Theoretical Performance Evaluation and Optimization of UW-OFDM

Heidi Steendam, *Senior Member, IEEE*

Abstract—Unique-word (UW) OFDM is a new OFDM variant that offers several advantages over cyclic prefix OFDM, such as better bit-error-rate (BER) performance in frequency selective channels and much lower out-of-band radiation. However, all results with respect to the BER performance are based on Monte-Carlo simulations only, and offer little insight in the behaviour of the system performance as function of its parameters. In this paper, we theoretically analyse the performance of the UW-OFDM system for both a deterministic channel and a fading channel, and offer a systematic way to generate the UW-OFDM signal to optimize the performance of the system. When the fading channel is known at the transmitter, we show that we can obtain a diversity order that is of the same order of magnitude as the theoretically maximum diversity order, but when the channel is not known at the transmitter, it will be very hard to achieve maximum diversity order, although the obtained diversity order is larger than one.

Index Terms—Unique-word OFDM, diversity, bit error rate, Euclidean distance, performance optimization, fading channel.

I. INTRODUCTION

MULTICARRIER communication has become an important transmission technique for frequency selective channels, witness the large number of standards adopting one of the multicarrier communication variants. The most widely spread of these variants is the cyclic prefix orthogonal frequency division multiplexing (CP-OFDM) technique [1]. However, the use of the cyclic prefix as a guard interval to avoid intercarrier and intersymbol interference involves some drawbacks. Firstly, the extension of the symbol length by inserting a cyclic prefix introduces an intolerable amount of out-of-band radiation, which can only be suppressed with a considerable amount of effort. Secondly, the randomness of the cyclic prefix – as its content is determined by the transmitted data – implies that the cyclic prefix cannot be used efficiently for particular needs such as synchronization and/or

system parameter estimation. Thirdly, as the cyclic prefix is thrown away at the receiver in order to detect the transmitted data symbols, channel diversity will be lost unless additional diversification of the transmitted signals is incorporated such as precoding or channel coding.

Recently, a new variant of OFDM has been introduced, i.e., unique-word (UW) OFDM [2]. In this technique, the guard interval is part of the time-domain interval corresponding to the inverse discrete Fourier transform (DFT) output, i.e., the presence of the guard interval does not extend the symbol duration, and contains a known sequence (the unique word) of length N_u, implying the last N_u samples of the time-domain sequence at the output of the inverse DFT are known and do not depend on the transmitted data. As a result, the content of the UW can be used for synchronization and parameter estimation. In order to allow for the insertion of a known sequence within the DFT interval, redundancy must be added in the frequency domain by means of a generator matrix. Because of this redundancy, the authors in [3] have shown that UW-OFDM outperforms CP-OFDM with respect to the bit error rate (BER) performance in frequency selective channels. Moreover, the addition of the redundancy has no adverse effect on the data throughput: both UW-OFDM and CP-OFDM have essentially the same throughput efficiency. Furthermore, in [4], it is shown that the out-of-band radiation is considerably lower in UW-OFDM than in CP-OFDM, and by sacrificing a few carriers [5], this out-of-band radiation can be made essentially zero. Consequently, UW-OFDM is highly spectrally efficient and is therefore a viable solution for e.g. cognitive radio applications.

Although the authors in [2], [3], [6], [7] considered the BER performance evaluation of UW-OFDM, their results were based on Monte-Carlo simulations only. However, simulations do not allow a deep understanding of the effect of the system parameters on the BER performance, and the optimization of the system parameters using simulations only is practically infeasible. Therefore, in this paper, we theoretically investigate the BER performance of UW-OFDM in (deterministic and random) frequency selective channels. Based on this theoretical analysis, we are able to identify the constraints required to achieve optimal BER performance, and using these constraints, we propose a systematic construction method to obtain the optimal UW-OFDM signal structure. Our results explain the simulation results from [2], [3], [6], [7], and show that for Rayleigh fading channels, the UW-OFDM system inherently has a higher diversity order than standard[1] CP-OFDM.

[1]Standard in this context means without additional precoding or channel coding.

The paper is organized as follows. In Section II, we describe a construction method for the UW-OFDM generator matrix that automatically incorporates the constraints of the UW-OFDM signal, and still offers a large amount of freedom to optimize the signal. Further, we inspect the expression for BER performance and derive a necessary condition for the generator matrix to achieve maximum diversity order. Based on the results of Section II, Section III explores the optimization of the error rate and derives the constraints for the generator matrix. Taking into account the conditions derived in Section III, Section IV proposes a systematic construction method to obtain optimal performance. Numerical results are given in Section V and Section VI concludes the paper.

II. UW-OFDM BIT ERROR RATE PERFORMANCE

A. System Description

In the following, we use the two-step approach from [3] to construct the UW-OFDM signal, where first the time-domain sequence corresponding to the data contribution is generated and the inverse DFT output contains zeros at the positions of the unique word, and afterwards the samples from the unique word are added. This method requires a lower transmitted energy than the direct construction method, where the unique word samples are generated directly at the inverse DFT output [3]. Assuming the unique word length equals N_u samples, the time-domain signal at the inverse DFT output, corresponding to the data part must contain N_u zeros at the end. The redundancy that needs to be added in the frequency domain in order to obtain the block of zeros in the time domain, limits the number of data symbols that can be transmitted: the number N_r of redundant symbols is lower bounded by N_u, i.e. $N_r \geq N_u$. As a result, assuming the DFT size equals N, the maximum number of data symbols equals $N - N_r \leq N - N_u$. Because of the presence of guard bands, or the introduction of additional redundancy in the frequency domain in order to reduce the out-of-band radiation or peak-to-average power [2], [5], [8], the number of data symbols is generally smaller than this maximum $N - N_r$. Let us assume $N_m \leq N$ carriers are used. In that case, the number of data symbols equals $N_d = N_m - N_r \leq N - N_u$. The frequency domain symbols that are modulated on the carriers can be expressed as

$$\mathbf{z} = \mathbf{BGx}_d \qquad (1)$$

where $\mathbf{x}_d$ contains the N_d data symbols, the $N_m \times N_d$ generator matrix $\mathbf{G}$ introduces the redundancy in the frequency domain, and the $N \times N_m$ mapping matrix $\mathbf{B}$ is a reduced version of the identity matrix, where the columns corresponding to the unmodulated carriers[2] are deleted, and which maps the N_m modulated carriers on the inverse DFT inputs. The frequency domain vector $\mathbf{z}$ is fed to the inverse DFT, resulting in the time-domain samples

$$\mathbf{y} = \mathbf{F}_N^H \mathbf{z} = \begin{pmatrix} * \\ \mathbf{0} \end{pmatrix} \qquad (2)$$

[2]This corresponds e.g. to the presence of guard bands or notch bands. For example, if g carriers at each side of the frequency band are not modulated as a guard band, the number of modulated carriers equals $N_m = N - 2g$.

where $\mathbf{F}_N^H$ is the inverse DFT matrix with $(\mathbf{F}_N)_{k,\ell} = \frac{1}{\sqrt{N}} e^{-j2\pi \frac{k\ell}{N}}$. Taking into account that the mapping matrix $\mathbf{B}$ is determined by the spectral band requirements of the system, the only freedom lies in the selection of the generator matrix $\mathbf{G}$. Let us define $\tilde{\mathbf{F}}$ as the $N_u \times N_m$ matrix containing the last N_u rows of $\mathbf{F}_N^H \mathbf{B}$. Imposing that the last N_u elements of $\mathbf{y}$ must be zero implies $\tilde{\mathbf{F}}\mathbf{G} = \mathbf{0}$, i.e., the columns of $\mathbf{G}$ belong to the null space of $\tilde{\mathbf{F}}$.

Proposition II.1. The $N_u \times N_m$ matrix $\tilde{\mathbf{F}}$, with $N_u < N_m$, has full rank.

The proof of this proposition is given in Appendix A. Because of the rank-nullity theorem, it follows that $N_m - N_u$ linearly independent null vectors exist. Assume the $N_m \times 1$ vectors $\mathbf{u}_1, \ldots, \mathbf{u}_{N_m - N_u}$ form an orthonormal basis for this null space[3]. Hence, the columns of $\mathbf{G}$ must consist of linear combinations of these basis vectors. Denote $\mathbf{g}_k$ as the k-th column of $\mathbf{G}$:

$$\mathbf{g}_k = \sum_{\ell=1}^{N_m - N_u} w_{k,\ell} \mathbf{u}_\ell, \qquad k = 1, \ldots, N_d \qquad (3)$$

or in matrix notation

$$\mathbf{G} = \mathbf{UW} \qquad (4)$$

where $\mathbf{U} = (\mathbf{u}_1 \ldots \mathbf{u}_{N_m - N_u})$ is the $N_m \times (N_m - N_u)$ matrix containing the basis vectors so that $\tilde{\mathbf{F}}\mathbf{U} = \mathbf{0}$, and $\mathbf{W}$ the $(N_m - N_u) \times N_d$ matrix containing the weight coefficients $\mathbf{W}_{\ell,k} = w_{k,\ell}$; this matrix $\mathbf{W}$ can be selected freely, provided that $N_d \leq N_m - N_u$ in order to avoid ambiguity between distinct data sequences. Throughout this paper, we will use this decomposition of $\mathbf{G}$ to simplify the derivations.

The UW-OFDM time-domain signal is transmitted over the channel. We assume that this channel is modelled as a tapped delay channel with $L + 1$ taps, i.e.,

$$\mathbf{h} = [h(0) \ldots h(L)]^T. \qquad (5)$$

The channel adds white Gaussian noise $\mathbf{w}$ of which the components have variance $N_0/2$ per real dimension [9]. We assume the guard interval, i.e., the unique word length, exceeds the channel length: $N_u \geq L$, implying we can avoid intersymbol interference between successive OFDM symbols. At the receiver, we apply a DFT to the received time-domain samples, resulting in

$$\mathbf{r} = \mathbf{F}_N \mathbf{H} \mathbf{F}_N^H \mathbf{z} + \mathbf{F}_N \mathbf{w}. \qquad (6)$$

In this expression, we neglected the presence of the unique word, which can easily be subtracted from the signal if the channel is known. The channel matrix $\mathbf{H}$ in (6) is a circulant matrix. Hence, the matrix product $\tilde{\mathbf{H}} = \mathbf{F}_N \mathbf{H} \mathbf{F}_N^H$ is a diagonal matrix with as diagonal elements the frequency response of the channel:

$$\tilde{\mathbf{H}} = \text{diag}\left(\mathcal{H}\left(j2\pi \frac{k}{N} \right) \right) \qquad (7)$$

[3]Such a basis can easily be found using the singular value decomposition of $\tilde{\mathbf{F}}$.

with $\mathcal{H}\left(j2\pi\frac{k}{N}\right) = \sum_{\ell=0}^{L} h(\ell)\mathrm{e}^{j2\pi\frac{k\ell}{N}}$. Therefore, the received samples at the DFT outputs can be rewritten as

$$\mathbf{r} = \tilde{\mathbf{H}}\mathbf{z} + \mathbf{F}_N\mathbf{w} = \mathbf{v} + \mathbf{F}_N\mathbf{w}, \tag{8}$$

where $\mathbf{v} = \tilde{\mathbf{H}}\mathbf{B}\mathbf{G}\mathbf{x}_d$. We want to design $\mathbf{G}$ so that the error rate performance is optimized.

B. Euclidean Distance

In this paper, we will find that optimizing the error rate performance is equivalent to optimizing the minimum Euclidean distance. Depending on whether the channel is known or not, we will show that the minimum Euclidean distance at the receiver or the transmitter side has to be maximized. Let us look closer at the Euclidean distance between the code words at the transmitter and receiver side. At the receiver, the squared Euclidean distance between two code words $\mathbf{v} = \tilde{\mathbf{H}}\mathbf{B}\mathbf{G}\mathbf{x}_d$ and $\mathbf{v}' = \tilde{\mathbf{H}}\mathbf{B}\mathbf{G}\mathbf{x}'_d$ is given by

$$d^2(\mathbf{v}, \mathbf{v}') = \mathbf{e}^H\mathbf{G}^H\mathbf{B}^H\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\mathbf{B}\mathbf{G}\mathbf{e} \triangleq \mathbf{e}^H\mathbf{A}_R\mathbf{e}, \tag{9}$$

where $\mathbf{e} = \mathbf{x}_d - \mathbf{x}'_d$ is the error vector. The matrix $\mathbf{A}_R$ depends on the channel taps $\mathbf{h}$ through the diagonal matrix $\tilde{\mathbf{H}}$. The squared minimum Euclidean distance between the code words at the receiver is defined as

$$d^2_{\min, R} = \min_{\mathbf{e}} d^2(\mathbf{v}, \mathbf{v}') = \min_{\mathbf{e}} \mathbf{e}^H\mathbf{A}_R\mathbf{e}. \tag{10}$$

Similarly, the squared Euclidean distance between the code words at the transmitter equals

$$d^2(\mathbf{z}, \mathbf{z}') = \mathbf{e}^H\mathbf{G}^H\mathbf{B}^H\mathbf{B}\mathbf{G}\mathbf{e} \triangleq \mathbf{e}^H\mathbf{A}_T\mathbf{e}, \tag{11}$$

where $\mathbf{z}$ is defined in (1), and the squared minimum Euclidean distance at the transmitter:

$$d^2_{\min, T} = \min_{\mathbf{e}} d^2(\mathbf{z}, \mathbf{z}') = \min_{\mathbf{e}} \mathbf{e}^H\mathbf{A}_T\mathbf{e}. \tag{12}$$

Both $\mathbf{A}_R$ and $\mathbf{A}_T$ are positive semi-definite Hermitian $N_d \times N_d$ matrices. Further, they are related to the transmitted and received energy:

$$\begin{aligned} E_T &= E_s\mathrm{trace}(\mathbf{A}_T) \\ E_R &= E_s\mathrm{trace}(\mathbf{A}_R). \end{aligned} \tag{13}$$

where E_s is the energy per data symbol. In [6], it was shown that the MSEs of the BLUE and LMMSE data detectors in a frequency-flat channel are minimized when $\mathbf{A}_T = c\mathbf{I}_{N_d}$. Later in this paper, we will show that, if the channel is known, the generator matrix $\mathbf{G}$ must be selected so that $\mathbf{A}_R = c\mathbf{I}_{N_d}$. When $c = 1$, these conditions correspond to the normalization of the transmitted and received energy, respectively.

C. Theoretical Error Performance

In [10], the authors considered the ML detection of the transmitted data symbols in OFDM-based systems. Assuming the channel is perfectly known, the pairwise error probability (PEP) of a data vector $\mathbf{x}_d$ and the detected vector $\mathbf{x}'_d \neq \mathbf{x}_d$, given a channel realization $\mathbf{h}$, is given by $Pr(\mathbf{x}'_d \neq \mathbf{x}_d|\mathbf{h})$. Taking into account the Gaussian character of the noise, the PEP is given by

$$Pr(\mathbf{x}'_d \neq \mathbf{x}_d|\mathbf{h}) = Q\left(\sqrt{\frac{d^2(\mathbf{v}, \mathbf{v}')}{2N_0}}\right) \tag{14}$$

where $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^{+\infty}\mathrm{e}^{-t^2/2}dt$ and $d^2(\mathbf{v}, \mathbf{v}')$ is defined in (9). The authors in [10] used the Chernoff bound [11] to upper bound this PEP and derive the average of the upper bound over the channel characteristics assuming the channel taps are uncorrelated Rayleigh fading distributed with $E[\mathbf{h}\mathbf{h}^H] \sim \mathbf{I}_{L+1}$ in order to derive the diversity order and coding gain. Further they showed that the maximum achievable diversity order $\delta = L + 1$ can only be reached when the minimum Euclidean distance $d_{\min, R} \geq (L+1)\sqrt{E_s}$. A condition to satisfy $d_{\min, R} \geq (L+1)\sqrt{E_s}$ is that $\mathrm{rank}(\mathbf{BG}) = N_d$, and $N_d \geq L + 1$. However, this maximum diversity order can only be achieved when the transmitter contains sufficient redundancy so that the signals are sufficiently diversified and the receiver is designed to exploit this redundancy.

Let us look closer at the condition that the $N_m \times N_d$ matrix $\mathbf{BG}$ must have full rank. Consider the decomposition $\mathbf{G} = \mathbf{UW}$ (4). The matrix $\mathbf{B}$ has full rank N_m, with $N_m - N_u \geq N_d$. Further, also the $N_m \times (N_m - N_u)$ matrix $\mathbf{U}$ has full rank as its columns consist of orthonormal basis vectors. In order to be able to achieve maximum diversity, the $(N_m - N_u) \times N_d$ matrix $\mathbf{W}$, determining the redundancy, needs to have full rank N_d. Hence, the weight coefficients $w_{k,\ell}$ must be selected so that the N_d columns of $\mathbf{W}$ are linearly independent. As it is possible to make $\mathbf{W}$ full rank, theoretically the maximum diversity order can be reached in UW-OFDM, provided the matrix $\mathbf{G}$ offers sufficient diversification. Let us look at the rank of the generator matrix of two standard constructions of UW-OFDM, i.e., systematic and non-systematic UW-OFDM [12].

Example 1: Systematic UW-OFDM

Let us consider the case of systematic UW-OFDM. In that case, we can decompose the matrix $\mathbf{G}$ as [12]

$$\mathbf{G} = \beta\mathbf{P}\begin{bmatrix}\mathbf{I}_{\mathbf{N_d}} \\ \mathbf{T}\end{bmatrix}, \tag{15}$$

where $\mathbf{P}$ is a $N_m \times N_m$ permutation matrix determining the data and redundant carrier positions, and $\mathbf{T}$ is the linear transformation determining the relationship between the data and redundant symbols. The factor β is selected to normalize the transmitted energy: $\mathrm{trace}(\mathbf{G}^H\mathbf{B}^H\mathbf{B}\mathbf{G}) = N_d$.

Proposition II.2. The code generator matrix $\mathbf{G}$ from (15) has full rank whenever the energy transmitted on the redundant carriers is finite.

Proof: The matrix $\mathbf{P}$ has full rank N_m as $\mathbf{P}^H\mathbf{P} = \mathbf{I}_{N_m}$. To determine the rank of $\mathbf{G}$, we consider the Gramian matrix $\mathbf{G}^H\mathbf{G} = |\beta|^2(\mathbf{I}_{N_d} + \mathbf{T}^H\mathbf{T})$. When $\mathbf{I}_{N_d} + \mathbf{T}^H\mathbf{T}$ has no eigenvalues equal to zero, it has full rank N_d. In [13], it is shown

that in order to have finite redundant energy, the eigenvalues of $\mathbf{T}^H\mathbf{T}$ must be strictly positive. Hence, under this last condition, and taking into account that $N_d \leq N_m - N_u$, the matrix $\mathbf{G}$ will automatically have full rank, independent of the positions of the redundant carriers, so that for this type of UW-OFDM the condition to allow maximum diversity order is satisfied. ∎

Example 2: Non-Systematic UW-OFDM

Taking into account the decomposition (4) of the code generator matrix $\mathbf{G}$, where the matrix $\mathbf{W}$ can be selected freely, it is clear that a non-systematic UW-OFDM system has many degrees of freedom. In [6], the authors used this degree of freedom to optimize the matrix $\mathbf{G}$ so that the MSE of the BLUE (best linear unbiased) or LMMSE (linear minimum mean squared error) data detector in an AWGN channel was minimized. Through an iterative procedure and simulations, they obtained the constraint $\mathbf{G}^H\mathbf{B}^H\mathbf{B}\mathbf{G} = \alpha\mathbf{I}_{N_d}$. The analytical proof of this constraint is given in Appendix B.

Proposition II.3. The code generator matrix $\mathbf{G}$ minimizing the MSE of the BLUE or LMMSE data detector has full rank.

Proof: Taking into account the condition $\mathbf{G}^H\mathbf{B}^H\mathbf{B}\mathbf{G} = \alpha\mathbf{I}_{N_d}$ and the results from Appendix B, it follows that $\mathbf{G}$ has full rank, which concludes the proof. ∎

Based on these observations, it turns out that it is quite simple to find a matrix $\mathbf{G}$ that has full rank, implying the maximum diversity order in a fading channel can theoretically be achieved in UW-OFDM provided that the generator matrix has sufficient redundancy to counteract deep fades. In the remainder of the paper, we restrict our attention to the case where the generator matrix is full rank.

III. Minimizing the Error Rate

In the previous section, the pairwise error probability, which is related to the symbol error rate, was derived. When the channel is known, the error probability (14) is a function of the Euclidean distance between code words at the receiver, which indicates that the error probability is determined by the minimum Euclidean distance. Hence, to minimize the error rate, this minimum Euclidean distance needs to be maximized. When the channel is known at the transmitter, optimizing this Euclidean distance turns out to be realizable. However if the channel is not known at the transmitter, the minimum Euclidean distance at the receiver cannot be maximized. In that case, we will propose an alternative solution that offers good performance.

A. Channel Known

Let us look closer at the minimum Euclidean distance $d_{\min,R}$ (10) at the receiver. We assume that the error vector $\mathbf{e} \in \mathbb{C}^{N_d \times 1}$ is normalized, i.e., $\mathbf{e}^H\mathbf{e} = 1$. Let us consider the eigenvalue decomposition of the positive semi-definite Hermitian matrix $\mathbf{A}_R$: $\mathbf{A}_R = \mathbf{V}_R^H\mathbf{\Lambda}_R\mathbf{V}_R$, where $\mathbf{V}_R$ is an orthogonal matrix and the eigenvalues $\lambda_{R,\ell}$ are positive real-valued. These eigenvalues correspond to the different modes of the channel. Taking into account the expression for the PEP (14) and noting that

$\mathbf{e}$ and $\mathbf{V}_R\mathbf{e}$ have the same energy, the average PEP over the different modes of the channel given the channel $\mathbf{h}$ can be approximated by

$$PEP_{avg} = \frac{1}{N_d}\sum_{\ell=1}^{N_d} Q\left(\sqrt{\frac{\lambda_{R,\ell}d^2(\mathbf{x}_d,\mathbf{x}_d')}{2N_0}}\right). \quad (16)$$

Due to the steep decrease of $Q(\cdot)$ as function of its argument, the PEP will be dominated by the term with the smallest argument. Because $\mathbf{e}^H\mathbf{e} = 1$, $d^2(\mathbf{x}_d,\mathbf{x}_d') = 1$ indicating the average PEP is dominated by the smallest of the eigenvalues $\lambda_{R,\ell}$. To find the smallest of the eigenvalues $\lambda_{R,\ell}$, we revert to the Rayleigh quotient of a Hermitian matrix $\mathbf{A}_R$, i.e. $(\mathbf{e}^H\mathbf{A}_R\mathbf{e})/(\mathbf{e}^H\mathbf{e})$, which is bounded by the minimum and maximum eigenvalue of the matrix $\mathbf{A}_R$ [14]. Using this Rayleigh quotient, it is clear that the minimum eigenvalue equals the squared minimum Euclidean distance at the receiver (10). Therefore, to minimize the average PEP, we must make this squared minimum Euclidean distance as large as possible.

Proposition III.1. A sufficient condition for the squared minimum Euclidean distance (9) to be maximized is that the generator matrix $\mathbf{G}$ is selected so that

$$\mathbf{A}_R = \mathbf{G}^H\mathbf{B}^H\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\mathbf{B}\mathbf{G} = \mathbf{I}_{N_d}. \quad (17)$$

Proof: Using the properties of the Rayleigh quotient of the matrix $\mathbf{A}_R$, maximizing the squared minimum Euclidean distance is equivalent with the maximization of the minimum eigenvalue of $\mathbf{A}_R$. ∎

In order to maximize the minimum eigenvalue, we apply Gerschgorin's theorem [14]. Given a Hermitian matrix $\mathbf{A}_R$, this theorem states that the real-valued eigenvalues lie in the interval $(\mathbf{A}_R)_{k,k} - r_k \leq \lambda_k \leq (\mathbf{A}_R)_{k,k} + r_k$ where

$$r_k = \sum_{\substack{\ell=1 \\ \ell \neq k}}^{N_d} |(\mathbf{A}_R)_{k,\ell}|. \quad (18)$$

In the following, we assume that $(\mathbf{A}_R)_{k,k} = 1$, which corresponds to the normalization of the received energy per data symbol (13); this implies that all data symbols have the same error performance [15], [16], resulting in the lowest error rate performance if we average over the data symbols. Under this assumption, the sum of the eigenvalues $\lambda_{R,k}$ is constant: $\text{trace}(\mathbf{A}_R) = \sum_{k=1}^{N_d}\lambda_{R,k} = 1$, and the boundaries on the eigenvalues reduce to $1 - r_k \leq \lambda_{R,k} \leq 1 + r_k$. A sufficient condition to ensure that the smallest of the eigenvalues is as large as possible is to make the upper bound equal to the lower bound, i.e., $r_k = 0$, which implies that all eigenvalues are equal to one, inferring $\mathbf{A}_R$ is the identity matrix. ∎

The above proposition states that we have to select the code generator matrix $\mathbf{G}$ so that the matrix $\mathbf{A}_R$ (17) is the identity matrix. In that case, the average PEP (16) reduces to

$$PEP_{avg} = Q\left(\sqrt{\frac{d^2(\mathbf{x}_d,\mathbf{x}_d')}{2N_0}}\right). \quad (19)$$

This result indicates that we can make the error probability of each data symbol equal irrespective of the used channel mode. When the channel is known, corresponding e.g., with the case of a fixed wired channel or a wireless link with a slowly varying channel, where the channel is estimated and fed back to the transmitter, finding a generator matrix satisfying the above condition is straightforward. In Section IV, we will discuss a systematic construction method for the code generator matrix $\mathbf{G}$ if the channel is known.

B. Channel Unknown

When no channel state information is available at the transmitter side, selecting the code generator matrix that results in $\mathbf{A}_R = \mathbf{I}_{N_d}$ is not obvious. Therefore, we consider a different approach. Define the matrix $\mathbf{B}_e$ as $\mathbf{B}_e \mathbf{h} \triangleq \tilde{\mathbf{H}} \mathbf{B} \mathbf{G} (\mathbf{x}_d - \mathbf{x}'_d) = \tilde{\mathbf{H}} \mathbf{B} \mathbf{G} \mathbf{e}$, i.e., $(\mathbf{B}_e)_{k,\ell} = e^{j2\pi \frac{k\ell}{N}} (\mathbf{B} \mathbf{G} \mathbf{e})_k$, so that $d^2(\mathbf{v}, \mathbf{v}') = \mathbf{h}^H \mathbf{B}_e^H \mathbf{B}_e \mathbf{h}$. The matrix $\mathbf{B}_e^H \mathbf{B}_e$ is a positive semi-definite Hermitian Toeplitz matrix with elements

$$(\mathbf{B}_e^H \mathbf{B}_e)_{\ell,\ell'} = \sum_{k=0}^{N-1} e^{j2\pi \frac{k(\ell-\ell')}{N}} |(\mathbf{B} \mathbf{G} \mathbf{e})_k|^2 \triangleq a_{\ell-\ell'}, \quad (20)$$

and depends on the error vector $\mathbf{e}$. Assuming the matrix $\mathbf{B}_e^H \mathbf{B}_e$ has full rank, the error probability is determined by the coding gain, which is shown in [10] to be equal to

$$\gamma = \min_{\mathbf{e} \neq \mathbf{0}} \quad \alpha_L \left[\det(\mathbf{B}_e^H \mathbf{B}_e) \right]^{\frac{1}{L+1}}, \quad (21)$$

where $\alpha_L = E[\mathbf{h}^H \mathbf{h}]$ is the energy of the channel impulse response and $\det(\mathbf{X})$ is the determinant of the matrix $\mathbf{X}$. Hence, to compute the coding gain, the determinant of the matrix $\mathbf{B}_e^H \mathbf{B}_e$ has to be minimized over the error vector $\mathbf{e}$. To optimize the error rate performance, we will use the degrees of freedom available in the selection of the generator matrix $\mathbf{G}$ through the decomposition (4). In order to find the minimum (over the error vector $\mathbf{e}$) of $\det(\mathbf{B}_e^H \mathbf{B}_e)$, we rewrite the elements a_m (20):

$$a_m = \mathbf{e}^H \mathbf{G}^H \mathbf{B}^H \mathbf{X}_m \mathbf{B} \mathbf{G} \mathbf{e} \quad (22)$$

where $\mathbf{X}_m = \text{diag}(e^{j2\pi \frac{km}{N}})$, $k = 0, \ldots, N-1$. Although [17] offers a recursive analytical expression for the computation of the determinant of Toeplitz matrices, the resulting analytical expression turns out to be too complex to optimize the result over the error vector $\mathbf{e}$. Therefore, we consider the following upper and lower bounds on the determinant of a positive definite Hermitian $m \times m$ matrix $\mathbf{D}$:

$$\det(\mathbf{D}) \leq \left(\frac{\text{trace}(\mathbf{D})}{m} \right)^m \quad (23)$$

and

$$(\lambda_{\min})^m \leq \det(\mathbf{D}) \leq (\lambda_{\max})^m \quad (24)$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are the minimum and maximum eigenvalue of the matrix $\mathbf{D}$. The first inequality (23) is a direct consequence of the arithmetic mean – geometric mean (AM-GM) inequality [18], and corresponds to Hadamard's inequality [19]. Equality occurs when all eigenvalues of $\mathbf{D}$ are equal. The inequalities in (24) result from the fact that the determinant is the product of the eigenvalues, which are all positive real values as the matrix $\mathbf{D}$ is Hermitian positive definite. Again, equality occurs when all eigenvalues are equal.

We will now apply the inequalities (23) and (24) to find the optimal generator matrix $\mathbf{G}$ that minimizes the error rate. Let us take a closer look at the inequality (23) for the matrix $\mathbf{B}_e^H \mathbf{B}_e$:

$$\det(\mathbf{B}_e^H \mathbf{B}_e) \leq \left(\frac{\text{trace}(\mathbf{B}_e^H \mathbf{B}_e)}{L+1} \right)^{L+1} = a_0^{L+1}. \quad (25)$$

Bearing in mind that $a_0 = \mathbf{e}^H \mathbf{A}_T \mathbf{e} = \mathbf{e}^H \mathbf{G}^H \mathbf{G} \mathbf{e}$, it follows that the upper bound still is a function of the error vector. Further, taking into account the squared Euclidean distance at the transmitter (11), $d^2(\mathbf{z}, \mathbf{z}') = a_0$, it follows that the upper bound in (25) is a function of this squared Euclidean distance. Minimizing the upper bound (25) over the error vector $\mathbf{e}$, in order to obtain an upper bound on the coding gain γ (21), and assuming $\mathbf{e}^H \mathbf{e} = 1$, the smallest value of a_0, and thus of the upper bound, equals the smallest eigenvalue of $\mathbf{G}^H \mathbf{G}$. Now we select the code generator matrix $\mathbf{G}$ to maximize the minimum upper bound. This corresponds to maximizing the minimum eigenvalue of $\mathbf{G}^H \mathbf{G}$, and results in the maximization of the minimum Euclidean distance at the transmitter. Taking into account the proof of Proposition III.1, this implies that all eigenvalues should be equal, or equivalently, $\mathbf{G}^H \mathbf{G} \sim \mathbf{I}_{N_d}$.

The above condition $\mathbf{G}^H \mathbf{G} \sim \mathbf{I}_{N_d}$ still leaves us several degrees of freedom to select the code generator matrix. Therefore, we consider the inequalities (24). The determinant of $\mathbf{B}_e^H \mathbf{B}_e$ is lower bounded by the $(L+1)$th power of its smallest eigenvalue $\lambda_{e,\min}$. Let us consider the worst case error vector $\mathbf{e}$ resulting in the smallest minimum eigenvalue $\lambda_{e,\min}$. We have to select $\mathbf{G}$ so that $\lambda_{e,\min}$ is maximized, taking into account that $\sum_{\ell=0}^{L} \lambda_{e,\ell} = \text{trace}(\mathbf{B}_e^H \mathbf{B}_e) = (L+1)a_0$. Hence, the eigenvalues are in the range $0 < \lambda_{e,\ell} < (L+1)a_0$. If the eigenvalues can take all values within this interval, we can easily prove through Lagrange optimization that a sufficient condition to maximize the minimum eigenvalue corresponds to the case where all eigenvalues are the same, i.e., $\lambda_{e,\ell} = a_0, \forall \ell$ (see the proof of Proposition III.1). This results in the condition $\mathbf{B}_e^H \mathbf{B}_e = a_0 \mathbf{I}_{L+1}$, implying the inequalities in (24) and (25) become equalities. However, we have to check if this condition is realizable in practice: is it possible to find a matrix $\mathbf{G}$ so that $\lambda_{e,\ell} = a_0, \forall \ell$, i.e., so that the matrix $\mathbf{B}_e^H \mathbf{B}_e$ is diagonal for the worst case $\mathbf{e}$ and thus also for all other error vectors $\mathbf{e}$? Unfortunately, according to the following Proposition, this is not possible in general.

Proposition III.2. The matrix $\mathbf{B}_e^H \mathbf{B}_e$ can be made diagonal for all possible $\mathbf{e}$ if $N_d \leq N/2$ but not when $N_d > N/2$, when $\mathbf{G}$ is assumed to have full rank.

Proof: See appendix C $\qquad\blacksquare$

In practice, to obtain a high data throughput, the number of data symbols transmitted per UW-OFDM block should be as large as possible, implying $N_d > N/2$. For these high data throughputs, the matrix $\mathbf{B}_e^H \mathbf{B}_e$ cannot be made diagonal for all

error vectors $\mathbf{e}$. If we cannot make the matrix $\mathbf{B}_e^H \mathbf{B}_e$ diagonal, the next best thing would be to select the matrix $\mathbf{G}$ so that the eigenvalues $\lambda_{e,\ell}$ of $\mathbf{B}_e^H \mathbf{B}_e$ are as close as possible to each other, i.e., the difference $\lambda_{e,\max} - \lambda_{e,\min}$ must be as small as possible. In other words, the resulting matrix $\mathbf{B}_e^H \mathbf{B}_e$ must be diagonally dominant, which requires [14]:

$$|a_0| \geq \sum_{m=1}^{L} |a_m|. \tag{26}$$

In that case, the bounds in (24) are still tight. To evaluate the eigenvalues $\lambda_{e,\ell}$, we take a closer look at the elements a_m of $\mathbf{B}_e^H \mathbf{B}_e$. Taking into account the decomposition $\mathbf{G} = \mathbf{U}\mathbf{W}$, the elements a_m can be rewritten as

$$a_m = \mathbf{e}^H \mathbf{W}^H \mathbf{U}^H \mathbf{B}^H \mathbf{X}_m \mathbf{B} \mathbf{U} \mathbf{W} \mathbf{e}$$
$$= \mathbf{e}'^H \mathbf{U}^H \mathbf{B}^H \mathbf{X}_m \mathbf{B} \mathbf{U} \mathbf{e}', \tag{27}$$

where $\mathbf{e}' = \mathbf{W}\mathbf{e}$. The matrices $\mathbf{U}^H \mathbf{B}^H \mathbf{X}_m \mathbf{B} \mathbf{U}$, $m = 0, \ldots, L$ are defined by the system parameters and are therefore fixed. Only the matrix $\mathbf{W}$ can be selected freely. Earlier in this section, we derived the condition $\mathbf{G}^H \mathbf{G} = c\mathbf{I}_{N_d}$, which implies $\mathbf{W}^H \mathbf{W} = c\mathbf{I}_{N_d}$. In the following, we assume $c = 1$, which corresponds to the normalization of the transmitted energy E_T (13). When $\mathbf{W}$ is a square matrix, i.e., when $N_d = N_m - N_u$, this condition states that $\mathbf{W}$ is orthogonal. Hence, the error vectors $\mathbf{e}$ and $\mathbf{e}'$ have the same statistical properties. As a consequence, we can (without loss of generality) use the second line in (27) to evaluate the matrix $\mathbf{B}_e^H \mathbf{B}_e$ and its eigenvalues, and eliminate the effect of the matrix $\mathbf{W}$. Through simulations, we investigated the magnitude of the elements a_m for randomly selected error vectors $\mathbf{e}'$, where $\mathbf{e}'^H \mathbf{e}' = 1$. In Fig. 1, we show an example of these simulations and the average over $\mathbf{e}'$ of the resulting $|a_m|$ for 100 randomly selected $\mathbf{e}'$. In general, it turns out that $|a_0|$ will be much larger than the other elements $|a_m|$, $m = 1, \ldots, L$, but the requirement (26) will not be fulfilled. Hence, the matrix $\mathbf{B}_e^H \mathbf{B}_e$ is not strictly diagonally dominant in the mathematical sense. However, let us look at the eigenvalues of the matrix $\mathbf{B}_e^H \mathbf{B}_e$. To obtain Fig. 2, we generated 100 random error vectors $\mathbf{e}'$ with $\mathbf{e}'^H \mathbf{e}' = 1$, constructed the matrix $\mathbf{B}_e^H \mathbf{B}_e$, computed the eigenvalues and sorted them in ascending order. As expected, due to the random error vectors, there is some variation on the values of the eigenvalues, although the variance with respect to the average (over $\mathbf{e}'$) of the eigenvalues is reasonably small. The main issue in this example is the presence of an eigenvalue that is much smaller than the other eigenvalues, which are more or less of the same magnitude. We observed in our simulations that in general, one or a few eigenvalues will be at least one order of magnitude smaller than the other eigenvalues. Hence, irrespective of the selected matrix $\mathbf{W}$, the matrix $\mathbf{B}_e^H \mathbf{B}_e$ is rather ill-conditioned. This implies that, although $\mathbf{G}$ is full rank, maximum diversity can only be achieved when the signal-to-noise ratio is extremely high, provided that $\mathbf{G}$ offers sufficient diversification. At practical, intermediate values of the signal-to-noise ratio, the diversity order will be reduced according to the number of small eigenvalues. Further, as $\mathbf{e}$ and $\mathbf{e}'$ have the same statistical properties, this analysis implies that for square matrices $\mathbf{W}$, the selected matrix $\mathbf{W}$ has almost no impact on the error performance.
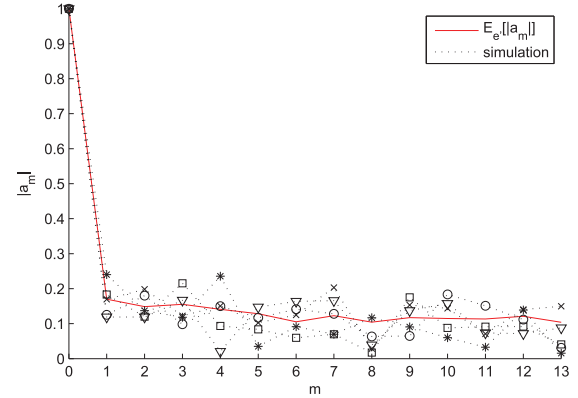

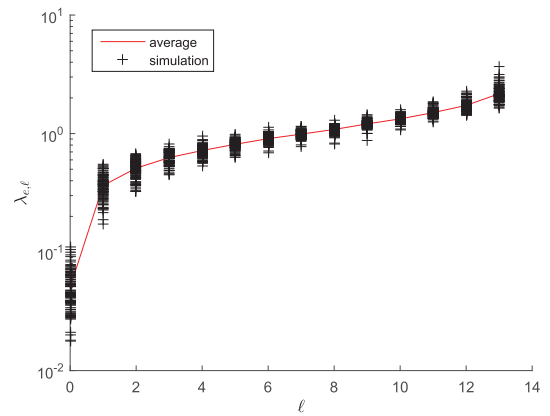Fig. 1. Magnitude of the elements a_m, $m = 0, \ldots, L$, $N = 128$, $N_m = 114$, $N_u = L = 13$, $N_d = 101$.

Fig. 2. Eigenvalues of $\mathbf{B}_e^H \mathbf{B}_e$, $N = 128$, $N_m = 114$, $N_u = L = 13$, $N_d = 101$.

In the above analysis we assumed that $\mathbf{W}$ is a square matrix. In that case, both $\mathbf{e}'$ and $\mathbf{e}$ can take all values in the space $\mathbb{C}^{N_m - N_u}$. When $N_d < N_m - N_u$, i.e., when $\mathbf{W}$ is no longer square, this is no longer the case as $\mathbf{e}'$ now belongs to a N_d-dimensional subspace of $\mathbb{C}^{N_m - N_u}$. Taking into account the condition $\mathbf{W}^H \mathbf{W} = c\mathbf{I}_{N_d}$, it follows that $\mathbf{W}$ forms a tight frame [20], [21]. To optimize the error rate, we could use the redundancy of the tight frame to make $\mathbf{B}_e^H \mathbf{B}_e$ diagonally dominant. However, the amount of redundancy in the tight frame is limited in practice: it is defined by $N_m - N_u - N_d$. Taking into account that, on the average, all non-diagonal elements in $\mathbf{B}_e^H \mathbf{B}_e$ have essentially the same magnitude (see Fig. 1), it is impossible to design $\mathbf{W}$ to reduce the magnitude of all non-diagonal elements for all possible error vectors $\mathbf{e}$. Hence, also when $\mathbf{W}$ is not square, its effect on the system performance is expected to be marginal.

IV. OPTIMIZING THE EUCLIDEAN DISTANCE

In the previous section, we showed that to optimize the error rate performance, the minimum Euclidean distance at the receiver needs to be maximized. However, unless the channel is known at the transmitter, selecting the code generator matrix resulting in the lowest error rate is a complex problem. As an alternative, we showed that, when the channel is not known, a

viable solution is to maximize the minimum Euclidean distance at the transmitter.

A. Euclidean Distance at Transmitter

The last hurdle to take is to find the matrix $\mathbf{G}$ so that $\mathbf{G}^H \mathbf{B}^H \mathbf{B} \mathbf{G} = \mathbf{I}_{N_d}$. To simplify this search, we use the decomposition (4): $\mathbf{G} = \mathbf{U} \mathbf{W}$. By enforcing the restriction $\mathbf{A}_T = \mathbf{I}_{N_d}$, we find the following restriction on $\mathbf{W}$: $\mathbf{W}^H \mathbf{U}^H \mathbf{B}^H \mathbf{B} \mathbf{U} \mathbf{W} = \mathbf{I}_{N_d}$, as the matrices $\mathbf{U}$ and $\mathbf{B}$ are determined by the system parameters, and thus known. To find $\mathbf{W}$, we notice that $\mathbf{B}^H \mathbf{B} = \mathbf{I}_{N_m}$ and $\mathbf{U}^H \mathbf{U} = \mathbf{I}_{N_m - N_u}$. Hence, the condition on $\mathbf{W}$ reduces to $\mathbf{W}^H \mathbf{W} = \mathbf{I}_{N_d}$. When $N_d = N_m - N_u$, i.e., when $\mathbf{W}$ is a square matrix, this implies that $\mathbf{W}$ must be an orthogonal matrix, whereas when $N_d < N_m - N_u$, $\mathbf{W}$ must be a tight frame [20], [21]. In both cases, many possible solutions for the matrix $\mathbf{W}$ exist. This flexibility can be used to optimize other performance measures.

B. Euclidean Distance at Receiver

To optimize the error rate performance, the minimum Euclidean distance $d_{\min, R}$ at the receiver must be maximized. In that case, we proved that the code generator matrix $\mathbf{G}$ has to be selected so that $\mathbf{A}_R = \mathbf{G}^H \mathbf{B}^H \tilde{\mathbf{H}}^H \tilde{\mathbf{H}} \mathbf{B} \mathbf{G} = \mathbf{I}_{N_d}$. Because of the dependency of the matrix $\mathbf{A}_R$ on the channel taps $\mathbf{h}$, finding $\mathbf{G}$ is not straightforward, unless the channel is known. Therefore, we only consider the case where the channel matrix $\tilde{\mathbf{H}}$ is known at the transmitter. To obtain the matrix $\mathbf{G}$ that results in $\mathbf{A}_R = \mathbf{I}_{N_d}$, we again decompose the matrix $\mathbf{G}$ into $\mathbf{G} = \mathbf{U} \mathbf{W}$. To find $\mathbf{W}$, we use the eigenvalue decomposition of the $(N_m - N_u) \times (N_m - N_u)$ matrix $\mathbf{U}^H \mathbf{B}^H \tilde{\mathbf{H}}^H \tilde{\mathbf{H}} \mathbf{B} \mathbf{U} = \mathbf{V}_{HBU}^H \mathbf{\Lambda}_{HBU} \mathbf{V}_{HBU}$, where the eigenvalues are real-valued and non-negative, and the matrix $\mathbf{V}_{HBU}$ is a unitary matrix. Assuming the rank of $\mathbf{B}^H \tilde{\mathbf{H}}^H \tilde{\mathbf{H}} \mathbf{B}$ is at least N_d (i.e., the channel does not contain too many spectral nulls), and taking into account that the $N_m \times (N_m - N_u)$ matrix $\mathbf{U}$ has orthogonal columns, we can find at least N_d non-zero eigenvalues. We select out of the non-zero eigenvalues a subset of N_d eigenvalues, i.e., λ_{HBU, n_ℓ}, $\ell = 1, \ldots, N_d$ with $n_\ell \in \{1, \ldots, N - N_u\}$, and define $\mathbf{v}_{HBU, n_\ell}$ as the n_ℓ-th row of $\mathbf{V}_{NBU}$. We construct the $(N_m - N_u) \times N_d$ matrix $\mathbf{W}$ as follows:

$$\mathbf{W} = \mathbf{Z} \mathbf{\Gamma}, \qquad (28)$$

where the $(N_m - N_u) \times N_d$ matrix $\mathbf{Z} = (\mathbf{v}_{HBU, n_1}^H, \ldots, \mathbf{v}_{HBU, n_{N_d}}^H)$, and $\mathbf{\Gamma}$ is a $N_d \times N_d$ diagonal matrix: $\mathbf{\Gamma} = \operatorname{diag}(\lambda_{HBU, n_\ell}^{-1/2})$. With this construction for $\mathbf{W}$, the product $\mathbf{G}^H \mathbf{B}^H \tilde{\mathbf{H}}^H \tilde{\mathbf{H}} \mathbf{B} \mathbf{G} = \mathbf{I}_{N_d}$. When the number of non-zero eigenvalues exceeds the number N_d of data symbols, many combinations to select the eigenvalues exist. Similarly as for the case of the optimization of the minimum Euclidean distance at the transmitter, the degree of freedom can be used to optimize other performance measures, e.g., in [22], this freedom is used to minimize the transmitted energy.

V. NUMERICAL RESULTS

A. Deterministic Channel

In this section, we consider the case where the channel is deterministic and known at the transmitter side. This could correspond to the case of transmission over a cable, where the channel transfer function is measured and made available to the transmitter. Hence, the code generator matrix $\mathbf{G}$ can be constructed to maximize the minimum Euclidean distance at the transmitter (Case 1) or the receiver side (Case 2) (see section IV). In this construction, we still have some degree of freedom, as we can select in Case 1 any matrix $\mathbf{W}$ that forms a tight frame and in Case 2 out of the set of non-zero eigenvalues of the matrix $\mathbf{U}^H \mathbf{B}^H \tilde{\mathbf{H}}^H \tilde{\mathbf{H}} \mathbf{B} \mathbf{U}$, a subset of N_d eigenvalues. In the following, we select in Case 2 $\mathbf{W}$ so that the transmit energy is minimized [22]. Note that Case 1 and Case 2 correspond to the normalization of the transmitted energy $E_T = N_d E_s$ and received energy $E_R = N_d E_s$, respectively (see (13)). We evaluate the bit error rate (BER) performance of the UW-OFDM system for two channel models:

$$\text{channel a} : h(\ell) = \nu$$
$$\text{channel b} : h(\ell) = \nu e^{-\ell}$$

where ν is a constant of normalization so that $\mathbf{h}^H \mathbf{h} = 1$. In [23], [24] it is shown that the BER performance of UW-OFDM depends on the used data detector, because of the redundancy available in the frequency domain. In this paper, we consider the sphere decoder [25] for our Monte-Carlo simulations, as it yields essentially optimal data detection. The simulated BER is shown in Fig. 3, assuming QPSK transmission, $N = 16$, $N_u = N_r = L = 2$ and $g = 1$ carrier at each side of the frequency band is not modulated as a guard band. Hence, the number of modulated carriers equals $N_m = N - 2g = 14$, and the number of data symbols transmitted per UW-OFDM block equals $N_d = N_m - N_r = 12$. Further, bearing in mind that the PEP is approximated by (16), we also plot the error probability:

$$BER_{th} = \frac{1}{N_d} \sum_{i=1}^{N_d} Q\left(\sqrt{\eta(\mathbf{h}) \lambda_i \frac{E_s}{2 N_0}} \right), \qquad (29)$$

where λ_i are the eigenvalues of the matrix $\mathbf{G}^H \mathbf{B}^H \tilde{\mathbf{H}}^H \tilde{\mathbf{H}} \mathbf{B} \mathbf{G}$, and $\eta(\mathbf{h})$ is a factor to take into account the transceiver side where the energy is normalized: $\eta(\mathbf{h}) = 1$ for Case 1, and $\eta(\mathbf{h}) = \frac{E_R}{E_T}$ for Case 2[4]. When the received energy is normalized (this corresponds to the maximization of the minimum Euclidean distance at the receiver), all eigenvalues are equal, so that the BER reduces to $Q\left(\sqrt{\eta(\mathbf{h}) \frac{E_s}{2 N_0}} \right)$. Fig. 3 shows that the approximated theoretical BER (29) well matches the simulated BER, which indicates that the BER performance of a UW-OFDM system in the presence of a known channel can

---

[4]Note that the ratio $\eta(\mathbf{h})$ in Case 2 is not equal to one, even though the channel is normalized, i.e., $\mathbf{h}^H \mathbf{h} = 1$. This can be explained as follows. Because $g > 0$, not all carriers are modulated, so that part of the energy of the channel cannot be used as it is connected to the unmodulated carriers. As a result, the received energy will always be smaller than the transmitted energy, as was shown in eqs. (11) and (12) in [22]. The reduction of the energy depends on the channel realization $\mathbf{h}$, so that also the ratio $\eta(\mathbf{h})$ depends on the channel.
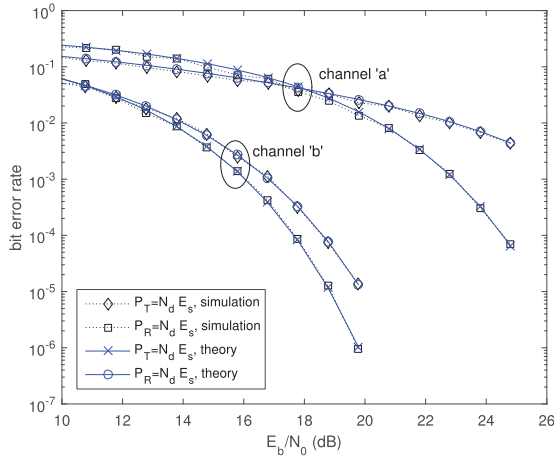
Fig. 3. Bit error probability, QPSK symbols, $N = 16$, $N_u = N_r = L = 2$, $g = 1$.

Fig. 4. Bit error probability, Rayleigh fading, channel known at transmitter, QPSK symbols, $N = 16$, $N_u = L = 4$, $g = 2$, $E_R = N_d E_s$.

easily be evaluated using (29). Further, we observe in Fig. 3 that the BER of Case 2 outperforms that of Case 1, for large signal-to-noise ratios E_s/N_0. This was expected, as in Case 2, the minimum Euclidean distance at the receiver side is maximized, which corresponds to the minimization of the error rate. The difference between the curves is smaller in channel 'b' than for channel 'a'. This can be explained by evaluating the frequency response of the channel: in channel 'b', the frequency response is reasonably flat, whereas in channel 'a', the channel is more frequency selective. Hence, the matrix $\tilde{\mathbf{H}}^H \tilde{\mathbf{H}}$ in channel 'b' will be 'closer' to the identity matrix, so that the eigenvalue spread of $\mathbf{G}^H \mathbf{B}^H \tilde{\mathbf{H}}^H \tilde{\mathbf{H}} \mathbf{B} \mathbf{G}$ will be smaller than for channel 'a'.

B. Rayleigh Fading Channel

In the case of a fading channel, the BER (29) must be averaged over the channel statistics. The BER depends on the channel through the factor $\eta(\mathbf{h})$ and the eigenvalues λ_i of the matrix $\mathbf{G}^H \mathbf{B}^H \tilde{\mathbf{H}}^H \tilde{\mathbf{H}} \mathbf{B} \mathbf{G}$. Hence, to obtain the average BER in a Rayleigh fading channel, the distribution of the eigenvalues and/or $\eta(\mathbf{h})$ must be known.

Let us first consider the case of a slow fading channel where the channel characteristics are estimated and fed back to the transmitter. We select the generator matrix $\mathbf{G}$ to maximize the minimum Euclidean distance at the receiver. Hence, the BER for QPSK transmission for a single channel realization equals $BER_{|\mathbf{h}} = Q\left(\sqrt{\eta(\mathbf{h}) \frac{E_s}{2N_0}}\right)$, where $\eta(\mathbf{h}) = N_d (\text{trace}(\mathbf{U}^H \mathbf{B}^H \tilde{\mathbf{H}}^H \tilde{\mathbf{H}} \mathbf{B} \mathbf{U})^{-1})^{-1}$. We simulated the BER through Monte-Carlo simulation, where we randomly generated the channels $\mathbf{h}$, with the real and imaginary part of each tap of the $(L+1)$-tap channel i.i.d. Gaussian distributed so that $E[\mathbf{h}\mathbf{h}^H] = \frac{1}{L+1}\mathbf{I}_{L+1}$. Further, we randomly selected data symbols and detect them using the sphere decoder. At the same time, we computed the BER in a semi-analytical way, where we consider 2.10^8 random channel realizations and compute the corresponding $\eta(\mathbf{h})$, with which we average $BER_{|\mathbf{h}}$. The results are shown in Fig. 4. As can be observed, the semi-analytical results well correspond to the Monte-Carlo
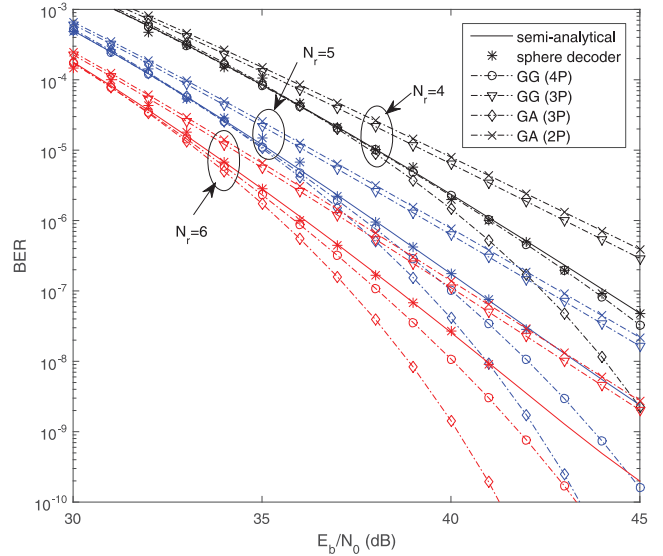
simulations, indicating the theoretical expression (29) can be used to find the BER in a fading channel[5].

Although the semi-analytical BER requires less simulation time than the Monte-Carlo simulations, it is still more complex than a simple theoretical expression for the BER. However, this latter approach requires the knowledge of the distribution of $\eta(\mathbf{h})$. As $\eta(\mathbf{h}) = N_d (\text{trace}(\mathbf{U}^H \mathbf{B}^H \tilde{\mathbf{H}}^H \tilde{\mathbf{H}} \mathbf{B} \mathbf{U})^{-1})^{-1}$, finding the distribution of $\eta(\mathbf{h})$ in an analytical way is not straightforward. However, in the literature, e.g. [26]–[29], it is shown that the distribution of the fading can be approximated by a (generalized) gamma distribution. Taking into account that the transmission of the UW-OFDM signal over a fading channel can be compared to the transmission of signals over correlated fading channels, we expect that the (generalized) gamma distribution approximation will also be valid for $\eta(\mathbf{h})$. In order to find the distribution of $\eta(\mathbf{h})$, we simulate 250000 random channel realizations, compute the corresponding $\eta(\mathbf{h})$ and use EasyFit to fit the values of $\eta(\mathbf{h})$ to four distributions, i.e. the gamma distribution with two (GA(2P)) and three (GA(3P)) parameters, i.e.,

$$f_{GA}(x) = \frac{(x - \gamma)^{k-1}}{\theta^k \Gamma(k)} e^{-\frac{x-\gamma}{\theta}}, x \geq \gamma,$$

where $\gamma = 0$ for the 2-parameter gamma distribution, and the generalized gamma distribution with three (GG(3P)) and four (GG(4P)) parameters, i.e.,

$$f_{GG}(x) = \frac{\mu(x - \gamma)^{\mu k - 1}}{\theta^{\mu k} \Gamma(k)} e^{-(\frac{x-\gamma}{\theta})^\mu}, x \geq \gamma,$$

where $\gamma = 0$ for the 3-parameter generalized gamma distribution. EasyFit estimates the parameters k, θ, γ and μ of the distributions in an iterative way, based on the maximum likelihood principle. In our simulations, we compared the goodness

---

[5]Note that this result is not obvious, as (29) is not necessarily accurate for small values of $\eta(\mathbf{h})$, because of the union-bound approximation used in the derivation of (29) and the fact that sequences at a larger Euclidean distance are neglected.

$N/L/N_r/g$	SA	GA(2P)	GA(3P)	GG(3P)	GG(4P)
16/2/2/1	2.8	1.7612	1.9007	1.8584	2.5577
16/4/4/2	3.3	2.6531	2.6581	2.7186	3.1789
16/4/5/2	3.7	3.1334	3.0755	3.1944	3.6014
16/4/6/2	4.1	3.4483	3.335	3.5114	3.9928
32/6/6/3	3.5	3.1348	3.1721	3.1758	3.9454
32/6/8/3	4.1	4.1003	3.9769	4.1568	4.8476
32/6/10/3	5.1	4.7072	4.4798	4.7521	5.0781
32/8/8/4	3.5	3.8628	3.9311	3.8675	4.2706

$N/L/N_r/g$	GA(2P)	GA(3P)	GG(3P)	GG(4P)
16/4/4/2	0.55185	0.5519	0.55255	0.5518
16/4/5/2	0.7558	0.75585	0.75635	0.75575
16/4/6/2	0.94795	0.9479	0.9485	0.9479

of fit (i.e., the Kolmogorov-Smirnov, Anderson-Darling and Chi-Squared statistics) to determine which distribution best fits the simulated values of $\eta(\mathbf{h})$. In all considered cases, the generalized gamma distribution with four parameters resulted in the best fit, whereas the gamma distribution in most cases had the worst fit. Based on the parameters of the fitted (generalized) gamma distribution, we can obtain an analytical expression for the BER by averaging (29) over the resulting (generalized) gamma distribution. The accuracy of the analytical expression can only be guaranteed if the distribution well fits in the tails, especially for small values of $\eta(\mathbf{h})$ as the diversity order is determined by the behaviour of the distribution for small values of $\eta(\mathbf{h})$. Therefore, we compare the parameters k (for the gamma distribution) and μk (for the generalized gamma distribution) with the diversity order extracted from the BER curves obtained with the semi-analytical (SA) results. The results are shown in Table I. From this table, we observe that the diversity order obtained with the gamma distribution is not sufficiently accurate to obtain an accurate analytical expression for the BER. This is explained as the gamma distribution is known to be less accurate in the tails of the distribution. On the other hand, the generalized gamma distribution with four parameters in many cases well fits the diversity order obtained through simulations. However, for some cases (i.e., for the parameter sets $N/L/N_r/g$ 32/6/6/3, 32/6/8/3, 32/8/8/4), the diversity order is far from the real diversity order. We looked if there is a relationship between the accuracy of the diversity order and the goodness of fit of the distribution, but found no correlation. Hence, we are not able to predict if the fitted diversity order is accurate. We numerically computed the analytical expression for the BER based on the fitted (generalized) gamma distribution, and added the results in Fig. 4. As can be observed, the GA(2P) and GG(3P) distributions result in a BER that is far from the real BER. This is because the diversity orders of the fitted distributions (see Table I) are considerably smaller than the real diversity order. The GA(3P) and GG(4P) distributions better follow the real BER, especially in the lower E_b/N_0 region, but deviate as soon as $\gamma \frac{E_b}{N_0} > 1$. Hence, none of the considered distributions is able to accurately predict the real BER for large E_b/N_0.

Note that the diversity order in Fig. 4 in general is smaller than the maximum diversity order $L + 1$ predicted by the theory. This can be explained as follows. In order to satisfy the restrictions in the time-domain signal, the minimum amount of redundancy that needs to be added in the UW-OFDM system equals $N_r = N_u$. Hence, when $N_r = N_u$, there is not much flexibility in selecting the generator matrix $\mathbf{G}$. Therefore,

although the generator matrix offers some signal diversity, the diversification offered by the system is insufficient to obtain maximum diversity order. When $N_r - N_u$ increases, the degrees of freedom in selecting the generator matrix increases, so that the diversity order rises, as the weakest modes, corresponding to the smallest eigenvalues λ_{HBU} (see section IV-B), can be circumvented.

Although the main goal of Fig. 4 is to evaluate the accuracy of the theoretical expressions for the BER, we can also draw some conclusions about the amount of redundancy that should be added in practice. In Fig. 4, we assumed the received energy per data symbol is constant and given by E_s, i.e. the signal-to-noise ratio E_b/N_0 corresponds to the received energy per bit. To convert the signal-to-noise ratio in Fig. 4 to the ratio of the transmitted energy per bit/symbol to the noise level, the value of $E_s(E_b)$ in Fig. 4 must be divided by $\eta(\mathbf{h})$. This factor will result in a shift of the BER curves, and a good indication for the resulting shift is the average (over all channel realizations) of $\eta(\mathbf{h})$. The average $E[\eta(\mathbf{h})]$ is given in Table II for the three cases shown in Fig. 4, based on the fitted parameters for the (generalized) gamma distribution. We observe that $E[\eta(\mathbf{h})]$ increases when N_r increases. This can be explained as by increasing N_r, we have the flexibility to avoid the weakest modes of the channel, which correspond to the smallest eigenvalues of the matrix $\mathbf{U}^H \mathbf{B}^H \tilde{\mathbf{H}}^H \tilde{\mathbf{H}} \mathbf{B} \mathbf{U}$. This will improve the energy efficiency of the system: by using the strongest modes only, less energy will be lost in the transmission of the data symbols. Simulations have shown that for Rayleigh fading channels, a channel will typically have only a few weak modes (i.e. eigenvalues that are at least an order of magnitude smaller than the average of the eigenvalues of $\mathbf{U}^H \mathbf{B}^H \tilde{\mathbf{H}}^H \tilde{\mathbf{H}} \mathbf{B} \mathbf{U}$), whereas the other eigenvalues have comparable magnitude. Therefore, when N_r further increases, the additional gain in energy efficiency will become smaller. However, by increasing N_r, although the energy efficiency improves, the bandwidth efficiency reduces, as less data symbols are transmitted per OFDM block. In practice, to reduce the loss in overall data rate, the DFT size N is increased, so that the guard interval length N_u becomes much smaller than N. In that case, a small increase of the redundancy N_r [6] will only result in a negligible reduction of the bandwidth efficiency, whereas the energy efficiency noticeably improves.

Until now, we assumed that the channel is known at the transmitter. However, when the channel is not known at the transmitter, the generator matrix cannot be selected to optimize the Euclidean distance at the receiver. Earlier in this paper, we have shown that as an alternative, the generator matrix can be selected to optimize the minimum Euclidean distance at the transmitter. In that case, $\eta(\mathbf{h}) = 1$, and the BER given

---

[6] where $N_r - N_u$ is smaller than or equal to the number of eigenvalues of $\mathbf{U}^H \mathbf{B}^H \tilde{\mathbf{H}}^H \tilde{\mathbf{H}} \mathbf{B} \mathbf{U}$ that are at least an order of magnitude smaller than the average of the eigenvalues.

the channel equals $BER_{|\mathbf{h}} = \frac{1}{N_d} \sum_{i=1}^{N_d} Q\left(\sqrt{\lambda_i \frac{E_s}{2N_0}}\right)$, where λ_i are the eigenvalues of the matrix $\mathbf{G}^H \mathbf{B}^H \tilde{\mathbf{H}}^H \tilde{\mathbf{H}} \mathbf{B} \mathbf{G}$. Hence, we need to find the distribution of the eigenvalues λ_i. Taking into account that the matrix $\tilde{\mathbf{H}}$ contains the Fourier transform of the Rayleigh faded channel taps, and the fact that the relationship between the elements of a matrix and the eigenvalues of the matrix is non-linear, deriving closed-form expressions for the distribution of the eigenvalues is not straightforward. Although solutions are available through random matrix theory [30] – stating that for a $n \times p$ matrix $\mathbf{X}$ of which the (real and imaginary parts of the) rows of $\mathbf{X}$ are randomly drawn from a p-variate Gaussian distribution, the eigenvalues of $\mathbf{X}^H \mathbf{X}$ are distributed according to the Wishart distribution, which is a generalization of the gamma distribution – this result cannot be applied directly to the problem at hand. Because of the redundancy in the frequency domain, the rows of the matrix $\tilde{\mathbf{H}} \mathbf{B} \mathbf{G}$ are not independent. However, we expect that the distribution of the eigenvalues can be approximated by a (generalized) gamma distribution. Similarly as the case where the channel is known at the transmitter, we carried out Monte-Carlo simulations, and our results showed that the distribution of the (unordered) eigenvalues approximates a (generalized) gamma distribution. However, similarly as for the distribution of $\eta(\mathbf{h})$, the (generalized) gamma distribution is not sufficiently accurate to provide a fully analytical solution based on the fitted distribution parameters.

When the channel is not known at the transmitter, the generator matrix $\mathbf{G}$ cannot be optimized to the channel. However, in this paper, we have shown that, using the decomposition $\mathbf{G} = \mathbf{U}\mathbf{W}$, the generator matrix offers several degrees of freedom as the matrix $\mathbf{W}$ can freely be selected provided it forms a tight frame. In the following, we evaluate the BER performance for two realizations of the tight frame, i.e., in Case 1, we consider $\mathbf{W}$ to be the generalized harmonic tight frame with $(\mathbf{W})_{k,\ell} = \frac{1}{\sqrt{N_m - N_u}} e^{-j2\pi \frac{k\ell}{N_m - N_u}}$, and in Case 2, the matrix $\mathbf{W}$ is a truncated identity matrix, i.e. $(\mathbf{W})_{k,\ell} = \delta_{k,\ell}$. Fig. 5 shows the BER obtained in a semi-analytical way, where 2.10^8 channels are randomly generated, and the corresponding eigenvalues λ_i are computed in order to average $BER_{|\mathbf{h}} = \frac{1}{N_d} \sum_{i=1}^{N_d} Q\left(\sqrt{\lambda_i \frac{E_s}{2N_0}}\right)$ over λ_i. Comparing the two possible realizations of the matrix $\mathbf{W}$, it can be observed in the figure that the truncated identity matrix (Case 2) offers a slightly better BER performance, although the difference is small, and for the case where $N_r = L$, the curves for the two cases overlap. This confirms the conclusion from Section III-B stating that the error rate performance is (essentially) independent of the selected matrix $\mathbf{W}$ for given system parameters. Hence, the truncated identity matrix, which is simple to implement, is a proper choice unless the matrix $\mathbf{W}$ is selected to optimize other performance measures. Further, similarly as for the case where the Rayleigh fading channel is known at the transmitter, the diversity order increases when the amount of redundancy is increased by increasing N_r, although the effect is modest. The obtained diversity order is smaller than for the case where the channel is known at the transmitter, and also smaller than the theoretical maximum achievable diversity order, although the diversity order is larger than one. This can be explained as the matrix $\mathbf{G}$ cannot be optimized to avoid the
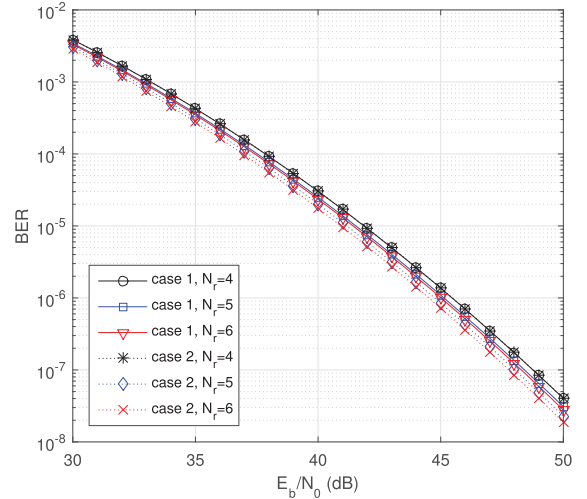

Fig. 5. Bit error probability, Rayleigh fading, channel unknown at transmitter, QPSK symbols, $N = 16$, $N_u = L = 4$, $g = 2$.

weakest modes of the channel and offer a higher amount of diversification, as the transmitter does not know where these weakest modes are. As a conclusion, when the channel is not known at the transmitter, it will be very hard if not impossible to obtain maximum diversity order if we maximize the minimum Euclidean distance at the transmitter.

VI. CONCLUSIONS

In this paper, we theoretically analyse the error rate performance of a UW-OFDM system. Based on this analysis, we derive the constraints to optimize the generator matrix of the UW-OFDM system, and propose a systematic construction method to satisfy these constraints. We show that the UW-OFDM system offers a diversity order larger than one when used in a frequency selective fading channel. Hence, in contrast with standard CP-OFDM, where no additional precoding or channel coding is used, the diversity order of standard UW-OFDM, where no channel coding is used, is considerably larger, although in practice the maximum diversity order is not reached when $N_r = N_u$, which is the minimum amount of redundancy required. We also showed that we can increase the diversity order of UW-OFDM by adding a larger amount of redundancy, corresponding to increasing $N_r - N_u$. In the case where the channel is known at the transmitter, and the generator matrix is optimized to maximize the minimum Euclidean distance at the receiver, the resulting diversity order will approximate the maximum diversity order with growing N_r. However, in case the channel is not known at the transmitter, we showed that the error rate performance is essentially independent of the generator matrix and the increase of the diversity order as function of N_r is moderate, so that the maximum diversity order is very hard to obtain in this case.

APPENDIX A
PROOF OF THE FULL RANK OF THE MATRIX $\tilde{\mathbf{F}}$

In the proof, we will use the following proposition:

Proposition A.1. Assume $d - 1$ consecutive elements of a size-N time-domain vector $\mathbf{c}$ are zero. Denote $\mathbf{C}$ the discrete

Fourier transform of the vector $\mathbf{c}$. Then the vector $\mathbf{C}$ contains at least d non-zero components.

This proposition and its proof are similar to the theorem on the minimum weight of a BCH code, given in [31](on p. 187). The only difference is the kernel of the Fourier transform used: for the BCH code, the kernel of the Fourier transform is a primitive element of the Galois field, whereas in this paper, the kernel is the N-th root of unity, i.e. $e^{j2\pi/N}$.

Now we show that the matrix $\tilde{\mathbf{F}}$ has full rank.

Proof: Let us define S_m as the set of indices of N_m modulated carriers. Hence, the components of the matrix $\tilde{\mathbf{F}}$ yield

$$(\tilde{\mathbf{F}})_{k,m} = \frac{1}{\sqrt{N}} e^{j2\pi \frac{(k+N-N_u)m}{N}}, \qquad (30)$$

with $k = 0, \ldots, N_u - 1$ and $m \in S_m$. Further, the Gramian matrix $\tilde{\mathbf{F}}\tilde{\mathbf{F}}^H$ equals

$$(\tilde{\mathbf{F}}\tilde{\mathbf{F}}^H)_{k,k'} = \frac{1}{N} \sum_{m \in S_m} e^{j2\pi \frac{(k-k')m}{N}}. \qquad (31)$$

Taking into account that the ranks of a matrix and its Gramian are equal, it is sufficient to show that the Gramian matrix $\tilde{\mathbf{F}}\tilde{\mathbf{F}}^H$ has full rank. Let us look closer at the matrix $\tilde{\mathbf{F}}\tilde{\mathbf{F}}^H$. It is a positive semi-definite Hermitian Toeplitz matrix, as $\mathbf{x}^H \tilde{\mathbf{F}}\tilde{\mathbf{F}}^H \mathbf{x} = ||\mathbf{x}\tilde{\mathbf{F}}||^2 \geq 0$. Hence, the eigenvalues of the matrix $\tilde{\mathbf{F}}\tilde{\mathbf{F}}^H$ are real-valued and non-negative. ∎

Let us assume the matrix $\tilde{\mathbf{F}}\tilde{\mathbf{F}}^H$ is not full rank. In that case, at least one of the eigenvalues of $\tilde{\mathbf{F}}\tilde{\mathbf{F}}^H$ equals zero. This implies a vector $\mathbf{x} \neq 0$, corresponding to the eigenvector of the zero eigenvalue, exists for which $\mathbf{x}^H \tilde{\mathbf{F}}\tilde{\mathbf{F}}^H \mathbf{x} = 0$. Let us rewrite $\mathbf{x}^H \tilde{\mathbf{F}}\tilde{\mathbf{F}}^H \mathbf{x}$:

$$\mathbf{x}^H \tilde{\mathbf{F}}\tilde{\mathbf{F}}^H \mathbf{x} = \sum_{m \in S_m} \left| \frac{1}{\sqrt{N}} \sum_{k=0}^{N_u-1} x_k e^{-j2\pi \frac{(k+N-N_u)m}{N}} \right|^2$$

$$= \sum_{m \in S_m} \left| \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} x'_k e^{-j2\pi \frac{km}{N}} \right|^2$$

$$= \sum_{m \in S_m} |X'_m|^2 \qquad (32)$$

where $\mathbf{X}' = \{X'_0 \ldots X'_{N-1}\}$ is the discrete Fourier transform of $\mathbf{x}' = \{x'_0 \ldots x'_{N-1}\}$ with

$$x'_k = \begin{cases} 0 & k = 0, \ldots, N - N_u - 1 \\ x_{k-N+N_u} & k = N - N_u, \ldots, N - 1 \end{cases}. \qquad (33)$$

Hence, if $\mathbf{x}^H \tilde{\mathbf{F}}\tilde{\mathbf{F}}^H \mathbf{x} = 0$, the N_m components of $\mathbf{X}'$ at positions $m \in S_m$ must be zero. However, the vector $\mathbf{x}'$ has $N - N_u$ consecutive components equal to zero, i.e., the first $N - N_u$ elements of the vector. Taking into account Proposition A.1, as $d - 1 = N - N_u$ successive components of $\mathbf{x}'$ are zero, its Fourier transform $\mathbf{X}'$ must have at least $d = N - N_u + 1$ non-zero components, implying the maximum number of zero components in $\mathbf{X}'$ equals $N_u - 1$. Assuming the number of transmitted data symbols is larger than zero, i.e., $N_d > 0$, and

taking into account that $N_d = N_m - N_r \leq N - N_u$, it follows that $N_u < N_m$. Hence, not all N_m components X'_m with $m \in S_m$ can be zero, implying no eigenvalue of $\tilde{\mathbf{F}}\tilde{\mathbf{F}}^H$ equal to zero exists, so the matrix $\tilde{\mathbf{F}}$ is full rank.

APPENDIX B
PROOF OF THE RESTRICTION $\mathbf{G}^H \mathbf{B}^H \mathbf{B} \mathbf{G} = \alpha \mathbf{I}_{N_d}$

In this appendix, we show that the minimization of the MSE of the BLUE or LMMSE data detector in an AWGN channel results in the requirement $\mathbf{G}^H \mathbf{B}^H \mathbf{B} \mathbf{G} = \alpha \mathbf{I}_{N_d}$.

Proof: In [6], it is shown that the cost functions to be optimized for the BLUE and LMMSE data detector in an AWGN channel are given by

$$J_{BLUE} = C_1 \text{trace}(\mathbf{G}^H \mathbf{G}) \text{trace}(\mathbf{G}^H \mathbf{G})^{-1} \qquad (34)$$

$$J_{LMMSE} = C_2 \text{trace} \left(\frac{C_3(\mathbf{G}^H \mathbf{G})}{\text{trace}(\mathbf{G}^H \mathbf{G})} + \mathbf{I}_{N_d} \right)^{-1} \qquad (35)$$

where C_1, C_2 and C_3 are constants irrelevant for the optimization. Assuming that the transmitted energy E_T (13) per UW-OFDM block equals $\alpha N_d E_s$ and taking into account that $\mathbf{B}^H \mathbf{B} = \mathbf{I}_{N_m}$, it follows that $\text{trace}(\mathbf{G}^H \mathbf{G}) = \alpha N_d$. Hence, the optimization problem reduces to

$$\min_{\mathbf{G}} \{J_{BLUE}, J_{LMMSE}\}$$
$$\text{s.t.trace}(\mathbf{G}^H \mathbf{G}) = \alpha N_d. \qquad (36)$$

Further, the squared Euclidean distance at the transmitter (11) equals $d^2(\mathbf{z}, \mathbf{z}') = \mathbf{e}^H \mathbf{G}^H \mathbf{B}^H \mathbf{B} \mathbf{G} \mathbf{e} = \mathbf{e}^H \mathbf{G}^H \mathbf{G} \mathbf{e}$, implying $\mathbf{G}^H \mathbf{G}$ is a positive-definite Hermitian matrix. Let us consider the eigenvalue decomposition of $\mathbf{G}^H \mathbf{G} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H$; the eigenvalues λ_k are positive real-valued and the eigenvector matrix $\mathbf{V}$ is a unitary matrix. Applying this eigenvalue decomposition to the cost functions J_{BLUE} and J_{LMMSE}, we obtain

$$J_{BLUE} = C_1 \alpha N_d \text{trace}(\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H)^{-1}$$
$$= C_1 \alpha N_d \text{trace}[(\mathbf{V}^H \mathbf{V})^{-1}\boldsymbol{\Lambda}^{-1}]$$
$$= C_1 \alpha N_d \sum_{k=1}^{N_d} \frac{1}{\lambda_k} \qquad (37)$$

and similarly

$$J_{LMMSE} = C_2 \sum_{k=1}^{N_d} \frac{1}{\frac{C_3}{\alpha N_d}\lambda_k + 1}, \qquad (38)$$

which have to be optimized subject to $\sum_{k=1}^{N_d} \lambda_k = \alpha N_d$. Using the Lagrange formalism, it straightforwardly follows that both cost functions are minimized when all eigenvalues are equal, i.e., $\lambda_k = \alpha$, implying that $\mathbf{G}^H \mathbf{G} = \mathbf{G}^H \mathbf{B}^H \mathbf{B} \mathbf{G} = \alpha \mathbf{I}_{N_d}$. ∎

APPENDIX C
PROOF OF PROPOSITION III.2

Proof: Let us assume a matrix $\mathbf{G}$ exists so that for all possible $\mathbf{e}$, and thus also for the worst case $\mathbf{e}$ resulting in

the smallest minimum eigenvalue, we are able to make $\mathbf{B}_e^H \mathbf{B}_e$ diagonal. This implies

$$a_m = \sum_{k=0}^{N-1} e^{j2\pi \frac{km}{N}} |(\mathbf{BGe})_k|^2 = 0, \qquad \text{for } m > 0. \quad (39)$$

From the definition of a_m, it follows that it is the Fourier transform of $|(\mathbf{BGe})_k|^2$. The requirement that $a_m = a_0 \delta_m$ can only be fulfilled when $|(\mathbf{BGe})_k|^2$ is a constant, i.e. $|(\mathbf{BGe})_k|^2 = a_0, \forall \mathbf{e}$, or equivalently $\mathbf{BGe} = \sqrt{a_0} e^{j\theta_e} \mathbf{1}_N$, where θ_e is a rotation angle and $\mathbf{1}_N$ is the vector of ones of length N. The solution of $\mathbf{BGe} = \sqrt{a_0} e^{j\theta_e} \mathbf{1}_N$ is given by

$$\mathbf{e} = \mathbf{e}_0 + \sum_{\ell=1}^{N-N_d} w_{BG,\ell} \mathbf{u}_{BG,\ell} \quad (40)$$

with $\mathbf{e}_0 = \sqrt{a_0} e^{j\theta_e} (\mathbf{BG})^\dagger \mathbf{1}_N$, where $(\mathbf{BG})^\dagger$ is the Penrose-Moore pseudo-inverse of $\mathbf{BG}$, the vectors $\mathbf{u}_{BG,\ell}$ form an (orthogonal) basis for the null space of the matrix $\mathbf{BG}$, and $w_{BG,\ell}$ are weighting coefficients. We assumed that $\mathbf{BG}$ has full rank, i.e., its rank is N_d. Therefore, $\mathbf{BG}$ has $N - N_d$ null vectors. If we want the solution (40) to represent all possible realizations of $\mathbf{e} \in \mathbb{C}^{N_d \times 1}$ (with $\mathbf{e}^H \mathbf{e} = 1$), this implies that the null space must span the space $\mathbb{C}^{N_d \times 1}$. When $N_d \leq N/2$, it follows that $N_d \leq N - N_d$, so that it is possible to make the matrix $\mathbf{B}_e^H \mathbf{B}_e$ diagonal for all error vectors $\mathbf{e} \in \mathbb{C}^{N_d \times 1}$. However, when $N_d > N/2$, it follows $N - N_d < N/2 < N_d$, which implies there are not enough null vectors to span the whole space $\mathbb{C}^{N_d \times 1}$. ∎

REFERENCES

[1] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, vol. 28, no. 5, pp. 5–14, May 1990.

[2] M. Huemer, C. Hofbauer, and J. B. Huber, "The potential of unique words in OFDM," in *Proc. 15th Int. OFDM Workshop (InOWo'10)*, Hamburg, Germany, Sep. 2010, pp. 140–144.

[3] A. Onic and M. Huemer, "Direct vs. two-step approach for unique word generation in UW-OFDM," in *Proc. 15th Int. OFDM-Workshop (InOWo'10)*, Hamburg, Germany, Sep. 2010, pp. 145–149.

[4] M. Rajabzadeh, H. Steendam, and H. Khoshbin, "Power spectrum characterization of systematic coded UW-OFDM systems," in *Proc. Veh. Technol. Conf. (VTC'13)*, Las Vegas, NV, USA, Sep. 2–5, 2013, pp. 1–5.

[5] M. Rajabzadeh, N. Noels, H. Khoshbin, and H. Steendam, "Sidelobe suppression for non-systematic coded UW-OFDM in cognitive radio networks," in *Proc. Eur. Wireless Conf. (EW'14)*, Barcelona, Spain, May 2014, pp. 1–6.

[6] M. Huemer, C. Hofbauer, and J. Huber, "Non-systematic complex number RS coded OFDM by unique word prefix," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 285–299, Jan. 2012.

[7] W. Haselmayr, C. Hofbauer, B. Etzlinger, A. Springer, and M. Huemer, "Iterative detection for unique word OFDM," in *Proc. GLOBECOM*, Austin, TX, USA, Dec. 2014, pp. 3261–3266.

[8] J. B. Huber, J. Rettelbach, M. Seidl, and M. Huemer, "Signal shaping for unique word OFDM by selected mapping," in *Proc. Eur. Wireless Conf. (EW'12)*, Poznan, Poland, Apr. 2012, pp. 1–8.

[9] H. L. Van Trees, *Detection, Estimation and Modulation Theory*. Hoboken, NJ, USA: Wiley, 1968.

[10] Z. Wang and G. B. Giannakis, "Complex-field coding for OFDM over fading wireless channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 1–13, Mar. 2003.

[11] S.-H. Chang, P. C. Cosman, and L. B. Milstein, "Chernoff-type bounds for the Gaussian error function," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 2939–2944, Nov. 2011.

[12] M. Huemer, C. Hofbauer, A. Onic, and J. B. Huber, "Design and analysis of UW-OFDM signals," *Int. J. Electron. Commun.*, vol. 68, no. 10, pp. 958–968, 2014.

[13] H. Steendam, "Analysis of the redundant energy in UW-OFDM," *IEEE Trans. Commun.*, vol. 60, no. 6, pp. 1692–1701, Jun. 2012.

[14] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*. Philadelphia, PA, USA: SIAM, 2000.

[15] Q.-T. Ngo, O. Berder, and P. Scalart, "General minimum Euclidean distance-based precoder for MIMO wireless systems," *EURASIP J. Adv. Signal Process.*, vol. 2013, no. 39, 12 pp., Mar. 2013.

[16] X.-L. Huang, G. Wang, and F. Hu, "Minimal Euclidean distance-inspired optimal and suboptimal modulation schemes for vector OFDM system," *Int. J. Commun. Syst.*, vol. 24, pp. 553–567, Aug. 2011.

[17] H.-C. Li, "On calculating the determinants of Toeplitz matrices," *J. Appl. Math. Bioinformat.*, vol. 1, no. 1, pp. 55–64, May 2011.

[18] J. M. Steele, The Cauchy-Schwarz Master Class:An Introduction to the Art of Mathematical Inequalities. Cambridge, U.K.: *Cambridge Univ. Press*, 2004.

[19] T. M. Cover and J. A. Thomas, "Determinant inequalities via information theory," *SIAM J. Matrix Anal. Appl.*, vol. 9, no. 3, pp. 384–392, Jul. 1988.

[20] P. G. Casazza and N. Leonhard, "Classes of finite equal norm Parseval frames," *Contemp. Math.*, vol. 451, pp. 11–31, 2008.

[21] P. G. Casazza and G. Kutyniok, "A generalization of Gram-Schmidt orthogonalization generating all Parseval frames," *J. Adv. Comput. Math.*, vol. 27, pp. 65–78, 2007.

[22] H. Steendam, "Design and analysis of the UW-OFDM signal," in *Proc. 6th Int. Symp. Commun. Control Signal Process. (ISCCSP'14)*, Athens, Greece, May 21–23, 2014, pp. 511–515.

[23] M. Huemer, A. Onic, and C. Hofbauer, "Classical and Bayesian linear data estimators for unique word OFDM," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6073–6085, Dec. 2011.

[24] A. Onic, "Receiver concepts for unique word OFDM," *Ph.D dissertation, Inst. Netw. Embedded Syst., Alpen-Adria-Universität Klagenfurt, Klagenfurt*, Austria, Nov. 2013, [Online]. Available: http://uwofdm.jku.at/publications.

[25] A. Onic, A. Schenk, M. Huemer, and J. Huber, "Soft-output sphere detection for coded unique word OFDM," in *Proc. Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, Nov. 4–7, 2012, pp. 138–142.

[26] S. Al-Ahmadi and H. Yanikomeroglu, "On the approximation of the generalized-K distribution by a gamma distribution for modeling composite fading channels," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 706–713, Feb. 2010.

[27] V. A. Aalo, "Performance of maximal-ratio diversity systems in a correlated Nakagami-fading environment," *IEEE Trans. Commun.*, vol. 43, no. 8, pp. 2360–2369, Aug. 1995.

[28] E. W. Stacy, "A generalization of the gamma distribution," *Ann. Math. Statist.*, vol. 33, no. 3, pp. 1187–1192, Sep. 1962.

[29] O. Gazi, "Bounds for generalized gamma distributed fading channels," *IEEE Commun. Lett.*, vol. 15, no. 12, pp. 1347–1349, Dec. 2011.

[30] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. Cambridge, U.K: Cambridge Univ. Press, 2011.

[31] G. C. Clark and J. Bibb Cain, *Error-Correction Coding for Digital Communications*. New York, NY, USA:Springer, 1981.

Heidi Steendam (M'01–SM'06) received the M.Sc. degree in electrical engineering and the Ph.D. degree in applied sciences from Ghent University, Gent, Belgium, in 1995 and 2000, respectively. Since September 1995, she has been with the Digital Communications (DIGCOM) Research Group, Department of Telecommunications and Information Processing (TELIN), Faculty of Engineering, Ghent University, first in the framework of various research projects, and since October 2002, as a full time Professor in the area of digital communications. She is the author of more than 140 scientific papers in international journals and conference proceedings, for which several best paper awards were received. Her research interests include statistical communication theory, carrier and symbol synchronization, bandwidth-efficient modulation and coding, cognitive radio, and cooperative networks and positioning. Since 2002, she has been an Executive Committee Member of the IEEE Communications and Vehicular Technology Society Joint Chapter, Benelux Section, and since 2012, the Vice Chair. She has been active in various international conferences as a Technical Program Committee Chair/Member and Session Chair. In 2004 and 2011, she was the conference Chair of the IEEE Symposium on Communications and Vehicular Technology in the Benelux. She is an Associate Editor of the IEEE Transactions on Communications, *EURASIP Journal on Wireless Communications and Networking*, and *Journal on Communications and Network*.