ACHIEVABLE DATA RATE OF WIDEBAND OFDM WITH DATA-AIDED CHANNEL ESTIMATION
Niels Hadaschik, Gerd Ascheid, Heinrich Meyer
Institute for Integrated Signal Processing Systems, RWTH Aachen University,
Templergraben 55, 52056 Aachen, Germany, hadaschik@iss.rwth-aachen.de

ABSTRACT
The achievable data rate of an OFDM system using data-aided channel estimation in the high bandwidth regime is evaluated under the assumption of a frequency selective, continuously fading channel. As previous results propose, the achievable data rate depends on the LMMSE channel estimate, for which a convenient representation is introduced here. The mean square estimation error is derived from this representation, allowing for an analysis with respect to the optimum amount and distribution of the pilot symbols in the wideband regime. The results on the data rate show good accordance to previous results based on non-data aided channel prediction especially in the interesting bandwidth range.

I. INTRODUCTION
Data transmission in the wideband regime has been a strongly discussed topic in the past few years. The research was driven by three events: firstly, the FCC regulation and the connected standardization under the name IEEE 802.15.3a for wireless personal area networks (WPAN). Another driving element has been the publication of promising results on the capacity in the wideband or low power spectral density region. Verdú [1], Medard and Gallagher [2] as well as Subramian and Hajek [3] have introduced variations of the fourth moment of the input distribution as measure for the suitability of transmission schemes in the wideband domain.

OFDM has already been selected for the MB-OFDM standard. In [4] it has been shown, that OFDM using constant power on all subcarriers, i.e. independent PSK modulation, is suboptimum in the extreme wideband regime. A further result of [4] has been that depending on the channel dynamics given via a Doppler frequency, there is a maximum achievable data rate when increasing the bandwidth of the OFDM transmission system. However, [4] also demonstrates that the suboptimality of OFDM in the extreme wideband regime is of no significance for common OFDM systems and even for UWB with bandwidths up to 7.5 GHz, as the deterioration only occurs for bandwidths orders of magnitude beyond the ones in use.

For evaluating the achievable data rate, [4] presumes a decision directed, non-data-aided (NDA) channel prediction that is computationally complex and suffers from decoding errors and latencies, e.g. if Turbo codes are considered. Ref. [5] assumes a completely non-coherent transmission for evaluating the channel capacity which is beyond our scope. In our work, we consider a purely data-aided (DA) channel estimation. As it is based on a smaller amount of observations, it delivers degraded results compared to the NDA estimation in the absence of decoding errors. But, it is less complex by an order of magnitude and insensitive to decoding errors, and it allows us to find an optimum amount and distribution of training data.

Another approach to optimizing the training data for OFDM assumes block fading [6] and follows similar argumentation like in [7] for MIMO systems. However, block fading fails to model reality in terms of the channel dynamics. Thus in our work, a continuously fading channel is assumed. This work follows the discussion in [8], [9] for the flat fading case, and respectively [10] for MIMO systems in frequency selective fading. It is based on the presumption that the receiver has imperfect channel state information (CSI) as obtained from the DA channel estimation, but, there is no CSI present at the transmitter, since channel feedback is not considered. The channel estimation is extended compared to [10] supporting an arbitrary number of pilots per OFDM symbol allowing for a simpler optimization scheme.

In the following, the channel and the signal model are introduced in Section II. In Section III, the LMMSE channel estimator is derived before the achievable data rate is evaluated in Section IV. Afterwards, some results are presented in Section V demonstrating the potential of the scheme and the work is concluded with Section VI.

II. CHANNEL AND SIGNAL MODEL
A discrete time domain channel with the maximum delay $\tau_{\text{max}}$ and the system bandwidth $B$ is assumed i.e. the channel consists of $\nu + 1 = [B \tau_{\text{max}}] \approx B \tau_{\text{max}}$ independent channel taps in the CIR vector $c_n = (c_n^{(0)} \ldots c_n^{(\nu)})^T$. We extend the CIR vector with stuffer zeros to the length $K$ vector $\tilde{c}_n = (c_n^{(0)} 0 \ldots 0)^T$. Applying a $K$-DFT on it, in vector-matrix notation

$$h_n = \sqrt{K} F_K \tilde{c}_n = \sqrt{K} F_K^{(\nu+1)} c_n$$  \hspace{1cm} (1)$$
results for the discrete channel transfer vector of a $K$-dimensional OFDM transmitter. The $K$-point DFT-matrix, therein, is defined as

$$F_K = \frac{1}{\sqrt{K}} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & f_K & \cdots & f_K^{K-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & f_K^{K-1} & \cdots & f_K^{(K-1)(K-1)} \end{pmatrix}$$

with $f_K = \exp(-j \frac{2 \pi}{K})$. The shortcut $F_K^{(\nu+1)}$ reduces this matrix, taking just its first $\nu + 1$ columns. The according IDFT operation is performed by multiplying with the inverse matrix $F_K^{-1} = F_K^H$. Note, however, that in accordance with the common definition of DFT and IDFT this symmetric (in terms of output power) definition of DFT- and IDFT-matrix necessitates additional scaling of the results: $\sqrt{K}$ for the DFT and $1/\sqrt{K}$ for the IDFT.

The channel taps $c_n^{(0)}$ are independent, identically (iid) gaussian distributed with zero mean and powers $\rho_i$. Therefore, the channel coefficients of the subcarriers $h_n^{(k)}$ are correlated and are zero mean gaussian with power $\sigma_h^{(k)}$.
It is assumed that covariances in time and frequency of the continuous channel are independent
\[ R_c(\Delta \omega, \Delta t) = R_h(\Delta \omega) R_c(\Delta t). \] (2)
The indices \( h \) and \( c \) distinguish frequency and time domain covariance of the channel. The time domain covariance \( R_h \) is connected to the Doppler spectrum \( S_c(\omega) = \mathcal{F}\{R_c(\Delta t)\} \). For this work, the Doppler spectrum is assumed uniformly distributed with magnitude \( 1/2 f_D \) on the interval \([-f_D, f_D]\) and zero otherwise. The uniform distribution serves as a worst case according to [4]. The matrix \( R_c \) represents the accordant temporal covariances of one tap.

Similarly, the frequency domain covariance \( R_h \) is related to the power delay profile (PDP) \( p_\tau(\tau) = \mathcal{F}^{-1}\{R_h(\Delta \omega)\} \). In this text, we will assume a constant PDP in the range \([0, \tau_{\text{max}}]\). This time continuous measure directly translates via \( \sum_{\tau=-\infty}^{\infty} p_\tau = \sigma_h^2 \) to the discrete case: the taps of the uniform PDP are \( p_\tau = \sigma_h^2/(\nu + 1) \). They are assembled in the diagonal PDP matrix \( \rho = \text{diag}(\rho_0, \cdots, \rho_\nu) \). Note, that the channel power is incorporated in the tap domain covariance.

For this work, disjoint symbol sets for training \( T \) and for information transmission \( D \) are supposed. The training data is additionally subdivided into a preamble \( T_{\text{pre}} \) and pilots \( T_{\text{plt}} \). The latter consist of known symbols in dedicated pilot carriers that are equidistantly spaced in the subcarrier space by \( D_{t} \). These pilot carriers comprise pilots each \( D_{t} \)-th OFDM symbol. Each burst consists of \( M \) OFDM symbols, the preamble is composed of \( M_{\text{pre}} = \lfloor T/|D| \rfloor/K \) OFDM symbols such that in total \( MK = |T| + |D| \) subcarrier symbols are transmitted per burst. \(|A|\) thereby denotes the cardinality of data set \( A \). Thus, each burst contains \( |D| = (M - M_{\text{pre}})(K - K_{\text{plt}}/D_t) \) data and \(|T| = MK - |D| \) training symbols. \( K_{\text{plt}} \) is the number of pilot carriers and \( N_P = (M - M_{\text{pre}})/D_t \) pilot carrying OFDM symbols per burst are considered.

Perfect interleaving is assumed breaking down channel correlation on the received data. A function \( d_l \) with \( 0 \leq l \leq |D| - 1 \) maps the received data on the outgoing, deinterleaved data symbols – interleaving is performed by \( d_l^{-1} \).

A transmitted OFDM symbol consists of \( N \) samples including the \( K \leq N \) samples emerging from the \( K \)-point IDFT of the \( K \) subcarrier signals and the \( N_{\text{gd}} = N - K \) guard interval samples. The overhead is expressed by a spectral (in) efficiency factor \( \eta_{\text{OH}} = N/K = B_c T_s \), wherein \( B_c = B/K \) is the subcarrier bandwidth and \( T_s = N/B \) is the OFDM symbol duration.

The transmission equation for a single OFDM symbol at the symbol time instance \( n \) in the frequency domain becomes
\[ y_n = \text{diag}\{h_n\} \cdot x_n + w_n = \text{diag}\{x_n\} R_h \cdot x_n + w_n \] (3)

In the following, e.g. \( X_n = \text{diag}\{x_n\} \) indicates the diagonal matrix with the elements of vector \( x_n \) on its diagonal. \( x_n \) and \( w_n \) denote the \( K \)-dimensional vectors of transmitted symbols, of received symbols and of iid white gaussian noise with respective powers \( \sigma_x^2 \), \( \sigma_w^2 \) and \( \sigma_w^2 \).

For the pilot symbols as subset of all symbols within a single OFDM symbol, the equation is equivalently denoted as
\[ y_{p,n} = H_{p,n} \cdot x_{p,n} + w_{p,n} = X_{p,n} \cdot h_{p,n} + w_{p,n}. \] (4)

The column vector \( x_{p,n}, y_{p,n}, h_{p,n} \) and \( w_{p,n} \) of dimension \( K_{\text{plt}} < K \) assemble the received and the transmitted symbols, the channel transfer function and the AWGN samples of the \( n \)-th OFDM symbol at the pilot positions. The channel transfer function with respect to the pilot symbols is defined as the decimated version of (1)
\[ h_{p,n} = \sqrt{K_{\text{plt}}} F_{K_{\text{plt}}}^{(p+1)} c_n. \]

The transmission equation for the full set of OFDM symbols within one burst becomes
\[ y = H \cdot x + w = X \cdot h + w. \] (5)

The \( MK \times 1 \) column vectors \( y, h, x \) and \( w \) therein are constructed by stacking the respective \( M \) symbol vectors e.g. \( y = (y_0^T, y_1^T, \cdots, y_{M-1}^T)^T \). The transmission equation for the pilot components of the above vector equation is rewritten as
\[ y_p = H_p \cdot x_p + w_p = X_p \cdot h_p + w_p \] (6)
with equivalently the stacked components e.g. \( y_p = (y_{p,0}^T, y_{p,1}^T, \cdots, y_{p,N-1}^T)^T \).

III. LMMSE CHANNEL ESTIMATION

The MIMO LMMSE channel estimator as proposed in [10] is modified to the case of oversampling in subcarrier domain, such that it can make use of more pilot carriers than, there are degrees of freedom in frequency dimension, \( K_{\text{plt}} > \nu + 1 \). The common pilot based LMMSE channel estimator for the used notation is found to be
\[ h_n = E\{h_n| x_p, y_p\} = R_{h_{\text{plt}}^T y_{\text{plt}}^T} R_{y_{\text{plt}} y_{\text{plt}}}^{-1} y_p. \] (7)

For the complete set of pilot symbols the transmission equation in (6) becomes
\[ y_p = \sqrt{K_{\text{plt}}^2} \cdot X_p (I_{N_p} \otimes F_{K_{\text{plt}}}^{(p+1)}) c_p + w_p \] (8)
with \( c_p = (c_p^T, c_p^T, \cdots, c_p^T_{N_p-1})^T \). Here, \( \otimes \) illustrates the Kronecker matrix product. The following paragraphs introduce decompositions of the covariance matrices. Their derivation follows straightforward from the transmission equation (6). For the channel autocorrelation matrix of the pilot subcarriers this results in
\[ R_h \cdot h_n = K_{\text{plt}} (I_{N_p} \otimes F_{K_{\text{plt}}}^{(p+1)}) (R_{c_p} \otimes \rho) (I_{N_p} \otimes F_{K_{\text{plt}}}^{(p+1)})^H \] (9)
and the crosscovariance with the data carriers evaluates as
\[ R_{h, y_p} = \sqrt{K_{\text{plt}}^2} K_{\text{plt}} (I_{N_p} \otimes F_{K_{\text{plt}}}^{(p+1)}) R_{c_p c_p} (I_{N_p} \otimes F_{K_{\text{plt}}}^{(p+1)})^H. \] (10)

Thus, the autocorrelation vector of the channel observations is gained by adaptation to the signal model
\[ R_{y_p, y_p} = K_{\text{plt}} \cdot X_p (I_{N_p} \otimes F_{K_{\text{plt}}}^{(p+1)}) (R_{c_p} \otimes \rho) (I_{N_p} \otimes F_{K_{\text{plt}}}^{(p+1)})^H X_p^H + \sigma_w^2 I_{K_{\text{plt}} N_p}. \] (11)

Similarly, the crosscovariance matrix yields
\[ R_{h, y_p} = \sqrt{K_{\text{plt}}^2} K_{\text{plt}} (I_{N_p} \otimes F_{K_{\text{plt}}}^{(p+1)}) R_{c_p c_p} (I_{N_p} \otimes F_{K_{\text{plt}}}^{(p+1)})^H X_p^H. \] (12)
Now, inserting (11) and (12) into (7) the LMMSE estimator is rewritten as

\[
\hat{h}_n = \sqrt{\frac{K_{\text{ph}}}{K_{\text{ph}}}} \cdot F_{K}^{(n+1)} R_{c,c} \cdot (I_{N_p} \otimes F_{K_{\text{ph}}}^{(n+1)})^H X_p^H \cdot F_{K_{\text{ph}}} \cdot X_p \cdot (I_{N_p} \otimes F_{K_{\text{ph}}}^{(n+1)} \cdot (R_{c,c} \otimes \rho))^{-1} (I_{N_p} \otimes F_{K_{\text{ph}}}^{(n+1)} \cdot R_{c,c} \otimes \rho)^{-1} (I_{N_p} \otimes F_{K_{\text{ph}}}^{(n+1)})^H X_p^H \cdot y_p.
\]

(13)

Invoking the matrix inversion lemma twice on (13) a shorter and more intuitive structure of the estimator results in

\[
\hat{h}_n = \frac{1}{\sigma_p^2} \left[ \frac{K}{K_{\text{ph}}} F_{K}^{(n+1)} R_{c,c} \cdot (R_{c,c} \otimes \rho) + \frac{\sigma_w^2}{K_{\text{ph}} \sigma_p^2} I_{(n+1)N_p} \right]^{-1} (I_{N_p} \otimes F_{K_{\text{ph}}}^{(n+1)})^H X_p^H \cdot y_p.
\]

(14)

with the variance of a pilot symbol \(\sigma_p^2\). The permuted Wiener estimator for the channel taps in time domain is identified in (14) between DFT and IDFT

\[
B_c = R_{c,c} \cdot (R_{c,c} \otimes \rho) + \frac{\sigma_w^2}{K_{\text{ph}} \sigma_p^2} I_{(n+1)N_p}^{-1}.
\]

Hence, the Wiener filter \(B_c\) in (14) can be decomposed into an IDFT of the demodulated pilot symbols. Wiener filtering of the single channel taps and a permuted DFT in order to gain the channel parameters in subcarrier domain. The mean square estimation error (MSEE) of the \(I\)-th delay tap is expressed as

\[
\tilde{\sigma}_{\epsilon,I}^2 = \rho I - \rho I_{c,c,n} \left( \rho R_{c,c} + \frac{\sigma_w^2}{K_{\text{ph}} \sigma_p^2} I_{N_p} \right)^{-1} R_{c,c,n} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \rho_s \left( e^{j\omega} \right) \left[ 1 - \frac{\rho \rho_s \left( e^{j\omega} \right)}{\rho_s \left( e^{j\omega} \right) + \frac{\sigma_w^2}{K_{\text{ph}} \sigma_p^2}} \right] d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} \rho_s \left( e^{j\omega} \right) \frac{\sigma_w^2}{K_{\text{ph}} \sigma_p^2} d\omega
\]

(15)

after reordering \(B_c\) with respect to a single channel tap and considering spectral analysis of Wiener filters [11]. We assess channel estimator performance for the uniform Doppler spectrum with relative bandwidth \(F_D = f_D T_s\). from the integral (15) as

\[
\tilde{\sigma}_{\epsilon,I}^2 = \frac{2 F_D D_p \rho s \rho_s^2}{2 F_D D_p \sigma_w^2 + \rho I_{c,c,n} \sigma_p^2}
\]

(16)

Thus, the subcarrier channel MSEE after the DFT yields

\[
\sigma_n^2 = \sum_{i=0}^{\nu} \frac{2 F_D D_p \rho s \rho_s^2}{2 F_D D_p \sigma_w^2 + \rho I_{c,c,n} \sigma_p^2}
\]

(17)

\[
= \frac{2 F_D D_p \sigma_h^2}{2 F_D D_p + \frac{\sigma_h^2}{\sigma_w^2} K_{\text{ph}} \sigma_p^2 \nu + 1}
\]

\[
= \frac{\sigma_h^2}{1 + \frac{\sigma_h^2}{\sigma_w^2} \beta_n \beta_i}.
\]

(18)

The parameters \(\beta_n = \frac{K_{\text{ph}}}{\nu + 1}\) and \(\beta_i = \frac{D_p n}{2 F_D D_p}\) represent channel oversampling factors in the subcarrier respectively tap domain. One can easily show from (18) that if the minimum sampling rate in tap (Nyquist) and subcarrier dimension is adhered to – i.e. equidistant pilot spacings \(D_s \leq 1/2 F_D\) and \(D_c \leq K/\nu + 1\) –, it does not matter how additional training data is spent. With regard to a constant total amount of training data the MSEE \(\sigma_n^2\) is independent of the distribution of extra training data in time or frequency dimension. For a latter optimization of the achievable data rate this means, it is sufficient to optimize the total amount of training data in terms of \(\beta = \beta_n \beta_i \leq \frac{1}{2 F_D T_s n_{\text{train}}}\) yielding the MSEE

\[
\sigma_n^2 = \frac{\sigma_h^2}{1 + \frac{\sigma_h^2}{\sigma_w^2} \beta_n \beta_i}.
\]

(19)

Note, however that the computations are based on the assumptions of an infinite block length i.e. marginal effects at the edge of the burst and of the used spectrum – where in real systems the estimation error rises – have not been taken into account.

IV. ACHIEVABLE DATA RATE

Starting from the vector of all symbols \(x\) and \(y\), we gather all transmitted data symbols and all received data symbols in a vector \(x_D\) and \(y_D\), respectively, and additionally, the pilot symbols in the vectors \(x\) and \(y\). The mutual information \(I(y; x)\) considers that \(x_D\) is known to the receiver. Hence, it is expressed according to the chain rule as

\[
I(y_D, y_P; x_D | x_P) = I(y_D; x_D | x_P) + I(y_P; x_D | y_D, x_P)
\]

(20)

\[
= I(y_D; x_D | x_P)
\]

(21)

\[
= I(y_D; y_P; x_P)
\]

(22)

\[
= h(y_D | x_D, y_D) - h(y_D | x_D, x_P, y_P)
\]

(23)

where \(h(x)\) denotes the differential entropy of a continuous random variable \(x\). In (22) the additional condition can be inserted, as \(y_P\) does not convey additional information on \(y_D\) or \(x_D\). Since the set of pilot symbols and the set of data symbols are disjoint, \(y_D\) depends neither on \(x_P\), nor on \(y_P\) and, thus, the entropy of the received data symbols simplifies to \(h(y_D | x_P, y_P) = h(y_D)\). Furthermore, the derived channel estimator (7) directly emerges [8] from

\[
E(y_D | x_D, x_P, y_P) = X_D E(h_D | x_D, y_P),
\]

(24)

which is the LMMSE channel estimate multiplied by the data symbols. The random variables \(y_D\) and \(y_P\) are independent, although they contain a multiplication of correlated channel coefficients \(h_n^{(m,z)}\) and \(h_n^{(m,y)}\). But the transmitted iid symbols decorrelate these received signals. By replacing conditioning on the received and transmitted pilot data with the finally employed data set of estimated channel coefficients, it is

\[
I(y_D; x_D | x_P, y_P) = I(y_D; x_D | h).
\]

(25)

This step is somehow misleading as in previous progress no presumptions with respect to the channel have been made. Considering the ergodic mutual information, it is more likely
to have (25) replaced by its expectation with respect to $\hat{h}$
\[
I(y_D; x_D | x_P, y_P) = E_{\hat{h}} \{ h(y_D | \hat{h}) - h(y_D | x_D, \hat{h}) \} = E_{\hat{h}} \left\{ \sum_{k=0}^{l-1} h(y_{d_k} | y_{d_{k+1}}, \ldots, y_{d_{l}}, \hat{h}) - \sum_{k=0}^{l-1} h(y_{d_k} | y_{d_{k+1}}, \ldots, y_{d_{l}}, x_D, \hat{h}) \right\}. \tag{26}
\]
The last step is a corollary of the chain rule for entropies. If it is assumed that the symbols $x_{d_k}$ are perfectly interleaved and as a consequence the $y_{d_k}$ are perfectly deinterleaved, the correlations between the channel realizations are broken down, and the received signal $y_{d_k}$ becomes independent of $y_{d_l}$ with $l \neq k$. Hence, these $y_{d_k}$ have no impact on the above entropies in (26). The same holds true for the channel estimates $\hat{h}_{d_k}$ and the other transmit symbols $x_{d_k}$, with $l \neq k$. Hence, a simplified version of (26) results in
\[
I(y_D; x_D | x_P, y_P) = E_{\hat{h}} \left\{ \sum_{k=0}^{l-1} h(y_{d_k} | \hat{h}_{d_k}) - \sum_{k=0}^{l-1} h(y_{d_k} | x_{d_k}, \hat{h}_{d_k}) \right\}. \tag{27}
\]
In order to evaluate (27), the statistical properties of the summands are analyzed. Observing the gaussian distribution $p(y_{d_k} | \hat{h}_{d_k})$ the first and the second moment are sufficient. The received symbol can be expressed as
\[
y_{d_k} = h_{d_k} x_{d_k} + w_{d_k} = \hat{h}_{d_k} x_{d_k} + (h_{d_k} - \hat{h}_{d_k}) x_{d_k} + w_{d_k} = \hat{h}_{d_k} x_{d_k} + \varepsilon_{d_k} x_{d_k} + w_{d_k} \tag{28}
\]
introducing an estimation error $\varepsilon_{d_k} = h_{d_k} - \hat{h}_{d_k}$. From that observation, it is possible to compute the conditional mean of the received symbol
\[
E \{ y_{d_k} | \hat{h}_{d_k} \} = E \{ \hat{h}_{d_k} x_{d_k} + \varepsilon_{d_k} x_{d_k} + w_{d_k} | \hat{h}_{d_k} \} = \hat{h}_{d_k} x_{d_k} \tag{29}
\]
and the conditional variance
\[
\text{Var} \{ y_{d_k} | \hat{h}_{d_k} \} = |\hat{h}_{d_k}|^2 \sigma_x^2 + \sigma_w^2 \tag{30}
\]
with respect to the channel estimate. If these results are additionally conditioned on the transmitted signal, the expectation of the received signal becomes
\[
E \{ y_{d_k} | x_{d_k}, \hat{h}_{d_k} \} = \hat{h}_{d_k} x_{d_k} \tag{31}
\]
and the autocovariance evaluates as
\[
\text{Var} \{ y_{d_k} | x_{d_k}, \hat{h}_{d_k} \} = \sigma_x^2 |x_{d_k}|^2 + \sigma_w^2 = \sigma_x^2 \sigma_x^2 + \sigma_w^2. \tag{32}
\]
The final step, therein, results from PSK signalling that is essentially sufficient in the low SNR domain. With these components it is possible to rewrite (27) as
\[
I(x_D; y_D | x_P, y_P) = E_{\hat{h}} \left\{ \sum_{k=0}^{l-1} \log \left[ 1 + \frac{|\hat{h}_{d_k}|^2 \sigma_x^2}{\sigma_x^2 \sigma_x^2 + \sigma_w^2} \right] \right\}. \tag{33}
\]
where $\log(a)$ is the logarithm of $a$ to the base 2. When realizing a blockwise transmission of sufficient size in order to approximate perfect interleaving and in order to guarantee near infinite dilution of the channel estimator input, the blockwise mutual information results in
\[
I(y_D; x_D | x_P, y_P) = |D| \cdot E_{\hat{h}_{d_k}} \left\{ \log \left[ 1 + \frac{|\hat{h}_{d_k}|^2 \sigma_x^2}{\sigma_x^2 \sigma_x^2 + \sigma_w^2} \right] \right\} \tag{34}
\]
bearing in mind that $|D| = (M - M_{\text{pre}}) \cdot (K - K_{\text{pl}} / D_{\text{f}})$. If $|D|$ is related to $MK$, the total number of transmitted symbols, the training data penalty on the achievable data rate yields
\[
\frac{|D|}{MK} = \frac{M - M_{\text{pre}}}{M} \left( 1 - 2 \beta f_D T_s \tau \right). \tag{35}
\]
One can easily see that irrespective of the actual, fixed value of $M_{\text{pre}}$ the impact of a preamble vanishes in the limit of high burst lengths. If the achievable data rate is then derived by letting $M$ approach infinity, it becomes
\[
R_{\text{OFDM}} = \lim_{M \rightarrow \infty} \frac{1}{MT_s} I(y_D; x_D | x_P, y_P) = \frac{K}{T_s} \left( 1 - 2 \beta f_D T_s \tau \right) \log(2 \exp(1/\theta) E_{\text{I}}(1/\theta)). \tag{36}
\]
Eq. (37) clarifies that increasing the Doppler frequency $f_D$ influences the achievable data rate in the same way as increasing the maximum channel spread $\tau$. As proposed in [4], we will thus, just review their product $f_D T_s$. If $f_D T_s$ rises and $\beta$ is held constant, $\frac{|D|}{MK}$ is decreased and with it the rate. In case if $|D|$ remains constant, the oversampling $\beta$ is reduced. According to (19) an increasing MSE $\sigma_x^2$, finally also reduces the achievable rate. Therefore, no matter how the system reacts on rising $f_D T_s$, the rate diminishes. This term can also be expressed analytically as e.g. proposed in [7] for block fading. I.e. the rate from (37) can then be expressed by defining the resulting SNR $\theta = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_w^2}$ as
\[
R_{\text{OFDM}} = \frac{K}{T_s} (1 - 2 \beta f_D T_s \tau) \log(2 \exp(1/\theta) E_{\text{I}}(1/\theta)). \tag{38}
\]
Therein, $E_{\text{I}}(x) = \int_x^\infty \frac{\exp(u)}{u} du$ denotes the exponential integral.

An upper bound for the achievable data rate follows from (22) employing Jensens inequality and the concavity of the logarithm
\[
R_{\text{OFDM}} \leq \frac{K}{T_s} \left( 1 - 2 \beta f_D T_s \tau \right) \log [1 + \theta]. \tag{39}
\]

V. RESULTS

Due to numerical problems in the evaluation of (38) for high bandwidths i.e. tiny $\theta$, eq. (38) is evaluated by means of Monte Carlo simulation and is compared to the bound (39). The statistical attributes of the channel estimates $\hat{h}_{d_k}$ are taken from Wiener estimation theory [11]:
\[
E \{ \hat{h}_{d_k} \} = 0 \tag{40}
\]
\[
\text{Var} \{ \hat{h}_{d_k} \} = \sigma_h^2 - \sigma_e^2. \tag{41}
\]
Since the channel coefficient $h_{dk}$ itself has zero expectation, (40) rapidly follows and, according to [11], (41) is valid. Since the channel estimates $\hat{h}_{dk}$ are computed as weighted sums of Gaussian random variables (7) they are themselves Gaussian.

The system is composed according to the MB-OFDM standard, i.e. the overhead is $\tau_{\text{OH}} = 165/128$, the symbol duration $T_s = 312.5\,\text{ns}$ and the subcarrier spacing $B_{sc} = 4.125\,\text{MHz}$. The bandwidth is varied according to multiples of the subcarrier spacing $B = KB_{sc}$. The energy of the sampled noise is $\sigma^2 = B_{sc}N_0 = 1.709 \cdot 10^{-14}\,\text{W}$, the signal power is $\sigma^2 = \frac{P}{\tau B_{sc}}$, and the channel loss is $90\,\text{dB}$ i.e. $\sigma^2 = 10^{-9}$. The transmit power $P$ is fixed for reasons of comparability, although UWB is limited in power spectral density. In fig. 1 at the bandwidth $B$ varies with the bandwidth $D$.

The bandwidth is varied according to multiples of the subcarrier spacing $B = KB_{sc}$. The energy of the sampled noise is $\sigma^2 = B_{sc}N_0 = 1.709 \cdot 10^{-14}\,\text{W}$, the signal power is $\sigma^2 = \frac{P}{\tau B_{sc}}$, and the channel loss is $90\,\text{dB}$ i.e. $\sigma^2 = 10^{-9}$. The transmit power $P$ is fixed for reasons of comparability, although UWB is limited in power spectral density. In fig. 1 at the bandwidth $B$ varies with the bandwidth $D$.

Since the channel coefficient $h_{dk}$ itself has zero expectation, (40) rapidly follows and, according to [11], (41) is valid. Since the channel estimates $\hat{h}_{dk}$ are computed as weighted sums of Gaussian random variables (7) they are themselves Gaussian.

The system is composed according to the MB-OFDM standard, i.e. the overhead is $\tau_{\text{OH}} = 165/128$, the symbol duration $T_s = 312.5\,\text{ns}$ and the subcarrier spacing $B_{sc} = 4.125\,\text{MHz}$. The bandwidth is varied according to multiples of the subcarrier spacing $B = KB_{sc}$. The energy of the sampled noise is $\sigma^2 = B_{sc}N_0 = 1.709 \cdot 10^{-14}\,\text{W}$, the signal power is $\sigma^2 = \frac{P}{\tau B_{sc}}$, and the channel loss is $90\,\text{dB}$ i.e. $\sigma^2 = 10^{-9}$. The transmit power $P$ is fixed for reasons of comparability, although UWB is limited in power spectral density. In fig. 1 at the bandwidth $B$ varies with the bandwidth $D$.

The amount of training data needed to optimize the rate for varying transmit powers ($10\,\text{mW}, 1\,\text{mW}, 0.1\,\text{mW}$) and different products $\hat{f}D_{\tau_{\text{max}}}$ ($10^{-4}, 10^{-5}$) is presented in fig. 2 vs. the bandwidth. The curves are truncated at $B_{\text{crit}}$. Obviously, for each $\hat{f}D_{\tau_{\text{max}}}$ approximately the same percentage of all symbols (and the same $\beta$) used for training, maximizes the rate at the critical bandwidth. Furthermore, $B_{\text{crit}}$ scales with the power $P$, implying that the channel MSEE $\sigma^2$ at $B_{\text{crit}}$ is constant for a fixed $\hat{f}D_{\tau_{\text{max}}}$.

Although not displayed here, evaluation shows that as in the MIMO context [7], the rate is maximized for very large bandwidths, i.e. for low subcarrier SNR, if 50% of all symbols are spent for training.

VI. CONCLUSION

We have extended a methodology for optimizing the training data with respect to the achievable data rate for wideband OFDM systems in frequency selective, continuous fading scenarios in terms of one system parameter. Furthermore, it was possible to verify prior results based on different transmission models or channels.

This methodology can easily be extended covering pilot boosting and multiple antennas, as well as different power delay profiles of the channel and correlated channel taps.

REFERENCES