A Unified Approach to Reduced-Redundancy Transceivers: Superfast Linear and Block-Iterative Generalized Decision Feedback Equalizers

Ricardo Merched, Senior Member, IEEE

Abstract—This paper shows, under general input data models, how block memoryless equalizers should be formulated considering reduced-redundancy transmissions for superfast detection. We propose linear and DFE-based, both multicarrier (MC) and single-carrier-frequency-domain (SC-FD) transceivers, along with efficient methods for the equalizer calculation, in a unified manner. We argue that, under a one-tap block decision feedback, transmitted redundancy can be reduced below the minimum $\frac{(L-1)}{2}$ samples allowed in the linear case, where $L$ is the channel length, even down to zero-redundancy, with improved BER performance. This is quantified in light of the optimal reconstruction delay set for a minimum-norm zero-forcing feedforward matrix in terms of the channel zeros location. The proposed MC and SC-FD block DFEs do not cancel inter-block-interference (IBI) via zeros-jamming; instead, it removes IBI completely, in part by decision-feedback, and in part by zero-padding, which allows for much lower redundancy transmissions. The remaining ISI is further eliminated through a one-step block-iterative-generalized-DFE (BI-GDFE) obtained in light of the optimal reconstruction delay set for a minimum-mean-square-error (MMSE) sense. Unlike computationally demanding block DFEs that eliminate ISI via successive cancelation, the proposed DFE schemes are as efficient as a superfast block-linear equalizer, requiring at most 3 receive branches to realize the order-$M$ feedforward matrices in $O(M \log M)$ operations.

Index Terms—Displacement structure, least-squares, MMSE, superfast algorithms.

I. INTRODUCTION

MATRIX inversions are at the heart of every single application one can think of. In modern communications employing block-based transceivers, least-squares (LS) and minimum-mean-square-error (MMSE) solutions are common instances where employing exact formulas is strictly prohibited, mostly due to computation inversion steps. In these scenarios, practical implementations widely resort to simplified, albeit suboptimal solutions, as palliative to the computational burden inherent to exact optimal formulas. For instance, orthogonal-frequency-division-multiplex (OFDM) and single-carrier frequency domain (SC-FD) systems [1] are popular examples of block memoryless equalizers, in which a particular choice of a DFT-based channel precoder yields very simple receivers [2]. This is due to the simple matrix algebra associated with circulant and skew-circulant matrices.

With the recent advances in VLSI design, the theory of structured matrices has evolved to an era where fast algorithms are no longer considered solely for their mathematical elegance, but rather by its true potential to replace simplified or approximate solutions with their original exact optimal formulas, at a reduced computational cost. In this context, the notion of displacement discussed in [6] forms the basis for very efficient matrix inversion and algorithms applied to signal processing and communications. It allows for another interpretation of the so-called Gohberg-Semencul formula [7], which further gave rise to superfast methods for matrix-vector multiplications. Superfast $O(M \log^2 M)$ representation of matrices ($p \leq 3$) refers to the solution of a displacement equation of Toeplitz structures with respect to factor circulant operators [7]. The solution yields efficient DFT-based expressions for Toeplitz inverses which are now widely known. In digital communications, the first mentions on the use of these efficient representations are given in [8] in the context of channel estimation and equalization in high-mobility environments. In the latter, it is shown that the general result of [7] can be directly applied to pilot-aided channel estimation problems within several scenarios where Toeplitz structures are induced. The advantage of this simple observation lies in the fact that the inversion step in MMSE or LS channel estimation formulas can be performed offline, while the pilot information needed for channel recovery can be easily stored in the transform-domain. The same goes for the equalization step, just by interchanging the roles of data and channel in the corresponding linear transmission model [9]. As a byproduct, while a low displacement rank allows us to represent an equalizer efficiently, we can also make use of this theory to compute them efficiently as well. For example, it is shown in [10] that MIMO decision feedback equalizers (DFE) can be calculated with the help of fast transversal (FT) recursive least squares (RLS) recursions, or via lattice cascades in the case of SISO block DFEs for shift data structures (see also [9], [11], [13]).

Contributions: A recent approach in [6], [15] describes a unified way of representing structured inverse covariances of a data matrix $H$, defined as $P = (H^*H)^{-1}$, by solving a displacement equation of the form

$$P - \Phi_\theta \Phi_\zeta^* = [I]_T,$$

for arbitrary operators that can be written as $\{\Phi_\theta, \Phi_\zeta\} = \{Z_\theta^{-1}\Psi, \Phi_\zeta = Z_\zeta^{-1}\Psi\}$, where $\Psi$ is a matrix that relates two successive regressors (rows) of $H$, and the choice of certain companion matrices $\{Z_\theta, Z_\zeta\}$ is free and designer.
induced. If the displacement operators are suitably chosen, the right-hand-side of (1) is said to have a low displacement rank. As we argued in [6], and unlike normally assumed from a mathematical point of view, one is not required to search for the operator that will fit a certain data structure, ultimately leading to low displacement-ranks; On the contrary, in signal processing and communications, it is rather the operator that acts on the data by changing its basis of representation, so that any arbitrarily structured operator will do. The efficiency of a certain inverse or inverse covariance representation is a result of proper choices of basis functions jointly with the free companion forms \( \{ Z_\theta, Z_c \} \). In other words, from the viewpoint of systems realizations, the role of the operator is to induce structure in the data, i.e., \( \mathbf{H} \), and therefore, the displacement rank of \( \mathbf{P} \) is at most 4, regardless of structure, and irrespective of the operators. The number of rank-one factors in the displacement equation is simply a function of the number of abrupt breakpoints seen along the rows and columns of \( \mathbf{H} \), which can be pre-windowed or post-windowed (meaning with a top lower or bottom upper triangular structure); for example, it is equal to 2 for doubly windowed data matrices; it is equal to 3, for a pre-windowed setup, and it is equal to 4, for non-pre-windowed data structures with an exponentially weighted window shape. A pure matrix inverse \( \mathbf{K} = \mathbf{H}^{-1} \) also yields a rank-2 displacement of the form \( \mathbf{K} - \mathbf{\Phi}_d \mathbf{K} \mathbf{\Phi}_e^* = [1, 1]^T \).

Following the results of [8], [9], the authors in [18]-[21] borrowed the superfast DFT-representation formulas of [7] with the intention to apply them to multicarrier (MC) and single carrier-frequency-domain (SC-FD) equalization, under the so-called minimum redundancy (MR) transmission scenario [16]. The conclusions regarding the performance of these particular systems in the superfast context are unfortunately flawed, as can be verified by slightly modifying their experiments parameters—see Section V. The performance of MR systems still lags behind the ones of standard MC and SC-FD systems by far, even for the carefully channels selected in [18]. The settings to which these ideas were particularized are such that their schemes presents the worst BER performance and the highest computational requirements compared to standard cyclic prefix based schemes. The motivation in these papers is solely based on the benefit of having halved the block redundancy at extremely high loss in BER. While the reasons for the deteriorated behavior of these reduced redundancy transceivers are well understood from simple equalization and linear algebra arguments, the authors in [18] forget to consider the fact that there exist an optimal reconstruction delay when implementing any equalizer, which is based on the zeros location of the channel impulse response (CIR). In this sense, linear minimum-redundancy transceivers represent the worst case scenario where the designer is not given the opportunity to set the right delay that will minimize the output noise power. This is why these systems do not work, except in some “pathological” circumstances (as the ones carefully chosen in their references).

In this paper, we rely on the concept of optimal reconstruction delay in order to motivate the construction of superfast reduced-redundancy transceivers. To this end, we make use of a unified polynomial Vandermonde representation of structured matrices developed in [6], in connection to fast Kalman recursions for general basis functions [14]. Specifically, our contributions consist in the following:

1) We show how the theory of fast and superfast algorithms is directly connected to the computation and realization of non-adaptive equalizers and channel estimators in the context of general bases. This is pursued in block memoryless MC and SC-FD equalization scenarios, where we identify fast decompositions in symbol and channel estimates, by writing the model accordingly. The superfast formulas are obtained by solving the displacement equation with respect to suitable user-designed operators [6]. The operators consist of a composition of companion matrices \( \{ Z_\theta, Z_c \} \), which do not compromise data throughput, and a precoder, which can be interpreted as a change of basis with no additional complexity. The equalizer parameters consist of the displacement generators, and we show how they are computed through a fast transversal (FT) algorithm intended to general models, as the one in [14];

2) We show how to select the transmitted redundancy in zero-forcing (ZF) transceivers such that the output noise power is minimized. This is motivated by the fact that an optimal minimum-norm ZF equalizer is implemented by only two superfast receive branches, which corresponds to the displacement rank of the MMSE covariance. As a fallout, it suggests a criterion to reduce redundancy in the case of MMSE receivers as well;

3) It is verified that under a one-tap block decision feedback, redundancy can be reduced below the \( \lfloor (L - 1)/2 \rfloor \) minimum number of samples allowed in the linear case, even down to zero, with improved BER performance. The amount of redundancy is quantified in light of the optimal delay set for a minimum-norm-ZF feedforward matrix in terms of the channel zeros location. The proposed one-block-tap DFE uses sufficient statistics in symbol demodulation, and does not cancel inter-block-interference (IBI) via zero-jamming (ZJ); Instead, it removes IBI in part by decision feedback, and in part by zero-padding, which allows for much lower redundancy transmissions. Note that zero-redundancy DFE based equalizers have been proposed in [22], [23], where IBI is completely cancelled via DF. While in [22] the receiver after IBI removal is linear, the receiver in [23] utilizes a second DFE to remove ISI via successive cancelation1. In all these instances, the complexity for computing and realizing such schemes are excessively high, and therefore impractical. Here, we show that the proposed DFE requires at most 3 receive branches in order to implement the feedforward matrix in superfast complexity. The one-tap DFE intended to IBI removal is combined with a one-step MMSE block-iterative generalized DFE (BI-GDFE) which removes the remaining ISI. We verify that a single re-estimation improves the performance of a reduced-redundancy schemes significantly, achieving the same performance of a full \( L - 1 \)

1A reduced redundancy block DFE based on [23] was also proposed in [25]; however, in their scheme the remaining IBI is not canceled, but only minimized by properly optimizing the precoder and feedback matrices. Although other references have considered a block DFE for reduced redundancy, these aim only inter-symbol-interference (ISI) elimination and exhibit inferior BER performance since relevant information of the received block is discarded through zero-jamming.
redundancy DFE that does not employ re-estimations. This further motivates us to pursue exact DFT-based MC and SC-FD versions of the DFE and the BI-GDFE, which are obtained here more generally under the polynomial Vandermonde decompositions of [6];

4) In [18]–[21], the authors made use of well known DFT factorizations for the purpose of block linear equalization in the MR transmission scenario. As a consequence of the discussions on the optimal reduced-redundancy transmissions, and as verified in our experiments using slightly modified scenarios of [18]–[21], we show that MR schemes offer no advantage over standard schemes, regardless of the block size, even when power loading is considered. Our simulations contradict the conclusions in [18]–[21], and verify on the other hand, that MR-DFEs outperform the corresponding MC and SC-FD linear MR counterparts significantly, under much lower redundancy transmissions, and similar superfast computational complexity.

Notation: We denote by $\{ \cdot \}^*$ the complex conjugate transposition. An $(N \times M)$ matrix will be generally denoted by $H_{M,N}$ unless a simplified notation is locally used for convenience. The notation $\mathcal{H}(a:b,c:d)$ captures a submatrix of $\mathcal{H}$ extending from row $a$ to row $b$ and column $c$ to $d$. The use of $\text{diag}(\mathbf{v})$ refers to a diagonal matrix with elements defined by the vector $\mathbf{v}$. We use the same notation to capture the diagonal elements of a matrix into a vector $\mathbf{v}$. We write $\text{Diag}(A)$ in order to define $\text{diag}(\text{diag}(A))$. The operator $[-]$ rounds the argument to the nearest integer towards infinity, while $\lfloor \phi \rfloor$ is the phase of $\phi$. The expectation operator is denoted by $\mathbb{E}$.

II. SUPERFAST MEMORYLESS BLOCK EQUALIZATION

Consider a discrete linear time invariant (LTI) single-input-single-output (SISO) channel $H(z)$ of length $L$, described as a block-based one via a $P \times P$ pseudocirculant matrix

$$
H(z) \triangleq H_0 + H_1 z^{-1},
$$

for transmitted vectors of size $P > L$. (see e.g., [2]). The coefficient matrix $H_0$, with first row given by the channel samples $h[0] h[1] \cdots h[L-1] 0 \cdots 0$, represents ISI, while $H_1$, represents the interblock interference (IBI).

Let $\mathbf{x}_i = [x(i) x(i-1) \cdots x(i-P+L)]^T$ be the $P \times 1$ transmitted vector at time $i$, so that the received block is written as

$$
\mathbf{y}_i = H \mathbf{x}_i + \mathbf{v}_i, \quad \text{where} \quad H = [H_0 H_1(:,0:L-1)].
$$

For the sake of generality, we consider a block affine precoding transmission [26], i.e., $\mathbf{x}_i = \mathbf{A}_i \mathbf{s}_i + \mathbf{t}_i$, where $\mathbf{s}_i = [s(Mi) s(Mi-1) \cdots s(Mi-M+1)]^T$ is the information vector and $\mathbf{t}_i = [t[Mi] t[Mi-1] \cdots t[Mi-M+1]]^T$ is the superimposed training vector used for estimating the channel within the $i$-th transmitted block. The role of the $P \times M$ matrix $\mathbf{A}_i$ is threefold:

(i) To control the level $\delta$ of symbol redundancy by transmitting $P = M + \delta$ symbols; (ii) to promote power loading; and (iii) in the light of [6], to perform a change of basis, each case targeting complexity and optimality interests. In block memoryless equalization, redundancy eliminates IBI via zero-padding (ZP) or zero-jamming (ZJ) with or without cyclic prefixing, or in the more general case, through a hybrid form of these schemes (ZP-ZJ) [16]. That is, assume that the channel state information (CSI) is available, and define the restriction matrices

$$
T_\delta = \begin{bmatrix} 0_{\delta \times M} \\ I_M \end{bmatrix}
$$

$$
\tilde{T}_\delta^T = \begin{bmatrix} I_P \\ L+1+\delta \\ 0_{P-L+1+\delta} \end{bmatrix} \mathcal{J}(L-1-\delta),
$$

where $\delta \in \{0, L-1\}$. The matrix (3) is multiplied by the transmitted vector so as to perform ZP, while (4) is multiplied by the output block for the purpose of ZJ. In this way, defining the precoder as $\mathbf{A}_i = T_\delta \mathbf{A}_i$ and assuming an additive noise vector $\mathbf{v}_i$ with power $R_v = \sigma_v^2 I$, the received block after IBI removal is given by

$$
\mathbf{y}_i = (\tilde{T}_\delta^T H_0 \mathcal{J}_\delta) \mathbf{A}_i \mathbf{s}_i + \tilde{T}_\delta^T \mathbf{v}_i = H_0 \mathbf{A}_i \mathbf{s}_i + \mathbf{v}_i,
$$

with first row given by $[h(\delta) \cdots h(L-1) 0 \cdots 0]$, and first column $[h(\delta) \cdots h(0) 0 \cdots 0]^T$. The extreme cases of $\delta = 0$ and $\delta = L-1$ correspond to the full ZJ and ZP schemes respectively. Choices between these values are said of reduced redundancy [16]. The case when $\delta = \lfloor (L-1)/2 \rfloor$ zeros are padded and discarded at the receiver has been referred to as a minimum-redundancy system, where the reminiscent ISI is given by a square Toeplitz matrix, which we shall assume invertible, in principle. Now, by making use of the polynomial Vandermonde factorization framework of [6], block SC-FD and MC type schemes can be promptly envisioned, as we shall explain; For simplicity, we assume $\mathbf{A}_i = I$ and drop the block index $i$ from here on.

A. Displacement Structure in Signal Processing

Our goal in this section is to highlight an important distinction between the displacement theory approach, addressed from a purely mathematical perspective, and its formulation under a specific signal processing application. To see this, we first formally introduce the concept of displacement of an arbitrarily structured matrix.

Definition 1: A matrix $\mathbf{A}$ is said to have a displacement structure with respect to the operator matrices $\{ \Phi, \Gamma \}$, if it satisfies the Stein and/or Sylvester displacement equations

$$
\nabla_{\{ \Phi, \Gamma \}^*} \mathbf{A} \triangleq \mathbf{A} - \Phi \mathbf{A}^* \Gamma = LM^*,
$$

$$
\nabla_{\{ \Phi, \Gamma \}^*} \mathbf{A} \triangleq \Phi \mathbf{A} - \mathbf{A}^* \Gamma = L'M^*
$$


2A pseudocirculant matrix is a basically a circulant matrix where the elements strictly below the diagonal are multiplied by a constant (here by $z^{-1}$).
where \(\{L, M\}\) are \(M \times t\) matrices whose columns are referred to as the generators of \(A\). The cardinal \(t\) is called the displacement rank of \(A\), where \(t \ll M\).

The type of operators \(\{\Phi, \Gamma\}\) that yield a low rank r.h.s. of (7) are normally chosen according to a given structure \(A\). For example, Toeplitz and Hankel matrices have displacements ranks with respect to factor circulant operators \(\{\Phi = Z_\gamma, \Gamma = Z_\gamma\}\), which does not exceed 2 [see, e.g., (23) further ahead]; Cauchy and the so-called polynomial Vandermonde matrices have displacement ranks with respect to diagonals \(\{\Phi = D_x, \Gamma = D_t\}\) and diagonal/Hessenberg matrices \(\{\Phi = D_x, \Gamma = \Psi\}\) which does not exceed 1 [12]. While these results can be proven for such specific structures, defining displacement operators for arbitrarily structured matrices is not an easy task. In particular, we are interested in the class of operators that will produce a lowrank representation of a covariance \(A = P_M\) such that \(\nabla\{s, r^\p\}(A)\) is induced by any given first-order data model.

The displacement structure of a certain matrix can be exploited implicitly or explicitly, in different scenarios. The Extended Generalized Sliding-Window Fast Transversal Filter (EGSWFTF) algorithm of [14] is an example where the displacement generators are used to update the solution of a LS problem by replacing the direct operations with the coefficient matrix \(P_M\), with the ones involving its generators instead. This is seen from the fast array version of the EGSWFTF. In this sense, the EGSWFTF performs the displacement decomposition implicitly.

A second way to exploit structure, is to solve the displacement equations (either in its Stein or Sylvester forms) of (7) for \(A\). Depending on the choice of the operator, the solution may be represented efficiently, and used explicitly, for example, in the realization of a LS or a MMSE formula for a certain signal processing application. Observe that while the former makes use of the displacement of \(P_M\) in an adaptive scenario, the latter can be seen as a non-adaptive, block realization of \(P_M\). Moreover, since the parameters of this decomposition have an exact interpretation as normalized Kalman and prediction vectors, the computation of the generators can be accomplished by an EGSWFTF algorithm as well.

In the above contexts, the central results of this paper rely on the fact that one is not required to search for a specific operator that will lead to a low displacement-rank, and consequently to an efficient representation of a certain inverse or inverse covariance matrix, usually desired in signal processing and communications applications. On the contrary, in these contexts, it is rather the operator that acts on the data, redefining its structure; in this sense, a low rank factorization holds regardless of the operator, the relevant question here is therefore how one should pick a suitable basis that will induce an alternative representation useful for a certain purpose.

In the following, we specify the class of operators \(\{\Phi, \Gamma\}\) that will produce a low rank r.h.s. of (7) where the generators of \(A = P_M\) [as mentioned in (1)] are explicitly defined, regardless of data structure. Extension to more general non-Hermitian cross-variances is straightforward.

**B. SC-FD Equalization**

Consider a transversal system realization based on arbitrary basis functions \(\{Q_m(z)\}\) as illustrated in Fig. 1.

![Fig. 1. Transversal realization based on general basis.](image1)

We organize the input regressors \(\{u_{M,n}\}\) of this network into a structured \(R \times M\) data matrix \(H_{M,N}\), i.e., \(H_{M,N} = [u_{M,N-1} + 1, \ldots, u_{M,N}]\). When \(\{Q_m(z)\}\) are constructed from recurrence-related polynomials, these induce a fixed relation between two successive rows \(\{u_{M,n}\}\), i.e., \(u_{M,n} = \Psi u_{M,n-1}\), where \(u_{M,n} = [u_{M-1,n}, u_{M-1,n-1}]\), and \(u_{M,n+1} = [u_0(n + 1), u_{M-1,n+1}]\). In this case, it can be shown that \(\Psi = B_Q = I\).

Similarly to \(H_{M,N}\), define the \(R \times M\) transformed data matrix

\[
\hat{H}_{M,N} = \begin{bmatrix}
\vec{u}_{M,N}^{-R+1} \\
\vdots \\
\vec{u}_{M,N}^{-1} \\
\vec{u}_{M,N}^0
\end{bmatrix} = \begin{bmatrix}
\hat{H}_0 B_Q^* \\
\vdots \\
\hat{r}_{M,N}
\end{bmatrix}
\]

where by virtue of the delay line of Fig. 2, \(\hat{H}_0\) exhibits a Toeplitz-like structure.

Now, let us return to the linear model of (5). The LS estimate of \(s\) is given by \(\hat{s}_{LS} = K_{LS} y\), where

\[
K_{LS} = (\mu I + H_0^* H_0)^{-1} H_0^* = B_Q (\hat{H}_{M,N}^{-1} + H_{B} H_B)^{-1} B_Q H_0^* \quad \hat{P}_M
\]

where \(\hat{H}_{M,N}^{-1} = \mu B_Q B_Q^*\), and where we make the identifications \(H_0 \rightarrow H_3\), and \(H_R \rightarrow H_{M,N}\) of (9). Because \(H_0\) in general does not exhibit an upper or lower triangular structure, it can be shown from [14], that the following displacement equation for \(\hat{P}_M\) holds, in connection to its defining fast Kalman recursion variables:

\[
\nabla\{\Phi_s, \Psi_s\}(\hat{P}_M) = \hat{P}_M - \Phi_s \hat{P}_M \Psi_s^* = \hat{P}_s \hat{M}_N \hat{k}_M \Psi_s^* + \hat{w}_M^* \hat{k}_M \hat{M}_N \Psi_s^* - \hat{z}_M^* \hat{w}_{M-1,N-1} \hat{k}_{M-1,N-1} \Psi_s^* - \hat{k}_{M-1,N} \hat{M}_N \Psi_s^*.
\]
where \( \{ \tilde{w}_{M-1,N}, \tilde{w}_{1,N-1}, \tilde{k}_{M,N}, \tilde{k}_{M-1,N} \} \) correspond to normalized backward and forward prediction vectors, and the Kalman gains associated to data breakpoints at the first and last row of \( H_0 \). The matrices \( \{ Z_\theta, Z_\zeta \} \) have companion forms, with last columns given by the vectors \( \{ \theta = [\theta_0 \cdots \theta_{M-1}]^T, \zeta = [\zeta_0 \cdots \zeta_{M-1}]^T \} \):

\[
Z_\theta = \begin{bmatrix}
0 & 0 & \cdots & \theta_0 \\
1 & 0 & \cdots & \theta_1 \\
0 & 1 & \cdots & \theta_2 \\
0 & 0 & \cdots & 1 \theta_{M-1}
\end{bmatrix},
\]

(11)

\[
Z_\zeta = \begin{bmatrix}
0 & 0 & \cdots & \zeta_0 \\
1 & 0 & \cdots & \zeta_1 \\
0 & 1 & \cdots & \zeta_2 \\
0 & 0 & \cdots & 1 \zeta_{M-1}
\end{bmatrix},
\]

(12)

A key result of [6] is that a low displacement rank (in the above example, of 4), can always be satisfied as long as the operators \( \{ \tilde{\phi}_\theta, \tilde{\phi}_\zeta \} \) are chosen in connection to the basis functions that generate the data in \( H_0 \) defined in (10) as \( \{ \tilde{\phi}_\theta = Z_\theta^{-1}, \tilde{\phi}_\zeta = Z_\zeta^{-1} \phi_0 \} \). Hence, by solving (11), we are able to find a general representation for \( \hat{P}_M \) in terms of the eigenvectors of the constructed operators \( \{ \tilde{\phi}_\theta, \tilde{\phi}_\zeta \} \). We next summarize the main result of [6] and provide a brief background on how this is interrelated with the construction of superfast receivers and their computation. Since these represent quite involved results, we encourage the reader to refer to [6] for more details.

**Theorem 1 (Polynomial Vandermonde Representation of Covariance Bezoutians):** Let \( \hat{P}_M \) be the inverse covariance matrix arising in a generalized window least-squares formulation for an arbitrary recurrence related polynomial basis \( \{ Q_k(z) \} \). Let the vector \( z_1 = [z_1(0) \cdots z_1(M - 1)]^T \) contain the distinct eigenvalues of

\[
\tilde{\Phi}_\theta = Z_\theta^{-1} \Sigma M = V_Q^{-1}(z_1) A_z, V_Q(z_1),
\]

satisfying its characteristic polynomial \( \Omega_\theta(z) \), where

\[
\hat{V}_Q(z) = \begin{bmatrix}
Q_1(z_1) & Q_2(z_1) & \cdots & \tilde{\Sigma}_M(z_1) \\
Q_1(z_2) & Q_2(z_2) & \cdots & \tilde{\Sigma}_M(z_2) \\
\vdots & \vdots & \ddots & \vdots \\
Q_1(z_{M-1}) & Q_2(z_{M-1}) & \cdots & \tilde{\Sigma}_M(z_{M-1})
\end{bmatrix}
\]

(13)

is a polynomial Vandermonde matrix. Let \( \mathcal{Q}_m(z), m = 1, \ldots, M - 1 \), define its \( m \)-th column. Given the free choice of the master polynomial \( \tilde{\Sigma}_M(z) = z^{-1} \tilde{\Sigma}_M(z) \), defining the following matrix-valued polynomial:

\[
\Omega_\theta(\tilde{\Phi}_\theta) = \phi_0 I + \sum_{m=1}^{M} \phi_m \tilde{\Phi}_\theta^{-m}
\]

(14)

with \( \theta = \Sigma_M \hat{V}_Q^{-1}(z_1) A_z^{-1} \mathcal{Q}_M(z_1) \), the following slightly changed version with zeros \( z'_1 \)

\[
\Omega'_\theta(\tilde{\Phi}_\theta) = \phi_0 I + \sum_{m=1}^{M} \phi_m \tilde{\Phi}_\theta^{-m}
\]

(16)

obtained by replacing the DC coefficient \( \phi_0 \) with \( \phi_0 \). As an abuse of notation, we denote by \( 1/z \) an entrywise inversion of \( z \). Set \( z_2 = [z_2(0) \cdots z_2(M - 1)]^T = 1/z_1' \), in a way that the coefficients of the \( M \) highest powers of \( \Omega_\theta(z) \) and \( \Omega'_\theta(z) \) coincide, where \( \{ z_1, z_2 \} \in \mathbb{C} \). Let \( \tilde{V}_Q(z_2) \) be the eigenvector matrix that corresponds to \( \hat{V}_Q \). Then, assuming that \( \phi_0 \neq \phi_0^* \),

\[
\hat{P}_M = \frac{\phi_M}{(\phi_0 - \phi_0^*)} \sum_{k=1}^{4} i_k \tilde{V}_Q^{-1}(z_1) A_{V_1, b_k, s} V_Q(z_1) \tilde{V}_Q^*(1/z_2^*)
\]

(17)

where \( I^* \) reverses the order of the entries of a vector,

\[
\tilde{V}_Q(z_2) = D^{-1}(1/z_2^*) A_{V_2, b_4, s} \tilde{A}_{V_2, b_4, s} \tilde{A}_{V_2, b_4, s}^{(M-1)*},
\]

(18)

\[
A_{V_1, b_k, s} = \text{diag}(\tilde{V}_Q(z_1) b_k, s),
\]

(19)

\[
D(z) = \text{diag}(d_1^0, \ldots, d_{M-1}^0),
\]

(20)

and where for compactness of notation we denote \( b_1, \theta \triangleq \tilde{M}_{M,N}, b_2, \theta \triangleq \tilde{M}_{1,N}^{-1} Z_1^{-1} \tilde{w}_{M-1,N}, b_4, \theta \triangleq \tilde{M}_{M-1,N}^{-1}, b_4, \zeta \triangleq \tilde{M}_{M-1,N}^{-1} z_{M-1}^{-1} \tilde{w}_{M-1,N} \), and \( b_4, \zeta \triangleq (M_{M-1,N}, \text{with } t_1 = t_2 = -1, t_4 = 1. \text{ Analogous definitions hold with respect to } \{ V_Q(z_2), A_{z_2}, Z_\zeta \} \text{—See Table 1 in Section IV, for a displacement rank-3 example.}

The precise definition of these generators in terms of Kalman vectors allows us to calculate these parameters recursively and exactly, as we shall explain in the sequel (See Section IV). As a result, equalizers that rely on inverse covariances will naturally have its parameters obtained through an efficient (extended fast transversal) algorithm, as long as the input basis functions are generated by recurrence relations.

Fig. 3 summarizes the unification of the theory of fast and superfast decompositions with the applications proposed in this paper. The approach on structured matrices in the more general adaptive case encompasses the development of the Extended Generalized Sliding Window Fast Transversal Filter (EGSWFTF) and the solution of the displacement equation of the corresponding data covariance for arbitrary operators [14], [6]. While the EGSWFTF recursions are adaptive, they exert direct impact on the computation of non-adaptive scalar and block transmission equalization techniques, which can be formulated under arbitrary basis functions. The usefulness of changing basis representation stems from compactness of models and efficient superfast realizations, for which the computation of the displacement generators in connection with the Kalman recursions in both cases was unavailable, even for tapped-delay-line models. The choice of free companion structures along with recurrence related basis representation yields an exact polynomial Vandermonde based decomposition, from the solution of the corresponding displacement equation. As a result, proper choices for the pair \( \{ \tilde{\Phi}_\theta, \tilde{\Phi}_\zeta \} \) lead to representations of highly structured inverses, extending the standard DFT formulas to other signal transformations. Next, we show
how these general receivers collapse to the ones employing DFT and DCT matrices.

**DFT Based Superfast Receivers:** Fig. 4 illustrates the regularized least-squares (or MMSE for the same matter) receiver of (10) that makes use of the representation (17). The displacement rank of 4 of the covariance of $H_0$ implies that the receiver requires 4 branches, regardless of the basis representation.

The well known DFT-representation is just a special case of the above formula, considering the zeros of the master polynomials $\Omega_0(z) = \phi_0 + z^{-M}$, and $\Omega_0(z) = \gamma_0 + z^{-M}$ calculated at \{z_1, 1/z_1^*\}. That is, in the latter, $z_1(m) = \phi e^{j\frac{2\pi m}{M}}$, where $\phi = \phi_0^{-1/M} e^{j\frac{2\pi m}{M}}$, and $z_2(m) = \gamma_0 e^{j\frac{2\pi m}{M}}$, with $\gamma = 1/\sqrt{M} 1^{j\frac{2\pi m}{M}}$. Defining $D_\phi \overset{\triangle}{=} \text{diag}(\phi^{-m} |_{m=0}^{M-1})$, this results in the DFT filterbanks $V_P(z_1) = \sqrt{M} D_\phi$, $V_P(1/z_1^*) = \sqrt{M} D_{1/\phi}$, and $D^{-1}(1/z_1^*) = M g^{(M-1)} \text{diag}(e^{j\frac{2\pi m}{M}} |_{m=0}^{M-1})$, where $F$ is the DFT matrix. The representation of $\tilde{P}_M$ then becomes

$$\tilde{P}_M = \frac{D_\phi F^*}{(\phi_0 - \bar{\phi}_0^*)} \sum_{k=1}^M \bar{A}_{V_k, \ell, k} \bar{A}_{V_k, \ell, k}^* \bar{F} \bar{D}_{\delta, \ell}^* \bar{F} \bar{D}_{\delta, \ell} \bar{A}_{V_k, \ell, k} \bar{F} \bar{D}_{\delta, \ell} \bar{A}_{V_k, \ell, k}$$

(22)

where $\bar{A}_{V_k, \ell, k} = \phi^{(M-1)} \bar{A}_{V_k, \ell, k} \text{diag}(e^{-j\frac{2\pi m}{M}} |_{m=0}^{M-1})$. Alternatively, we can factor $D^{-1}(1/z_1^*)/M$ into the definition of
Fig. 5. SC-FD DFT Decomposition.

Remark: Two particular transceivers that rely on the inversion of Toeplitz matrices are of special interest:

1) ZF Receiver: In this case, a portion of size $M$ of the received vector $y$ is captured, so that the resulting linear model relies on a simple inversion of a square Toeplitz matrix. Since Toeplitz inverses have a displacement rank of 2 with respect to factor circulants, we can represent it as follows:

$$
\begin{align*}
(\hat{T}_d H_0)^{-1} &= \frac{1}{1 - \varphi_d / \varphi_0} D_{1/\varphi} D^* F D_{1/\varphi}^* F^* \Lambda_{\varphi} \Lambda_{\varphi}^* \alpha_2 \\
&= \Lambda_{\varphi} \Lambda_{\varphi}^* F^* D_{1/\varphi} D F^* D_{1/\varphi}^* \alpha_2
\end{align*}
$$

where $d$ is a suitable delay chosen to minimize the noise power at the receiver. Unlike the rank-4 case, the diagonal matrices $\Lambda_{\varphi}$ depend only on two prediction (generating) vectors $\{\hat{w}_1, \hat{w}_2\}$ (see details in [9], [6]).

2) Full $(L - 1)$ Redundancy ZP Receiver: It is well known that higher redundancy results in better BER performance. Moreover, besides superiority in detection, ZP schemes also allow for less complex representations, since when $\delta = L - 1$, $H_0$ exhibits a doubly-windowed structure, and so $H_0^* H_0$ in this case becomes symmetric Toeplitz. Hence, its inverse is represented via 2 branches only, except that here symmetry implies computation of a single generating vector. This fact was already used in [8], [9] for channel estimation in a high Doppler OFDM setup.

Trigonometric Transforms Based Superfast Receivers: There are some unsolved issues which prevent the use of equalization formulas for arbitrary transforms, mostly due to the lack of a unified treatment of this subject. For instance, DCT-based fast convolution techniques [4] have only been defined for symmetric channels, and recent works even imply that DCT-OFDM schemes require the said symmetry to be feasible [5]. In fact, as can be verified in [15], there is an alternative reason for the appearance of a certain desired transform, as a consequence of the choice of the input basis representation. The polynomial Vandermonde decomposition of Theorem 1 allows us to choose other suitable operators based on alternative transforms. For instance, considering a ZF receiver, when $B_Q$ is Chebyshev induced, we obtain the following DCT-based decomposition [15]

$$
(\hat{T}_d H_0)^{-1} = \alpha B_Q C M
$$

where $\alpha = \frac{\varphi_d \varphi_0}{M^2 - \varphi_0^2}$, and $C_M$ is the DCT-III matrix. These decompositions are particularly useful for real data constellations, since they rely solely on real transforms, which naturally arise without any symmetry constrains. Fig. 6 illustrates the SC-FD DCT-decomposition.

Remark: Since our goal in this paper is to verify the performance of superfast receivers in new reduced redundancy linear and DFE based contexts, we shall focus mostly on DFT transceivers. The performance of SC-FD receivers is identical for other eigenvector transformations, given that the structure of the equalizer is unchanged in this case. MC transceivers based on other polynomial Vandermonde transformations on the other hand, will be investigated in a future work, since a change to a possibly non-orthonormal basis requires careful, non-equipower loading.

C. Optimal Redundancy for Minimum-Norm Zero-Forcing SC-FD Equalization

The goal of a ZF scheme is to invert a submatrix of $H_0$, when $\delta \geq \frac{(L - 1)}{2}$, so that the receiver is implemented with only 2 branches. That is, define $K_{ZF} = (\hat{T}_d H_0)^{-1}$ where

$$
\hat{T}_d = \begin{bmatrix}
0_{M \times d} & I_M & 0_{M \times 2\delta - L + 1} \cdot d \in \{0, 2\delta - L + 1\}
\end{bmatrix}
$$
Two important issues arise in the receiver design: 1) Which level of redundancy must be introduced. For example, from the perspective of bandwidth efficiency, a MC obtained with \( \delta = \lfloor (L - 1) / 2 \rfloor \) is appealing. This is however, a naive choice, since it fixes only one possibility for matrix inversion, implying the highest probability of noise amplification, for arbitrary channels. Because each choice of \( \delta \) results in different conditioning for \( \mathbf{H}_0 \), and so does the choice of \( d \), we are led to a more relevant question: 2) What is the optimum amount of redundancy \( \delta \) and delay \( d \) such that the noise power is minimized at the output? According to our context, the problem is to set \( d = d_0 \) such that

\[
d_0 = \arg \min_d \left\| \mathbf{K}_{ZF} \right\| \quad \text{s.t.} \quad \mathbf{K}_{ZF}^\dagger \mathbf{T}_d^\dagger \mathbf{T}_d^* \mathbf{H}_0 \mathbf{I}_d^d \mathbf{I}_d - \mathbf{I}
\]

and \( d \in [0, 2b - L + 1] \) (25)

Interestingly, a similar problem stated also in this mathematical form was solved for \( \delta = L - 1 \) in a famous paper by Scaglione et al. [3] (although within a different context), building on a reasoning that yielded the scalar counterpart solution (see the references therein). For \( M \gg 1 \), the solution \( d_0 \) in [3] is given by the number of minimum-phase zeros of \( \mathbf{H}(z) \), and in particular,

\[
\lim_{M \to \infty} \frac{1}{M} \left\| \mathbf{T}_d^\dagger \mathbf{H}_0 \right\|^{-1} = \begin{cases} \infty, & d \neq d_0 \ 
0, & d = d_0 
\end{cases}
\]

(26)

Note that any choice of \( \delta < L - 1 \) implies a smaller set of possible choices for \( d \), which means that \( \mathbf{H}_0 \) may not contain the submatrix that will correspond to the optimal delay given by (25). This suggests that we must pick the redundancy optimally as \( \delta = d_0 \), which coincides with the optimal delay when \( \delta = L - 1 \) as given by (25), and then choose \( d_0 = 0 \) with respect to \( \mathbf{H}_0 \). In other words, we can match the optimal delay in equalizing a full ZP convolution matrix, with its optimal cut pattern given by \( \mathbf{H}_0 \) in (6), which arises in the the reduced-redundancy scenario. Hence, given the optimal delay, we immediately know what is the optimal ZP-ZF redundancy scheme such that the ZF matrix has minimum-norm. Combining these two pieces of information, we simply set

\[
\delta_{opt} = \# \text{maximum-phase zeros of } \mathbf{H}(z)
\]

(27)

An optimal delay in the context of a minimum-norm-ZF equalizer thus suggests that we can also combine it with a MMSE (LS) receiver, and still obtain a reduced-redundancy scheme, with some optimality. That is, let \( \mathbf{H}_{0, \delta} \triangleq \mathbf{T}_d^\dagger \mathbf{H}_0 \mathbf{I}_d \). Since when \( M \to \infty \), \( \mathbf{H}_{0, \delta} \) is approximatively square, we can borrow the result of (27) again, and because \( \mathbf{H}_{0, \delta} \) is in general a tall matrix, conditioning is improved. The receiver, however, will comprise 4 branches in general. That is, redundancy can be decreased at the expense of increased computational complexity, while maintaining optimality up to a certain level.

\[\text{D. Multicarrier Based Equalization}\]

Strictly speaking, a MC scheme is such that a (square) matrix derived from the (block) channel model is exactly diagonalized by the pair of transforms used at transmission and reception. This is possible by setting \( \delta = 0 \), so that under cyclic prefixing,

\[\text{III. SUPERFAST REDUCED REDUNDANCY BLOCK DFE TRANSCEIVERS}\]

Differently from the ZP-ZF approach, inter-block-interference can be removed via decision-directed receivers, and more importantly, without introducing any form of redundancy, through a simple one-tap block DFE. It is shown [22] that such DFE-OFDM transceiver outperforms a conventional OFDM system in terms of both symbol error rate and mutual information for indoor wireless networks, given that the former uses sufficient statistics in symbol demodulation, while the latter discards relevant received samples on which IBI exists. That is, once the channel is estimated, IBI is removed as

\[
\mathbf{y}_i = \mathbf{y}_i - \mathbf{H}_1 \mathbf{s}_{i-1} = \mathbf{H}_0 \mathbf{A} \mathbf{s}_i + \mathbf{v}_i
\]

and the role of the receiver is to deal with the remaining ISI represented by \( \mathbf{H}_0 \mathbf{A} \). For example, in [23], after IBI removal, the remaining ISI is removed by another DFE as illustrated in Fig. 7. Although this is optimal in a symbol-by-symbol detection sense when compared to a linear receiver, the computational complexity for obtaining the DFE matrices and realizing the corresponding block equalization are excessively high for practical purposes.

The framework of superfast solutions thus provides us a suitable implementation for the DFE matrices \( \{ \mathbf{K}, \mathbf{B} \} \), considering exact MC and SC-FD transceivers. Moreover, for each of these
schemes, we shall consider two possible configurations with distinct levels of complexity, yet both exhibiting superfast implementations.

**A. Superfast IBI Removal and Linear Block ISI Equalization**

Assume first, $B = 0$ and $A = I$. Because $\mathcal{H}_c$ is upper triangular, it can be verified that one pair of generators of the inverse which defines (23), collapses to $\tilde{w}_1 = \zeta^{-1/2}(N)e_0$, and $\tilde{w}_2 = \zeta^{-1/2}e_{M-1}$, where $\zeta = -w^T(1 - w)[1 - w]^T$ is the minimum cost associated with the forward LS prediction vector $w_0$ of order $M - 1$ w.r.t. the first column of $\mathcal{H}_0$, and $\tilde{u}$ is the first row of $\mathcal{H}_0$. The vectors $\{e_0, e_{M-1}\}$ are the pinning vectors corresponding to the zero-th and $(M - 1)$-th canonical bases. In this case, two of the transforms in the representation can be dropped, so that $K$ is given by

$$K = \frac{1}{2} \left[ \zeta^{-1/2} F^* \Lambda^*_c \tilde{w}_2 F + \zeta^{-1/2} D_{1/2} F^* \Lambda \tilde{z}_c \tilde{e}_c^2 F D_{1/2} \right], \tag{29}$$

while the convolution with $\mathcal{H}_1$ is easily done via circulant embedding.

This scheme shows that we could envision a “multicarrier” version of the above scheme, by moving the output transform before the slicer, to the input node, defined as the precoder $A = F^*$. The receiver is thus written as

$$K_{MC} = \frac{1}{2} \left[ \zeta^{-1/2} \Lambda^*_c \tilde{w}_2 + \zeta^{-1/2} D_{1/2} F^* \Lambda \tilde{z}_c \tilde{e}_c^2 F D_{1/2} \right], \tag{30}$$

which is illustrated in Fig. 8. Observe that we also include the channel estimation step, whose output will be passed to the equalizer computation phase, illustrated in shaded boxes. The goal of the former is to efficiently obtain the CSI, from which the generating vectors $\tilde{w}_1, \tilde{w}_2$ are computed by a fast algorithm. In the general Toeplitz case, they will be DFT-tranformed into the diagonal matrices $\{\Lambda, \Lambda_c\}$, appearing in (23). Here, due to the lower triangular form of $\mathcal{H}_0$, only one generator is required. We shall exemplify this procedure in Section IV.

We readily observe that while any linear reduced redundancy scheme requires 6 FFTs for Toeplitz inversion, the exact DFE-OFDM requires the same complexity, however, introducing zero redundancy. Note further that in general, both schemes are ill-conditioned, since no optimal delay is allowed in these cases. However, considering correct IBI cancelation, from [3], if the channel is minimum-phase, the zero-redundancy scheme is optimal in the minimum-norm ZF sense, which is the only case where such scheme is reliable. A LS solution thus becomes much more interesting both in terms of performance and complexity, the latter due to the post-windowed structure of $\mathcal{H}_0$.

The forms of IBI cancelation seen in the above block DFE and in (linear) reduced redundancy schemes, therefore suggest a more powerful combination of these two schemes in a single one, with enhanced detection performance. That is, note that after padding with $\delta$ zeros at transmission, instead of discarding $L - 1 - \delta$ samples at reception, we may opt to cancel these remaining IBI samples by decision feedback. Specifically, write

$$y_i^\delta = \mathcal{H}_c T^\delta A_i s_i + \mathcal{H}_1 T^\delta A_i s_{i-1} + v_i, \tag{31}$$

$$= H_0 A_i s_i + H_1 A_i s_{i-1} + v_i, \tag{32}$$

with the $(M + \delta) \times M$ blocks in (32) given by

$$H_0 = \begin{bmatrix} \delta + 1 \\ L - 1 \end{bmatrix}, \quad H_1 = \begin{bmatrix} \delta \\ L - 1 - \delta \end{bmatrix}, \tag{33}$$

so that

$$y_{i}^\delta = y_i^\delta - \mathcal{H}_1 T^\delta A_i s_{i-1} = H_0 A_i s_i + v_i. \tag{33}$$

This further suggests 2 slightly different DFE structures with useful features:

1. Assuming correct decisions at detection, a block DFE employing $\delta = \delta_0 \in [0, L - 1]$ optimum redundancy. This would provide improved performance compared to other scheme without DF, since besides using sufficient statistics in symbol demodulation, in principle, a reduced size feedback matrix guarantees a smaller error propagation effect, when compared to the case of full IBI cancelation via one-tap DF. Observe that the optimal redundancy in this case can be smaller than the one considered for the linear receiver, since the cut-pattern of $H_0$ is such that $\delta_0$ is allowed to be smaller than $\lceil(L - 1)/2\rceil$, without further discarding rows via ZJ (and thus always left-invertible).

2. A MR transceiver employing less redundant samples, such that when compared to the linear MR scheme, the MMSE or LS estimation is not degraded.

We conclude from 1. and 2. that compared to linear MR schemes, a block DFE employing $\delta_0 < \lceil(L - 1)/2\rceil$ is still expected to result in superior BER performance, as long as redundancy is not so small that ill-conditioning and error propagation become important; in this way, one could seek a balance between a minimum, zero-redundancy scheme, and one that employs more redundant samples, with superior performance against the linear MR scheme.

More importantly, in both cases, because the reminiscent model consists of a tall $(M + \delta) \times M$ matrix with a post-windowed structure, it can be verified that the Kalman gain vector $\tilde{k}_{M,N}$ in (11) vanishes, so that at most 3 feedforward branches are required to implement a MMSE or LS solution (in the case of a full $(L - 1)$ redundancy, only 2 branches are needed). The resulting MC scheme is illustrated in Fig. 9. The analogous SC-FD scheme is simply obtained by moving the DFT precoder back to the receiver end.

Fig. 9 also displays the channel estimation step, whose output will be passed to the equalizer computation phase, similarly to the ZF receiver. Here, the goal is to obtain the CIR estimate, from which this time, three generating vectors, $\tilde{w}_1, \tilde{w}_2, \tilde{w}_{M-1,N}$, and $\tilde{k}_{M,1,N}$ must be computed by a fast algorithm. They will be DFT-tranformed into the diagonal matrices $\{\Lambda_x, \Lambda_c\}$, $j = 1, 2, k = 1, 2, 3$, appearing in the three receive branches. To this end, in Section IV, we propose an efficient procedure for obtaining the equalizer parameters. It is described for general basis functions, and collapses to the FFT algorithm for shift data structures in the monomial basis case of Fig. 9. Block equalizers that rely on 2 or 4 receive branches
B. IBI Removal and MMSE-BI-GDFE

We now consider the DFE structure of Fig. 7, where IBI is eliminated in a hybrid, reduced-redundancy fashion. In a BI-GDFE [24], the formulation relies on a prior block decision \( \hat{s}_m \) which is first obtained according to a ZF or a MMSE linear equalizer. Then, instead of assuming correct decisions through the traditional assumption \( E\hat{s}_m^*\hat{s}_m = \sigma^2 \), the BI-GDFE relies on the “soft” assumption \( E\hat{s}_m^*\hat{s}_m = \rho_m \sigma^2 \), where \( \rho_m \) represents the input-decision-correlation (IDC) coefficient that reflects the reliability of the decisions taken at the \( m \)-th iteration of a re-estimation process. Unlike the approach in [24], where optimization is with respect to the signal-to-interference-plus-noise ratio (SINR), here we obtain improved estimates of \( \hat{s} \) by solving

\[
\min_{\mathbf{G}, \mathbf{B}} \mathbb{E}[\| \mathbf{s} - \mathbf{Gy} + \mathbf{B}s_{m-1} \|^2].
\]  

(34)

The concept of assigning uncertainty to the decisions of an exact block DFE employing successive cancelation was first proposed by this author in [11], by interpreting the uncertainty as an energy-constrained problem. Although the first estimate \( \hat{s}_0 \) could rely on a similar DFE, the resulting matrices turn out to depend on the Cholesky factors of the received vector covariance, due to the causal structure of \( \mathbf{B} \). As a consequence, the solution DFE matrices \( \{\mathbf{G}_0, \mathbf{B}_0\} \) are such that their efficient multiplications by vectors are not straightforward. On the other hand, given the prior \( \hat{s}_0 \), the structure of the feedback matrix \( \mathbf{B}_m \) at the \( m \)-th iteration can be chosen as a full matrix, except for null diagonal entries. Since our goal is to keep complexity to a minimum when implementing these receivers, we shall assume that a single iteration is sufficient to obtain a reliable decision. For this reason, we set \( \rho_1 = 1 \), in which case the solutions \( \{\mathbf{G}_1, \mathbf{B}_1\} \) are given by the following theorem.

**Theorem 2 (MMSE-BI-GDFE):** Consider the channel model after IBI removal (33), and let \( \hat{s}_c \) be the MMSE or LS estimate obtained through the superfast receiver of Fig. 9. Let \( \Delta \triangleq \text{Diag}(\mathbf{A}^\dagger \mathbf{H}_0^* \mathbf{R}_v^{-1} \mathbf{H}_0 \mathbf{A}) \). Then

\[
\begin{cases}
\mathbf{G}_1 & = (\sigma^2 I + \Delta)^{-1} \mathbf{A}^\dagger \mathbf{H}_0^* \mathbf{R}_v^{-1} \\
\mathbf{B}_1 & = (\sigma^{-2} I + \Delta)^{-1} \mathbf{A}^\dagger \mathbf{H}_0^* \mathbf{R}_v^{-1} \mathbf{H}_0 \mathbf{A} + \sigma^{-2} I - I
\end{cases}
\]  

(35)

**Proof:** Set \( \rho_1 = 1 \) in (A-5) in the Appendix.

The above result reduces to a minimum-variance-unbiased (MVUE) solution if we regard \( \mathbf{s} \) as deterministic, which implies that \( \sigma^2 I \rightarrow 0 \). In this case, for \( \mathbf{R}_v = \sigma^2 I \), we obtain \( \mathbf{G}_1 = \Delta^{-1} \mathbf{A}^\dagger \mathbf{H}_0^* \mathbf{R}_v^{-1} \mathbf{H}_0 \mathbf{A} \), and \( \mathbf{B}_1 = \Delta^{-1} \mathbf{A}^\dagger \mathbf{H}_0^* \mathbf{R}_v^{-1} \mathbf{H}_0 \mathbf{A} - I \), for \( \Delta \triangleq \text{Diag}(\mathbf{A}^\dagger \mathbf{H}_0^* \mathbf{R}_v \mathbf{A}) \). Moreover, it admits a MC or SC-FD implementation in connection to the superfast implementation.
of Fig. 9, depending on the choice of \( \mathbf{A} \). Either one can be realized by embedding \( \mathbf{H}_0 \), appearing in both \( \mathbf{G}_1 \) and \( \mathbf{B}_1 \), into a larger circulant matrix, where transmission is easily accomplished by fast FFT convolution; When \( \mathbf{A} = \mathbf{I} \), the diagonal \( \Delta \) contains the partial energies of the channel vector, and therefore only \( M \) multiplications are necessary in order to calculate it. When \( \mathbf{A} = \mathbf{F} \), obtaining the diagonal elements requires: (i) Computing \( \mathbf{M} + \delta \) steps of a sliding-window FFT of size \( M \); (ii) computing the energies of the \( M \) columns of the resulting matrix \( \mathbf{H}_0 \).

### IV. FAST AND SUPERFAST EQUALIZER COMPUTATION

In [6], [14], I have explicitly shown the connection between the generators of a structured matrix and the EGSWFTF recursions. Although this was presented in the context of adaptive channel estimation, where a structured matrix contains the transmitted data, we can similarly make use of these fast recursions in order to compute the generators of structured channel matrices, and according to arbitrary basis representations. To this end, once the channel is estimated, all we need is to feed the corresponding channel samples to the desired filter realization that by itself implements \( \mathbf{B}_Q \).

The complexity of obtaining the equalizer parameter in the ZF case, is even simpler; The lower triangular structure of \( \mathbf{H}_0 \) implies that the backward prediction section in the FTF algorithm is no longer necessary, since \( \mathbf{w}_n^k = 0 \) for all \( n \), and therefore, (7)–(13) are eliminated. The overall DFE realization requires 9 FFTs, each with complexity roughly of \( \text{op} \) operations per block transmitted; The equalizer parameters require 6 FFTs whenever the channel is re-estimated.

Note that the structure of the channel estimator depends on how its corresponding linear model is written. If the transceiver architecture allows us to express these models independently, we can opt to estimate the channel impulse response before any IBI or ISI removal. Consider for example the former case where \( \mathbf{y} = \mathbf{H}_0 \mathbf{t} + \mathbf{v} \), with training vector \( \mathbf{t} \). By rewriting it equivalently as \( \mathbf{y} = \mathbf{T} \mathbf{h} + \mathbf{v} \), with the vector of channel coefficients \( \mathbf{h} = [h(1) \cdots h(L)]^T \), the matrix \( \mathbf{T} \) becomes Toeplitz-like, similarly to \( \mathbf{H}_0 \). Hence, we readily verify that the LS estimate \( \hat{\mathbf{h}} - (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \mathbf{y} \) is such that the covariance \( (\mathbf{T}^T \mathbf{T})^{-1} \) admits a superfast representation, just like \( \mathbf{P}_{M} \) in (22). The main advantage here is that the transform domain generators are easily computed offline.
V. SIMULATION RESULTS

In this section, we compare the performance of standard MC and SC-FD schemes with the ones of the corresponding DFT-based minimum and optimal linear and DFE reduced redundancy schemes for randomly generated and fixed channels. For the sake of comparison with the experimental results of [18]–[21], we assume exact CSI, and also considered a minimum variance diagonal power-loading for the MC-MR transceiver.

Experiment 1 (Linear MR power-loading × Optimal Redundancy and Standard MC and SC-FD): In order to illustrate the direct relation between the channel zeros and the choice for the optimal redundancy in a MC setup, we transmitted blocks of QAM-4 symbols of size $M = 32$ through a $L = 15$-tap channel given by $H(z) = (0.77 + 0.38j) + 0.58j z^{-5} - 0.58z^{-3} - 0.567z^{-10} + 2.7z^{-13} + 0.4z^{-14}$. The zeros plot for this channel is illustrated in Fig. 10(left). Since it contains 13 maximum-phase zeros, the optimal redundancy in this case is $\delta = 13$. Fig. 10(right) shows the BER as a function of the channel SNR for the ZF and MMSE with minimum (ZF-Min. Red, MMSE-Min. Red) and optimal (Min. Norm-ZF-opt. red, Min. norm-MMSE-opt. red) redundancies, which are compared with the standard OFDM schemes, denoted as ZF-OFDM and MMSE-OFDM. In this example, both ZF and MMSE curves appear on top of each other. Observe that power loading in the MSE sense yields no advantage compared with the standard OFDM with equally distributed power. The minimum-redundancy scheme is far from optimality, and provides meaningless BER. Also, note that $M \approx 2L$, showing that the conclusions of [18] regarding the fact that MR schemes would be beneficial, for channel and blocks of the same order, are misleading. Choosing $L_{\text{opt}} = 13$ on the other hand, yields outstanding performance.

Secondly, we replaced the fixed tap gains by randomly generated values. We have considered a fixed delay path profile for 10000 ensemble generated gains, where both real and imaginary parts are independently drawn from a white Gaussian process. The histogram for the maximum-phase zeros of the corresponding channels is shown in Fig. 11(a). That is, on average, the optimal redundancy for a model with a fixed delay path profile is $\delta = 9$. This is a more realistic situation which shows again that a minimum redundancy choice is not optimal, even when power-loading is applied, exhibiting a worse BER performance compared to standard OFDM transmissions. This is verified in Fig. 11(b). The optimally chosen reduced redundancy scheme yields again the best results.

Experiment 2 (Random path delays, longer channels): Fig. 12 illustrates the BER curves for the case when 9 delay paths are randomly located within $L = 41$-tap channels (ensemble of 50000) for MC and SC-FD schemes, with $M = 64$. 

Fig. 10. (Left) Zeros of the fixed channel of Example 1; (Right) BER for the MC scheme.

Fig. 11. (a) (\# channels) x (\# maximum-phase zeros) histogram; (b) Resulting BER for $M = 32$, $L = 15$. 
This is more reasonable in practice than considering all taps random, given known wireless communication standards. For example, the channel models for the Long Term Evolution (LTE) standard specify either 7 or 9 tap gains within a long impulse response. Here we pose an average, worse case scenario, and the optimally chosen redundancy MC system still reaches the same performance as standard schemes, with the advantage of providing higher throughput; both MC and the corresponding optimal-redundancy SC-FD equalizers outperform the MR transceivers. We further include full \((L - 1)\) redundancy transmissions for comparison. Again, observe that \(M\) and \(L\) are of the same order, showing that the conclusions of [18] are misleading.\(^5\)

\[\text{\textbullet\hspace{1cm}}\text{Experiment 3 (DFE-IBI removal and SF-Linear receiver):}\] Here, we compare the performance of standard OFDM and SC-FD schemes with the linear and MC-DFE and SC-FD DFT-based minimum and optimal reduced redundancy schemes. First, we have transmitted blocks of QAM-4 symbols of size \(M = 32\) through the fixed channel \(H(z) = [0.77 + 2.38j] + 1.58jz^{-8} - 0.358z^{-9} - 0.567jz^{-10} + 0.5z^{-13} + 0.1z^{-14}\), which in this case has a single maximum-phased root. This example was particularly chosen with some zeros close to the unit circle, in order to characterize the performance of cyclic prefix based schemes. The zeros location is illustrated in Fig. 13.

For this channel, in theory, the optimal redundancy for the minimum-norm ZF receiver is \(\epsilon_o = 1\), which can only be achieved with a DFE receiver. Observe that the MR in the linear case is equal to \((L - 1)/2 = 7\), which is far from optimal. In order to implement a (linear) minimum-norm ZF equalizer, we are thus required to transmit \(L - 2 = 13\) redundant samples. Using this same value to build optimal MMSE receivers, we verify from Fig. 14 that their corresponding BER is much worse compared to the ones of OFDM-DFE and SC-DFE transceivers. The standard ZF and MMSE-OFDM transmissions outperform any linear, MR schemes, and the optimal MMSE linear receiver constructed with the optimal redundancy value is, however, significantly better than the remaining linear schemes. Similar conclusion can be drawn for the SC-FD configurations, except that the ZF-SC is the most degraded due to nearness of the channel zeros to the unit circle. Observe that both optimal and zero-redundancy block DFE receivers present the best BER performance. Further improvement in BER is only achieved with increasing redundant samples, as can be seen in the case of the MMSE-SC-FD, employing \(L - 1\) padded zeros.

As in the previous examples, we transmitted blocks of \(M = 32\) through \(10^6\) ensemble generated channels of length \(L = 29\), except that this time we have randomly selected 9 nonzero tap gains within the impulse response, as white Gaussian variables. At every realization, we employed the optimal redundancy for

\[\text{\textbullet\hspace{1cm}}\text{Minimum-redundancy schemes rely on square transmission matrices, for which their left-inverses do not exhibit a null-space. As a result, the use of these schemes does not allow the designer to select the right delay corresponding to a given channel, and consequently, square channel matrices can drastically amplify the channel noise depending on the CIR. This was reinterpreted here with the purpose of setting the best reduced redundancy for ZF transmission. The authors in [18]–[21], observe that the performance of their MR scheme will depend on the CSI, but provide no analytical explanation to this fact. For instance, it is mentioned that when the block size and the channel length are of the same order, MR schemes are preferred over standard OFDM and SC systems. This is a misleading statement. As seen in the simulations, we ran several experiments where the channel and block data sizes are actually of the same order, and still the performance of the MR scheme is the worst among all standard schemes. The explanation is rather simple: The minimum \((L - 1)/L\) redundancy is optimal (in a minimum-norm-ZF sense) if it equals the number of stable zeros of the channel impulse response, considering the optimal redundancy result discussed. The longer the block size compared to the channel, the taller is the full convolution matrix, and the smaller the probability that its optimal subblock corresponds to the one of a MR system.}\]
the DFE and linear receivers. We also considered a DFE scheme with the same minimum redundancy of the linear case. As we can see in Fig. 15, the former outperforms all other schemes by far, except in the SC-FD case, which otherwise make use of full \( L - 1 \) cyclic redundancy (in this case DF is not necessary). We remark that although this utilizes maximum redundancy for IBI cancelation, the doubly windowed structure of \( \mathbf{H}_0 \) yields a much simpler receiver requiring 2 superfast branches, and the best BER performance. Any MMSE or LS reduced redundancy scheme would require an additional receive branch. A ZF receiver, on the other hand could use smaller redundancy with 2 receive branches only. Another important conclusion is that the standard OFDM and SC-FD systems always outperform the linear MR systems, even for \( M \) and \( L \) of the same order.

**Experiment 4 (DFE-IBI removal and BF-GDFE):** Finally, we repeat the experiment of Fig. 15, however, employing one symbol re-estimation via the matched receiver and feedback matrices \([\mathbf{G}_1, \mathbf{B}_1]\) of (A-7). As can be seen in Fig. 16, a single iteration is sufficient to improve the performance of \((L - 1)/2\)-redundancy and optimal redundancy systems significantly, when compared to the linear and standard MC and SC-FD transceivers that makes use of full \((L - 1)\)-redundancy.

**Remark (LTE channel):** We have considered the Extended Pedestrian A (EPA) LTE channel model during the experiments as well. The BER curves in this case shows a more favorable behavior when compared to an average of Gaussian channels, but were not included at this point, due to space limitation. The transmitted block size was set to \( M = 64 \), while the channel, \( L = 21 \) coefficients. It is possible to verify that the LTE channel has approximately 5–6 zeros outside the unit circle, which shows that approximately \( h = 5 \) or \( k = 6 \) redundancies are optimal, according to our discussions. This redundancy cannot be employed by a linear scheme via ZP-ZJ, which can only reduce redundancy down to the minimum of \( \delta = (L - 1)/2 = 10. \) In addition, we could verify that the MR schemes present constants BERs of 0.4 for a ZF receiver at any SNR, and \( 2 \cdot 10^{-3} \) for an SNR \( > 15 \) dB. In other words, linear MR systems do not work. Moreover, it is also possible to verify that the performance of a transceiver where IBI is removed by ZJ is significantly inferior to the one that removes IBI via decision feedback.

VI. CONCLUSION

In this work, we have applied the framework of fast algorithms and related superfast covariance decompositions of [6]
(A-4)

\[ \begin{align*}
\sigma_s^2 \left[ (1 - \rho_m^2) \mathbf{H}_0^* - \rho_m \mathbf{I} \right] - \left[ \mathbf{G}_m (\mathbf{R}_e + (1 - \rho_m^2) \mathbf{H}_0 \mathbf{H}_0^*) \sigma_s^2 \left( \mathbf{B}_m - \rho_m \mathbf{G}_m \mathbf{H}_0 \right) \right] - \left[ \rho_m \mathbf{\Sigma} \mathbf{H}_0^* \mathbf{\Sigma} \right]
\end{align*} \]

Fig. 16. Random delay Paths; BI-GDFE run with one re-estimation: (Left) MC-BI-GDFE and (Right) SC-FD-BI-GDFE; \( M = 32, L = 29 \).

to block memoryless receivers, showing that efficient representations of reduced redundancy systems naturally arise. The parameters of these decompositions are presented in exact correspondence to fast Kalman variables even in the case of extended models, and therefore admit efficient exact computations as well.

Today, a great deal of research papers are commonly written towards an end in themselves, for reasons sometimes more related to publication strategies, rather than commitment with quality. The articles \([18]–[21]\) present several claims, including originality on the underlying theory, which can be contested. These were initially exposed in this presentation; however, due to lack of space, and as suggested by one of the reviewers of this paper, they will be addressed elsewhere, as a separate publication.

While overwhelming attention was brought on superfast minimum-redundancy systems recently, the claim that they work well in the only case where the channel and transmitted blocks are of the same order is false, as we could verify from our simulations. Also the claim that power-loading can combat ill-conditioning in this context is misleading.

On the other hand, we have shown how reduced redundancy must be chosen in a ZF scenario in order to minimize the output noise power, for symbol estimation. This results in significant improvement over minimum redundancy schemes, without the need for numerical optimization techniques, and employing a 2-branch superfast receiver. Moreover, we have shown that under DF, not only smaller redundancy transmissions are allowed compared to the minimum one in the linear case, but also that such block DFE scheme is implemented with superfast efficiency. In this context, the proposed MMSE BI-GDFE scheme shows that a single re-estimation can be implemented with superfast efficiency as well, and improves the prior estimate significantly. In all cases, decompositions can be performed according to alternative basis, yielding trigonometric based transforms. Its use in the MC setup is however subject of future work, since proper care must be taken when power loading the transmission.

APPENDIX

Defining \( \mathbf{W} \triangleq \begin{bmatrix} \mathbf{G} & \mathbf{B} \end{bmatrix} \) the minimization (34) at the \( m \)-th iteration can be written as

\[ \min_{\mathbf{W}} E \left[ \mathbf{w} \right] = \begin{bmatrix} \mathbf{G} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{\hat{s}}_{m-1} \end{bmatrix} \right]^2 \]  

(A-1)

The solution \( \mathbf{W}_m = [\mathbf{G}_m \ \mathbf{B}_m] \) must satisfy the normal equations \( \mathbf{R}_{\mathbf{w} \mathbf{w}} - \mathbf{W}_m \mathbf{R}_e \mathbf{R}_e^{-1} = [0 \ \mathbf{\Sigma}] \), where \( \mathbf{\Sigma} \) is some diagonal matrix to be determined. Using the linear model, under the approximation that the entries of \( \mathbf{s} \) are i.i.d., with equal energy \( \sigma_s^2 \) \([24]\), and the fact that \( \mathbb{E} \mathbf{s} \mathbf{s}^T \mathbf{m} = \rho_m \sigma_s^2 \mathbf{I} \), we have

\[ \begin{align*}
\sigma_s^2 \mathbf{H}_0^* - \rho_m \sigma_s^2 \mathbf{I} & \quad - \begin{bmatrix} [\mathbf{G}_m \ \mathbf{B}_m] \\

\mathbf{R}_e + \sigma_s^2 \mathbf{H}_0 \mathbf{H}_0^* & - \rho_m \sigma_s^2 \mathbf{H}_0 \\

- \rho_m \sigma_s^2 \mathbf{H}_0 \mathbf{H}_0^* & \sigma_s^2 \mathbf{I} \end{bmatrix} \begin{bmatrix} 0 \\

[0 \ \mathbf{\Sigma}] \end{bmatrix}
\end{align*} \]

(A-2)

Using a standard block factorization, yields

\[ \begin{align*}
\sigma_s^2 \mathbf{H}_0^* - \rho_m \mathbf{I} & \quad - \begin{bmatrix} [\mathbf{G}_m \ \mathbf{B}_m] \\

\mathbf{R}_e + (1 - \rho_m^2) \sigma_s^2 \mathbf{H}_0 \mathbf{H}_0^* \\

0 \end{bmatrix} \begin{bmatrix} \mathbf{I} \\

0 \end{bmatrix} \begin{bmatrix} \mathbf{I} \\

0 \end{bmatrix} \\

- \rho_m \mathbf{H}_0 \mathbf{H}_0^* & \sigma_s^2 \mathbf{I} \begin{bmatrix} 0 \\

[0 \ \mathbf{\Sigma}] \end{bmatrix}
\end{align*} \]

(A-3)

so that [see (A-4) at the top of the page], which gives [see (A-5) at the top of the next page]. For compactness of notation, let \( \alpha = \sigma_s^2 (1 - \rho_m^2) \). Substituting the expression of \( \mathbf{G}_m \) into \( \mathbf{B}_m \), and using the matrix inversion lemma, we get

\[ \begin{align*}
\mathbf{B}_m + \rho_m \mathbf{I} + \sigma_s^{-2} \mathbf{\Sigma} = \rho_m [\alpha \mathbf{I} - \rho_m \mathbf{\Sigma}] \\
\times \left[ \alpha^{-1} \mathbf{I} - \alpha^{-2} (\alpha^{-1} \mathbf{I} + \mathbf{H}^* \mathbf{R}_e \mathbf{H})^{-1} \right]
\end{align*} \]
\[ \begin{align*}
G_m &= \left[ \sigma_n^2 \left(1 - \rho_m^2 \right) I - \rho_m \Sigma \right] H_0^* (R_v + \left(1 - \rho_m^2 \right) \sigma_n^2 H_0 H_0^*)^{-1} \\
B_m &= \rho_m G_m H_0 - \rho_m I - \alpha_s^{-2} \Sigma
\end{align*} \]

(A-5)

Defining \( \tilde{\Sigma} \triangleq \text{Diag}(\alpha^{-1} I + H^* R_v H) \) \(-1 \), we conclude that

\[ \Sigma = \rho_m \alpha^{-1} \tilde{\Sigma} \left( \left( \sigma_n^2 + \rho_m \alpha^{-1} \right) I - \rho_m \alpha^{-2} \tilde{\Sigma} \right)^{-1} \]

(A-6)

so that [see (A-7) at the top of the page].

REFERENCES


Ricardo Merched received the B.S. and the M.Sc. degrees from the Federal University of Rio de Janeiro (UFRJ), Brazil, and the Ph.D. degree from the University of California, Los Angeles (UCLA), in 2001. He became Professor with the Department of Electrical and Computer Engineering, UFRJ, in 2002. He was a visiting professor with the University of California, Irvine, and Unik, University Graduate Center in Oslo, during 2006 and 2007. His current main interests include adaptive filtering algorithms, multirate systems, efficient digital signal processing techniques for MIMO equalizer architectures in wireless and wireline communications, RADAR imaging, and biomedical imaging applications, as well as low complexity solutions for these problems. He holds six US patents on efficient digital signal processing algorithms for channel estimation and equalization. Dr. Merched was an Associate Editor of the IEEE TRANSACTIONS ON Circuits and Systems I, the IEEE TRANSACTIONS ON SIGNAL PROCESSING LETTERS, and the EURASIP European Journal on Advances in Signal Processing.