can be kept very small in most practical situations [6], and take the channel mean square error

(MSE) as the optimality criterion. It is shown in [2], [4] that for the (C)CE-BEM, the optimal pilots must be uniformly distributed with equal power.¹ However, it is not yet clear what the optimal pilot structure is for other BEMs, e.g., the DPS-BEM, the P-BEM or the (O)CE-BEM, which are of more practical significance in view of their superior modeling capacity [6].

In this letter, we will optimize the TM training scheme under a general BEM assumption. An MSE-related cost function will be formulated for the LS channel estimator, and translated into a mixed-integer non-linear programming problem [8]. We show that the resulting optimal pilot structure can produce a lower channel MSE than the classical equi-distant/powered pilots if a general BEM assumption is adopted.

Notation: We use upper (lower) bold face letters to denote matrices (column vectors). $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ represent conjugate, transpose and complex conjugate transpose (Hermitian), respectively. \mathbf{I}_m stands for an $m \times m$ identity matrix. $\mathcal{D}\{\mathbf{x}\}$ stands for a diagonal matrix with \mathbf{x} as its diagonal, and $\mathcal{D}\{\mathbf{X}_1, \mathbf{X}_2, \dots\}$ for a block-wise diagonal matrix with $\mathbf{X}_1, \mathbf{X}_2, \dots$ as its diagonal blocks. $\text{tr}\{\mathbf{X}\}$ represents the trace of \mathbf{X} . Further, we use $[\mathbf{X}]_{m,n}$ to indicate the $(m+1, n+1)$ st entry of the matrix \mathbf{X} , and $[\mathbf{X}]_{m,:}$ to indicate the $(m+1)$ st row of \mathbf{X} .

II. DATA MODEL FOR TIME-SELECTIVE CHANNEL ESTIMATION

Let us consider a communication system over a time-selective channel. The I/O relationship in discrete form can be expressed as $y_n = h_n^{(t)} s_n + w_n$, where y_n , s_n , and w_n denote respectively the received signal, the transmitted signal and the noise at the n th time instant; $h_n^{(t)}$ stands for the “true” channel gain, which varies with time $h_n^{(t)} \neq h_m^{(t)}$ for $m \neq n$. To approximate this time-variation with a BEM, let us choose the time instants $n = 0, \dots, N-1$ as an observation window and collect all the channel gains within this observation window in the vector $\mathbf{h}^{(t)} := [h_0^{(t)}, \dots, h_{N-1}^{(t)}]^T$. By selecting a proper scale Q for the BEM, we can closely fit the channel's time variation within the observation window with $Q+1$ BEM coefficients

$$\mathbf{h}^{(t)} = \mathbf{B}\mathbf{h} + \boldsymbol{\delta} \quad (1)$$

where \mathbf{B} is an $N \times (Q+1)$ BEM matrix, and \mathbf{h} a $(Q+1)$ -long vector containing the related BEM coefficients. $\boldsymbol{\delta}$ represents the BEM modeling error. With its entries independent of the channel, a BEM reduces the system scale from N to Q ($N \gg Q$) at the price of a modeling error, which is nonetheless

¹The optimal pilots proposed in [2], [4] are based on a linear minimum mean square error estimator under certain statistical assumptions, but the optimality remains also valid for the least squares (LS) estimator. We will deal with an LS estimator since it is more robust to a possible mismatch in the channel statistics.

Manuscript received October 19, 2006; revised January 17, 2007. This work was supported in part by NWO-STW under the VICI program (DTC.5893) and the VIDi program (DTC.6577). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Gerald Matz.

The authors are with Delft University of Technology-Fac. EEMCS, 2628 CD Delft, The Netherlands (e-mail: Z.Tang@TUDelft.NL; G.Leus@TUDelft.NL).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/LSP.2007.896146

ignored in this paper for the design of the channel estimator. This is because the modeling error can be kept very small by choosing an appropriate BEM [6]. As a result, its impact at a practical signal-to-noise ratio (SNR) can be almost neglected. Furthermore, it is usually difficult to gauge the modeling error when the channel itself is unknown. We also note that the idea of a BEM is reflected in some other applications without inducing a modeling error, e.g., the frequency-domain channel in an OFDM system [9] can be model by a (C)CE-BEM whereas the frequency-domain channel in an OFDM system with virtual subcarriers [10] can be modeled by an (O)CE-BEM. For these cases, the approach discussed in this letter can be also applied.

To estimate the BEM coefficients \mathbf{h} , we resort to a TM training scheme, in which the pilots and the data will be interleaved in the time domain. Suppose there are K pilots transmitted during the observation window, whose positions are collected in the set $\mathcal{G} := \{g_0, g_1, \dots, g_{K-1}\}$. Furthermore, we assume that the k th pilot s_{g_k} has phase ϕ_k and power P_k , i.e., $s_{g_k} = \sqrt{P_k}e^{j\phi_k}$ with the total power for training $P = \sum_k P_k$. With the aid of these notations and neglecting the modeling error, we can express the I/O relationship that results from the pilots in matrix/vector form as

$$\mathbf{y}_p = \Phi_p \sqrt{\mathcal{P}_p} \mathbf{B}_p \mathbf{h} + \mathbf{w}_p \quad (2)$$

where \mathbf{y}_p is a $K \times 1$ vector collecting the received samples at the pilot-related positions, i.e., $\mathbf{y}_p := [y_{g_0}, \dots, y_{g_{K-1}}]^T$; \mathbf{w}_p is similarly defined as \mathbf{y}_p ; \mathcal{P}_p and Φ_p are both $K \times K$ diagonal matrices with $[P_0, \dots, P_{K-1}]$ and $[e^{j\phi_0}, \dots, e^{j\phi_{K-1}}]$ as their diagonals, respectively; \mathbf{B}_p is a $K \times (Q+1)$ matrix consisting of K rows carved out of \mathbf{B} corresponding to the pilot positions. An LS estimate from (2) can be obtained as $\hat{\mathbf{h}} = (\mathbf{B}_p^H \mathcal{P}_p \mathbf{B}_p)^{-1} \mathbf{B}_p^H \sqrt{\mathcal{P}_p} \Phi_p^H \mathbf{y}_p$ resulting in a channel MSE equal to $\mathcal{E}\{\|(\mathbf{B}_p^H \mathcal{P}_p \mathbf{B}_p)^{-1} \mathbf{B}_p^H \sqrt{\mathcal{P}_p} \Phi_p^H \mathbf{w}_p\|^2\}$. Assuming the noise w_p to be white with variance σ^2 , we can express the MSE as

$$\text{MSE} = \sigma^2 \text{tr} \left\{ (\mathbf{B}_p^H \mathcal{P}_p \mathbf{B}_p)^{-1} \right\}. \quad (3)$$

From the above, we remark that the phase of the pilots has no impact on the channel MSE, and we only need to focus on the powers and positions of the pilots. The optimization problem can thus be formulated as

$$\begin{aligned} \arg \min_{\{\mathcal{G}, \mathcal{P}_p\}} \text{MSE}(\mathcal{G}, \mathcal{P}_p) &= \sigma^2 \text{tr} \left\{ (\mathbf{B}_p^H \mathcal{P}_p \mathbf{B}_p)^{-1} \right\} \\ \text{s.t. } \mathbf{B}_p &= [\mathbf{B}]_{\mathcal{G},:}, \mathcal{G} \subset \{0, \dots, N-1\} \text{ and } |\mathcal{G}| = K, \\ \mathcal{P}_p &\text{ positive diagonal with } \text{tr}\{\mathcal{P}_p\} = P. \end{aligned} \quad (4)$$

III. OPTIMIZATION ALGORITHM

The above formulation is a mixed-integer non-linear optimization problem, where we have posed restrictions on the total number and power of the pilots. [8] provides an algorithm known as the generalized Benders decomposition (GBD). In a nutshell, the GBD iteratively projects the minimization problem onto the \mathcal{P}_p —(primal problem) and the \mathcal{G} —(relaxed master problem). Since the constraints in (4) and (5) are separable in \mathcal{G}

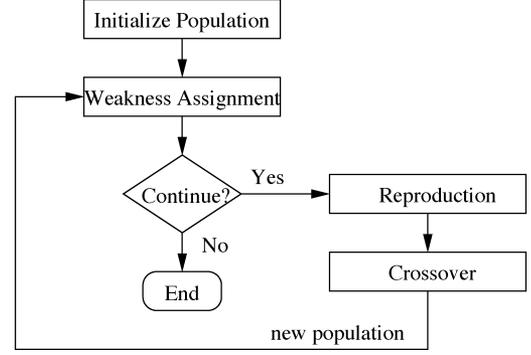


Fig. 1. Flowchart of GA.

and \mathcal{P}_p , the primal problem becomes simply the search for the optimal \mathcal{P}_p subject to a fixed \mathcal{G} , and the relaxed master problem becomes the search for the optimal \mathcal{G} subject to a fixed \mathcal{P}_p . We summarize the GBD algorithm as follows.

- 1) Let the superscript (k) denote the iteration index. To initialize, set $k = 0$, and assume $\mathcal{G}^{(-1)} = \{0, \dots, K/2 - 1, N - K/2, \dots, N - 1\}$ and $\mathcal{P}_p^{(0)} = (P/K)\mathbf{I}_K$.
- 2) For $\mathcal{P}_p^{(k)}$, solve $v_{\text{master}}^{(k)} = \min_{\{\mathcal{G}\}} \text{MSE}(\mathcal{G}, \mathcal{P}_p^{(k)})$ subject to (4) resulting in $\mathcal{G}^{(k)}$.
- 3) For $\mathcal{G}^{(k)}$, solve $v_{\text{primal}}^{(k)} = \min_{\{\mathcal{P}_p\}} \text{MSE}(\mathcal{G}^{(k)}, \mathcal{P}_p)$ subject to (5) resulting in $\mathcal{P}_p^{(k+1)}$.
- 4) For a predetermined ϵ , if $|v_{\text{primal}}^{(k)} - v_{\text{master}}^{(k)}| \leq \epsilon$ then terminate. Otherwise, set $k = k + 1$ and return to Step 2.

The primal problem at Step 3 is a nonlinear programming problem (NLP), which can be solved analytically if $K = Q + 1$ [10]. In other cases where $K > Q + 1$, we can resort to the MATLAB built-in function *fmincon*(\cdot). The solution corresponds to the global optimum due to the following lemma.

Lemma 1: The function $\text{MSE}(\mathcal{G}, \mathcal{P}_p) = \sigma^2 \text{tr}\{(\mathbf{B}_p^H \mathcal{P}_p \mathbf{B}_p)^{-1}\}$ is strictly convex on the positive diagonal matrix \mathcal{P}_p for a given \mathcal{G} .

To prove the above lemma, we realize that the function $\text{tr}\{\mathbf{A}^{-1}\}$ is strictly convex on a positive-definite Hermitian matrix \mathbf{A} [11]. In our case, $\mathbf{A} = \mathbf{B}_p^H \mathcal{P}_p \mathbf{B}_p$ is linear in \mathcal{P}_p for a given \mathcal{G} , which preserves the convexity.

However, the master problem at Step 2, which is essentially a binary programming problem, is more problematic. Especially due to the fact that the MSE is not a convex function on the pilot positions despite Lemma 1. In [12] and [13], this is solved by relaxing the binary problem to a non-binary problem such that the convexity can be still called upon. Unfortunately, this approach, when applied to our problem, does not facilitate a fast convergence due to a very large pilot position space. In this letter, we resort to a combined genetic algorithm (GA) [14], whose working principle is illustrated by the flowchart in Fig. 1, with:

- “Population” representing a set of candidate (pilot position) solutions.
- “Weakness” representing the cost function value defined in (3) for each candidate. Those candidates that do not satisfy

condition (4) will be penalized by a weakness value equal to infinity.

- “Reproduction” representing the operation that copies the population except for the candidates that either have the largest weakness or the smallest weakness. The former will be discarded while the latter will be copied twice since it has the lowest MSE at the moment.
- “Crossover” representing the operation applied on the candidate pool produced by the reproduction: at the l th iteration, all the candidates are randomly grouped into pairs (parents). Let us take one such pair for example: suppose “Parent1” is the m th candidate $[g_0^{(l,m)}, \dots, g_{K-1}^{(l,m)}]$ and “Parent2” the n th candidate $[g_0^{(l,n)}, \dots, g_{K-1}^{(l,n)}]$ with $m \neq n$. Here, $g_k^{(l,i)}$ stands in our context for the position of the k th pilot that corresponds to the i th candidate obtained at the l th iteration. “Mating” these two parents results in two new candidates (children) with “Child1” denoted as $[g_0^{(l+1,m)}, \dots, g_{K-1}^{(l+1,m)}]$ and “Child2” as $[g_0^{(l+1,n)}, \dots, g_{K-1}^{(l+1,n)}]$. In generating each entry of the children, either of two operations will take place: replication or swapping with a probability of 0.5, respectively. For the k th entry for instance, in case of replication $g_k^{(l+1,m)} = g_k^{(l,m)}$, and $g_k^{(l+1,n)} = g_k^{(l,n)}$. In case of swapping, we first express the k th entry of “Parent1” and “Parent2” in binary-form. These two binary strings are then divided at the same arbitrary place, from which the right-hand part of “Parent2” will be concatenated to the left-hand part of “Parent1” and vice versa. In this way, two new binary strings are created and converted back in decimal form.²

If the observation window size N or pilot number K is large, the GA must abide with a large population size. This results in a higher complexity and lower convergence rate, which in practice, often leads to a “near-optimal” solution. It is thus helpful if we could equip the GA with some *a priori* knowledge about the solution. For the considered case (both N and K are assumed to be even), we could constrain the pilot structure to be symmetric with respect to the center of the observation window, e.g., we let $g_i = N - g_{K-i-1} - 1$ and $P_i = P_{K-i-1}$. This constraint is introduced due to the following properties.

Property 1: Without loss of generality, we can design the BEM in a particular way. For the CE-BEM, we take $[\mathbf{B}]_{n,q} = e^{j(2\pi/\kappa N)(n - ((N-1)/2))(q - (Q/2))}$, where $\kappa = 1$ for the (C)CE-BEM as in [2], [4], and $\kappa > 1$ for the (O)CE-BEM as in [1], [3]. For the P-BEM, we take $[\mathbf{B}]_{n,q} = (n - ((N-1)/2))^q$. Finally, for the DPS-BEM, we take as the columns of \mathbf{B} the $Q + 1$ most significant eigenvectors of a kernel matrix $[\mathbf{C}]_{n,q} = 2\nu_D \text{sinc}(2\pi(n - q)\nu_D)$, where ν_D stands for the normalized Doppler spread.

The BEMs taking the above expressions admit some symmetric structure: if we use \mathbf{B}_u and \mathbf{B}_d to denote the first and second half of \mathbf{B} , respectively, i.e., $\mathbf{B} = [\mathbf{B}_u^T, \mathbf{B}_d^T]^T$, they are related to each other as $\mathbf{B}_d = \mathbf{A}\mathbf{B}_u^*\mathbf{\Xi}$, with \mathbf{A} being an $(N/2) \times (N/2)$ permutation matrix, which has only zero elements ex-

²For instance, suppose the k th entry of “Parent1” is $g_k^{(l,m)} = 2$ and the k th entry of “Parent2” is $g_k^{(l,n)} = 5$. If swapping takes place, we first find the binary expression for $g_k^{(l,m)} = [0, 1, 0]$ and $g_k^{(l,n)} = [1, 0, 1]$. Suppose we cut randomly at the first bit, then $g_k^{(l+1,m)} = [0, 0, 1]$, and $g_k^{(l+1,n)} = [1, 1, 0]$, which in decimal-form equal 1 and 6, respectively.

cept for $[\mathbf{A}]_{k,N/2-k-1} = 1$; $\mathbf{\Xi} := \mathbf{I}_{Q+1}$ for the CE-BEM, and $\mathbf{\Xi} := \mathcal{D}\{[(-1)^0, \dots, (-1)^Q]^T\}$ for the P-BEM and DPS-BEM.

Property 2: If $(\mathcal{G}_{\text{opt}}, \mathcal{P}_{p,\text{opt}})$ denotes the optimal pilot structure with $\mathcal{P}_{p,\text{opt}} = \mathcal{D}\{[P_{0,\text{opt}}, \dots, P_{K-1,\text{opt}}]\}^T$, and $\mathcal{G}_{\text{opt}} = \{g_{0,\text{opt}}, \dots, g_{K-1,\text{opt}}\}$, then we can always find another pilot structure $(\tilde{\mathcal{G}}_{\text{opt}}, \tilde{\mathcal{P}}_{p,\text{opt}})$ with $\tilde{\mathcal{P}}_{p,\text{opt}} = \mathcal{D}\{[P_{K-1,\text{opt}}, \dots, P_{0,\text{opt}}]\}^T$ and $\tilde{\mathcal{G}}_{\text{opt}} = \{N-1-g_{0,\text{opt}}, \dots, N-1-g_{K-1,\text{opt}}\}$, such that $\text{MSE}(\tilde{\mathcal{G}}_{\text{opt}}, \tilde{\mathcal{P}}_{p,\text{opt}}) = \text{MSE}(\mathcal{G}_{\text{opt}}, \mathcal{P}_{p,\text{opt}})$.

Proof: Let us absorb the effect of \mathcal{P}_p and \mathcal{G} in a larger $N \times N$ diagonal matrix \mathcal{P} , whose diagonal consists of zeros except for the pilot positions, i.e., $[\mathcal{P}]_{i,i} = 1$ if $i \in \mathcal{G}$. By this means, we can rewrite (3) as $\text{MSE}(\mathcal{P}) = \sigma^2 \text{tr}\{(\mathbf{B}^H \mathcal{P} \mathbf{B})^{-1}\}$.

Suppose \mathcal{P}_{opt} corresponds to the optimal pilots $(\mathcal{G}_{\text{opt}}, \mathcal{P}_{p,\text{opt}})$. We divide it into the left-upper and right-bottom half, which are represented by $\mathcal{P}_{u,\text{opt}}$ and $\mathcal{P}_{d,\text{opt}}$, respectively, i.e., $\mathcal{P}_{\text{opt}} = \mathcal{D}\{\mathcal{P}_{u,\text{opt}}, \mathcal{P}_{d,\text{opt}}\}$. Obviously, the counterpart $\tilde{\mathcal{P}}_{\text{opt}}$ of the pilot structure $(\tilde{\mathcal{G}}_{\text{opt}}, \tilde{\mathcal{P}}_{p,\text{opt}})$ can be related to \mathcal{P}_{opt} as $\tilde{\mathcal{P}}_{\text{opt}} = \mathcal{D}\{\mathbf{A}^H \mathcal{P}_{d,\text{opt}} \mathbf{A}, \mathbf{A}^H \mathcal{P}_{u,\text{opt}} \mathbf{A}\}$. By Property 1, we have

$$\begin{aligned} \text{MSE}(\mathcal{P}_{\text{opt}}) &= \sigma^2 \text{tr} \left\{ \left(\mathbf{B}_u^H \mathcal{P}_{u,\text{opt}} \mathbf{B}_u + \mathbf{\Xi}^H \mathbf{B}_u^T \mathbf{A}^H \mathcal{P}_{d,\text{opt}} \mathbf{A} \mathbf{B}_u^* \mathbf{\Xi} \right)^{-1} \right\}. \end{aligned}$$

Likewise, for $\tilde{\mathcal{P}}_{\text{opt}}$, we obtain

$$\begin{aligned} \text{MSE}(\tilde{\mathcal{P}}_{\text{opt}}) &= \sigma^2 \text{tr} \left\{ \left(\mathbf{B}_u^H \mathbf{A}^H \mathcal{P}_{d,\text{opt}} \mathbf{A} \mathbf{B}_u \right. \right. \\ &\quad \left. \left. + \mathbf{\Xi}^H \mathbf{B}_u^T \mathcal{P}_{u,\text{opt}} \mathbf{B}_u^* \mathbf{\Xi} \right)^{-1} \right\} \\ &\stackrel{a}{=} \sigma^2 \text{tr} \left\{ \left(\mathbf{\Xi}^H \mathbf{B}_u^H \mathbf{A}^H \mathcal{P}_{d,\text{opt}} \mathbf{A} \mathbf{B}_u \mathbf{\Xi} \right. \right. \\ &\quad \left. \left. + \mathbf{B}_u^T \mathcal{P}_{u,\text{opt}} \mathbf{B}_u^* \right)^{-1} \right\} \\ &\stackrel{b}{=} \sigma^2 \text{tr} \left\{ \left(\mathbf{\Xi}^H \mathbf{B}_u^T \mathbf{A}^H \mathcal{P}_{d,\text{opt}} \mathbf{A} \mathbf{B}_u^* \mathbf{\Xi} \right. \right. \\ &\quad \left. \left. + \mathbf{B}_u^H \mathcal{P}_{u,\text{opt}} \mathbf{B}_u \right)^{-1} \right\} \\ &= \text{MSE}(\mathcal{P}_{\text{opt}}) \end{aligned}$$

where $\stackrel{a}{=}$ holds because $\mathbf{\Xi} \mathbf{\Xi} = \mathbf{\Xi}^H \mathbf{\Xi}^H = \mathbf{I}_{Q+1}$, and $\stackrel{b}{=}$ holds because $\text{tr}\{\mathbf{A}^{-1}\} = \text{tr}\{\mathbf{A}^{*-1}\}$ for a Hermitian matrix \mathbf{A} . $\square\square\square$

Property 2 suggests that the optimal pilot structure exists in pairs. Should there be a unique global minimum, this would imply that the optimal pilot structure ought to be symmetric. Since we have observed in applications that multiple global minima could exist, imposing a symmetric constraint upon the GA would lead to an MSE degradation in theory. However, as shown in the simulation part, the symmetric algorithm inflicts a much lower complexity and renders a performance very close to (or even better than) the non-symmetric algorithm. This is probably due to the fact that the channel MSE is not a convex function of the pilot positions and has a large number of local minima. By enforcing symmetry, the search space is reduced and hence the global minimum might never be reached. On the other hand, a smaller search space is beneficial in avoiding local minima and thus increases the convergence rate.

TABLE I
PILOT STRUCTURE COMPARISON FOR $K = 6$

	symmetric NLP / symmetric GA	non-symmetric NLP / non-symmetric GA	equi-powered/ equi-distant	equi-powered/ symmetric GA	symmetric NLP / equi-distant
$g_0(p_0)$	1(0.8184)	1(1.4643)	1(1)	1(1)	1(0.6234)
$g_1(p_1)$	48(1.2214)	51(0.3162)	44(1)	44(1)	44(1.0457)
$g_2(p_2)$	126(0.9157)	53(1.8032)	87(1)	105(1)	87(1.0733)
$g_3(p_3)$	131(0.9157)	138(0.3162)	130(1)	152(1)	130(0.6099)
$g_4(p_4)$	209(1.2214)	210(0.31622)	173(1)	213(1)	173(1.2571)
$g_5(p_5)$	256(0.8184)	253(0.31622)	216(1)	256(1)	216(1.1890)

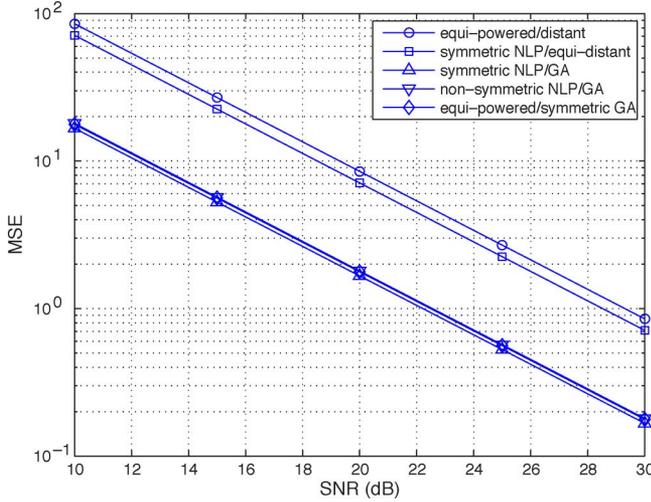


Fig. 2. MSE versus SNR.

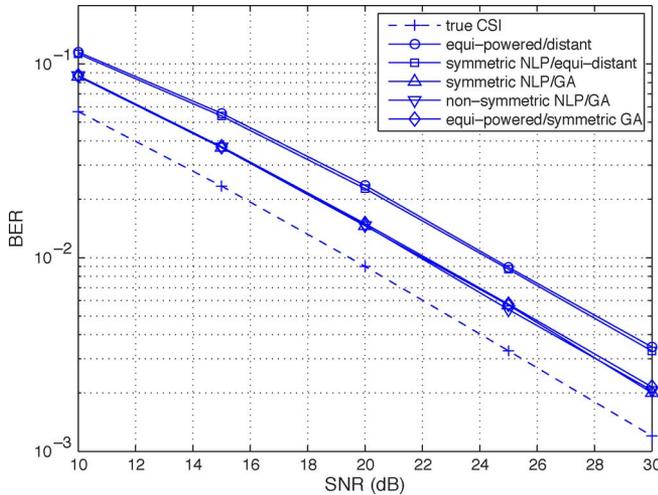


Fig. 3. BER versus SNR.

IV. NUMERICAL EXAMPLES

We generate time-selective channels as prescribed in [15] for a normalized Doppler spread $\nu_D = 0.002$. The (O)CE-BEM assumption will be adopted to approximate the channel's time-variation, though other BEMs are also applicable but will not be examined here due to space restrictions. Following the definition given in Property 1, we set $N = 256$, $Q = 4$ and $\kappa = 4$.

We compare different solutions for $K = 6$ pilots listed in Table I. The symmetric GA is equipped with a population pool size of 100 and iterates 100 times, and the non-symmetric GA is equipped with a population pool size of 250 and iterates 1000 times.

The corresponding MSE performances are plotted in Fig. 2, and the BER performances resulting from a maximum likelihood equalizer are plotted in Fig. 3. From the figures, one can observe that the performance of the symmetric algorithm is slightly better than that of the non-symmetric algorithm, even though the former requires a much smaller complexity. It is also noteworthy that only optimizing the pilot positions but assuming equal power, the resulting pilot structure can already improve the performance considerably.

V. CONCLUSION

This letter shows how to search for the optimal pilot structure numerically, which can be applied to time-selective channel estimation based on a general BEM assumption. Note that to implement the proposed optimization algorithms, it is up to the transmitter to choose a proper BEM.

REFERENCES

- [1] T. A. Thomas and F. W. Vook, "Multi-user frequency-domain channel identification, interference suppression, and equalization for time-varying broadband wireless communications," in *Proc. 2000 IEEE Sensor Array and Multichannel Signal Processing Workshop*, Mar. 2000, pp. 444–448.
- [2] X. Ma, G. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly-selective fading channels," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1351–1366, May 2003.
- [3] G. Leus, "On the estimation of rapidly time-varying channels," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Sep. 2004, pp. 2227–2230.
- [4] A. P. Kannu and P. Schniter, "MSE-optimal training for linear time-varying channels," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Mar. 2005.
- [5] D. K. Borah and B. D. Hart, "Frequency-selective fading channel estimation with a polynomial time-varying channel model," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 862–873, Jun. 1999.
- [6] T. Zemen and C. F. Mecklenbräuer, "Time-variant channel estimation using discrete prolate spheroidal sequences," *IEEE Trans. Signal Process.*, vol. 53, no. 9, pp. 3597–3607, Sep. 2005.
- [7] D. Kincaid and W. Cheney, *Numerical Analysis*. St. Paul, MN: Brooks/Cole Publishing Company, 1991.
- [8] C. A. Floudas, *Nonlinear and Mixed-Integer Optimization: Fundamentals and Applications*. Oxford, U.K.: Oxford Univ. Press, 1995.
- [9] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Trans. Consumer Electron.*, vol. 44, pp. 1122–1128, Aug. 1998.
- [10] M. Ghogho, "On optimum pilot design for OFDM systems with virtual carriers," in *Proc. IEEE EURASIP Int. Symp. Control, Communications, and Signal Processing*, Mar. 2006.
- [11] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1999.
- [12] S. Boyd and L. Vandenberghe, *Convex Optimization*. : CAM, 2004.
- [13] A. Dua, K. Medepalli, and A. J. Paulraj, "Receive antenna selection in MIMO systems using convex optimization," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 2353–2357, Sep. 2006.
- [14] K. Deb and M. Goyal, "A combined genetic adaptive search (GeneAs) for engineering design," *Comput. Sci. and Informatics*, vol. 26, no. 4, pp. 30–45, 1996.
- [15] Y. R. Zheng and C. Xiao, "Simulation models with correct statistical properties for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 51, no. 6, pp. 920–928, Jun. 2003.