2D Graph-Based Soft Channel Estimation for MIMO-OFDM

Christopher Knievel, Zhenyu Shi and Peter Adam Hoeher
Faculty of Engineering
University of Kiel
Kaisersr. 2, 24143 Kiel, Germany
Email: {chk,zs,ph}@tf.uni-kiel.de

Gunther Auer
DOCOMO Euro-Labs
80687 Munich, Germany
Email: auer@docomolab-euro.com

Abstract—We address joint channel estimation and data detection based on factor graphs. The considered graph-based approach utilizes reliability information of channel estimates to facilitate soft-output data detection, and in-turn reliability information about the data symbols is taken into account for channel estimation. In this paper graph-based soft channel estimation and detection is extended to an OFDM based air interface, where the channel response varies in two dimensions; time and frequency. Initial channel estimates obtained by training symbols are conveyed by a two dimensional (2D) factor-graph in time and frequency with only a linear increase in complexity. The required training overhead for the proposed 2D graph-based soft channel estimation scheme may be substantially reduced by taking the redundancy introduced by the channel coding into account.

I. INTRODUCTION

To facilitate coherent detection, distributed training symbols (pilots) known at the receiver side are commonly used [1], [2]. In OFDM systems, the channel response at the positions of the unknown data symbols can be reconstructed by means of 2D interpolation and filtering [3], [4]. The main drawback of training-based channel estimation is the induced overhead, which compromises both power efficiency and bandwidth efficiency. For reliable channel estimation the spacing of training symbols in frequency- and time-selective fading channels is upper bounded by the sampling theorem. This problem is exacerbated if transmitters and receivers are equipped with multiple antennas (multiple-input multiple-output, MIMO). While MIMO-OFDM offers promises diversity and/or capacity gains [5], [6], the required training overhead grows proportionally with the number of transmit antennas [7], [8].

Receivers that jointly estimate the channel coefficients and detect the information-carrying data symbols are able to reduce the required training overhead [9], [10] and were adapted to OFDM in [11]. Unfortunately, joint estimation and detection techniques typically incur high computational cost. Iterative receivers that utilize the turbo principle may solicit channel estimation by reliably detected data symbols that serve as auxiliary training symbols [12]–[14]. However, these iterative receivers require tentative channel estimates based on training symbols and are therefore unable to reduce the training overhead. More recently, soft channel estimation based on iterative message-passing over a general factor graph has been applied to MIMO operating in block-fading channels [15], and was shown to attain a substantial reduction in training overhead at modest complexity.

In this paper graph-based soft channel estimation [15] is applied to MIMO-OFDM over time-varying, frequency-selective channels, by extending the underlying factor graph to two dimensions with only linear increase in complexity. Initial channel estimates at training symbol locations are dispersed through the factor graph by message exchange over time and frequency dimensions via so-called Δ-transfer nodes. We observe that in conjunction with channel coding the training pattern are not bound by the sampling theorem, hence, training overhead may be substantially reduced. Simulation results demonstrate that only one training symbol per transmit antenna in the frequency domain is sufficient for the proposed algorithm to converge.

The remainder of this paper is organized as follows: Sec. II introduces the system and channel model. The graph-based iterative receiver as well as the novel Δ-transfer nodes are discussed in Sec. III and simulation results are presented in Sec. IV. Finally Sec. V draws the conclusions.

II. SYSTEM AND CHANNEL MODEL

A MIMO-OFDM system with \( N \) subcarriers and \( K \) OFDM symbols per frame is considered. The equivalent discrete-time model of a MIMO channel with \( N_T \) transmit (Tx) and \( N_R \) receive (Rx) antennas is given by

\[
y_n[l,k] = \sum_{m=1}^{N_T} h_{n,m}[l,k] \cdot x_m[l,k] + w_n[l,k],
\]

\[
h_{n,m}[l,k] = h_{n,m}[l,k] + \sum_{i \neq m}^{N_T} h_{n,i}[l,k] \cdot x_i[l,k] + w_n[l,k] + \text{MAI} + \text{AWGN}
\]

(1)

where \( l \in \{0,1,\ldots,L-1\} \) is the subcarrier index and \( k \in \{0,1,\ldots,K-1\} \) the time index. The received signal \( y_n[l,k] \in \mathbb{C} \) consists of superimposed signals from \( N_T \) transmit antennas. \( h_{n,m}[l,k] \in \mathbb{C} \) represents the channel coefficient of the subchannel connecting the \( n \)th Rx antenna and the \( m \)th Tx antenna. The channel coefficients in this case are...
assumed to be wide sense stationary (WSS), complex Gaussian variables with zero mean. \( x_m[l,k] \in \{ \pm 1 \} \) is the channel input symbol at the \( m \)th Tx antenna. \( w_n[l,k] \) is an additive white Gaussian noise sample with zero mean and variance \( \sigma^2_{w} \). The desired signal from Tx antenna \( m \) is superimposed by \( (N_T-1) \) signals, termed multi-antenna interference (MAI), and the additive white Gaussian noise (AWGN) sample.

A frequency-selective and time-varying fading channel is considered, which means that the channel response changes between adjacent subcarriers and OFDM symbols. Binary antipodal training symbols are periodically inserted, according to the training pattern displayed in Fig. 1. Training symbols associated to different transmit antennas are orthogonally separated in time and frequency, i.e. when one antenna emits a training symbol all other antennas remain silent. \( D_f \) hereby represents the spacing of training symbols in frequency, whereas \( D_t \) is the spacing of training symbols in time.

III. GRAPH-BASED ITERATIVE RECEIVER

Factor graphs are efficient graphical models that have been adopted to a variety of problems in digital communications [16], [17]. In combination with the sum-product algorithm for message exchange, powerful channel estimators have been derived in [15], [18], [19].

A. Receiver Structure

The objective of joint channel estimation and data detection is to estimate two variables, the data symbols \( x_m[l,k] \) and the channel coefficients \( h_{n,m}[l,k] \). Provided that encoded data symbols are sufficiently interleaved, adjacent data symbols will be independent in time and frequency domain. The corresponding factor graph for two transmit and two receive antennas is displayed in Fig. 2 with the unknown data symbols and received symbols (observation nodes) in circles and the channel coefficients (coefficient nodes) in rectangles. Without loss of generality the time and subcarrier index are omitted in Fig. 2, thus representing one subcarrier at one time index.

By connecting the channel coefficients of the neighboring subcarriers via so-called \( \Delta \)-transfer nodes, messages can be exchanged over time and frequency domain throughout the graph. Fig. 3 shows the connections of the channel coefficient nodes with the \( \Delta \)-transfer nodes. The structure of the graph

is independent of the channel characteristics. The functions inside the \( \Delta \)-transfer nodes can be adapted from time-invariant flat-fading channels to fast-fading frequency-selective channels. Note that each domain usually needs its own transfer function depending on the fading characteristic.

If channel coding is applied in time and/or frequency domain the corresponding connections of the data symbols can be either (a) directly implemented as a part of the factor graph or (b) separated from the factor graph. For method (a) Fig. 2 would also incorporate connections to code nodes, whereas the factor graph remains unchanged for method (b). A large variety of channel codes have been successfully implemented in factor graphs [20].

B. Soft Channel Estimation

Define an effective noise sample \( v_{n,m}[l,k] \) of the observation \( y_n[l,k] \) in (1), with respect to (w.r.t.) the transmitted symbol \( x_m[l,k] \) [15]:

\[
v_{n,m}[l,k] = y_n[l,k] - h_{n,m}[l,k]x_m[l,k].
\]

Furthermore, the effective noise sample \( v_{n,m}[l,k] \) and the channel coefficient \( h_{n,m}[l,k] \) are assumed to be Gaussian distributed. \( h_{n,m}[l,k] \) is then fully characterized by \( h_{n,m}[l,k] \sim \mathcal{CN}(\mu_{h_{n,m}[l,k]}, \sigma^2_{h_{n,m}[l,k]}) \). The mean and variance for this channel coefficient can be calculated according

\[
\mu_{\Delta h_{n,m}[l,k]} = \mu_{y_n[l,k]} - \mu_{x_m[l,k]}h_{n,m}[l,k]
\]
to [15]:

\[
\mu_{h_{n,m}[l,k]} = (y_n[l,k] - \mu_{v_{n,m}[l,k]}) (P_{m+1}[l,k] - P_{m-1}[l,k]),
\]

\[
\sigma^2_{h_{n,m}[l,k]} = \sigma^2_{v_{n,m}[l,k]} + 4P_{m+1}[l,k]P_{m-1}[l,k] |y_n[l,k] - \mu_{v_{n,m}[l,k]}|^2,
\]

with \(P_{m+1}[l,k] \doteq P(x_{m}[l,k]) = +1\), \(P_{m-1}[l,k] \doteq P(x_{m}[l,k]) = -1\) being the a-priori information of the data symbols, which is obtained during initialization by training symbols and during iterations by the decoder feedback. The mean \(\mu_{h_{n,m}[l,k]}\) of the channel coefficient represents the estimated value of the coefficient, whereas the variance \(\sigma^2_{h_{n,m}[l,k]}\) can be seen as reliability information of the estimated value. A large variance means a less reliable estimate of the coefficient and vice versa.

C. \(\Delta\)-Transfer Node

Information about channel coefficients are exchanged via transfer nodes throughout the graph. Two neighboring coefficients are connected through a transfer node, which exchange their information according to the sum-product algorithm [16]. We introduce \(\Delta\)-transfer nodes to describe the relation of neighboring coefficients that convert the message of a channel coefficient \(h[l,k]\) to \(h[l+l',k+k']\):

\[
\Delta[l',k'] \doteq h[l,k] - \omega h[l+l',k+k'].
\]

(4)

Only neighboring coefficients of one domain are allowed to connect, that is when \(l' = \pm 1 \Rightarrow k' = 0\) and when \(k' = \pm 1 \Rightarrow l' = 0\). The \(\Delta\)-transfer function is approximated by a Gaussian pdf

\[
\Delta[l',k'] \sim \mathcal{N}(0, \sigma^2_{h}[l',k']).
\]

(5)

The complex factor \(\omega = e^{ij\tau}\) in (4) may be tuned such that the variance \(\sigma^2_{h}[l',k']\) is minimized.

Based on the \(\Delta\)-transfer function (4), information between adjacent channel coefficients is exchanged as follows:

\[
\mu_{h}[l+l',k+k'] = \omega \mu_{h}[l,k],
\]

\[
\sigma^2_{h}[l+l',k+k'] = \sigma^2_{h}[l,k] + \Delta^2_{h}[l',k'].
\]

(7)

As previously described, the mean value represents the estimate of a coefficient and the variance may be interpreted as its reliability information. With the added variance of the \(\Delta\)-transfer function, the reliability of the mean value of the neighboring coefficient decreases. For initialization of the factor graph, the exchange of information of channel coefficients is started at training positions. Their mean values are then distributed throughout the graph, decreasing the reliability of the estimated coefficients with increasing distance to the training position. That implies that mean values of coefficients closer to training are more reliable than those further away. The variance of the \(\Delta\)-transfer function is given by:

\[
\sigma^2_{h}[l',k'] = \mathbb{E}\{|h[l,k] - \omega h[l+l',k+k']|^2\}
\]

\[
=2-2\text{Re}\{\omega \mathbb{E}\{h^*[l,k] \cdot h[l+l',k+k']\}\}
\]

(8)

where \(\mathbb{E}\{ \cdot \}\) denotes the expectation operator.

1) Frequency Domain: Information exchange between channel coefficient nodes located on adjacent subcarriers, \(h[l,k]\) and \(h[l \pm 1,k]\), is facilitated by the transfer nodes \(\Delta_f\) in Fig. 3.

Having no a-priori information about the delay power profile, a uniform distribution within \([0, \tau_{\text{max}}]\) may be assumed, where \(\tau_{\text{max}}\) is the maximum propagation delay. In this case, the variance of the \(\Delta\)-transfer function between adjacent subcarriers \(l\) and \(l \pm 1\) amounts to

\[
\sigma^2_{\Delta,f} = 2 - 2\text{sinc}(\tau_{\text{max}} F) \cdot \text{Re}\{\omega_f e^{j2\pi \tau_{\text{max}} F}\}.
\]

(9)

where \(F\) is the subcarrier spacing. The variance (9) is minimized when its real part is maximized. The imaginary part in (9) diminishes by setting

\[
\omega_f = e^{-j2\pi \tau_{\text{max}} F} = e^{-j\tau_{\text{max}} F}
\]

(10)

which gives the optimum value that minimizes \(\sigma^2_{\Delta,f}\). The minimum variance \(\sigma^2_{\Delta,f}\) then yields

\[
\sigma^2_{\Delta,f} = 2 - 2 \cdot \text{sinc}(\tau_{\text{max}} F).
\]

(11)

Setting \(\omega_f = e^{-j\tau_{\text{max}} F}\) is equivalent to shifting the delay power profile by a factor \(-\tau_{\text{max}}/2\) [3], [4], as depicted in Fig. 4 for a uniform distribution. This shift is optimizing both variance and single-sided “bandwidth” in the case of a uniform distribution, whereas the focus should always be on minimizing the variance.

![Fig. 4. Equivalent shift of propagation delays in time domain.](image)

2) Time Domain: For time-varying channels the \(\Delta\)-transfer function \(\Delta_t\), depicted in Fig. 3, connects channel coefficients of adjacent OFDM symbols, \(h[l,k]\) and \(h[l,k \pm 1]\).

Assuming a uniform Doppler power spectral density between \([-f_{D,\text{max}}, f_{D,\text{max}}]\), the variance of \(\Delta_t\) results in

\[
\sigma^2_{\Delta,t} = 2 - 2 \cdot \text{sinc}(2f_{D,\text{max}} T_s),
\]

(12)

where \(f_{D,\text{max}}\) denotes the maximum Doppler frequency. For a Jakes power spectral density the variance of \(\Delta_t\) yields

\[
\sigma^2_{\Delta,t} = 2 - 2J_0(2\pi f_{D,\text{max}} T_s),
\]

(13)

where \(J_0\) denotes the Bessel function of the first kind and order zero. As both (12) and (13) are real valued, the variance of \(\Delta_t\) is minimized by setting the tuning factor in (4) to \(\omega=1\).

D. Information Exchange at Coefficient Nodes

A coefficient node is connected with two \(\Delta\)-transfer nodes for each domain\(^2\) and one observation node as illustrated in the left part of Fig. 5. Suppose a channel coefficient receives the

\(^2\)At the beginning and the end of each burst only one connection for one or both domains is possible, see also Fig. 3.
messages $p_i \sim \mathcal{CN} \left( \mu_i, \sigma_i^2 \right)$, $i = 1, \ldots, 5$. The messages leaving a coefficient node are generated as depicted in the right part of Fig. 5. Note that the product of Gaussian PDFs results in another Gaussian PDF:

$$\prod_{j=1, j \neq i}^N p_j(h) : \mathcal{CN} \left( \mu_i, \sigma_i^2 \right), \quad (14)$$

with

$$\mu_i = \frac{\sum_{j=1, j \neq i}^{N} \frac{\mu_j}{\sigma_j^2}}{\sum_{j=1, j \neq i}^{N} \frac{1}{\sigma_j^2}}, \quad (15)$$

$$\sigma_i^2 = \frac{1}{\sum_{j=1, j \neq i}^{N} \frac{1}{\sigma_j^2}}, \quad (16)$$

where $N$ represents the number of connected nodes to the channel coefficient.

E. Soft Data Detection

With the updated messages received from a channel coefficient node $h_{n,m}[l,k] \sim \mathcal{CN} \left( \mu_{h_{n,m}[l,k]}, \sigma_{h_{n,m}[l,k]}^2 \right)$ an observation node $y_{n}[l,k]$ can calculate the LLR value of the data symbol $x_{n}[l,k]$ with the following equation [15]:

$$\text{LLR} \left( x_{n}[l,k] \right) \doteq \ln \frac{p( y_{n}[l,k]| x_{n}[l,k] = +1 )}{ p( y_{n}[l,k]| x_{n}[l,k] = -1 )} = 4 \cdot \Re \left\{ \mu_{h_{n,m}[l,k]} ( y_{n}[l,k] - \mu_{v_{n,m}[l,k]} ) \right\} \left( \frac{\sigma_{h_{n,m}[l,k]}^2 + \sigma_{w_{n}[l,k]}^2}{\alpha_{h_{n,m}[l,k]}} \right), \quad (17)$$

It can be seen from equation (17) that the soft information from the channel coefficients are taken into account with $\sigma_{h_{n,m}[l,k]}^2$ in the denominator. An unreliable channel estimate will further reduce the log-likelihood ratio for the corresponding data symbol.

F. Training Pattern

The graph-based iterative receiver is not limited to a specific training pattern. However, training symbols are most effective when they are not concentrated within one burst, but spread throughout it, thus keeping the variance of the estimated channel coefficients almost constant for the whole burst. Since interpolation is not needed, in conjunction with channel coding the spacing of training is not bound to the sampling theorem, which gives a maximum training distance in order to correctly sample the fading process and enable a good channel estimation [1]. However, if channel coding is not taken into account during channel estimation, the spacing of training in two dimensions is given by $D_T f_{D,\text{max}} T_s \leq 1/2$ and $D_T \tau_{\text{max}} T \leq 1$ [3], [4].

G. Message Exchange Scheduling

Message exchange is first started with initial channel estimation for the positions of the training symbols. Their mean and variance values are then distributed throughout the factor graph with the help of the $\Delta$-transfer nodes. From here message exchange rules stay the same for all following iterations. The updated channel coefficients are passed to the observation nodes and a LLR value for the data symbols is calculated. The LLR value is then sent to the symbol nodes. If channel coding is applied, the LLR values are processed accordingly and fed back into the factor graph, where they are used as a-priori information of the data symbols in order to refine the channel coefficient estimates. Information exchanged between nodes are assumed to be extrinsic.

The order of the message exchange for the two domains (time and frequency) has an influence on BER or MSE performance in the lower SNR region. Messages from the domain with the lower variance should be exchanged first as the neighboring coefficient will also receive a lower variance.

IV. NUMERICAL RESULTS

The performance of the proposed 2D graph-based estimator is evaluated by means of MSE and bit error performance. An OFDM system with $N = 128$ subcarriers, $K = 6$ symbols, $N_T = 4$ transmit and $N_R = 4$ receive antennas is implemented. BPSK modulation and a repetition code with code rate 1/4 are used. The WINNER C2 NLOS channel model [21] is applied. The WINNER C2 channel model is characterized by an exponential decaying delay power profile and a Jakes power spectral density. A velocity of 60 km/h at a carrier frequency of 4 GHz is chosen, resulting in a normalized maximum Doppler shift $f_{D,\text{max}} T_s \approx 0.01587$. The maximum propagation delay $\tau_{\text{max}}$ is set to 1845 ns. Given these parameters, according to the sampling theorem the maximum spacing of training results in $D_T \leq 36$ (for an uniform delay power profile) and $D_T \leq 31$. The variances for the $\Delta$-transfer functions were chosen according to equation (9) and (12), hence the only a-priori information needed are the maximum Doppler frequency $f_{D,\text{max}}$ and the maximum propagation delay $\tau_{\text{max}}$.

In order to minimize $\sigma_{h_{n,m}[l,k]}^2$ for an exponential delay power profile, the tuning factor is chosen to be $\omega_T = e^{-j 2 \pi \frac{\tau_{\text{max}}}{T}}$. The BER and MSE performances depending on the spreading of training symbols are shown in Fig. 6 and Fig. 7, respectively. The sampling theorem is violated for the last two curves in the figures. With only one training symbol in the frequency domain the MSE curve roughly converges at 10 dB. The BER performance is improved by using at least two training symbols in the frequency domain ($D_T = 64$). By lowering the code rate the spacing of training could be increased.
of the receiver grows only linear with the number of Tx and Rx antennas, the number of OFDM subcarriers and OFDM symbols, as well as the code rate and the number of dimensions. The proposed channel estimator can nicely be extended to an arbitrary number of further dimensions.

REFERENCES


Fig. 6. Bit error performance for different spacings of training symbols with N=128, K=6, R=1/4 and BPSK modulation.

Fig. 7. MSE performance for different spacings of training symbols with N=128, K=6, R=1/4 and BPSK modulation. The performance of linear interpolation is plotted for comparison with the same spacings.

without performance degradation, however, there is a trade-off between data rate and training overhead. Furthermore, one can observe that MSE and BER performance with $D_t = 16$ and $D_f = 32$ can not be further improved by lowering the spacing of training. Training overhead in this case is reduced to only 3%. The BER performance at $10^{-4}$ with perfect channel knowledge is about 2.7 dB better compared to the performance with channel estimation and a training spacing of $D_t = 16$ and $D_f = 32$. MSE results for linear interpolation are shown for comparison in Fig. 7. It can be seen that graph-based channel estimation outperforms linear interpolation for the complete SNR range. An error floor appears at high SNRs and large training spacings when linear interpolation is used, whereas the performance of graph-based channel estimation is not affected by the spacing of training.

V. CONCLUSION

A graph-based iterative soft channel estimation and data detection scheme for time- and frequency-selective channels is presented in this paper. Training overhead can be significantly reduced by message exchange with novel $\Delta$-transfer nodes. As a-priori information, only the maximum propagation delay and the maximum Doppler shift are needed. The complexity