Information Processing in Soft–Output Decoding

Simon Huettinger[1], Johannes Huber[1], Rolf Johannesson[2], and Robert Fischer[1]

[1]Chair of Information Transmission
University Erlangen-Nürnberg
Cauerstr. 7/NT
D-91058 Erlangen, Germany
{huettinger,huber,fischer}@LNT.de

[2]Department of Information Technology
Lund University
Box 118, S-221 00 Lund
Sweden
rolf@it.lth.se

Abstract

Recent literature presents methods for the analysis of concatenated coding schemes by solely characterizing the behavior of the component codes [4, 12, 14, 9, 7]. Component codes are analyzed either analytically using unique properties of special component codes, e.g., single–parity–check code or accumulator, or via simulations.

The goals of this paper are to find fundamental insights into concatenated codes by analyzing the input–output relation of their components from an information–theoretic point of view. We derive a Information Processing Characteristic (IPC), which completely characterizes the behavior of a coding scheme comprising encoder, code and decoder for the general class of linear codes. For time invariant convolutional codes it is studied how the IPC can be obtained in practice.

1 Introduction

Since the invention of turbo–codes in 1993 [2] unveiling the iterative decoding algorithm has been a vivid field of research. However, it is still unclear what properties of a component code are necessary for a power–efficient concatenated coding scheme. For example, applying systematic constituent encoders as in a classical turbo–coding scheme [2], less iterations are necessary for convergence (or final failure), while in some cases nonsystematic constituent encoders offer a higher power efficiency at the expense of a largely increased number of iterations [11].

In the present paper, we try to analyze the coding schemes from an information–theoretic point of view. Especially, we study the suboptimality of codes and decoders used as components in concatenated coding schemes. Therefore we separate between the suboptimality introduced by codes of finite (constraint) length and the information loss due to symbol–by–symbol decoding.

2 Information Processing Characteristic of a Coding Scheme

We investigate the *coded* transmission of a block of K binary information symbols U_j, $j \in \{1, 2, \cdots, K\}$, which are independent identically distributed with $\Pr(U_j = 1) =$

$\Pr(U_j = 0) = 0.5$, hence, the entropy $H(U) = 1$. The *encoder* maps the information vector $\vec{U}$ to a codeword $\vec{X}$ which consists of N symbols X_n, $n \in \{1, 2, \cdots, N\}$. The rate of the code is $R = K/N$ measured in bit per channel symbol. The codeword $\vec{X}$ is transmitted over a memoryless channel that corrupts the message by *substitution* errors, e.g., the binary symmetric channel (BSC) or the additive white Gaussian noise channel (AWGN Channel). Insertions or deletions of the binary symbols X_n are not allowed.

The corrupted received sequence $\vec{Y}$ is processed by the decoder. The decoder output is the vector of soft–output values $\vec{V}$. Usually $\vec{V}$ consists of real–valued elements V_i, $i \in \{1, 2, \cdots, S\}$, where we do not require $S = K$. The soft–output $\vec{V}$ can be used to calculate the hard–output, where each symbol estimation $\hat{U}_j$ corresponds to a single input symbol U_j.

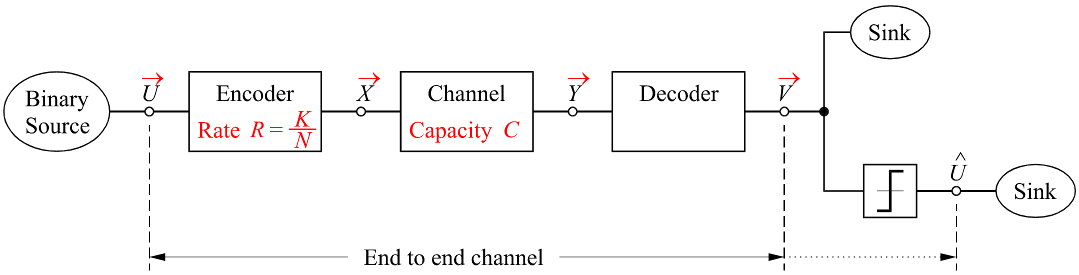

Figure 1: System model.

With respect to hard–output decoding a coding scheme is usually described by its average bit error ratio versus channel capacity or equivalent measures. The average bit error ratio is given by

$$\mathrm{BER} = \frac{1}{K} \sum_{j=1}^{K} \mathrm{BER}_j = E_j\{\mathrm{BER}_j\} \tag{1}$$

with $\mathrm{BER}_j = \Pr(\hat{U}_j \neq U_j)$.

In order to characterize the performance of a coding scheme w.r.t. soft–output decoding in a compact form we introduce a Information Processing Characteristic (IPC) of a coding scheme. It is the characterization of coding schemes (i.e., code, encoder, decoder) via end–to–end average mutual information per source symbol versus channel capacity C

$$\mathrm{IPC}(C) \stackrel{\mathrm{def}}{=} \frac{1}{K} I(\vec{U}; \vec{V}). \tag{2}$$

Note, $C \in [0, C_{\max}]$, e.g., for discrete, M–ary equiprobable signaling $C_{\max} = \log_2(M)$.

Example: Repetition Code of Rate $R = 1/2$

For a repetition code of rate–1/2 the IPC can easily be calculated analytically. As the information block length is $K = 1$ vector–wise and symbol–wise encoder input is equal, i.e., $\vec{U} = U$. Hence, $I(\vec{U}; \vec{V}) = I(U; \vec{V})$ can be calculated from the two probability density functions $p(\vec{V}|U = 1)$ and $p(\vec{V}|U = 0)$.

For some channels the analysis can be simplified even more. If the physical channel is a BSC with bit error ratio ϵ, applying a rate–1/2 repetition code results in a binary error–and–erasure channel with error probability ϵ^2 and erasure probability $2(1-\epsilon)\epsilon$. Hence, IPC(C) is easily expressed as the capacity of the equivalent channel

$$\mathrm{IPC}(C) = 1 - 2\mathrm{e}_2(\epsilon) + \mathrm{e}_2\left[(1-\epsilon)^2 + \epsilon^2\right] \text{ with } \epsilon = \mathrm{e}_2^{-1}(1-C). \tag{3}$$

$\mathrm{e}_2(\cdot)$ denotes the binary entropy function $\mathrm{e}_2(x) := -x\log_2(x) - (1-x)\log_2(1-x))$, $x \in (0,1)$, e_2^{-1} is its inverse for $x \in (0, 1/2)$.

The equivalent channel model for binary antipodal signaling (e.g., BPSK with signal points ± 1) over the AWGN channel ($C = C_{\mathrm{BPSK}}\left(1/\sigma_n^2\right) = C_{\mathrm{BPSK}}\left(2E_{\mathrm{s}}/\mathcal{N}_0\right)$ [13]) is even simpler. The optimum "decoder" performs maximum ratio combining [3] and the rate–1/2 repetition code leads to a doubling of signal energy. Here, the IPC(C) reads

$$\mathrm{IPC}(C) = C_{\mathrm{BPSK}}\left(\frac{2}{\sigma_n^2}\right) = C_{\mathrm{BPSK}}\left(4\frac{E_{\mathrm{s}}}{\mathcal{N}_0}\right). \tag{4}$$

Fig. 2 shows the resulting IPC(C) for coded transmission, using a rate–1/2 repetition code, over a BSC and over an AWGN channel with binary antipodal modulation.

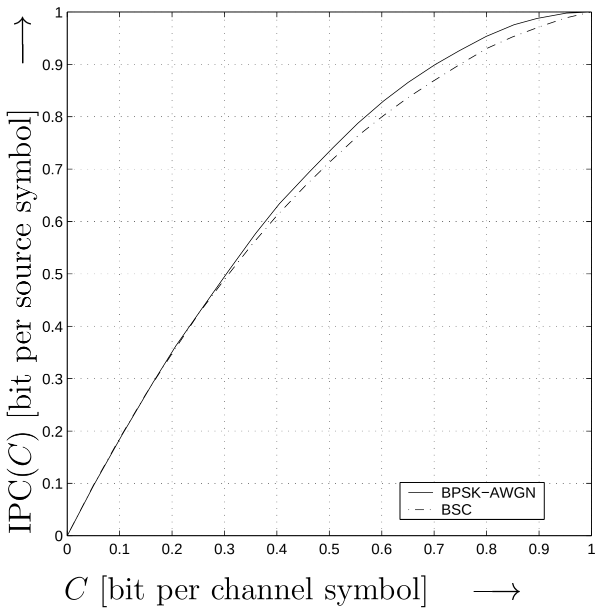

Figure 2: IPC(C) for a rate–1/2 repetition code over a BPSK–AWGN and BSC channel.

We observe, that the IPC is almost independent of the applied channel. This is due to the fact, that entropy and mutual information are the information theoretic characterizations of a probability distribution. Regarding an M–ary random variable, this data compression from space $[0,1]^{M-1}$ to the interval $[0, \log_2(M)]$ is lossless. A single real number represents sufficient information to characterize a probability distribution for information representation and transmission if blocks of N symbols are regarded and $N \to \infty$.

Obviously, the IPC offers a good characterization of a coding scheme, only specifying the channel by C (almost independent of the actual channel) even for $N = 2$.

3 IPC of Ideal Codes

In the following we define an *ideal coding scheme* and derive its IPC. As long as $C \geq R$, reliable communication is possible, and in this region IPC(C) $= 1$ holds for an ideal coding scheme. For $R \geq C$ we can derive the IPC of an ideal coding scheme by upper and lower bounding of $I(\vec{U};\vec{V})$. Rates above capacity are of interest for cancatenated coding schemes, as the component codes operate in this region.

Data processing theorem [6] tells us that the capacity of the end–to–end channel can not be larger than the capacity of the physical channel, i.e.,

$$I(\vec{U};\vec{V}) \leq I(\vec{X};\vec{Y}) \leq NC. \tag{5}$$

As memory increases mutual information [19], the symbol–wise mutual information between U_j and corresponding soft–output value V_j can not exceed the vector–wise one.

$$I(\vec{U};\vec{V}) \geq \sum_{j=1}^{K} I(U_j;V_j). \tag{6}$$

With the definition of mean mutual information

$$\bar{I}(U;V) \stackrel{\text{def}}{=} \frac{1}{K}\sum_{j=1}^{K} I(U_j;V_j) \tag{7}$$

we find an upper bound on the symbol–wise mutual information:

$$\bar{I}(U;V) \leq \frac{1}{K} I(\vec{U};\vec{V}) \leq \frac{N}{K} C = C/R. \tag{8}$$

The hard limiter that converts the soft–output values V_j into the hard–output values $\hat{U}_j$ is, at least in most cases, a lossy data processor. Hence, the symbol–wise mutual information is further decreased for the end–to–end channel that additionally comprises the hard limiter:

$$\bar{I}(U;\hat{U}) \stackrel{\text{def}}{=} \frac{1}{K}\sum_{j=1}^{K} I(U_j;\hat{U}_j) \leq \bar{I}(U;V). \tag{9}$$

Combining (9), (8) and (5) yields a chain of inequalities:

$$\bar{I}(U;\hat{U}) \leq \bar{I}(U;V) \leq \frac{1}{K} I(\vec{U};\vec{V}) \leq \frac{1}{K} I(\vec{X};\vec{Y}) \leq C/R. \tag{10}$$

Next, we apply Fano's inequality [8] for binary symbols which reads $\mathrm{e}_2(\mathrm{BER}_j) \geq H(U_j|\hat{U}_j)$ to the whole block of K information symbols

$$\frac{1}{K}\sum_{j=1}^{K} H(U_j|\hat{U}_j) \leq \frac{1}{K}\sum_{j=1}^{K} \mathrm{e}_2(\mathrm{BER}_j) = E_j\{\mathrm{e}_2(\mathrm{BER}_j)\} \leq \mathrm{e}_2\left(E_j\{\mathrm{BER}_j\}\right) = \mathrm{e}_2\left(\mathrm{BER}\right). \tag{11}$$

Therby, Jensen's inequality $\mathrm{E}\{f(x)\} \leq f\left(\mathrm{E}\{x\}\right)$ for convex $\cap$ functions $f(x)$ has been used. We see that the average entropy of U having observed $\hat{U}$ is smaller or equal to the binary entropy function of the average bit error ratio. The average mutual information between U and $\hat{U}$ is given by:

$$\bar{I}(U;\hat{U}) = \frac{1}{K}\sum_{j=1}^{K}\Big(H(U_j) - H(U_j|\hat{U}_j)\Big) = 1 - \frac{1}{K}\sum_{j=1}^{K} H(U_j|\hat{U}_j) \geq 1 - \mathrm{e}_2\left(\mathrm{BER}\right). \tag{12}$$

Since $e_2(x)$ is strictly monotonic in the interval $(0, 0.5)$, the bit error ratio of any coding scheme providing an end–to–end average mutual information $\bar{I}(U;\hat{U})$ has to be larger than or equal to the one given by assuming the hard–output $\hat{U}$ is memoryless:

$$\mathrm{BER} \geq e_2^{-1}\left(1 - \bar{I}(U;\hat{U})\right). \tag{13}$$

However, rate–distortion–theory [16] postulates, that if an end–to–end average bit error ratio $\mathrm{BER_T} < 0.5$ is tolerated, a code with rate R and appropriate decoding rule exists which achieves an average bit error ratio $\mathrm{BER} \leq \mathrm{BER_T}$ as long as $R \leq \frac{C}{1-e_2(\mathrm{BER_T})}$ and $N \to \infty$. Hence,

$$\frac{C}{R} \geq 1 - e_2(\mathrm{BER}) \stackrel{\mathrm{def}}{=} C_{\mathrm{BSC}}(\mathrm{BER}) \tag{14}$$

holds.

An *ideal coding scheme* is a coding scheme (i.e., code, encoder, decoder) with rate $R = \frac{C}{1-e_2(\mathrm{BER_T})}$ whose average bit error ratio does not exceed a tolerated bit error ratio $\mathrm{BER_T}$: $\mathrm{BER} \leq \mathrm{BER_T}$. As (10) and (11) require $\mathrm{BER} \geq \mathrm{BER_T}$, we get a chain of equalities for the ideal coding scheme.

$$1 - e_2(\mathrm{BER}) = \bar{I}(U;\hat{U}) = \frac{1}{K} I(\vec{U};\vec{V}) = C/R = 1 - e_2(\mathrm{BER_T}). \tag{15}$$

From (2) and (15) follows $\mathrm{IPC}(C) = C/R$ for $C \leq R$, i.e., if an average of $C \cdot \frac{1}{R} < 1$ bit per $\frac{1}{R}$ channel uses is collected, a BSC with $\mathrm{BER} = e_2^{-1}(1 - C/R)$ results. Fig. 3 shows the $\mathrm{IPC}(C)$ for an ideal coding scheme.

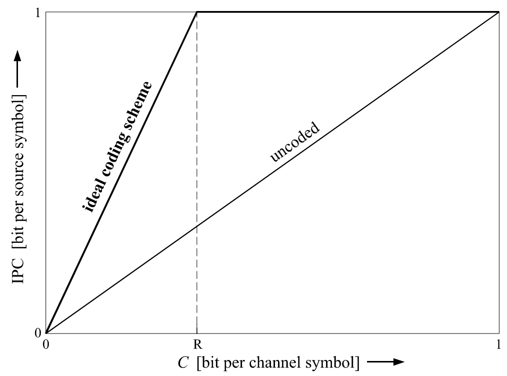

Figure 3: IPC of an ideal coding scheme with rate R.

A trivial lower bound is also given in Fig. 3. If transmission is *uncoded* the end–to–end channel and the physical channel are the same and hence, have the same capacity. The IPC of uncoded transmission is $\mathrm{IPC}(C) = C$.

We can conclude that for an ideal coding scheme the end–to–end hard decision channel is a *memoryless* BSC achieving minimum possible bit error ratio. If we look at the coding scheme from a point of serially concatenated codes we would combine the inner ideal coding scheme, that results in a bit error ratio $\mathrm{BER_T}$, with an outer code. From $\bar{I}(U;\hat{U}) = \frac{1}{K} I(\vec{U};\hat{\vec{U}})$ we see, that interleaving has no impact on the quality of the decoder output. Furthermore, as $\bar{I}(U;\hat{U}) = \bar{I}(U;V)$, soft–output has no benefit over hard–output and all symbols $\hat{U}_j$ are equally reliable. The single parameter post–decoding error probability $\mathrm{BER_T}$ entirely specifies data reliability for all decoder output symbols, i.e., soft–output decoding is important for *non–ideal* coding schemes, only.

4 Optimum Soft–Output Decoding

We restrict to encoders that perform a bijective encoding $\vec{U} \longleftrightarrow \vec{X}$ which are lossless data processors and $H(\vec{U}|\vec{X}) = 0$ holds. As $H(U) = 1$ is assumed, the a–priori probability for selecting codeword $\vec{x}_j$ is $\Pr(\vec{X} = \vec{x}_j) = 2^{-K}, \quad \forall j \in \{1, 2, \ldots, 2^K\}$, i.e., $I(\vec{U}; \vec{X}) = H(\vec{U}) = H(\vec{X}) = K$. Then, the condition for lossless soft–output decoding is:

$$I(\vec{U}; \vec{V}) \stackrel{!}{=} I(\vec{X}; \vec{Y}). \tag{16}$$

A decoder that performs a bijective mapping $\vec{Y} \longleftrightarrow \vec{V}$ would fulfill (16). An example for a nontrivial optimum soft–output decoder is a decoder that provides the complete list of a–posteriori code word probabilities. Because of the above assumptions

$$\Pr(\vec{X} = \vec{x}_i|\vec{Y}) = \Pr(\vec{U} = \vec{u}_i|\vec{Y}) = \frac{\Pr(\vec{Y}|\vec{x}_i)}{\underbrace{\sum_{j=1}^{2^K} \Pr(\vec{Y}|\vec{x}_j)}_{\text{irrelevant normaliziation factor}}} \tag{17}$$

holds. Noteworthy, the probability $\Pr(\vec{Y}|\vec{x}_i)$ fully characterizes the channel $\vec{X} \rightarrow \vec{Y}$.

The list of a–posteriori codeword probabilities is a *code property* independent of the particular encoding. Hence, $\mathrm{IPC}(C) \stackrel{\text{def}}{=} \frac{1}{K} I(\vec{U}; \vec{V})$, which can be derived from this list, characterizes the *code* w.r.t. the applied channel. For increasing K, $\mathrm{IPC}(C)$ becomes more and more independent of the applied channel and fully characterizes the code.

A practical way to calculate $\mathrm{IPC}(C)$ is found via the chain rule of mutual information:

$$I(\vec{U}; \vec{Y}) = I(U_1; \vec{Y}) + I(U_2; \vec{Y}|U_1) + I(U_3; \vec{Y}|(U_1, U_2)) + \cdots \tag{18}$$

For $R > C$, i.e., $I(\vec{U}; \vec{Y}) \leq K$, reliable communication is possible by application of additional outer binary codes C_i with rate $R_i = I(U_i; \vec{Y}|(U_1 \cdots U_{i-1}))$ and thus lowering the total rate below C. This scheme is depicted in Fig. 4. The chain rule leads immediately to the multilevel structure of generalized concatenated codes [21] designed w.r.t. a capacity criterion [18]. Furthermore we see, that multistage decoding is optimum in principle, as it has the potential to achieve capacity.

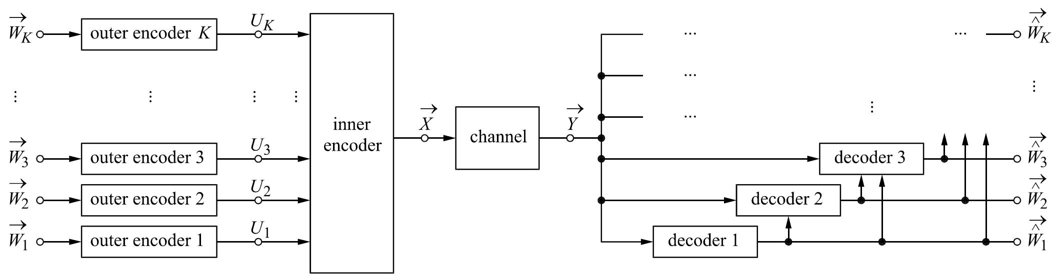

Figure 4: Generalized concatenated code with multistage decoding.

As mentined above $I(\vec{U}; \vec{Y})$ is independent of the particular encoding used, but $I(U_i; \vec{Y}|(U_1 \cdots U_{i-1}))$ depends on the encoding. Thus, the rate design of a generalized concatenated code is question of the encoder mapping of the inner code. Optimality of such a coding scheme together with multistage decoding does, however, not depend on the particular encoding.

For a linear binary, time invariant convolutional code with linear encoding over memoryless symmetric channels, as, e.g., the BSC or binary antipodal signaling over the AWGN channel, $I(U_i;\vec{Y}|(U_1\cdots U_{i-1}))$ does not depend on the particular choice of $U_1\cdots U_{i-1}$. Hence, without loss of generality we can assume that the all–zero information word and due to linear encoding also the all–zero codeword has been transmitted:

$$I(U_i;\vec{Y}|(U_1\cdots U_{i-1})) = I(U_i;\vec{Y}|(\underbrace{00\cdots 00}_{i-1})) \tag{19}$$

Let us consider a trellis representation of the encoder, such that U_i is the i^{th} input bit to the encoder. In multistage decoding, every U_i known from lower stages is used for successive thinning of the code space. While a Viterbi decoder [17] or a BCJR decoder [1] has to compare all paths of the trellis, the optimum multistage decoder always starts from a given state. Due to linearity, the decoding situation is always the same as for the very first bit and starting from the all–zero state:

$$I(U_i;\vec{Y}|(U_1\cdots U_{i-1})) = I(U_1;\vec{Y}); \quad I(\vec{U};\vec{Y}) = K\cdot I(U_1;\vec{Y}). \tag{20}$$

The mutual information $I(U_1;\vec{Y})$ between channel output $\vec{Y}$ and a single symbol U_1 is the same as the mutual information between the soft–output V_1 generated by an optimum symbol–by–symbol decoding algorithm as the BCJR decoder [1] and U_1. Hence, $\mathrm{IPC}(C) = I(U_1;\vec{Y}) = I(U_1;V_1)$ is accessible via simulations. Firstly, the probability density $\mathrm{p}(V_1|U_1=0)$ of the soft–output V_1 for the first bit U_1 has to be determined. Soft–Output values V_1 can be obtained from BCJR decoding. Due to the linear code, linear encoding and symmetric channel, $\mathrm{p}(V_1|U_1=1)$ is symmetric to $\mathrm{p}(V_1|U_1=0)$. If the soft–output V_1 is given as probabilities $\mathrm{p}(V_1|U_1=1) = \mathrm{p}((1-V_1)|U_1=0)$ is valid, whereas for log–likelihood–ratios (LLR) [10] $\mathrm{p}(V_1|U_1=1) = \mathrm{p}(-V_1|U_1=0)$ holds. Mutual information $I(U_1;\vec{Y})$ can be calculated from the probability density $\mathrm{p}(V_1|U_1=0)$ via:

$$\begin{aligned}\mathrm{IPC}(C) &= I(U_1;\vec{Y}) = I(U_1;V_1) \\ &= \int_{-\infty}^{\infty} \mathrm{p}(V_1|U_1=0)\log_2\left(\frac{2\mathrm{p}(V_1|U_1=0)}{\mathrm{p}(V_1|U_1=0)+\mathrm{p}(V_1|U_1=1)}\right)\mathrm{d}V_1.\end{aligned} \tag{21}$$

Exemplary the IPC(C) for rate–1/2 maximum free distance convolutional codes [13] of memory $\nu=2$, $\nu=3$, $\nu=4$ and $\nu=5$ for a BPSK–AWGN channel are given in Fig. 5, and for a BSC in Fig. 6, respectively.

We note the following observations: For $C<R$ these simple codes perform astonishingly close to the limit of ideal coding schemes and the results are very similar for both channels. The remaining difference vanishes for increased memory ν. But for $C\to 1$ the difference to the ideal coding scheme is quite large. Hence, it is obvious, that convolutional codes can be applied more successfully in the region $C<R$, i.e., as component codes in concatenated coding schemes, than for establishing highly reliable communication .

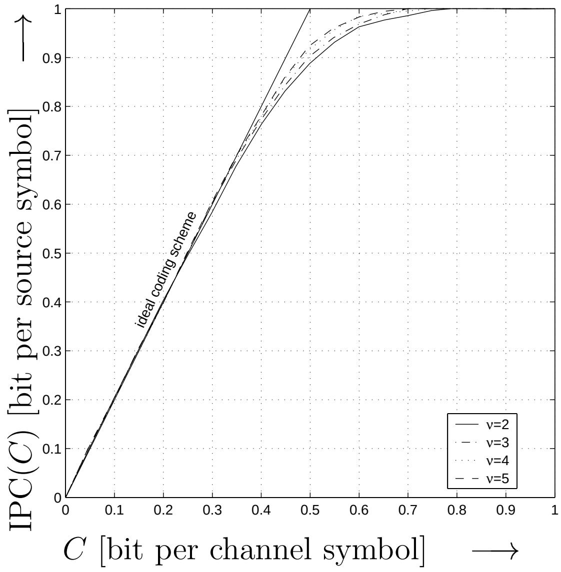

Figure 5: IPC(C) for rate–1/2 Convolutional Codes over a BPSK–AWGN channel.

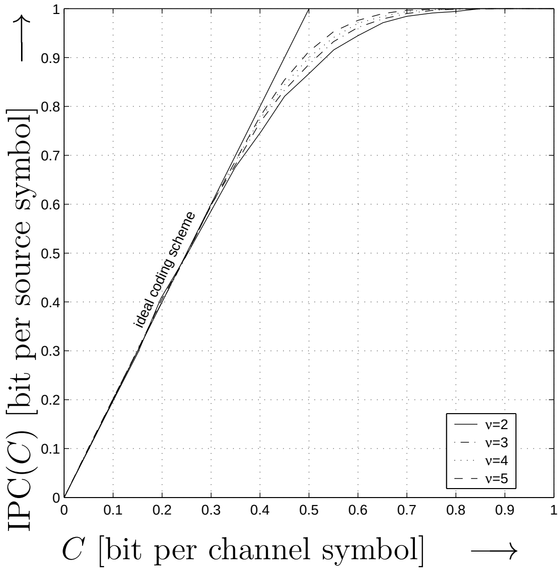

Figure 6: IPC(C) for rate–1/2 Convolutional Codes over a BSC.

5 Optimum Symbol–by–Symbol Decoding

In order to analyze optimum symbol–by–symbol soft–output decoding we extend the transmission model. Additionally we introduce an (infinite) interleaver π_∞ before encoding that converts the end–to–end channel between $\vec{U}$ and $\vec{V}$ to a memoryless channel.

The soft–output w.r.t. symbol U_i is the a–posteriori probability taking the received vector $\vec{Y}$ and code constraints into account, i.e.,

$$V_i \stackrel{\text{def}}{=} \Pr\left(U_i = 0|\vec{Y}\right). \tag{22}$$

After symbol–by–symbol decoding and interleaving an end–to–end channel with mutual information

$$\bar{I}(U;V) = \frac{1}{K}\sum_{j=1}^{K} I(U_j;V_j) \tag{23}$$

is observed.

If we compare this result to the optimum soft–out decoder, via chain rule we observe that the knowledge on previous symbols is discarded by symbol–by–symbol estimation. Mathematically, we have

$$K \cdot \bar{I}(U;V) = I(U_1;V_1) + I(U_2;V_2|\cancel{U_1}) + I(U_3;V_3|(\cancel{U_1,U_2})) + \cdots \leq I(\vec{U};\vec{V}). \tag{24}$$

Symbol–by–symbol soft–output decoding corresponds to parallel decoding of levels [15] or Bit Interleaved Coded Modulation (BICM) [20] in coded modulation. In these systems the conditioning term $U_1 \cdots U_{i-1}$ is also neglected.

$\bar{I}(U;V)$ is a property of *code* and *encoding*. Hence, the Information Processing Characteristic for symbol–by–symbol decoding and Interleaving

$$\text{IPC}_\text{I}(C) \stackrel{\text{def}}{=} \bar{I}(U;V) \tag{25}$$

characterizes code and encoding with respect to symbol–by–symbol decoding.

If we compare symbol–by–symbol decoding with optimum decoding, we see, that whereas the optimum decoder, e.g., outputs the complete list of post–decoding code word probabilities $(\Pr(\vec{U} = \vec{u}_1|\vec{Y}),\ \Pr(\vec{U} = \vec{u}_2|\vec{Y})\dots)$ that can be represented in a vector of length $2^K - 1$, the symbol–by–symbol decoder only provides the post–decoding probabilities for the information symbols $(\Pr(U_1 = 0|\vec{Y}),\ \Pr(U_2 = 0|\vec{Y})\dots)$, i.e., a vector of length K. Hence, symbol–by–symbol decoding may be interpreted as a (possibly) lossy data compression from the space $[0,1]^{2^K-1}$ to $[0,1]^K$. The soft–output of the symbol–by–symbol decoder can directly be calculated from the list of post–decoding code word probabilities:

$$V_j = \sum_{\forall \vec{u}_i \text{ with } u_{ij}=0} \Pr(\vec{U} = \vec{u}_i|\vec{Y}). \tag{26}$$

Fig. 7 and Fig. 8 show the $\mathrm{IPC_I}(C)$ for rate–1/2 convolutional codes for systematic respectively nonsystematic encoding of the same codes.

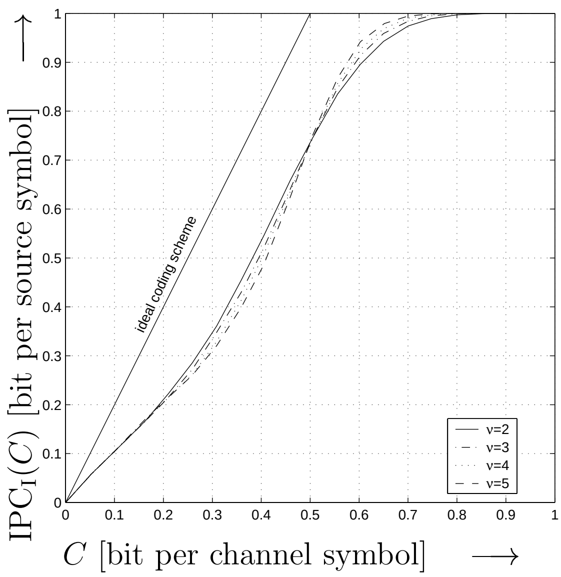

Figure 7: $\mathrm{IPC_I}(C)$ for rate–1/2 *systematic* convolutional codes.

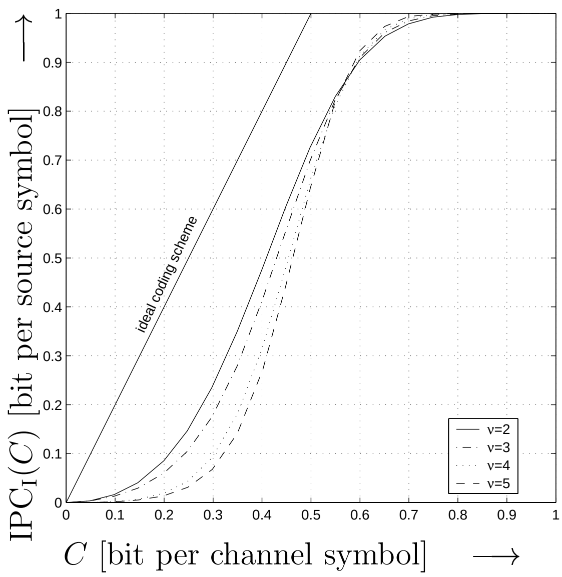

Figure 8: $\mathrm{IPC_I}(C)$ for rate–1/2 *nonsystematic* convolutional codes.

We can observe a huge loss of *optimum* symbol–by–symbol soft–output decoding, when compared to optimum soft–output decoding, especially for $C < R$. An important exception is the repetition code. As the information block length $K = 1$, symbol–by–symbol decoding is optimum $(I(\vec{U};\vec{Y}) = \bar{I}(U;V))$.

Furthermore we see, that different encoders lead to the best symbol–by–symbol soft–output $\mathrm{IPC_I}(C)$ for different channel capacities. We observe a common intersection point at $C = R$ for all examined systematic convolutional codes. Systematic convolutional codes with less memory are superior for $C < R$, whereas for $C > R$ systematic codes with more memory offer a higher end–to–end mutual information when applying symbol–by–symbol soft–output decoding. The decision whether to apply systematic or nonsystematic encoding also depends on the code memory. But always systematic encoding is superior for $C < R$ and nonsystematic encoding approaches $\mathrm{IPC_I}(C) \to 1$ faster for $C \to 1$. By this, optimum encoding and optimum choice of code memory for application of a symbol–by–symbol soft–output decoding algorithm unfortunately is a question of channel quality.

6 Conclusion

The analysis of codes and decoders w.r.t soft–output decoding shows that the code properties of convolutional codes are almost ideal for $R > C$. Hence, convolutional codes are well suited for concatenated coding schemes, where the component codes operate at rates above capacity. Furthermore we have seen, that the IPC of long codes does not significantly vary with the channel used. Hence, as for random codes, a single code construction is sufficient to approach channel capacity for a class of channels, e.g., for all binary input, symmetric channels.

Optimum symbol–by–symbol decoding is shown to be suboptimal, as data processing within the decoder is lossy. But, the degree of suboptimality depends on code memory and encoding. Hence, different concatenation parameters require different choices of component codes. Using the characteristic function $\mathrm{IPC_I}$ this decision can be made without any simulations of the entire concatenated coding scheme.

References

[1] L. Bahl, J. Cocke, F. Jelinek, J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate". *IEEE Trans. Inform. Theory*, vol. IT–20, no. 2, pp. 284–287, 1974.

[2] C. Berrou, A. Glavieux, and P. Thitimashima, "Near Shannon Limit Error–Correcting Coding and Decoding: Turbo–Codes". *Proceedings of ICC '93*, pp. 1064-1070, 1993.

[3] D. G. Brennan, "Linear diversity combining techniques". *Proceedings of the IRE*, vol. 47, pp. 1075–1102, Jun. 1959.

[4] S. ten Brink, "Convergence of iterative decoding", *IEE Electronics Letters*, vol. 35, no. 10, pp. 806–808, May 1999.

[5] S. ten Brink, "Designing Iterative Decoding Schemes with the Extrinsic Information Transfer Chart". *AEÜ Int. J. Electron. Commun.*, no. 6, pp. 389–398, Dec. 2000

[6] T. Cover J. Thomas, "Elements of Information Theory". *Wiley*, New York, 1991.

[7] D. Divsalar, S. Dolinar, F. Pollara, "Low Complexity Turbo–like Codes". *Proceedings of International Symposium on Turbo Codes*, pp. 73-80, Brest, France, Sept. 2000.

[8] R. M. Fano, "Transmission of Information: A Statistical Theory of Communication". *Wiley*, New York, 1961.

[9] A. Hammons, H. Gamal, "On the Convergence of the Turbo Decoder", *Proceedings of CISS '00*, Princeton university, Princeton, NY, Mar. 2000.

[10] J. Hagenauer, E. Offer, L. Papke, "Iterative Decoding of Binary Block and Convolutional Codes", *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 429–445, 1996.

[11] S. Huettinger, S. ten Brink, J. Huber, "Turbo–Code representation of RA–codes and DRS–codes for reduced decoding complexity", *Proceedings of CISS '01*, Johns Hopkins University, Baltimore, Maryland, pp. 118–123, Mar. 2001.

[12] H. Jin, and R. McEliece, "RA Codes Achieve AWGN Channel Capacity". *13th AAECC Proceedings*, Springer Publication 1719. pp. 10–18, 1999.

[13] J. Proakis, "Digital Communications". *McGraw Hill*, 4th Edition, 2001.

[14] T. Richardson and R. Urbanke, "The capacity of low-density parity check codes under message-passing decoding". *submitted to IEEE Transactions on Information Theory*, 1998.

[15] Schramm, P., "Multilevel Coding with Independent Decoding on Levels for Efficient Communication on Static and Interleaved Fading Channels", *In Proceedings of PIMRC '97*, pp. 1196–1200, Helsinki, Finland, Sept. 1997.

[16] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion". *IRE National Convention Record*, Part 4, pp. 142–163, 1959.

[17] A. Viterbi, "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm". *IEEE Trans. Inform. Theory*, vol. IT-13, no. 2, pp. 260–269, Apr. 1967.

[18] U. Wachsmann, R. Fischer, J. Huber, "Multilevel Codes: Theoretical Concepts and Practical Design Rules". *IEEE Transactions on Information Theory*, pp. 1361-1391, July 1999.

[19] J. Wolfowitz, "Memory increases capacity". *Inf. and Control*, 11:pp. 423–428, 1967.

[20] Zehavi, E., "8–PSK Trellis Codes for a Rayleigh Channel", *IEEE Transactions on Communications* , vol. 40, pp. 873–884, May 1992.

[21] V. Zyablov, S. Shavgulidze and M. Bossert, "An Introduction to Generalized Concatenated Codes". *European Transactions on Telecommunications*, vol. 10, no. 6, 1999.