A Graph-Based Message Passing Approach for Joint Source-Channel Coding via Information Bottleneck Principle

Shayan Hassanpour, Dirk Wübben, and Armin Dekorsy
Department of Communications Engineering
University of Bremen, 28359 Bremen, Germany
Email: {hassanpour, wuebben, dekorsy}@ant.uni-bremen.de

***Abstract*—In this paper, we focus on an extended version of noisy source coding wherein the compressed data shall be transmitted over an imperfect (forward) channel for further processing. As the quantization design framework, we deploy the Information Bottleneck principle and propose a novel treatment by successful exploitation of a quite generic and highly flexible graph-based clustering routine known as Affinity Propagation. We also provide simulation results regarding a typical digital transmission setup to compare the performance of our proposed treatment with a state-of-the-art routine from literature.**

I. Introduction

The subject of *joint source-channel coding (JSCC)* is considered. Explicitly, we assume the case in which the aim is to quantize an observed signal (e.g., at an access channel's output) from a given source with the side-knowledge that the compressed signal shall be transmitted over a non-ideal forward channel to be further processed at a distant unit. In fact, this is the underlying scenario on a broad range of practical applications, among others, cooperative transmission via relaying over noisy links with *quantize-and-forward* strategy [1], distributed inference sensor networks with imperfect links to the fusion center [2], and Centralized Radio Access Networks (C-RANs) with non-ideal fronthaul connections [3]. On such occasions, the impacts of imperfect forwarding have to be encompassed into the quantizer design setup. A rather straightforward approach to address this problem is to treat the observed signal as a *virtual* source. One can then adapt the conventional techniques from *Rate-Distortion (RD)* theory [4] by expanding the *distortion measure function* such that the effects of imperfect forward channel(s) are taken into account [5], [6]. Basically, in such procedures, the actual source is not explicitly brought into the design setup and there is no general way to *systematically* obtain the appropriate distortion measure for any particular case of interest. These facts are indeed incentives to think of an alternative framework for quantization. As a quite interesting choice, the *Information Bottleneck (IB)* method [7] can be deployed. It was primarily proposed in the context of *machine learning* in which the intended purpose was to extract a certain feature from a typically huge dataset via smart clustering [8]. Applying this type of *dimensionality reduction* is an indispensable task in a wide variety of practical fields which exploit *statistical analysis* to process data [9]. To acquire a general picture about the IB paradigm and a number of pertinent routines, interested readers are referred to [10]–[13].

Inspired by the original IB philosophy, instead of minimizing the average distortion w.r.t. a certain distortion measure, one may think of maximizing the *Mutual Information (MI)* between the source signal and forward channel's output. Contrary to the conventional methods from RD theory, this brings about a *symmetric* design structure in which both *complexity* and *precision* of the resultant outcome are quantified by MI terms. As a result, the quantization task becomes purely statistical and irrelevant to the specific realizations of the variable to be compressed that makes it fundamentally different from the well-established approaches. For a given input statistics, the obtained quantizer maximizes the end-to-end transmission rate that is definitely desired for (almost) all communication schemes. As will be shown later, for such quantizers attaining the *globally* optimal solution through a tractable complexity is rather demanding. Accordingly, one may pragmatically resort to some efficient heuristics that aim at addressing the design problem at least *locally*.

Here, we tackle this task from a new perspective taking advantage of a highly flexible and generic message passing based clustering routine called *Affinity Propagation (AP)* [14]. AP is an efficient tool designed for *exemplar-based* grouping. It is fed by a matrix containing the pairwise *similarities* among the bunch of *articles* to be clustered and yields a high-quality clustering result. To be able to utilize this powerful tool, the case-specific grouping task must be translated into an equivalent exemplar-based clustering problem with clear specification of the appropriate measure of pairwise similarities. In this study, we address this translation in case of IB-based JSCC and point out what exactly are the corresponding articles to be clustered and how to obtain the appropriate measure of pairwise similarities among them.

To that end, we introduce the system model for JSCC along with a state-of-the-art algorithm called *Channel-Optimized Information Bottleneck (Ch-Opt-IB)* [15] in Section II. In Section III we present the basic comprehension of AP and demonstrate how to attain an equivalent exemplar-based clustering problem for IB-based JSCC setup. Finally, we provide performance results in Section IV and a wrap-up containing the salient points in Section V.

978-1-5386-7048-4/18/$31.00 ©2018 IEEE

II. Joint Source-Channel Coding Setup

A. Presumed System Model

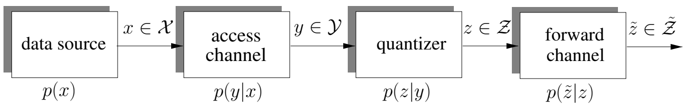

Fig. 1. System model for joint source-channel compression/coding

For the considered system model illustrated in Fig. 1 we presume a discrete memoryless source x (with realizations $x \in \mathcal{X}$) having the a-priori distribution $p(x)$ followed by a Discrete Memoryless Channel (DMC) that is described by the transition probability distribution $p(y|x)$. The observed signal y (with realizations $y \in \mathcal{Y}$) at the access channel's output shall be compressed to the signal z (with realizations $z \in \mathcal{Z}$) before being transmitted over the forward DMC that is characterized via the conditional distribution $p(\tilde{z}|z)$. Moreover, we presume that $\mathsf{x} \leftrightarrow \mathsf{y} \leftrightarrow \mathsf{z} \leftrightarrow \tilde{\mathsf{z}}$ introduces a *Markov chain* and the joint distribution $p(x,y) = p(x)\,p(y|x)$ and also the forward channel probabilities $p(\tilde{z}|z)$ are given. To design a quantizer $p(z|y)$ which maximizes the overall transmission rate in this setup, the following problem has to be addressed[1]:

$$p^{\star}(z|y) = \underset{p(z|y)}{\operatorname{argmax}}\, I(\mathsf{x};\tilde{\mathsf{z}}) \;\; \text{for } |\mathcal{Z}| \le M, \tag{1}$$

in which M is the maximum number of quantization levels and $|\cdot|$ denotes the cardinality of a given set. To acquire an impression about the type of optimization task, we investigate (1) in more details. It can be shown that for a given $p(x)$, the objective function in (1) is convex w.r.t. the conditional distribution $p(\tilde{z}|x)$ [4]. In addition, the relation among $p(\tilde{z}|x)$ and the quantizer mapping $p(z|y)$ is established as

$$p(\tilde{z}|x) = \sum_{y \in \mathcal{Y}} \sum_{z \in \mathcal{Z}} p(\tilde{z}|z) p(z|y) p(y|x), \tag{2}$$

which is of *affine* type preserving convexity. Thus, it is inferred that $I(\mathsf{x};\tilde{\mathsf{z}})$ is also convex w.r.t. the quantizer mapping $p(z|y)$. For each specific receive signal $y \in \mathcal{Y}$ it applies $\sum_{z \in \mathcal{Z}} p(\mathsf{z}{=}z|\mathsf{y}{=}y) = 1$, that defines a $(|\mathcal{Z}|-1)$-dimensional *probability simplex*. Consequently, the overall search space in (1) is obtained by the Cartesian product of $|\mathcal{Y}|$ of such simplices leading to a closed convex polytope in the space of dimensionality $|\mathcal{Y}| \times (|\mathcal{Z}|-1)$. Altogether, the optimization in (1) boils down to maximizing a convex function over a closed and convex set. In optimization theory, this task is referred to as *convex maximization* or *concave optimization* proven to be *NP-hard* in general [16]. Resorting to a well-known proposition which asserts that a convex function attains its *global* maximum over a closed convex set at its *extreme points*, it is immediately deducible that the optimal solution in (1) is obtained by *deterministic* mappings, i.e., $p(z|y) \in \{0,1\}$ for all pairs $(y,z) \in \mathcal{Y} \times \mathcal{Z}$. To see this, one shall note that extreme points of a polytope translate into its vertices and for the search space polytope in (1), each vertex corresponds to the Cartesian product of vertices of its constituent simplices.

Since the naive *brute-force* search over all vertices of the event space in (1) leads to an exponential complexity w.r.t. $|\mathcal{Y}|$, plainly it cannot be considered as a promising strategy to achieve the desired mapping $p(z|y)$ in practice. This, in fact, is the incentive behind the emergence of heuristics aiming at treating (*locally*) the design problem (1) in an efficient fashion. In the following part, we concisely present the Ch-Opt-IB routine as an instance of state-of-the-art techniques to address the design problem in (1). The provided discussion there paves the way towards perceiving the IB-based JSCC as an exemplar-based clustering task. This, indeed, opens up the chance of utilizing AP as a quite novel treatment (see Section III).

B. Channel-Optimized Information Bottleneck (Ch-Opt-IB)

In [15], the authors have considered an analogous setup to the one depicted in Fig. 1 and developed an iterative algorithm that yields a *vector quantizer* that maximizes $I(\underline{\mathsf{x}};\tilde{\mathsf{z}})$, wherein $\underline{\mathsf{x}}$ denotes a vector of length K comprising elements produced by the source x. Here, we restrict ourselves to the *scalar quantizer* design, i.e., $K{=}1$. Following a round of derivations [13] the objective function in (1) can be rewritten as

$$I(\mathsf{x};\tilde{\mathsf{z}}) = I(\mathsf{x};\mathsf{y}) - I(\mathsf{x};\mathsf{y}|\tilde{\mathsf{z}}) \;. \tag{3}$$

As $I(\mathsf{x};\mathsf{y})$ is fixed, $I(\mathsf{x};\mathsf{y}|\tilde{\mathsf{z}})$ has to be minimized. Introducing[2] $C(\mathsf{y}=y,\tilde{\mathsf{z}}=\tilde{z}) \triangleq D_{\mathrm{KL}}\big(p(x|y)\|p(x|\tilde{z})\big)$, one can write

$$I(\mathsf{x};\mathsf{y}|\tilde{\mathsf{z}}) = \mathbb{E}_{\mathsf{y}}\big\{\mathbb{E}_{\tilde{\mathsf{z}}}\{C(\mathsf{y},\tilde{\mathsf{z}})|\mathsf{y}\}\big\}, \tag{4}$$

in which the conditional expectation term is calculated by

$$\mathbb{E}_{\tilde{\mathsf{z}}}\{C(\mathsf{y},\tilde{\mathsf{z}})|\mathsf{y}{=}y\} = \sum_{z \in \mathcal{Z}} p(z|y) \sum_{\tilde{z} \in \tilde{\mathcal{Z}}} p(\tilde{z}|z) C(\mathsf{y}{=}y,\tilde{\mathsf{z}}{=}\tilde{z}) \;. \tag{5}$$

As the inner sum term in (5) is constant for a certain $z \in \mathcal{Z}$, to minimize the conditional expectation (5) for every $y \in \mathcal{Y}$, the quantizer mapping has to be chosen as $p(z|y) = \delta_{z,z^{\star}(y)}$, wherein the optimum cluster is

$$z^{\star}(y) = \underset{z}{\operatorname{argmin}} \sum_{\tilde{z} \in \tilde{\mathcal{Z}}} p(\tilde{z}|z) C(\mathsf{y}=y,\tilde{\mathsf{z}}=\tilde{z}) \;. \tag{6}$$

Accordingly, (4) is minimized for a given $C(\mathsf{y},\tilde{\mathsf{z}})$. Besides, it is inferable that $p(\tilde{z}|y) = \sum_{z \in \mathcal{Z}} p(\tilde{z}|z)p(z|y) = p(\tilde{z}|z^{\star}(y))$. To develop an iterative routine which maximizes $I(\mathsf{x};\tilde{\mathsf{z}})$, one shall update $C(\mathsf{y},\tilde{\mathsf{z}})$ at the end of each iteration. This is done via updating $p(x|\tilde{z})$ by

$$p(x|\tilde{z}) = \frac{\sum\limits_{y \in \mathcal{Y}_z} p(x,y)p(\tilde{z}|y)}{\sum\limits_{y \in \mathcal{Y}_z} p(\tilde{z}|y)p(y)}, \tag{7}$$

wherein $\mathcal{Y}_z$ denotes the subset of $\mathcal{Y}$ for which all elements are allotted to the cluster z. Specifically, Ch-Opt-IB is initialized by a *random* choice of $C(\mathsf{y},\tilde{\mathsf{z}})$ and then iterates over the resultant mapping by (6) (*assignment* phase) and the recalculated version of $C(\mathsf{y},\tilde{\mathsf{z}})$ acquired from (7) (*update* phase) till converging to a *local* optimum.

[1]The MI between discrete random variables a and b with marginal and joint distributions $p(a)$, $p(b)$ and $p(a,b)$, respectively is defined as $I(\mathsf{a};\mathsf{b}) \triangleq \sum\limits_{a}\sum\limits_{b} p(a,b)\log\frac{p(a,b)}{p(a)p(b)}$.

[2]$D_{\mathrm{KL}}(\cdot\,\|\,\cdot)$ is *Kullback-Leibler* (KL) divergence that is defined for probability distributions $p(a)$ and $q(a)$ over the same event space $\mathcal{A}$ of the random variable a as $D_{\mathrm{KL}}\big(p(a)\|q(a)\big) \triangleq \sum_{a \in \mathcal{A}} p(a)\log\frac{p(a)}{q(a)}$ [4]. The relation among MI and KL divergence is established via $I(\mathsf{a};\mathsf{b}) = D_{\mathrm{KL}}\big(p(a,b)\|p(a)p(b)\big)$.

III. Affinity Propagation (AP)

A. Description and Discussion

Principally, AP is a recursive message passing algorithm devised for exemplar-based grouping [14]. An *exemplar* is the representative of a certain cluster which is chosen from the primary set of *articles* to be clustered, i.e., the *mother-set*. AP treats all primary articles as nodes of a network being delineated with a *fully-connected* graph in which there is an edge between every pair of articles. These articles communicate with each other over the edges in a recursive fashion (bilateral messages per edge) to gradually decide about the most suitable *exemplar-set*. The *responsibility* $r(i,k)$ is the transmitted message from *article* i to the *potential exemplar* k indicating to which extent i conceives k to be responsible for serving it as an exemplar. Conversely, the *availability* $a(i,k)$ is the transmitted message from the *potential exemplar* k towards an *article* i signifying to which extent k conceives itself to be available as an exemplar for i. At initial stage, all N different articles are presumed to be potential exemplars. Therefore, every article i must engender a responsibility message $r(i,k)$ for the article k with $i,k \in \{1,2,...,N\}$. The primary round of responsibilities are solely generated from the pairwise *similarities* $s(i,j)$ for $i,j \in \{1,2,...,N\}$, inserted as the input of AP. A fascinating feature of AP which brings about a great deal of flexibility is the fact that the *mother-set* does not have to be embedded in a *metric* or *continuous* or even *ordinal* space. Thus, the pairwise similarities do not have to be calculated based on a metric, i.e., they do not need to be *symmetric* and even they are not required to satisfy the *triangle inequality*. On the next phase of AP dynamism the availabilities $a(i,k)$ are engendered from the received responsibilities $r(i,k)$. The aforementioned bilateral message passing procedure is perpetuated until a highly satisfactory set of exemplars and subsequently the respective clusters emerge.

Another intriguing feature of AP is about its astonishingly intuitive and simple update rules. Specifically, the update equation for *responsibility* calculations is given as

$$r(i,k) = s(i,k) - \max_{j \neq k} \{a(i,j) + s(i,j)\}, \quad (8)$$

and the respective update rules for *availabilities* are

$$a(i,k) = \min\Big\{0, r(k,k) + \sum_{j \notin \{i,k\}} \max\{0, r(j,k)\}\Big\} \text{ for } i \neq k \quad (9)$$

and in case of *self-availability*

$$a(k,k) = \sum_{j \neq k} \max\{0, r(j,k)\} \ . \quad (10)$$

The presented update equations are derived by exploitation of *Max-Product* algorithm [17] for approximating the marginals of the *global function* pertinent to the specific *factor graph* illustrated in Fig. 2. The *factor nodes* f_i and the *variable nodes* c_i with $i \in \{1,...,N\}$ represent the *coherency check* and the *chosen exemplar* for the article i, respectively. By coherency check f_i, it is stipulated that if any other article $j \neq i$ decides in favor of the article i as its exemplar, then the article i must be the chosen exemplar for itself as well. Regarding the *singleton* factor nodes in Fig. 2, it is rather straightforward to

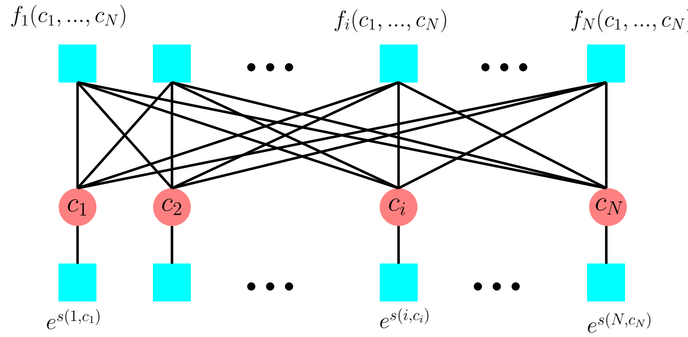

Fig. 2. The factor graph on which AP is developed, the *factor nodes* f_i and the *variable nodes* c_i with $i \in \{1,...,N\}$ represent the *coherency check* and the *chosen exemplar* for the article i, respectively

perceive that the purpose of computation over such a factor graph is to figure out the particular *coherent configuration* of variables that maximizes the sum of overall similarities among all the articles and their respective exemplars. The immediate use of *Max-Product* algorithm requires *vector-valued* messages. Nevertheless, applying some clever mathematical tricks as stated in [18], the update equations are shrunk to the *scalar-valued* versions (8)-(10) which leads to the complexity of $O(N^2)$ per iteration. At initial stage of AP, to treat all the articles equitably, i.e., providing equal chance of being chosen as an exemplar, all the availabilities $a(i,j)$ in (8) are set to zero. Nonetheless, after a while if an article gets allotted to another one, its availability becomes negative, which immediately influences the pertaining *effective similarity* (the second term) in (8), pushing it out of the ongoing exemplarship rivalry. Concentrating on the availability update (9) for the article k, it is principally calculated as the sum of all the positive responsibilities from the other articles and its *self-responsibility* $r(k,k)$. The negative responsibilities are disregarded as those belong to the articles for which the article k is not a proper exemplar and for a decent exemplar, it suffices to represent some and not all of the articles effectively.

Another compelling characteristic of AP is about its so-called *automatic model selection* capability [18]. One shall note that the cardinality of the *exemplar-set* is not fixed a-priori but it rather emerges naturally for every certain choice of the common *self-similarity*, which in AP terminology is designated as the *Common Preference (CP)*. Generally, the preference values are not required to be identical for all the articles and the larger the preference $s(i,i)$, the higher the chance of article i to be selected as an exemplar. Accordingly, when a CP is present, it is essentially the very parameter that can be tuned to affect the resultant *exemplar-set* cardinality. The underlying reason is the fact that the sum of preferences can be reckoned a penalty term (w.r.t. the complexity) being existent at the objective function of AP factor graph to avoid, e.g., the case wherein every article is treated as an exemplar.

It is a common observation that applying *belief propagation* over loopy graphs may cause an unstable behavior. Thus, as a crucial implementation detail, the messages which are going to be transmitted over the *fully-connected* AP graph of Fig. 2 must be dampened. Therefore, irrespective of being either responsibility or availability, the messages are calculated as the weighted combination of their current and previous values. Eventually, it must be mentioned that the allotted cluster for article i at the end of each round of message exchange is estimated as $\hat{c}_i = \text{argmax}_k \left\{a(i,k) + r(i,k)\right\}$.

B. IB-Based JSCC Utilizing AP

In this section, we establish the connection among the IB-based JSCC problem in (1) and the generic exemplar-based grouping task behind AP. Accordingly, we propose exploiting AP as a novel treatment regarding the quantization design problem at hand. Commencing with (3) and having (4) and (5) in mind, it is immediately inferred that maximization of the end-to-end transmission rate, $I(\mathsf{x};\tilde{\mathsf{z}})$, is equivalent to maximization of the average term

$$\begin{aligned}&\mathbb{E}_{\mathsf{y}}\{\mathbb{E}_{\tilde{\mathsf{z}}}\{-C(\mathsf{y},\tilde{\mathsf{z}})|\mathsf{y}\}\} \qquad (11)\\ &= \sum_{y\in\mathcal{Y}} - p(y)\sum_{z\in\mathcal{Z}} p(z|y)\sum_{\tilde{z}\in\tilde{\mathcal{Z}}} p(\tilde{z}|z)\cdot D_{\text{KL}}\big(p(x|y)\|p(x|\tilde{z})\big)\ .\end{aligned}$$

As already discussed, the *hard* mapping is favorable for the present *convex maximization* problem. Hence, presuming a *deterministic* quantization and denoting the selected z for each specific y as $z^\star(y)$, (11) can be rewritten as

$$\sum_{y\in\mathcal{Y}} - p(y)\sum_{\tilde{z}\in\tilde{\mathcal{Z}}} p\big(\tilde{z}|z^\star(y)\big)\cdot D_{\text{KL}}\big(p(x|y)\|p(x|\tilde{z})\big)\ . \qquad (12)$$

The structure of (12) is reminiscent of the global objective function of AP, $\sum_i s(i,c_i)$, that is the overall sum of similarities among individual articles i and their respective exemplars c_i. In principle, by regarding the realizations $y^{(i)}$ and $y^{(j)}$ of the observed random variable y in Fig. 1 as *articles* i and j, respectively, and by defining the *pairwise similarities*

$$s(i,j) = -p(i)\cdot\sum_{\tilde{z}\in\tilde{\mathcal{Z}}} p(\tilde{z}|j) D_{\text{KL}}\big(p(x|i)\|p(x|\tilde{z})\big), \qquad (13)$$

the global objective function of AP coincides with (12). Since the goal of AP is to find out the coherent configuration which maximizes its objective function, it is totally in line with the IB-based JSCC design criterion that boils down to maximizing the expectation term in (11). As already mentioned, at initial stage of AP all articles are treated as potential exemplars. This means, primarily, it is assumed that $\mathcal{Z} = \mathcal{Y}$. Therefore, to be able to compute all the pairwise similarities as suggested in (13), the transition probabilities of the *forward channel* must be available for all realizations of the observed variable y. This requirement yields a fundamental difference between the proposed AP-based treatment and the other state-of-the-art approaches. Contrary to other methods, the AP-based solution starts with an *augmented* forward channel matrix and after quantization shrinks that matrix to the one solely containing the forward transition probabilities of the chosen exemplars. This, indeed, can be interpreted as an *automatic forward channel selection*. To obtain the conditional probability $p(x|\tilde{z})$ which is required for computation of pairwise similarities in (13), applying Bayes' rule, one can write

$$p(x|\tilde{z}) = \frac{p(x)p(\tilde{z}|x)}{\sum\limits_{x'\in\mathcal{X}} p(x')p(\tilde{z}|x')}\ . \qquad (14)$$

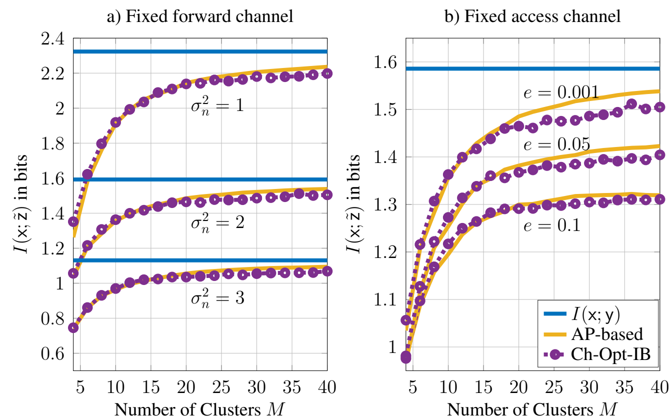

Fig. 3. Overall transmission rate $I(\mathsf{x};\tilde{\mathsf{z}})$ vs. number of clusters M, 16-QAM signaling: a) fixed forward channel with reliability $e = 0.001$, AWGN access channel with $\sigma_n^2 = 1,2,3$ and b) fixed AWGN access channel with $\sigma_n^2 = 2$, forward channel with reliabilities $e = 0.1, 0.05, 0.001$

Since $p(x)$ is known, only the conditional probability $p(\tilde{z}|x)$ must be calculated. Employing the Markovian property

$$p(\tilde{z}|x) = \sum_{y\in\mathcal{Y}}\sum_{z\in\mathcal{Z}} p(\tilde{z}|z)p(z|y)p(y|x)\ . \qquad (15)$$

Primarily, $\mathcal{Z}=\mathcal{Y}$, and therefore one can rewrite (15) as

$$p(\tilde{z}|x)=\sum_{y\in\mathcal{Y}}\sum_{y'\in\mathcal{Y}} p(\tilde{z}|y')p(y'|y)p(y|x) = \sum_{y\in\mathcal{Y}} p(\tilde{z}|y)p(y|x) \qquad (16)$$

by exploiting the fact that $p(y'|y)=1$ only for $y'=y$ and *zero* otherwise. Eventually, it can be seen that by knowing the *access* and *forward* channel transition probabilities, one can calculate the pairwise similarities and let AP decide in favor of the coherent configuration which maximizes the end-to-end transmission rate $I(\mathsf{x};\tilde{\mathsf{z}})$ for a given input statistics $p(x)$.

IV. Simulation Results

In this section, we investigate the performance of a typical digital transmission setup for equiprobable 16-QAM signals with $\sigma_x^2 = 10$. To simulate the *access* channel, we generate $N = 200$ samples from an AWGN channel with different noise variances ($\sigma_n^2 = 1,2,3$). As the *forward* channel, in case of AP-based approach we consider an N-ary symmetric model being characterized by the reliability parameter e (for each symbol the correct reception occurs with probability $1-e$ and the erroneous reception to every other symbol occurs with probability $\frac{e}{N-1}$). For a fair comparison, in case of Ch-Opt-IB we assume an $M\times N$ model with the same reliability. As performance indicator, we calculate the resultant end-to-end MI, $I(\mathsf{x};\tilde{\mathsf{z}})$, over the varying maximum number of output bins for two distinct scenarios to investigate the effects of both DMCs in our presumed system model. Firstly, we keep the forward channel constant (with reliability parameter $e = 0.001$) and vary the noise variance of the access channel. Secondly, we fix the access channel (with the noise variance $\sigma_n^2 = 2$) and vary the reliability parameter e of the forward channel. The pertinent plots are depicted in Figs. 3. To obtain these curves, the *bisection method* [18] is exploited for the AP-based treatment to determine the required CP corresponding to each specific number of output bins. Moreover, to avoid getting

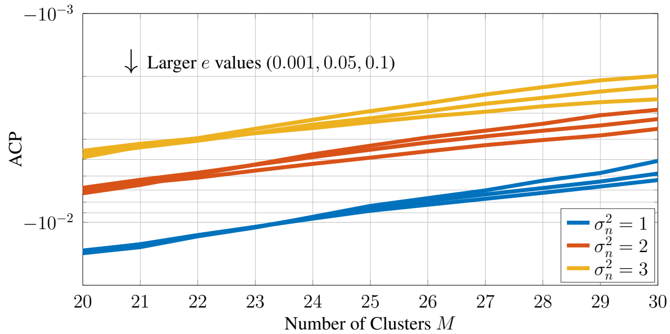

Fig. 4. ACP vs. number of clusters M, 16-QAM signaling, AWGN access channel with $\sigma_n^2 = 1, 2, 3$, forward channel with $e = 0.1, 0.05, 0.001$

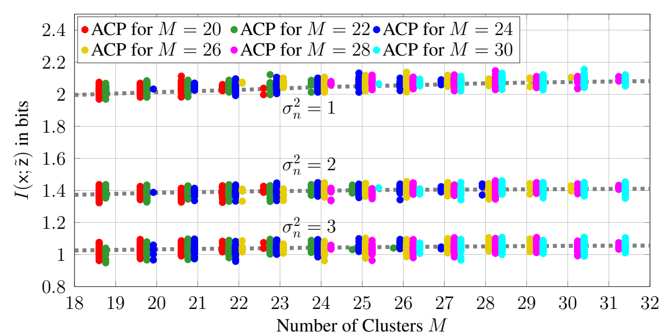

Fig. 5. Overall transmission rate $I(\mathsf{x}; \tilde{\mathsf{z}})$ vs. number of clusters M, 16-QAM signaling, AWGN access channel with $\sigma_n^2 = 1, 2, 3$, forward channel with $e = 0.05$, bisection result (····), slight misalignment for overlap prevention

stuck into bad local optima, Ch-Opt-IB was rerun 100 times with the best outcome taken. Regarding both figures, it is seen that irrespective of the specific choice of model parameters, our proposed treatment outperforms the result achieved by Ch-Opt-IB for a wide range of output bins. Besides, the observed behavior of acquired end-to-end transmission rate, $I(\mathsf{x}; \tilde{\mathsf{z}})$, w.r.t. different choices of σ_n^2 and e is justified following the line of argumentation provided in [13]. Principally, as $I(\mathsf{x}; \tilde{\mathsf{z}})$ is upper-bounded by the *minimum* capacity among both DMCs at the system model in Fig. 1, its observed increment either by decreasing σ_n^2 or e is quite natural.

Next, to obviate the exploitation of bisection method which results in a substantial reduction on the required computational load of our proposed AP-based treatment, we calculated the *Approximate Common Preference (ACP)* for any number of output cardinality as illustrated in Fig. 4. To obtain these curves, for each certain σ_n^2 and e, we repeated the aforementioned transmission 100 times and stored the respective CPs (using bisection). The acquired ACP for every value of M is the arithmetic mean of the lowest and the highest values. To test the usability of such a preference guideline (that can be calculated once offline), we repeated the transmission 250 times more and for every new trial, we inserted the ACP values (taken from Fig. 4) for six choices of M as the chosen CP and executed the AP-based routine only once. The attained results are depicted in Fig. 5. Irrespective of the specific choice of M and σ_n^2, we observed that the *one-shot* result had the output cardinality in the range of $M \pm 2$ for at least 85% of the trials. This means that, in cases for which the strict upholding of a specific output cardinality is not a must and a rather small variation range can be tolerated, only *one run* of our proposed AP-based approach (applying the precalculated ACPs) yields a quite promising result.

V. Summary

In this paper, we focused on the joint source-channel coding setup. Instead of conventional techniques from *Rate-Distortion* theory, we deployed the *Information Bottleneck* paradigm. As the key contribution, we proposed a novel treatment to address the design problem exploiting the *Affinity Propagation* that is an efficient message passing approach for clustering. Eventually, we investigated the performance of our proposed treatment w.r.t. the SotA Ch-Opt-IB and showed that the novel approach can be taken as a quite competitive alternative.

Acknowledgment

This work was partly funded by the German ministry of education and research (BMBF) under grant 16KIS0720 (TACNET 4.0).

References

[1] G. C. Zeitler, "Low-Precision Quantizer Design for Communication Problems," Ph.D. dissertation, TU Munich, Germany, 2012.
[2] B. Chen, L. Tong, and P. K. Varshney, "Channel-Aware Distributed Detection in Wireless Sensor Networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 16–26, July 2006.
[3] D. Wübben, P. Rost, J. Bartelt, M. Lalam, V. Savin, M. Gorgoglione, A. Dekorsy, and G. Fettweis, "Benefits and Impact of Cloud Computing on 5G Signal Processing," *Special Issue "The 5G Revolution" of the IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 35–44, Nov. 2014.
[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, 2006.
[5] A. Kurtenbach and P. Wintz, "Quantizing for Noisy Channels," *IEEE Trans. on Communication Technology*, vol. 17, no. 2, pp. 291–302, 1969.
[6] N. Farvardin and V. Vaishampayan, "Optimal Quantizer Design for Noisy Channels: An Approach to Combined Source-Channel Coding," *IEEE Trans. on Information Theory*, vol. 33, no. 6, pp. 827–838, 1987.
[7] N. Tishby, F. C. Pereira, and W. Bialek, "The Information Bottleneck Method," in *37th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, pp. 368–377, Sep. 1999.
[8] N. Slonim, "The Information Bottleneck: Theory and Applications," Ph.D. dissertation, Hebrew University of Jerusalem, Israel, 2002.
[9] C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
[10] S. Hassanpour, D. Wübben, and A. Dekorsy, "Overview and Investigation of Algorithms for the Information Bottleneck Method," in *11th Int. Conference on Systems, Communications and Coding (SCC)*, Hamburg, Germany, Feb. 2017.
[11] S. Hassanpour, D. Wübben, A. Dekorsy, and B. M. Kurkoski, "On the Relation Between the Asymptotic Performance of Different Algorithms for Information Bottleneck Framework," in *IEEE Int. Conference on Communications (ICC)*, Paris, France, May 2017.
[12] S. Hassanpour, D. Wübben, and A. Dekorsy, "On the Equivalence of Double Maxima and KL-Means for Information Bottleneck-Based Source Coding," in *IEEE Wireless Communications and Networking Conference (WCNC 2018)*, Barcelona, Spain, Apr. 2018.
[13] ——, "On the Equivalence of Two Information Bottleneck-Based Routines Devised for Joint Source-Channel Coding," in *25th Int. Conference on Telecommunication (ICT 2018)*, Saint-Malo, France, June 2018.
[14] B. J. Frey and D. Dueck, "Clustering by Passing Messages between Data Points," *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007.
[15] A. Winkelbauer, G. Matz, and A. Burg, "Channel-Optimized Vector Quantization with Mutual Information as Fidelity Criterion," in *Proc. Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, pp. 851–855, Nov. 2013.
[16] R. Horst and P. M. Pardalos, *Handbook of Global Optimization*. Springer Science & Business Media, 2013, vol. 2.
[17] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor Graphs and the Sum-Product Algorithm," *IEEE Trans. on Information Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
[18] D. Dueck, "Affinity Propagation: Clustering Data by Passing Messages," Ph.D. dissertation, University of Toronto, Canada, 2009.