

Reachability of the Capacity in Parallel Quantised Channels

Christoph Schmitz
UMIC Research Centre
RWTH Aachen University
52056 Aachen, Germany
schmitz@umic.rwth-aachen.de

Anke Schmeink
UMIC Research Centre
RWTH Aachen University
52056 Aachen, Germany
schmeink@umic.rwth-aachen.de

Abstract—Based upon profound knowledge about wireless communication systems that was gathered during the past decades, modelling biological systems analogously has become more and more popular. Although the assumptions about the latter are sometimes less reliable than those about technical systems, some models have proven useful in explaining the behaviour of biological communication systems for example when affected by diseases like Alzheimer. In this work, such a model based upon parallel quantised channels is examined further. The key aspect is to figure out whether the proven capacity bounds are reachable, and under which conditions.

I. INTRODUCTION

In [1], an approach for modelling biological communication systems by a set of parallel channels, together with quantisation and information fusion, is presented. Some real-valued signal X is transmitted via n channels. In each channel i the signal is disturbed by an additive noise term W_i . Both X and the W_i follow a zero-mean Gaussian distribution with variances σ_X^2 and σ_W^2 , respectively. The sums $V_i = X + W_i$ are then quantised by a function $q : \mathbb{R} \rightarrow \{0, \dots, m-1\}$, and a function $u : \{0, \dots, m-1\}^n \rightarrow \{0, \dots, m-1\}$ is employed to merge the (possibly different) values $Y_i = q(V_i)$ to a decision $U = u(Y_1, \dots, Y_n)$. This model is depicted in Figure 1. The authors show that the capacity of this channel, expressed by the mutual information, is bounded by

$$I(X; U) \leq \min \left\{ \log m, \frac{1}{2} \log \left(1 + n \frac{\sigma_X^2}{\sigma_W^2} \right) \right\}, \quad (1)$$

whereupon the term $\log m$ reflects the maximal amount of information that can be transmitted by a quantisation with m steps. The second term refers to the capacity of n parallel AWGN channels.

In this work, we take a closer look at the question if this bound is reachable. First, the capacity of the quantised channel with real-valued input X will be examined, with particular attention paid to the capacity restriction induced by the quantisation. Then, quantised channels with discrete input will be considered. Based on input distributions proposed in [2] and [3], the capacity achievable in our channel model will be calculated. Third, numerical results for the mutual information of a quantised AWGN channel will be compared to the upper bound that originates from the data processing inequality. At the end, simulation results for a proposed

information fusion algorithm will be presented, which show the efficiency of the algorithm.

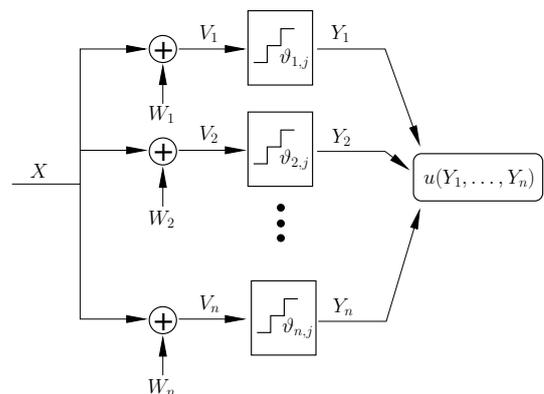


Fig. 1. Model for Parallel Channels with Quantisation

II. RESULTS AND DISCUSSION

A. Identifying the Limiting Factor

Assuming optimal quantisation and information fusion, it can be shown that the capacity bound in Equation 1 can be reached. In this case, the limiting factor is either the capacity of the parallel AWGN channels, or the amount of information that can be preserved during the quantisation. A closer look onto Equation 1 shows that the quantisation is the limiting factor if $m < \sqrt{1 + n\sigma_X^2/\sigma_W^2}$, otherwise the capacity is limited by the AWGN channels. On the one hand, this means that, given the number of channels n and the signal-to-noise ratio (SNR) σ_X^2/σ_W^2 , it is possible to ensure that the quantisation is not the limiting factor, by increasing the number of steps m above this bound. On the other hand, even when an arbitrary low SNR is given, increasing the number of parallel channels n leads to a point where the capacity of the AWGN channels is not the limiting factor anymore.

B. Quantisation with Inadequate Thresholds

As seen before, one of the limiting factors for the capacity of the whole channel is the quantisation step $Y_i = q(V_i)$. As Y_i is a discrete random variable with m outcomes,

$$I(V_i; Y_i) = H(Y_i) - H(Y_i|V_i) = H(Y_i)$$

is bounded by the value $\log m$. To achieve this bound, the quantising function q has to ensure that Y_i is uniformly distributed on $\{0, \dots, m-1\}$. With knowledge about the distribution of X and W_i , this is easily possible. For stochastically independent $X \sim N(0, \sigma_X^2)$ and $W_i \sim N(0, \sigma_W^2)$, the sum $X + W_i$ is known to be $N(0, \sigma_X^2 + \sigma_W^2)$ -distributed. Choosing the threshold values

$$\vartheta_j = \sqrt{\sigma_X^2 + \sigma_W^2} \Phi^{-1} \left(\frac{j}{m} \right), \quad 0 \leq j \leq m,$$

the quantising function

$$q(v) = \sum_{j=0}^{m-1} j \mathbb{I}\{\vartheta_j < v \leq \vartheta_{j+1}\}$$

defined in [1] provides the desired result, as it was also shown in [4]. Φ^{-1} denotes the inverse of the cumulative distribution function Φ of the standard Gaussian distribution. Note that $\vartheta_0 = -\infty$ and $\vartheta_m = \infty$.

The situation changes when the quantiser is not aware of the presence of the noise term, or has no information about its distribution. In the following we assume that the quantising function is optimised for the distribution of X , but applied to the random variable $V_i = X + W_i$. Thus, it utilises the thresholds

$$\vartheta_j = \sigma_X \Phi^{-1} \left(\frac{j}{m} \right), \quad 0 \leq j \leq m,$$

to quantise a $N(0, \sigma_X^2 + \sigma_W^2)$ -distributed random variable. The resulting distribution of Y_i can be described by

$$\begin{aligned} P(Y_i = j) &= \Phi \left(\frac{\vartheta_{j+1}}{\sqrt{\sigma_X^2 + \sigma_W^2}} \right) - \Phi \left(\frac{\vartheta_j}{\sqrt{\sigma_X^2 + \sigma_W^2}} \right) \\ &= \Phi \left(\frac{\sigma_X \Phi^{-1} \left(\frac{j+1}{m} \right)}{\sqrt{\sigma_X^2 + \sigma_W^2}} \right) - \Phi \left(\frac{\sigma_X \Phi^{-1} \left(\frac{j}{m} \right)}{\sqrt{\sigma_X^2 + \sigma_W^2}} \right). \end{aligned}$$

Note that it only depends on m and on

$$\frac{\sigma_X}{\sqrt{\sigma_X^2 + \sigma_W^2}} = \sqrt{\frac{\sigma_X^2}{\sigma_X^2 + \sigma_W^2}},$$

a function of the ratio between the variances of X and $X + W_i$. If this ratio is one, we have again

$$\begin{aligned} P(Y_i = j) &= \Phi \left(\Phi^{-1} \left(\frac{j+1}{m} \right) \right) - \Phi \left(\Phi^{-1} \left(\frac{j}{m} \right) \right) \\ &= \frac{j+1}{m} - \frac{j}{m} = \frac{1}{m}, \end{aligned}$$

in this case Y_i is uniformly distributed. There exists no closed-form expression for the entropy

$$H(Y_i) = - \sum_{j=0}^{m-1} P(Y_i = j) \log P(Y_i = j)$$

as a function of the ratio between the variances, but of course it can easily be computed numerically.

Some results for different values of m are shown in Figure 2, in all cases the entropy is calculated using the dyadic

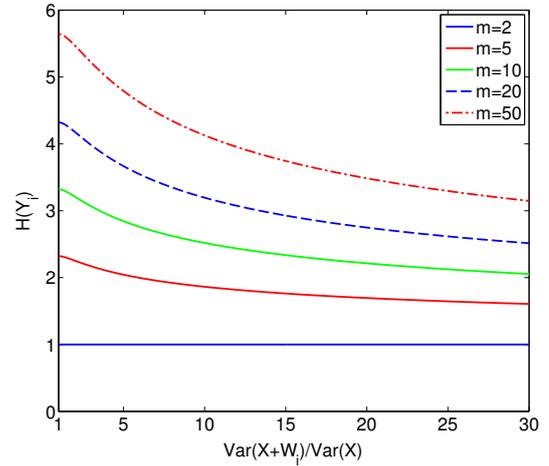


Fig. 2. Entropy of Y_i under Inadequate Quantisation

logarithm. Thus, the maximal entropy is one for $m = 2$, but higher for greater values of m . In all cases the maximal entropy is achieved when the ratio of the variances is one. This means, in other words, that the variance of W_i is zero, so $W_i \equiv 0$ and $X + W_i = X$. As $\text{Var}(W_i)$ increases, the entropy decreases in all cases except $m = 2$. In the latter case, each $N(0, \sigma_X^2 + \sigma_W^2)$ -distributed random variable is optimally quantised, as the threshold values $\vartheta_0 = -\infty$, $\vartheta_1 = 0$ and $\vartheta_2 = \infty$ always yield a uniform distribution on $\{0, 1\}$ for Y_i .

C. Discrete Input

Until now, the common input X of the parallel channels was assumed to be continuous, following a zero-mean Gaussian distribution with variance σ_X^2 . Although this seems reasonable for biological sensor systems which measure for example physical quantities like brightness or temperature, the model might also be examined in conjunction with a discrete input. As this combination is highly relevant for wireless communication systems, it has already been widely studied. Typically the noise term W_i is still assumed to be continuous in this context.

It should be kept in mind that a discrete input introduces another predicament in achieving the capacity. As seen in Equation 1, the capacity is bounded by the capacity of the AWGN channels and by the entropy achieved by the quantisation. In case of a continuous input the latter can be maximal in all cases, as long as the distribution of V_i is known to the quantiser, this was shown in Section II-B. Thus, the input distribution can be chosen to maximise the capacity of the AWGN channels, which results in a Gaussian distribution. A discrete input, on the other hand, should be close to a uniform distribution in case of high SNR, to maximise the entropy $H(Y_i)$. In case of low SNR, this counteracts the effort to achieve the capacity of the AWGN channels, so a compromise has to be found.

To examine the behaviour of our model with discrete input, some different types of input distributions will now be considered. All of them are zero-mean discrete distributions

with m equidistant possible values x_j and variance σ_X^2 . The uniquely determined uniform distribution with these properties is given by $X \sim U(\{x_0, \dots, x_{m-1}\})$ with

$$x_j = \frac{3 \sigma_X (2j + 1 - m)}{\sqrt{3(m^2 - 1)}}$$

for $0 \leq j \leq m - 1$. Thus, the thresholds for the ML quantisation are

$$\begin{aligned} \vartheta_0 &= -\infty, \quad \vartheta_m = \infty \quad \text{and} \\ \vartheta_j &= \frac{3 \sigma_X (2j - m)}{\sqrt{3(m^2 - 1)}}, \quad 1 \leq j \leq m - 1, \end{aligned}$$

yielding the transition probabilities

$$P(Y_i = 0|X = x_j) = \Phi \left(\frac{3 \sigma_X (1 - 2j)}{\sigma_W \sqrt{3(m^2 - 1)}} \right),$$

$$P(Y_i = m - 1|X = x_j) = 1 - \Phi \left(\frac{3 \sigma_X (2(m - j) - 3)}{\sigma_W \sqrt{3(m^2 - 1)}} \right)$$

and

$$\begin{aligned} P(Y_i = k|X = x_j) &= \Phi \left(\frac{3 \sigma_X (2(k - j) + 1)}{\sigma_W \sqrt{3(m^2 - 1)}} \right) \\ &\quad - \Phi \left(\frac{3 \sigma_X (2(k - j) - 1)}{\sigma_W \sqrt{3(m^2 - 1)}} \right) \end{aligned}$$

for $1 \leq k \leq m - 2$. For a high SNR σ_X^2/σ_W^2 , the input value $X = x_j$ induces the output value $Y_i = j$ with high probability, thus the mutual information $I(X; Y_i)$ and the entropy $H(Y_i)$ are nearly maximal.

As shown in [3], the capacity-achieving discrete input distribution is not a uniform distribution anymore when the SNR of the channel is very low. Instead, it looks like an approximation of the Gaussian distribution, which is known as the capacity-achieving *continuous* input distribution. This leads to the idea of employing a symmetric binomial distribution on m equidistant possible values x_j as input distribution. A zero-mean distribution of this kind with variance σ_X^2 is uniquely determined, it is described by

$$P(X = x_j) = \binom{m-1}{j} \frac{1}{2^{m-1}}$$

with

$$x_j = \frac{\sigma_X (2j + 1 - m)}{\sqrt{m - 1}}$$

for $0 \leq j \leq m - 1$. It should be noted that this distribution is identical to the above-mentioned uniform distribution in the case $m = 2$, and that for higher values of m the spacing of the input values x_j is wider compared to the uniform distribution, given the same variance σ_X^2 . The thresholds for the ML quantisation are

$$\begin{aligned} \vartheta_0 &= -\infty, \quad \vartheta_m = \infty \quad \text{and} \\ \vartheta_j &= \frac{\sigma_X (2j - m)}{\sqrt{m - 1}}, \quad 1 \leq j \leq m - 1, \end{aligned}$$

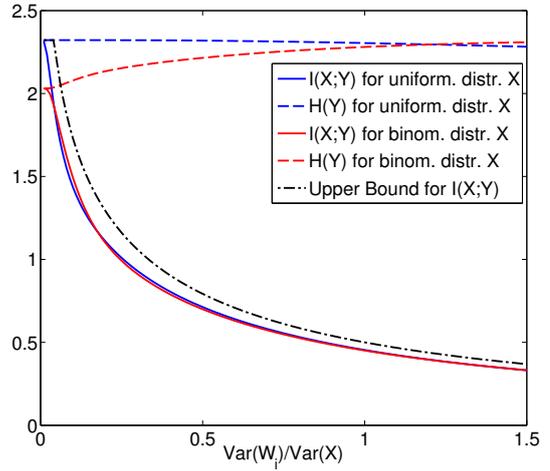


Fig. 3. The Mutual Information and Entropy of Y_i for Discrete Input

however it should be noted that they are not identical to those of the maximum a posteriori (MAP) quantisation anymore, as the input distribution differs from the uniform distribution. The transition probabilities are

$$P(Y_i = 0|X = x_j) = \Phi \left(\frac{\sigma_X (1 - 2j)}{\sigma_W \sqrt{m - 1}} \right),$$

$$P(Y_i = m - 1|X = x_j) = 1 - \Phi \left(\frac{\sigma_X (2(m - j) - 3)}{\sigma_W \sqrt{m - 1}} \right)$$

and

$$\begin{aligned} P(Y_i = k|X = x_j) &= \Phi \left(\frac{\sigma_X (2(k - j) + 1)}{\sigma_W \sqrt{m - 1}} \right) \\ &\quad - \Phi \left(\frac{\sigma_X (2(k - j) - 1)}{\sigma_W \sqrt{m - 1}} \right) \end{aligned}$$

for $1 \leq k \leq m - 2$.

Numerical results for both approaches are shown in Figure 3. X and Y_i both have $m = 5$ possible values, and the value $\text{Var}(W_i)/\text{Var}(X)$ which is the inverse of the SNR varies between 0 and 1.5. The depicted upper bound for the mutual information

$$I(X; Y_i) \leq \min \left\{ \frac{1}{2} \log \left(1 + \frac{\sigma_X^2}{\sigma_W^2} \right), \log m \right\}$$

will be explained in detail in Section II-D. As expected, the uniform input distribution performs better for a very high SNR where $H(Y_i|X)$ trends to zero, as it achieves the maximal value $H(Y_i) = \text{ld}(5)$ in this case. With decreasing SNR, the entropy of Y_i remains near this high level, but of course the mutual information $I(X; Y_i)$ decreases significantly as $H(Y_i|X)$ increases.

The binomial input distribution induces a similar distribution for Y_i when the SNR is very high, thus $H(Y_i)$ and also $I(X; Y_i)$ stay well below the theoretical limit in this case. With decreasing SNR, the changing output distribution leads to an increase of the entropy $H(Y_i)$, until it even exceeds the output entropy delivered by the uniform input distribution.

Of course, the mutual information $I(X; Y_i)$ decreases with decreasing SNR. Interestingly the mutual information is more or less equal for both approaches over nearly the whole SNR range, the second approach does not outperform the first one for higher SNR values, as one might have expected. On the other hand, both approaches nearly reach the upper bound given by the capacity of a channel with continuous input distribution.

D. Mutual Information between X and Y_i

The theoretical bound given in Equation 1 was derived using the data processing inequality to obtain

$$I(X; U) \leq \min\{I(X; \mathbf{V}), I(\mathbf{V}; \mathbf{Y}), I(\mathbf{Y}; U)\},$$

see [1]. The same argument also yields the bound

$$\begin{aligned} I(X; Y_i) &\leq \min\{I(X; V_i), I(V_i; Y_i)\} \\ &\leq \min\left\{\frac{1}{2} \log\left(1 + \frac{\sigma_X^2}{\sigma_W^2}\right), \log m\right\}, \end{aligned}$$

but of course it is questionable if this bound is tight. The attempt to calculate the desired mutual information directly yields

$$\begin{aligned} I(X; Y_i) &= H(Y_i) - H(Y_i|X) \\ &= H(Y_i) - \int_{-\infty}^{\infty} H(Y_i|X = x) f_X(x) dx \end{aligned}$$

with $H(Y_i) = -\sum_{j=0}^{m-1} q_j \log q_j$ and

$$H(Y_i|X = x) = -\sum_{j=0}^{m-1} q_{j|x} \log q_{j|x}.$$

Here the $q_j := P(Y_i = j)$ denote the probabilities for Y_i , and the $q_{j|x} := P(Y_i = j|X = x)$ denote the probabilities for Y_i given $X = x$. As it is known that $V_i = X + W_i$ is $N(0, \sigma_X^2 + \sigma_W^2)$ -distributed if no condition applies, and that it is $N(x, \sigma_W^2)$ -distributed under the condition $X = x$, these probabilities can be computed numerically for arbitrary $x \in \mathbb{R}$ and a quantisation given by thresholds $\vartheta_0, \dots, \vartheta_m$. We know that

$$q_j = \Phi\left(\frac{\vartheta_{j+1}}{\sqrt{\sigma_X^2 + \sigma_W^2}}\right) - \Phi\left(\frac{\vartheta_j}{\sqrt{\sigma_X^2 + \sigma_W^2}}\right)$$

and

$$q_{j|x} = \Phi\left(\frac{\vartheta_{j+1} - x}{\sigma_W}\right) - \Phi\left(\frac{\vartheta_j - x}{\sigma_W}\right).$$

For a $k' \in \mathbb{N}$, thresholds $-\infty = \varkappa_0 \leq \dots \leq \varkappa_{k'} = \infty$ and arbitrarily chosen $x'_0, \dots, x'_{k'-1}, x'_k \in [\varkappa_k, \varkappa_{k+1}]$, an approximation

$$\begin{aligned} H(Y_i|X) &\approx \sum_{k=0}^{k'-1} H(Y_i|X = x'_k) \int_{\varkappa_k}^{\varkappa_{k+1}} f_X(x) dx \\ &= \sum_{k=0}^{k'-1} H(Y_i|X = x'_k) \left(\Phi\left(\frac{\varkappa_{k+1}}{\sigma_X}\right) - \Phi\left(\frac{\varkappa_k}{\sigma_X}\right) \right) \end{aligned}$$

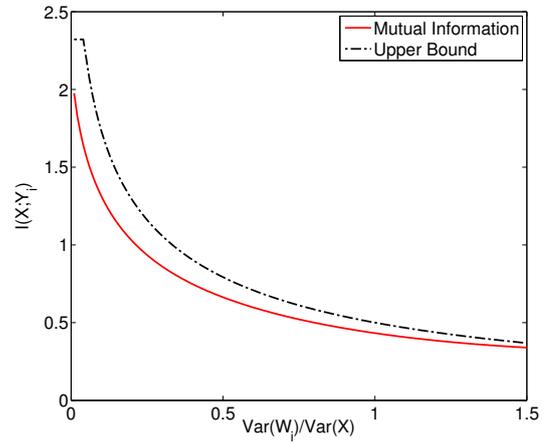


Fig. 4. The Mutual Information between X and Y_i

of the entropy can be computed, and it converges to the true value with increasing k' if the thresholds \varkappa_k are chosen in an appropriate way, for example such that

$$\int_{\varkappa_k}^{\varkappa_{k+1}} f_X(x) dx = \Phi\left(\frac{\varkappa_{k+1}}{\sigma_X}\right) - \Phi\left(\frac{\varkappa_k}{\sigma_X}\right) = \frac{1}{k'}$$

for all $0 \leq k \leq k' - 1$.

Some results for $m = 5$ are given in Figure 4, together with the theoretical upper bound which is given by

$$I(X; Y_i) \leq \min\left\{\frac{1}{2} \text{ld}\left(1 + \frac{\sigma_X^2}{\sigma_W^2}\right), \text{ld } m\right\}.$$

The value $\text{Var}(W_i)/\text{Var}(X) = \sigma_W^2/\sigma_X^2$ on the horizontal axis is the inverse of the SNR. These results were computed by dividing the whole range of real numbers into $k' = 10000$ intervals $[\varkappa_k, \varkappa_{k+1}]$ with equal probabilities for the outcome of X , and the values x'_k were determined by the rule

$$\Phi\left(\frac{x'_k}{\sigma_X}\right) = \frac{2k+1}{2k'}, \quad 0 \leq k \leq k' - 1.$$

The thresholds ϑ_j for the quantisation were chosen such that the entropy $H(Y_i)$ is maximal. While the upper bound remains at $\text{ld}(5)$ for very high values of the SNR, the actual mutual information begins to decrease immediately as the SNR decreases (or, in other words, σ_W^2/σ_X^2 increases), but it stays close to the bound and even approaches it again for lower SNR values. Of course both the bound as well as the mutual information tend to zero when the SNR goes to zero.

E. Information Fusion

It has been stated in [1] that the information fusion step is a hard problem in general. Given some identically distributed random variables Y_1, \dots, Y_n , described by the probabilities $q_j = P(Y_i = j)$, the problem is to find a mapping

$$\begin{aligned} u : \{0, \dots, m-1\}^n &\rightarrow \{0, \dots, m-1\}, \\ (Y_1, \dots, Y_n) &\mapsto u(Y_1, \dots, Y_n) =: U, \end{aligned}$$

that maximises the entropy $H(U)$. As the Y_i are functions of a common input X and independent noise terms W_i , they are

dependent to some degree. For a given entropy $H(Y_i)$ of the Y_i , an upper bound for the entropy of U is given by

$$\begin{aligned} H(U) &= H(u(Y_1, \dots, Y_n)) \leq H(Y_1, \dots, Y_n) \\ &\leq H(Y_1) + \dots + H(Y_n) = n \cdot H(Y_i). \end{aligned}$$

If the Y_i are uniformly distributed on $\{0, \dots, m-1\}$, there is a rather counterintuitive, but simple possibility to reach the bound $H(U) = \log m$. This can be achieved by defining $u(Y_1, \dots, Y_n) = Y_k$ for some arbitrary $1 \leq k \leq n$. Although the amount of information represented by the other Y_i , $i \neq k$, is wasted, this method still suffices to utilise the maximal amount of information that U can represent. Of course this result is related to the fact that the random variable U can not represent more information than one single Y_i , as they are all discrete random variables with support $\{0, \dots, m-1\}$.

In general, the distribution of U is characterised by the probabilities $P(U = j) = P(u(\mathbf{Y}) = j) = P(\mathbf{Y} \in u^{-1}(j))$, where $\mathbf{Y} = (Y_1, \dots, Y_n)$ denotes the random vector formed by the random variables Y_i , and

$$u^{-1}(j) = \{\mathbf{Y} \in \{0, \dots, m-1\}^n \mid u(\mathbf{Y}) = j\}$$

denotes the preimage of j under the mapping u . So the problem of finding the best mapping u is equivalent to the task of partitioning the set $\{0, \dots, m-1\}^n$ into m parts, such that the probabilities for \mathbf{Y} falling into the different parts are as uniform as possible. A potential approach using a greedy heuristic method is proposed in [1]. The algorithm starts with the whole set $\{0, \dots, m-1\}^n$ and the related probabilities. In each step, the two elements with the smallest probabilities are merged, until there are exactly m elements left. These form the desired partitioning of the original set.

Some simulation results for different values of m are shown in Figure 5. Here, the distribution of the Y_i was taken from the inadequate quantisation example in Section II-B, and the most extreme value $(\sigma_X^2 + \sigma_W^2)/\sigma_X^2 = 30$ for the ratio between the variances was chosen. To facilitate the computation, the Y_i are now assumed to be stochastically independent. As the common input X is quite small in relation to the independent noise terms W_i , this approximation seems acceptable. The range for the number of channels n is chosen such that the number of elements in the set $\{0, \dots, m-1\}^n$, which of course amounts to $N := m^n$, does not exceed one billion.

As mentioned before, the Y_i are already uniformly distributed in the case $m = 2$. In this case the algorithm preserves the maximal entropy when performing the information fusion. For higher values of m , the entropy of the Y_i is a lot smaller than the maximum given by the dyadic logarithm $\text{ld}(m)$. This can easily be seen in the figure, as we know that $H(U) = H(Y_i)$ for $n = 1$. When the number of channels n increases, the entropy of U quickly approaches the maximum. As the simulation shows, $H(U)$ is already nearly optimal for $n \geq 2$ in the case $m = 5$, and for $n \geq 3$ in the other cases reviewed here. Due to the heuristic concept of the algorithm, the entropy does not increase monotonically with increasing n . In the case $m = 5$ for example, the algorithm performs

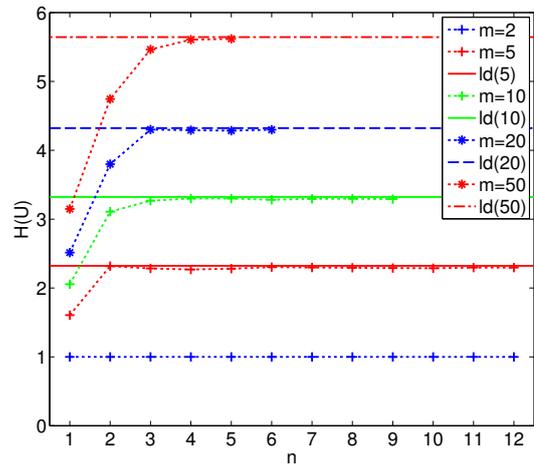


Fig. 5. Entropy of U in Relation to the Number of Channels n

better for $n = 2$ than for every other n considered, although the difference is indeed very small.

III. CONCLUSION

The bound for the capacity of our channel model given in Equation 1 can also be expressed as

$$I(X; U) \leq \min\{I(X; \mathbf{V}), I(\mathbf{V}; \mathbf{Y}), I(\mathbf{Y}; U)\},$$

which means that the capacity of the whole system is bounded by the capacities of the individual stages. While $I(X; \mathbf{V})$ is already covered by known results about AWGN channels, we have shown that $I(V_i; Y_i)$ can always be maximised by an appropriate choice of the quantisation, as long as the distribution of the V_i is known. Additionally, we examined $I(X; Y_i)$ directly and found that the bound

$$I(X; Y_i) \leq \min\left\{\frac{1}{2} \log\left(1 + \frac{\sigma_X^2}{\sigma_W^2}\right), \log m\right\}$$

is almost reachable, for continuous as well as for discrete input. For $I(Y_i; U)$, we showed that the proposed algorithm does a good job of maximising the entropy of U , even when the entropy of the Y_i is low.

ACKNOWLEDGMENT

This work was supported by the UMIC Research Centre at RWTH Aachen University and by DFG grant SCHM 2643/4-1.

REFERENCES

- [1] R. Mathar and A. Schmeink, "A bio-inspired approach to condensing information," in *IEEE International Symposium on Information Theory (ISIT)*, Saint-Petersburg, Aug. 2011, pp. 2524–2528.
- [2] J. Huang and S. Meyn, "Characterization and computation of optimal distributions for channel coding," *Information Theory, IEEE Transactions on*, vol. 51, no. 7, pp. 2336–2351, July 2005.
- [3] J. Singh, O. Dabeer, and U. Madhow, "On the limits of communication with low-precision analog-to-digital conversion at the receiver," *IEEE Transactions on Communications*, vol. 57, no. 12, pp. 3629–3639, 2009.
- [4] C. Novak, P. Fertl, and G. Matz, "Quantization for soft-output demodulators in bit-interleaved coded modulation systems," in *Proceedings of the 2009 IEEE international conference on Symposium on Information Theory - Volume 2*, ser. ISIT'09, 2009, pp. 1070–1074.
- [5] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley and Sons, Inc., 1991.