Coding Approaches for Multiple Antenna Transmission in Fast Fading and OFDM

Stefan H. Müller-Weinfurtner, *Member, IEEE*

***Abstract*—Multiple-antenna channel coding for orthogonal frequency-division multiplexing (OFDM) transmission over dispersive channels is reconsidered because with frequency interleaving, the effective channel characteristic across subcarriers is rather fast fading. The channel does not comply with the quasistatic model widely assumed for space-time trellis codes (STCs). For that reason, we first study the ideal fast-fading multiple transmit and receive antenna channel and then compare the performance of STCs with that of bit-interleaved coded modulation in fast fading. Mutual information of the ergodic channel is evaluated for numerous modulation scenarios, and capacity comparisons generate guidelines on how to jointly adjust coding rate and modulation cardinality. Bit-based coding offers large flexibility in rate adaptation, and simulation results show that it outperforms STCs in ideal fast fading and, finally, in a realistic OFDM application as well.**

***Index Terms*—BICM, bit interleaving, capacity, fast fading, multiple antennas, OFDM, space-time channel coding, STC, wireless.**

I. Introduction

CHANNEL coding for wireless communications equipment with transmit and receive diversity is a fairly new field of research [1]–[3]. Performance limits in terms of outage capacity are derived in [4] for the quasistatic fading channel, whereas [5] also covers the capacity for the ergodic fast fading channel.

Space-time trellis codes (STCs) [6]–[8] are well suited for multiple antenna transmission systems with a quasistatic fading (i.e., block-fading) environment, but our ultimate aim is to apply coding in orthogonal frequency-division multiplexing (OFDM) systems [9]–[13]. We need to realize that the channel characteristic across subcarriers in OFDM does not comply with the widely used quasistatic channel model. Fig. 1 illustrates the magnitude of the channel transfer function $H[\nu]$ versus the subcarrier number ν for two independent realizations of a multipath channel of length 10 in an OFDM system with 256 subcarriers. Even for this rather short channel, we experience significant fading behavior for coding across subcarriers, i.e., in the frequency direction. For longer channels, variation across frequency becomes even "faster." Adjacent channel coefficients are not independent, but together with frequency interleaving within one OFDM symbol, the resulting channel characteristics can be approximated by an independent fast fading channel model.

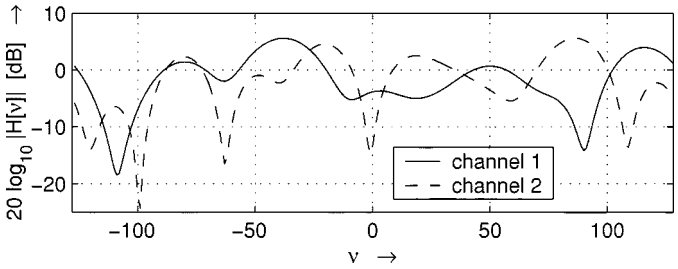

Fig. 1. Magnitude of channel transfer function $H[\nu]$ versus subcarrier index ν for two independent multipath channels of length 10 with linearly decaying power delay profile in a 256-subcarrier OFDM system.

Obviously, we need to reconsider the channel coding problem in OFDM to make best use of exploitable frequency diversity of the transformed multipath channel. Multidimensional signal sets are introduced in [14] to increase the diversity factor in OFDM. Recent progress has been made for the multiple antenna case by optimizing STCs for fast fading [15]–[17] or by applying the idea of I–Q (inphase and quadrature component) interleaving to STCs [18], [19]. We will not investigate the latter method because it also increases diversity in bit-based coding [20]. The question remains as to whether or not STCs are an appropriate channel coding class for fast(er) fading scenarios or for OFDM.

We compare STCs and bit-interleaved coded modulation (BICM) [21], [22] for multiple antenna scenarios with fast—or at least faster than quasistatic—fading conditions. Our objective is to provide extensive link-level capacity results for a variety of modulation schemes and to present interesting comparisons. They in turn provide new insights, which allow the joint adaptation of the rate for the underlying binary codes and the modulation scheme in BICM. No thoughts are given to "full diversity order," and interestingly enough, the BICM schemes perform very well. From simulation results, we later conclude that BICM outperforms STCs because BICM relies on binary codes instead of being signal-space codes like STCs. With comparable code trellis complexity, BICM achieves larger Hamming distances [20], which are beneficial in fast fading scenarios. Further, a larger flexibility in rate adaptation is achieved, which is desirable in packet data communication.

II. Channel Model

For the theoretical study, we put OFDM aside and consider ideal fast fading single-carrier transmission with n_t transmit an-

Manuscript received May 14, 2001; revised April 23, 2002. The associate editor coordinating the review of this paper and approving it for publication was Prof. Georgios B. Giannakis.

The author was with the Wireless Systems Research Department, AT&T Labs-Research, Middletown, NJ 07748 USA. He is now with Philips Semiconductors TCMC, Nürnberg, Germany (e-mail: smw@ieee.org).

Publisher Item Identifier 10.1109/TSP.2002.803351.

tennas (Tx) and n_r receive antennas (Rx). The signal constellation used for modulation in the Tx branch j is $\mathcal{A}_j$, and the transmitted signal points are equiprobable in either of the n_t transmit antennas. The independently chosen signal points are transmitted simultaneously. In each frame, L symbols are transmitted from each antenna. Each symbol is separated in time by the modulation interval T. For time step ℓ, the transmitted symbols are collected in the hypersymbol ($n_t \times 1$ vector of constituent symbols) $\boldsymbol{a}[\ell] = [a_1[\ell], \ldots, a_{n_t}[\ell]]^{\mathsf{T}}$. Hence, $\boldsymbol{a}[\ell] \in \mathcal{A}_1 \times \cdots \times \mathcal{A}_{n_t}$, where $\times$ denotes the Cartesian product of constituent signal constellations. The hypersymbol is transmitted via the noisy channel to obtain

$$\boldsymbol{r}[\ell] = \boldsymbol{H}[\ell]\boldsymbol{a}[\ell] + \boldsymbol{n}[\ell], \quad 0 \le \ell \le L-1. \tag{1}$$

Obviously, $\boldsymbol{H}[\ell]$ is the $n_r \times n_t$ matrix with the channel coefficients $H_{ij}[\ell]$, which describe the transmission characteristic between the jth transmit and ith receive antenna. The $n_r \times 1$ vector $\boldsymbol{r}[\ell] = [r_1[\ell], \ldots, r_{n_r}[\ell]]^{\mathsf{T}}$ represents the received samples at time ℓ. The noise samples in the $n_r \times 1$ noise vector $\boldsymbol{n}[\ell]$ are assumed to be mutually independent zero-mean complex Gaussian variates so that the covariance matrix is $\mathcal{E}_{\boldsymbol{n}[\ell_1],\boldsymbol{n}[\ell_2]}\{\boldsymbol{n}[\ell_1]\boldsymbol{n}[\ell_2]^{\mathsf{H}}\} = \delta[\ell_1 - \ell_2]\sigma_n^2 \boldsymbol{I}_{n_r}$, where

$\delta[\cdot]$ Kronecker delta;
$\sigma_n^2 = N_0/T$ variance per complex dimension;
$\boldsymbol{I}_n$ $n \times n$ identity matrix.

N_0 is the one-sided power spectral density of the white noise. We further introduce $\mathcal{E}\{|\boldsymbol{a}[\ell]|^2\} = \mathcal{E}\{\boldsymbol{a}^{\mathsf{H}}[\ell]\boldsymbol{a}[\ell]\} = E_a/T$, where E_a is the average energy per hypersymbol, i.e., the average total energy transmitted per time step. Together with $\mathcal{E}\{|H_{ij}[\ell]|^2\} = 1, \forall\, i,j$, and independent $H_{ij}[\ell]$, we have $\mathcal{E}\{|\boldsymbol{H}[\ell]\boldsymbol{a}[\ell]|^2\} = n_r E_a/T$, and the average signal-to-noise ratio (SNR) per receive antenna is E_a/N_0.

The following comparisons will be based on finite and discrete signal sets like they are used in practical transmission systems. We introduce the total number of bits carried in one hypersymbol $\boldsymbol{a}[\ell]$ as

$$\Lambda = \sum_{j=1}^{n_t} \log_2 |\mathcal{A}_j| \tag{2}$$

which is equivalent to the number of bit levels in the transmission scheme. $|\mathcal{A}_j|$ is the cardinality of the constituent signal constellation $\mathcal{A}_j$. For digital transmission, we map Λ bits b^{λ} into one hypersymbol, and the binary vector $[b^0, \ldots, b^{\Lambda-1}]$ is the so-called label of the hypersymbol.

III. Coding Architectures

We give a brief overview of two channel coding architectures for multiple antenna transmission from the recent literature, which we will investigate.

A. Space-Time Trellis Codes

STCs are presented in [6] and further discussed in [7], [8]. The coding architecture is sketched in Fig. 2. Binary information $b[\iota]$ enters the STC, and in each time step, a complex-valued symbol for each antenna is generated according to a code trellis

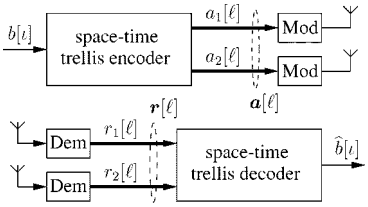

Fig. 2. Transmitter and receiver in a 2×2 space-time coding transmission system.

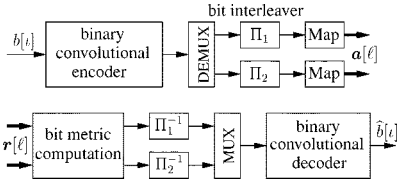

Fig. 3. Transmitter and receiver in a 2×2 bit-interleaved transmission.

in such a fashion that the diversity and/or coding gain is maximized. The signal constellations used in all transmit antennas are identical and denoted by $\mathcal{A}$. In our view of channel coding, all two-antenna STCs in [6] are effectively rate-1/2 codes, which means that for each 2×1 output vector $\boldsymbol{a}[\ell] \in \mathcal{A} \times \mathcal{A}$ (which could carry $2\log_2|\mathcal{A}|$ bits), only $\log_2|\mathcal{A}|$ bits are entering the STC. The space-time decoder directly operates on the vector of received signal-space samples to estimate the most likely transmitted information sequence of binary decisions $\hat{b}[\iota]$.

STCs are signal-space codes, which already indicates that they might perform poorly in fast fading channels with much time diversity. Furthermore, their inflexible effective coding rate (in most cases $1/n_t$) is undesirable if one aims to achieve a fine-grained link adaptation to make best use of the available channel.

B. Bit-Interleaved Coded Modulation

The basic idea of BICM [22] can also be expanded to multiple antenna transmission [5], [23] to obtain advantages in fast fading channels. Fig. 3 illustrates the coding architecture that is used in [5], where again, bold lines indicate complex values, whereas finer lines represent binary values or metrics for them. A single convolutional code (CC) is used to encode the information bits $b[\iota]$. The coded bits are cyclically demultiplexed into the Tx branches, where they are bit-interleaved by different interleavers Π_1 and Π_2 and mapped via Gray labeling onto the signal constellation $\mathcal{A}$ used in both Tx branches and transmitted. In the receiver, bit metrics are calculated independently for each bit, ignoring the values of other bits in this vector. The bit metrics are deinterleaved and multiplexed into one stream, which is decoded by a conventional soft-input Viterbi algorithm.

The complexity of the bit metric computation is $\sim |\mathcal{A}|^{n_t}$ and rises exponentially with n_t. Via the underlying and well-known convolutional codes, the effective coding rate can be adapted in fine-grained steps by the use of actual rate-k/m CCs or by puncturing of a mother code of rate $1/m$.

IV. Link-Level Capacity

First, we review expressions for mutual information achieved with arbitrary Tx signal constellation mixtures and with perfect

knowledge of the channel in the receiver only. Their numerical evaluation then forms the foundation for extensive capacity results and interesting comparisons. Further, the metric used in our simulations is given for easy reference.

A. Mutual Information

We investigate the conditional mutual information [5]

$$I(\boldsymbol{a};\boldsymbol{r}\mid\boldsymbol{H}) = H(\boldsymbol{a}) - H(\boldsymbol{a}\mid\boldsymbol{r},\boldsymbol{H}) \tag{3}$$

$$= \mathcal{E}_{\boldsymbol{a}}\{-\log_2\Pr\{\boldsymbol{a}\}\} - \mathcal{E}_{\boldsymbol{a},\boldsymbol{r}}\{-\log_2\Pr\{\boldsymbol{a}\mid\boldsymbol{r},\boldsymbol{H}\}\} \tag{4}$$

$$= \Lambda - \mathcal{E}_{\boldsymbol{a},\boldsymbol{r}}\left\{\log_2\frac{\sum_{\tilde{\boldsymbol{a}}\in\mathcal{A}_1\times\cdots\times\mathcal{A}_{n_t}} p_{\boldsymbol{r}}(\boldsymbol{r}\mid\tilde{\boldsymbol{a}},\boldsymbol{H})}{p_{\boldsymbol{r}}(\boldsymbol{r}\mid\boldsymbol{a},\boldsymbol{H})}\right\} \tag{5}$$

which is measured in bits per hypersymbol and gives the reduction in the uncertainty (entropy $H(\boldsymbol{a})$) of $\boldsymbol{a}$ due to the knowledge of $\boldsymbol{r}$ for one specific known channel realization $\boldsymbol{H}$. The last step followed from assuming equiprobable transmit hypersymbols $\boldsymbol{a}$, i.e., $\Pr\{\boldsymbol{a}\} = 1/\prod_{j=1}^{n_t}|\mathcal{A}_j| = 2^{-\Lambda}$ [note (2)]. For the pdf of $\boldsymbol{r}$, which is required in (5), we assume independent white Gaussian noise and have

$$p_{\boldsymbol{r}}(\boldsymbol{r}\mid\boldsymbol{a},\boldsymbol{H}) = \frac{1}{(\pi\sigma_n^2)^{n_r}}\exp\left(-\frac{|\boldsymbol{r}-\boldsymbol{H}\boldsymbol{a}|^2}{\sigma_n^2}\right). \tag{6}$$

B. Bit Probabilities and Mutual Information on Bitlevels

We obtain the mutual information for the λ-th bitlevel addressing the vector $\boldsymbol{a}[\ell]$, assuming that all other bit positions in the label of $\boldsymbol{a}[\ell]$ are unknown, as [5]

$$I(b^\lambda;\boldsymbol{r}\mid\boldsymbol{H}) = H(b^\lambda) - H(b^\lambda\mid\boldsymbol{r},\boldsymbol{H}) \tag{7}$$

$$= \mathcal{E}_{b^\lambda}\left\{-\log_2\Pr\{b^\lambda\}\right\} - \mathcal{E}_{b^\lambda,\boldsymbol{r}}\left\{-\log_2\Pr\{b^\lambda\mid\boldsymbol{r},\boldsymbol{H}\}\right\} \tag{8}$$

$$= 1 - \mathcal{E}_{b^\lambda,\boldsymbol{r}}\left\{\log_2\frac{\sum_{\tilde{b}^\lambda\in\{0,1\}} p_{\boldsymbol{r}}(\boldsymbol{r}\mid\tilde{b}^\lambda,\boldsymbol{H})}{p_{\boldsymbol{r}}(\boldsymbol{r}\mid b^\lambda,\boldsymbol{H})}\right\}. \tag{9}$$

The last step followed from the assumption of equiprobable bit values b^λ, i.e., $\Pr\{b^\lambda\} = 1/2$. For the conditional pdf of $\boldsymbol{r}$, which is required in (9), we have

$$p_{\boldsymbol{r}}(\boldsymbol{r}\mid b^\lambda,\boldsymbol{H}) = \frac{1}{|\mathcal{A}(b^\lambda)|}\sum_{\tilde{\boldsymbol{a}}\in\mathcal{A}(b^\lambda)} p_{\boldsymbol{r}}(\boldsymbol{r}\mid\tilde{\boldsymbol{a}},\boldsymbol{H}) \tag{10}$$

where $\mathcal{A}(b^\lambda)$ is that subset of the hypersymbol constellation that complies with the bit value at the λth label position as demanded by b^λ. Clearly, (6) is used to calculate the probabilities in (10).

C. Channel With Ideal Fast Fading

We assume a fast-fading channel model so that the channel matrix $\boldsymbol{H}[\ell]$ is random and takes independent values for each ℓ. The matrix entries $H_{ij}[\ell]$, which are the channel coefficients between Tx antenna j and Rx antenna i, are mutually uncorrelated zero-mean complex Gaussian random variables. Hence, the magnitude of each entry follows a Rayleigh distribution. The channel parameter is perfectly known for each ℓ, and we are interested in the average mutual information (AMI) for coded modulation (CM). Due to fast fading, the channel can be assumed to be ergodic so that the desired AMI is obtained by averaging over the given channel statistic. This yields [22]

$$I_{\mathrm{CM}} = \mathcal{E}_{\boldsymbol{H}}\{I(\boldsymbol{a};\boldsymbol{r}\mid\boldsymbol{H})\}. \tag{11}$$

In perfect correspondence, we introduce the bitlevel AMI of the ergodic channel

$$I^\lambda = \mathcal{E}_{\boldsymbol{H}}\{I(b^\lambda;\boldsymbol{r}\mid\boldsymbol{H})\} \tag{12}$$

which is related to the overall AMI of BICM schemes via [22]

$$I_{\mathrm{BICM}} = \sum_{\lambda=0}^{\Lambda-1} I^\lambda. \tag{13}$$

I_{BICM} can never be larger than I_{CM} because the bitlevels are treated as independent binary channels without exploiting known bits from other levels when decoding the current level.

D. Near-Optimum Bit Metrics

For Viterbi channel decoding of the bit-based coding architectures, we always use the log-likelihood metric

$$L(b^\lambda\mid\boldsymbol{r},\boldsymbol{H}) = \ln\frac{\Pr\{b^\lambda=1\mid\boldsymbol{r},\boldsymbol{H}\}}{\Pr\{b^\lambda=0\mid\boldsymbol{r},\boldsymbol{H}\}} \tag{14}$$

$$\approx \ln\frac{\max_{\tilde{\boldsymbol{a}}\in\mathcal{A}(b^\lambda=1)} p_{\boldsymbol{r}}(\boldsymbol{r}\mid\tilde{\boldsymbol{a}},\boldsymbol{H})}{\max_{\tilde{\boldsymbol{a}}\in\mathcal{A}(b^\lambda=0)} p_{\boldsymbol{r}}(\boldsymbol{r}\mid\tilde{\boldsymbol{a}},\boldsymbol{H})} \tag{15}$$

$$= \frac{1}{\sigma_n^2}\left(\min_{\tilde{\boldsymbol{a}}\in\mathcal{A}(b^\lambda=0)}|\boldsymbol{r}-\boldsymbol{H}\tilde{\boldsymbol{a}}|^2 - \min_{\tilde{\boldsymbol{a}}\in\mathcal{A}(b^\lambda=1)}|\boldsymbol{r}-\boldsymbol{H}\tilde{\boldsymbol{a}}|^2\right) \tag{16}$$

which has a reasonable complexity after the log-sum approximation, as only Euclidean distances in a multidimensional space need to be calculated, and the nearest representative for the respective bit value needs to be found. Nonetheless, near-optimum performance is achieved by this simplified metric [22].

E. Link-Level Capacity Evaluation in Fast Fading

The link-level results are obtained by randomly generating matrix channels and evaluating (11)–(13) in a Monte-Carlo integration fashion. The curves converge very fast and can be considered to be fairly exact. As an initial example, Fig. 4 shows results for the mutual information achievable in fast fading with transmission schemes for a maximum of $\Lambda = 4$ bits per channel use. This can be done by either transmitting 16QAM from one antenna or by transmitting 4PSK from each of two transmit antennas. To have a fair comparison, both schemes use either only one receive antenna or optimum receive diversity with two receive antennas.

Let us first consider the case with two receive antennas. We see two curves for each of the transmission schemes. One is for CM, and the other is for BICM, which strongly relies on the Gray labeling of the constituent signal constellations. For 16QAM from 1 Tx, we see the well-known result that the curves for CM and BICM with Gray labeling nearly coincide for large

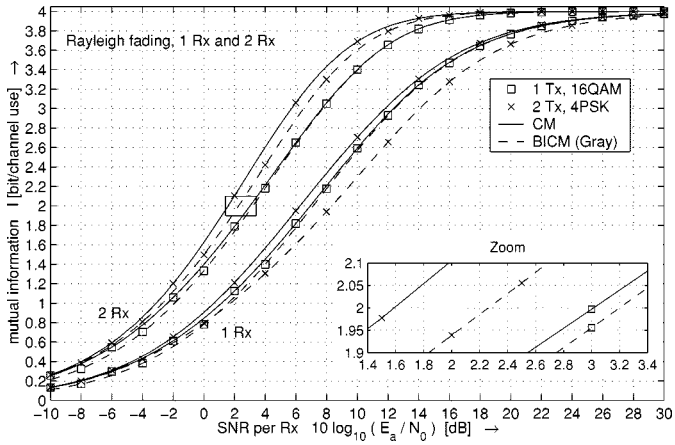

Fig. 4. Mutual information versus average SNR per Rx antenna for transmission schemes with $\Lambda = 4$ bits per channel use. 16QAM transmitted from one Tx and 4PSK from two Tx are shown.

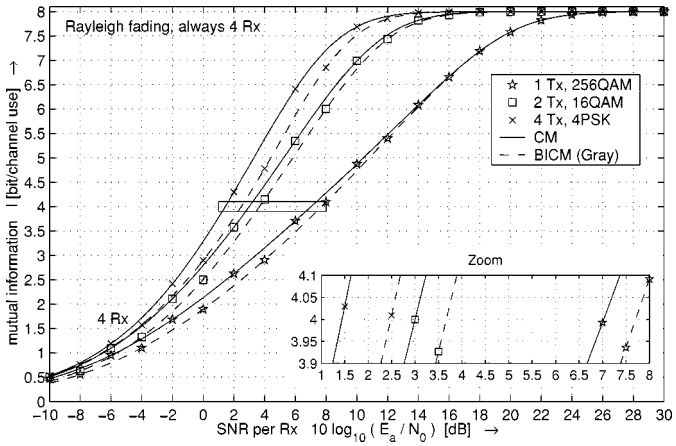

Fig. 5. Mutual information versus average SNR per Rx antenna for transmission schemes with $\Lambda = 8$ bits per channel use. 256QAM transmitted from one Tx, 16QAM from two Tx, and 4PSK from four Tx are shown.

SNRs while exhibiting a negligible gap at lower SNRs [22]. To be able to compare BICM to STCs, which all operate with an effective $R = 1/2$ coding rate, we have to take a close look at the horizontal line at spectral efficiency 2. The zoomed area reveals that the SNR loss due to using BICM instead of CM is 0.2 dB for 16QAM. In contrast, for 4PSK from 2 Tx, the gap between CM and BICM is larger [5], and we find a 0.65-dB loss at spectral efficiency 2. Nonetheless, an SNR advantage of 0.95 dB remains when moving from 1 Tx 16QAM BICM to 2 Tx 4PSK BICM. Even better, a 1.6-dB SNR advantage exists when we compare 1 Tx BICM with 2 Tx CM. Clearly, this is a capacity comparison only, but the BICM coding architecture translates these predicted gains into an appropriately shifted error rate curve, as we will see in Section V. We further point out that the SNR advantage of multiple transmit antennas over a single transmit antenna is even more obvious at higher spectral efficiencies so that the use of coding schemes with $R > 1/2$ is an interesting direction to be investigated.

In terms of capacity, one should abstain from operating 2 Tx 4PSK in fast fading with one Rx antenna only. It is alarming that the capacity of 2 Tx 4PSK CM is only slightly better than 1 Tx 16QAM CM, and for BICM, 2 Tx 4PSK is even worse than 1 Tx 16QAM. It appears that in this fast fading scenario, pure receive diversity is more beneficial than pure transmit diversity if finite discrete signal sets are used instead of Gaussian inputs. The reason for this is that the combination of 2 Tx with 1 Rx leads for some channel realizations to ambiguous hypersymbol constellations so that (e.g., for $H_{11} = \pm H_{12}$) the total number of distinguishable signal points in the received hypersymbol set cannot be guaranteed to be 2^{Λ}. The use of a second Rx ensures it will be much less likely that that this ambiguity will occur. Hence, transmit diversity schemes with $n_t = 2$ in fast fading should always be operated with $n_r \geq 2$, as otherwise, they might be outperformed by higher order modulation systems with one Tx (i.e., well-defined signal constellation), which have even further advantages in terms of interference cancellation and, hence, in overall system capacity of cellular systems [24].

Fig. 5 illustrates the mutual information achievable with transmission schemes for a maximum of 8 bits per channel use. This is done by either transmitting 256QAM from one antenna, 16QAM from two antennas, or by transmitting 4PSK from each of four transmit antennas. Again, we aim at a fair comparison and investigate the situation with four branch receive diversity in all schemes. Here, the differences are much larger than in Fig. 4 and at spectral efficiency 4, we gain 5.2 and 4 dB for 4 Tx and 2 Tx, respectively, if we consider the BICM schemes. The SNR gain for 4 Tx 4PSK CM over 1 Tx 256QAM BICM is even 6.2 dB, but it needs closer investigation as to whether or not the 1.2-to-1.5-dB additional gain of the 4×4 4PSK system over the 2×4 16QAM system justifies doubling the number of required Tx power amplifiers. Further, the former requires estimation of 16 channel coefficients, whereas the latter needs "only" 8 while enabling further system capacity improvements by interference suppression due to $n_r > n_t$. Especially in Fig. 5, it becomes apparent that the advantage of transmit diversity in fast fading is increasingly visible at spectral efficiencies larger than $\Lambda/2$ for which coding rates around $R = 3/4$ would be interesting. It appears that a large target spectral efficiency is required to make best use of the benefits of transmit diversity.

These figures indicate the benefits of transmit diversity to achieve one and the same maximum value of spectral efficiency but does not yet give a design guideline for the combination of coding rate and modulation types to obtain a given overall target spectral efficiency. For this purpose, we compare the AMI of various signal sets in a 2×2 system in Fig. 6. Let us first fix three bits per channel use as the target in the small zoomed area. Given the larger metric computation complexity as well as higher accuracy requirements for channel estimation in 8PSK systems when compared with 4PSK, it might be interesting to use 2 Tx 4PSK with a rate 3/4 code instead of the 2 Tx 8PSK with a rate 1/2 code, especially if BICM is considered, where the capacity gap is 0.5 dB. If we now consider the second zoomed area for 4 bits per channel use, the same reasoning applies to the comparison of 2 Tx 16QAM with rate 1/2 and 2 Tx 8PSK with rate 2/3. Here, we have an SNR gap of 0.9 dB for CM, which reduces to 0.5 dB for BICM. Note also the large performance gap between BICM and CM, which is 1.5 dB for 16QAM and 1.2 dB for 8PSK. This penalty is lowered if more Rx antennas are used, like in Fig. 7, which shows results for the same transmitter

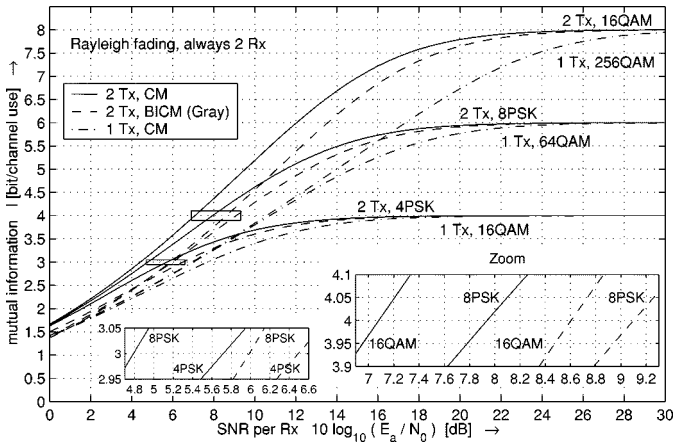

Fig. 6. Mutual information versus average SNR per Rx antenna (using 2 Rx) for transmission schemes with one and two Tx for various maximum spectral efficiencies Λ.

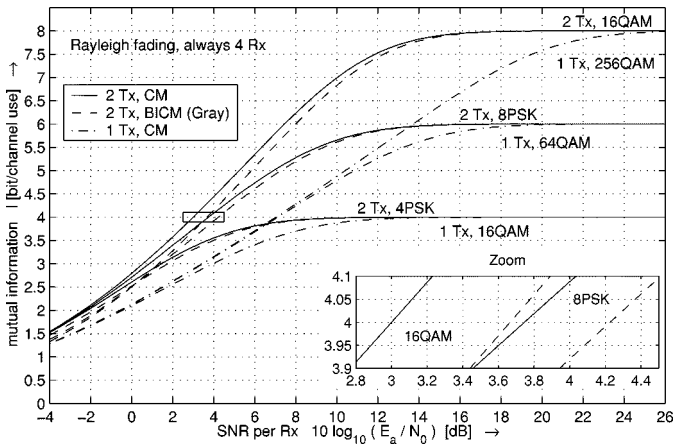

Fig. 7. Mutual information versus average SNR per Rx antenna (using 4 Rx) for transmission schemes with one and two Tx for various maximum spectral efficiencies Λ.

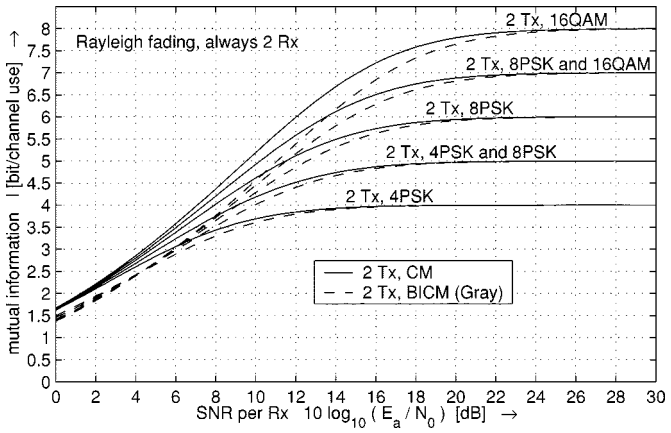

Fig. 8. Mutual information versus average SNR per Rx antenna (using 2 Rx) for transmission schemes with two Tx for various spectral efficiencies. The signal constellations in both antennas have identical average power.

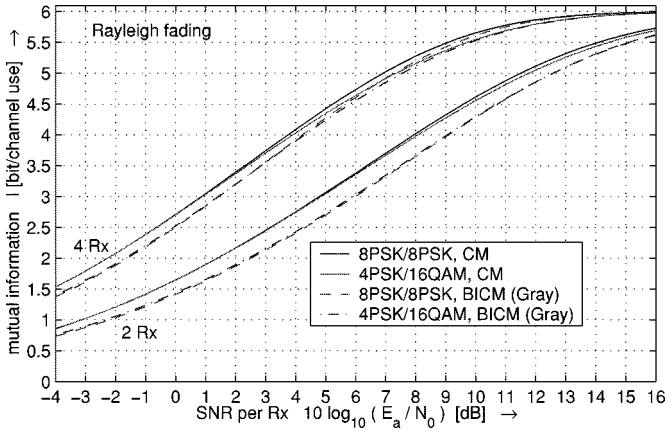

Fig. 9. Mutual information versus average SNR per Rx antenna (using 2 and 4 Rx) for transmission with 2 Tx using 8PSK/8PSK or 4PSK/16QAM to achieve spectral efficiency 6. The signal constellations in both antennas have identical average power.

scenario but now with 4 Rx. We again concentrate on 4 bits per channel use and see that in comparison with Fig. 6, the performance penalty for using BICM instead of CM is lowered to 0.65 and 0.45 dB for 16QAM and 8PSK, respectively. Hence, a larger number of receive than transmit antennas, i.e., $n_r > n_t$, reduces the capacity gap between BICM and CM. The SNR gap at 4 bit per channel use for BICM 16QAM with rate 1/2 and 8PSK with rate 2/3 is with 0.55 dB almost the same as the 0.5 dB in Fig. 6.

Especially with larger numbers of Tx antennas n_t, the spectral efficiencies provided by transmission schemes with equal signal sets in all Tx antennas increases in steps of n_t, which might be undesirable if the data rate needs to be adjusted in fine-grained steps. The aforementioned combination of a smaller hypersymbol set with higher rate coding is attractive in transmit diversity schemes, especially in terms of the complexity for metric computation in BICM. Hence, it is reasonable to consider nonequal signal constellations to be transmitted from the different Tx antennas. Fig. 8 shows results for *mixed signal constellations* with equal average power in a 2×2 system, enabling a bit-wise adaptation of spectral efficiencies for the underlying signal constellations during joint design of modulation cardinality and channel coding rate.

The mixing of signal constellations can be driven to extremes. Let us consider a transmission system with a maximum spectral efficiency of $\Lambda = 6$ bits per channel use with 2 Tx antennas. Naturally, this can be accomplished by the use of 2 Tx antennas with 8PSK. For successive interference cancellation schemes like in vertical Bell Laboratories layered space-time (V-BLAST) [1], it might also be interesting to combine 4PSK with 16QAM, which also leads to $\Lambda = 6$ bits per channel use. In this sense, the 4PSK could be decoded first, using the decisions to cancel its interference for the final decoding of the 16QAM stream. Clearly, only CM reflects use of decisions from other bitlevels. BICM does not include exploitation of other bit decisions and is investigated for completeness only. In Fig. 9, we evaluate the average mutual information achieved by these two systems with 2 and 4 Rx antennas, and amazingly, almost no difference can be seen at 3 bit per channel use. Only at spectral efficiencies larger than 3.5 can an increasing advantage of the "balanced" 8PSK/8PSK scheme over the "unbalanced" 4PSK/16QAM scheme be seen for CM and, in addition, with 4 Rx for BICM; the curves for BICM with 2 Rx virtually coincide. Hence, it seems that for minor imbalance of bit content transmitted from different antennas, almost no performance penalty

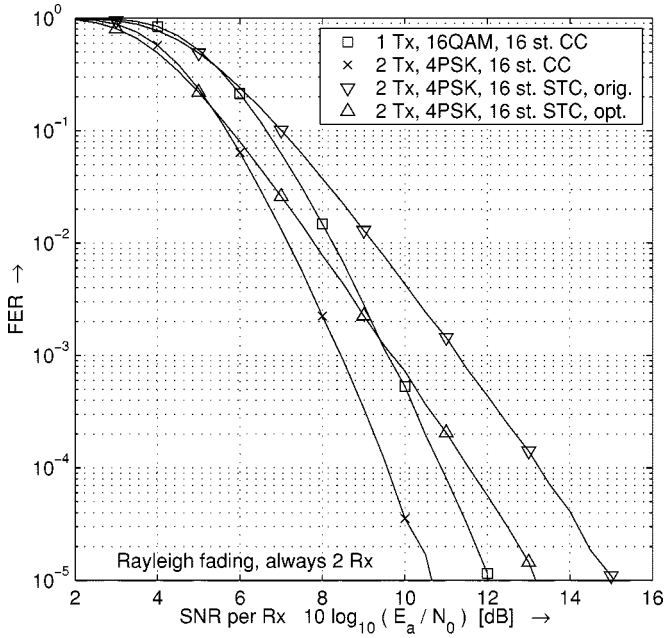

Fig. 10. Frame error rate versus average SNR per Rx antenna for space-time coding and BICM with rate-1/2 convolutional codes and 2 Rx.

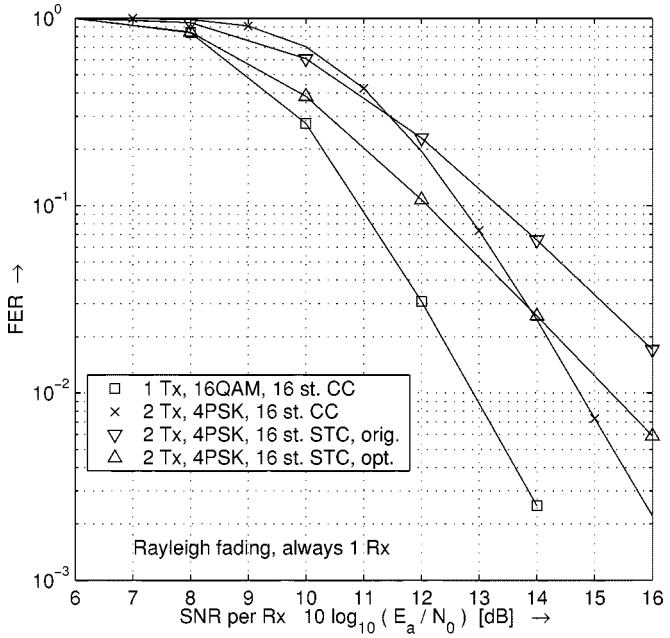

Fig. 11. Frame error rate versus average SNR per Rx antenna for space-time coding and BICM with rate-1/2 convolutional codes and 1 Rx.

needs to be paid in terms of overall link-level capacity, but this scheme might prove more robust when operated with nonperfect coding. Fig. 9 reveals again that $n_r > n_t$ also lowers the gap between BICM and CM for the case of mixed signal constellations.

In Section V, we want to check the above capacity comparisons against the "reality" of simulation results.

V. Simulation Results in Ideal Fast Fading

The previous capacity comparisons allowed interesting performance predictions, which we now confirm by simulation results, in which the BICM scheme is implemented with sufficient interleaving, and expression (16) is used as metric for Viterbi decoding. The matrices with channel coefficients are randomly generated for each time step to simulate the independent fast fading channel and are known perfectly in the receiver.

A. Spectral Efficiency of 2 Bit/s/Hz

Fig. 10 shows frame error rate (FER) performance versus average SNR per Rx antenna in various transmission systems with 2 Rx and with the identical target spectral efficiency of 2 bits per channel use. One frame is $L = 130$ hypersymbols long so that a total of 260 information and termination bits are transmitted in such a frame. In the codes with 16 states, four bits are used for trellis termination, leaving 256 information bits per frame. We first consider the two 16-state 4PSK STC (space-time codes). The fast-fading optimized STC [15], [16] (marker $\triangle$) provides a 1 to 1.8 dB improvement over the original STC [6] (marker ∇). The latter is not intended for the fast-fading channel. For FER below 10^{-3} (which might not be interesting for some applications), both STC are even outperformed by a single Tx 16QAM BICM scheme with rate-1/2 16 state CC with octal generators (23, 35) and minimum free distance $d_{\text{free}} = 7$ [25], which is equal to the maximum achievable diversity [22], but for larger tolerated FER, the optimized STC performs better than the single Tx system. At FER below 10^{-1}, the 2 Tx 4PSK BICM system with rate-1/2 16 state CC outperforms all other benchmark systems. To summarize these results, we note that the slopes of the FER of all BICM schemes are the same, whereas the slope of the STC systems is significantly less steep, indicating their failure to produce a comparatively high order of diversity like BICM. It is worth mentioning that the respective spacing of the FER curves of the bit-based systems agrees nicely with the AMI curves in Fig. 4. We actually see the 1-dB shift from 1 Tx 16QAM BICM to 2 Tx 4PSK BICM. Hence, the gains predicted by capacity considerations actually translate into an appropriately shifted frame error rate performance, justifying the use of the capacity measure to compare different BICM schemes with multiple antennas. At an FER of 10^{-3}, the simple 16-state CC operates almost constantly with a 6.2-dB gap to the respective capacity limit observable in Fig. 4.

To show that the use of $n_r < n_t$ leads to severe performance penalties, Fig. 11 summarizes the results obtained for one Rx antenna. From Fig. 4, a large capacity gap between BICM and CM follows, which leads to the superior performance of space-time codes over 2 Tx BICM at lower SNRs. At higher SNRs, the larger diversity order of the bit-based systems becomes dominant in the frame error rate. The most important point in this comparison is that actually *all* considered schemes with 2 Tx and 1 Rx are easily outperformed by the simple single transmitter single receiver 16QAM BICM system. Hence, if the 1 Rx case is intended to be the most prevalent mode of operation in fast fading channels, one should abstain from using 2 Tx 4PSK transmit diversity solutions and opt for 16QAM modulation with 1 Tx. We observe a 1.8-dB loss in the BICM simulation for 2 Tx 4PSK compared against 1 Tx 16QAM, which is in

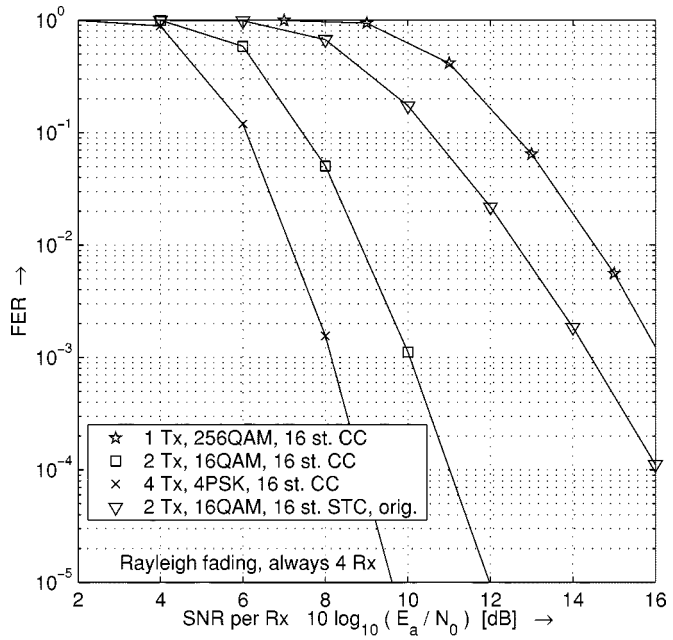

Fig. 12. Frame error rate versus average SNR per Rx antenna for space-time coding and BICM with rate-1/2 convolutional codes and with 4 Rx.

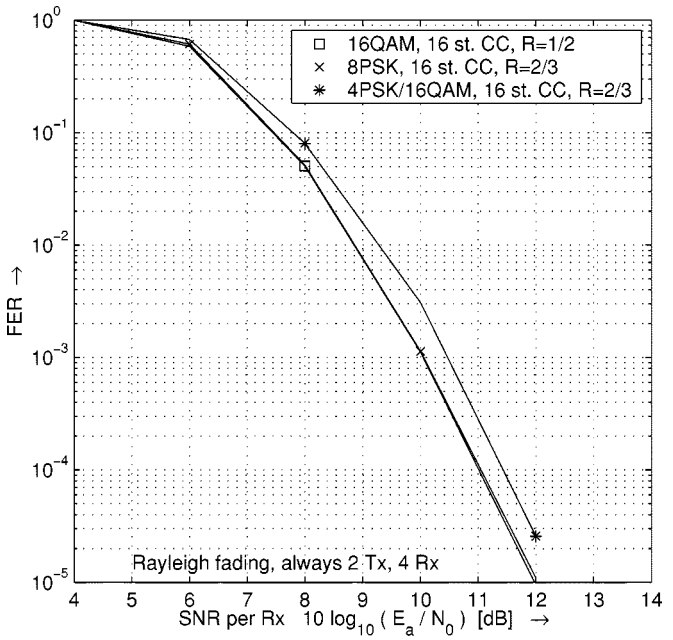

Fig. 13. Frame error rate versus average SNR per Rx antenna for BICM with rate-1/2 and rate-2/3 convolutional codes for 2 Tx (identical average power) and 4 Rx systems.

good correspondence with the capacity results in Fig. 4, where we have a gap of 1.5 dB at 2 Bit/s/Hz (1 Rx).

B. Spectral Efficiency of 4 Bit/s/Hz

We again consider frames of $L = 130$ hypersymbols so that now, a total of 520 information and termination bits are transmitted in such a frame. An impressive demonstration of the usefulness of transmit diversity in terms of link-level capacity has been given in Fig. 5, and we want to demonstrate that the promised gains can actually be achieved by BICM. In Fig. 12, we compare the FER versus average SNR per Rx antenna in various transmission systems with 4 Rx and identical target spectral efficiency of 4 bits per channel use. A 1 Tx 256QAM BICM scheme is compared with 2 Tx 16QAM BICM and 4 Tx 4PSK BICM, all with rate-1/2 16 state CC with octal generators (23, 35) and $d_{\text{free}} = 7$ [25]. We clearly see the gains predicted by theory in the respective shift of the frame error rates. The 16-state, 16QAM STC for 2 Tx from [6] operated with 4 Rx manages to outperform the 256QAM scheme from a single transmitter but exhibits a large performance gap to the bit-based coding schemes from 2 Tx. Clearly, this is due to the fact that this STC is not intended to be used in fast fading. We did not use the 16QAM STC like it is given in [7] with natural labeling of the signal points because this leads to an approximately 2-dB loss when compared with the 16QAM labeling, as given in [6]. The reason for the very poor performance of STCs with large signal sets in fast fading is that they require a large number of states, which in turn does *not* translate into a large Hamming distance in the time direction to exploit the time diversity. This clearly needs to be blamed on STCs being signal-space codes instead of binary codes.

Now, let us take another look at transmission with 4 bits per channel use. In Fig. 7, we compared the 2 Tx 8PSK and 16QAM constellations at this spectral efficiency and found a difference of only 0.55 dB. Further, from Fig. 9, we saw that the mixed constellation 4PSK/16QAM shows negligible difference from 8PSK/8PSK in terms of capacity. We want to confirm those similarities by the simulation results in Fig. 13, where we use a rate-1/2 CC with octal generators (23, 35) and $d_{\text{free}} = 7$ [25] for 16QAM/16QAM, and a real (nonpunctured) rate-2/3 CC with octal generators (27, 75, 72) and $d_{\text{free}} = 5$ [25] for 8PSK/8PSK and 4PSK/16QAM to obtain the spectral efficiency of 4 bits per channel use. Especially for the balanced signal constellations paired with different coding rates, the closeness of the error rate results is striking. The mixed-signal constellation performs slightly worse because the convolutional code has trouble averaging over three strongly different bitlevel capacity values I^{λ}. The balanced transmission schemes only have two different capacity values on the bitlevels so that the convolutional code does not need as much effort to average over them. Expanding the signal constellation in fading 1 Tx/1 Rx links [26] has large benefits that are no longer that visible in the multiple antenna scenario. Nonetheless, the good correspondence of the simulation and capacity results again justifies the validity of the previous capacity comparisons.

VI. Application in OFDM

We studied multiple antenna coding approaches in ideal fast fading, and now, we apply multiple antenna BICM to OFDM [9]–[13] to show that much of the earlier conclusions are applicable in this correlated fading case with interleaving.

A. OFDM Channel Model With Multiple Antennas

The transmission link from Tx antenna j to Rx antenna i is represented by a baud-spaced multipath channel, which is characterized by its finite discrete-time channel impulse response

$h_{ij}[\kappa]$, $0 \le \kappa < K$, $1 \le i \le n_r$, $1 \le j \le n_t$, i.e., all channels are of length K. The channel taps are zero-mean complex Gaussian random variables, and they are mutually uncorrelated in time and also across the antennas. We assume the same average power delay profile for all Tx–Rx links with average tap power $p[\kappa] = \mathcal{E}\left\{|h_{ij}[\kappa]|^2\right\}$, $\forall i, j$. As a simulation model, we use the exponential average power delay profile

$$p[\kappa] = \begin{cases} \frac{1-e^{-1/K_{\exp}}}{1-e^{-K/K_{\exp}}} e^{-\kappa/K_{\exp}}, & 0 \le \kappa < K \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

where $K_{\exp}$ is a parameter that characterizes the exponential decay of the average echo power over the channel impulse response length K. From the special scaling, it follows directly that the average sum power is normalized to $\sum_{\kappa=0}^{K-1} p[\kappa] = 1$. The parameter $K_{\exp}$ is loosely related to the rms delay spread.

For OFDM transmission, it is natural to consider coding across subcarriers because of delay constraints for data transmission. OFDM has a natural blocking of data so that delay of one OFDM symbol is always present in the link. Channel coding over a large number of OFDM symbols would result in mostly unacceptable overall delays. Further, in a sufficiently scattered multipath environment with low mobility, the channel fluctuation in time is usually smaller than the change of channel conditions in frequency so that a higher degree of diversity results by coding across subcarriers. Hence, the discrete-time variable ℓ from Section II is now associated with the frequency, and via the L-point discrete Fourier transform (DFT), we obtain

$$H_{ij}[\ell] = \sum_{\kappa=0}^{K-1} h_{ij}[\kappa] e^{-j2\pi\ell\kappa/L}. \tag{18}$$

We choose the length of the transmitted signal block L to be a power of 2 in order to implement the DFT with a fast algorithm. The $n_r \times n_t$ channel matrix across the frequency axis ℓ is given by $\boldsymbol{H}[\ell] = [H_{ij}[\ell]]$. With $K < L$ ($K \ll L$), adjacent channel matrices are (strongly) correlated and frequency-domain interleaving of the transmitted hypersymbols $\boldsymbol{a}[\ell]$ is beneficial for STC and BICM. There are still correlations after interleaving, but the resulting channel experienced by STC and BICM with limited code memory is well approximated by fast fading.

B. Simulation Results in OFDM for 2 Bit/s/Hz

In [27], the use of STC in OFDM is proposed. To have a close relationship to the results in Section VI-A, where we investigated block lengths of 130, we choose the DFT size $L = 128$. Further, we use a channel with decay parameter $K_{\exp} = 4.0$ and length $K = 8$, which is fairly short when compared with L and should lead to a conservative estimate, as to whether or not the diversity advantages of bit-based channel coding schemes over STC, which we observed in ideal fast fading, still hold for the OFDM channel properties. A frequency block interleaver of depth 11 for complex symbols is used for all coding approaches to transform the correlated frequency selectivity of the OFDM system into a virtually fast(er) fading characteristic. Fig. 14 shows the frame error rate performance of the optimized 4PSK STC [15], [16] and rate-1/2 4PSK BICM. Results with and without interleaving are shown for both schemes to demonstrate the necessity of appropriate frequency-domain interleaving to destroy the fading correlations between adjacent subcarriers.

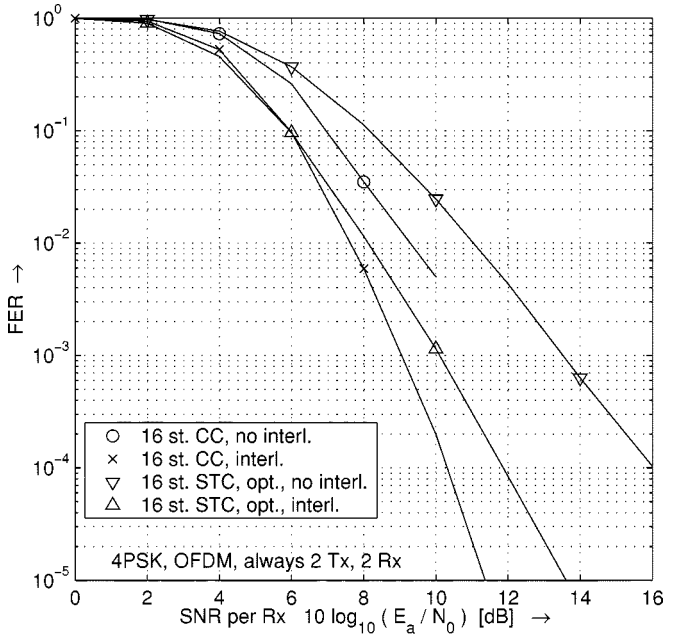

Fig. 14. Frame error rate versus average SNR per Rx antenna for space-time coding and BICM coding schemes for 2 Tx and 2 Rx in an OFDM system with $L = 128$ subcarriers and a multipath channel with $K = 8$ and $K_{\exp} = 4.0$. Results with and without frequency interleaving are shown.

Comparing Figs. 10 and 14, we see that the performance of both schemes with interleaving in OFDM is slightly worse than in the ideally fast fading channel, but the performance advantage of bit-based channel coding is, even though reduced, still remarkable. For signal constellations larger than 4PSK, the performance advantage of bit-interleaving is larger, as observed in Section V-B in Fig. 12, and this is expected to be also true in OFDM.

VII. Conclusions

Binary convolutional coding with appropriate bit-interleaving is a widely accepted way to do channel coding with higher order modulation in fading channels with a single transmit antenna. The results in this paper show that bit-based coding architectures also lead to flexible coding schemes for the multiple Tx antenna case while enabling reasonable performance in fast fading. When a convolutional code exhibits a given distance from the single-Tx capacity limit, it approximately retains this distance from the respective multiple-Tx capacity limit, indicating the usefulness of the bit-based schemes. A very interesting possibility lies in the use of higher rate coding with smaller signal sets, paving the way to extremely interesting combinations like the 2 Tx 8PSK scheme with $R = 2/3$ for a spectral efficiency of 4 bit/s/Hz shown in Fig. 13. The large order of diversity in multiple antenna systems encourages the use of higher rate codes, which keeps the modulation schemes small. This is interesting in practical systems, when nonlinear transceiver effects and channel estimation errors limit constellation sizes.

REFERENCES

[1] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. Int. Symp. Signals, Syst., Electron.*, 1998, pp. 295–300.

[2] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, pp. 311–335, 1998.

[3] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," in *Proc. Allerton Conf. Commun., Contr., Comput.*, 2001.

[4] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 6, pp. 41–59, 1996.

[5] E. Biglieri, G. Taricco, and E. Viterbo, "Bit-Interleaved time-space codes for fading channels," in *Proc. Conf. Inform. Sci. Syst.*, Princeton, NJ, 2000, pp. WA4.1–WA4.6.

[6] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-Time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Apr. 1998.

[7] A. F. Naguib, V. Tarokh, N. Seshadri, and A. R. Calderbank, "A space-time coding modem for high-data-rate wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1459–1478, Aug. 1998.

[8] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Space-Time codes for high data rate wireless communication: Performance criterion in the presence of channel estimation errors, mobility, and multiple paths," *IEEE Trans. Commun.*, vol. 47, pp. 199–207, Feb. 1999.

[9] S. B. Weinstein and P. M. Ebert, "Data transmission by frequency-division multiplexing using the discrete fourier transform," *IEEE Trans. Commun.*, vol. COM-19, pp. 628–634, 1971.

[10] L. J. Cimini Jr., "Analysis and simulation of a digital mobile channel using orthogonal frequency division mutiplexing," *IEEE Trans. Commun.*, vol. COM-33, pp. 665–675, 1985.

[11] M. Alard and R. Lassalle, "Principle of modulation and channel coding for digital broadcasting for mobile receivers," *EBU Rev.—Tech.*, vol. 224, pp. 168–190, 1987.

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

[13] H. Sari, G. Karam, and I. Jeanclaude, "Transmission techniques for digital terrestrial TV broadcasting," *IEEE Commun. Mag.*, vol. 33, no. 2, pp. 100–109, 1995.

[14] D. L. Goeckel, "Coded modulation with nonstandard signal sets for wireless OFDM systems," in *Proc. IEEE Int. Conf. Commun.*, 1999, pp. 791–795.

[15] R. S. Blum, "New analytical tools for designing space-time convolutional codes," in *Proc. Conf. Inform. Sci. Syst.*, Princeton, NJ, 2000, pp. WP3.1–WP3.6.

[16] R. S. Blum, Y. (Geoffrey) Li, J. H. Winters, and Q. Yan, "Improved space-time coding for MIMO-OFDM wireless communications," *IEEE Trans. Commun.*, 2001, to be published.

[17] W. Firmanto, B. S. Vucetic, and J. Yuan, "Space-Time TCM with improved performance on fast fading channels," *IEEE Commun. Lett.*, vol. 5, pp. 154–156, Apr. 2001.

[18] S. A. Zummo and S. A. Al-Semari, "Space-Time coded QPSK for rapid fading channels," in *Proc. Int. Symp. Pers., Indoor Mobile Radio Commun.*, London, U.K., 2000, pp. 504–508.

[19] ——, "A decoding algorithm for I–Q space-time coded systems in fading environments," in *Proc. IEEE Veh. Technol. Conf.*, Boston, MA, 2000, pp. 331–335.

[20] A. Chindapol and J. A. Ritcey, "Bit-Interleaved coded modulation with signal space diversity in Rayleigh fading," in *Conf. Rec. 33rd Asilomar Conf. Signals, Syst., Comput.*, 1999, pp. 1003–1007.

[21] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, pp. 873–884, May 1992.

[22] G. Caire, G. Taricco, and E. Biglieri, "Bit-Interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 44, pp. 927–946, June 1998.

[23] ——, "Recent results on coding for multiple-antenna transmission systems," in *Proc. 6th Int. Symp. Spread-Spectrum Tech. Appl.*, 2000, pp. 117–121.

[24] R. S. Blum and J. H. Winters, "On the capacity of cellular systems with MIMO," presented at the Int. Contr. Conf., 2002.

[25] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1995.

[26] U. Hansson and T. Aulin, "Channel symbol expansion diversity," *Electron. Lett.*, vol. 31, no. 18, pp. 1545–1546, 1995.

[27] D. Agrawal, V. Tarokh, A. Naguib, and N. Seshadri, "Space-Time coded OFDM for high data-rate wireless communications over wideband channels," in *Proc. IEEE Veh. Technol. Conf.*, Ottawa, ON, Canada, 1998, pp. 2232–2236.

Stefan H. Müller-Weinfurtner (S'96–M'01) received the Dipl.-Ing. degree in 1996 from University of Erlangen-Nürnberg, Germany, with a thesis on soft-output equalization algorithms for GSM, incorporating soft-decision feedback. In 2000, he authored the thesis "Orthogonal Frequency-Division Multiplexing (OFDM) for Wireless Communications," covering peak-power reduction techniques, receiver windowing, and signal structures and algorithms for frame and frequency synchronization, for which he received the Dr.-Ing. degree from the same university.

From 2000 to 2002, he was with the Wireless Systems Research Department at AT&T Labs—Research, Middletown, NJ, studying multiple antenna channel coding for wideband OFDM as well as equalization in IEEE 802.11b wireless LAN receivers. Since June 2002, he has been with Philips Semiconductors TCMC, Nürnberg, working on third-generation wireless system implementation. His research interests are in the design of wireless digital communications systems, OFDM, synchronization, peak-power reduction, adaptive equalization, and multiple antenna technologies.