A Bio-Inspired Approach to Condensing Information

Rudolf Mathar
Inst. for Theoretical Information Technology
RWTH Aachen University
D-52056 Aachen, Germany
mathar@ti.rwth-aachen.de

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

***Abstract*— In this paper, we consider a class of models that describe parallel observations of a single source by many noisy sensors, lossy quantization at each sensor, and finally information fusion of the quantized data. Certain phenomena in biophysics and neural information processing, but also in detection networks and modern communications can be elucidated by these models. Mutual information is used as an analytical measure of information exchange. We characterize the optimum information fusion rule by maximum entropy of the corresponding output distribution. For discrete input distributions, this problem can be reduced to a generalized Knapsack problem, which is hard to solve in general. We suggest a heuristic that minimizes the decrease of entropy in each step, and show that for binary information fusion the true optimum is attained for dyadic distributions. The problem of finding optimum quantization rules is an essential part of the model and treated analogously. For input distributions with a density, optimality is achieved by determining appropriate quantization thresholds. Finally, by applying the data processing inequality, an upper bound for the mutual information of arbitrary stochastic pooling channels is found. This bound provides interesting insight into the resilience of parallel noisy information processing in biological systems.**

I. Introduction and Motivation

Present wireline and wireless communication systems are highly optimized. Transceivers, coding and multiaccess schemes are designed to achieve reliable communication at highest possible rates. By defining the concept of capacity, Shannon has provided an upper bound which cannot be exceeded by whatever practical implementation. As a matter of fact, modern digital communication systems come very close to the Shannon bound.

Biological communication and information systems in contrast have evolved over millions of years. They have been optimized by evolution subject to completely different design criteria. Biological information channels will use a rate which ensures survival and proliferation of the species. Signaling and communiciation between cells, within the brain or between entities of a species are developed by evolution to a degree such that creatures are able to cope with environmental challenges. Speed and capacity are not the primary concerns, if both are sufficient they will no more be an objective of evolutionary optimization.

On the other hand, for biological systems communication and information exchange has to be extremely reliable in a wide range of situations. If some information sources or channels are not operational, their role should be taken over by others, still functional. Massive parallelism seems to be the solution to this problem in biological systems. The retina in the human eye, inner hair cells in the auditory cochlea, and the semicircular canals of the human ear, e.g., process information in parallel and convey quantized signals to the brain. Nerve tracts serve as channels and information is mainly processed in the brain, often after local quantization and compression. Low energy consumption paired with simplicity, efficiency and adaptability are further important objectives for information exchange in living organisms.

In this work, we study bio-inspired communication by a certain class of models. A common signal is observed by many sensors, each afflicted by noise. At each sensor, observations are quantized to a finite number of states and reported to a central unit. The measurement of each node is then combined into a single decision, which is expected to represent the original input signal. Three stages are typical for this class of models: 1. multiple sensors which make stochastic observations, 2. lossy compression by quantization at each sensor and 3. information fusion at the central processing unit.

Such models are a subclass of so called *stochastic pooling networks*, a denotation first coined by [1] for a binary detection problem. This class of networks has turned out as useful model for many applications and has induced a whole series of publications. An overview of potential applications is given in [2]. Artificial sensor networks, digitized beamforming, stochastic resonance, biological neurons, cochlear implants and also complex social networks are prominent examples and widely investigated. An important background reference for the present work is [3].

The effect that the presence of noise can enhance the detection of weak signals is called *stochastic resonance* in [4]. A model of the type used in this paper is employed in the work [5] to explore this effect analytically. A common Gaussian input signal X is observed by n sensors, each subject to independent Gaussian noise, 0-1-quantization applies with equal threshold values $\vartheta = \mathrm{E}(X)$. The number of 1's is then added to form the output signal. Fig. 2 with binary quantization and $U = u(Y_1, \dots, Y_n) = \sum_{i=1}^{n} Y_i$ specifies this model. Simulation and numerical computations in [5] demonstrate in

978-1-4577-0594-6/11/$26.00 ©2011 IEEE

concert, that mutual information between input and output is enlarged by the presence of noise, particularly as the number of sensors increases and signals are mainly above thresholds. The effect is called *suprathreshold stochastic resonance* in [5]. It is of particular interest for bio-inspired, highly robust communication channels.

In this paper, we approach the problem from an information theoretic point of view. Mutual information, $I(X;U)$, between a certain stochastic input X and output U is used to describe the amount of information a channel is able to convey. We investigate mutual information stagewise between (i) the input and its noisy observations, (ii) the observations and the lossily quantized version hereof, and finally (iii) between quantized input and the final output of the processing unit. The main novel contribution of our work is to find the optimum information preserving quantizer on one hand, and the optimum information fusion rule on the other. The latter problem leads to a generalized Knapsack problem, which is hard to solve in general. We provide a greedy heuristic by minimizing the decrease of entropy in each iteration step. Amazingly, the resulting algorithm may be organized like the Huffman tree for finding binary codes. Optimum quantizers for absolutely-continuous random variables are demonstrated to generate a discrete uniform distribution as output. Finally, the data processing inequality allows us to interpret reliability and performance of data transmission in a stochastic pooling channel in accordance with what is observed for biological systems. We commence with preparing some information theoretic prerequisites in the next section.

II. Mathematical Prerequisites

To clarify notation we briefly introduce mutual information between two random variables X and Z as

$$I(X;Z) = H(X) - H(X \mid Z) = H(Z) - H(Z \mid X).$$

$H(X)$ denotes the entropy of random variable X and $H(X \mid Z)$ the conditional entropy of X given Z, cf. [6]. Mutual information can be interpreted as the reduction of uncertainty about X when Z is given, or equivalently as the amount of information about X provided by Z.

Let Y be a discrete random variable with values in some finite set $\mathcal{Y} = \{y_1, \dots, y_N\}$, the support, and distribution $\boldsymbol{p} = (p_1, p_2, \dots, p_N)$ with $p_i = P(Y = y_i)$, $i = 1, \dots, N$. We synonymously write $H(Y) = H(p_1, \dots, p_N)$.

Let $\mathcal{U}$ denote the set of functions

$$u : \mathcal{Y} \to \{0, 1, \dots, m-1\}.$$

Each function u of this type is uniquely characterized by the partitioning $\mathcal{Y}_0, \mathcal{Y}_1, \dots, \mathcal{Y}_{m-1}$ of $\mathcal{Y}$ where

$$\mathcal{Y}_j = u^{-1}(\{j\}) = \{y \in \mathcal{Y} \mid u(y) = j\},$$

u^{-1} denoting the preimage. Vice versa, each partitioning $\mathcal{Y}_0, \mathcal{Y}_1, \dots, \mathcal{Y}_{m-1}$ of $\mathcal{Y}$ defines some function u via

$$u(y) = j \text{ for all } y \in \mathcal{Y}_j, \quad j = 0, \dots, m-1.$$

For a given partitioning $\mathcal{Y}_0, \mathcal{Y}_1, \dots, \mathcal{Y}_{m-1}$ of $\mathcal{Y}$ define

$$q_j = \sum_{y_k \in \mathcal{Y}_j} p_k, \quad j = 0, \dots, m-1.$$

We next characterize functions $u \in \mathcal{U}$ which maximize the mutual information $I\big(Y; u(Y)\big)$.

Theorem 1: Let Y be a discrete random variable with support $\mathcal{Y} = \{y_1, \dots, y_N\}$ and distribution $\boldsymbol{p} = (p_1, p_2, \dots, p_N)$. Then

$$\max_{u \in \mathcal{U}} I\big(Y; u(Y)\big) \tag{1}$$

is attained at function u^* if the corresponding partitioning $\mathcal{Y}_0^*, \mathcal{Y}_1^*, \dots, \mathcal{Y}_{m-1}^*$ solves

$$\max_{\mathcal{Y}_0, \dots, \mathcal{Y}_{m-1} \subseteq \mathcal{Y}} H\big(q_0, \dots, q_{m-1}\big)$$

over all partitionings $\mathcal{Y}_0, \dots, \mathcal{Y}_{m-1} \subseteq \mathcal{Y}$.

Proof. We first write

$$I\big(Y; u(Y)\big) = H\big(u(Y)\big) - H\big(u(Y) \mid Y\big) = H(q_0, \dots, q_{m-1}),$$

since $H\big(u(Y) \mid Y\big) = 0$ and random variable $u(Y)$ is governed by distribution $(q_0, \dots, q_{m-1})$. Maximizing the left hand side over all functions $u \in \mathcal{U}$ hence means to seek a partitioning $\mathcal{Y}_0, \dots, \mathcal{Y}_{m-1}$ such that the corresponding values $q_0, \dots, q_{m-1}$ maximize the entropy on the right. ∎

¿From Theorem 1 we conclude that the solution of (1) is given by the solution of the following 0-1 integer optimization problem

$$\begin{aligned} &\max\ H(q_0, \dots, q_{m-1}) \\ &\text{such that } \ q_i = \sum_{j=1}^{N} p_j x_{ij},\ i = 0, \dots, m-1 \\ &\qquad \sum_{i=0}^{m-1} x_{ij} = 1,\ j = 1, \dots, N \\ &\qquad x_{ij} \in \{0, 1\}. \end{aligned}$$

A. Special case: a knapsack problem

Solving (1) for the binary case $m = 2$ leads to a so called *knapsack problem*. In general, the knapsack problem may be written as a 0-1 integer linear program as follows. Assume there are N items of weight w_i and value v_i each, $i = 1, \dots, N$. The objective is to collect items such that the total value is maximized and a given weight threshold W is not exceeded, cf. [7],

$$\max \sum_{i=1}^{N} v_i x_i \ \text{ such that } \sum_{i=1}^{N} w_i x_i \le W,\ x_i \in \{0, 1\}. \tag{2}$$

Since entropy is a Schur convex function, see [8], the maximum in (1) is attained at (q_0, q_1) whenever $|q_1 - q_0|$ is minimum. A solution is given by a solution of knapsack problem (2) with $w_i = v_i = p_i$, $i = 1, \dots, N$ and $W = 1/2$.

$$\max \sum_{i=1}^{N} p_i x_i \ \text{ such that } \sum_{i=1}^{N} p_i x_i \le \frac{1}{2},\ x_i \in \{0, 1\}, \tag{3}$$

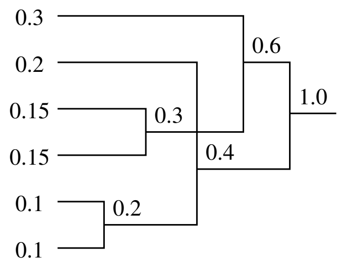

Fig. 1. Successive greedy partitioning by a hierarchical tree.

setting $q_0 = \sum_{i=1}^{N} p_i x_i$ and $q_1 = 1 - q_0$.

The case of a uniform distribution is particularly important for modeling many sensor systems where each sensor and subchannel carries the same characteristics, see Section III. In the case of a uniform distribution and dichotomous information fusion, i.e., $p_i = 1/N$, $i = 1, \dots, N$ and $m = 2$, the solution follows easily from (3) as

$$q_0 = \lfloor N/2 \rfloor / N, \quad q_1 = 1 - q_0,$$

with $\lfloor N/2 \rfloor$ elements contained in $\mathcal{Y}_0$ and $\lceil N/2 \rceil$ elements contained in $\mathcal{Y}_1$.

In general, the knapsack problem is NP hard. However, there is a dynamic programming approach for the 0-1 knapsack problem (3) that runs in pseudo-polynomial time, see [7]. We proceed by developing a heuristic which aims at minimizing the decrease in each pooling step.

B. A greedy heuristic

The heuristic presented in this section relies on the following postulate for entropy. Denote by $q_{N-1} = p_{N-1} + p_N$ for some stochastic vector $(p_1, \dots, p_{N-1}, p_N)$. It holds that

$$\begin{aligned} &H(p_1, \dots, p_{N-2}, q_{N-1}) \\ &= H(p_1, \dots, p_N) - q_{N-1} H(p_{N-1}/q_{N-1}, p_N/q_{N-1}). \end{aligned} \tag{4}$$

The basic idea of the algorithm consists of successively forming subsets by adding the two smallest probabilities until exactly m remaining probabilities and subsets are left. In each step entropy is decreased, as can be easily seen from equation (4). However agglomerating the smallest probabilities results in the least reduction of entropy, which is demonstrated by the following theorem. The algorithm hence utilizes the greedy principle of minimizing the decrease of entropy in each iteration step.

For a stochastic vector $\boldsymbol{p} = (p_1, \dots, p_N)$ denote by $\boldsymbol{p}^{[k,l]}$, $1 \le k \ne l \le N$, the stochastic vector of dimension $N-1$ evolving from $\boldsymbol{p}$ by adding components k and l, explicitly for $k < l$,

$$\boldsymbol{p}^{[k,l]} = (p_1, \dots, p_k + p_l, \dots, p_{l-1}, p_{l+1}, \dots, p_N).$$

Theorem 2: For any stochastic vector $\boldsymbol{p} = (p_1, \dots, p_N)$ with $p_1 \ge \dots \ge p_{N-1} \ge p_N$ it holds that

$$H(\boldsymbol{p}^{[k,l]}) \le H(\boldsymbol{p}^{[N-1,N]}) \text{ for all } 1 \le k, l \le N, \tag{5}$$

stating that entropy is largest if the two smallest probabilities are added.

Proof. We first fix $k \in \{1, \dots, N\}$, let $1 \le r \le l \le N$ and consider $\boldsymbol{p}^{[k,l]}$. ¿From (4) it follows that

$$H(\boldsymbol{p}^{[k,l]}) = H(\boldsymbol{p}) - (p_k + p_l) H\Big(\frac{p_k}{p_k + p_l}, \frac{p_l}{p_k + p_l}\Big).$$

Hence,

$$\begin{aligned} &H(\boldsymbol{p}^{[k,l]}) - H(\boldsymbol{p}^{[k,r]}) \\ &= (p_k + p_r) H\Big(\frac{p_k}{p_k + p_r}, \frac{p_r}{p_k + p_r}\Big) \\ &\qquad - (p_k + p_l) H\Big(\frac{p_k}{p_k + p_l}, \frac{p_l}{p_k + p_l}\Big) \\ &= -p_r \log p_r + (p_k + p_r) \log(p_k + p_r) \\ &\qquad + p_l \log p_l - (p_k + p_l) \log(p_k + p_l) \\ &\ge 0, \end{aligned}$$

since $g(p) = p \log p - (p_k + p) \log(p_k + p)$ is a monotonically decreasing function of $p \in (0, 1)$, as may be seen from its first derivative.

For any pair of k, l it now follows from the above that

$$\begin{aligned} H(\boldsymbol{p}^{[k,l]}) &\ge H(\boldsymbol{p}^{[k,N]}) = H(\boldsymbol{p}^{[N,k]}) \\ &\ge H(\boldsymbol{p}^{[N,N-1]}) = H(\boldsymbol{p}^{[N-1,N]}) \end{aligned}$$

which completes the proof. ∎

Theorem 2 suggests a greedy algorithm for determining a partitioning into two subsets ($m = 2$).

> *Successively merge the two smallest probabilities until two probabilities are left. The indices which correspond to the addends of each determine the final partitioning.* (6)

In a greedy fashion, in each step the decrease of entropy is minimized, aiming at retaining maximum entropy in the final partitioning.

The algorithm may be organized analogously to the Huffman code tree. Figure 1 demonstrates the principle. This example also shows that the global optimum may not be achieved. The final partition into two subset by the greedy algorithm is $\mathcal{Y}_0 = \{1, 3, 4\}$ and $\mathcal{Y}_1 = \{2, 5, 6\}$ with probabilities $q_0 = 0.6$ and $q_1 = 0.4$. The optimum partitioning however is $\mathcal{Y}'_0 = \{1, 2\}$ and $\mathcal{Y}'_1 = \{3, 4, 5, 6\}$ with probabilities $q'_0 = q'_1 = 0.5$.

One might conjecture that the greedy algorithm yields the optimum partitioning in case of a uniform distribution $p_0 = p_1 = \dots = p_{N-1}$. Obviously this holds true if $N = 2^k$ for some integer k. However, the maximum mutual information is not attained in general. The smallest case where it fails to find the maximum for uniformly distributed input is $N = 6$. The partitioning found is $\mathcal{Y}_0 = \{y_1, y_2, y_3, y_4\}$ and $\mathcal{Y}_1 = \{y_5, y_6\}$, while the optimum is attained at $\mathcal{Y}^*_0 = \{y_1, y_2, y_3\}$ and $\mathcal{Y}^*_1 = \{y_4, y_5, y_6\}$ with $q^*_0 = q^*_1 = 0.5$.

If the probabilities are of the form $p_i = 2^{-k_i}$, $k_i \in \mathbb{N}$, $i = 1, \dots, N$, and $m = 2$, greedy heuristic (6) will find the partitioning which maximizes mutual information (1). In order to prove this, let $m = 2$ and

$$p_i = 2^{-k_i}, \; k_1 \le \dots \le k_N \in \mathbb{N} \tag{7}$$

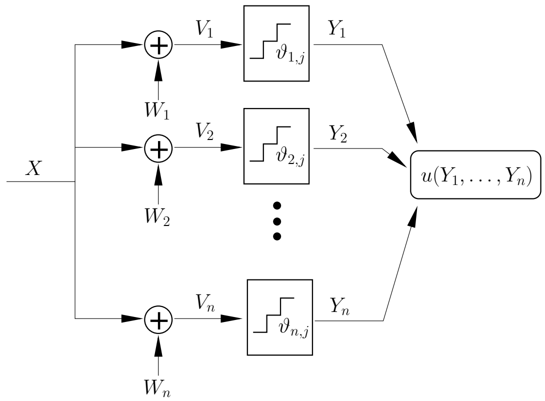

Fig. 2. Condensing Information from a many-sensor system.

such that $p_1 \geq \cdots \geq p_N$.

Proposition 3: If (7) holds, then necessarily $p_{N-1} = p_N$.

Proof. We assume on contrary that $k_{N-1} < k_N$ and use the representation $k_i = k_{N-1} - r_i$ for certain integers r_i. Then $2^{-k_i} = 2^{r_i}\, 2^{-k_{N-1}}$, $i = 1, \dots, N-1$, and

$$1 = \sum_{i=1}^{N} 2^{-k_i} = \Big(\sum_{i=1}^{N-1} 2^{r_i}\Big) 2^{-k_{N-1}} + 2^{-k_N}.$$

Abbreviating $s_{N-1} = \sum_{i=1}^{N-1} 2^{r_i}$ yields

$$2^{-k_N} = 1 - s_{N-1} 2^{-k_{N-1}} \geq 2^{-k_{N-1}},$$

since the term in the middle is positive. Hence $k_N \leq k_{N-1}$, a contradiction. It follows that $k_N = k_{N-1}$. ∎

Now, if the probabilities are of the form $p_i = 2^{-k_i}$, then by Proposition 3 algorithm (6) will terminate with two probabilities which are equal to $\frac{1}{2}$, corresponding to the maximum achievable entropy. This is because in each iteration step of (6), by addition of the smallest equal probabilities a stochastic vector is generated with dyadic entries of the type $2^{-\ell_i}$.

III. Modeling Many-Sensor Systems

A stochastic signal X is observed by many sensors labeled $1, 2, \dots, n$. Each observation is subject to additive random noise $W_1, W_2, \dots, W_n$, which may be due to random haziness of the sensor or disturbance of the information perceived by the sensor. Each sensor i applies m-ary quantization with thresholds $\vartheta_{i,1} \leq \cdots \leq \vartheta_{i,m-1}$, $i = 1, \dots, n$, to yield m-ary output random variables $Y_1, Y_2, \dots, Y_n$. The purpose of function

$$u : \{0, \dots, m-1\}^n \to \{0, 1, \dots, m-1\}$$

is to condense the constituent information about the jointly observed signal X into a single decision $U = u(Y_1, \dots, Y_n)$. For $m = 2$ the decision will be dichotomous, but we also allow for discrimination into m different classes. The basic model is depicted in Fig. 2. The joint distribution of the random variables X and $W_1, \dots, W_n$ will be specified later.

In this model, information is conveyed in three steps, from X onto the noisy versions $V_1, \dots, V_n$, quantization from $V_1, \dots, V_n$ to $Y_1, \dots, Y_n$ and finally information condensation from $Y_1, \dots, Y_n$ into $U = u(Y_1, \dots, Y_n)$. The corresponding mutual information $I\big(X; (V_1, \dots, V_n)\big)$, $I\big(V_i; Y_i\big)$, $i = 1, \dots, n$, and $I\big((Y_1, \dots, Y_n); U\big)$ will be considered in the following. We use boldface characters to denote random vectors $\boldsymbol{V} = (V_1, \dots, V_n)$ and $\boldsymbol{Y} = (Y_1, \dots, Y_n)$.

A. The n-look channel

This type of channel is well investigated for the case of Gaussian input X with variance σ_X^2 and independent identically distributed Gaussian noise $W_1, \dots, W_n$ with variance σ_W^2, see [3] and [6]. Direct application of the formula for jointly n-dimensional distributed Gaussian random variables yields

$$I\big(X, \boldsymbol{V}\big) = \frac{1}{2} \log \Big(1 + n\, \frac{\sigma_X^2}{\sigma_W^2}\Big).$$

B. The best quantizer

The key problem here is to determine the threshold values $\vartheta_1 \leq \vartheta_2 \leq \cdots \leq \vartheta_{m-1}$. We omit index i and describe the quantizer for each sensor by some quantizing function q with image $\{0, 1, \dots, m-1\}$ and threshold values $\vartheta_1, \dots, \vartheta_{m-1}$. Hence,

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

with $\vartheta_0 = -\infty$ and $\vartheta_m = \infty$. Let $Y = q(V)$. The problem now is to determine $\vartheta_1, \dots, \vartheta_m$ as to

$$\text{maximize } I(V; Y) = I\big(V, q(V)\big).$$

over all quantizing functions q.

Similar to the proof of Theorem 1 it follows that

$$\begin{aligned} I(V; Y) &= H(Y) - H(Y \mid V) \\ &= H\big(q(V)\big) = H(q_0, \dots, q_{m-1}), \end{aligned}$$

with $q_k = P(Y = k)$, $k = 0, \dots, m-1$. If V has density $f(v)$ then the maximum $I(V; Y)$ is attained for a uniform distribution. Hence the optimum threshold values ϑ_k are such that

$$q_k = \int_{\vartheta_k}^{\vartheta_{k+1}} f(v) dv = P(Y = k) = \frac{1}{m}$$

for all $k = 0, \dots, m-1$, where $\vartheta_0 = -\infty$ and $\vartheta_m = \infty$. Mutual information for the optimum quantizer q has the value $I(V; Y) = \log m$.

The same information-theoretic principle for optimum quantization is applied to coded modulation systems in [9].

C. Optimum information fusion

This is the most difficult step, it concerns the selection of a decision function u and needs the preparatory work from Section II. The authors [3], e.g., employ $u(y_1, \dots, y_n) = \sum_{i=1}^{n} y_i$ to compress information into a single value in $\{0, 1, \dots, n\}$. This choice may not be optimum as will be shown later. The general objective is to obtain information about the outcome of X most reliably.

Under the above model, $\boldsymbol{Y} = (Y_1, \dots, Y_n)$ is a discrete random vector with values in $\{0, 1, \dots, m-1\}^n$. Assume its distribution is described by the stochastic vector $\boldsymbol{p} = (p_1, p_2, \dots, p_N)$ with $N = m^n$. We seek to maximize mutual information between $\boldsymbol{Y}$ and $u(\boldsymbol{Y})$ over all functions u which map $\boldsymbol{Y}$ onto a set of decisions $\{0, 1, \dots, m-1\}$.

The optimum u is characterized by Theorem 1. By addition, $p_1, \dots, p_N$ are grouped into probabilities $q_0, \dots, q_{m-1}$ such that $H(q_0, \dots, q_{m-1})$ is maximized. As mentioned before, this leads to a generalized knapsack problem whose solution is hard to achieve in general. For $m = 2$ greedy heuristic (6) can be used to reach a near optimal solution.

D. Successive data processing

Using the data processing inequality (see [6]) mutual information $I(X; U)$ can be upper bounded by

$$I(X; U) \le \min\big\{I(X; \boldsymbol{V}), I(\boldsymbol{V}; \boldsymbol{Y}), I(\boldsymbol{Y}; U)\big\}. \tag{8}$$

This holds since $U = u\big(\boldsymbol{q}(\boldsymbol{V})\big)$ is a function of $\boldsymbol{V}$ such that $X \to \boldsymbol{V} \to U$ forms a Markov chain, see [6], p. 32. Here, $\boldsymbol{q}$ denotes the function which componentwise applies some quantizer q as described in Section III-B. Hence

$$I(X; U) \le \min\{I(X; \boldsymbol{V}), I(\boldsymbol{V}; U)\}$$

holds. Furthermore, $\boldsymbol{V} \to \boldsymbol{Y} \to U$ forms a Markov chain such that

$$I(\boldsymbol{V}; U) \le \min\{I(\boldsymbol{V}; \boldsymbol{Y}), I(\boldsymbol{Y}; U)\},$$

entailing inequality (8).

We now consider the case of a Gaussian input X with variance σ_X^2 and i.i.d. Gaussian noise variables $W_1, \dots, W_n$ with variance σ_W^2. The result is

$$I(X; \boldsymbol{V}) = \frac{1}{2} \log\Big(1 + n\,\frac{\sigma_X^2}{\sigma_W^2}\Big).$$

We furthermore assume that the total channel is designed in an optimal way in the sense that optimum quantizers and optimum information fusion are applied. In this case $I(V_i; Y_i) = \log m$ holds for all $i = 0, \dots, m-1$. Iteratively applying the chain rule for information (see [6]) yields

$$\log m = I(V_i; Y_i) \le I(\boldsymbol{V}; \boldsymbol{Y}) \le n \log m.$$

Finally, optimum information fusion means to group the discrete uniform distribution with probabilities m^{-n} into m probabilities m^{-1}. This is obviously possible yielding $I(\boldsymbol{Y}, U) = \log m$, where the right hand side is the trivial bound by the logarithm of the support cardinality.

In summary, by (8)

$$I(X, U) \le \min\Big\{\log m,\ \frac{1}{2}\,\log\Big(1 + n\,\frac{\sigma_X^2}{\sigma_W^2}\Big)\Big\} \tag{9}$$

follows. Since optimum quantization and information fusion in the sense of maximum mutual information is assumed, any other type of quantization and fusion is upper bounded by (9).

The upper bound in equation (9) as a function of n, $B(n)$, say, gives some interesting insight into the resilience of the many-sensor channel with information combining for Gaussian input and Gaussian noise. Obviously mutual information is bounded by the number of bits to represent the discrete image U of input X. This refers to the first term, $\log m$, in the above minimum.

Resilience of the total system is represented by the second term considered as a function of n. This is a logarithmic function, which is concave and increases very slowly. Hence, if $\log m$ is the active bound, or if a large number of subchannels is available, the failure of only a few does not change the performance of the whole system, or deteriorates it only very little. On the other hand, if only a few subchannels are left, the failure of some leads to a drastic degradation of the whole system. This is an interesting effect which is also observed in biological information processing. Observing the typical cause of Alzheimer disease teaches us that in the beginning disrupted information channels have only minor effect on the patients abilities, but passing a certain threshold leads to a drastic loss of cognition, reaction and situational awareness.

IV. Conclusions

The main contributions of this paper are characterizations of optimal quantization and information fusion rules for a bio-inspired parallel channel model, called stochastic pooling network. In the case of discrete input distribution this leads to a knapsack problem, which is hard to solve in general. A greedy heuristic, similar to the Huffman coding tree, is developed. For input distributions with density f the problem is easily solved by generating discrete uniformly distributed output. These findings are used to combine separate observations by noisy sensors into a single optimal m-ary decision.

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

References

[1] S. Zozor, P.-O. Amblard, and C. Duchene, "On pooling networks and fluctuation in suboptimal detection framework," *Fluctuation and Noise Letters*, vol. 7, no. 1, pp. L39–L60, 2007.

[2] M. D. McDonnell, P.-O. Amblard, and N. G. Stocks, "Stochastic pooling networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 1, pp. 1–18, January 2009.

[3] ——, "Bio-inspired communication: Performance limits for information transmission and compression in stochastic pooling networks with binary quantizing nodes," *Journal of Computational and Theoretical Nanoscience*, vol. 7, no. 5, pp. 876–883, May 2010.

[4] K. Wiesenfeld and F. Moss, "Stochastic resonance and the benefits of noise: from ice ages to crayfish and SQUIDs," *Nature*, vol. 373, pp. 33–36, jan 1995.

[5] N. Stocks, "Suprathreshold stochastic resonance in multilevel threshold systems," *Physical Review Letters*, vol. 84, no. 11, pp. 2310–2313, March 2000.

[6] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York, NY, USA: Wiley-Interscience, 1991.

[7] S. Martello and T. Paolo, *Knapsack Problems: Algorithms and Computer Implementations*. New York, NY, USA: Wiley, 1990.

[8] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic Press, 1979.

[9] C. Novak, P. Fertl, and G. Matz, "Quantization for soft-output demodulators in bit-interleaved coded modulation systems," in *International Symposium on Information Theory, ISIT 2009*, Seoul, Korea, 2009, pp. 1070–1074.