On Low-Complexity Decodable Universally Good Linear Codes

Todd P. Coleman
University of Illinois at Urbana-Champaign
Urbana, Il 61801
colemant@uiuc.edu

Muriel Médard
Massachusetts Institute of Technology
Cambridge, MA 02139
medard@mit.edu

***Abstract*— Here we discuss the universal block decoding problem for memoryless systems and focus on figures of merit and linear code constructions that facilitate the analysis and construction of low-complexity decoding algorithms. We discuss the properties of 'universally good codes' and how such codes lie on the Gilbert-Varshamov bound. We discuss analogues of the minimum-distance criterion and develop conditions for universal decoding success. We illustrate that universal decoding over general linear codes is NP-complete. Next we show the fragility of the pessimistic complexity result by considering bipartite graph code constructions and illustrating that with large enough fixed degree, linear codes based on graphs and inner codes that are universally good become universally good aggregately. This creates opportunities for universal decoding algorithms with polynomial complexity and exponential error probability.**

I. Introduction

In the information theory literature there has been discussion on *universal* coding, where encoders and decoders are constructed that operate *without knowledge of the underlying probability distribution*. From an ontological perspective, there has been much success - it has been shown that for numerous settings [1], [2], [3], there exist block encoders and decoders that can attain the same error exponent (exponential rate of decay in probability of error) as that of the random coding exponent corresponding to maximum-likelihood decoding. Such universal decoding algorithms have also served as subcomponents of other multiterminal communication systems - for instance statistical inference problems under data compression constraints [4], [5]. As in the typical channel coding case, the encoding situation is not nearly as difficult as the decoding situation. Indeed, the proposed universal decoding algorithms' nonlinearities and difficulty of implementation have obfuscated the desire for people to construct *practical* code constructions and decoding algorithms. This apparent intrinsic difficulty in universal decoding manifests itself in the consideration of other decoding algorithms [6, Sec. I]: "Theoretically, one can employ universal decoding; however, in many applications, it is ruled out by complexity considerations."

However, we take a fresh perspective by looking back at how key advances manifested themselves in the traditional coding literature:

- Linear codes have been known to be sufficient for many channel coding problems to attain all achievable rates and the random coding exponent. However, ML decoding for general linear codes has been shown [7] to be an intrinsically complex (NP-complete) problem.
- A 'divide-and-conquer' approach has been employed by coding theorists since the beginnings of information theory to construct large linear codes from smaller good components with polynomial complexity decoding algorithms whose performance is empirically good [8] as well as provably good [9], [10].

and try to walk an analogous path to address practical code constructions and decoding algorithms for the universal setting:

- Linear codes have already been known to *universally* attain all achievable rates with exponential error probability decay. In this chapter we show that universal decoding general linear codes is also a provably complex (NP-complete) problem.
- We employ a 'divide-and-conquer' graph-based approach to show that large linear codes constructed from smaller 'universally good' component codes are aggregately provably good under optimal universal decoding.

Although not discussed here, these approaches can be directly extended to multiterminal settings, including Slepian-Wolf near-lossless distributed data compression and certain types of multiple-access channel coding. Hopefully these results will contribute to forming a basis for efficient, practical code and decoding designs for the universal setting.

II. The General Universal Decoding Problem for a Single Source

A. Model and Definitions

Throughout this discussion we will consider a discrete memoryless source (DMS) U over $\mathcal{U} = \{0, 1, \ldots, Q-1\}$. The set of all probability distributions on $\mathcal{U}$ is given by $\mathcal{P}(\mathcal{U})$. For a length-N sequence $\underline{u} = (u_1, u_2, \ldots, u_N) \in \mathcal{U}^N$, the type $P_{\underline{u}} \in \mathcal{P}(\mathcal{U})$ is the probability distribution defined by $P_{\underline{u}}(a) = \frac{1}{N}\sum_{i=1}^{N} 1_{\{u_i=a\}}$, for all $a \in \mathcal{U}$. We denote by W^N the pmf induced on $\mathcal{U}^N$ by N independent drawings according to W. We denote by $\mathcal{P}_N(\mathcal{U})$ the subset of $\mathcal{P}(\mathcal{U})$ consisting of the possible types of sequences $\underline{u} \in \mathcal{U}^N$. For any type $\mathbb{P} \in \mathcal{P}_N(\mathcal{U})$, the type class $T(\mathbb{P})$ is the set of all $\underline{u} \in \mathcal{U}^N$

such that $P_{\underline{u}} = \mathbb{P}$. To summarize, we have that

$$\begin{aligned}
\mathcal{P}(\mathcal{U}) &= \left\{ P = \left(\{P_a\}_{a\in\mathcal{U}}\right) : P \geq \underline{0}, \sum_{a\in\mathcal{U}} P_a = 1 \right\} \\
P_{\underline{u}} &= \left(\left\{ \frac{1}{N} \sum_{i=1}^{N} 1_{u_i=a} \right\}_{a\in\mathcal{U}} \right) \text{ for } \underline{u} \in \mathcal{U}^N \qquad (1) \\
\mathcal{P}_N(\mathcal{U}) &= \left\{ P \in \mathcal{P}(\mathcal{U}) : P = P_{\underline{u}} \text{ for some } \underline{u} \in \mathcal{U}^N \right\} \\
T(\mathbb{P}) &= \left\{ \underline{u} \in \mathcal{U}^N \mid P_{\underline{u}} = \mathbb{P} \right\}.
\end{aligned}$$

For a random variable U with probability distribution W we will denote its entropy as $H(U)$ which is a function of W. When we instead want to explicitly speak of the entropy as a function of some $P \in \mathcal{P}_N(\mathcal{U})$, then we will denote this as $h(P)$. For two random variables with conditional and marginal distributions given by $P_{U|V}$ and P_V, we will denote the conditional entropy $H(U|V)$ explicitly interms of $P_{U|V}$ and P_V as $h\left(P_{U|V}|P_V\right)$.

From [11] we note the following:

$$\begin{aligned}
|\mathcal{P}_N(\mathcal{U})| &\leq (N+1)^{|\mathcal{U}|} \qquad (2) \\
|T(\mathbb{P})| &\leq 2^{Nh(\mathbb{P})} \qquad (3) \\
W^n(\underline{u}) &= 2^{-N\left[h\left(P_{\underline{u}}\right)+D\left(P_{\underline{u}}\|W\right)\right]} \quad \forall\, \underline{u} \in \mathcal{U}^N \qquad (4)
\end{aligned}$$

Thus ***the number of types is polynomial in*** N.

B. The General Problem

In this discussion we consider code constructions for fixed block length universal coding for the two dual settings of data compression and channel coding. The compression scenario mentioned could be relevant, for instance, in a wireless sensor network where the following two points apply:

1) Time-varying nature of field makes knowledge of field being sensed, the probability distribution on the data is not completely accurately modeled,
2) Complexity, memory, and energy constraints make a universal fixed-to-fixed length algebraic compression approach more viable than a universal fixed-to-variable length compression approach (such as Lempel-Ziv [12], [13] or Burrows-Wheeler [14]) that requires dictionaries and table-lookups.

Similarly, due to the time-varying and multipath effects of the wireless channel, the universal channel coding scenario could be relevant where phase information cannot be accurately tracked.

More specifically, we take interest in universal decoding for discrete memoryless settings, where the decoder does not have knowledge of the probability distribution to aid in decoding. Consider a DMS U with probability distribution $W \in \mathcal{P}(\mathcal{U})$. Without loss of generality, we assume that $\mathcal{U} = \{0, 1, \ldots, Q-1\}$ where $Q = 2^t$ for some integer $t \geq 1$. Thus we may assume that U takes on values in $\mathbb{F}_{2^t}$. Our goal is to design a fixed-rate universal code that permits a decoding algorithm that is blind to W to have provably good performance. We consider the case where a linear mapping

$$H = \begin{bmatrix} -H_1'- \\ -H_2'- \\ \vdots \\ -H_M'- \end{bmatrix} : \mathcal{U}^N \to \mathcal{U}^M$$

is used to map $\underline{u} \in \mathcal{U}^N$ to $\underline{s} \in \mathcal{U}^M$ via

$$\underline{s} = H\underline{u} \qquad (5)$$

where $M < N$ and U is memoryless with probability distribution $W \in \mathcal{P}(\mathcal{U})$. We will denote the rate R as

$$R = t\frac{M}{N} \qquad (6)$$

and note that this corresponds to rate in a data compression sense and *not* in a channel coding sense (which would correspond to $t - R$). Throughout the rest of this chapter we will speak of rate in a data compression sense. The decoder knows that $\underline{u}$ must be consistent with $\underline{s}$, in other words it must lie in the coset

$$\mathrm{Co}(H, \underline{s}) = \{\underline{u} \mid H\underline{u} = \underline{s}\}, \qquad (7)$$

and selects $\hat{\underline{u}}$ as the 'best' coset member (in a universal sense). This encompasses two settings:

a) Fixed-to-fixed length near-lossless data compression, where $\underline{u}$ is identified as the sourceword and $\underline{s}$ is the syndrome, the output of the compression operation.
b) An additive noise channel $\underline{y} = \underline{x} \oplus \underline{u}$. By using a linear code $\mathcal{C}$ for $\underline{x}$, and identifying the parity check matrix H with $\mathcal{C}$ as

$$\mathcal{C} = \{\underline{x} : H\underline{x} = \underline{0}\}, \qquad (8)$$

then we have that a sufficient statistic for decoding is

$$H\underline{y} = H\underline{u} = \underline{s}.$$

Successfully decoding for the noise vector $\underline{u}$ is equivalent to successfully decoding for the transmitted codeword $\underline{x}$:

$$\hat{\underline{x}} = \hat{\underline{u}} \oplus \underline{y}.$$

We assume that the rate R is achievable (i.e. $t\frac{M}{N} > H(U)$). It has been known in the information theory literature for quite a while [1], [2] that in the *universal* setting, *linear codes* still suffice to attain all achievable rates and can the same error exponent as the random coding exponent. Note that for any $\underline{u} \in \mathcal{U}^N$, we have that

$$P(\underline{u}) = 2^{-N\left[D\left(P_{\underline{u}}\|W\right)+h\left(P_{\underline{u}}\right)\right]}.$$

Thus an ML decoder with knowledge of W operates by selecting

$$\hat{\underline{u}} \in \arg \min_{\underline{u}\in \mathrm{Co}(H,\underline{s})} D\left(P_{\underline{u}}\|W\right) + h\left(P_{\underline{u}}\right)$$

Csiszár's 'minimum-entropy' decoder selects as the source reconstruction the coset's entropy minimizer

$$\hat{\underline{u}} \in \arg \min_{\underline{u}\in \mathrm{Co}(H,\underline{s})} h\left(P_{\underline{u}}\right). \qquad (9)$$

In [2], Csiszár shows that not only do there exist linear codes such whose rates can be arbitrarily close to $H(U)$ when such a decoder is applied, but also that minimum entropy decoding achieves the same error exponent as the optimal maximum-likelihood (ML) decoder. Another interpretation of the universal decoding paradigm is that it is a manifestation of Occam's razor: "Find the explanation most easy to accept." Since the entropy function measures the inherent uncertainty or difficulty in explaining something, selecting the lowest entropy candidate consistent with observations is the same as selecting the easiest to accept candidate consistent with observations.

III. Universally Good Linear Codes

Csiszár's lemma specifying good encoders [2, Sec. III] illustrates the existence of linear mappings $H: \mathcal{U}^N \to \mathcal{U}^M$ such for any joint type $\mathbb{P} \in \mathcal{P}_N\left(\mathcal{U}^2\right)$ with the definitions

$$\mathcal{N}_H\left(\mathbb{P}\right) \triangleq \left|\left\{(\underline{u} \in \mathcal{U} \;\middle|\; \begin{array}{c} H\underline{u} = H\tilde{\underline{u}} \\ P_{\underline{u},\tilde{\underline{u}}} = \mathbb{P} \end{array} \text{ for some } \tilde{\underline{u}} \neq \underline{u}\right\}\right| \quad (10)$$

every joint type $\mathbb{P} \in \mathcal{P}_N\left(\mathcal{U}^2\right)$ satisfies:

$$a) \qquad \mathcal{N}_H\left(\mathbb{P}\right) \leq 2^{-N(R-h(\mathbb{P})-\delta_N)} \quad (11)$$

$$b) \quad \text{if } h\left(\mathbb{P}_{U-\tilde{U}}\right) \leq R-\delta_N \text{ then } \mathcal{N}_H\left(\mathbb{P}\right) = 0 \quad (12)$$

where $\delta_N \to 0$ as $N \to \infty$. We will denote such codes as *universally good*. Note that the bound (11) can be strengthened to:

$$\begin{aligned} \mathcal{N}_H\left(\mathbb{P}\right) &\leq 2^{-N(R-h(\mathbb{P})-\delta_N)} \\ &= 2^{N\left(h(\mathbb{P}_U)-\left(R-h\left(\mathbb{P}_{\tilde{U}|U}|\mathbb{P}_U\right)-\delta_N\right)\right)} \\ \Rightarrow \mathcal{N}_H\left(\mathbb{P}\right) &\leq 2^{N\left(h(\mathbb{P}_U)-\left|R-h\left(\mathbb{P}_{\tilde{U}|U}|\mathbb{P}_U\right)-\delta_N\right|^+\right)} \quad (13) \\ &\leq 2^{N\left(h(\mathbb{P}_U)-\left|R-h\left(\mathbb{P}_{\tilde{U}}\right)-\delta_N\right|^+\right)} \end{aligned}$$

where (13) follows because by the definition of $\mathcal{N}_H\left(\mathbb{P}\right)$, $\mathcal{N}_H\left(\mathbb{P}\right) \leq \left|T\left(\mathbb{P}_U\right)\right| \leq 2^{Nh(\mathbb{P}_U)}$.

A. The Gilbert-Varshamov Distance

One important property of any linear code $\mathcal{C}$ with associated parity-check matrix H is its minimum distance

$$d_{\min}(H) = \min_{\underline{u} \in \mathrm{Co}(H,\underline{0}) \setminus \underline{0}} w_h\left(\underline{u}\right) \quad (14)$$

where $w_h\left(\cdot\right)$ is the Hamming distance. It is well known that the larger the minimum distance of a code, the larger the number of errors it can guarantee to correct:

$$w_h\left(\underline{u}\right) < \frac{1}{2} d_{\min}(H) \Rightarrow w_h\left(\underline{u}\right) < w_h\left(\underline{u}+\tilde{\underline{u}}\right) \forall \tilde{\underline{u}} \in \mathcal{C} \setminus \underline{0}. \quad (15)$$

Here we briefly note how condition (12) of universally good codes relates to a standard bound on good linear codes. It has been well known that random linear codes with parity-check matrix H have minimum distance lying on the Q-ary Gilbert-Varshamov distance bound with high probability [15, p. 42-43]:

$$d_{\min}(H) \geq N\left(h_Q^{-1}(R)-\epsilon\right)$$

where $h_Q(\alpha)$ for $0 < \alpha \leq \frac{Q-1}{Q}$ is given by

$$h_Q(x) = x\log(Q-1) - x\log x - (1-x)\log(1-x).$$

Lemma 3.1: Universally Good Linear Codes lie on the Gilbert-Varshamov bound.

Proof: Setting $\tilde{\underline{u}} = \underline{0}$ we have from condition (12) of universally good codes that any $\underline{u} \in \mathrm{Co}\left(H,\underline{0}\right) \setminus \underline{0}$ satisfies $h\left(P_{\underline{u}}\right) \geq R-\epsilon_N$, where $\epsilon_N \to 0$ as $N \to \infty$. Now if we see what this means in terms of Hamming distance, we can perform the following minimization:

$$\begin{aligned} \min \quad & w_h\left(\underline{u}\right) \\ s.t. \quad & h\left(P_{\underline{u}}\right) \geq R-\epsilon_N. \end{aligned}$$

From the concavity of the entropy function $h\left(\cdot\right)$, $\underline{u}^*$ will in $(1-\delta)N$ positions be 0 and in $\frac{\delta}{Q-1}N$ positions be a, for each $a \in \mathcal{U} \setminus 0$. Thus we have that

$$\begin{aligned} R-\epsilon_N &= h\left(P_{\underline{u}^*}\right) \\ &= -(1-\delta)\log(1-\delta) - (Q-1)\frac{\delta}{Q-1}\log\left(\frac{\delta}{Q-1}\right) \\ &= -(1-\delta)\log(1-\delta) - \delta\log\left(\frac{\delta}{Q-1}\right) \\ &= -(1-\delta)\log(1-\delta) - \delta\log\delta + \delta\log(Q-1) \\ &= h_Q(\delta). \end{aligned}$$

■

B. Guarantees on Universal Decoding Success

Here we discuss some conditions for guarantees on universal decoding success. We present these to fall in analogy with previously well-established conditions for minimum-distance decoding. Analogous to the use of (14) for ML decoding on general linear codes, we will speak to condition (12) for universal decoding on general linear codes. Define the *universal rate* R_{univ} associated with matrix H to be the largest R such that condition (12) holds for H with $\delta_N = 0$.

Lemma 3.2: Consider any linear matrix H, that is used to map $\underline{u}$ to $\underline{s}$ according to (5), and its associated R_{univ}. If $h\left(P_{\underline{u}}\right) < \frac{1}{2}R_{\text{univ}}$ then $\underline{u}$ is the unique solution to

$$\min_{\tilde{\underline{u}} \in \mathrm{Co}(H,\underline{s})} h\left(P_{\tilde{\underline{u}}}\right).$$

Proof: We proceed with a proof by contradiction. Suppose $H\tilde{\underline{u}} = H\underline{u}$ and $h\left(P_{\tilde{\underline{u}}}\right) \leq h\left(P_{\underline{u}}\right)$. Note that

$$\begin{aligned} h\left(P_{\underline{u}-\tilde{\underline{u}},\underline{u},\tilde{\underline{u}}}\right) &= h\left(P_{\underline{u},\tilde{\underline{u}}}\right) + h\left(P_{\underline{u}-\tilde{\underline{u}}|\underline{u},\tilde{\underline{u}}}|P_{\underline{u},\tilde{\underline{u}}}\right) \\ &= h\left(P_{\underline{u},\tilde{\underline{u}}}\right) \\ h\left(P_{\underline{u}-\tilde{\underline{u}},\underline{u},\tilde{\underline{u}}}\right) &= h\left(P_{\underline{u}-\tilde{\underline{u}}}\right) + h\left(P_{\underline{u},\tilde{\underline{u}}|\underline{u}-\tilde{\underline{u}}}|P_{\underline{u}-\tilde{\underline{u}}}\right) \\ \Rightarrow h\left(P_{\underline{u}-\tilde{\underline{u}}}\right) &= h\left(P_{\underline{u},\tilde{\underline{u}}}\right) - h\left(P_{\underline{u},\tilde{\underline{u}}|\underline{u}-\tilde{\underline{u}}}|P_{\underline{u}-\tilde{\underline{u}}}\right) \\ &\leq h\left(P_{\underline{u},\tilde{\underline{u}}}\right) \\ &\leq h\left(P_{\underline{u}}\right) + h\left(P_{\tilde{\underline{u}}}\right) \\ &\leq 2h\left(P_{\underline{u}}\right) \\ &< R_{\text{univ}}. \end{aligned}$$

But by (12) there can be no such $\tilde{\underline{u}} \neq \underline{u}$ with $H\tilde{\underline{u}} = H\underline{u}$ and thus we have a contradiction. ■

IV. The Complexity of Universal Decoding with Linear Codes

Now we discuss the computational complexity of universal 'minimum-entropy' decoding for general linear codes. We restrict ourselves to the binary case here, although these results can be extended to any alphabet size that is a power of 2. Consider a binary linear code $\mathcal{C}$ specified by its parity check matrix $H \in \{0,1\}^M \times \{0,1\}^N$, given by (8). Then we have that H maps a $\underline{u} \in \{0,1\}^N$ to $\underline{s} \in \{0,1\}^M$ via (5), where $M < N$ and U is memoryless with $P(U_i = 1) = p$. Taking $\underline{s}$ as its input along with knowledge of H, the decoder knows that $\underline{u}$ must lie in the coset $\mathrm{Co}\,(H, \underline{s})$ given by (7). In the case of ML decoding, if the decoder knew that $p < \frac{1}{2}$, then it selects $\hat{\underline{u}}$ as the coset's smallest hamming weight member - termed *coset leader*:

$$\hat{\underline{u}} = \arg \min_{\underline{u} \in \mathrm{Co}(H, \underline{s})} w_h\left(\underline{u}\right). \tag{16}$$

In a universal setting, the decoder is unaware of the sign of $p - \frac{1}{2}$, and selects $\hat{\underline{u}}$ as the coset's empirical entropy minimizer, given by (9).

It has been shown in [7] that ML decoding for general linear codes - performing (16) for a general matrix H - is NP-complete. Thus it is not surprising the following theorem holds, but we state it here for the sake of completeness, a solid foundation, and motivation for future code-on-graph based decoding techniques:

Theorem 4.1: The algorithm **MINIMUM-ENTROPY**$[H, \underline{s}]$ for general binary linear codes is NP-complete.

Proof: We prove this by means of a reduction, a common technique for proving theorems in complexity theory. Suppose we are given an instance of the NP-complete problem **COSET-LEADER**$[H, \underline{s}]$, which performs (16). We would like to reduce this to minimum-entropy decoding by showing that if there exists a polynomial-time algorithm for **MINIMUM-ENTROPY**$[H, \underline{s}]$, which performs (9), then it can be used to solve any instance of **COSET-LEADER**$[H, \underline{s}]$. Consider the coset $\mathrm{Co}\left(\tilde{H}, \tilde{\underline{s}}\right)$ where

$$\tilde{H} = \begin{bmatrix} H & 0 \\ 0 & I_N \end{bmatrix}, \quad \tilde{\underline{s}} = \begin{bmatrix} \underline{s} \\ \underline{0} \end{bmatrix},$$

where I_N is the $N \times N$ identity matrix. Consider any $\tilde{\underline{u}} = \begin{bmatrix} \underline{u} \\ \underline{u}' \end{bmatrix} \in \mathrm{Co}\left(\tilde{H}, \tilde{\underline{s}}\right)$ and note that $\tilde{\underline{u}}$ must satisfy

$$\underline{u}' = 0, \quad \underline{u} \in \mathrm{Co}\,(H, \underline{s}). \tag{17}$$

Furthermore, note that any $\tilde{\underline{u}} \in \mathrm{Co}\left(\tilde{H}, \tilde{\underline{s}}\right)$ satisfies

$$\frac{1}{2N} w_h\left(\tilde{\underline{u}}\right) \leq \frac{1}{2} \quad \Leftrightarrow \quad P_{\tilde{\underline{u}}}(0) \leq \frac{1}{2}. \tag{18}$$

Since the binary entropy function is monotonically increasing on $[0, \frac{1}{2})$, we have that the two optimization problems

$$\textbf{MINIMUM-ENTROPY}\left[\tilde{H}, \tilde{\underline{s}}\right]: \quad \min_{\tilde{\underline{u}} \in \mathrm{Co}\left(\tilde{H}, \tilde{\underline{s}}\right)} h\left(P_{\tilde{\underline{u}}}\right),$$

$$\textbf{COSET-LEADER}\left[\tilde{H}, \tilde{\underline{s}}\right]: \quad \min_{\tilde{\underline{u}} \in \mathrm{Co}\left(\tilde{H}, \tilde{\underline{s}}\right)} w_h\left(\tilde{\underline{u}}\right)$$

have the same optimal solution(s). Let $\tilde{\underline{u}}^* = \begin{bmatrix} \underline{u}^* \\ \underline{u}'^* \end{bmatrix}$ be an optimal solution from above. Then from (17) and we have that $\underline{u}^*$ is an optimal solution to **COSET-LEADER**$[H, \underline{s}]$. ■

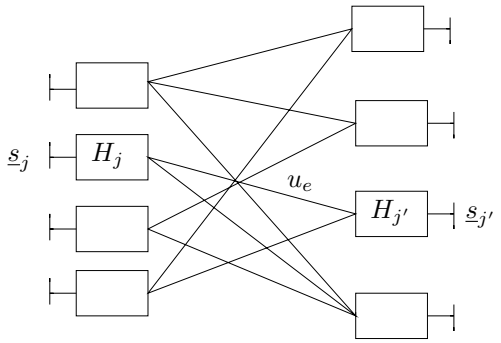

Fig. 1. Graphical representation of a linear system representing $\mathrm{Co}(H, \underline{s})$

V. Codes on Graphs

Considering how we found in the previous section that minimum-entropy decoding is NP-complete, we now discuss code constructions based upon bipartite graphs. In particular, we concern ourselves with graphical realizations of linear systems to represent the coset $\mathrm{Co}\,(H, \underline{s})$. We adhere to Forney's "normal graph" [16, Sec VIII.B] realizations:

- A graph $G = (V, E \cup \bar{E})$ with $|E| = N$ two-sided edges and $\left|\bar{E}\right|$ one-sided edges. For a vertex $j \in V$ we denote $\Gamma\left(j\right)$ as the set of edges $e \in E$ adjacent to j and $\tilde{\Gamma}\left(j\right)$ as the set of edges $\bar{e} \in \bar{E}$ adjacent to j.
- Each local constraint $\mathcal{C}_j$ is represented by a vertex $j \in V$
- The state variable $u_e \in \mathbb{F}_{2^t}$ corresponds to a two-sided edge $e \in E$ and is involved in the two local constraints corresponding to the vertices adjacent to e. The set of all state variables is represented as the vector $\underline{u}$. For any $j \in V$ we denote $\{u_e\}_{e \in \Gamma(j)}$ as $\underline{u}_j$.
- The symbol variable $s_{\bar{e}} \in \mathbb{F}_{2^t}$ corresponds to a one-sided 'leaf-edge' $\bar{e} \in \bar{E}$ and is involved in one local constraint corresponding to the vertex adjacent to $\bar{e}$. For any $j \in V$, we abbreviate $\{s_{\bar{e}}\}_{\bar{e} \in \tilde{\Gamma}(j)}$ as $\underline{s}_j$.
- Any $j \in V$ j and its associated code C_j imposes the constraint that

$$H_j \underline{u}_j + \underline{s}_j = 0 \;\Leftrightarrow\; H_j \underline{u}_j = \underline{s}_j \;\Leftrightarrow\; \underline{u}_j \in \mathrm{Co}\left(H_j, \underline{s}_j\right). \tag{19}$$

The coset $\mathrm{Co}\,(H, \underline{s})$ can be expressed as

$$\mathrm{Co}\,(H, \underline{s}) = \left\{\underline{u} \;\middle|\; \underline{u}_j \in \mathrm{Co}\left(H_j, \underline{s}_j\right), \; \forall\, j \in V\right\}. \tag{20}$$

For a particular graph G, we denote $\mathrm{COG}\left(G, \{H_j\}, \{\underline{s}_j\}\right)$ as the way in which we specify $\mathrm{Co}\,(H, \underline{s})$ in terms of (20).

A. Universal Goodness of Bipartite Graph Codes

We now denote UCOG $\left(G_{\Delta,n},\{H_j\},\{\underline{s}_j\}\right)$ as a COG $\left(G,\{H_j\},\{\underline{s}_j\}\right)$ system where $G_{\Delta,n}$ is a Δ-regular bipartite graph with n vertices and each local H_j is universally good. We are interested in the universal goodness of such codes where where n grows and Δ is a large but fixed constant. We first consider the following lemma that will be useful for our analysis:

Lemma 5.1: Consider a set A where $|A|=n$ and suppose that $\left\{\mathbb{P}^j \in \mathcal{P}_\Delta\left(\mathcal{U}^2\right)\right\}_{j\in A}$, $\mathbb{P}\in\mathcal{P}_{n\Delta}\left(\mathcal{U}^2\right)$. Then:

$$\sum_{\frac{1}{n}\sum_{j\in A}\mathbb{P}^j=\mathbb{P}}\prod_{j\in A}2^{-\Delta\left(R_a-h\left(\mathbb{P}^j\right)-\delta_\Delta\right)} \le 2^{-n\Delta\left(R_a-h(\mathbb{P})-\epsilon'_\Delta\right)}$$

where $\epsilon'_\Delta\to 0$ as $\Delta\to\infty$.

Proof:

$$\begin{aligned}
&\sum_{\frac{1}{n}\sum_{j\in A}\mathbb{P}^j=\mathbb{P}}\prod_{j\in A}2^{-\Delta\left(R_a-h\left(\mathbb{P}^j\right)-\delta_\Delta\right)}\\
&= \sum_{\frac{1}{n}\sum_{j\in A}\mathbb{P}^j=\mathbb{P}} 2^{-n\Delta\left(R_a-\frac{1}{n}\sum_{j\in A}h\left(\mathbb{P}^j\right)-\delta_\Delta\right)} \quad (21)\\
&\le \sum_{\frac{1}{n}\sum_{j\in A}\mathbb{P}^j=\mathbb{P}} 2^{-n\Delta\left(R_a-h\left(\sum_{j\in A}\frac{1}{n}\mathbb{P}^j\right)-\delta_\Delta\right)} \quad (22)\\
&= \sum_{\frac{1}{n}\sum_{j\in A}\mathbb{P}^j=\mathbb{P}} 2^{-n\Delta\left(R_a-h(\mathbb{P})-\delta_\Delta\right)}\\
&\le \left|\mathcal{P}_\Delta\left(\mathcal{U}^2\right)\right|^n 2^{-n\Delta\left(R_a-h(\mathbb{P})-\delta_\Delta\right)}\\
&\le (\Delta+1)^{n|\mathcal{U}|^2}\,2^{-n\Delta\left(R_a-h(\mathbb{P})-\delta_\Delta\right)} \quad (23)\\
&= 2^{-n\Delta\left(R_a-h(\mathbb{P})-\delta_\Delta-|\mathcal{U}|^2\frac{\log(\Delta+1)}{\Delta}\right)}\\
&= 2^{-n\Delta\left(R_a-h(\mathbb{P})-\epsilon'_\Delta\right)}
\end{aligned}$$

where (22) follows from the concavity of entropy, and (23) follows from (2). ∎

This allows us to state the following lemma.

Lemma 5.2: For large but fixed Δ, codes defined in terms of UCOG $\left(G_{\Delta,n},\{H_j\},\{\underline{s}_j\}\right)$ are universally good.

Proof: Let $N=n\Delta$ and $\mathbb{P}\in\mathcal{P}_N\left(\mathcal{U}^2\right)$ correspond to the joint type of any length-N pair $(\underline{u},\underline{\tilde{u}})$ satisfying $H\underline{u}=H\underline{\tilde{u}}$. Define $\mathbb{P}^j\in\mathcal{P}_\Delta\left(\mathcal{U}^2\right)$ to correspond to the joint type of any *local* length-Δ pair $(\underline{u}',\underline{\tilde{u}}')$ satisfying $H_j\underline{u}'=H_j\underline{\tilde{u}}'$. Then we have:

$$\begin{aligned}
\mathcal{N}_H(\mathbb{P}) &\le \sum_{\frac{1}{n}\sum_{j\in A}\mathbb{P}^j=\mathbb{P}}\prod_{j\in A}\mathcal{N}_{H_j}\left(\mathbb{P}^j\right)\\
&\le \sum_{\frac{1}{n}\sum_{j\in A}\mathbb{P}^j=\mathbb{P}}\prod_{j\in A}2^{-\Delta\left(R_a-h\left(\mathbb{P}^j\right)-\delta_\Delta\right)} \quad (24)\\
&\le 2^{-N\left(R_a-h(\mathbb{P})-\epsilon'_\Delta\right)} \quad (25)
\end{aligned}$$

where $\epsilon'_\Delta\to 0$ as $\Delta\to\infty$, (24) follows from (11), (25) follows from Lemma 5.1. Thus it follows that H becomes universally good for large Δ, when thought of having rate $R'=R_a-\epsilon'_\Delta$. ∎

VI. Discussion

We have illustrated the intrinsic difficulty - by means of computational complexity - in doing exact universal decoding for general linear codes. As is the case with other problems, however, we showed the fragility of NP-completeness in robustly characterizing the intrinsic difficulty of universal decoding. By illustrating certain properties of universally good codes as well as constructing graphical universal codes, we have developed the opportunity to create polynomial complexity decoding algorithms [17] - including ones analogous to linear programming relaxations [18] and iterative symbol-flipping expander graph based decoding [19] for ML decoding- with provably good performance.

References

[1] V. D. Goppa, "Universal decoding for symmetric channels," *Probl. Peredachi Inform.*, vol. 11, no. 1, pp. 15–22, 1975, (In Russian).

[2] I. Csiszár, "Linear codes for sources and source networks: Error exponents, universal coding," *IEEE Transactions on Information Theory*, vol. 28, no. 4, pp. 585–592, 1982.

[3] M. Feder and A. Lapidoth, "Universal decoding for channels with memory," *IEEE Transactions on Information Theory*, vol. 44, no. 5, pp. 1726–1745, 1998.

[4] T. S. Han and S. Amari, "Parameter estimation with multiterminal data compression," *IEEE Transactions on Information Theory*, vol. 41, pp. 1802–1833, 1995.

[5] R. Jornsten, "Data compression and its statistical implications, with an application to the analysis of microarray images," Ph.D. dissertation, University of California, Berkeley, Berkeley, CA, December 2001.

[6] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai Shitz, "On information rates for mismatched decoders," *IEEE Transactions on Information Theory*, vol. 40, no. 6, pp. 1953–1967, 1994.

[7] E. R. Berlekamp, R. J. McEliece, and H. C. A. van Tilborg, "On the intractability of certain coding problems," *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 384–386, 1978.

[8] R. Gallager, "Low-density parity-check codes," *IRE Transactions on Information Theory*, vol. 8, pp. 21–28, Jan. 1962.

[9] G. D. Forney, "Concatendated codes," Ph.D. dissertation, MIT, Cambridge, MA, June 1965.

[10] ——, "Generalized minimum distance decoding," *IEEE Transactions on Information Theory*, vol. 12, no. 2, pp. 125–131, 1966.

[11] I. Csiszár, "The method of types," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2205–2523, 1998.

[12] A. Lempel and J. Ziv, "A universal algorithm for sequential data compression," *IEEE Transactions on Information Theory*, pp. 337–343, 1977.

[13] ——, "Compression of individual sequences via variable-rate coding," *IEEE Transactions on Information Theory*, pp. 530–536, 1978.

[14] M. Effros, K. Visweswariah, S. R. Kulkarni, and S. Verdu, "Universal lossless source coding with the Burrows Wheeler transform," *IEEE Transactions on Information Theory*, vol. 48, no. 5, pp. 1061–1081, May 2002.

[15] S. Wicker and S. Kim, *Fundamentals of Codes, Graphs, and Iterative Decoding*. Norwell, MA: Kluwer Academic Publishers, 2003.

[16] G. D. Forney, "Codes on graphs: Normal realizations," *IEEE Transactions on Information Theory*, pp. 101–112, 2001.

[17] T. P. Coleman, "Low-complexity approaches to distributed data dissemination," Ph.D. dissertation, MIT, Cambridge, MA, 2005.

[18] J. Feldman, "Decoding error-correcting codes via linear programming," PhD Dissertation, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, September 2003.

[19] A. Barg and G. Zémor, "Error exponents of expander codes," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1725–1729, 2002.