

Turbo modulation and coding: Design and evaluation of iterative bit-interleaved coded modulation methods for wireless systems beyond 3G[†]

Gerhard Bauch^{1*}, Frank Schreckenbach², Christoph Hausl² and Tetsushi Abe¹

¹DoCoMo Euro-Labs, Landsberger Strasse 312, D-80687 Munich, Germany

²Institute for Communications Engineering (LNT), Munich University of Technology (TUM), D-80290 Munich, Germany

SUMMARY

We give an overview of iterative techniques for bit-interleaved coded modulation. We start with simple methods where a convolutional code is concatenated with a standard QAM modulation scheme. The mapping is carefully chosen different from Gray mapping such that significant gains can be obtained by an iterative (turbo) receiver. The idea can also be applied to transmit antenna diversity as well as spatial multiplexing (MIMO). We then consider various powerful iterative coding schemes which are under discussion in standardisation for wireless systems beyond 3G such as serial or parallel concatenated convolutional codes, repeat accumulate codes and LDPC codes. We give a comparison in terms of complexity and performance. The main intention is to provide an illustrative overview over relevant applications of the turbo principle in demodulation and decoding. It is a summary of results which have been obtained in collaboration between the Institute of Communications Engineering, TUM and DoCoMo Euro-Labs. Copyright © 2007 John Wiley & Sons, Ltd.

1. INTRODUCTION

In upcoming mobile communication systems, high data rates need to be transmitted with high reliability. This requires an efficient combination of high order modulation schemes with powerful channel coding techniques. In addition, the system needs to be highly adaptive in the sense that it is easy to adapt the data rate to the channel characteristics and the computational complexity to the available resources.

Under that perspective, bit-interleaved coded modulation (BICM) is a well suited bandwidth efficient scheme. BICM was proposed in Reference [1] and a thorough investigation was done in Reference [2]. BICM is the serial concatenation of an encoder, interleaver and mapper, as shown in Figure 1. The coded and interleaved bits $c_{n,m}$, $m = 1, \dots, M$ are mapped at time instant n to a signal point x_n from a 2^M -ary signal constellation (e.g. QPSK, 16QAM)

according to the chosen bit-to-symbol mapping (e.g. Gray, Set Partitioning [3]). The performance of BICM can be greatly improved through iterative information exchange according to the turbo principle [4] between the demapper of a high order modulation scheme and the decoder. This system, introduced in References [5, 6], is usually referred to as BICM with iterative decoding (BICM-ID). It was soon recognized that the bit-to-symbol mapping has to be well chosen to achieve significant gains over the demapping and decoding iterations.

Another approach is to use powerful channel codes that circumvent the need of iterative demapping and decoding. Candidates for these strong codes include parallel concatenated convolutional codes ('turbo' codes) [7], serial concatenated convolutional codes [8] and various versions of low density parity check (LDPC) codes [9].

A powerful tool to analyse these different approaches is the EXIT chart [10]. The EXIT chart visualises the density

* Correspondence to: Gerhard Bauch, DoCoMo Euro-Labs, Landsberger Strasse 312, D-80687 Munich, Germany. E-mail: bauch@docomolab-euro.com

[†] Dedicated to Prof. Dr.-Ing. Dr.-Ing. E. h. Joachim Hagenauer on the Occasion of his 65th Birthday.

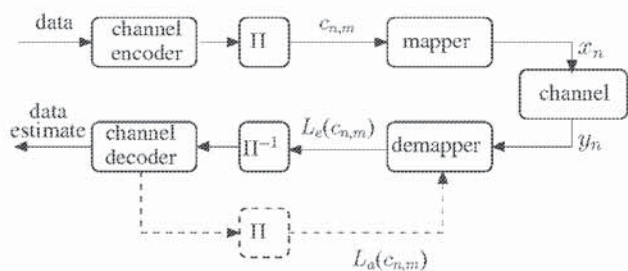


Figure 1. Bit-interleaved coded modulation with optional iterative demapping and decoding: System model.

evolution of the extrinsic information over the iterations. EXIT functions describe the input/output behaviour of the receiver components using the mutual information between the coded bits at the transmitter and the log-likelihood ratios at the receiver as single characteristic variable.

This paper is organised as follows: In Section 2, we describe variants of iterative BICM schemes, where extrinsic information is exchanged between the QAM demapper and the decoder of a simple convolutional code. In those schemes, the choice of the mapping of bits to QAM symbols is essential. Then, in Section 3, we describe and compare BICM schemes with powerful codes such as parallel and serial concatenated convolutional codes or LDPC codes instead of simple convolutional codes.

2. BICM WITH SIMPLE CODES

With a simple convolutional code in a bit-interleaved coded modulation scheme, special mappings of bits to QAM constellation points different from Gray mapping have to be

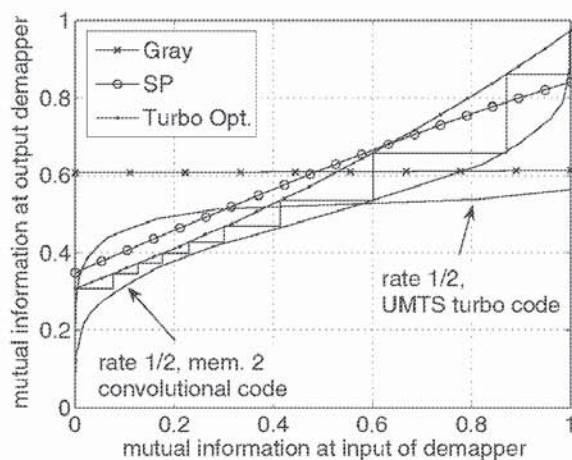


Figure 2. EXIT chart for iterative decoding and demapping with 16-QAM and Gray, Set Partitioning (SP) [3] and turbo optimised M16a mapping from [11]; $10 \log_{10}(E_s/N_0) = 7\text{dB}$.

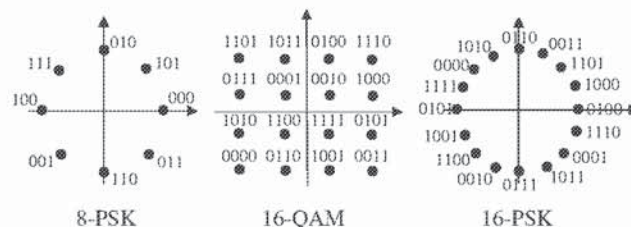


Figure 3. Optimised mappings for iterative demapping.

used to allow substantial turbo gains. This can be explained using EXIT charts (see Figure 2). With Gray mapping, the EXIT curve of the demapper is a horizontal line which indicates that no gains can be obtained by turbo iterations. The trajectory will get stuck after the first iteration. With mappings which are optimised for turbo processing, we obtain a non-horizontal curve which allows turbo gains. Optimisation of the mappings is a non-trivial problem. Example for 16-QAM modulation, there exist $16! \approx 2 \cdot 10^{13}$ possible mappings. For optimisation, we can use the EXIT chart. Our design criterion is to maximise the y-coordinate of the end point of the EXIT curve. This yields the lowest possible error floor for a given signal constellation.

In order to find a mapping which maximises the end point of the EXIT curve, we apply a method which is known from assignment of bits in vector quantisation: the binary switching algorithm [12]. For details, we refer to [11, 13]. Examples for optimised mappings are given for 8-PSK, 16-PSK and 16-QAM in Figure 3. The respective BER is depicted in Figure 4 for 16-QAM with memory-2 convolutional codes and the UMTS turbo code.

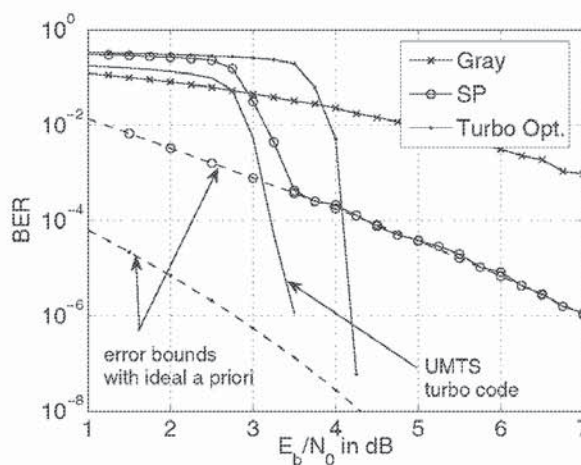


Figure 4. BER for iterative demapping with 16-QAM and Gray, Set Partitioning (SP) [3] and turbo optimised M16a mapping from [11]. (AWGN channel, rate-1/2 codes, information frame length 5000).

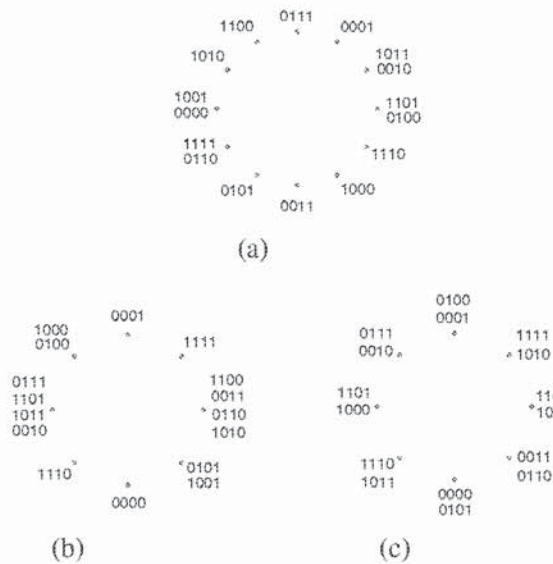


Figure 5. Optimised non-unique PSK mappings. (a) 12PSK, (b) 8PSK (1), (c) 8PSK (2).

We included also the analytical bounds for the error floor when perfect extrinsic information is available. The error floor is unavoidable if no additional recursive inner code is used.

Playing with mappings and the signal constellation offers another possibility to reduce the error floor. We can e.g. use non-unique mappings [14, 15]: we reduce the number of constellation points without decreasing the number of bits per constellation point. This results in non-unique mappings, where multiple bit sequences are mapped to the same constellation point. Examples are given in Figure 5. The multiplicities can be resolved in the iterative process provided that sequences which are mapped to the same constellation point differ in at least two positions. Mappings which differ in only one bit cannot be distinguished based on fed back extrinsic information. This is best explained by assuming perfect extrinsic information in which case all the bits are known except for the bit on which the current decision has to be made.

Non-unique mappings can provide a lower error floor than is achievable with unique mappings since with perfect extrinsic information, the Euclidean distance between two relevant constellation points can be increased compared to a constellation with unique mappings and consequently more constellation points. However, it turns out that the gain in the error floor region comes on the expense of convergence at a significantly higher SNR. This is explained as follows: Signal points with a larger number of associated bit sequences are more probable than signal points with a

lower number of associated bit sequences. The specific bit-to-symbol mapping does not change the capacity. However, the probabilities of the signal points do affect the capacity. Non-uniform mappings designed to lower the error floor usually reduce the capacity.

Non-unique mappings can be designed to provide a capacity gain: By assigning more bit sequences to signal points with low energy and less bit sequences to signal points with high energy, signal points with less energy are used more frequently than points with high energy. Such approaches are usually referred to as signal shaping. The non-uniform signaling will reduce the average number of bits per symbol and thus the bit rate. However, the energy savings may more than compensate for this loss of bit rate. The system is optimized by carefully choosing the probability distribution of the signal points and the specific bit-to-symbol mapping and by using EXIT charts to adapt the decoder to the demapper EXIT curve. Then, for 16ASK modulation as example, we achieve a BER of 10^{-5} at the capacity of the uniform distribution and an iterative decoding threshold 0.7 dB below the capacity of the uniform distribution [16].

An interesting observation is that the gap between the EXIT curve of the demapper and the curve of the decoder corresponds to the channel capacity normalised to one bit per channel use. Consequently, we wish to minimise this gap in order to reduce the SNR loss compared to the theoretical limit. Using irregular mappings provides a simple strategy in order to fit the curve of the demapper to the curve of the convolutional decoder [17]. Irregular mapping means the use of different modulation formats and/or different mappings within one coded block. Given knowledge on the average channel quality, that is the SNR, we can determine for which fraction of the block we should apply a certain modulation format and mapping. The actual position of the respective symbols in the block does not matter in principle, only the fraction is important. Such an irregular mapping allows to adapt the modulation scheme very accurately and with fine grid to the current channel state.

The iterative demapping schemes which have been discussed so far should be combined with relatively simple outer FEC codes. In the examples above, we used convolutional codes with memory 2. The EXIT curve of those codes has a steepness which fits well to the demapper curve, see Figure 2. The complexity of the detection scheme is mainly determined by the iterations rather than the components within one iteration. Now, the question arises, if it makes sense to combine powerful FEC codes such as turbo codes with iterative demapping. More precisely, should we do iterations within the turbo decoder and then feed

back extrinsic information to the demapper for a second, outer turbo loop. The EXIT curve of a turbo code is close to the curve of an optimum code: It is very steep in the beginning and at the end, but almost horizontal in between. Consequently, the decoder curve crosses the steep curve of a demapper with the optimised mappings which have been proposed above. Hence, turbo iterations will not be effective. Moreover, the curve with Gray mapping fits best with an outer turbo code. It can be concluded that turbo optimised mappings should be used with weak outer codes. However, when powerful codes such as turbo codes are applied, we should use Gray mapping and perform no iterations between FEC decoder and demapper.

The idea of iterative demapping with optimised mappings can be extended to multiple-input multiple-output (MIMO) transmission schemes. A simple MIMO technique is space-time block coding. A space-time block code together with linear combining at the receiver provides a diversity gain by transforming a MIMO channel into an effective equivalent single-input single-output (SISO) channel with reduced variance of the SNR compared to a single antenna system. Consequently, the mappings which have been presented above can be directly applied to be used with space-time block codes [18]. On the other hand, in spatial multiplexing, we transmit independent data simultaneously from multiple antennas in order to increase the data rate rather than provide diversity. Using the same modulation format and mapping at all transmit antennas already allows to improve the achievable turbo gains. However, space offers another dimension for optimisation. The mapping to a hypersymbol consisting of the symbols which are transmitted simultaneously from different antennas should be jointly optimised [19, 20]. Note that the symbols belonging to one hypersymbol may alternatively be transmitted sequentially over a single antenna [21].

3. BICM WITH POWERFUL CODES

Parallel concatenated convolutional codes known as turbo codes [7] allow transmission close to the Shannon limit with manageable complexity. Ten years after their invention, they have found their way into standards such as UMTS. Since turbo codes perform close to the theoretical limit, no significant improvements can be expected by using other coding schemes. In fact, the performance of powerful codes which are currently discussed for standardisation of wireless systems differ only by fractions of a dB. Today, the points of interest are practically relevant issues such as complexity, ease of parallel implementation and

flexibility. Future systems will be highly adaptive. In order to provide reliable transmission and high throughput, code rate and modulation will be regularly adapted according to the current channel state and QoS requirements. In this sense, flexibility means a coding scheme, the rate of which can easily be changed. We prefer to use rate-compatible punctured codes [22], where different code rates are obtained by puncturing code bits of a low-rate mother code. This is also advantageous for hybrid automatic repeat request (HARQ) schemes, since punctured bits can be transmitted as incremental redundancy if an error was detected at the receiver.

Another problem which becomes severe for the envisaged high data throughput of 100 Mbit/s–1 Gbit/s is hardware limitation for the decoder throughput. Throughput refers to the number of information bits which the decoder can process per second. In order to provide the required throughput, we need decoder structures which allow parallel implementation. This is not straight forward with currently used turbo codes. The main problems are the decoders for the constituent codes which have to be decoded sequentially and require a forward and a backward recursion. Moreover, the interleaver which is currently used in UMTS limits the degree of parallel implementation due to memory access collisions.

Candidate coding schemes that meet those requirements are besides parallel concatenated convolutional codes (PCCC, turbo codes) serial concatenated convolutional codes (SCCC) [8] and various versions of low density parity check (LDPC) codes [9]. Serial concatenated convolutional codes are interesting since they can achieve lower error floor than parallel concatenated convolutional codes. Furthermore, it has been claimed in Reference [23] that serial concatenated codes can be implemented with significantly lower complexity than PCCC or LDPC codes. The main argument for LDPC codes, which are currently very fashionable, is that parallel implementation is straight forward.

All candidate coding schemes which are considered rely on iterative decoding. Since turbo codes are already part of some standards, we use them as reference scheme.

First, we consider SCCC. The proposal in Reference [23] is intended for satellite applications where a very low frame error rate of 10^{-7} and hence a low error floor is required. Since PCCC tend to have a higher error floor than SCCC, the comparison in Reference [23] is between PCCC with memory 4 constituent codes and SCCC with memory 2 constituent codes. Since complexity grows exponential with the memory of the constituent codes, the comparison clearly yields a complexity advantage for SCCC. However,

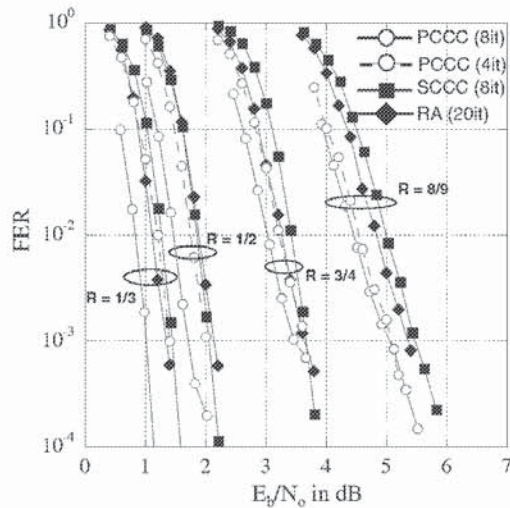


Figure 6. FER of PCCC, SCCC and RA codes. AWGN channel, QPSK, 1000 bits per information frame.

for wireless communications, we can allow a higher frame error rate of 10^{-2} , because a fast feedback channel allows to request an HARQ retransmission in case that the first transmission was not decoded successfully. We compare the SCCC scheme proposed in Reference [23] to PCCC with memory 3 constituent codes which is used in UMTS. Performance results are given in Figure 6. It can be observed that the difference in frame error rate (FER) is within 0.5 dB. However, the complexity per iteration is higher for the considered PCCC since the memory of the constituent codes is higher. Moreover, we can observe from Figure 6 that PCCC achieves comparable or even better performance as SCCC with a lower number of iterations (4 instead of 8 iterations). Consequently, at comparable performance, PCCC shows even lower complexity than SCCC. For details, we refer to [24]. Rate-compatible puncturing for SCCC is more complicated than for PCCC, because the more important information bits, which should not be punctured, are interleaved and can only be localised by taking into account the interleaver. In Reference [23], a long puncturing pattern which repeats every 300 code bits is proposed. As a conclusion, we prefer PCCC over SCCC for application in B3G systems.

Another possible class of coding schemes are LDPC codes. LDPC codes are often represented by a Tanner graph. Each of the so-called variable nodes on the left hand side corresponds to a received code bit. Several variable nodes are connected *via* the interleaver network with check nodes. The modulo 2 sum of the code bits which are connected to a certain check node is zero. LDPC codes can be decoded by a message passing algorithm, where extrinsic information

is sent forth and back between the variable nodes and the check nodes. In principle, all decoder operations at the check nodes and all decoder operations at the variable nodes, respectively, can be performed simultaneously. Hence, LDPC codes enable implementation with a high degree of parallelism. In practice, the achievable degree of parallelisation is limited by the chip area.

Disadvantages which are often associated with LDPC codes are high encoding complexity and low flexibility. For non-structured LDPC codes, the encoding complexity grows quadratically with the block length. As far as flexibility is concerned, different codes have to be designed for different block length. Moreover, variable code rates are often achieved by using different codes rather than puncturing. This does not allow to use HARQ with incremental redundancy.

We addressed these problems and considered a certain class of LDPC codes, so called repeat accumulate (RA) codes [25, 26]. Here the encoding complexity grows linearly with the block size. We designed the unpunctured mother code with code rate $R = 1/3$ according to [27] with EXIT charts. We chose the degree of the check nodes to be four. The structure of the edges between variable and check nodes are determined with the progressive edge growth algorithm [28]. In order to obtain higher code rates, we punctured the redundancy bits regularly.

Performance results are included in Figure 6. It can be seen that the degradation compared to PCCC is 0.2–0.5 dB. More important, the complexity in terms of number of operations is about 244% of that required for the UMTS PCCC. The high complexity is due to the fact that a large number of iterations is necessary for decoding of RA codes even though the complexity of each individual iteration is relatively small. Using scheduling algorithms such as shuffling algorithm proposed in Reference [29] may reduce the complexity gap to PCCC.

4. CONCLUSIONS

We have presented an overview over bit-interleaved coded modulation methods with iterative decoding. We first explained schemes with simple convolutional codes where the iterations are done between the decoder and the QAM demapper. Those schemes provide turbo gains only for higher order modulation and require special mappings different from Gray mapping. The convolutional code should have low memory. Substantial gains can be achieved using this simple method but an error floor is unavoidable. On the other hand, we discussed powerful outer codes which

should be combined with Gray mapping and no iterations to the demapper. For application in wireless communications, parallel concatenated convolutional codes turn out to be preferable over serial concatenated convolutional codes and LDPC codes both in terms of performance and complexity.

REFERENCES

1. Zehavi E. 8-PSK trellis codes for a Rayleigh channel. *IEEE Transactions on Communications* 1992; **40**(5):873–884.
2. Caire G, Taricco G, Biglieri E. Bit-interleaved coded modulation. *IEEE Transactions on Information Theory* 1998; **44**(3):927–946.
3. Ungerböck G. Channel coding with multilevel/phase signals. *IEEE Transactions on Information Theory* 1982; **28**(1):55–67.
4. Hagenauer J. The turbo principle: Tutorial introduction and state of the art. In *International Symposium on Turbo-Codes*, Brest, France, 1997.
5. Li X, Ritecy J. Bit-interleaved coded modulation with iterative decoding using soft feedback. *Electronic Letters* 1998; **34**(10):942–943.
6. ten Brink S, Speidel J, Yan R. Iterative demapping and decoding for multilevel modulation. In *Proceedings of IEEE Globecom Conference*, 1998; Sydney, 579–584.
7. Berrou C, Glavieux A, Thitimajshima P. Near shannon limit error-correcting coding and decoding: Turbo-codes. In *IEEE International Conference on Communications (ICC)*, 1993; Geneva, Switzerland, 1064–1070.
8. Benedetto S, Divsalar D, Montorsi G, Pollara F. Serial concatenation of interleaved codes: Performance analysis, design, and iterative decoding. *IEEE Transactions on Information Theory* 1998; **44**(3):909–926.
9. Gallager RG. Low-density parity-check codes. *IRE Transactions on Information Theory* 1962; **8**:21–28.
10. ten Brink S. Convergence behavior of iteratively decoded parallel concatenated codes. *IEEE Transactions on Communications* 2001; **49**(10):1727–1737.
11. Schreckenbach F, Görtz N., Hagenauer J, Bauch G. Optimization of symbol mappings for bit-interleaved coded modulation with iterative decoding. *IEEE Communications Letters* 2003; **7**(12):593–595.
12. Zeger K, Gersho A. Pseudo-Gray coding. *IEEE Transactions on Communications* 1990; **38**:2147–2158.
13. Schreckenbach F, Görtz N, Hagenauer J, Bauch G. Optimized symbol mappings for bit-interleaved coded modulation with iterative decoding. In *Proceedings of IEEE Globecom Conference*, San Francisco, 2003; 3316–3320.
14. Clevorn T, Vary P. Iterative decoding of BICM with non-regular signal constellation sets. In *5th International ITG-Conference on Source and Channel Coding (SCC)*, Erlangen, Germany, 2004.
15. Bauch G, Sethuraman P, Schreckenbach F. Partially unique mappings for bit-interleaved coded modulation with iterative detection. In *Proceedings of IEEE Globecom Conference*, St. Louis, USA, 2005.
16. Schreckenbach F, Henkel P. Signal shaping using non-unique symbol mappings. In *43rd Annual Allerton Conference on Communication, Control, and Computing*, Monticello, USA, 2005.
17. Schreckenbach F, Bauch G. Bit-interleaved coded irregular modulation. *European Transactions on Telecommunications (ETT)* 2006; **17**(2):269–282.
18. Bauch G, Schreckenbach F. How to obtain turbo gains in coherent and non-coherent orthogonal transmit diversity. In *IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, Beijing, China, 2003.
19. Båro S. Turbo detection for MIMO systems: Bit labeling and pre-coding. *European Transactions on Telecommunications* 2004; **15**:1–8.
20. Muhammad N, Speidel J. Design of multidimensional mappings for iterative MIMO detection with minimized bit error floor. In *6th International Symposium on Turbo Codes*, Munich, Germany, 2006.
21. Simoons F, Wymeersch H, Moeneclaey M. Multi-dimensional mapping for bit-interleaved coded modulation with BPSK/QPSK signaling. *IEEE Communications Letters* 2005; **9**(5):453–455.
22. Hagenauer J. Rate-Compatible Punctured Convolutional Codes (RCPC Codes) and their Applications. *IEEE Transactions on Communications* 1988; **36**(4):389–400.
23. Benedetto S, Garelo R, Montorsi G, et al. MHOMS: High-Speed ACM Modem for Satellite Applications. *IEEE Wireless Communications Magazine* 2005; **12**(2):66–77.
24. Abe T, Bauch G, Hausl C. A comparison of rate compatible pccc and secc for next generation wireless communication systems. In *Proceedings of 3rd International Symposium on Wireless Communication Systems (ISWCS)*, Valencia, Spain, 2006.
25. Divsalar D, Jin H, McEliece RJ. Coding theorems for 'turbo-like' codes. In *Proceedings of 36th Annual Allerton Conference on Communications*, 201–210, Monticello, Urbana-Champaign, 1998.
26. Jin H, Kandeekar A, McEliece R. Irregular repeat-accumulate codes. In *Proceedings of 2nd International Symposium on Turbo codes and Related Topics*, 1–8, Brest, France, 2000.
27. ten Brink S, Kramer G. Design of repeat-accumulate codes for iterative detection and decoding. *IEEE Transactions on Signal Processing* 2003; **51**(11):2764–2772.
28. Hu XY, Eleftheriou E, Arnold DM. Progressive edge-growth tanner graphs. In *Proceedings of IEEE Globecom Conference*, 995–1001, San Antonio, 2001.
29. Zhang J, Fossorier MPC. Shuffled iterative decoding. *IEEE Transactions on Communications* 2005; **53**(2):209–213.