Sampling Rate Distortion

Vinay Praneeth Boda, *Student Member, IEEE*, and Prakash Narayan, *Fellow, IEEE*

Abstract—Consider a discrete memoryless multiple source with m components of which $k \leq m$ possibly different sources are sampled at each time instant and jointly compressed in order to reconstruct all the m sources under a given distortion criterion. A new notion of sampling rate distortion function is introduced, and is characterized first for the case of fixed-set sampling. Next, for independent random sampling performed without knowledge of the source outputs, it is shown that the sampling rate distortion function is the same regardless of whether or not the decoder is informed of the sequence of sampled sets. Furthermore, memoryless random sampling is considered with the sampler depending on the source outputs and with an informed decoder. It is shown that deterministic sampling, characterized by a conditional point-mass, is optimal and suffices to achieve the sampling rate distortion function. For memoryless random sampling with an uninformed decoder, an upper bound for the sampling rate distortion function is seen to possess a similar property of conditional point-mass optimality. It is shown by example that memoryless sampling with an informed decoder can outperform strictly any independent random sampler, and that memoryless sampling can do strictly better with an informed decoder than without.

Index Terms—Discrete memoryless multiple source, independent random sampler, memoryless random sampler, random sampling, rate distortion, sampling rate distortion function.

I. INTRODUCTION

CONSIDER a set $\mathcal{M}$ of m discrete memoryless sources with a known joint probability mass function. Subsets of $k \leq m$ sources are sampled "spatially" at each time instant, and jointly processed with the objective of reconstructing *all* the m sources as compressed representations, within a specified level of distortion. How should the sampler optimally sample the sources in a causal manner to yield the best compression rate for a given distortion level? What are the tradeoffs – under optimal processing – among the sampling procedure, compression rate and distortion level? This paper is our preliminary attempt at answering these questions.

Such questions are motivated by various applications. An instance is in "dynamic thermal management," which is

the process of controlling surges in the operating temperature of a multicore processor chip during runtime, based on measurements by a limited number of on-chip thermal sensors. Strategic sensor placement and processing are needed to estimate temperatures at grid points over the entire chip [20], [34]. Another typical application involves "in-network computation" [12], [11] in which a subset of a network of collocated sensors use only their own measurements to estimate an aggregate function of the entirety of distributed and correlated measurements, e.g., overall average parameter values in environmental monitoring. In such settings, the mechanisms for (spatial) sampling, compression and estimation are collocated, with the latter two being aware of the sampler realizations.

The study of problems of combined sampling and compression has a rich and varied history in diverse contexts. Highlights include: classical sampling and processing, rate distortion theory, multiterminal source coding, wavelet-based compression, and compressed sensing, among others. Rate distortion theory [1] rules the compression of a given sampled signal and its reconstruction within a specified distortion level. On the other hand, compressed sensing [10] provides a random linear encoding of nonprobabilistic analog sources marked by a sparse support, with lossless recovery as measured by a block error probability (with respect to the distribution of the encoder). Upon placing the problem of lossless source coding of analog sources in an information theoretic setting, with a probabilistic model for the source that need not be encoded linearly, Rényi dimension is known to determine fundamental performance limits [32] (see also [16], [31]). Several recent studies consider the compressed sensing of a signal with an allowed detection error rate or quantization distortion [25], [30]; of multiple signals followed by distributed quantization [26], including a study of scaling laws [14]; or of sub-Nyquist rate sampled signals followed by lossy reconstruction [17].

Closer to the line of our work, the rate distortion function has been characterized when multiple Gaussian signals from a random field are sampled and quantized (centralized or distributed) in [21]–[23]. Also, in a series [27]–[29] (see also [6], [13], [24]), various aspects of random field-sampling and reconstruction for special models are considered. In a setting of distributed acoustic sensing and reconstruction, centralized as well as distributed coding schemes and sampling lattices are studied, and their performance is compared with corresponding rate distortion bounds [18]. In [15], considering a Gaussian random field on the interval $[0, 1]$ and i.i.d. in time, reconstruction of the entire field from compressed versions of k sampled sequences under the mean-squared

error distortion criterion is studied. In a different formulation, for the case of $m = 2$ sources, each of which is sampled for a fixed proportion of time, the rate distortion function and associated sampling mechanism are characterized in [19].

Our work differs materially from the approaches above. Sampling is spatial rather than temporal, unlike in most of the settings above. Furthermore, we introduce new forms of randomized sampling that can depend on the observed source values. Such randomized samplings, albeit of increased complexity, are shown to yield clear gains in performance. It bears emphasizing that we deal with centralized – and not distributed – processing of the sources. Moreover, no sparsity assumption is made on the sources.

Our contributions are as follows. We consider a new formulation involving a *sampling rate distortion*[1] *function* (SRDf), which combines a sampling of sources and lossy compression, to address the questions posed at the outset. As a basic ingredient, the sampling rate distortion function is characterized for a fixed sampling set of size $k \leq m$. This characterization is a consequence of prior work by Dobrushin-Tsybakov [9] (see also Berger [1], [2] and Yamamoto and Itoh [33]) on the rate distortion function for a "remote" source-receiver model in which the encoder and receiver lack direct access to the source and decoder outputs, respectively. For the special case of the probability of error distortion criterion, we show that the optimal procedure can be simplified to a rate distortion code for the sampled sources followed by maximum a posteriori estimation of the remaining sources.

Best fixed-set sampling can be strictly inferior to random sampling. Considering an independent random sampler, in which the sampling does not depend on the source outputs and is independent (but not necessarily identically distributed) in time, we show that the corresponding SRDf remains the same regardless of whether or not the decoder is provided information regarding the sequence of sampled sets. This surprising property does not hold for any causal sampler, in general. Next, we consider a generalization, namely a memoryless random sampler whose output can depend on the source values at each time instant. The associated formula for SRDf is used now to study the structural characteristics of the optimal sampler. Specifically, we show when the decoder too is aware of the sequence of sampled sets that the optimal sampler is characterized by a conditional point-mass; this has the obvious benefit of a reduction in the search space for an optimal sampler. We also show that such a memoryless sampler can outperform strictly a random sampler that lacks access to source values. Finally, in a setting in which the decoder is unaware of the sampled sequence, an upper bound for the SRDf is seen to have an optimal conditional point-mass sampler.

Our models are described in Section II. The main results, along with examples, are stated in Section III. Section IV contains the proofs. Presented first are the achievability proofs that are built successively in the order of increasing complexity

of the samplers. The converse proofs follow in reverse order in a unified manner.

II. PRELIMINARIES

Let $\mathcal{M} = \{1, \ldots, m\}$ and $X_{\mathcal{M}} = (X_1, \ldots, X_m)$ be a $\mathcal{X}_{\mathcal{M}} = \overset{m}{\underset{i=1}{\times}} \mathcal{X}_i$-valued rv where each $\mathcal{X}_i$ is a finite set. It will be convenient to use the following compact notation. For a nonempty set $A \subseteq \mathcal{M}$, we denote by X_A the rv $(X_i, i \in A)$ with values in $\underset{i \in A}{\times} \mathcal{X}_i$, and denote n repetitions of X_A by $X_A^n = (X_i^n, i \in A)$ with values in $\mathcal{X}_A^n = \underset{i \in A}{\times} \mathcal{X}_i^n$, where $X_i^n = (X_{i1}, \ldots, X_{in})$ takes values in the n-fold product space $\mathcal{X}_i^n = \mathcal{X}_i \times \cdots \times \mathcal{X}_i$. For $1 \leq k \leq m$, let $\mathcal{A}_k = \{A : A \subseteq \mathcal{M}, |A| = k\}$ be the set of all k-sized subsets of $\mathcal{M}$ and let $A^c = \mathcal{M} \setminus A$. All logarithms and exponentiations are with respect to the base 2.

Consider a discrete memoryless multiple source (DMMS) $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$ consisting of i.i.d. repetitions of the rv $X_{\mathcal{M}}$ with given pmf $P_{X_{\mathcal{M}}}$ of assumed full support $\mathcal{X}_{\mathcal{M}}$. Let $\mathcal{Y}_{\mathcal{M}} = \overset{m}{\underset{i=1}{\times}} \mathcal{Y}_i$, where $\mathcal{Y}_i$ is a finite reproduction alphabet for X_i.

Definition 1: A *k-random sampler* (k-RS), $1 \leq k \leq m$, collects causally at each $t = 1, \ldots, n$, random samples[†] X_{S_t} from $X_{\mathcal{M}t}$, where S_t is a rv with values in $\mathcal{A}_k$ with (conditional) pmf $P_{S_t|X_{\mathcal{M}}^t S^{t-1}}$, with $X_{\mathcal{M}}^t = (X_{\mathcal{M}1}, \ldots, X_{\mathcal{M}t})$ and $S^{t-1} = (S_1, \ldots, S_{t-1})$. Such a k-RS is specified by a (conditional) pmf $P_{S^n|X_{\mathcal{M}}^n}$ with the requirement

$$P_{S^n|X_{\mathcal{M}}^n} = \prod_{t=1}^{n} P_{S_t|X_{\mathcal{M}}^t S^{t-1}}. \tag{1}$$

The output of a k-RS is (S^n, X_S^n) where $X_S^n = (X_{S_1}, \ldots, X_{S_n})$. Successively restrictive choices of a k-RS in (1) corresponding to

$$P_{S_t|X_{\mathcal{M}}^t S^{t-1}} = P_{S_t|X_{\mathcal{M}t}}, \quad t = 1, \ldots, n \tag{2}$$

and

$$P_{S_t|X_{\mathcal{M}}^t S^{t-1}} = P_{S_t}, \quad t = 1, \ldots, n \tag{3}$$

will be termed the *k-memoryless* and the *k-independent random samplers* and denoted by k-MRS and k-IRS, respectively.

Definition 2: An n-length block code with k-RS for a DMMS $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$ with alphabet $\mathcal{X}_{\mathcal{M}}$ and reproduction alphabet $\mathcal{Y}_{\mathcal{M}}$ is a triple $(P_{S^n|X_{\mathcal{M}}^n}, f_n, \varphi_n)$ where $P_{S^n|X_{\mathcal{M}}^n}$ is a k-RS as in (1), and (f_n, φ_n) are a pair of mappings where the encoder f_n maps the k-RS output (S^n, X_S^n) into some finite set $\mathcal{J} = \{1, \ldots, J\}$ and the decoder φ_n maps $\mathcal{J}$ into $\mathcal{Y}_{\mathcal{M}}^n$. We shall use the compact notation $(P_{S|X_{\mathcal{M}}}, f, \varphi)$, suppressing n. The rate of the code with k-RS $(P_{S|X_{\mathcal{M}}}, f, \varphi)$ is $\frac{1}{n} \log J$.

Remark: An encoder that uses a deterministic estimate of $X_{S^c}^n$ from (S^n, X_S^n) in its operation is subsumed by the definition above.

For a given (single-letter) finite-valued distortion measure $d : \mathcal{X}_{\mathcal{M}} \times \mathcal{Y}_{\mathcal{M}} \to \mathbb{R}^+ \cup \{0\}$, an n-length block code with k-RS $(P_{S|X_{\mathcal{M}}}, f, \varphi)$ will be required to satisfy the expected

---

[1]This apt terminology has been used also in an earlier work on compressed sensing with error tolerance [25].

[†]With an abuse of notation, we write $X_{S_t t}$ simply as X_{S_t}.

fidelity criterion (d, Δ), i.e.,

$$\mathbb{E}\left[d\left(X_\mathcal{M}^n, \varphi\big(f(S^n, X_S^n)\big)\right)\right]$$
$$\triangleq \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^n d\left(X_{\mathcal{M}t}, \left(\varphi\big(f(S^n, X_S^n)\big)\right)_t\right)\right] \le \Delta. \quad (4)$$

We shall consider also the case where the decoder is informed of the sequence of sampled sets S^n. Denoting such an *informed decoder* by φ_S, the expected fidelity criterion (4) will use the augmented $\varphi_S\big(S^n, f(S^n, X_S^n)\big)$ instead of $\varphi\big(f(S^n, X_S^n)\big)$. The earlier decoder (that is not informed) will be termed an *uninformed decoder*.

Definition 3: A number $R \ge 0$ is an achievable k-sample coding rate at average distortion level Δ if for every $\epsilon > 0$ and sufficiently large n, there exist n-length block codes with k-RS of rate less than $R + \epsilon$ and satisfying the expected fidelity criterion $(d, \Delta + \epsilon)$; and (R, Δ) will be termed an achievable k-sample rate distortion pair. The infimum of such achievable rates is denoted by $R^I(\Delta)$ for an informed decoder, and by $R^U(\Delta)$ for an uninformed decoder. We shall refer to $R^I(\Delta)$ as well as $R^U(\Delta)$ as the *sampling rate distortion function* (SRDf), suppressing the dependence on k.

Remarks: (i) In the setting of an informed decoder, the sampling mechanism has two means of conveying information regarding $X_\mathcal{M}^n$ to the decoder: via the encoder output as well as by embedding it implicitly in S^n.

(ii) Clearly, $R^I(\Delta) \le R^U(\Delta)$, and both are nonincreasing in k.

(iii) For a DMMS $\{X_{\mathcal{M}t}\}_{t=1}^\infty$, the requirement (2) on the sampler renders $\{(X_{\mathcal{M}t}, S_t)\}_{t=1}^\infty$ and thereby also $\{(X_{S_t}, S_t)\}_{t=1}^\infty$ to be memoryless sequences.

III. RESULTS

We state our results in the order of increasing complexity of the sampling mechanism. Concomitant improvements in performance thereby become evident.

Single-letter characterizations of the SRDfs in this paper involve, as an ingredient, a characterization of $R^I(\Delta)$ with $S_t = A$, $t = 1, \ldots, n$, where $A \subseteq \mathcal{M}$ is a *fixed* set with $|A| = k$. Denote the corresponding $R^I(\Delta)$ by $R_A(\Delta)$ (with an abuse of notation). The *fixed-set* SRDf $R_A(\Delta)$, in effect, is the (standard) rate distortion function for the DMMS $\{X_{At}\}_{t=1}^\infty$ using a modified distortion measure $d_A : \mathcal{X}_A \times \mathcal{Y}_\mathcal{M} \to \mathbb{R}^+ \cup \{0\}$ defined by

$$d_A(x_A, y_\mathcal{M}) = \mathbb{E}[d(X_\mathcal{M}, y_\mathcal{M})|X_A = x_A]. \quad (5)$$

Proposition 1: For a DMMS $\{X_{\mathcal{M}t}\}_{t=1}^\infty$, the fixed-set SRDf for $A \subseteq \mathcal{M}$ is

$$R_A(\Delta) = \min_{\substack{X_{A^c} \, -\!\!\circ\!\!- \, X_A \, -\!\!\circ\!\!- \, Y_\mathcal{M} \\ \mathbb{E}[d(X_\mathcal{M}, Y_\mathcal{M})] \le \Delta}} I\big(X_A \wedge Y_\mathcal{M}\big) \quad (6)$$

for $\Delta_{\min, A} \le \Delta \le \Delta_{\max}$, *and equals* 0 *for* $\Delta \ge \Delta_{\max}$, *where*

$$\Delta_{\min, A} = \mathbb{E}\left[\min_{y_\mathcal{M} \in \mathcal{Y}_\mathcal{M}} d_A(X_A, y_\mathcal{M})\right],$$
$$\Delta_{\max} = \min_{y_\mathcal{M} \in \mathcal{Y}_\mathcal{M}} \mathbb{E}[d(X_\mathcal{M}, y_\mathcal{M})]$$
$$= \min_{y_\mathcal{M} \in \mathcal{Y}_\mathcal{M}} \mathbb{E}[d_A(X_A, y_\mathcal{M})]. \quad (7)$$

Corollary 1: With $\mathcal{X}_\mathcal{M} = \mathcal{Y}_\mathcal{M}$, *for the probability of error distortion measure*

$$d(x_\mathcal{M}, y_\mathcal{M}) = \mathbb{1}(x_\mathcal{M} \ne y_\mathcal{M})$$
$$= 1 - \prod_{i=1}^m \mathbb{1}(x_i = y_i), \quad x_\mathcal{M}, y_\mathcal{M} \in \mathcal{X}_\mathcal{M} \quad (8)$$

we have

$$R_A(\Delta) = \begin{cases} \min I\big(X_A \wedge Y_A\big), & \Delta_{\min} \le \Delta \le \Delta_{\max} \\ 0, & \Delta \ge \Delta_{\max}, \end{cases} \quad (9)$$

where the minimum in (9) *is subject to*

$$\mathbb{E}[\alpha(X_A)\mathbb{1}(X_A \ne Y_A)] \le \Delta - (1 - \mathbb{E}[\alpha(X_A)]) \quad (10)$$

with

$$\alpha(x_A) = \max_{\tilde{x}_{A^c} \in \mathcal{X}_{A^c}} P_{X_{A^c}|X_A}(\tilde{x}_{A^c}|x_A)$$

and

$$\Delta_{\min} = 1 - \mathbb{E}[\alpha(X_A)], \quad \Delta_{\max} = 1 - \max_{x_\mathcal{M} \in \mathcal{X}_\mathcal{M}} P_{X_\mathcal{M}}(x_\mathcal{M}).$$

Remarks: (i) The minimum in (6) exists by virtue of the continuity in $P_{X_\mathcal{M} Y_\mathcal{M}}$ of $I(X_A \wedge Y_\mathcal{M})$ over the compact set $\{P_{X_\mathcal{M} Y_\mathcal{M}} : X_{A^c} -\!\!\circ\!\!- X_A -\!\!\circ\!\!- Y_\mathcal{M}, \ \mathbb{E}[d(X_\mathcal{M}, Y_\mathcal{M})] \le \Delta\}$.

(ii) The corollary relies on showing that the minimum in (6) is attained in this particular instance by a pmf $P_{X_\mathcal{M} Y_\mathcal{M}}$ under a longer Markov chain

$$X_{A^c} -\!\!\circ\!\!- X_A -\!\!\circ\!\!- Y_A -\!\!\circ\!\!- Y_{A^c}.$$

Interestingly, the achievability proof entails in a first step a mapping of x_A^n in $\mathcal{X}_A^n$ into its codeword y_A^n, from which in a second step a reconstruction $y_{A^c}^n$ of $x_{A^c}^n$ is obtained as a maximum a posteriori (MAP) estimate.

The k-IRS affords a more capable mechanism than the fixed-set sampler of Proposition 1, with the sampling sets possibly varying in time. Surprisingly, our next result shows that the SRDf for a k-IRS, displayed as $R_i(\Delta)$, remains the same regardless of whether or not the decoder is provided information regarding the sequence of sampled sets. As seen from its proof in Section IV, this is enabled by the lack of dependence of the sampling sequence on the DMMS realization.

Theorem 2: For a k-IRS,

$$R_i^I(\Delta) = R_i^U(\Delta) = R_i(\Delta) = \min I\big(X_S \wedge Y_\mathcal{M}|S\big) \quad (11)$$

for $\Delta_{\min} \le \Delta \le \Delta_{\max}$, *where the minimum is with respect to* $P_{X_\mathcal{M} S Y_\mathcal{M}} = P_{X_\mathcal{M}} P_S P_{Y_\mathcal{M}|S X_S}$ *and* $\mathbb{E}[d(X_\mathcal{M}, Y_\mathcal{M})] \le \Delta$, *with*

$$\Delta_{\min} = \min_{A \in \mathcal{A}_k} \mathbb{E}\left[\min_{y_\mathcal{M} \in \mathcal{Y}_\mathcal{M}} d_A(X_A, y_\mathcal{M})\right]$$

and $\Delta_{\max}$ *as in* (7).

A convenient equivalent expression for $R_i(\Delta)$ in (11) is given by

Proposition 3: For a k-IRS,

$$R_i(\Delta) = \min \sum_{A \in \mathcal{A}_k} P_S(A) R_A(\Delta_A), \quad \Delta_{\min} \le \Delta \le \Delta_{\max},$$

$$(12)$$

where the minimum is with respect to

$$P_S, \ \left\{ \Delta_A \geq \Delta_{\min,A}, \ A \in \mathcal{A}_k : \sum_{A \in \mathcal{A}_k} P_S(A)\Delta_A \leq \Delta \right\}. \quad (13)$$

Proof: For every $\Delta_{\min} \leq \Delta \leq \Delta_{\max}$, in (11),

$$\min_{\substack{P_{X_{\mathcal{M}}} P_S P_{Y_{\mathcal{M}}|SX_S} \\ \mathbb{E}[d(X_{\mathcal{M}}, Y_{\mathcal{M}})] \leq \Delta}} I(X_S \wedge Y_{\mathcal{M}}|S)$$

$$= \min_{\substack{P_{X_{\mathcal{M}}} P_S P_{Y_{\mathcal{M}}|SX_S} \\ \mathbb{E}[d(X_{\mathcal{M}}, Y_{\mathcal{M}})] \leq \Delta}} \sum_{A \in \mathcal{A}_k} P_S(A) I(X_A \wedge Y_{\mathcal{M}}|S=A)$$

$$= \min_{\substack{P_S, \Delta_A: \\ \sum_{A \in \mathcal{A}_k} P_S(A)\Delta_A \leq \Delta}} \sum_{A \in \mathcal{A}_k} P_S(A)$$

$$\times \min_{\substack{P_{Y_{\mathcal{M}}|S=A, X_A} \\ \mathbb{E}[d(X_{\mathcal{M}}, Y_{\mathcal{M}})|S=A]=\Delta_A}} I(X_A \wedge Y_{\mathcal{M}}|S=A)$$

$$= \min_{\substack{P_S, \Delta_A: \\ \sum_{A \in \mathcal{A}_k} P_S(A)\Delta_A \leq \Delta}} \sum_{A \in \mathcal{A}_k} P_S(A) R_A(\Delta_A), \quad (14)$$

where $P_{Y_{\mathcal{M}}|S=A, X_A}$ is used to denote $P_{Y_{\mathcal{M}}|S, X_S}(\cdot|A, \cdot)$ for compactness. The validity of (14) follows by the introduction of the Δ_As and observing that the order of the minimization does not alter the value of the minimum. The last step obtains upon noting that the value of the inner minimum in (14) is the same upon replacing the equality in $\mathbb{E}[d(X_{\mathcal{M}}, Y_{\mathcal{M}})|S = A] = \Delta_A$ with "$\leq$". $\square$

Remark: By Proposition 3, the SRDf for a k-IRS is the lower convex envelope of the set of SRDfs $\{R_A(\Delta), \ A \in \mathcal{A}_k\}$ and thus is convex in $\Delta \geq \Delta_{\min}$. Furthermore,

$$R_i(\Delta) \leq \min_{A \in \mathcal{A}_k} R_A(\Delta).$$

Additionally, a k-IRS can outperform strictly the best fixed-set sampler. For instance, if there is no fixed-set SRDf for any $A \in \mathcal{A}_k$ that is uniformly best for all Δ, then the previous inequality can be strict. This is illustrated by the following example.

Example 1: With $\mathcal{M} = \{1, 2\}$, $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{Y}_2 = \{0, 1\}$, and $\mathcal{Y}_1 = \{0, 1, e\}$, let X_1, X_2 be i.i.d. Bernoulli(0.5) rvs, and

$$d\big((x_1, x_2), (y_1, y_2)\big) = d_1(x_1, y_1) + d_2(x_2, y_2)$$

with

$$d_1(x_1, y_1) = \begin{cases} 0, & \text{if } x_1 = y_1 = 0; \ x_1 = y_1 = 1 \\ 1, & \text{if } x_1 = 0, 1, \ \ y_1 = e \\ \infty, & \text{if } x_1 = 0, \ y_1 = 1; \ x_1 = 1, \ y_1 = 0, \end{cases}$$

$$d_2(x_2, y_2) = \mathbb{1}(x_2 \neq y_2).$$

For $k = 1$,

$$R_{\{1\}}(\Delta) = 1.5 - \Delta, \quad 0.5 \leq \Delta \leq 1.5,$$
$$R_{\{2\}}(\Delta) = 1 - h(\Delta - 1), \quad 1 \leq \Delta \leq 1.5$$

whereas

$$R_i(\Delta) = \begin{cases} 1.5515 - 1.103\Delta, & 0.5 \leq \Delta \leq 1.318, \\ 1 - h(\Delta - 1), & 1.318 \leq \Delta \leq 1.5, \end{cases}$$

where $h(\cdot)$ is the binary entropy function. Clearly, $R_i(\Delta)$ is strictly smaller than $\min\{R_{\{1\}}(\Delta), R_{\{2\}}(\Delta)\}$ for
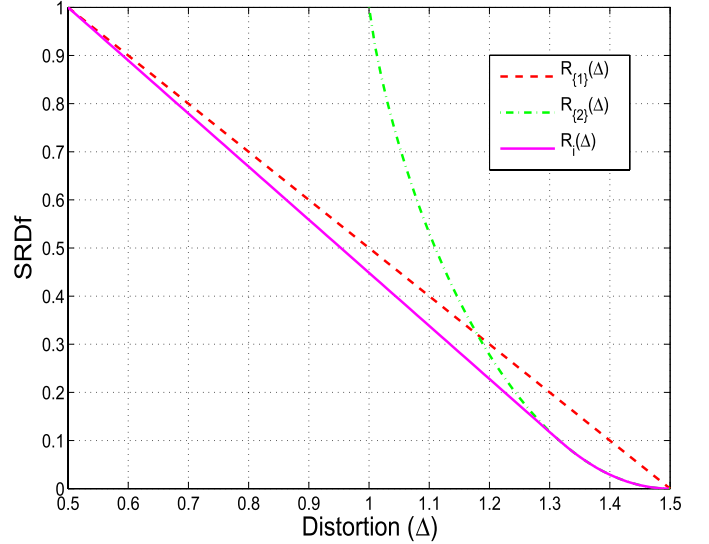

Fig. 1.   SRDfs for k-IRS vs. fixed-set sampler.

$0.5 < \Delta < 1.318$; see Fig. 1. Note that while the distortion measure d in Definition 2 is taken to be finite-valued, the event $\{d(X_1, Y_1) = \infty\}$ above is accommodated by assigning (optimally) zero probability to it.

$\square$

A k-MRS is more powerful than a k-IRS in that sampling with the former at each time instant can depend on the current DMMS realization. The SRDf for a k-MRS can improve with an informed decoder unlike for a k-IRS.

Theorem 4: For a k-MRS with informed decoder, the SRDf is

$$R_m^I(\Delta) = \min I\big(X_S \wedge Y_{\mathcal{M}}|S, U\big) \quad (15)$$

for $\Delta_{\min} \leq \Delta \leq \Delta_{\max}$, *where the minimum is with respect to* $P_{U X_{\mathcal{M}} S Y_{\mathcal{M}}} = P_U P_{X_{\mathcal{M}}} P_{S|X_{\mathcal{M}} U} P_{Y_{\mathcal{M}}|SX_S U}$ *and* $\mathbb{E}[d(X_{\mathcal{M}}, Y_{\mathcal{M}})] \leq \Delta$, *with*

$$\Delta_{\min} = \min_{P_{S|X_{\mathcal{M}}}} \mathbb{E}\left[\min_{P_{Y_{\mathcal{M}}|SX_S}} \mathbb{E}[d(X_{\mathcal{M}}, Y_{\mathcal{M}})|S, X_S] \right], \quad (16)$$

$$\Delta_{\max} = \min_{P_{S|X_{\mathcal{M}}}} \mathbb{E}\left[\min_{y_{\mathcal{M}}} \mathbb{E}[d(X_{\mathcal{M}}, y_{\mathcal{M}})|S] \right], \quad (17)$$

and U *being a* $\mathcal{U}$-*valued rv with* $|\mathcal{U}| \leq 3$.

Remark: Analogously as in Proposition 3, the SRDf $R_m^I(\Delta)$ can be expressed as

$$R_m^I(\Delta) = \min_{\substack{P_U, \ \Delta_u: \\ \sum_u P_U(u)\Delta_u \leq \Delta}} \sum_u P_U(u)$$

$$\times \min_{\substack{P_{S|X_{\mathcal{M}}, U=u}, P_{Y_{\mathcal{M}}|SX_S, U=u} \\ \mathbb{E}[d(X_{\mathcal{M}}, Y_{\mathcal{M}})|U=u]=\Delta_u}} I(X_S \wedge Y_{\mathcal{M}}|S, U=u)$$

$$(18)$$

and thereby equals a lower convex envelope of functions of Δ.

The optimal sampler that attains the SRDf in Theorem 4 has a simple structure. It is easy to see that each of $\Delta_{\min}$ and $\Delta_{\max}$ in (16) and (17), respectively, is attained by a sampler for which $P_{S|X_{\mathcal{M}}}$ takes the form of a conditional point-mass.

Such samplers, in fact, are optimal for every distortion level $\Delta_{\min} \leq \Delta \leq \Delta_{\max}$ and will depend on Δ, in general.

Definition 4: Given a mapping $h : \mathcal{X}_{\mathcal{M}} \times \mathcal{U} \to \mathcal{A}_k$, the (conditional point-mass) mapping $\delta_{h(\cdot)} : \mathcal{A}_k \to \{0, 1\}$ is defined by

$$\delta_{h(x_{\mathcal{M}}, u)}(s) \triangleq \begin{cases} 1, & s = h(x_{\mathcal{M}}, u) \\ 0, & \text{otherwise,} \end{cases} \quad (19)$$

for $(x_{\mathcal{M}}, u) \in \mathcal{X}_{\mathcal{M}} \times \mathcal{U}, \; s \in \mathcal{A}_k$.

The following reduction of Theorem 4 shows the optimality of conditional point-mass samplers for a k-MRS which will be seen to play a material role in the achievability proof of Theorem 4.

Theorem 5: For a k-MRS with informed decoder,

$$R_m^I(\Delta) = \min I(X_S \wedge Y_{\mathcal{M}} | S, U) \quad (20)$$

for $\Delta_{\min} \leq \Delta \leq \Delta_{\max}$, with $\Delta_{\min}$ and $\Delta_{\max}$ as in (16) and (17), respectively, where the minimum is with respect to $P_{U X_{\mathcal{M}} S Y_{\mathcal{M}}}$ of the form $P_U P_{X_{\mathcal{M}}} \delta_{h(\cdot)} P_{Y_{\mathcal{M}} | S X_S U}$ with $\mathbb{E}[d(X_{\mathcal{M}}, Y_{\mathcal{M}})] \leq \Delta$, where the (time-sharing) rv U takes values in $\mathcal{U}$ with $|\mathcal{U}| \leq 3$.

The structure of the optimal sampler in Theorem 5 implies that the search space for minimization now can be reduced to the corner points of the simplexes of the conditional pmfs $P_{S | X_{\mathcal{M}} U}(\cdot | x_{\mathcal{M}}, u), \; (x_{\mathcal{M}}, u) \in \mathcal{X}_{\mathcal{M}} \times \mathcal{U}$. The SRDf in (20) is thus the lower convex envelope of the SRDfs for conditional point-mass samplers. In general, time-sharing between such samplers will be seen to achieve the best compression rate for a given distortion level.

Finally, for a k-MRS with uninformed decoder, we provide an upper bound for the SRDf $R_m^U(\Delta)$.

Theorem 6: For a k-MRS with uninformed decoder,

$$R_m^U(\Delta) \leq \min I(S, X_S \wedge Y_{\mathcal{M}}) \quad (21)$$

for $\Delta_{\min} \leq \Delta \leq \Delta_{\max}$, where the minimum is with respect to $P_{X_{\mathcal{M}} S Y_{\mathcal{M}}} = P_{X_{\mathcal{M}}} P_{S | X_{\mathcal{M}}} P_{Y_{\mathcal{M}} | S X_S}$ and $\mathbb{E}[d(X_{\mathcal{M}}, Y_{\mathcal{M}})] \leq \Delta$, with $\Delta_{\min}$ and $\Delta_{\max}$ being as in (16) and (7).

Remark: Clearly, when (S, X_S) in (21) determines $X_{\mathcal{M}}$, we have $R_m^U(\Delta) = R(\Delta) =$ the (standard) rate distortion function for the DMMS $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$.

The (achievability) proof of Theorem 6 is along the lines of Proposition 1. The lack of a converse is due to the inability to prove or disprove the convexity of the right-side of (21) in Δ. (Convexity would imply equality in (21).) The optimal sampler, however, can be shown to be a conditional point-mass sampler (19) along the lines of Theorem 5. Note that the same conditional point-mass sampler need not be the best in (15) and (21).

Strong forms of the k-MRS and k-IRS are obtained by allowing time-dependence in sampling. Specifically, (2) and (3) can be strengthened, respectively, to

$$P_{S_t | X_{\mathcal{M}}^t S^{t-1}} = P_{S_t | X_{\mathcal{M}t} S^{t-1}} \quad (22)$$

and

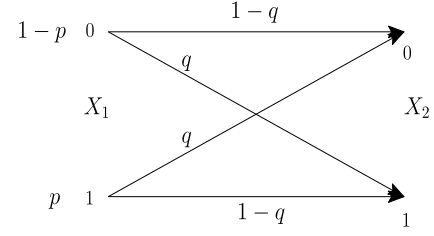$$P_{S_t | X_{\mathcal{M}}^t S^{t-1}} = P_{S_t | S^{t-1}}. \quad (23)$$

Fig. 2. BSC (q).

Surprisingly, this does not improve SRDf for the k-MRS (with decoder informed) or the k-IRS.

Proposition 7: For a strong k-MRS in (22) and a strong k-IRS in (23), the corresponding SRDfs $R_{ms}^I(\Delta)$ and $R_{is}(\Delta)$ equal the right-sides of (15) and (11), respectively.

Finally, standard properties of the SRDf for the fixed-set sampler, k-IRS and k-MRS with informed decoder are summarized in the following

Lemma 8: For a fixed $P_{X_{\mathcal{M}}}$, the right-sides of (6), (11) and (15) are finite-valued, nonincreasing, convex, continuous functions of Δ.

We close this section with an example showing that (i) the SRDf for a k-MRS with informed decoder can be strictly smaller than that of a k-IRS; and (ii) furthermore, unlike for a k-IRS, a k-MRS with informed decoder can outperform strictly that with an uninformed decoder, uniformly for all feasible distortion values.

Example 2: With $\mathcal{M} = \{1, 2\}$ and $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$, consider a DMMS with $P_{X_1 X_2}$ represented by a virtual binary symmetric channel (BSC) shown in Figure 2. Fix $p \leq 0.5$ and $q = 0.5$, i.e., X_1 and X_2 are independent. Let d correspond to the probability of error criterion, i.e., $d(x_{\mathcal{M}}, y_{\mathcal{M}}) = \mathbb{1}(x_{\mathcal{M}} \neq y_{\mathcal{M}})$.

(i) Considering a k-MRS, $k = 1$, with informed decoder, we obtain by Theorem 5 that $\Delta_{\min} = 0$, $\Delta_{\max} = p$, and the (conditional point-mass) sampler

$$P_{S | X_{\mathcal{M}}}(s | x_{\mathcal{M}}) = \begin{cases} 1, & s = 1, \; x_{\mathcal{M}} = 00 \text{ or } 11 \\ 1, & s = 2, \; x_{\mathcal{M}} = 01 \text{ or } 10 \\ 0, & \text{otherwise} \end{cases} \quad (24)$$

is uniformly optimal for all $0 \leq \Delta \leq p$, and

$$R_m^I(\Delta) = h(p) - h(\Delta), \quad 0 \leq \Delta \leq p.$$

To obtain $R_i(\Delta)$, the SRDfs for fixed-set samplers (6) are

$$R_{\{1\}}(\Delta) = h(p) - h(2\Delta - 1), \quad \tfrac{1}{2} \leq \Delta \leq \frac{1 + p}{2},$$
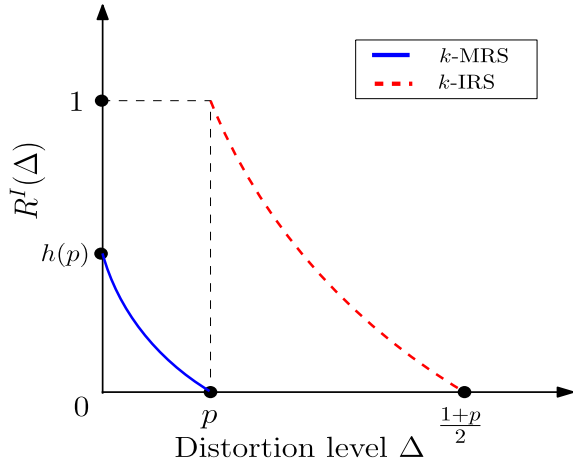
and

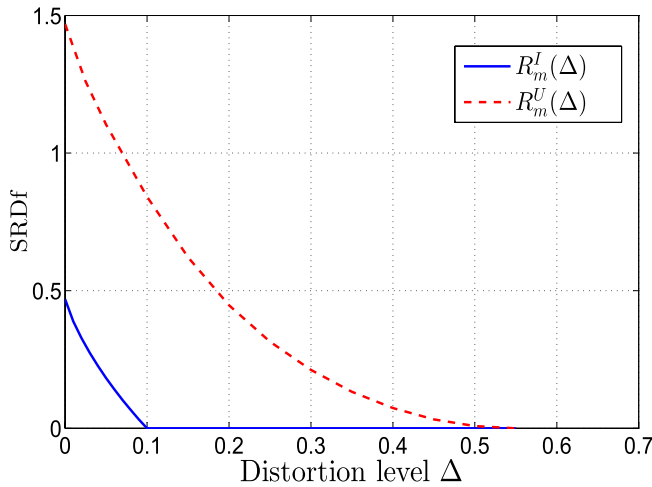$$R_{\{2\}}(\Delta) = h\left(\frac{1}{2}\right) - h\left(\frac{\Delta - p}{1 - p}\right), \quad p \leq \Delta \leq \frac{1 + p}{2}.$$

Since $R_{\{2\}}(\Delta) \leq R_{\{1\}}(\Delta)$ for all Δ, it is a simple exercise to show that

$$R_i(\Delta) = R_{\{2\}}(\Delta).$$

Clearly, $R_m^I(\Delta) \leq R_i(\Delta)$, with $\Delta_{\max}$ for the former being $\Delta_{\min}$ for the latter, as shown in Figure 3.

Fig. 3. SRDf for k-MRS vs. k-IRS.

Fig. 4. SRDf for k-MRS.

Fig. 5. $R_m^U(\Delta)$ and $R_i(\Delta)$.

(ii) The conditional pmf $P_{S|X_\mathcal{M}}$ in (24) represents a 1-1 map between the values of $X_\mathcal{M}$ and (S, X_S), and can be seen also to be the optimal choice in the right-side of (21) for all $0 \leq \Delta \leq \frac{1+p}{2}$. The remaining minimization in (21), with respect to $P_{Y_\mathcal{M}|SX_S}$, renders the right-side to be convex in Δ. Consequently, as observed in the passage following Theorem 6, the bound in (21) is tight. For $p = 0.1$, the resulting values of $R_m^I(\Delta)$ and $R_m^U(\Delta)$ are plotted in Figure 4, and of $R_m^U(\Delta)$ and $R_i(\Delta)$ in Figure 5. Figure 4 illustrates the benefit of decoder information for a k-MRS, while Figure 5 shows the compression gain achieved by providing source knowledge to the sampler.

IV. PROOFS

A. Achievability Proofs

Our achievability proofs successively build upon each other in the order: fixed-set sampler, k-IRS and k-MRS. The achievability proof of Proposition 1 for a fixed-set sampler forms a basic building block for subsequent application. Relying on this, the SRDf for a k-IRS is shown to be achieved in Theorem 2 without the decoder being informed of the sequence of sampled sets. Next, for a k-MRS with informed decoder, we prove first Theorem 5 which shows that the optimal sampler is deterministic in that the corresponding $P_{S|X_\mathcal{M}}$ is a
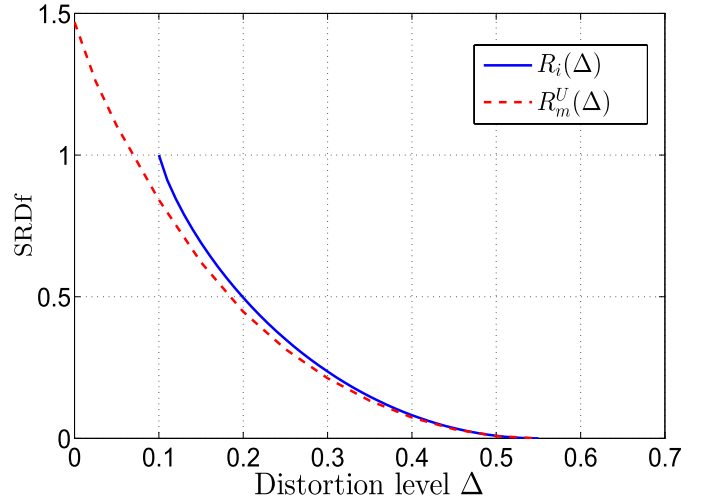
point-mass. This structure enables an achievability proof of Theorem 4 which builds on that of Proposition 1. Lastly, for a k-MRS with uninformed decoder, the achievability proof of Theorem 6 rests on the preceding proofs.

Proposition 1: The achievability proof below can be obtained directly from [9], but is given here for completeness. Observe first that

$$
\begin{aligned}
\Delta_{\min, A} &= \min_{X_{A^c} \; \multimap \; X_A \; \multimap \; Y_\mathcal{M}} \mathbb{E}[d(X_\mathcal{M}, Y_\mathcal{M})] \\
&= \min_{X_{A^c} \; \multimap \; X_A \; \multimap \; Y_\mathcal{M}} \mathbb{E}\big[\mathbb{E}[d(X_\mathcal{M}, Y_\mathcal{M})|X_A]\big] \\
&= \min_{P_{X_A Y_\mathcal{M}}} \mathbb{E}[d_A(X_A, Y_\mathcal{M})] \\
&\quad \text{by (5) and since } X_{A^c} \; \multimap \; X_A \; \multimap \; Y_\mathcal{M} \\
&= \mathbb{E}\Big[\min_{y_\mathcal{M} \in \mathcal{Y}_\mathcal{M}} d_A(X_A, y_\mathcal{M}) \Big]
\end{aligned}
$$

and

$$
\begin{aligned}
\Delta_{\max} &= \min_{\substack{X_{A^c} \; \multimap \; X_A \; \multimap \; Y_\mathcal{M} \\ P_{X_A Y_\mathcal{M}} = P_{X_A} P_{Y_\mathcal{M}}}} \mathbb{E}[d(X_\mathcal{M}, Y_\mathcal{M})] \\
&= \min_{P_{X_\mathcal{M}} P_{Y_\mathcal{M}}} \mathbb{E}[d(X_\mathcal{M}, Y_\mathcal{M})] \\
&= \min_{y_\mathcal{M} \in \mathcal{Y}_\mathcal{M}} \mathbb{E}[d(X_\mathcal{M}, y_\mathcal{M})].
\end{aligned}
$$

Next, note that for every $\Delta_{\min, A} \leq \Delta \leq \Delta_{\max}$,

$$
\begin{aligned}
&\min_{\substack{X_{A^c} \; \multimap \; X_A \; \multimap \; Y_\mathcal{M} \\ \mathbb{E}[d(X_\mathcal{M}, Y_\mathcal{M})] \leq \Delta}} I(X_A \wedge Y_\mathcal{M}) \\
&= \min_{\mathbb{E}[d_A(X_A, Y_\mathcal{M})] \leq \Delta} I(X_A \wedge Y_\mathcal{M}).
\end{aligned}
$$

Clearly every feasible $P_{X_\mathcal{M} Y_\mathcal{M}} = P_{X_{A^c} X_A Y_\mathcal{M}}$ on the left-side above gives a feasible $P_{X_A Y_\mathcal{M}}$ on the right-side. Similarly every feasible $P_{X_A Y_\mathcal{M}}$ on the right-side leads to a feasible $P_{X_\mathcal{M} Y_\mathcal{M}}$ on the left-side of the form $P_{X_\mathcal{M} Y_\mathcal{M}} = P_{X_{A^c}|X_A} P_{X_A Y_\mathcal{M}}$.

Given $\epsilon > 0$, consider a (standard) rate distortion code (f, φ) for the DMMS $\{X_{At}\}_{t=1}^\infty$ with distortion measure d_A, of rate $\frac{1}{n} \log \|f\| \leq R_A(\Delta) + \epsilon$ and with expected distortion $\mathbb{E}\big[d_A\big(X_A^n, \varphi(f(X_A^n))\big)\big] \leq \Delta + \epsilon$ for all $n \geq N_A(\epsilon)$, say.

The code (f, φ) also satisfies

$$\mathbb{E}[d(X_{\mathcal{M}}^n, Y_{\mathcal{M}}^n)]$$
$$= \frac{1}{n}\mathbb{E}\left[\sum_{t=1}^n \mathbb{E}\left[d\left(X_{\mathcal{M}t}, (\varphi(f(X_A^n)))_t\right)\Big|X_A^n\right]\right]$$
$$= \frac{1}{n}\mathbb{E}\left[\sum_{t=1}^n \mathbb{E}\left[d\left(X_{\mathcal{M}t}, (\varphi(f(X_A^n)))_t\right)\Big|X_{At}\right]\right]$$
$$= \mathbb{E}\left[d_A\left(X_A^n, \varphi(f(X_A^n))\right)\right]$$
$$\leq \Delta + \epsilon,$$

thereby yielding achievability in the proposition.

Turning to the corollary, for every $P_{X_{\mathcal{M}}Y_{\mathcal{M}}}$ satisfying the constraints in (6), consider the pmf $Q_{X_{\mathcal{M}}Y_{\mathcal{M}}}$ defined for $x_{\mathcal{M}}, y_{\mathcal{M}} \in \mathcal{X}_{\mathcal{M}}$

$$Q_{X_{\mathcal{M}}Y_{\mathcal{M}}}(x_{\mathcal{M}}, y_{\mathcal{M}})$$
$$\triangleq P_{X_{\mathcal{M}}Y_A}(x_{\mathcal{M}}, y_A)\mathbb{1}\left(y_{A^c} = MAP\ (y_A)\right), \quad (25)$$

where

$$MAP\ (y_A) = \arg\max_{\tilde{y}_{A^c} \in \mathcal{Y}_{A^c}} P_{X_{A^c}|X_A}(\tilde{y}_{A^c}|y_A) \quad (26)$$

is the maximum a posteriori estimate of y_{A^c} given y_A according to $P_{X_{A^c}|X_A}$. Observe that $Q_{X_{\mathcal{M}}Y_{\mathcal{M}}}$ satisfies

$$Q_{X_{A^c}} \multimap Q_{X_A} \multimap Q_{Y_A} \multimap Q_{Y_{A^c}} \quad (27)$$

and

$$\mathbb{E}_P[d(X_{\mathcal{M}}, Y_{\mathcal{M}})]$$
$$= P(X_{\mathcal{M}} \neq Y_{\mathcal{M}})$$
$$= P(X_A \neq Y_A)$$
$$\quad + P(X_A = Y_A)P(X_{A^c} \neq Y_{A^c}|X_A = Y_A)$$
$$= Q(X_A \neq Y_A)$$
$$\quad + Q(X_A = Y_A)P(X_{A^c} \neq Y_{A^c}|X_A = Y_A)$$
$$\geq Q(X_A \neq Y_A)$$
$$\quad + Q(X_A = Y_A)Q(X_{A^c} \neq Y_{A^c}|X_A = Y_A)$$
$$= Q(X_{\mathcal{M}} \neq Y_{\mathcal{M}}) = \mathbb{E}_Q[d(X_{\mathcal{M}}, Y_{\mathcal{M}})],$$

where the inequality is by (25), (26) and the optimality of the MAP estimator. Also, it is readily checked that

$$\mathbb{E}_Q[d(X_{\mathcal{M}}, Y_{\mathcal{M}})]$$
$$= 1 - \mathbb{E}[\alpha(X_A)] + \mathbb{E}[\alpha(X_A)\mathbb{1}(X_A \neq Y_A)]. \quad (28)$$

Furthermore,

$$I_Q(X_A \wedge Y_{\mathcal{M}}) = I_Q(X_A \wedge Y_A)$$
$$= I_P(X_A \wedge Y_A) \leq I_P(X_A \wedge Y_{\mathcal{M}}). \quad (29)$$

Putting together (25) - (29) and comparing with (6) establishes the corollary.

It is interesting to note that the form of (6)

$$\min_{\substack{X_{A^c} \multimap X_A \multimap Y_{\mathcal{M}} \\ P(X_{\mathcal{M}} \neq Y_{\mathcal{M}}) \leq \Delta}} I(X_A \wedge Y_{\mathcal{M}})$$
$$= \min_{\mathbb{E}[\alpha(X_A)\mathbb{1}(X_A \neq Y_A)] \leq \Delta - (1 - \mathbb{E}[\alpha(X_A)])} I(X_A \wedge Y_A)$$

leads to a simpler and direct proof of achievability of the corollary. Specifically, for a given Δ, first x_A^n is mapped into (only) its corresponding codeword y_A^n but under a modified distortion measure $\tilde{d}(x_A, y_A) \triangleq \alpha(x_A)\mathbb{1}(x_A \neq y_A)$ and a corresponding reduced threshold as indicated by (10). Next, the codewords y_A^n serve as sufficient statistics from which (the unsampled) $x_{A^c}^n$ is reconstructed as $y_{A^c}^n = MAP\ (y_A^n)$ under $P_{X_{A^c}^n|X_A^n}$; the corresponding estimation error coincides with the reduction in the threshold. $\qquad\square$

Theorem 2: The equivalent expression for $R_i(\Delta)$ given by Proposition 3 suggests an achievability scheme using a concatenation of fixed-set sampling rate distortion codes from Proposition 1. Let P_S and $\{\Delta_A,\ A \in \mathcal{A}_k\}$ yield the minimum in Proposition 3. A sequence of sampling sets S^n are constructed *a priori* with $S_t = A$ repeatedly for approximately $nP_S(A)$ time instants, for each A in $\mathcal{A}_k$. Correspondingly, sampling rate distortion codes of blocklength $\cong nP_S(A)$ – with distortion $\cong \Delta_A$ and of rate $\cong R_A(\Delta_A)$ – are concatenated. This predetermined selection of sampling sets does not require the decoder to be additionally informed.

For a fixed $\Delta_{\min} \leq \Delta \leq \Delta_{\max}$, let P_S and $\{\Delta_A,\ A \in \mathcal{A}_k\}$ attain the minimum in (12). Fix $\epsilon > 0$ and $0 < \epsilon' < \epsilon$. Order (in any manner) the elements of $\mathcal{A}_k$ as A_i, $i \in \mathcal{M}_k \triangleq \{1, \ldots, M_k\}$, with $M_k = \binom{m}{k}$. For $i \in \mathcal{M}_k$ and $n \geq 1$, define the "time-sets" ν_{A_i} for $A_i \in \mathcal{A}_k$ as

$$\nu_{A_i} = \left\{t : \lceil n\sum_{j=1}^{i-1} P_S(A_j)\rceil + 1 \leq t \leq \lceil n\sum_{j=1}^{i} P_S(A_j)\rceil\right\}.$$

The time-sets cover $\{1, \ldots, n\}$, i.e.,

$$\bigcup_{i \in \mathcal{M}_k} \nu_{A_i} = \{1, \ldots, n\}$$

and satisfy

$$\left|\frac{|\nu_{A_i}|}{n} - P_S(A_i)\right| \leq \frac{1}{n}, \quad i \in \mathcal{M}_k.$$

Now, a k-IRS is chosen with a deterministic sampling sequence $S^n = s^n$ according to

$$S_t = s_t = A_i, \quad t \in \nu_{A_i}, \ A_i \in \mathcal{A}_k.$$

By Proposition 1, for each A_i in $\mathcal{A}_k$, there exists a code (f_{A_i}, φ_{A_i}), $f_{A_i} : \mathcal{X}_{A_i}^{\nu_{A_i}} \to \{1, \ldots, J_{A_i}\}$ and $\varphi_{A_i} : \{1, \ldots, J_{A_i}\} \to \mathcal{Y}_{\mathcal{M}}^{\nu_{A_i}}$ of rate $\frac{1}{|\nu_{A_i}|}\log J_{A_i} \leq R_{A_i}(\Delta_{A_i}) + \frac{\epsilon'}{2}$ and with

$$\mathbb{E}\left[d\left(X_{\mathcal{M}}^{\nu_{A_i}}, \varphi_{A_i}\left(f_{A_i}(X_{A_i}^{\nu_{A_i}})\right)\right)\right]$$
$$= \mathbb{E}\left[d_{A_i}\left(X_{A_i}^{\nu_{A_i}}, \varphi_{A_i}\left(f_{A_i}(X_{A_i}^{\nu_{A_i}})\right)\right)\right]$$
$$\leq \Delta_{A_i} + \frac{\epsilon'}{2}$$

for all $|\nu_{A_i}| \geq N_{A_i}\left(\frac{\epsilon'}{2}\right)$ (cf. proof of Proposition 1).

Consider a (composite) code (f, φ) as follows. For the deterministic sampling scheme defined above, the encoder f

consists of a concatenation of encoders defined for $x^n \in \overset{M_k}{\underset{i=1}{\times}} \mathcal{X}_{A_i}^{\nu_{A_i}}$ by

$$f(S^n, x^n) = \left(f_{A_1}\left(x_{A_1}^{\nu_{A_1}} \right), \ldots, f_{A_{M_k}}\left(x_{A_{M_k}}^{\nu_{A_{M_k}}} \right) \right),$$

which maps the output of the k-IRS into the set $\mathcal{J} \triangleq \overset{M_k}{\underset{i=1}{\times}} \{1, \ldots, J_{A_i}\}$. The decoder φ is given by

$$\varphi\left(j_1, \ldots, j_{M_k} \right) \triangleq \left(\varphi_{A_1}(j_1), \ldots, \varphi_{A_{M_k}}(j_{M_k}) \right),$$

for $(j_1, \ldots, j_{M_k}) \in \mathcal{J}$, and is aware of the sampling sequence without being informed additionally of it.

The rate of the code is

$$\frac{1}{n} \log \|\mathcal{J}\| = \frac{1}{n} \sum_{i=1}^{M_k} \log J_{A_i}$$

$$\leq \frac{1}{n}\left(\sum_{i=1}^{M_k} |\nu_{A_i}| \left(R_{A_i}(\Delta_{A_i}) + \frac{\epsilon'}{2} \right) \right)$$

$$\leq \sum_{i=1}^{M_k} \left(\left(P_S(A_i) + \frac{1}{n} \right) \left(R_{A_i}(\Delta_{A_i}) + \frac{\epsilon'}{2} \right) \right)$$

$$\leq \sum_{i=1}^{M_k} P_S(A_i) R_{A_i}(\Delta_{A_i}) + \epsilon' < R_i(\Delta) + \epsilon, \quad (30)$$

where the previous inequality holds for all n large enough. Denoting the decoder output by

$$Y_{\mathcal{M}}^n \triangleq \varphi\left(f(S^n, X_S^n) \right),$$

we have that

$$\mathbb{E}[d(X_{\mathcal{M}}^n, Y_{\mathcal{M}}^n)]$$

$$= \mathbb{E}\left[\frac{1}{n} \sum_{t=1}^{n} d(X_{\mathcal{M}t}, Y_{\mathcal{M}t}) \right]$$

$$= \frac{1}{n} \sum_{i=1}^{M_k} |\nu_{A_i}| \mathbb{E}\left[d\left(X_{\mathcal{M}}^{\nu_{A_i}}, \varphi_{A_i}\left(f_{A_i}(X_{A_i}^{\nu_{A_i}}) \right) \right) \right]$$

$$\leq \sum_{i=1}^{M_k} \left(P_S(A_i) + \frac{1}{n} \right) \left(\Delta_{A_i} + \frac{\epsilon'}{2} \right)$$

$$= \sum_{i=1}^{M_k} P_S(A_i) \Delta_{A_i} + \frac{M_k}{n} \left(\frac{\epsilon'}{2} + \Delta_{\max} \right) + \frac{\epsilon'}{2}$$

$$\leq \Delta + \epsilon \quad (31)$$

by (13) and for all n large enough. The proof is completed by noting that (30) and (31) hold simultaneously for all n large enough. □

Next, we establish Theorem 5. The structure of the conditional point-mass sampler therein will be used next in the achievability proof of Theorem 4 to follow.

Theorem 5: Denoting the minima in (15) and (20) by $q(\Delta)$ and $r(\Delta)$, respectively, clearly

$$q(\Delta) \leq r(\Delta), \quad \Delta_{\min} \leq \Delta \leq \Delta_{\max}. \quad (32)$$

In fact, equality will be shown to hold, thereby proving the theorem. First, since $q(\Delta)$ and $r(\Delta)$ are convex in Δ by Lemma 8, by [8, Lemma 8.1] they can be expressed in terms of their Lagrangians as

$$q(\Delta) = \max_{\lambda \geq 0} G_q(\lambda) - \lambda\Delta \quad \text{and} \quad r(\Delta) = \max_{\lambda \geq 0} G_r(\lambda) - \lambda\Delta,$$

where $G_q(\lambda)$ and $G_r(\lambda)$ are the respective minima of

$$I(X_S \wedge Y_{\mathcal{M}}|S, U) + \lambda\mathbb{E}[d(X_{\mathcal{M}}, Y_{\mathcal{M}})] \quad (33)$$

over $\left(P_U, P_{S|X_{\mathcal{M}}U}, P_{Y_{\mathcal{M}}|SX_SU} \right)$ and $\left(P_U, \delta_{h(\cdot)}, P_{Y_{\mathcal{M}}|SX_SU} \right)$. By the conditional version of Topsøe's identity [8, Lemma 8.5], the expression in (33) equals

$$\min_{Q_{Y_{\mathcal{M}}|SU}} D\left(P_{Y_{\mathcal{M}}|SX_SU} \| Q_{Y_{\mathcal{M}}|SU} \middle| P_{SX_SU} \right) + \lambda\mathbb{E}[d(X_{\mathcal{M}}, Y_{\mathcal{M}})]. \quad (34)$$

In $G_q(\lambda)$, the minimum of the expression in (34) also over $\left(P_U, P_{S|X_{\mathcal{M}}U}, P_{Y_{\mathcal{M}}|SX_SU} \right)$ is not altered by changing the order of minimization with $P_{S|X_{\mathcal{M}}U}$ being the innermost. Using this fact, it is shown in Appendix that the minimizing $P_{S|X_{\mathcal{M}}U}$ is of the form $\delta_{h(\cdot)}$, whereby

$$G_q(\lambda) = G_r(\lambda). \quad (35)$$

Hence, equality holds in (32).

□

Theorem 4: By (18), using the result of Theorem 5,

$$R_m^I(\Delta) = \min_{\substack{P_U, \ \Delta_u: \\ \sum_u P_U(u)\Delta_u \leq \Delta}} \sum_{u \in \mathcal{U}} P_U(u)\tilde{R}(\Delta_u) \quad (36)$$

where

$$\tilde{R}(\Delta_u) = \min_{\substack{P_{X_{\mathcal{M}}}\delta_{h_u(\cdot)}P_{Y_{\mathcal{M}}|SX_S} \\ \mathbb{E}[d(X_{\mathcal{M}}, Y_{\mathcal{M}})] \leq \Delta_u}} I(X_S \wedge Y_{\mathcal{M}}|S) \quad (37)$$

for $\Delta_{\min} \leq \Delta, \Delta_u \leq \Delta_{\max}$, with the pmf $P_{X_{\mathcal{M}}}\delta_{h_u(\cdot)}P_{Y_{\mathcal{M}}|SX_S}$ being understood as $P_{X_{\mathcal{M}}}\delta_{h(\cdot,u)}P_{Y_{\mathcal{M}}|SX_S, U=u}$. To simplify notation, the conditioning on $U = u$ will be suppressed except when needed. It suffices to show the existence of a code of rate $\cong \tilde{R}(\Delta_u)$ with distortion $\mathbb{E}[d(X_{\mathcal{M}}, Y_{\mathcal{M}})] \overset{\sim}{\leq} \Delta_u$. A concatenation of such codes indexed by $u \in \mathcal{U}$ yields, in effect, suitable time-sharing among them, leading to the achievability of (36). By Theorem 5, in view of the optimality of point-mass samplers, concatenating fixed-set sampling rate distortion codes for conditional sources $P_{X_{\mathcal{M}}|S=A}$, $A \in \mathcal{A}_k$, will suffice.

Given any $\Delta_{\min} \leq \Delta_u \leq \Delta_{\max}$, for the minimizer in (37), consider the corresponding

$$P_{S|X_{\mathcal{M}}} = \delta_{h_u(\cdot)}, \quad \Delta_{A_i} \triangleq \mathbb{E}[d(X_{\mathcal{M}}, Y_{\mathcal{M}})|S = A_i] \quad \text{and}$$
$$I(X_{A_i} \wedge Y_{\mathcal{M}}|S = A_i), \quad i \in \mathcal{M}_k.$$

The associated $\{(S_t, X_{S_t})\}_{t=1}^{\infty}$ is an i.i.d. sequence (cf. Remark (ii) following Definition 3). The sampling sets characterized by the conditional point-mass sampler above and the DMMS realizations $x_{\mathcal{M}}^n$, are denoted as $s^n(x_{\mathcal{M}}^n) \triangleq \left(s(x_{\mathcal{M}1}), \ldots, s(x_{\mathcal{M}n}) \right)$, and hence $S^n = s^n(X_{\mathcal{M}}^n)$.

The idea behind the remainder of the proof below for each $U = u$ is the following. We collect all those time instants at

which a particular A_i in $\mathcal{A}_k$ is sampled, with the objective of applying a fixed-set sampling rate distortion code. Since the size of this time-set will vary according to $x_{\mathcal{M}}^n$ in $\mathcal{X}_{\mathcal{M}}^n$, the rate of such a code, too, will vary accordingly. However, since we seek fixed rate codes (rather than codes with a desired average rate), we apply fixed-set sampling codes to subsets of predetermined lengths from among typical sampling sequences in $\mathcal{A}_k^n$.

Fix $\epsilon > 0$ and $0 < \epsilon' < \epsilon$. Ordering the elements of $\mathcal{A}_k$ as in the proof of Theorem 2, for $n \geq 1$, the sets $\tau_{s^n}(A_i) \triangleq \{t : 1 \leq t \leq n,\ s_t = A_i\}$, $i \in \mathcal{M}_k$, cover $\{1, \ldots, n\}$; denote the set of the first $\max\{\lceil n(P_S(A_i) - \epsilon')\rceil, 0\}$ time instants in $\tau_{s^n}(A_i)$ by ν_{A_i}. For the (typical) set

$$\mathcal{T}_{\epsilon'}^{(n)} \triangleq \left\{ s^n \in \mathcal{A}_k^n : \left| \frac{|\tau_{s^n}(A_i)|}{n} - P_S(A_i) \right| \leq \epsilon',\ i \in \mathcal{M}_k \right\},$$

$P\left(S^n \in \mathcal{T}_{\epsilon'}^{(n)}\right) \geq 1 - \frac{\epsilon'}{2}$ for all $n \geq N_1(\epsilon')$, say.

Along the lines of proof of Theorem 2, for each DMMS with (conditional) pmf $P_{X_{\mathcal{M}}|S=A_i}$, $i \in \mathcal{M}_k$, there exists a code (f_{A_i}, φ_{A_i}), $f_{A_i} : \mathcal{X}_{A_i}^{\nu_{A_i}} \to \{1, \ldots, J_{A_i}\}$ and $\varphi_{A_i} : \{1, \ldots, J_{A_i}\} \to \mathcal{Y}_{\mathcal{M}}^{\nu_{A_i}}$ of rate $\frac{1}{|\nu_{A_i}|} \log J_{A_i} \leq I(X_{A_i} \wedge Y_{\mathcal{M}}|S = A_i) + \frac{\epsilon'}{2}$ and with

$$\mathbb{E}\left[d\left(X_{\mathcal{M}}^{\nu_{A_i}}, \varphi_{A_i}\left(f_{A_i}(X_{A_i}^{\nu_{A_i}}) \right) \right) \middle| S^{\nu_{A_i}} = A_i^{\nu_{A_i}} \right] \leq \Delta_{A_i} + \frac{\epsilon'}{2}$$

for all $|\nu_{A_i}| \geq N_{A_i}\left(\frac{\epsilon'}{2}\right)$.

A (composite) code (f, φ_S), with f taking values in $\mathcal{J} \triangleq \overset{M_k}{\underset{i=1}{\times}} \{1, \ldots, J_{A_i}\}$ is constructed as follows. The encoder f consists of a concatenation of encoders defined by

$$f\left(s^n(x_{\mathcal{M}}^n); x_{s_1}, \ldots, x_{s_n} \right)$$
$$= \begin{cases} \left(f_{A_1}\left(x_{A_1}^{\nu_{A_1}} \right), \ldots, f_{A_{M_k}}\left(x_{A_{M_k}}^{\nu_{A_{M_k}}} \right) \right), & s^n(x_{\mathcal{M}}^n) \in \mathcal{T}_{\epsilon'}^{(n)}, \\ (1, \ldots, 1), & s^n(x_{\mathcal{M}}^n) \notin \mathcal{T}_{\epsilon'}^{(n)}. \end{cases}$$

For $t = 1, \ldots, n$, and $(j_1, \ldots, j_{M_k}) \in \mathcal{J}$, the informed decoder φ_S is given by

$$\left(\varphi_S\left(s^n, (j_1, \ldots, j_{M_k}) \right) \right)_t$$
$$= \begin{cases} \left(\varphi_{A_i}\left(x_{A_i}^{\nu_{A_i}} \right) \right)_t, & s^n \in \mathcal{T}_{\epsilon'}^{(n)} \text{ and } t \in \nu_{A_i}, i \in \mathcal{M}_k, \\ y_1, & \text{otherwise}, \end{cases}$$

where y_1 is a fixed but arbitrary symbol in $\mathcal{Y}_{\mathcal{M}}$.

The rate of the code is

$$\frac{1}{n} \log \|\mathcal{J}\| = \frac{1}{n} \sum_{i=1}^{M_k} \log J_{A_i}$$
$$\leq \frac{1}{n} \left(\sum_{i=1}^{M_k} |\nu_A| \left(I(X_{A_i} \wedge Y_{\mathcal{M}}|S = A_i) + \frac{\epsilon'}{2} \right) \right)$$
$$\leq \sum_{i=1}^{M_k} P_S(A_i) \left(I(X_{A_i} \wedge Y_{\mathcal{M}}|S = A_i) + \frac{\epsilon'}{2} \right)$$
$$= \sum_{i=1}^{M_k} P_S(A_i) I(X_{A_i} \wedge Y_{\mathcal{M}}|S = A_i) + \frac{\epsilon'}{2}$$
$$\leq I(X_S \wedge Y_{\mathcal{M}}|S) + \epsilon. \tag{38}$$

Defining $d_{\max} \triangleq \max_{(x_{\mathcal{M}}, y_{\mathcal{M}}) \in \mathcal{X}_{\mathcal{M}} \times \mathcal{Y}_{\mathcal{M}}} d(x_{\mathcal{M}}, y_{\mathcal{M}})$, and with $Y_{\mathcal{M}}^n$ denoting the output of the decoder, we have

$$\mathbb{E}[d(X_{\mathcal{M}}^n, Y_{\mathcal{M}}^n)] = \mathbb{E}\left[\mathbb{E}\left[d(X_{\mathcal{M}}^n, Y_{\mathcal{M}}^n)|S^n \right] \right]$$
$$= \sum_{s^n \in \mathcal{T}_{\epsilon'}^{(n)}} P_{S^n}(s^n) \mathbb{E}\left[d(X_{\mathcal{M}}^n, Y_{\mathcal{M}}^n)|S^n = s^n \right]$$
$$\quad + \sum_{s^n \notin \mathcal{T}_{\epsilon'}^{(n)}} P_{S^n}(s^n) \mathbb{E}\left[d(X_{\mathcal{M}}^n, Y_{\mathcal{M}}^n)|S^n = s^n \right]$$
$$\leq \sum_{s^n \in \mathcal{T}_{\epsilon'}^{(n)}} P_{S^n}(s^n)$$
$$\quad \times \sum_{i=1}^{M_k} \frac{|\nu_{A_i}|}{n} \mathbb{E}\left[d\left(X_{\mathcal{M}}^{\nu_{A_i}}, \varphi_{A_i}\left(f_{A_i}(X_{A_i}^{\nu_{A_i}}) \right) \right) \middle| S^{\nu_{A_i}} = A_i^{\nu_{A_i}} \right]$$
$$\quad + \frac{1}{n} \sum_{s^n \in \mathcal{T}_{\epsilon'}^{(n)}} P_{S^n}(s^n)$$
$$\quad \times \sum_{i=1}^{M_k} \sum_{t \in \tau_{s^n}(A_i) \setminus \nu_{A_i}} \mathbb{E}\left[d(X_{\mathcal{M}}, Y_{\mathcal{M}t})|S^n = s^n \right]$$
$$\quad + \sum_{s^n \notin \mathcal{T}_{\epsilon'}^{(n)}} P_{S^n}(s^n) d_{\max}$$
$$\leq \sum_{i=1}^{M_k} P_S(A_i) \left(\Delta_{A_i} + \frac{\epsilon'}{2} \right) + \left(1 - \frac{\sum_{i=1}^{M_k} |\nu_{A_i}|}{n} \right) d_{\max}$$
$$\quad + \frac{\epsilon'}{2} d_{\max}$$
$$\leq \Delta_u + \frac{\epsilon'}{2} + \left(2M_k\epsilon' \right) d_{\max} + \frac{\epsilon'}{2} d_{\max}$$
$$< \Delta_u + \epsilon \tag{39}$$

for all n large enough. The proof is completed by noting that for n large enough (38) and (39) hold simultaneously and time-sharing between the codes corresponding to $U = u$, $u \in \mathcal{U}$, completes the proof. $\qquad \square$

Theorem 6: The proof is similar to that of Proposition 1 with the i.i.d. sequence $\{X_{At}\}_{t=1}^\infty$ replaced by the i.i.d. sequence $\{S_t, X_{S_t}\}_{t=1}^\infty$ with joint pmf P_{SX_S} obtained from (21) and a modified distortion measure $\tilde{d}((s, x_s), y_{\mathcal{M}}) \triangleq \mathbb{E}[d(X_{\mathcal{M}}, y_{\mathcal{M}})|S = s, X_S = x_s]$. The details, identical to those in the achievability proof of Proposition 1, are omitted. $\qquad \square$

B. Converse proofs

Separate converse proofs can be provided for Proposition 1 and Proposition 7. However, in order to highlight the underlying ideas economically, we develop the proofs in a unified manner. Specifically, in contrast with the achievability proofs above, our converse proofs are presented in the order of weakening power of the sampler, viz., k-MRS, k-IRS and fixed-set sampler. We begin with the proof of Lemma 8 followed by pertinent technical results before turning to Proposition 1 and Proposition 7.

Lemma 8: We need to prove only that the right-sides of (6), (11) and (15) are convex and continuous, since they are evidently finite-valued and nonincreasing in Δ. The convexity

of the right-side of (6) on $[\Delta_{\min,A}, \Delta_{\max}]$ is a standard consequence of the convexity of

$$I(X_A \wedge Y_{\mathcal{M}}) = I\left(P_{X_A}, P_{Y_{\mathcal{M}}|X_A}\right)$$

in $P_{Y_{\mathcal{M}}|X_A}$ and the convexity of the constraint set in (6). The convexity of the right-sides of (11) and (15) is immediate by the remarks following Proposition 3 and Theorem 4, and their continuity for $\Delta > \Delta_{\min}$ is a consequence. Continuity at $\Delta = \Delta_{\min}$ in (6), (11) and (15) holds, for instance, as in [8, Lemma 7.2]. □

Lemma 9: Let the finite-valued rvs A^n, B^n, C^n, D^n be such that (A_t, B_t), $t = 1, \ldots, n$, are mutually independent and satisfy

$$B^n \;-\!\circ\!-\; A^n, C^n \;-\!\circ\!-\; D^n \qquad (40)$$

and

$$C_t \;-\!\circ\!-\; A_t, B_t, C^{t-1} \;-\!\circ\!-\; A^{n\backslash t}, B^{n\backslash t}, \quad t = 1, \ldots, n, \quad (41)$$

where $A^{n\backslash t} = A^n \backslash A_t$. Then, the following hold for $t = 1, \ldots, n$:

$$I(A^t, B^t, C^t \wedge A_{t+1}^n, B_{t+1}^n) = 0; \qquad (42)$$

$$A_t, B_t \;-\!\circ\!-\; C^t \;-\!\circ\!-\; A^{n\backslash t}, B^{n\backslash t}, C_{t+1}^n; \qquad (43)$$

and

$$B_t \;-\!\circ\!-\; A_t, C^t \;-\!\circ\!-\; D_t. \qquad (44)$$

Proof: First, (42) is true by the following simple observation: for $t = 1, \ldots, n$,

$$
\begin{aligned}
&I(A^t, B^t, C^t \wedge A_{t+1}^n, B_{t+1}^n) \\
&\quad = I(A^t, B^t \wedge A_{t+1}^n, B_{t+1}^n) \\
&\qquad + I(C^t \wedge A_{t+1}^n, B_{t+1}^n | A^t, B^t) \\
&\quad = 0
\end{aligned} \qquad (45)
$$

where the first term in the sum above is zero by the mutual independence of (A_t, B_t), $t = 1, \ldots, n$, and the second term equals zero by (41). Next, the claim (42) and the Markov property (41) imply that for $t = 1, \ldots, n$,

$$
\begin{aligned}
&I(A_t, B_t, C_t \wedge A^{t-1}, B^{t-1} | C^{t-1}) \\
&\quad = I(A_t, B_t \wedge A^{t-1}, B^{t-1} | C^{t-1}) \\
&\qquad + I(C_t \wedge A^{t-1}, B^{t-1} | A_t, B_t, C^{t-1}) \\
&\quad = 0.
\end{aligned} \qquad (46)
$$

The claim (43) now follows, since

$$
\begin{aligned}
&I(A_t, B_t \wedge A^{n\backslash t}, B^{n\backslash t}, C_{t+1}^n | C^t) \\
&\quad = I(A_t, B_t \wedge A^{t-1}, B^{t-1} | C^t) \\
&\qquad + I(A_t, B_t \wedge A_{t+1}^n, B_{t+1}^n | A^{t-1}, B^{t-1}, C^t) \\
&\qquad + I(A_t, B_t \wedge C_{t+1}^n | A^{n\backslash t}, B^{n\backslash t}, C^t) \\
&\quad = 0
\end{aligned}
$$

where the first term in the sum above is zero by (46), and the latter two terms are zero by (45) and (41), respectively.

Now using (40),

$$
\begin{aligned}
0 &= I(B^n \wedge D_t | A^n, C^n) \\
&= \sum_{t=1}^{n} I(B_t \wedge D_t | B^{t-1}, A^n, C^n) \\
&= \sum_{t=1}^{n} I(B_t \wedge D_t | B^{t-1}, A_t, A^{n\backslash t}, C^t, C_{t+1}^n) \\
&= \sum_{t=1}^{n} \Big[I(B_t \wedge B^{t-1}, A^{n\backslash t}, C_{t+1}^n, D_t | A_t, C^t) \\
&\qquad - I(B_t \wedge B^{t-1}, A^{n\backslash t}, C_{t+1}^n | A_t, C^t) \Big] \\
&= \sum_{t=1}^{n} I(B_t \wedge B^{t-1}, A^{n\backslash t}, C_{t+1}^n, D_t | A_t, C^t) \text{ by (43)} \\
&\geq \sum_{t=1}^{n} I(B_t \wedge D_t | A_t, C^t),
\end{aligned}
$$

so that the claim (44) follows. □

We now prove Proposition 7 which, in effect, implies the converse proofs for Theorem 4, Theorem 2 and Proposition 1. Specifically, a converse is fashioned for $R_{ms}^I(\Delta)$, with those for $R_{is}(\Delta)$ and $R_A(\Delta)$ emerging along the way.

Let $\left(\{P_{S_t|X_{\mathcal{M}t}S^{t-1}}\}_{t=1}^n, f, \varphi_S\right)$ be an n-length strong k-MRS block code with decoder output $Y_{\mathcal{M}}^n = \varphi_S\left(S^n, f(S^n, X_S^n)\right)$ and satisfying $\mathbb{E}\left[d\left(X_{\mathcal{M}}^n, Y_{\mathcal{M}}^n\right)\right] \leq \Delta$. The hypothesis of Lemma 9 with $A^n = X_S^n$, $B^n = X_{S^c}^n$, $C^n = S^n$ and $D^n = Y_{\mathcal{M}}^n$ is met since

$$X_{S^c}^n \;-\!\circ\!-\; S^n, X_S^n \;-\!\circ\!-\; Y_{\mathcal{M}}^n$$

and by (22),

$$P_{S_t|X_{\mathcal{M}}^n S^{t-1}} = P_{S_t|X_{\mathcal{M}t}S^{t-1}}.$$

Then by Lemma 9, for $t = 1, \ldots, n$,

$$I(S^{t-1} \wedge X_{\mathcal{M}t}) = 0, \qquad (47)$$
$$X_{\mathcal{M}t} \;-\!\circ\!-\; S^t \;-\!\circ\!-\; X_{\mathcal{M}}^{n\backslash t}, X_S^{n\backslash t}, S_{t+1}^n \qquad (48)$$

and

$$X_{S_t^c} \;-\!\circ\!-\; S^t, X_{S_t} \;-\!\circ\!-\; Y_{\mathcal{M}t}. \qquad (49)$$

Denoting by $||f||$ the cardinality of the range space of the encoder f, the rate R of the code satisfies

$$
\begin{aligned}
nR &= \log||f|| \geq H\left(f(S^n, X_S^n)\right) \\
&\geq H\left(\varphi_S\left(S^n, f(S^n, X_S^n)\right)|S^n\right) = H\left(Y_{\mathcal{M}}^n|S^n\right) \\
&= H\left(Y_{\mathcal{M}}^n|S^n\right) - H\left(Y_{\mathcal{M}}^n|S^n, X_S^n\right) \\
&= I\left(X_S^n \wedge Y_{\mathcal{M}}^n|S^n\right) \\
&= \sum_{t=1}^{n} \Big(H(X_{S_t}|S^t, S_{t+1}^n, X_S^{t-1}) \\
&\qquad - H(X_{S_t}|S^t, S_{t+1}^n, X_S^{t-1}, Y_{\mathcal{M}}^n) \Big) \\
&\geq \sum_{t=1}^{n} \Big(H(X_{S_t}|S^t) - H(X_{S_t}|S^t, Y_{\mathcal{M}t}) \Big) \qquad (50) \\
&= \sum_{t=1}^{n} I(X_{S_t} \wedge Y_{\mathcal{M}t}|S^t) \qquad (51)
\end{aligned}
$$

where (50) follows from (48). Denote $\mathbb{E}[d(X_{\mathcal{M}t}, Y_{\mathcal{M}t})]$ by Δ_t.

For the strong k-MRS code above, in (51) using (47) and (49), we get

$$I(X_{S_t} \wedge Y_{\mathcal{M}t}|S^t)$$
$$\geq \min I(X_{S_t} \wedge Y_{\mathcal{M}t}|S_t, S^{t-1}) \tag{52}$$
$$\geq \min_{\substack{P_{U_t} P_{X_{\mathcal{M}t}} P_{S_t|X_{\mathcal{M}t}U_t} P_{Y_{\mathcal{M}t}|S_t X_{S_t}U_t} \\ \mathbb{E}[d(X_{\mathcal{M}t}, Y_{\mathcal{M}t})]\leq\Delta_t}} I(X_{S_t} \wedge Y_{\mathcal{M}t}|S_t, U_t), \tag{53}$$

where the minimum in (52) is with respect to $P_{X_{\mathcal{M}t} S^t Y_{\mathcal{M}t}} = P_{S^{t-1}} P_{X_{\mathcal{M}t}} P_{S_t|X_{\mathcal{M}t}S^{t-1}} P_{Y_{\mathcal{M}t}|S_t X_{S_t}S^{t-1}}$ and $\mathbb{E}[d(X_{\mathcal{M}t}, Y_{\mathcal{M}t})] = \Delta_t$, and where U_t is a rv taking values in a set of cardinality $|\mathcal{A}_k|^{t-1}$. The existence of the minima in (52) and (53) comes from the continuity of the conditional mutual information terms over compact sets of pmfs.

By the Carathéodory theorem [7], every point in the convex hull of the set

$$\mathcal{C} = \left\{ \big(\mathbb{E}[d(X_{\mathcal{M}}, Y_{\mathcal{M}})], I(X_S \wedge Y_{\mathcal{M}}|S)\big) : \right.$$
$$\left. X_{\mathcal{M}} \multimap S, X_S \multimap Y_{\mathcal{M}} \right\} \subset \mathbb{R}^2$$

can be represented as a convex combination of at most three points in $\mathcal{C}$. Hence, to describe every element in the set

$$\left\{ \big(\mathbb{E}[d(X_{\mathcal{M}t}, Y_{\mathcal{M}t})], I(X_{S_t} \wedge Y_{\mathcal{M}t}|S_t, U_t)\big) : \right.$$
$$\left. P_{U_t X_{\mathcal{M}t} S_t Y_{\mathcal{M}t}} = P_{U_t} P_{X_{\mathcal{M}t}} P_{S_t|X_{\mathcal{M}t}U_t} P_{Y_{\mathcal{M}t}|S_t X_{S_t}U_t} \right\},$$

it suffices to consider a rv U_t with support of size three. (For $t = 1$, this assertion is straightforward.) Consequently, the right-side of (53) equals $R_m^I(\Delta_t)$ (cf. (15)). Using the convexity of $R_m^I(\Delta)$ in Δ, we get from (51) that

$$nR \geq \sum_{t=1}^{n} R_m^I(\Delta_t)$$
$$\geq nR_m^I\left(\frac{1}{n}\sum_{t=1}^{n}\Delta_t\right)$$
$$\geq nR_m^I(\Delta), \tag{54}$$

i.e., $R \geq R_m^I(\Delta)$, $\Delta \geq \Delta_{\min}$, thereby completing the converse proof for a strong k-MRS and Theorem 4.

Next, an n-length strong k-IRS code and fixed-set sampler code can be viewed as restrictions of the strong k-MRS code above. Specifically, the strong k-IRS and fixed-set sampler respectively entail replacing $P_{S_t|X_{\mathcal{M}t}S^{t-1}}$ by $P_{S_t|S^{t-1}}$ and $P_{S_t} = \mathbb{1}(S_t = A)$. Counterparts of (52) and (53) hold with the mentioned replacements. For a strong k-IRS, upon replacing $P_{S_t|X_{\mathcal{M}t}S^{t-1}}$ with $P_{S_t|S^{t-1}}$, we observe that the right-side of (53), viz.

$$\min_{\substack{P_{U_t} P_{X_{\mathcal{M}t}} P_{S_t|U_t} P_{Y_{\mathcal{M}t}|S_t X_{S_t}U_t} \\ \mathbb{E}[d(X_{\mathcal{M}t}, Y_{\mathcal{M}t})]\leq\Delta_t}} I(X_{S_t} \wedge Y_{\mathcal{M}t}|S_t, U_t)$$

is now the lower convex envelope of the SRDf for a k-IRS, already convex in distortion, and hence, equals $R_i(\Delta_t)$ itself. Thus, (53) becomes

$$I(X_{S_t} \wedge Y_{\mathcal{M}t}|S^t) \geq \min_{\substack{P_{X_{\mathcal{M}t}} P_{S_t} P_{Y_{\mathcal{M}t}|S_t X_{S_t}} \\ \mathbb{E}[d(X_{\mathcal{M}t}, Y_{\mathcal{M}t})]\leq\Delta_t}} I(X_{S_t} \wedge Y_{\mathcal{M}t}|S_t)$$
$$= R_i(\Delta_t). \tag{55}$$

Combining (51) and (55), we get along the lines of (54) that $R \geq R_i(\Delta)$, $\Delta \geq \Delta_{\min}$, which gives the converse proof for a strong k-IRS and Theorem 2.

In a manner analogous to a strong k-IRS, for a fixed-set sampler the convexity of $R_A(\Delta)$ in Δ implies that the counterpart of the right-side of (53), with $P_{S_t|X_{\mathcal{M}t}U_t}$ replaced by $\mathbb{1}(S_t = A)$, simplifies to $R_A(\Delta_t)$. As in (54), it follows that $R \geq R_A(\Delta)$, $\Delta \geq \Delta_{\min,A}$, which gives the converse for Proposition 1.

$\square$

V. CONCLUSION

Our new framework of sampling rate distortion describes the centralized sampling of fixed-size subsets of the components of a DMMS, followed by encoding and lossy reconstruction of the full DMMS. Specifically, we examine the tradeoffs between sampling strategy, optimal encoding rate and distortion in reconstruction as characterized by a sampling rate distortion function. Three sampling strategies are considered: fixed-set sampling, independent random sampling and memoryless random sampling; in the latter two settings, the decoder may or may not be informed of the sampling sequence.

Single-letter characterizations of the SRDf are provided for the sampling strategies above but for a memoryless random sampler with uninformed decoder. In the last case, an achievability proof yields an upper bound for the SRDf whose tightness is unknown. This upper bound in Theorem 6 can be convexified by means of a time-sharing random variable whereupon the modified bound becomes tight. However, it remains open whether such time-sharing is necessary for convexification.

APPENDIX
PROOF OF (35)

We show that the minimum of (34) with respect to $\big(P_U, P_{S|X_{\mathcal{M}}U}, P_{Y_{\mathcal{M}}|SX_SU}\big)$ is attained by $P_{S|X_{\mathcal{M}}U}$ of the form $\delta_{h(\cdot)}$. The Lagrangian is

$$G_q(\lambda) = \min_{\substack{P_U, P_{S|X_{\mathcal{M}}U}, \\ P_{Y_{\mathcal{M}}|SX_SU}, Q_{Y_{\mathcal{M}}|SU}}} D\big(P_{Y_{\mathcal{M}}|SX_SU}\big|\big|Q_{Y_{\mathcal{M}}|SU}\big|P_{SX_SU}\big)$$
$$+ \lambda\mathbb{E}\big[d(X_{\mathcal{M}}, Y_{\mathcal{M}})\big]$$
$$= \min_{\substack{P_U, Q_{Y_{\mathcal{M}}|SU}, \\ P_{Y_{\mathcal{M}}|SX_SU}}} \sum_{u, x_{\mathcal{M}}} P_{UX_{\mathcal{M}}}(u, x_{\mathcal{M}})$$
$$\times \min_{P_{S|X_{\mathcal{M}}U}} \sum_{s\in\mathcal{A}_k} P_{S|X_{\mathcal{M}}U}(s|x_{\mathcal{M}}, u)\eta(u, x_{\mathcal{M}}, s)$$

where

$$\eta(u, x_{\mathcal{M}}, s) = \mathbb{E}\left[\log \frac{P_{Y_{\mathcal{M}}|SX_SU}(Y_{\mathcal{M}}|s, x_s, u)}{Q_{Y_{\mathcal{M}}|SU}(Y_{\mathcal{M}}|s, u)}\right.$$
$$\left. + \lambda d(x_{\mathcal{M}}, Y_{\mathcal{M}})\Big|S = s, X_S = x_s, U = u\right]$$

with the expectation being over $P_{Y_{\mathcal{M}}|S=s, X_S=x_s, U=u}$. Since $\eta = \eta(u, x_{\mathcal{M}}, s)$ we get

$$G_q(\lambda) = \min_{\substack{P_U, Q_{Y_{\mathcal{M}}|SU}, \\ P_{Y_{\mathcal{M}}|SX_SU}}} \sum_{u, x_{\mathcal{M}}} P_{UX_{\mathcal{M}}}(u, x_{\mathcal{M}})$$
$$\min_{s\in\mathcal{A}_k} \eta(u, x_{\mathcal{M}}, s)$$

$$
\begin{aligned}
&= \min_{\substack{P_U, Q_{Y_{\mathcal{M}}|SU}, \\ P_{Y_{\mathcal{M}}|SX_SU}}} \sum_{u, x_{\mathcal{M}}} P_{UX_{\mathcal{M}}}(u, x_{\mathcal{M}}) \\
&\quad \times \min_{\delta_{h(\cdot)}} \sum_{s \in \mathcal{A}_k} \delta_{h(x_{\mathcal{M}}, u)}(s)\eta(u, x_{\mathcal{M}}, s) \\
&= \min_{\substack{P_U, \delta_{h(\cdot)}, Q_{Y_{\mathcal{M}}|SU}, \\ P_{Y_{\mathcal{M}}|SX_SU}}} D\left(P_{Y_{\mathcal{M}}|SX_SU} \| Q_{Y_{\mathcal{M}}|SU} \big| P_{SX_SU}\right) \\
&\quad + \lambda \mathbb{E}\big[d(X_{\mathcal{M}}, Y_{\mathcal{M}})\big] \\
&= G_r(\lambda).
\end{aligned}
$$

ACKNOWLEDGEMENT

REFERENCES

[1] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression.* Englewood Cliffs, NJ, USA: Prentice-Hall, 1971.

[2] T. Berger, "Multiterminal source coding," *Inf. Theory Approach Commun.*, vol. 229, pp. 171–231, Jul. 1977.

[3] V. P. Boda and P. Narayan, "Sampling rate distortion," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2014, pp. 3057–3061.

[4] V. P. Boda and P. Narayan, "Memoryless sampling rate distortion," in *Proc. 53rd Annu. Allerton Conf. Commun., Control, Comput.*, Oct. 2015, pp. 919–923.

[5] V. P. Boda and P. Narayan, "Independent and memoryless sampling rate distortion," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 2968–2972.

[6] Z. Chen, J. Ranieri, R. Zhang, and M. Vetterli, "DASS: Distributed adaptive sparse sensing," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2571–2583, May 2015.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* Hoboken, NJ, USA: Wiley, 2012.

[8] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* Cambridge, U.K.: Cambridge Univ. Press, 2011.

[9] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," *IRE Trans. Inf. Theory*, vol. 8, no. 5, pp. 293–304, Sep. 1962.

[10] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[11] N. M. Freris, H. Kowshik, and P. R. Kumar, "Fundamentals of large sensor networks: Connectivity, capacity, clocks, and computation," *Proc. IEEE*, vol. 98, no. 11, pp. 1828–1846, Nov. 2010.

[12] A. Giridhar and P. R. Kumar, "Computing and communicating functions over sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 23, pp. 755–764, Apr. 2005.

[13] A. Hormati, O. Roy, Y. M. Lu, and M. Vetterli, "Distributed sampling of signals linked by sparse filtering: Theory and applications," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1095–1109, Mar. 2010.

[14] P. Ishwar, A. Kumar, and K. Ramachandran, "On distributed sampling in dense sensor networks: A 'bit conservation principle,'" in *Proc. Int. Symp. Inf. Process. Sensor Netw.*, Apr. 2003, pp. 17–31.

[15] A. Kashyap, L. A. Lastras-Montano, C. Xia, and L. Zhen, "Distributed source coding in dense sensor networks," in *Proc. Data Compress. Conf. (DCC)*, Mar. 2005, pp. 13–22.

[16] T. Kawabata and A. Dembo, "The rate-distortion dimension of sets and measures," *IEEE Trans. Inf. Theory*, vol. 40, no. 5, pp. 1564–1572, Dec. 1994.

[17] A. Kipnis, A. J. GoldSmith, Y. C. Eldar, and T. Weissman, "Distortion rate function of sub-Nyquist sampled Gaussian sources," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 401–429, Jan. 2016.

[18] R. L. Konsbruck, E. Telatar, and M. Vetterli, "On sampling and coding for distributed acoustic sensing," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3198–3214, May 2012.

[19] X. Liu, O. Simeone, and E. Erkip, "Lossy computing of correlated sources with fractional sampling," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Sep. 2012, pp. 232–236.

[20] J. Long, S. O. Memik, G. Memik, and R. Mukherjee, "Thermal monitoring mechanisms for chip multiprocessors," *ACM Trans. Archit. Code Optim.*, vol. 5, no. 2, p. 9, Aug. 2008.

[21] D. L. Neuhoff and S. S. Pradhan, "Information rates of densely sampled Gaussian data," in *Proc. IEEE Int. Symp. Inf. Theory Process. (ISIT)*, Aug. 2011, pp. 2776–2780.

[22] D. L. Neuhoff and S. S. Pradhan, "Rate-distortion behavior at low distortion for densely sampled Gaussian data," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2012, pp. 358–362.

[23] S. S. Pradhan and D. L. Neuhoff, "Transform coding of densely sampled Gaussian data," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2007, pp. 1111–1114.

[24] J. Ranieri, A. Chebira, and M. Vetterli, "Near-optimal sensor placement for linear inverse problems," *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1135–1146, Mar. 2014.

[25] G. Reeves and M. Gastpar, "The sampling rate-distortion tradeoff for sparsity pattern recovery in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3065–3092, May 2012.

[26] J. Z. Sun and V. K. Goyal, "Intersensor collaboration in distributed quantization networks," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3931–3942, Sep. 2013.

[27] J. Unnikrishnan and M. Vetterli, "Sampling and reconstructing spatial fields using mobile sensors," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 3789–3792.

[28] J. Unnikrishnan and M. Vetterli, "On sampling a high-dimensional bandlimited field on a union of shifted lattices," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2012, pp. 1468–1472.

[29] J. Unnikrishnan and M. Vetterli, "Sampling high-dimensional bandlimited fields on low-dimensional manifolds," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2103–2127, Apr. 2013.

[30] C. Weidmann and M. Vetterli, "Rate distortion behavior of sparse sources," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 4969–4992, Aug. 2012.

[31] Y. Wu and S. Verdú, "Optimal phase transitions in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6241–6263, Oct. 2012.

[32] Y. Wu and S. Verdú, "Rényi information dimension: Fundamental limits of almost lossless analog compression," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3721–3748, Aug. 2010.

[33] H. Yamamoto and K. Itoh, "Source coding theory for multiterminal communication systems with a remote source," *IEICE Trans.*, vol. E63, no. 10, pp. 700–706, Oct. 1980.

[34] Y. Zhang, A. Srivastava, and M. Zahran, "On-chip sensor-driven efficient thermal profile estimation algorithms," *ACM Trans. Design Autom. Electron. Syst.*, vol. 15, no. 3, p. 25, May 2010.

Vinay Praneeth Boda received the Bachelor of Technology degree in Electrical Engineering and the Master of Technology degree in Communication Systems, both from the Indian Institute of Technology, Madras, India in 2011. He is currently a Ph.D. candidate at the University of Maryland, College Park, USA.

Prakash Narayan received the Bachelor of Technology degree in Electrical Engineering from the Indian Institute of Technology, Madras in 1976. He received and the M.S. degree in Systems Science and Mathematics in 1978, and the D.Sc. degree in Electrical Engineering, both from Washington University, St. Louis, MO.

He is Professor of Electrical and Computer Engineering at the University of Maryland, College Park, with a joint appointment at the Institute for Systems Research. His research interests are in multiuser information theory, communication theory, communication networks, cryptography, and information theory and statistics.