Pruned DFT-Spread FBMC: Low PAPR, Low Latency, High Spectral Efficiency

Ronald Nissel and Markus Rupp, *Fellow, IEEE*

***Abstract*— We propose a novel modulation scheme which combines the advantages of filter bank multi-carrier (FBMC)-offset quadrature amplitude modulation and single-carrier frequency-division multiple access (SC-FDMA). On the top of a conventional FBMC system, we develop a novel precoding method based on a pruned discrete Fourier transform (DFT) in combination with one-tap scaling. The proposed technique has the same peak-to-average power ratio as SC-FDMA but does not require a cyclic prefix and has much lower out-of-band emissions. Furthermore, our method restores complex orthogonality, and the ramp-up and ramp-down period of FBMC is dramatically decreased, allowing low latency transmissions. Compared to pure SC-FDMA, the computational complexity of our scheme is only two times higher. Simulations over doubly selective channels validate our claims, further supported by a downloadable MATLAB code. Note that pruned DFT-spread FBMC can equivalently be interpreted as a modified SC-FDMA transmission scheme. In particular, the requirements on the prototype filter are less strict than in conventional FBMC systems.**

***Index Terms*— FBMC, OQAM, DFT-s-OFDM, windowed OFDM.**

I. Introduction

FILTER Bank Multi-Carrier (FBMC) with Offset Quadrature Amplitude Modulation (OQAM), in short just FBMC, is an interesting modulation scheme for future wireless systems because it has much lower Out-Of-Band (OOB) emissions than Orthogonal Frequency Division Multiplexing (OFDM) [1]. This improves the performance in asynchronous transmissions and allows an efficient time-frequency allocation for different use cases [2]. Additionally, FBMC typically does not require a Cyclic Prefix (CP), further increasing the throughput. To fulfill the Balian-Low theorem [3], FBMC replaces the complex orthogonality condition with the less strict real orthogonality condition. This causes intrinsic interference, concentrated on the imaginary part, which makes channel estimation [4] and Multiple-Input and Multiple-Output (MIMO) [5], [6] more challenging. Several methods have been proposed to deal with those challenges [2], [7], [8]. For example, by spreading symbols in time or frequency complex orthogonality can be restored in FBMC, allowing to straightforwardly employ almost all detection methods from OFDM [2]. This works as long as the channel is approximately flat within the spreading length. The spreading itself can be based on Discrete Fourier Transform (DFT) spreading in time, as proposed in [6]. However, if the channel is approximately flat, our investigations in [9] and [10] indicate that Walsh-Hadamard spreading [5], [11] is a better option than DFT spreading because it perfectly restores complex orthogonality within one block and has a lower computational complexity. Nonetheless, DFT spreading has advantages when it comes to shaping the transmit signal in time and reducing the Peak-to-Average Power Ratio (PAPR).

Besides the intrinsic interference, nonlinearities such as a limited Digital-to-Analog-Converter (DAC) resolution or a nonlinear power amplifier impose an even greater challenge in practical systems because they destroy the superior spectral confinement of FBMC [2], [12]. Thus, FBMC is only useful if operated in a sufficiently linear regime. In multi-carrier systems this is hard to achieve because of the poor PAPR. To reduce the PAPR in OFDM, several techniques have been suggested such as selective mapping [13] or partial transmit sequences [14]. Those methods can be extended to FBMC as shown in [15]–[17]. However, all those techniques require a high computational complexity and side information. Those drawbacks explain why they are not employed in practical systems. Instead, Long Term Evolution (LTE) uses Single Carrier - Frequency-Division Multiple Access (SC-FDMA) in the uplink [18], essentially a DFT precoded OFDM system. The same technique will also be included as an additional option in the uplink of the Fifth Generation (5G) of mobile communication systems (besides CP-OFDM) [19].

Unfortunately, simply combining FBMC and a DFT, similar as in SC-FDMA, performs poorly in FBMC [20]–[22]. To improve the performance, Ihalainen *et al.* [20] propose precoding by a filter bank instead of a DFT. While such method reduces the PAPR it still does not perform as well as SC-FDMA and has the additional disadvantages of an increased overhead and a higher computational complexity. Na and Choi [22] showed that, in contrast to conventional FBMC,

Manuscript received November 21, 2017; revised March 2, 2018 and May 9, 2018; accepted May 11, 2018. Date of publication May 16, 2018; date of current version October 16, 2018. The financial support by the Austrian Federal Ministry of Science, Research and Economy and the National Foundation for Research, Technology and Development is gratefully acknowledged. This work was supported in parts by Inwite and by the Czech Science Foundation, Project No. 17-18675S "Future transceiver techniques for the society in motion." The associate editor coordinating the review of this paper and approving it for publication was S. Muhaidat. *(Corresponding author: Ronald Nissel.)*

R. Nissel is with the Institute of Telecommunications, TU Wien, 1040 Vienna, Austria, and also with the Christian Doppler Laboratory for Dependable Wireless Connectivity for the Society in Motion, TU Wien, 1040 Vienna, Austria (e-mail: ronald.nissel@gmail.com).

M. Rupp is with the Institute of Telecommunications, TU Wien, 1040 Vienna, Austria, and also with the Department of Radio Electronics, Brno University of Technology, 616 00 Brno, Czech Republic (e-mail: mrupp@nt.tuwien.ac.at).

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

Digital Object Identifier 10.1109/TCOMM.2018.2837130

the phase term has an influence on the PAPR performance of a simple DFT spread FBMC scheme. However, even by considering an optimal phase term the PAPR is still not as good as in SC-FDMA. Na and Choi [22] therefore propose a selection scheme. This, however, requires side information and increases the overall complexity as well as the latency. To overcome all those drawbacks we propose a novel modulation scheme based on a pruned DFT in combination with one-tap scaling. Our method even restores complex orthogonality in FBMC. The advantages and possible disadvantages of our method can be summarized as follows:

Advantages:

- Low PAPR, same as in SC-FDMA.
- Low OOB emissions, comparable to FBMC.
- The ramp-up and ramp-down period of FBMC is dramatically reduced, allowing low latency transmissions.
- Complex orthogonality is restored, enabling efficient multi-user uplink transmissions.
- Maximum symbol density, same as in FBMC.
- Low-complexity one-tap equalizers can be used.
- Relatively high robustness in doubly-selective channels.
- In contrast to conventional FBMC a better compatibility to MIMO, but only if the channel is approximately flat within the spreading interval, see disadvantages.

Possible Disadvantages:

- Slightly higher computational complexity, approximately two times that of SC-FDMA.
- Only quasi-orthogonal, that is, some small residual interference remains. This, however, is usually not a problem. Furthermore, an additional frequency CP can reduce this interference.
- Low-complexity Maximum Likelihood (ML) MIMO detection only works if the channel is approximately flat within the spreading interval (same drawback as in SC-FDMA).
- Throughput is usually slightly lower than in multi-carrier systems because of spreading (same drawback as in SC-FDMA).
- Alamouti's space time block code only works if the channel is approximately flat within the spreading interval.

There exist two equivalent interpretations of our novel transmission scheme (transmitter side):

1) Modified FBMC-OQAM: The complex-to-real transformation of a conventional FBMC-OQAM system is replaced by a pruned DFT in combination with one-tap scaling. Furthermore, the prototype filter is reduced to a time-length of approximately $\frac{1.5}{F}$, with F denoting the subcarrier spacing.
2) Modified SC-FDMA: Half of the input data symbols of a conventional SC-FDMA system are set to zero (pruned DFT) and one-tap scaling is applied on the other half of the input data symbols. Furthermore, the Inverse Fast Fourier Transform (IFFT) output, including the CP, is multiplied by an approximately $\frac{1.5}{F}$ length window function, and the time spacing between SC-FDMA symbols is reduced from $T = \frac{1}{F} + T_{\mathrm{CP}}$ to $T = \frac{1}{2F}$.

We will mainly consider the FBMC interpretation because it is well-known that one-tap equalizers often perform close to the optimum in FBMC, while in windowed OFDM without CP this is not clear.

A. Related Work

Pruned DFT spread FBMC has a lower PAPR than the method in [20]. Furthermore, compared with [22], our method does not require any side information. Similar as in [5], [6], [9], and [10] we perform precoding to restore complex orthogonality in FBMC. However, in contrast to [5], [6], [9], and [10], our method also reduces the PAPR and we consider equalization in the multi-carrier domain so that, in contrast to those previous papers, the channel must not necessarily be flat within the spreading length. Moreover, compared with [6], we spread in frequency instead of time, include pre-equalization, employ a modified prototype filter, and focus on the PAPR performance as well as the latency. Compared with [5], [9], and [10], we spread with a pruned DFT instead of a pruned Walsh-Hadamard transform, improving the PAPR. With respect to the pruned DFT, a related concept is also zero-tail DFT spread OFDM [23]. However, our method has much lower OOB emissions and typically does not require any overhead.

B. Outline

In Section II we provide a short overview of conventional FBMC and describe our transmission system model. The idea of pruned DFT spread FBMC is then presented in Section III, where we also discuss the OOB emissions, latency and the computational complexity. In Section IV we quantify the orthogonality approximation error and propose a frequency CP to reduce it. In Section V we investigate the effect of one-tap equalization in doubly-selective channels. Moreover, we discuss how spreading can be utilized to enable MIMO in FBMC. Finally, in Section VI we present Monte Carlo simulations and discuss the performance of the PAPR, the Bit Error Ratio (BER) and the throughput.

Notation: matrices are denoted by bold upper-case letters, vectors by bold lower-case letters and scalars by non-bold letters. The *i*-th row and *j*-th column element of matrix $\mathbf{M}$ is denoted by $[\mathbf{M}]_{i,j}$. Matrix $\mathbf{I}_N$ represents an identity matrix of size *N*, $\mathbf{0}_{N\times M}$ an all zero matrix of size $N \times M$, $\mathbf{W}_L$ a DFT matrix of size *L* and $\widetilde{\mathbf{W}}_{L\times L/2}$ a pruned DFT matrix of size $L \times L/2$, that is, a conventional DFT matrix where $L/2$ columns are canceled.

To support reproducibility our **MATLAB code** can be downloaded at **https://www.nt.tuwien.ac.at/downloads/**

II. FBMC-OQAM

We consider the transmission of *K* FBMC symbols in time, each consisting of *L* subcarriers. The transmitted signal in the time domain, $s(t)$, can then be expressed by [2]

$$s(t) = \sum_{k=1}^{K} \sum_{l=1}^{L} g_{l,k}(t)\, x_{l,k}, \tag{1}$$

where $x_{l,k}$ represents the transmitted symbol at subcarrier position l and time position k, and is usually chosen from a Pulse-Amplitude Modulation (PAM) signal constellation. Basis pulse $g_{l,k}(t)$ in (1),

$$g_{l,k}(t) = p(t - kT)\mathrm{e}^{\mathrm{j}2\pi lF(t-kT)}\mathrm{e}^{\mathrm{j}\frac{\pi}{2}(l+k)}, \tag{2}$$

is, essentially, a time and frequency shifted version of prototype filter $p(t)$, with T denoting the time spacing and F the frequency spacing (subcarrier spacing). We assume that prototype filter $p(t)$ is zero outside the time interval $-OT \leq t < OT$, where O represents the overlapping factor. Furthermore, in FBMC, the prototype filter must be a real-valued even function, $p(t) = p(-t)$, and orthogonal for a time-frequency spacing of $T \times F = TF = 2$. To improve the spectral efficiency in FBMC, the time-spacing as well as the frequency spacing are both reduced by a factor of two, leading to $TF = 0.5$. This causes intrinsic interference which, however, is concentrated on the imaginary part because of the phase term $\mathrm{e}^{\mathrm{j}\frac{\pi}{2}(l+k)}$ in (2). Thus, the interference can easily be canceled by taking only the real part. Note that only real-valued data symbols are transmitted in such a system and that two real-valued data symbols are required to transmit one complex-valued data symbol. Thus, the time-frequency spacing of $TF = 0.5$ corresponds to an equivalent time-frequency spacing of $TF = 1$ for complex-valued symbols. Very often, the real part of a complex-valued symbol is mapped to the first time-slot and the imaginary part to the second time-slot, thus the name offset-QAM. However, such self-limitation is not necessary. One can equivalently perform this mapping over subcarriers or directly consider PAM symbols instead of "staggered" QAM symbols.

To simplify analytical investigations we consider a matrix-based system model [2]. The basis pulses in (2) are sampled at rate $f_\mathrm{s} = 1/\Delta t = FN_\mathrm{FFT}$ and stacked in a basis pulse vector $\mathbf{g}_{l,k} \in \mathbb{C}^{N\times 1}$ according to

$$[\mathbf{g}_{l,k}]_i = \sqrt{\Delta t}\, g_{l,k}(t)\Big|_{t=(i-1)\Delta t-(O-1)T}, \tag{3}$$

for $i = 1, 2, \ldots, N$, where the total number of samples is given by $N = ON_\mathrm{FFT} + \frac{N_\mathrm{FFT}}{2}(K-1)$. The interpretation of overlapping factor O and Fast Fourier Transform (FFT) size $N_\mathrm{FFT} \geq L$ becomes more clear later in this section when we discuss an efficient FFT implementation. Note that practical systems never operate at a critically sampling rate ($N_\mathrm{FFT} = L$) because this would lead to large OOB emissions. This is even more true in FBMC, as a critically sampled FBMC system has the same poor OOB emissions as OFDM.

Utilizing (3) we define the basis pulse matrix at time position k by $\mathbf{G}_k = \begin{bmatrix}\mathbf{g}_{1,k} \ldots \mathbf{g}_{L,k}\end{bmatrix} \in \mathbb{C}^{N\times L}$, and the overall basis pulse matrix by $\mathbf{G} = \begin{bmatrix}\mathbf{G}_1 \ldots \mathbf{G}_K\end{bmatrix} \in \mathbb{C}^{N\times LK}$. The sampled transmit signal in (1), $\mathbf{s} \in \mathbb{C}^{N\times 1}$, can then be expressed by

$$\mathbf{s} = \sum_{k=1}^{K} \mathbf{G}_k\mathbf{x}_k = \mathbf{G}\mathbf{x}, \tag{4}$$

with $\mathbf{x}_k = \begin{bmatrix}x_{1,k} \ldots x_{L,k}\end{bmatrix}^\mathsf{T} \in \mathbb{C}^{L\times 1}$ denoting the transmitted symbols at time position k and $\mathbf{x} = \mathrm{vec}\{\begin{bmatrix}\mathbf{x}_1 \ldots \mathbf{x}_K\end{bmatrix}\} \in \mathbb{C}^{LK\times 1}$ all transmitted symbols in vectorized form. We model the transmission over a doubly-selective channel by a banded time-variant convolution matrix $\mathbf{H} \in \mathbb{C}^{N\times N}$, defined as $[\mathbf{H}]_{i,j} = h_\mathrm{conv.}[i, i-j]$ with time-variant impulse response $h_\mathrm{conv.}[i, m_\tau]$, together with an additive white Gaussian noise vector $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, P_\mathrm{n}\mathbf{I}_N)$ where P_n denotes the noise power in the time domain, so that the received signal $\mathbf{r} \in \mathbb{C}^{N\times 1}$ can be described by,

$$\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{n}. \tag{5}$$

The whole transmission system, after demodulation by $\mathbf{G}^\mathsf{H}$, can then be expressed as,

$$\mathbf{y} = \mathbf{G}^\mathsf{H}\mathbf{H}\mathbf{G}\mathbf{x} + \mathbf{G}^\mathsf{H}\mathbf{n} \tag{6}$$

$$\approx \mathrm{diag}\{\mathbf{h}\}\mathbf{G}^\mathsf{H}\mathbf{G}\mathbf{x} + \mathbf{G}^\mathsf{H}\mathbf{n}, \tag{7}$$

where $\mathbf{y} \in \mathbb{C}^{LK\times 1}$ denotes the received symbols. If the delay spread and the Doppler spread are sufficiently low, the channel induced interference can be neglected [2]. This allows us to factor out the channel in (6) according to (7), where $\mathbf{h} \in \mathbb{C}^{LK\times 1}$ describes the one-tap channel, that is, the diagonal elements of $\mathbf{G}^\mathsf{H}\mathbf{H}\,\mathbf{G}$. To be specific, the one-tap channel at subcarrier position l and time-position k is given by

$$h_{l,k} = \mathbf{g}_{l,k}^\mathsf{H}\mathbf{H}\mathbf{g}_{l,k} \approx H(kT, lF), \tag{8}$$

and can be interpreted as the sampled time-variant transfer function $H(kT, lF)$. Note that FBMC experiences imaginary interference, described by the off-diagonal elements of $\mathbf{G}^\mathsf{H}\mathbf{G}$ in (7), and only the real orthogonality condition holds true, that is, $\Re\{\mathbf{G}^\mathsf{H}\mathbf{G}\} = \mathbf{I}_{LK}$.

FBMC signals can be efficiently generated by an IFFT together with a polyphase network [24]. However, in contrast to the conventional multi-rate interpretation of a polyphase network [25], we consider a vector based description [2], [26], [27]. Such interpretation is important for understanding pruned DFT spread FBMC so that we briefly repeat our results from [2] with some small modifications.

Without loss of generality we consider only time-position $k = 0$ (slight abuse of notation) and $K = 1$ FBMC symbol. Any other time-position can easily be obtained by time-shifting this special case in combination with time-domain overlapping. The main idea for an efficient IFFT implementation is to factor out prototype filter $p(t)$ so that the sampled signal in (1) becomes,

$$s(i\Delta t) = p(i\Delta t)\sum_{l=1}^{L}\mathrm{e}^{\mathrm{j}2\pi l\frac{i}{N_\mathrm{FFT}}}x_{l,0}\mathrm{e}^{\mathrm{j}\frac{\pi}{2}(l+0)}, \tag{9}$$

for $i = -\frac{ON_\mathrm{FFT}}{2}, \ldots, \frac{ON_\mathrm{FFT}}{2} - 1$. The summation in (9) corresponds to an N_FFT point IFFT for which the input arguments are $\{0,\ x_{1,0}\,\mathrm{e}^{\mathrm{j}\frac{\pi}{2}(1+0)},\ \cdots,\ x_{L,0}\,\mathrm{e}^{\mathrm{j}\frac{\pi}{2}(L+0)},\ 0,\ 0,\ \cdots\}$. Furthermore, because l is an integer, the summation in (9) is N_FFT periodic with respect to i. Thus, the IFFT has to be calculated only for N_FFT samples. Those samples can then be copied O-times followed by an element-wise multiplication with the prototype filter $p(i\,\Delta t)$. Such copy operation also appears in windowed CP-OFDM [28], as illustrated in Figure 1. Thus, from a conceptual point of view, there is no difference in the signal generation between windowed CP-OFDM and FBMC.

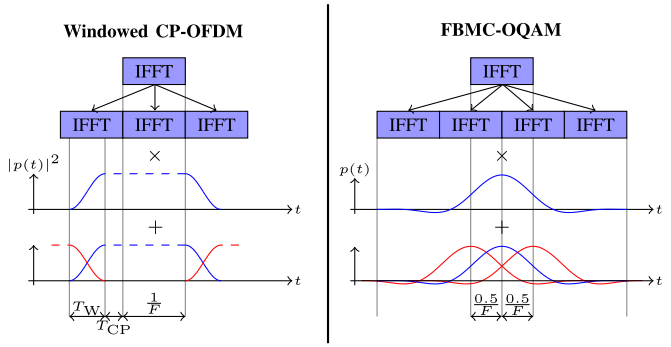

Fig. 1. From a conceptual point of view the signal generation in windowed OFDM and FBMC requires the same basic operations, namely, an IFFT, copying the IFFT output O-times, element-wise multiplication with the prototype filter and, finally, overlapping [26], [27]. The receiver works in a similar way, but in reversed order. ©2017 IEEE, [2].

One can easily transform OFDM into FBMC simply by, 1) Removing the CP. 2) Changing the window function. 3) Reducing the time-spacing from $T = T_{\mathrm{W}} + T_{\mathrm{CP}} + \frac{1}{F}$ to $T = \frac{0.5}{F}$. 4) Transmitting only real-valued data symbols and fulfilling the phase pattern $\mathrm{e}^{\mathrm{j}\frac{\pi}{2}(l+k)}$. This observation will later be used in our pruned DFT spread FBMC transmission scheme. Note that in Figure 1 we consider an overlapping factor of four (FBMC), a common value in literature. However, the overlapping factor must not necessarily be an integer, implying that at the edges only some samples of the IFFT output are copied. For example windowed CP-OFDM employs an overlapping factor of $O = 2\ T_{\mathrm{W}}F + T_{\mathrm{CP}}F + 1$. More generally, we utilize the overlapping factor to describe how much longer the basis pulses are relative to the reference time-period of $\frac{1}{F}$. The receiver works in a similar way as the transmitter, but in reversed order, that is, 1) Element-wise multiplication of the received signal by the prototype filter. 2) Summing up samples, corresponding to the copy operation at the transmitter. 3) An N_{FFT} point FFT.

III. Pruned DFT Spread FBMC

In Section III-A, we present the underlying idea of pruned DFT spread FBMC and provide an intuitive explanation why the PAPR is reduced and why complex orthogonality is (approximately) restored. Section III-B then includes a more rigorous mathematical description. In Section III-C, we discuss latency aspects and in Section III-D the computational complexity.

A. Basic Idea

The basic idea of pruned DFT spread FBMC can be best explained by the underlying basis pulses. Precoding by $\mathbf{C}$ transforms the basis pulses $g_{l,k}(t)$ into $\tilde{g}_i(t)$, described by $\tilde{\mathbf{G}} = \mathbf{G}\mathbf{C} = \left[\tilde{\mathbf{g}}_1\ \tilde{\mathbf{g}}_2\ \ldots\right]$. Note that the size of $\mathbf{C}$ and thus the number of new basis pulses $\tilde{g}_i(t)$ depends on the precoding method. Figure 2 shows the power of the basis pulses and illustrates a step by step construction of our method, starting from a conventional OFDM system. Figure 2 (a) shows OFDM [29] for $N_{\mathrm{FFT}} = 512$, $L = 16$ and $K = 1$. The underlying basis pulses are frequency shifted rectangular functions. In terms of

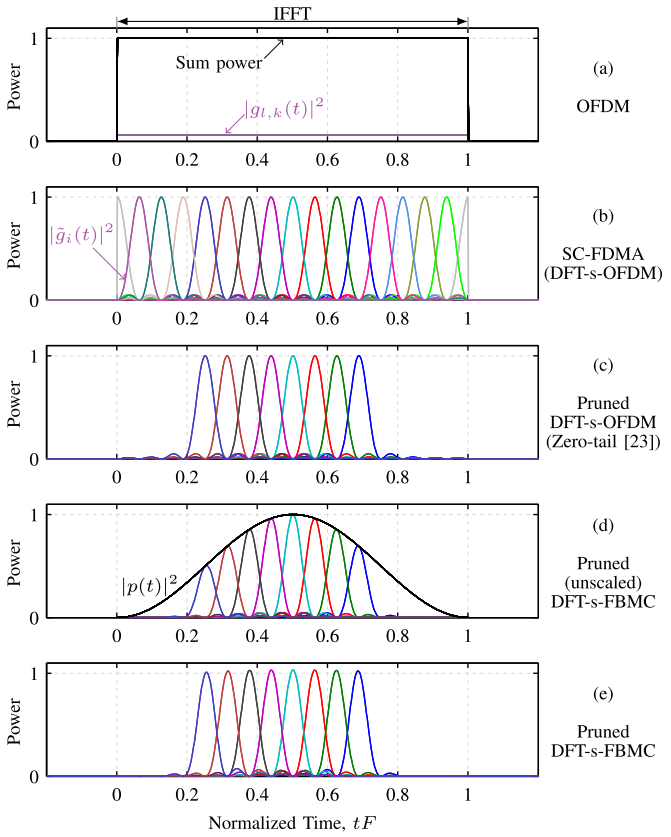

Fig. 2. Power of the underlying basis pulses in time, that is, $|g_{l,k}(t)|^2$ and $|\tilde{g}_i(t)|^2$ for $N_{\mathrm{FFT}} = 512$, $L = 16$ and $K = 1$; (a) conventional OFDM; (b) conventional SC-FDMA, that is, precoding by DFT matrix $\mathbf{W}_L$; (c) only $L/2 = 8$ basis pulses, close to the center, are utilized, that is, $\mathbf{W}_L$ is replaced by a pruned DFT matrix, $\widetilde{\mathbf{W}}_{L\times L/2}$; (d) multiplication by a window/prototype filter $p(t)$ so that OFDM transforms into FBMC; (e) one-tap scaling of the basis pulses so that the transmit power is approximately constant over time.

transmit power, however, a frequency shift has no influence so that only one basis pulse can be observed in Figure 2 (a). All the basis pulses are added together with some random weights (the data symbols), so that, according to the central limit theorem, the signal distribution at one time sample approaches a Gaussian distribution. This explains the poor PAPR of OFDM. In SC-FDMA [30], see Figure 2 (b), DFT precoding by $\mathbf{W}_L$ transforms the basis pulses of a conventional OFDM system in such a way that a single carrier transmission is emulated. In particular, the basis pulses are more localized in time and even though they still overlap the signal at one time sample is mainly determined by 1-2 basis pulses. Thus, as long as the data symbols are not Gaussian distributed but chosen from a limited symbol alphabet such as a Quadrature Amplitude Modulation (QAM) signal constellation, the PAPR will be better than in OFDM. Unfortunately, SC-FDMA has the same poor OOB emissions as OFDM. This can easily be deduced by considering the transmitted signal at the edge positions, that is, $tF = 0$ and $tF = 1$. Similar as in OFDM, the underlying rectangular pulse cuts through the signal so that, at the edges, the signal abruptly jumps to zero without a smooth transition. Only basis pulses close to the edge positions are affected by this cutting effect. Thus, setting

the edge basis pulses to zero reduces the OOB emissions and is indeed the basic idea of zero-tail DFT-spread-OFDM [23], [31]. However, Berardinelli *et al.* [23], [31] remove only a few basis pulses to keep the overhead low. We, on the other hand, remove $L/2$ basis pulses from the set, that is, DFT spreading matrix $\mathbf{W}_L$ is replaced by a pruned DFT matrix $\mathbf{W}_{L\times L/2}$. This step can also be interpreted as setting half the input samples of a conventional DFT to zero. In contrast to zero-tail DFT-spread-OFDM, our method does not impose any overhead because we also reduce the time spacing by a factor of two, as typically done in FBMC-OQAM. This also explains why we remove exactly $L/2$ basis pulses. The result of our approach is shown in Figure 2 (c). To combat multi-path delays, zero-tail DFT-spread-OFDM utilizes the zero-tail overhead in a similar way as the CP in OFDM. This reduces the spectral efficiency. Again, we choose a different approach, namely, we transform the OFDM system into an FBMC system so that the channel induced interference becomes very low and can often be neglected [2]. As discussed in Section II, an OFDM system can easily be transformed into an FBMC system simply by multiplying the IFFT output with a prototype filter $p(t)$, as shown in Figure 2 (d). In the last step, see Figure 2 (e), the individual basis pulses are scaled up so that the sum transmit power is approximately constant over the transmission time. This final step completes our novel pruned DFT spread FBMC transmission scheme. Figure 2 also explain why complex orthogonality is approximately restored. To be specific, DFT spreading reduces the time duration of the underlying basis pulses so that each basis pulse experiences an approximately flat prototype filter. Such system reflects a conventional SC-FDMA transmission and is clearly orthogonal. Orthogonality relies on the approximation that the prototype filter is flat over the duration of the basis pulse. This approximation becomes tight for $L \to \infty$ because the time duration of each individual basis pulse approaches zero. However, in practical systems, this will not be the case. In Section IV we calculate the orthogonality approximation error and propose an additional frequency CP to reduce it.

B. Mathematical Details

In Figure 2 (d) we employ a time domain root-raised cosine pulse with roll-off factor one, that is,

$$p_{\mathrm{tRRC}}(t)=\begin{cases}\sqrt{F\left(\cos(2\pi tF)+1\right)} & \text{if } -\frac{1}{2F}\leq t<\frac{1}{2F}\\ 0 & \text{otherwise,}\end{cases} \tag{10}$$

because it corresponds to an overlapping factor of $O = 1$, greatly simplifying the illustration. However, we can also employ a larger overlapping factor, implying that the IFFT output has to be copied similar as in Figure 1. From Figure 2 (c) we deduce that the overlapping factor should not be larger than $O \approx 1.5$ in order to avoid interference between symbols in time, caused by the IFFT repetition in FBMC. Thus, the PHYDYAS prototype filter [32] is not suited for our transmission scheme because of its poor time localization. Instead, we employ a truncated Hermite prototype filter. The Hermite prototype filter was suggested in [33] and is based on Hermite polynomials $H_i\{\cdot\}$. It can be expressed by

$$p_{\mathrm{Herm.}}(t) = \sqrt{F}\mathrm{e}^{-2\pi(tF)^2}\sum_{\substack{i=\{0,4,8,\\12,16,20\}}}\alpha_i H_i\{2\sqrt{\pi}tF\}, \tag{11}$$

where the coefficients α_i can be found in [2]. The Hermite pulse is based on a Gaussian function and therefore has a very good joint time-frequency localization of $\sigma_t\sigma_f = 1.02\times 1/4\pi$, almost as good as the bound of $\sigma_t\sigma_f \geq 1/4\pi$ (attained by the Gaussian pulse) and much better than the PHYDYAS prototype filter ($\sigma_t\sigma_f = 1.13 \times 1/4\pi$). This also explains why the Hermite pulse is more robust in doubly-selective channel [2] and, in particular, in time-variant channels [34]. As the overlapping factor should not be larger than $O \approx 1.5$, we set the Hermite pulse in (11) to zero after the first zero-crossing, that is,

$$p_{\mathrm{Herm.Trunc.}}(t) = \begin{cases}p_{\mathrm{Herm.}}(t) & \text{if } -\frac{1.56}{2F}\leq t<\frac{1.56}{2F}\\ 0 & \text{otherwise.}\end{cases} \tag{12}$$

In contrast to (10) and (11), employing a truncated Hermite prototype filter, see (12), no longer guarantees real orthogonality in FBMC, $\Re\{\mathbf{G}^{\mathsf{H}}\mathbf{G}\} \neq \mathbf{I}_{LK}$. Instead, one observes an Signal-to-Interference Ratio (SIR) of 28 dB. This, however, has no direct influence on our transmission scheme. Moreover, our transmission method can utilize any window function as long as the overlapping factor is between approximately 0.6 and 1.6. Of course, there exists a trade-off between latency, OOB emissions, robustness to doubly selective channels, and the achievable SIR. If not stated otherwise we will always employ a truncated Hermite prototype filter because our investigations have shown that it offers a good trade-off between the relevant factors. However, a more detailed discussion on the optimal trade-off for specific use-cases and the subsequent optimal filter design could be investigated in future works. For conventional FBMC we rely on the Hermite prototype filter, see (11).

Let us now discuss the optimal size of precoding matrix $\mathbf{C}$. Figure 2 already provides an intuitive explanation why only $L/2$ basis pulses (per time-position) are employed, namely, the reduction of the time-spacing in FBMC by a factor of two (compared with OFDM). However, there exists also a more formal explanation based on an eigenvalue decomposition. Similar as in [5], [6], [9], and [10], our goal is to restore complex orthogonality in FBMC, that is, $\mathbf{C}^{\mathsf{H}}\mathbf{G}^{\mathsf{H}}\mathbf{G}\mathbf{C} = \mathbf{I}$. As for the derivation of the MIMO channel capacity [35], the optimal precoding matrix can be found by an eigenvalue decomposition of $\mathbf{G}^{\mathsf{H}}\mathbf{G}$ in combination with water-filling. Ignoring any edge effects, $\mathbf{G}^{\mathsf{H}}\mathbf{G}$ has exactly $LK/2$ non-zero eigenvalues, each having the same value [2]. Thus, the optimal size of precoding matrix $\mathbf{C}$ is $LK \times \frac{LK}{2}$. Because we spread in frequency only, that is, $\mathbf{C} = \mathbf{I}_K \otimes \mathbf{C}_f$, the optimal frequency spreading matrix must have a size of $\mathbf{C}_f \in \mathbb{C}^{L\times\frac{L}{2}}$.

Figure 3 shows a block diagram of our pruned DFT spread FBMC transmission scheme. We spread $L/2$ complex-valued data symbols, $\tilde{\mathbf{x}}_k \in \mathbb{C}^{\frac{L}{2}\times 1}$, assumed to be uncorrelated and with unit power, over L subcarriers, so that the transmitted symbols for FBMC at time position k become

$$\mathbf{x}_k = \mathbf{C}_f\,\tilde{\mathbf{x}}_k, \tag{13}$$

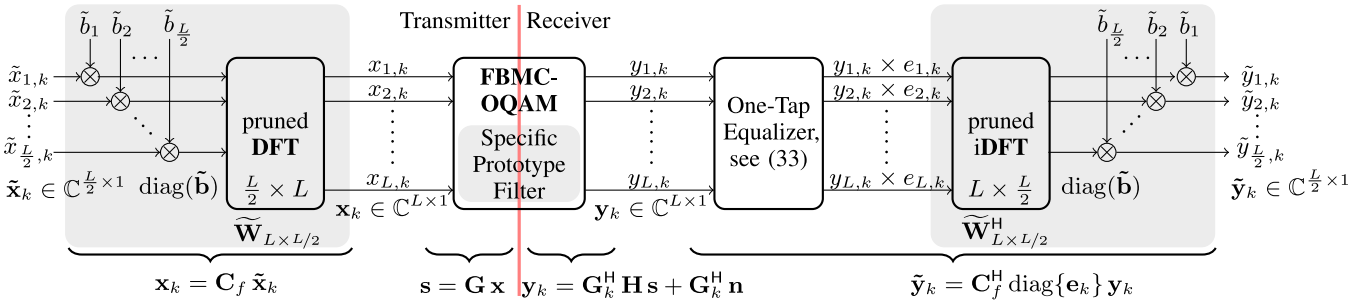

Fig. 3. Block diagram of pruned DFT spread FBMC at time position k. Compared to conventional FBMC-OQAM transmissions the complex-to-real transformation is replaced by precoding with $\mathbf{C}_f$, the real-to-complex transformation by $\mathbf{C}_f^{\mathsf{H}}$, and the protoype filter must be shorter in time than approximately $1.5/F$. Note that in conventional FBMC-OQAM the data symbols $\mathbf{x}_k \in \mathbb{R}^{L\times 1}$ are real-valued.

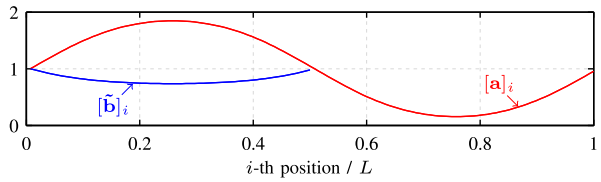

Fig. 4. Equalizing $L/2$ largest elements of $\mathbf{a}$ delivers the scaling values $\tilde{\mathbf{b}}$. These scaling values guarantee that the transmit power is approximately constant over time and that the diagonal elements of $\mathbf{C}_f^{\mathsf{H}}\mathbf{G}_k^{\mathsf{H}}\mathbf{G}_k\mathbf{C}_f$ are exactly one. Note that element $[\mathbf{a}]_i$ corresponds to the i-th column of $\mathbf{W}$.

with $\mathbf{C}_f \in \mathbb{C}^{L\times\frac{L}{2}}$ denoting the frequency spreading matrix. Note that, in contrast to conventional FBMC, the transmitted symbols are no longer real-valued but complex-valued. The received data symbols $\tilde{\mathbf{y}}_k \in \mathbb{C}^{\frac{L}{2}\times 1}$ are obtained by one-tap equalization of the received symbols with $\mathbf{e}_k \in \mathbb{C}^{L\times 1}$, followed by despreading according to

$$\tilde{\mathbf{y}}_k = \mathbf{C}_f^{\mathsf{H}}\,\mathrm{diag}\{\mathbf{e}_k\}\,\mathbf{y}_k. \tag{14}$$

For the derivation of spreading matrix $\mathbf{C}_f$ we assume an Additive White Gaussian Noise (AWGN) channel, $\mathbf{H} = \mathbf{I}_N$, for which no equalization is necessary. The ultimate goal is to restore complex orthogonality, that is,

$$\mathbf{C}_f^{\mathsf{H}}\mathbf{G}_k^{\mathsf{H}}\mathbf{G}_k\mathbf{C}_f \approx \mathbf{I}_{L/2}. \tag{15}$$

The approximation symbol in (15) indicates that a small residual interference remains, see Section IV, so that our system is only quasi orthogonal. However, in many cases this has no impact on the performance.

As already explained in Section III-A, frequency spreading matrix $\mathbf{C}_f \in \mathbb{C}^{L\times\frac{L}{2}}$ consists of a pruned DFT in combination with one-tap scaling, that is,

$$\mathbf{C}_f = \widetilde{\mathbf{W}}_{L\times L/2}\,\mathrm{diag}\{\tilde{\mathbf{b}}\}, \tag{16}$$

with pruned DFT matrix $\widetilde{\mathbf{W}}_{L\times L/2} \in \mathbb{C}^{L\times\frac{L}{2}}$ and one-tap scaling vector $\tilde{\mathbf{b}} \in \mathbb{R}^{\frac{L}{2}\times 1}$. To further describe $\widetilde{\mathbf{W}}_{L\times L/2}$ and $\tilde{\mathbf{b}}$, we utilize an auxiliary vector $\mathbf{a} \in \mathbb{R}^{L\times 1}$, defined as,

$$\mathbf{a} = \mathrm{diag}\{\mathbf{W}_L^{\mathsf{H}}\mathbf{G}_k^{\mathsf{H}}\mathbf{G}_k\mathbf{W}_L\}, \tag{17}$$

which implicitly assumes spreading and despreading by a full DFT matrix $\mathbf{W}_L \in \mathbb{C}^{L\times L}$. Figure 4 shows how $[\mathbf{a}]_i$ depends on position i. The i-th element of $\mathbf{a}$ corresponds to the i-th column of $\mathbf{W}_L$. The main idea of our transmission scheme is to utilize only those column vectors of $\mathbf{W}_L$ which correspond to the $L/2$ largest elements of $\mathbf{a}$. Thus, we only employ the first $L/2$ column vectors of $\mathbf{W}_L$, see Figure 4. Furthermore, we perform pre-equalization of $[\mathbf{a}]_i$. Pruned DFT matrix $\widetilde{\mathbf{W}}_{L\times L/2} \in \mathbb{C}^{L\times\frac{L}{2}}$ and scaling vector $\tilde{\mathbf{b}} \in \mathbb{R}^{\frac{L}{2}\times 1}$ in (16) can therefore be expressed as

$$[\tilde{\mathbf{b}}]_i = \sqrt{\frac{1}{[\mathbf{a}]_i}}, \quad \text{for} \quad i = 1\ldots\frac{L}{2}, \tag{18}$$

$$\widetilde{\mathbf{W}}_{L\times L/2} = \mathbf{W}_L\begin{bmatrix}\mathbf{I}_{\frac{L}{2}}\\ \mathbf{0}_{\frac{L}{2}}\end{bmatrix}. \tag{19}$$

Note that $\tilde{\mathbf{b}}$ in (18) guarantees that the diagonal elements of (15) are exactly one. Furthermore, (18) and (19) depend on the underlying IFFT (it makes a difference if the overlapping factor is even or odd, see Figure 1). Thus, (18) and (19) do not always correspond to the first $L/2$ positions and one might have to rely on auxiliary vector $\mathbf{a}$ to find the correct positions.

C. Latency

Figure 5 shows the expected transmit power in time for one FBMC symbol, $\mathbf{p}_k^{(t)} \in \mathbb{R}^{N\times 1}$, calculated by

$$\mathbf{p}_k^{(t)} = \mathrm{diag}\{\mathbf{G}_k\mathbf{C}_f\,\mathbf{C}_f^{\mathsf{H}}\mathbf{G}_k^{\mathsf{H}}\}. \tag{20}$$

Compared with Figure 2 we now employ $L = 128$ subcarriers. In conventional FBMC there exists a large overlapping of symbols in time and the transmission requires a long ramp-up and ramp-down period. In pruned DFT spread FBMC, on the other hand, precoding by $\mathbf{C}_f$ shapes the transmitted signal in such a way that the overlapping in time is very low and the ramp-up and ramp-down period dramatically reduced. This allows us to reduce the overlapping factor, for example to $O = 0.8$ in combination with a Tukey window (parameter $\beta = 0.6$). In conventional FBMC this is not feasible because such windowing operation would reduce the SIR to 13 dB. In pruned DFT spread FBMC, on the other hand, the main energy is concentrated within $O = 0.5$. Thus, reducing the overlapping factor to $O = 0.8$ has only a minor influence on the SIR. To be specific, the SIR is only 1 dB lower compared to $O = 1.56$. In Section IV we show how to calculate the SIR. Note that a similar precoding effect as shown in Figure 5 was

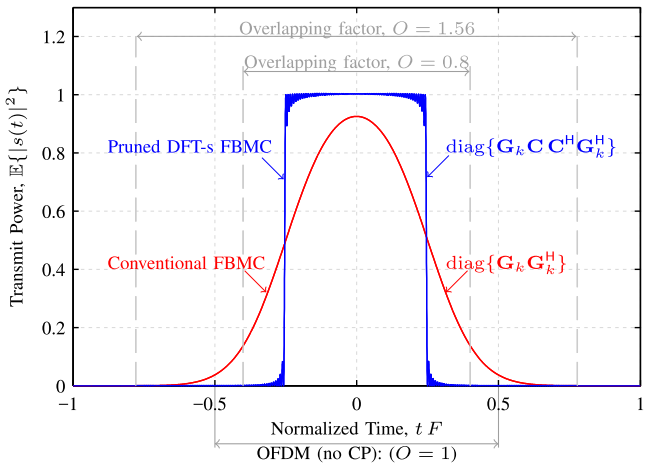

Fig. 5. Precoding matrix $\mathbf{C}_f$ shapes the transmitted signal in such a way, that the average transmit power shows an almost perfect rectangular shape with many beneficial properties.

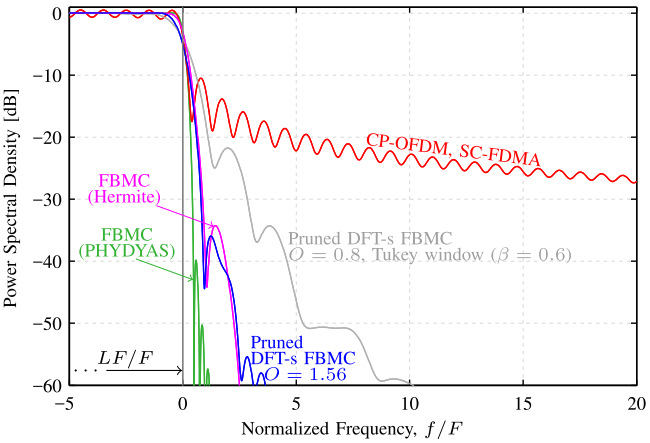

Fig. 6. The superior spectral confinement of FBMC is preserved in pruned DFT Spread FBMC. Reducing the overlapping factor, see Figure 5, reduces the latency but also increases the OOB emissions. Thus, there is a trade-off.

also observed in FFT-FBMC [36], but in the frequency domain instead of the time domain.

Reducing the overlapping factor increases the OOB emissions. To describe this effect we calculate the power spectral density, $\mathbf{p}_k^{(f)} \in \mathbb{R}^{N\times 1}$, by

$$\mathbf{p}_k^{(f)} = \mathrm{diag}\{\mathbf{W}_N\mathbf{G}_k\mathbf{C}_f\mathbf{C}_f^{\mathsf{H}}\mathbf{G}_k^{\mathsf{H}}\mathbf{W}_N^{\mathsf{H}}\}, \tag{21}$$

where we ignore scaling (we later normalize to 0 dB) and implicitly assume infinitely many repetitions in time. The frequency resolution in (21) is $\Delta f = \frac{f_s}{N}$. As shown in Figure 6 the OOB emissions of pruned DFT spread FBMC are relatively low, even for an overlapping factor of $O = 0.8$. For an overlapping factor of $O = 1.56$ the OOB emissions of our method are comparable to conventional FBMC transmissions (Hermite prototype filter) and much better than in OFDM. As a reference we also consider the PHYDYAS prototype filter which has the lowest OOB emissions, but also the worst time-localization.

Let us now quantify the latency in more detail, where we focus on the underlying pulse duration but ignore other sources of delays such as channel delay or processing delay. The transmission time of one FBMC symbol depends on overlapping factor O and requires a time length of $\frac{O}{F}$. However, one FBMC symbol only carries half the information of that of an OFDM symbol. Thus, we might need to include the second symbol, leading to an additional delay of $\frac{0.5}{F}$, that is, the time spacing. For example, an overlapping factor of $O = 0.8$ implies that the first half of the information is received 20% faster than in OFDM (no CP) while the second half needs 30% longer. A conventional FBMC transmission with $O = 4$, on the other hand, requires 350% longer than OFDM (no CP). A similar behavior can be observed by considering an FBMC transmission block, consisting of K symbols in time, for which the delay is given by

$$T_{\mathrm{Block}} = \frac{O}{F} + \frac{0.5}{F}(K-1). \tag{22}$$

In LTE, a block (subframe) has a duration of 1 ms and consists of $K = 14$ OFDM symbols ($T_{\mathrm{CP}} = \frac{1}{14F}$ and $F = 15\,\mathrm{kHz}$). Because each FBMC symbol only carries half the information of that of an OFDM symbol (same number of subcarriers), we require in total $K = 28$ FBMC symbols for a fair comparison to LTE. For an overlapping factor of $O = 0.8$, this implies that our method has a transmission time of $T_{\mathrm{Block}} \approx 0.95\,\mathrm{ms}$, faster than in LTE. For an overlapping factor of $O = 1.5$, the transmission time is exactly 1 ms, same as in LTE. Conventional FBMC ($O = 4$), on the other hand, performs relatively poor and requires 1.2 ms, 20% longer than LTE. Note that the ramp-up and ramp-down period in FBMC increases the latency but not necessarily the sum throughput of the whole system because different transmission blocks can overlap in time. This works as long as the phase pattern which shifts the intrinsic interference to the imaginary domain is fulfilled, as typically the case in downlink transmissions. However, in multi-user uplink transmissions, different users experience different phase shifts. Thus, the required phase pattern is violated and a guard time might be necessary. In such cases, the ramp-up and ramp-down period not only increases the latency but also reduces the sum throughput. In pruned DFT spread FBMC, on the other hand, this is not an issue because we restore complex orthogonality and are therefore not affected by any phase shifts.

D. Computational Complexity

The computational complexity of our transmission scheme is approximately two times higher than in conventional SC-FDMA. To be specific, the computational complexity relative to SC-FDMA (without CP) at the transmitter can be approximated by

$$\frac{2\left(\frac{L}{2} + L\log\frac{L}{2} + N_{\mathrm{FFT}}\log N_{\mathrm{FFT}} + ON_{\mathrm{FFT}}\right)}{L\log L + N_{\mathrm{FFT}}\log N_{\mathrm{FFT}}} \approx 2. \tag{23}$$

If we ignore channel equalization (23) also represents the computational complexity at the receiver. The term $N_{\mathrm{FFT}}\log N_{\mathrm{FFT}}$ corresponds to the IFFT, required for both, FBMC and OFDM. Additionally, FBMC requires an element-wise multiplication by the prototype filter, see Figure 1, leading to an additional complexity of ON_{FFT}. DFT spreading in

SC-FDMA has a complexity of $L \log L$, while one-tap scaling in combination with a pruned DFT requires approximately $\frac{L}{2} + L \log \frac{L}{2}$ multiplications [37]. Finally, the reduced time-spacing in FBMC implies that all the calculations have to be applied two times as often as in SC-FDMA (no CP), explaining the factor of two in the nominator. For example an LTE like setup with $N_{\text{FFT}} = 1024$ and $L = 600$ implies for $O = 1.56$ that the computational complexity of our scheme is approximately $^{34149}/_{15777} \approx 2.16$ times higher than in SC-FDMA. For $O = 0.8$, it is $^{32593}/_{15777} \approx 2.07$.

IV. Orthogonality Approximation

In Section III we have argued that complex orthogonality is approximately restored. In this section we quantify this approximation by considering the SIR.

In contrast to (15) we are not only interested in orthogonality within one time position, but in orthogonality within the whole block, that is,

$$\mathbf{C}^{\mathsf{H}}\mathbf{G}^{\mathsf{H}}\mathbf{G}\mathbf{C} \approx \mathbf{I}_{LK/2}, \tag{24}$$

for which overall spreading matrix $\mathbf{C} \in \mathbb{C}^{LK \times \frac{LK}{2}}$ is given by

$$\mathbf{C} = \mathbf{I}_K \otimes \mathbf{C}_f. \tag{25}$$

The Kronecker product $\otimes$ maps frequency spreading matrix $\mathbf{C}_f \in \mathbb{C}^{L \times \frac{L}{2}}$, see (16), to the correct time-positions, that is, $\text{vec}\{\mathbf{C}_f \left[\tilde{\mathbf{x}}_1 \ \ldots \ \tilde{\mathbf{x}}_K\right]\} = (\mathbf{I}_K \otimes \mathbf{C}_f)\tilde{\mathbf{x}}$. The diagonal elements of $\mathbf{C}^{\mathsf{H}}\mathbf{G}^{\mathsf{H}}\mathbf{G}\mathbf{C}$ in (24) are exactly one because of one-tap scaling with $\mathbf{b}$. However, the off-diagonal elements of $\mathbf{C}^{\mathsf{H}}\mathbf{G}^{\mathsf{H}}\mathbf{G}\mathbf{C}$ are not exactly zero implying that symbols interfere with each other. We quantify this interference with the SIR, given for time-position k by

$$\text{SIR}_k^{\text{Orth.Appr.}} = \frac{\frac{L}{2}}{||\mathbf{C}_f^{\mathsf{H}}\mathbf{G}_k^{\mathsf{H}}\mathbf{G}\mathbf{C}||_{\mathsf{F}}^2 - \frac{L}{2}}, \tag{26}$$

where $||\cdot||_{\mathsf{F}}$ denotes the Frobenius norm and $L/2$ reflect the sum power of the diagonal elements of $\mathbf{C}_f^{\mathsf{H}}\mathbf{G}_k^{\mathsf{H}}\mathbf{G}_k\mathbf{C}_f$. The SIR is the same for all time positions except those at the beginning and the end of the block. The blue curve in Figure 7 shows the SIR. Not all symbol experience exactly the same SIR so that we also include the maximum and the minimum SIR (within one time position), indicated by the dotted lines in Figure 7. For a large number of subcarriers the SIR is sufficiently high, allowing us to neglect the interference because it is dominated by noise. However, in some rare cases we might need a higher SIR. One can then, for example, reduce the subcarrier spacing while keeping the bandwidth constant, resulting in a higher number of subcarriers and thus a higher SIR. Another way of increasing the SIR is to employ a frequency CP,[1] that is, a cyclical extension of the signal in the frequency domain. The drawback is a small reduction in spectral efficiency. The idea of a frequency CP is based on a critically sampled system, that is, $N_{\text{FFT}} = L$. In pruned DFT spread FBMC a critically sampled system perfectly restores

[1]We call it frequency CP, even though it is a cyclic prefix and cyclic suffix.

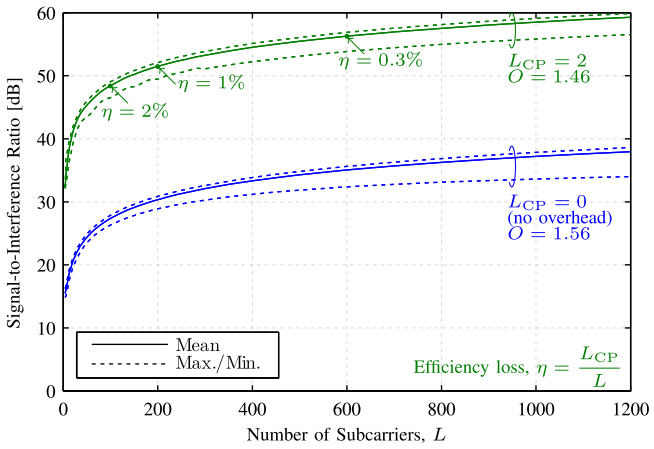

Fig. 7. Complex orthogonality is not perfectly restored, causing interference. We quantify this interference with the SIR. Note that the SIR is high enough so that interference can usually be neglected. Only if the number of subcarriers is very small, a frequency CP might be necessary, see Figure 8, slightly reducing the spectral efficiency.

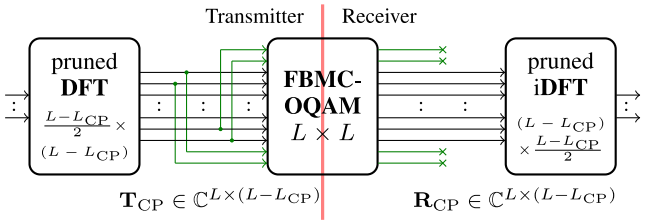

Fig. 8. A frequency CP emulates a critically sampled system by repeating a small part of the signal in the frequency domain. This greatly improves the SIR, see Figure 7. However, in many practically relevant cases a frequency CP is not necessary.

complex orthogonality in (24), that is, $\mathbf{C}^{\mathsf{H}}\mathbf{G}^{\mathsf{H}}\mathbf{G}\mathbf{C} = \mathbf{I}_{LK/2}$. This can easily be shown with Figure 1 and Figure 2. The IFFT in FBMC and DFT precoding perfectly cancel each other, that is, $\mathbf{W}_L^{\mathsf{H}}\mathbf{W}_L = \mathbf{I}_L$, so that each data symbol corresponds to one sample in time. With respect to a pruned DFT, we only utilize $L/2$ time samples in the center, see Figure 2 (c). If $O < 1.5$, the IFFT repetition in FBMC, see Figure 1, has no influence on the transmission. Furthermore, the scaling values $\mathbf{b}$ perfectly equalize the prototype filter. Thus, the overall system is characterized by transmitting one data symbol at one time sample, that is, $\mathbf{s} = \tilde{\mathbf{x}} = \tilde{\mathbf{y}}$ (back-to-back transmission, no noise and ignoring the "ramp-up" and "ramp-down" period). Of course such a system makes no sense in practice. However, one can emulate a critically sampled system by repeating the signal in the frequency domain. Because the basis pulses in FBMC are localized in frequency, it is usually sufficient to repeat only a few subcarriers at the edges. Figure 8 illustrates this process for a frequency CP length of $L_{\text{CP}} = 4$. If L_{CP} is an even number the frequency CP at the transmitter can be described by matrix $\mathbf{T}_{\text{CP}} \in \mathbb{C}^{L \times (L-L_{\text{CP}})}$, defined as,

$$\mathbf{T}_{\text{CP}} = \begin{bmatrix} \begin{bmatrix} \mathbf{0}_{(L_{\text{CP}}/2) \times (L-3L_{\text{CP}}/2)} & \mathbf{I}_{L_{\text{CP}}/2} \end{bmatrix} \\ \mathbf{I}_{L-L_{\text{CP}}} \\ \begin{bmatrix} \mathbf{I}_{L_{\text{CP}}/2} & \mathbf{0}_{(L_{\text{CP}}/2) \times (L-3L_{\text{CP}}/2)} \end{bmatrix} \end{bmatrix}, \tag{27}$$

and the CP reduction at the receiver by $\mathbf{R}_{\mathrm{CP}} \in \mathbb{C}^{L\times(L-L_{\mathrm{CP}})}$, defined as

$$\mathbf{R}_{\mathrm{CP}} = \begin{bmatrix} \mathbf{0}_{(L_{\mathrm{CP}}/2)\times(L-L_{\mathrm{CP}})} \\ \mathbf{I}_{L-L_{\mathrm{CP}}} \\ \mathbf{0}_{(L_{\mathrm{CP}}/2)\times(L-L_{\mathrm{CP}})} \end{bmatrix}. \tag{28}$$

Compared to the previous coding matrix in (16), the new precoding matrices, $\mathbf{C}_{f,\mathrm{TX}} \in \mathbb{C}^{L\times\frac{L-L_{\mathrm{CP}}}{2}}$ at the transmitter and $\mathbf{C}_{f,\mathrm{RX}} \in \mathbb{C}^{L\times\frac{L-L_{\mathrm{CP}}}{2}}$ at the receiver, change according to

$$\mathbf{C}_{f,\mathrm{TX}} = \mathbf{T}_{\mathrm{CP}}\widetilde{\mathbf{W}}_{\mathrm{CP}}\,\mathrm{diag}\{\tilde{\mathbf{b}}_{\mathrm{CP}}\} \tag{29}$$

$$\mathbf{C}_{f,\mathrm{RX}} = \mathbf{R}_{\mathrm{CP}}\widetilde{\mathbf{W}}_{\mathrm{CP}}\,\mathrm{diag}\{\tilde{\mathbf{b}}_{\mathrm{CP}}\}. \tag{30}$$

Furthermore, pruned DFT matrix $\widetilde{\mathbf{W}}_{\mathrm{CP}} \in \mathbb{C}^{(L-L_{\mathrm{CP}})\times\frac{L-L_{\mathrm{CP}}}{2}}$ and scaling vector $\tilde{\mathbf{b}}_{\mathrm{CP}} \in \mathbb{R}^{\frac{L-L_{\mathrm{CP}}}{2}\times 1}$ now also have different dimensions. The process of finding $\widetilde{\mathbf{W}}_{\mathrm{CP}}$ and $\tilde{\mathbf{b}}_{\mathrm{CP}}$, however, is the same as in Section III-B and depends on auxiliary vector

$$\mathbf{a}_{\mathrm{CP}} = \mathrm{diag}\{\mathbf{W}^{\mathsf{H}}_{(L-L_{\mathrm{CP}})}\mathbf{R}^{\mathsf{H}}_{\mathrm{CP}}\mathbf{G}^{\mathsf{H}}_{k}\mathbf{G}_{k}\mathbf{T}_{\mathrm{CP}}\mathbf{W}_{(L-L_{\mathrm{CP}})}\}. \tag{31}$$

The new coding matrices in (29) and (30) also change the SIR calculation in (26) according to

$$\mathrm{SIR}_{k}^{\mathrm{Orth.Appr.CP}} = \frac{\frac{L-L_{\mathrm{CP}}}{2}}{||\mathbf{C}^{\mathsf{H}}_{f,\mathrm{RX}}\mathbf{G}^{\mathsf{H}}_{k}\mathbf{G}(\mathbf{I}_K\otimes\mathbf{C}_{f,\mathrm{TX}})||^2_{\mathsf{F}} - \frac{L-L_{\mathrm{CP}}}{2}} \tag{32}$$

Figure 7 shows that a frequency CP of length $L_{\mathrm{CP}} = 2$ can significantly improve the SIR by approximately 20 dB, while the spectral efficiency loss, given by $\eta = \frac{L_{\mathrm{CP}}}{L}$, is relatively small. Note that for $L_{\mathrm{CP}} = 2$ we reduce the overlapping factor to $O = 1.46$ because it increases the SIR further and has almost no influence on the spectral confinement (only for values smaller than -50 dB in Figure 6).

While a frequency CP can improve the SIR, our investigations have shown that in most cases it is not necessary because interference is dominated by noise. Even if a frequency CP is employed, the efficiency loss is often less than 1%. Only in the rare case when the number of data subcarriers is very small, say for example 12, a frequency CP causes a relatively high spectral efficiency loss of $\eta = \frac{L_{\mathrm{CP}}}{L} = \frac{2}{12+2} = 14\,\%$. Conventional LTE only has a loss of 7%, caused by the CP. However, this statement only holds true for perfectly synchronized systems. In asynchronous transmissions guard subcarriers are often required. Suppose for example that the power spectral density must be below 30 dB of its maximum value outside the effective transmission bandwidth (including the guard band). For pruned DFT spread FBMC this implies that one guard subcarrier is required, see Figure 6, already fulfilled by the frequency CP so that the efficiency loss remains at 14 %. Universal-Filtered OFDM (UF-OFDM) [38], on the other hand, requires $L_{\mathrm{G}} = 8$ guard subcarriers, leading to an efficiency loss [2] of $\eta = TF + TF\frac{L_{\mathrm{G}}}{L} - 1 = 1.07 + 1.07\frac{8}{20} - 1 = 50\,\%$. Thus, in asynchronous transmissions pruned DFT spread FBMC is much more efficient than UF-OFDM, even if a frequency CP is employed. Only when compared to conventional FBMC transmissions the frequency CP imposes a minor issue, as conventional FBMC only requires one guard subcarrier, causing an efficiency loss of $\eta = \frac{L_{\mathrm{CP}}}{L} = \frac{1}{13} = 8\,\%$.

Because a frequency CP is often not necessary and to keep the notation consistent with the block diagram in Figure 3 we ignore the frequency CP in the next two subsections. However, by replacing $\mathbf{C}$ with $(\mathbf{I}_K \otimes \mathbf{C}_{f,\mathrm{TX}})$ at the transmitter and $\mathbf{C}^{\mathsf{H}}$ by $(\mathbf{I}_K \otimes \mathbf{C}_{f,\mathrm{RX}})^{\mathsf{H}}$ at the receiver, all equations are also valid for the frequency CP case.

V. One-Tap Equalizers

So far we have ignored the channel but will now include it into our considerations. With respect to channel equalization we restrict ourself to one-tap equalizers because of the low computational complexity. Note, however, that multi-tap equalizers [34] or a sliding window frequency domain equalizer [39] could improve the performance further. In Section V-A we derive a closed form solution for the Signal-to-Interference plus Noise Ratio (SINR) in doubly-selective channels. Section V-B then considers the special case of a "flat" channel, allowing to straightforwardly employ all MIMO methods.

A. SINR in Doubly-Selective Channels

One of the main advantages of DFT precoded OFDM and FBMC systems compared with conventional single carrier schemes is that the channel equalization can be performed in the multi-carrier domain. Thus, low-complexity one-tap equalizers can be used. This is a crucial aspect of pruned DFT spread FBMC so that we will discussion the effect of one-tap equalizers on the SINR in more detail.

We employ a scaled one-tap Minimum Mean Squared Error (MMSE) equalizer, given for subcarrier position l and time position k by,

$$e_{l,k} = \frac{h^*_{l,k}}{|h_{l,k}|^2 + P_{\mathrm{n}}} \; \frac{1}{\frac{1}{L}\sum_{l=1}^{L}\frac{1}{1+\frac{P_{\mathrm{n}}}{|h_{l,k}|^2}}}, \tag{33}$$

where $h_{l,k} = \mathbf{g}^{\mathsf{H}}_{l,k}\mathbf{H}\,\mathbf{g}_{l,k}$ denotes the one-tap channel. The first term in (33) is a conventional one-tap MMSE equalizer, while the second term is a scaling factor which guarantees that the estimated data symbols are approximately unbiased, that is, $\mathbb{E}\{\tilde{y}_{\tilde{l},k}|\tilde{x}_{\tilde{l},k}\} \approx \tilde{x}_{\tilde{l},k}$. This approximation becomes tight for perfectly orthogonal systems such as SC-FDMA in a time-invariant channel. Note that for a doubly-flat channel (all $h_{l,k}$ have the same value) or if the noise is zero ($P_{\mathrm{n}} = 0$), (33) becomes a zero-forcing equalizer, that is, $e_{l,k} = 1/h_{l,k}$.

We stack all equalizer elements of (33) in a vector $\mathbf{e} \in \mathbb{C}^{LK\times 1}$, defined as $[\mathbf{e}]_{l+L(k-1)} = e_{l,k}$. The input-output relationship of the whole transmission system can then be modeled by,

$$\tilde{\mathbf{y}} = \underbrace{\mathbf{C}^{\mathsf{H}}\mathrm{diag}\{\mathbf{e}\}\mathbf{G}^{\mathsf{H}}\mathbf{H}\mathbf{G}\mathbf{C}}_{\mathbf{D}}\,\tilde{\mathbf{x}} + \underbrace{\mathbf{C}^{\mathsf{H}}\mathrm{diag}\{\mathbf{e}\}\mathbf{G}^{\mathsf{H}}}_{\mathbf{\Gamma}}\,\mathbf{n}. \tag{34}$$

where $\tilde{\mathbf{y}}$ provides an estimate of the transmitted data symbols $\tilde{\mathbf{x}}$. The interference is described by the off-diagonal elements of matrix $\mathbf{D} \in \mathbb{C}^{\frac{L}{2}K\times\frac{L}{2}K}$ and the noise by matrix $\mathbf{\Gamma} \in \mathbb{C}^{\frac{L}{2}K\times N}$. For uncorrelated data symbols with unit power, $P_{\tilde{\mathrm{x}}} = \mathbb{E}\{|\tilde{x}_{\tilde{l},k}|^2\} = 1$, and uncorrelated noise samples with

power P_{n}, the SINR, conditioned on channel realization $\mathbf{H}$, can be calculated by a row-wise summation of the squared absolute elements of $\mathbf{D}$ and $\mathbf{\Gamma}$, according to,

$$\mathrm{SINR}_{\tilde{l},k}(\mathbf{H}) = \frac{1}{\sum\limits_{i=1}^{\frac{L}{2}K} |[\mathbf{D}-\mathbf{I}]_{\underline{\tilde{l}k},i}|^2 + P_{\mathrm{n}} \sum\limits_{i=1}^{N} |[\mathbf{\Gamma}]_{\underline{\tilde{l}k},i}|^2}. \quad (35)$$

In (35), we utilize the short notation $\underline{\tilde{l}k} = \tilde{l} + \frac{L}{2}(k-1)$ to describe the $\underline{\tilde{l}k}$-th row position, corresponding to code-position $\tilde{l} = 1, 2, \ldots, \frac{L}{2}$ and time-position k of our vectorized system model. Equation (35) includes the channel induced interference as well as the orthogonality approximation error. If both of those interference terms are sufficiently low, that is, $\mathbf{G}^{\mathsf{H}}\mathbf{H}\,\mathbf{G} \approx \mathrm{diag}\{\mathbf{h}\}\,\mathbf{G}^{\mathsf{H}}\mathbf{G}$ and $\mathbf{C}^{\mathsf{H}}\mathbf{G}^{\mathsf{H}}\mathbf{G}\mathbf{C} \approx \mathbf{I}_{LK/2}$, (35) can be approximated by

$$\mathrm{SINR}_{\tilde{l},k}^{\mathrm{Appr.}}(\mathbf{H}) = \frac{\frac{1}{L}\sum\limits_{l=1}^{L} \frac{1}{1+\frac{P_{\mathrm{n}}}{|h_{l,k}|^2}}}{1 - \frac{1}{L}\sum\limits_{l=1}^{L} \frac{1}{1+\frac{P_{\mathrm{n}}}{|h_{l,k}|^2}}}. \quad (36)$$

In Section VI-B we provide a numerical example of the approximation error. For a doubly-flat channel the approximation in (36) transforms to $\mathrm{SINR}_{\tilde{l},k}^{\mathrm{Appr.}} = |h_{\tilde{l},k}|^2/P_{\mathrm{n}}$, same as for conventional multi-carrier systems such as OFDM and FBMC (again ignoring any channel induced interference). Thus, if the delay spread and the bandwidth are sufficiently low, pruned DFT spread FBMC behaves like a conventional FBMC system. On the other hand, if the delay spread and the bandwidth are very high, our scheme shows a different behavior. In particular we observe a channel hardening effect. Let us consider the limit case of $L \to \infty$. For Rayleigh fading with unit power the averaging term in (36) can then be calculated by $\mathbb{E}_{|h|}\{1/(1+P_{\mathrm{n}}/|h|^2)\} = 1 - P_{\mathrm{n}}\mathrm{e}^{P_{\mathrm{n}}}\,\mathrm{E}_1\{P_{\mathrm{n}}\}$, where $\mathrm{E}_1\{\cdot\}$ denotes the exponential integral function. Thus, the SINR approximation in (36) becomes,

$$\mathrm{SINR}_{L\to\infty}^{\mathrm{Appr.}} = \frac{1}{P_{\mathrm{n}}}\left(\frac{\mathrm{e}^{-P_{\mathrm{n}}}}{\mathrm{E}_1\{P_{\mathrm{n}}\}} - P_{\mathrm{n}}\right), \quad (37)$$

and no longer depends on a specific channel realization.

B. "Flat" Channel: Enabling All MIMO Methods

The main goal of pruned DFT spread FBMC is to reduce the PAPR. This is different to the contribution of Lélé *et al.* [5], Zakaria and Le Ruyet [6], and our previous papers in [9] and [10], as the main motivation was to enable MIMO transmissions in FBMC by restoring complex orthogonality through spreading. Because pruned DFT spread FBMC also restores complex orthogonality our scheme can also be used for that purpose. In particular Alamouti's space time block code and ML MIMO detection become feasible. The drawback, however, is the same as in [5], [6], [9], and [10], namely, the channel has to be approximately flat within the spreading length. The main idea for enabling MIMO in FBMC is to despread before equalization, that is,

$$\tilde{\mathbf{y}} = \mathbf{C}^{\mathsf{H}}\mathbf{G}^{\mathsf{H}}\mathbf{H}\mathbf{G}\mathbf{C}\tilde{\mathbf{x}} + \tilde{\mathbf{n}} \approx \mathrm{diag}\{\tilde{\mathbf{h}}\}\tilde{\mathbf{x}} + \tilde{\mathbf{n}}. \quad (38)$$

If the approximation in (38) holds one obtains the same system model as in conventional OFDM transmissions. Thus, most known methods from OFDM can be straightforwardly employed. As already mentioned in Section III-A, precoding by $\mathbf{C}$ can be interpreted as transforming the underlying basis pulses according to $\tilde{\mathbf{G}} = \mathbf{G}\mathbf{C} = \left[\tilde{\mathbf{g}}_1 \cdots \tilde{\mathbf{g}}_{LK/2}\right]$. Thus, instead of modulating data symbols with $g_{l,k}(t)$, as in (1), we modulate them with $\tilde{g}_i(t)$. In contrast to conventional multicarrier systems, however, the transformed basis pulses $\tilde{g}_i(t)$ no longer employ all the same prototype filter $p(t)$. Instead, basis pulses have their own, unique, prototype filter $p_i(t)$ which makes a straightforward signal generation difficult. By interpreting $\tilde{\mathbf{G}}$ as a precoded FBMC system, however, the advantage of an efficient signal generation are preserved. Moreover, such interpretation offers a high flexibility.

In Section III-A we already presented an example of how spreading in frequency transforms the underlying basis pulses, see Figure 2 (e). In particular the transformed basis pulses become shorter in time and are therefore more robust in time-variant channels. On the other hand, the system becomes more sensitive to frequency-selective channels. Thus, if we want to employ low-complexity one-tap equalizers the delay spread must be very low, while the Doppler spread can be relatively high. For FFT-FBMC [6] (pruned DFT spreading in time) the opposite holds true, that is, the Doppler spread must be very low, while the delay spread can be relatively high.

To determine if the approximation in (38) holds Zakaria and Le Ruyet [6] assume that the transmission time is much shorter than the coherence time. Unfortunately such approach is not very accurate. Additionally, the approximation error also depends on the noise level. We therefore suggest a better measure by calculating the SIR in (38). As long as the Signal-to-Noise Ratio (SNR) is approximately 10 dB lower than the SIR, interference is dominated by noise and the approximation in (38) holds. Once the SNR approaches the SIR, the performance degeneration becomes equivalent to an SNR shift of approximately 3 dB. If the SNR is higher than the SIR, on the other hand, the approximation in (38) no longer holds and more sophisticated detection methods might be necessary, or the equalization must be performed in the multi-carrier domain, see Section V-A. The SIR in (38) is calculated in a similar way as in [2] and [10], leading to

$$\mathrm{SIR}_{\tilde{l},k}^{\mathrm{Appr.Flat}} = \frac{\mathbb{E}\{|\tilde{h}_{\tilde{l},k}\tilde{x}_{\tilde{l},k}|^2\}}{\mathbb{E}\{|\mathbf{c}_{\tilde{l},k}^{\mathsf{H}}\mathbf{G}^{\mathsf{H}}\mathbf{H}\mathbf{G}\mathbf{C}\tilde{\mathbf{x}} - \tilde{h}_{\tilde{l},k}\tilde{x}_{\tilde{l},k}|^2\}} \quad (39)$$

$$= \frac{[\mathbf{\Phi}]_{\underline{\tilde{l}k},\underline{\tilde{l}k}}}{\mathrm{tr}\{\mathbf{\Phi}\} - [\mathbf{\Phi}]_{\underline{\tilde{l}k},\underline{\tilde{l}k}}}, \quad (40)$$

where $\underline{\tilde{l}k}$ is the short notation for $\underline{\tilde{l}k} = \tilde{l} + \frac{L}{2}(k-1)$, vector $\mathbf{c}_{\tilde{l},k} \in \mathbb{C}^{LK\times 1}$ denotes the $\underline{\tilde{l}k}$-th column of $\mathbf{C}$ and matrix $\mathbf{\Phi} \in \mathbb{C}^{\frac{L}{2}K\times\frac{L}{2}K}$ is defined as

$$\mathbf{\Phi} = \left(\mathbf{C}^{\mathsf{T}}\mathbf{G}^{\mathsf{T}} \otimes \mathbf{c}_{\tilde{l},k}^{\mathsf{H}}\mathbf{G}^{\mathsf{H}}\right)\mathbf{R}_{\mathrm{vec}\{\mathbf{H}\}}\left(\mathbf{C}^{\mathsf{T}}\mathbf{G}^{\mathsf{T}} \otimes \mathbf{c}_{\tilde{l},k}^{\mathsf{H}}\mathbf{G}^{\mathsf{H}}\right)^{\mathsf{H}}. \quad (41)$$

Note that correlation matrix $\mathbf{R}_{\mathrm{vec}\{\mathbf{H}\}} = \mathbb{E}\{\mathrm{vec}\{\mathbf{H}\}\mathrm{vec}\{\mathbf{H}\}^{\mathsf{H}}\}$ depends only on the power delay profile and the Doppler spectral density. Thus, in contrast to Section V-A, the SIR

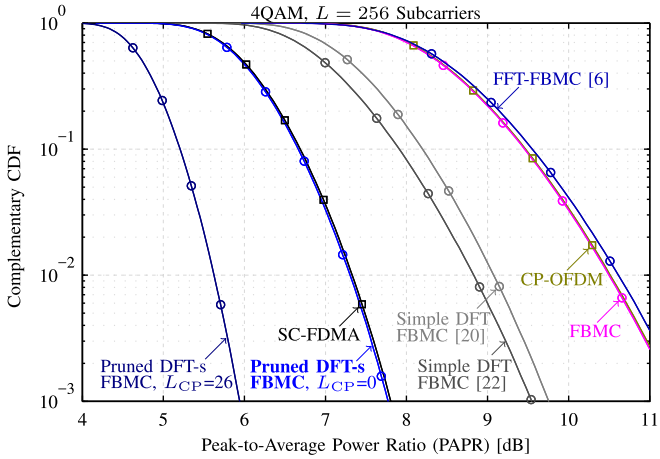

Fig. 9. Pruned DFT spread FBMC has the same PAPR as SC-FDMA, but the additional advantage of much lower OOB emissions. A frequency CP can further improve the PAPR at the expense of a reduced spectral efficiency.

in (39) no longer depends on a specific channel realization but only on the second order channel statistics.

VI. Simulation Results

To better show the potential of pruned DFT spread FBMC we perform Monte Carlo simulations. We first evaluate the PAPR in Section VI-A. Afterwards in Section VI-B we discuss the SINR, the BER and the throughput performance in doubly-selective channels. Finally, in Section VI-C, we consider the scenario of a very short delay spread and a small bandwidth (flat channel).

A. PAPR

Figure 9 shows the Complementary Cumulative Distribution Function (CCDF) of the PAPR for a 4-QAM signal constellation and $L = 256$ subcarriers. Conventional FBMC has the same poor PAPR as OFDM. Pruned DFT spread FBMC, on the other hand, performs as well as SC-FDMA and is approximately 3 dB better than OFDM and FBMC. Moreover, one can further reduce the PAPR by utilizing a frequency CP, although this comes at the expense of a lower spectral efficiency. Figure 2 (c) helps to explain why a frequency CP reduces the PAPR further. To be specific, the number of transformed basis pulses is reduced from $\frac{L}{2}$ to $\frac{L-L_{\mathrm{CP}}}{2}$ and the basis pulses are further apart from each other. This decreases the overlapping between pulses in time and consequently reduces the PAPR further.

Compared with FFT-FBMC [6], that is, pruned DFT spreading in time, described by $\mathbf{C} = \widetilde{\mathbf{W}}^{(1)}_{K\times K/2} \otimes \mathbf{I}_{L/2} \otimes \left[\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right] + \widetilde{\mathbf{W}}^{(2)}_{K\times K/2} \otimes \mathbf{I}_{L/2} \otimes \left[\begin{smallmatrix}0&0\\0&1\end{smallmatrix}\right]$, spreading in frequency offers a much lower PAPR. In fact the PAPR of FFT-FBMC is even worse than in conventional FBMC systems. This effect can again be explained by the underlying basis pulses. Similar to spreading in frequency, Figure 2 (d) can also be used to describe spreading in time, where we only have to replace "time" with "frequency" on the x-axis. In particular FFT-FBMC is nothing else than a subband filtered OFDM scheme, similar as UF-OFDM [38], but with the additional advantage that the signal can be generated more efficiently. From Figure 2 (d) we conclude that spreading in time can be interpreted as increasing the number of "subcarriers" by a factor of K, that is, $L \to LK$. As such, the PAPR becomes worse than in FBMC and OFDM which only employ L subcarriers.

Pruned DFT spread FBMC also outperforms other frequency spreading methods for FBMC. For example the simple DFT spreading scheme, proposed in [20], has a relatively poor PAPR. To improve the PAPR Na and Choi [22] recently proposed an optimal phase pattern for the simple DFT spreading scheme, that is, the phase $\mathrm{e}^{\mathrm{j}\frac{\pi}{2}(l+k)}$ in (2) is replaced by $\mathrm{e}^{\mathrm{j}\frac{\pi}{2}(l+k)}\mathrm{e}^{-\mathrm{j}\pi lk}$. However, as shown in Figure 9 an optimal phase pattern only slightly improves the PAPR. Thus, Na and Choi [22] further suggested a candidate selection scheme where they generate four different DFT spread FBMC signals and select the one with the lowest PAPR, resulting in a similar PAPR as in SC-FDMA (not shown in the figure). However, the main drawback of [22] is the necessity of side information. While Na and Choi [22] argue that the side information only consists of two bits and can thus be neglected, we would like to point out that side information causes additional challenges. For example the side information is very crucial for the detection process and must therefore be channel coded at low rate in order to guarantee robustness. The overhead is therefore much larger than just two bits. Furthermore, the side information cannot be directly transmitted within the same FBMC symbol because it must already be known before demodulation is possible. Thus, the side information must be transmitted on a separate transmission channel or computational demanding blind detection must be employed. All those drawbacks are avoided in pruned DFT spread FBMC because we do not require any side information. Furthermore, compared with [22] the computational complexity of our scheme is approximately 10%-80% lower, the latency is reduced, see Section III-C, and we restore complex orthogonality, allowing efficient uplink transmissions. On the other hand, the spectral confinement of [22] might be better than for our method because the overlapping factor is not restricted to be approximately $O \leq 1.5$. However, we can circumvent this drawback in pruned DFT spread FBMC by reducing the subcarrier spacing, for example by a factor of two, so that the power spectral density becomes very similar to conventional FBMC based on the PHYDYAS prototype filter, see Figure 6.

B. Performance in Doubly-Selective Channels

To describe wireless channels we consider the new 3GPP 38.900 channel model [40, Sec. 7.7.3]. To be specific we assume a Tapped Delay Line (TDL)-A power delay profile and a long delay spread of 300 ns. In future wireless systems we expect that beamforming and a small cell size will typically lead to a much shorter delay spread [2]. However, by considering such a long delay spread we are able to evaluate the robustness of our transmission scheme. As proposed by 3GPP time variations are modeled by a Jakes Doppler spectrum.

Let us first compare the true SINR in (35) to the approximated SINR in (36). As the SINR depends on a given channel

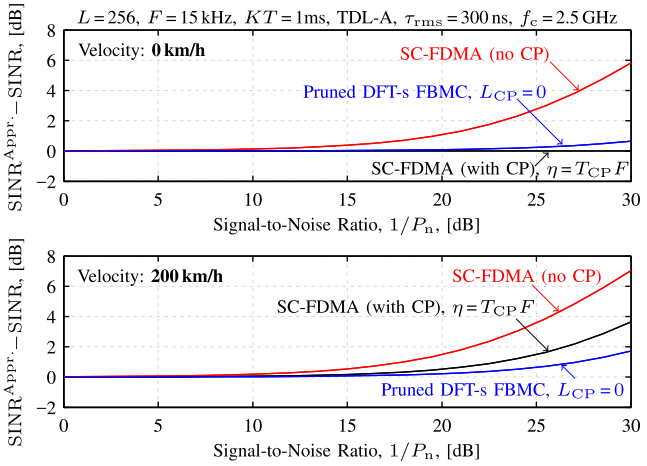

Fig. 10. The approximated SINR in (36) accurately describes the true SINR in (35) as long as the channel induced interference is dominated by noise.

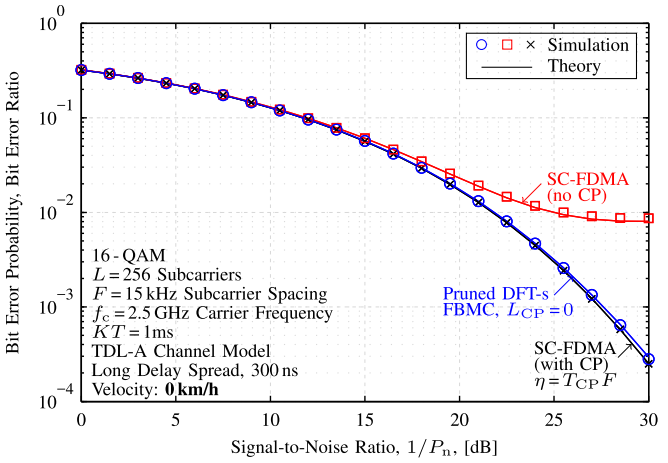

Fig. 11. Our method is relatively robust to large delay spreads and performs close to SC-FDMA (with CP), but has the additional advantage that no CP is necessary. Removing the CP in SC-FDMA causes severe interference, leading to a saturation effect of the BER. The closed-form BEP expression in (43) accurately describes the simulated BER.

realization we define the total SINR as the average signal power (which is one) divided by the average interference plus noise power,

$$\text{SINR} = \frac{1}{\frac{1}{K}\sum_{k=1}^{K}\frac{1}{L/2}\sum_{\tilde{l}=1}^{L/2}\mathbb{E}_{\mathbf{H}}\left\{\frac{1}{\text{SINR}_{\tilde{l},k}(\mathbf{H})}\right\}}. \quad (42)$$

The expectation $\mathbb{E}_{\mathbf{H}}\{\cdot\}$ in (42) is calculated by Monte Carlo evaluation. Figure 10 shows the error between the approximated SINR in (36) and the true SINR in (35). In low SNR regimes the channel induced interference is dominated by noise so that the approximation in (36) is tight. The approximation error can also be utilized to describe how much interference is caused by the channel. The reference curve is then 0 dB (no interference), achieved by SC-FDMA with CP for 0 km/h. However, this optimal performance of SC-FDMA comes at the expense of a reduced data rate, $T_{\text{CP}}F = \tau_{\max}F = 4\%$. If the CP is removed[2] in SC-FDMA, the performance severely starts to degenerate for SNR values larger than 15 dB, as shown Figure 10. Pruned DFT spread FBMC, on the other hand, is almost not affected by the large delay spread and performs close to SC-FDMA with CP. Moreover, pruned DFT spread FBMC performs well in high velocity scenarios, such as 200 km/h, and outperforms SC-FDMA with CP which is severely affected by inter-carrier-interference.

The SINR can also be utilized to calculate the Bit Error Probability (BEP), a more meaningful measure than the SINR. By assuming that the interference is Gaussian distributed we calculate the BEP by,

$$\text{BEP} = \mathbb{E}_{\mathbf{H}}\left\{\frac{1}{K}\sum_{k=1}^{K}\frac{1}{L/2}\sum_{\tilde{l}=1}^{L/2}\text{BEP}_{\text{AWGN}}\{\text{SINR}_{\tilde{l},k}(\mathbf{H})\}\right\}. \quad (43)$$

The function $\text{BEP}_{\text{AWGN}}\{\cdot\}$ represent the BEP for an AWGN channel and can easily be calculated. For example, in case of a 4-QAM signal constellation $\text{BEP}_{\text{AWGN}}^{\text{4QAM}}\{\text{SINR}\} = \mathcal{Q}\{\sqrt{\text{SINR}}\}$, with $\mathcal{Q}\{\cdot\}$ denoting the Q-function. For higher modulation orders, however, the function $\text{BEP}_{\text{AWGN}}\{\cdot\}$ becomes more evolved and includes many summations, see [41]. Figure 11 and Figure 12 show that the closed-form solution in (43) accurately describes the simulated BER, indicated by the markers. Only for high SNR values small deviations can be observed because the channel induced interference is not Gaussian distributed, violating the underlying assumption of (43). Figure 11 shows the BEP for a velocity of 0 km/h. Overall, the behavior is similar as already observed in Figure 10. SC-FDMA without CP is severely affected by the long delay spread, while pruned DFT spread FBMC performs close to the optimum, that is, SC-FDMA with CP. Figure 12 shows the BEP for a velocity of 200 km/h. Similar as for the SINR, SC-FDMA is severely affected by inter-carrier-interference. Pruned DFT spread FBMC, on the other hand, is more robust in doubly-selective channels and outperforms SC-FDMA.

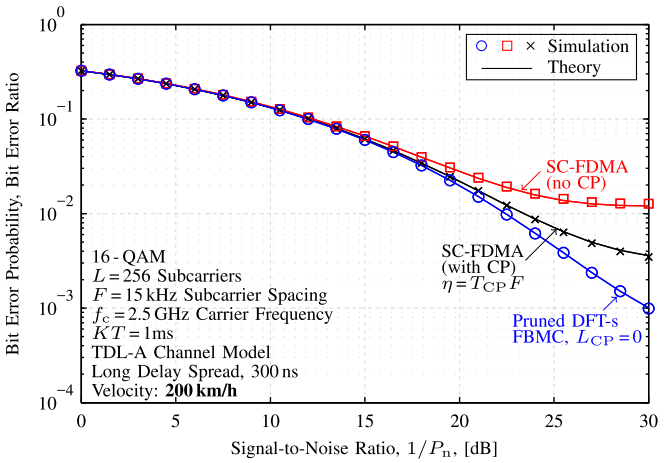

Fig. 12. Similar as Figure 11 but for a velocity of 200 km/h. Pruned DFT spread FBMC now even outperforms SC-FDMA with CP, which is severely affected by inter-carrier-interference.

[2]We consider SC-FDMA "without CP" because it has the same data rate as pruned DFT spread FBMC, allowing for a fair comparison.

In practical systems channel coding and link adaption are of utmost importance. For a fair comparison of different modulation schemes we therefore consider the throughput, a more meaningful measure than the BEP because it includes channel coding and link adaption. To derive an information theoretic upper bound of the throughput, the achievable rate *R*, we assume that the transmission of each data symbol can be considered as a separate AWGN channel, characterized by $\mathrm{SINR}_{\tilde{l},k}(\mathbf{H})$. The capacities of those separate transmission channels are then modeled by the Bit-Interleaved Coded Modulation (BICM) capacity, $\mathrm{BICM}_{\mathrm{AWGN}}\{\mathrm{SINR}\}$, see [2], [42]. Compared with the well-known channel capacity $\mathcal{C}_{\mathrm{AWGN}}\{\mathrm{SINR}\} = \log_2(1+\mathrm{SINR})$, the BICM capacity does not assume Gaussian distributed data symbols and instead allows for a QAM signal constellation in combination with bitwise interleaving. The upper bound of the throughput, achievable rate *R*, can then be calculated by

$$\mathrm{R} = \mathbb{E}_{\mathbf{H}} \left\{ \frac{1}{KT} \sum_{k=1}^{K} \sum_{\tilde{l}=1}^{L/2} \mathrm{BICM}_{\mathrm{AWGN}}\{\mathrm{SINR}_{\tilde{l},k}(\mathbf{H})\} \right\}, \quad (44)$$

where we implicitly assume infinitely many transmission blocks in time so that *KT* represents the average transmission time for one block. To simulate the throughput we employ turbo coding in combination with 15 different modulation and coding schemes ({4, 16, 64}-QAM and code rates between 0.08 and 0.93, as proposed by the LTE standard). We transmit the signal for all possible modulation and coding schemes and choose at the receiver the highest throughput, that is, the highest data rate for which all data bits were correctly detected, implicitly assuming perfect feedback. The Log-Likelihood Ratio (LLR), required for turbo decoding, is calculated by considering a per symbol AWGN channel, characterized by the approximated SINR in (36). Thus, we ignore any noise and interference correlation to keep the computational complexity low. Moreover, we utilize the approximated SINR instead of the true SINR in (35) because it only requires knowledge of the one-tap channels and the noise power, making it a more practical approach. Figure 13 shows the simulated throughput as well as the achievable rate, see (44). The simulated throughput is approximately 2dB SNR shifted when compared to the achievable rate. Such difference can be explained by an imperfect coder, a limited code length, and a limited number of code rates. Nonetheless, the achievable rate in (44) accurately captures the main properties. In particular FBMC is at most 13% better than pruned DFT spread FBMC and the SNR shift between both curves is at most 2.3 dB. Those performance differences are mainly caused by frequency spreading in combination with a long delay spread and a relatively large bandwidth. If, for example, the delay spread would be smaller, pruned DFT spread FBMC would perform closer to FBMC (not shown in the Figure). In the extreme case of a doubly-flat channel both schemes would show the same performance. On the other hand, if we consider the limit case of $L \to \infty$, the performance difference between FBMC and pruned DFT spread FBMC would be at most 27 % and the SNR shift at most 3.6 dB. Those values are calculated by comparing $\mathbb{E}_{|h|}\{\mathrm{BICM}_{\mathrm{AWGN}}\{|h|^2/P_{\mathrm{n}}\}\}$ with

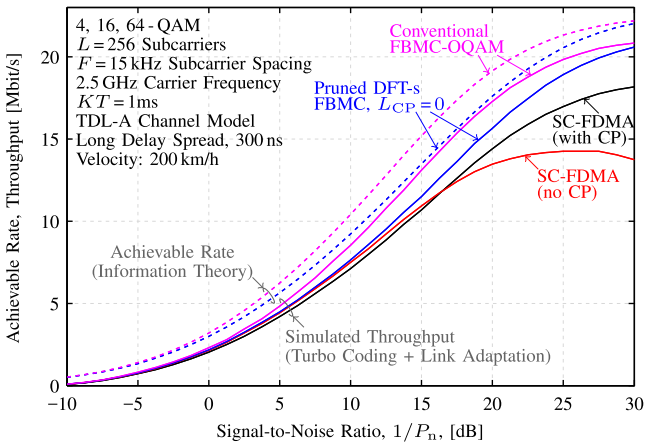

Fig. 13. Pruned DFT spread FBMC outperforms SC-FDMA because it is more robust in doubly-selective channels and does not require a CP. Spreading causes an averaging effect of the channel, reducing the throughput when compared to conventional multi-carrier schemes. The achievable rate in (44), an upper bound for the throughput, captures the main properties of our transmission system.

$\mathrm{BICM}_{\mathrm{AWGN}}\{\mathrm{SINR}^{\mathrm{Appr.}}_{L\to\infty}\}$, see (37). Thus, in Figure 13 we observe an intermediate performance degradation of pruned DFT spread FBMC when compared to conventional FBMC. For other settings, the difference could be higher ($L \to \infty$) or lower (doubly-flat channel). Similar as already observed for the SINR and the BEP, SC-FDMA is severely affected by inter-carrier interference. This causes a saturation effect of the throughput once the channel induced interference starts to dominate the noise. Pruned DFT spread FBMC, on the other hand, is more robust in time-variant channels so that the throughput is higher than in SC-FDMA. Moreover, our method does not need a CP, further increasing the throughput when compared to SC-FDMA with CP. Note that for SNR values smaller than 16.5 dB, SC-FDMA without CP outperforms SC-FDMA with CP. Furthermore, the throughput for SC-FDMA without CP decreases for high SNR values because our LLR calculation does not account for the channel induced interference, see (36), causing a mismatch.

C. MIMO in a "Flat" Channel

We now consider the special case of a very short delay spread in combination with a small bandwidth, as we expect for example in indoor Machine to Machine (M2M) communications. As discussed in Section V-B it is possible to despread before equalization. If the channel is sufficiently flat, the transmission can still be modeled by one-tap channels so that all MIMO methods known in OFDM can be straightforwardly employed in FBMC. To decide whether the channel can be considered as flat or not, we proposed the SIR in Section V-B. For a delay spread of 10 ns, a velocity of 3 km/h, and $L = 64$ subcarriers, the SIR is 25 dB, calculated by (39). Thus, for SNR values smaller than 15 dB we expect that the channel can be considered as flat. Figure 14 shows the throughput for 2×2 spatial multiplexing where we assume that both MIMO streams employ the same modulation and coding scheme. For SNR values smaller than 15 dB, the throughput

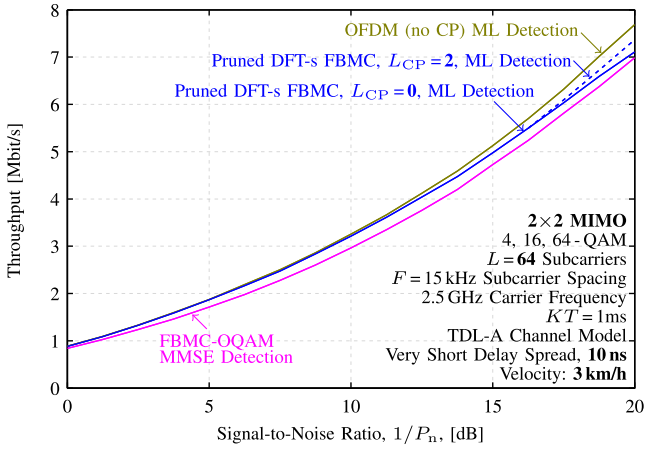

Fig. 14. If the channel is approximately frequency-flat, low-complexity ML-MIMO detection becomes feasible in pruned DFT spread FBMC because complex orthogonality is restored. Our scheme performs close to OFDM but has the additional advantage of much lower OOB emissions, see Figure 6.

difference between pruned DFT spread FBMC and OFDM is less that 3 %. Thus, our scheme shows approximately the same performance as OFDM and is up to 10% better than conventional FBMC, for which low-complexity ML detection is not feasible (we consider MMSE detection as a reference). For SNR values larger than 16 dB it becomes advantageous to include a frequency CP in pruned DFT spread FBMC because of the reduced self-interference. However, in contrast to Figure 7 the SIR is not 45 dB (as for a flat channel), but 34 dB because of the delay spread. To improve the SIR further the bandwidth has to be reduced.

VII. Conclusion

Our novel pruned DFT spread FBMC transmission scheme outperforms SC-FDMA in many aspects. It is more robust in doubly-selective channels, requires no CP and has much lower OOB emissions. Furthermore, if the channel is approximately frequency-flat, our method even outperforms conventional FBMC because MIMO can be straightforwardly employed. Potential applications of pruned DFT spread FBMC include uplink transmissions in wireless communications as well as M2M communications, where the good time-frequency localization guarantees that no sophisticated synchronization between users is necessary.

References

[1] B. Farhang-Boroujeny and H. Moradi, "OFDM inspired waveforms for 5G," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 2474–2492, 4th Quart., 2016.

[2] R. Nissel, S. Schwarz, and M. Rupp, "Filter bank multicarrier modulation schemes for future mobile communications," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 8, pp. 1768–1782, Aug. 2017.

[3] H. G. Feichtinger and T. Strohmer, *Gabor Analysis and Algorithms: Theory and Applications*. Basel, Switzerland: Birkhäuser, 1998. [Online]. Available: https://doi.org/10.1007/978-1-4612-2016-9

[4] D. Katselis, E. Kofidis, A. Rontogiannis, and S. Theodoridis, "Preamble-based channel estimation for CP-OFDM and OFDM/OQAM systems: A comparative study," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2911–2916, May 2010.

[5] C. Lélé, P. Siohan, and R. Legouable, "The Alamouti scheme with CDMA-OFDM/OQAM," *EURASIP J. Adv. Signal Process.*, Dec. 2010, Art. no. 703513. [Online]. Available: https://doi.org/10.1155/2010/703513

[6] R. Zakaria and D. Le Ruyet, "A novel filter-bank multicarrier scheme to mitigate the intrinsic interference: Application to MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 1112–1123, Mar. 2012.

[7] A. I. Pérez-Neira *et al.*, "MIMO signal processing in offset-QAM based filter bank multicarrier systems," *IEEE Trans. Signal Process.*, vol. 64, no. 21, pp. 5733–5762, Nov. 2016.

[8] M. Caus and A. I. Perez-Neira, "SDMA for filterbank with Tomlinson Harashima precoding," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2013, pp. 4571–4575.

[9] R. Nissel and M. Rupp, "Enabling low-complexity MIMO in FBMC-OQAM," in *Proc. IEEE Globecom Workshops (GC Wkshp)*, Dec. 2016, pp. 1–6.

[10] R. Nissel, J. Blumenstein, and M. Rupp, "Block frequency spreading: A method for low-complexity MIMO in FBMC-OQAM," in *Proc. IEEE 18th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2017, pp. 1–5.

[11] C. Lélé, P. Siohan, R. Legouable, and M. Bellanger, "CDMA transmission with complex OFDM/OQAM," *EURASIP J. Wireless Commun. Netw.*, Dec. 2007, Art. no. 748063. [Online]. Available: https://link.springer.com/article/10.1155/2008/748063

[12] A. Sahin, I. Guvenc, and H. Arslan, "A survey on multicarrier communications: Prototype filters, lattice structures, and implementation aspects," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1312–1338, 3rd Quart., 2014.

[13] R. W. Bäuml, R. F. H. Fischer, and J. B. Huber, "Reducing the peak-to-average power ratio of multicarrier modulation by selected mapping," *Electron. Lett.*, vol. 32, no. 22, pp. 2056–2057, Oct. 1996.

[14] S. H. Müller and J. B. Huber, "OFDM with reduced peak-to-average power ratio by optimum combination of partial transmit sequences," *Electron. Lett.*, vol. 33, no. 5, pp. 368–369, Feb. 1997.

[15] A. Skrzypczak, J.-P. Javaudin, and P. Siohan, "Reduction of the peak-to-average power ratio for the OFDM/OQAM modulation," in *Proc. 63rd IEEE Veh. Technol. Conf. (VTC Spring)*, vol. 4, May 2006, pp. 2018–2022.

[16] D. Qu, S. Lu, and T. Jiang, "Multi-block joint optimization for the peak-to-average power ratio reduction of FBMC-OQAM signals," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1605–1613, Apr. 2013.

[17] C. Ye, Z. Li, T. Jiang, C. Ni, and Q. Qi, "PAPR reduction of OQAM-OFDM signals using segmental PTS scheme with low complexity," *IEEE Trans. Broadcast.*, vol. 60, no. 1, pp. 141–147, Mar. 2014.

[18] S. Sesia, M. Baker, and I. Toufik, *LTE—The UMTS Long Term Evolution: From Theory to Practice*. Hoboken, NJ, USA: Wiley, 2011.

[19] *Study on New Radio Access Technology; Physical Layer Aspects; (Release 14)*, document TR 38.802, 3GPP, Mar. 2017. [Online]. Available: http://www.3gpp.org/DynaReport/38802.htm

[20] T. Ihalainen, A. Viholainen, T. H. Stitz, M. Renfors, and M. Bellanger, "Filter bank based multi-mode multiple access scheme for wireless uplink," in *Proc. IEEE Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2009, pp. 1354–1358.

[21] C. H. Yuen, P. Amini, and B. Farhang-Boroujeny, "Single carrier frequency division multiple access (SC-FDMA) for filter bank multicarrier communication systems," in *Proc. 5th IEEE Int. Conf. Cognit. Radio Oriented Wireless Netw. Commun. (CROWNCOM)*, Jun. 2010, pp. 1–5.

[22] D. Na and K. Choi, "Low PAPR FBMC," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 182–193, Jan. 2018.

[23] G. Berardinelli, F. M. L. Tavares, T. B. Sørensen, P. Mogensen, and K. Pajukoski, "Zero-tail DFT-spread-OFDM signals," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2013, pp. 229–234.

[24] P. Siohan, C. Siclet, and N. Lacaille, "Analysis and design of OFDM/OQAM systems based on filterbank theory," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1170–1183, May 2002.

[25] M. Renfors, X. Mestre, E. Kofidis, and F. Bader, *Orthogonal Waveforms and Filter Banks for Future Communication Systems*. New York, NY, USA: Academic, 2017.

[26] M. Terre. FBMC Modulation/Demodulation. MATLAB Central. Accessed: Oct. 19, 2017. [Online]. Available: https://www.mathworks.com/matlabcentral

[27] G. Matz, D. Schafhuber, K. Gröchenig, M. Hartmann, and F. Hlawatsch, "Analysis, optimization, and implementation of low-interference wireless multicarrier systems," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1921–1931, May 2007.

[28] Qualcomm Incorp., *Waveform Candidates*, document 3GPP TSG-RAN WG1 84b, Busan, South Korea, Apr. 2016.

[29] Y. G. Li and G. L. Stuber, *Orthogonal Frequency Division Multiplexing for Wireless Communications*. New York, NY, USA: Springer, 2006.
[30] H. G. Myung, J. Lim, and D. J. Goodman, "Single carrier FDMA for uplink wireless transmission," *IEEE Veh. Technol. Mag.*, vol. 1, no. 3, pp. 30–38, Sep. 2006.
[31] G. Berardinelli, K. I. Pedersen, T. B. Sorensen, and P. Mogensen, "Generalized DFT-spread-OFDM as 5G waveform," *IEEE Commun. Mag.*, vol. 54, no. 11, pp. 99–105, Nov. 2016.
[32] M. Bellanger *et al.*, "FBMC physical layer: A primer," *PHYDYAS*, vol. 25, no. 4, pp. 7–10, 2010.
[33] R. Haas and J.-C. Belfiore, "A time-frequency well-localized pulse for multiple carrier transmission," *Wireless Pers. Commun.*, vol. 5, no. 1, pp. 1–18, Jan. 1997.
[34] R. Nissel, M. Rupp, and R. Marsalek, "FBMC-OQAM in doubly-selective channels: A new perspective on MMSE equalization," in *Proc. IEEE 18th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2017, pp. 1–5.
[35] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Emerg. Telecommun. Technol.*, vol. 10, no. 6, pp. 585–595, 1999.
[36] R. Zakaria and D. Le Ruyet, "Theoretical analysis of the power spectral density for FFT-FBMC signals," *IEEE Commun. Lett.*, vol. 20, no. 9, pp. 1748–1751, Sep. 2016.
[37] F. Franchetti and M. Puschel, "Generating high performance pruned FFT implementations," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2009, pp. 549–552.
[38] F. Schaich, T. Wild, and Y. Chen, "Waveform contenders for 5G-suitability for short packet and low latency transmissions," in *Proc. IEEE 79th Veh. Technol. Conf. (VTC Spring)*, May 2014, pp. 1–5.
[39] J. Li, E. Bala, and R. Yang, "Sliding window-frequency domain equalization for multi-mode communication systems," in *Proc. Long Island IEEE Syst., Appl. Technol. Conf. (LISAT)*, May 2013, pp. 1–6.
[40] *Study on Channel Model for Frequency Spectrum Above 6 GHz; (Release 14)*, document TR 38.900, 3GPP, Dec. 2016. [Online]. Available: http://www.3gpp.org/DynaReport/38900.htm
[41] R. Nissel and M. Rupp, "OFDM and FBMC-OQAM in doubly-selective channels: Calculating the bit error probability," *IEEE Commun. Lett.*, vol. 21, no. 6, pp. 1297–1300, Jun. 2017.
[42] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.

Ronald Nissel received the bachelor's degree in electrical engineering from TU Wien in 2010, the bachelor's degree in economics from WU Wien in 2010, the Dipl.-Ing. degree (M.Sc. equivalent) in telecommunications from TU Wien in 2013, the Mag.rer.soc.oec. degree (M.Sc. equivalent) in economics from the University of Vienna in 2015, and the Dr.techn. degree (Ph.D. equivalent) from TU Wien in 2017. Since 2011, he has been with the Institute of Telecommunications, TU Wien. His research interests include multicarrier modulation, channel estimation, and testbed measurements.

Markus Rupp received the Dipl.Ing. degree from the University of Saarbrücken, Germany, in 1988, and the Dr.-Ing. degree from the Technische Universität Darmstadt, Germany, in 1993. Until 1995, he held a post-doctoral position with the University of California at Santa Barbara, Santa Barbara, CA, USA. From 1995 to 2001, he was with the Wireless Technology Research Department, Nokia Bell Labs, Holmdel, NJ, USA. Since 2001, he has been a Full Professor of digital signal processing in mobile communications with TU Wien.