Tradeoff Functions for Constrained Systems With Unconstrained Positions

T. Lei Poo, *Member, IEEE*, Panu Chaichanavong, *Member, IEEE*, and Brian H. Marcus, *Fellow, IEEE*

Abstract—We introduce a new method for analyzing and constructing combined modulation and error-correcting codes (ECCs), in particular codes that utilize some form of reverse concatenation and whose ECC decoding scheme requires easy access to soft information. We expand the work of Immink and Wijngaarden and also of Campello, Marcus, New, and Wilson, in which certain bit positions in the modulation code are deliberately left unconstrained for the ECC parity bits, in the sense that such positions can take on either bit value without violating the constraint. Our method of analysis involves creating a single graph that incorporates information on these unconstrained positions directly into the constraint graph without any assumptions of periodicity or sets of unconstrained positions, and is thus completely general. We establish several properties of the tradeoff function that relates the density of unconstrained positions to the maximum code rate. In particular, the tradeoff function is shown to be concave and continuous. Algorithms for computing lower and upper bounds for this function are presented. We also show how to compute the maximum possible density of unconstrained positions and give explicit values for the runlength-limited (RLL (d, k)) and maximum-transition-run (MTR (j, k)) constraints.

Index Terms—Bit-insertion schemes, bit-stuffing, combined modulation and error-correcting codes (ECCs), maximum transition run (MTR) system, run-length-limited (RLL) system, tradeoff functions, unconstrained positions.

I. INTRODUCTION

IN recording channels, an error-correcting code (ECC) and a modulation code are often used together to improve the detection performance. An ECC improves the minimum distance, while a modulation code imposes a constraint on the recorded sequences. Some well-known binary constraints include the runlength-limited (RLL (d, k)) system, which limits the run of 0 to be at least d and at most k, and the maximum transition run (MTR (j, k)) [3], which limits the run of 0 to be at most k and the run of 1 to be at most j. When there is no restriction on the runs of 0, we say that $k = \infty$ and, by tradition, we denote such a constraint by MTR (j) [12].

The ECC and the modulation code are usually concatenated so that the ECC is the outer code and the modulation code is the inner code. Since a modulation decoder is typically a hard decoder, the error-correcting decoder does not have the soft information from the channel, and hence the error-correction capability is degraded. Several schemes have been proposed to solve this limitation, such as soft decoding for modulation codes and reverse concatenation [8], [5]. In this paper, we are interested in the scheme presented by Wijngaarden and Immink [14], and Campello *et al.* [1], in which we design a modulation code so that some prespecified positions are "unconstrained." Such a position can take on any symbol without violating the constraints of the modulation code. Thus, we can employ a systematic ECC with appropriate rate and insert the parity symbols in these positions. At the receiver, the ECC is first decoded with no loss of soft information, and the modulation code is subsequently decoded with minimal error propagation.

Example 1 ([14], [1]): Consider a block code for MTR (2) with two codewords: $\{10101, 01101\}$. It is clear that if the third or fifth bit of either codeword is flipped, the constraint is still satisfied. Hence, we can map a user bit to $\{100, 010\}$ and then encode these words by a systematic rate-$3/5$ ECC to obtain an overall rate of $1/5$, see Fig. 1. We say that the period is 5 and the insertion rate is $2/5$. Moreover, by letting $1, 2, 3, 4, 5$ be the bit positions, the unconstrained set is defined to be the set of bit positions that can be flipped, which is $\{3, 5\}$.

The ECC design is beyond the scope of this paper. We will focus on computing the maximum overall code rate for a given constraint and insertion rate. We view the overall code rate as a function of insertion rate and call it the *insertion rate–code rate tradeoff function* for the constraint. For brevity, we will refer to it as the tradeoff function for the rest of this paper. Given a constraint, a period, and an unconstrained set, Campello *et al.* [1] introduced a graph presentation that can be used to compute the maximum rate. However, this approach is somewhat inefficient: we need to consider all unconstrained sets and the period may need to be large. Thus, we would need a presentation that does not depend on the unconstrained set and period; such a presentation is given in Section III.

The paper is organized as follows. In Section II, we give a short background in constrained coding. Following this, we lay out a new framework for understanding constrained systems with unconstrained positions in Section III. In Section IV, we state the general definitions of tradeoff functions for code rate and the maximum insertion rate. In Section V, we show that the maximum insertion rate is rational and equals the maximum cycle mean of the presentation constructed in Section III. In Section VI, we examine different properties of the code rate versus insertion rate tradeoff function in detail. In Section VII, we com-

T. L. Poo was with the Department of Electrical Engineering, Stanford University, Stanford, CA USA. She is now with Marvell Semiconductor Inc., Santa Clara, CA 95054 USA (e-mail: lpoo@marvell.com).

P. Chaichanavong was with the Center for Magnetic Recording Research, University of California, San Diego, CA 92093 USA. He is now with Marvell Semiconductor Inc., Santa Clara, CA 95054 USA (e-mail: panuc@marvell.com).

B. H. Marcus is with the Department of Mathematics, University of British Columbia, Vancouver, BC V6T 1Z2, Canada (e-mail: marcus@math.ubc.ca).
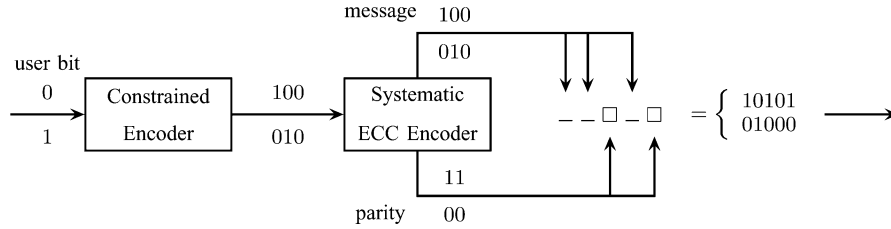
Fig. 1.   Encoding diagram of Example 1.

pare different tradeoff functions. Some lower and approximate upper bounds for the tradeoff function are presented in Section VIII. In Section IX, we consider further properties of the tradeoff functions for specific maximum transition run (MTR) and run-length-limited (RLL) constraints. Finally, in Section X, we summarize the main definitions and results of this paper.

II. DEFINITIONS AND BACKGROUND

In this section, we briefly summarize some background on constrained systems, graph presentations, and follower sets of constrained systems. For a more thorough treatment, the reader may consult [11].

A *labeled graph* $G = (V, E, L)$ consists of

- a finite set of states $V = V_G$;
- a finite set of edges $E = E_G$, where each edge e has an *initial state* $\sigma(e)$ and a *terminal state* $\tau(e)$, both in V;
- an edge labeling $L = L_G : E \to \Sigma$, where Σ is a finite alphabet.

For convenience, we sometimes use

$$u \xrightarrow{a} v$$

to denote an edge or transition from state u to state v with label a.

A *path* γ of length l in a graph G is a finite sequence of edges $e_1 e_2 \ldots e_l$ such that $\sigma(e_{i+1}) = \tau(e_i)$ for $i = 1, 2, \ldots, l - 1$. The state sequence of a path $e_1 e_2 \ldots e_l$ is the sequence $\sigma(e_1)\sigma(e_2)\ldots\sigma(e_l)\tau(e_l)$. A *cycle* in G is a path $e_1 e_2 \ldots e_l$ where $\sigma(e_l) = \tau(e_1)$. A cycle of length one is a *loop*. A *simple cycle* is a cycle in which the state sequence consists of distinct states except the initial and terminal states.

Two labeled graphs can be viewed as the same if there is a labeled graph isomorphism from one to the other, i.e., a bijection from states to states and a bijection from edges to edges which preserve initial states, terminal states and labels.

For a graph G, the *adjacency matrix* $A = A_G$ is the $|V_G| \times |V_G|$ matrix whose entries are indexed by the states of G and $A_{u,v}$ is the number of edges from u to v in G.

Formally, a *constrained system* or *constraint* $S = S(G)$ is the set of finite sequences obtained by reading the edge labels of a path in a labeled graph G. Such a graph is called a *presentation* of the constraint. An element in S is called a *word* or *codeword*. We will use ϵ to denote the empty word. A *homing word* for a state v in a labeled graph G is a word w such that all paths in G that generate w terminate at v. We use wz to denote the concatenation of two words w and z; w^n denotes the concatenation of n copies of w; $|w|$ denotes the length of w; and w_i denotes the ith symbol in w.

A reducible graph can be viewed as a graph consisting of many irreducible subgraphs, called the *irreducible components* with possible *transitions* or *transitional edges* from component to component. An irreducible component G_1 can be reached from an irreducible component G_2 if there is a path from a state in G_2 to a state in G_1. An irreducible component is *trivial* if it has only one state and no edge. A graph is *primitive* if there exists a positive integer p such that for every ordered pair (u, v) of states, there is a path of length p from u to v. A constrained system is said to be irreducible (primitive) if it has an irreducible (primitive) presentation.

A labeled graph is *deterministic* if at each state, all outgoing edges carry distinct labels.

A labeled graph has *finite memory* if there is an integer m such that all paths of length m with the same labeling terminate at the same state. The smallest m for which this holds is called the *memory* of the graph. A constrained system S is *finite-type* or has *finite memory* if there is an integer m such that, for any symbol $a \in \Sigma$ and any word $x \in S$ of length at least m, we have $xa \in S$ if and only if $ya \in S$, where y is the suffix of x of length m. The smallest m for which this holds is called the *memory* of S. It is well known that a constrained system S has finite memory if there is a presentation of S with finite memory. Moreover, the memory of S is the smallest memory of any presentation of S with finite memory. Many practical constrained systems, including RLL (d, k) and MTR (j, k), are primitive and have finite memory.

The *capacity* of a constrained system S, denoted by $\mathrm{cap}(S)$, is defined to be

$$\mathrm{cap}(S) = \lim_{q \to \infty} \frac{\log N(q)}{q}$$

where $N(q)$ is the number of words of length q in S. Let G be a deterministic presentation of S. The capacity can be computed by $\mathrm{cap}(S) = \log \lambda(A_G)$, where $\lambda(A)$ denotes the *spectral radius* of A, which is defined as the maximum of the absolute values of the eigenvalues of A.

Let G be a graph and u a state of G. The *follower set* of u in G, denoted by $\mathsf{F}(u) = \mathsf{F}_G(u)$, is defined to be the set of all finite words that can be generated from u in G. A labeled graph G is *reduced* if for any two distinct states u and v in G, $\mathsf{F}_G(u) \neq \mathsf{F}_G(v)$. For an irreducible constraint, there is a unique minimal (in terms of the number of states) deterministic presentation, called the *Shannon cover*. It is the unique presentation that is irreducible, deterministic, and reduced. Every state of the Shannon cover has a homing word.

We can also define the *follower set* of a finite word x in S as follows:

$$\mathcal{F}(x) = \mathcal{F}_S(x) = \{y \in S \ : \ xy \in S\}.$$

We allow x to be the empty word ϵ, in which case the follower set is all of S. Note that if x does not occur in S, then $\mathcal{F}(x)$ is empty. If G is a presentation of S, $\mathcal{F}_S(x) = \bigcup_u \mathsf{F}_G(u)$, where the union is over the terminal states u of all paths with label x. Since a constrained system is presented by a finite-state graph, it has finitely many follower sets.

The follower sets can be used to construct a special presentation G, called the *follower set graph*, of a constrained system S. The states of G are follower sets of all words in S and the transitions are given by

$$\mathcal{F}(x) \xrightarrow{a} \mathcal{F}(xa)$$

where $a \in \Sigma$ and $xa \in S$. Note in particular that the follower set graph is deterministic. Note also that whenever a word x is the label of a path in the follower set graph ending at state $\mathcal{F}(y)$, then $\mathcal{F}(y) \subseteq \mathcal{F}(x)$. The Shannon cover, defined earlier, turns out to be the subgraph of the follower set graph, determined by follower sets of homing words.

III. PRESENTATIONS WITH THE UNCONSTRAINED SYMBOL

One of the main contributions of this paper is introducing a constrained system and a graph with unconstrained symbol, which we denote by $\square$. This symbol represents the unconstrained position and can be thought of as a wildcard symbol, which is allowed to take on any symbol in the alphabet of the constraint. From now on, we will only consider binary constraints, whose alphabet is $\{0, 1\}$. Most results can be easily extended to nonbinary constraints.

We use $\mathcal{A}_2$ and $\mathcal{A}_3$ to denote the alphabets $\{0, 1\}$ and $\{0, 1, \square\}$, respectively. For an alphabet $\mathcal{A}$, we define $\mathcal{A}^*$ to be the set of all finite words over $\mathcal{A}$. For a given word $w \in \mathcal{A}_3^*$, we define the set of all *fillings* of w as

$$\Phi(w) = \{x \in \mathcal{A}_2^* : |x| = |w|, \ x_i = w_i \text{ if } w_i = 0 \text{ or } 1,$$
$$\text{and } x_i \in \{0, 1\} \text{ if } w_i = \square\}.$$

For example, if $w = 0\square1\square$ then

$$\Phi(w) = \{0010, 0011, 0110, 0111\}.$$

For convenience, we extend the domain of Φ to include right-infinite words. In particular, if w is a right-infinite word, $\Phi(w)$ is given by

$$\Phi(w) = \{x \in \mathcal{A}_2^{\mathbb{N}} : x_i = w_i \text{ if } w_i = 0 \text{ or } 1,$$
$$\text{and } x_i \in \{0, 1\} \text{ if } w_i = \square\}.$$

Let S be a binary constrained system. Define

$$\hat{S} = \{w \in \mathcal{A}_3^* : \Phi(w) \subseteq S\}.$$

This set $\hat{S}$ is a collection of words w such that if we replace each occurrence of $\square$ in w independently by 0 or 1, we obtain a word
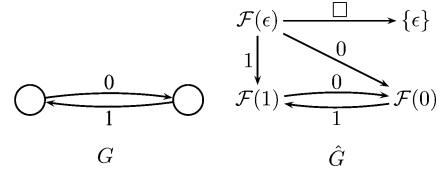

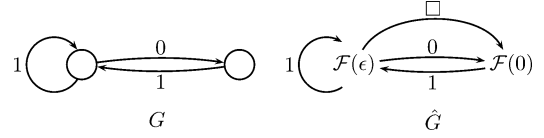Fig. 2.   The Shannon cover G and $\hat{G}$ for the constraint that forbids 00 and 11.

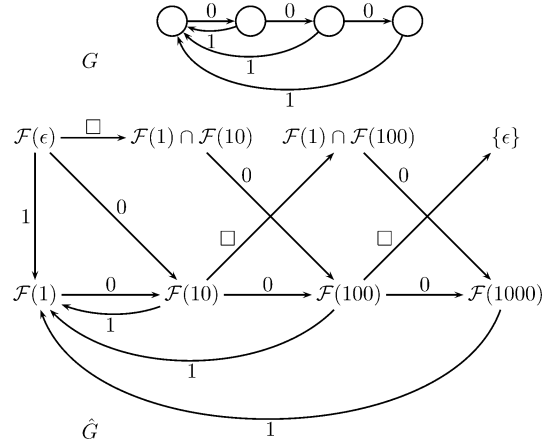Fig. 3.   The Shannon cover G and $\hat{G}$ for the RLL $(0, 1)$.

Fig. 4.   The Shannon cover G and $\hat{G}$ for the RLL $(1, 3)$.

in S. We will construct a graph $\hat{G}$ from S and show that $\hat{S}$ is the constrained system presented by $\hat{G}$.

We define the graph $\hat{G}$ from S as follows.
- **States:** All intersections of the follower sets of words in S.
- **Transitions:** The transitions are given by

$$\bigcap_{i=1}^{k} \mathcal{F}(x_i) \xrightarrow{0} \bigcap_{i=1}^{k} \mathcal{F}(x_i 0), \quad \text{if } x_i 0 \in S \text{ for all } 1 \leq i \leq k$$

$$\bigcap_{i=1}^{k} \mathcal{F}(x_i) \xrightarrow{1} \bigcap_{i=1}^{k} \mathcal{F}(x_i 1), \quad \text{if } x_i 1 \in S \text{ for all } 1 \leq i \leq k$$

$$\bigcap_{i=1}^{k} \mathcal{F}(x_i) \xrightarrow{\square} \bigcap_{b=0}^{1} \bigcap_{i=1}^{k} \mathcal{F}(x_i b), \quad \text{if } x_i 0 \text{ and } x_i 1 \in S$$

$$\text{for all } 1 \leq i \leq k.$$

It can be easily checked that the graph $\hat{G}$ is well defined. Figs. 2–4 show some examples of $\hat{G}$. It is clear from the figures that $\hat{G}$ may not be irreducible even if G is irreducible.

Theorem 2: $\hat{S}$ is the constrained system presented by $\hat{G}$.

Proof: Let γ be a path in $\hat{G}$ with label w. Let $\bigcap_{i=1}^{k} \mathcal{F}(x_i)$ be the initial state of γ. By the definition of transitions of $\hat{G}$, $x_i y \in S$ for all $y \in \Phi(w)$. It follows that every subword of $x_i y$ is also in S. Hence, $y \in S$ for all $y \in \Phi(w)$, which implies that $w \in \hat{S}$.

Conversely, let w be a word of length l in $\hat{S}$. Then there is a path from the state $\mathcal{F}(\epsilon) = S$ in $\hat{G}$ with label w. Specifically

$$\mathcal{F}(\epsilon) \xrightarrow{w_1} u_1 \xrightarrow{w_2} u_2 \xrightarrow{w_3} \cdots \xrightarrow{w_{l-1}} u_{l-1} \xrightarrow{w_l} u_l,$$

where $u_i = \bigcap_{x \in \Phi(w_1 \ldots w_i)} \mathcal{F}(x)$. ∎

In general, $\hat{G}$ can be exponentially large (in the number of states); but if the constraint S has enough structure, $\hat{G}$ may have a reasonable size. We remark that for two natural examples of constraints, the size of $\hat{G}$ is quadratic. The first example is the MTR (j, k) constraint. Let $x \in$ MTR (j, k). Then $\mathcal{F}(x)$ is determined by the last run of symbol of x, e.g., $\mathcal{F}(011000) = \mathcal{F}(000)$. Moreover, if $a \leq b$, then $\mathcal{F}(0^b) \subseteq \mathcal{F}(0^a)$ and $\mathcal{F}(1^b) \subseteq \mathcal{F}(1^a)$. Thus, for any collection of words $x_i \in S$, $1 \leq i \leq n$, $\bigcap_{i=1}^{n} \mathcal{F}(x_i) = \mathcal{F}(0^a) \cap \mathcal{F}(1^b)$ for some $0 \leq a \leq k$ and $0 \leq b \leq j$. This implies that the number of states of $\hat{G}$ is at most $(j+1)(k+1)$. To see that the number of states is exactly $(j+1)(k+1)$, it suffices to check that $\mathcal{F}(0^a) \cap \mathcal{F}(1^b) \neq \mathcal{F}(0^c) \cap \mathcal{F}(1^d)$ if $a \neq c$ or $b \neq d$. Without loss of generality, assume that $0 \leq a < c \leq k$. Then 0^{k-a} is in $\mathcal{F}(0^a) \cap \mathcal{F}(1^b)$ but not in $\mathcal{F}(0^c) \cap \mathcal{F}(1^d)$.

The second example is the RLL (d, k) constraint. We will only sketch the idea and leave out most of the details. First, it is easy to check that $\mathcal{F}(0^a) \cap \mathcal{F}(0^b) = \mathcal{F}(0^{\max\{a,b\}})$. It can be shown that if $a \leq b \leq c$, then $\mathcal{F}(10^a) \cap \mathcal{F}(10^c) \subseteq \mathcal{F}(10^b)$ [2]. With these observations, any intersection U of follower sets can be written as the intersection of at most three sets: $\mathcal{F}(0^a), \mathcal{F}(10^b), \mathcal{F}(10^c)$, for some a, b, c. Without loss of generality, we assume that $b \leq c$.

Case 1: $U = \mathcal{F}(10^b)$. There are $k+1$ possible choices of U in this case.

Case 2: $U = \mathcal{F}(10^b) \cap \mathcal{F}(10^c)$, $b < c$. If $d \leq b$, then $U = \mathcal{F}(10^c)$ which is in Case 1. If $d > b$, it can be shown that there are $d(k - d + 1)$ possible intersections.

Case 3: $U = \mathcal{F}(0^a) \cap \mathcal{F}(10^b)$. If $a \leq b$, then $U = \mathcal{F}(10^b)$ which is Case 1. If $a > b$, then $U = \mathcal{F}(10^a) \cap \mathcal{F}(10^b)$ which is Case 2.

Case 4: $U = \mathcal{F}(0^a) \cap \mathcal{F}(10^b) \cap \mathcal{F}(10^c)$. This case also falls in Cases 1 and 2 by the same argument in Case 3.

Case 5: $U = \mathcal{F}(0^a)$. If $a \geq d$, then $U = \mathcal{F}(10^a)$ which is in Case 1. Thus, there are d possible choices of U in this case.

From Cases 1, 2, and 5, the number of states of $\hat{G}$ for RLL (d, k) is $k + 1 + d(k - d + 1) + d = dk + k + 2d + 1 - d^2$.

In the remainder of this section, we establish some properties of $\hat{G}$.

Proposition 3: Let $\hat{G}$ be constructed from a constrained system S as described above. Then $\hat{G}$ is deterministic and reduced. If u and v are states of $\hat{G}$ and $u \subseteq v$, then $\mathsf{F}_{\hat{G}}(u) \subseteq \mathsf{F}_{\hat{G}}(v)$.

Proof: The graph $\hat{G}$ is deterministic by definition. Let u be a state of $\hat{G}$. Then

$$\mathsf{F}_{\hat{G}}(u) = \{w \in \mathcal{A}_3^* : \Phi(w) \subseteq u\}.$$

Hence, no two states of $\hat{G}$ have the same follower set, and $\mathsf{F}_{\hat{G}}(u) \subseteq \mathsf{F}_{\hat{G}}(v)$ if $u \subseteq v$. ∎

Proposition 4: If S has memory m, then $\hat{G}$ and $\hat{S}$ have memory m.

Proof: Let γ_1 and γ_2 be paths of length m in $\hat{G}$ with the same label w. Let

$$\bigcap_{i=1}^{k_1} \mathcal{F}(x_i) \quad \text{and} \quad \bigcap_{j=1}^{k_2} \mathcal{F}(y_j)$$

be the initial states of γ_1 and γ_2, respectively. Then the terminal states of the two paths are

$$\bigcap_{z \in \Phi(w)} \bigcap_{i=1}^{k_1} \mathcal{F}(x_i z) \quad \text{and} \quad \bigcap_{z \in \Phi(w)} \bigcap_{j=1}^{k_2} \mathcal{F}(y_j z).$$

Suppose that S has memory m. Then for any words x and y of length at least m in S, $\mathcal{F}(x) = \mathcal{F}(y)$ if x and y have the same suffix of length m. It follows that $\mathcal{F}(x_i z) = \mathcal{F}(y_j z)$ for all x_i, y_j, z. Thus, γ_1 and γ_2 have the same terminal state, which is $\bigcap_{z \in \Phi(w)} \mathcal{F}(z)$. This proves that $\hat{G}$ has memory at most m. This in turn implies that $\hat{S}$ has memory at most m.

Next we show that $\hat{G}$ has memory exactly m. Since S has memory m, there exist $x, y \in S$ of length at least m with the same suffix of length $m - 1$ such that $\mathcal{F}(x) \neq \mathcal{F}(y)$. In $\hat{G}$, there are two paths from $\mathcal{F}(\epsilon)$ to $\mathcal{F}(x)$ and $\mathcal{F}(y)$ with label x and y, respectively. The last portions of these paths have the same label, namely, the suffix of length $m - 1$ of x and y, but they do not terminate at the same state. Therefore, if $\hat{G}$ has finite memory, its memory is at least m. Since we have shown that $\hat{G}$ has memory at most m, we conclude that $\hat{G}$ must have memory m. To see that $\hat{S}$ has memory exactly m, it suffices to show that $\mathcal{F}_{\hat{S}}(x) \neq \mathcal{F}_{\hat{S}}(y)$. From the definition of $\hat{S}$, a binary word $xz \in \hat{S}$ if and only if $xz \in S$. Since $\mathcal{F}(x) \neq \mathcal{F}(y)$, it follows that $\mathcal{F}_{\hat{S}}(x) \neq \mathcal{F}_{\hat{S}}(y)$. ∎

Proposition 5: Let S be an irreducible constrained system and G the Shannon cover for S. Let $\hat{G}$ be constructed from S as described above. Then there is a unique subgraph H of $\hat{G}$ that is isomorphic to G.

Proof: Since G is the Shannon cover for S, there is a homing word for every state of G. Let u be a state of G and h its homing word. It is clear that $\mathcal{F}_S(h) = \mathsf{F}_G(u)$. This implies that for any homing words h and h' of the same state, $\mathcal{F}_S(h) = \mathcal{F}_S(h')$. Conversely, if h and h' are homing words of u and u', and $\mathcal{F}_S(h) = \mathcal{F}_S(h')$, then $u = u'$ because the Shannon cover G is reduced.

Let $u_1, \ldots, u_n$ be the states of G. For $1 \leq i \leq n$, let h_i be a homing word of u_i in G. Since G is reduced, $\mathcal{F}(h_i) = \mathsf{F}_G(u_i)$ are distinct. Thus $\mathcal{F}(h_i)$ are distinct states in $\hat{G}$. We will show that there is a subgraph H of $\hat{G}$ with $V_H = \{\mathcal{F}(h_i) : 1 \leq i \leq n\}$ that is isomorphic to G.

Suppose that there is an edge from u_i to u_j in G with label $b \in \mathcal{A}_2$. Then $h_i b \in S$ and, hence, there is a transition from $\mathcal{F}(h_i)$ to $\mathcal{F}(h_i b)$ with label b in $\hat{G}$. Since $h_i b$ is a homing word of u_j, $\mathcal{F}(h_i b) = \mathcal{F}(h_j)$. Thus, this transition indeed terminates at $\mathcal{F}(h_j)$. This defines the transitions of H.

To show that H is a unique subgraph of $\hat{G}$, it suffices to show that each h_i is a homing word of $\hat{G}$. First, there is a path from $\mathcal{F}(\epsilon)$ to $\mathcal{F}(h_i)$ with label h_i. Suppose there is another path from $\bigcap_{j=1}^{k} \mathcal{F}(z_j)$ with label h_i. Then its terminal state is $\bigcap_{j=1}^{k} \mathcal{F}(z_j h_i)$. Since every word $z_j h_i$ is a homing word of u_i in G, $\mathcal{F}(z_j h_i) = \mathcal{F}(h_i)$ and so the terminal state is indeed $\mathcal{F}(h_i)$. It follows that h_i is a homing word in $\hat{G}$. ∎

Lemma 6: Let w be the label of a cycle in $\hat{G}$. Then the right-infinite word w^∞ can be generated from H.

Proof: For simplicity, we will consider follower sets to include right-infinite words in this proof. Let G be the Shannon cover of the constrained system S. Define $\mathsf{F}_G^\infty(u)$ to be the set of right-infinite words that can be generated from a state u in V_G. We claim that there is a state $u \in V_G$ such that

$$\Phi(w^\infty) \subseteq \mathsf{F}_G^\infty(u) \tag{1}$$

and, hence, w^∞ is in the follower set of the corresponding state in H. To prove the claim, suppose that there is no such state. Since $w^\infty \in \hat{S}$, there is a set of states $V \subseteq V_G$ such that

a) $|V| \geq 2$,
b) $\Phi(w^\infty) \subseteq \bigcup_{v \in V} \mathsf{F}_G^\infty(v)$,
c) if V' is another subset of V_G such that

$$\Phi(w^\infty) \subseteq \bigcup_{v \in V'} \mathsf{F}_G^\infty(v)$$

then $|V'| \geq |V|$.

Pick any $v \in V$. Then there is a word $x \in \Phi(w^\infty)$ such that $x \notin \mathsf{F}_G(v)$. This implies that there is an integer n such that $x_1 \ldots x_{n|w|} \notin \mathsf{F}_G(v)$. Thus,

$$\Phi(x_1 \ldots x_{n|w|} w^\infty) \subseteq \bigcup_{u \in V \setminus \{v\}} \mathsf{F}_G^\infty(u). \tag{2}$$

Let T be the set of terminal states of the paths from $V \setminus \{v\}$ with label $x_1 \ldots x_{n|w|}$. Since G is deterministic, $|T| \leq |V| - 1$. From (2), $\Phi(w^\infty) \subseteq \bigcup_{u \in T} \mathsf{F}_G^\infty(u)$, which contradicts c); this proves the claim. ∎

Lemma 7: Let S be an irreducible constraint. Let γ be a cycle in $\hat{G}$ and let w be the label of γ. Then there is a binary word w' such that ww' is a homing word in $\hat{G}$ of a state in V_H.

Proof: Let G be the Shannon cover for S. Let V be the set of states $v \in V_G$ such that $\Phi(w^\infty) \subseteq \mathsf{F}_G(v)$. By (1), V is nonempty. Let x be a binary word and $U \subseteq V_G$. Define $T(x, U)$ to be the set of terminal states of all paths from U with label x. By convention, $T(\epsilon, U) = U$. For simplicity, we write $V_x = T(x, V)$. Observe the following properties of V and T.

a) $V_x \subseteq V$ for all $x \in \Phi(w)$.
b) If $U_1 \subseteq U_2$, then $T(x, U_1) \subseteq T(x, U_2)$.
c) $T(x, \bigcup_i U_i) = \bigcup_i T(x, U_i)$.
d) $T(xy, U) = T(y, T(x, U))$.

We claim that there are a nonnegative integer n and a word $y \in \Phi(w^n)$ such that

$$\bigcap_{x \in \Phi(w)} T(y, V_x) \neq \emptyset.$$

To prove the claim, assume that there are no such n and y. Define

$$N = \min_{m \geq 0} \min_{z \in \Phi(w^m)} \left| \bigcup_{x \in \Phi(w)} T(z, V_x) \right|.$$

From the definition of V, $T(z, V_x)$ is nonempty for all z and x. Therefore, $N \geq 1$. Suppose that m' and z' achieve N, i.e.,

$$N = \left| \bigcup_{x \in \Phi(w)} T(z', V_x) \right|.$$

By the assumption

$$\bigcap_{x \in \Phi(w)} T(z', V_x) = \emptyset.$$

Thus, there is some $x' \in \Phi(w)$ such that

$$|T(z', V_{x'})| < \left| \bigcup_{x \in \Phi(w)} T(z', V_x) \right| = N.$$

Now

$$\bigcup_{x \in \Phi(w)} V_x \subseteq V \qquad \text{(by a))}$$

$$T\left(x', \bigcup_{x \in \Phi(w)} V_x \right) \subseteq T(x', V) = V_{x'} \qquad \text{(by b))}$$

$$T\left(z', T\left(x', \bigcup_{x \in \Phi(w)} V_x \right) \right) \subseteq T(z', V_{x'}) \qquad \text{(by b))}$$

$$T\left(x'z', \bigcup_{x \in \Phi(w)} V_x \right) \subseteq T(z', V_{x'}) \qquad \text{(by d))}$$

$$\bigcup_{x \in \Phi(w)} T(x'z', V_x) \subseteq T(z', V_{x'}) \qquad \text{(by c))}$$

$$\left| \bigcup_{x \in \Phi(w)} T(x'z', V_x) \right| \leq |T(z', V_{x'})| < N$$

which contradicts the definition of N. This proves the claim.

Thus, there are a nonnegative integer n and a word $y \in \Phi(w^n)$ such that

$$\bigcap_{x \in \Phi(w)} T(y, V_x) \neq \emptyset.$$

Let $v \in \bigcap_{x \in \Phi(w)} T(y, V_x)$. Let $F = \bigcap_{x \in \Phi(w)} \mathcal{F}_S(xy)$. Then F is the terminal state in $\hat{G}$ of the path from $\mathcal{F}_S(\epsilon)$ with label wy. Now

$$F = \bigcap_{x \in \Phi(w)} \mathcal{F}_S(xy)$$

$$= \bigcap_{x \in \Phi(w)} \bigcup_{u \in T(xy, V_G)} \mathsf{F}_G(u)$$

$$\supseteq \bigcap_{x \in \Phi(w)} \bigcup_{u \in T(xy, V)} \mathsf{F}_G(u)$$

$$= \bigcap_{x \in \Phi(w)} \bigcup_{u \in T(y, V_x)} \mathsf{F}_G(u)$$

$$\supseteq \bigcap_{x \in \Phi(w)} \mathsf{F}_G(v) \quad (\text{since } v \in T(y, V_x) \text{ for all } x \in \Phi(w))$$

$$= \mathsf{F}_G(v).$$

Since G is the Shannon cover of an irreducible constraint, there is a homing word h in G from v to a state v'. It follows that $h \in \mathsf{F}_{\hat{G}}(F)$ and wyh is a homing word in $\hat{G}$ of a state in H. ∎

In Lemmas 6 and 7, we established a very interesting property of $\hat{G}$, which will lead to the main result in Proposition 10. Given

the label w of a cycle in $\hat{G}$, there are irreducible components G_1 and G_2 of $\hat{G}$ such that G_1 can reach H, H can reach G_2, $w \in S(G_1)$, and $w \in S(G_2)$. We also need the following two lemmas to establish Proposition 10.

Lemma 8: Let G be a graph and let S be the constrained system presented by G. Let w be a word in S such that $w^m \in S$ for all $m \in \mathbb{N}$. Then there exists a cycle in G with label w^k for some k.

Proof: Let $n = |V_G|$. Let π be a path in G with label w^n. This path can be written as a concatenation of n subpaths of the same length: $\pi = \alpha_1 \alpha_2 \ldots \alpha_n$. Let u_0 be the initial state of π and let u_i, $1 \leq i \leq n$, be the terminal states of α_i. There must be $j_1, j_2 \in \{0, 1, \ldots, n\}$ such that $j_1 < j_2$ and $u_{j_1} = u_{j_2}$. Then $\alpha_{j_1+1} \ldots \alpha_{j_2}$ is a cycle with label $w^{j_2 - j_1}$. ∎

Lemma 9: Let G be an irreducible graph. Let K be a graph such that the label of every cycle in G is in $S(K)$. Then there is an irreducible component K' of K such that $S(G) \subseteq S(K')$.

Proof: Let $K_1, \ldots, K_n$ be the irreducible components of K. Suppose that there is no K_i such that $S(G) \subseteq S(K_i)$. Then for each K_i, there is a word $z_i \in S(G)$ such that $z_i \notin S(K_i)$. Since G is irreducible, there are words $x_1, \ldots, x_n$ such that $w = x_1 z_1 \ldots x_n z_n$ is the label of a cycle in G. By Lemma 8, K has a cycle with label w^k for some k. This cycle must be in an irreducible component K_j. Thus, $x_1 z_1 \ldots x_n z_n \in S(K_j)$, which contradicts $z_j \notin S(K_j)$. Hence, there must be an irreducible component K' such that $S(G) \subseteq S(K')$. ∎

Proposition 10: Let S be an irreducible constraint. Let G be an irreducible component of $\hat{G}$. There exist irreducible components G_1 and G_2 of $\hat{G}$ such that

- G_1 can reach H,
- H can reach G_2,
- $S(G) \subseteq S(G_1)$,
- $S(G) \subseteq S(G_2)$.

Proof: Let K be the subgraph of $\hat{G}$ that can reach H. Let $\bar{K}$ be the subgraph of $\hat{G}$ that can be reached from H. By Lemmas 7 and 6, the label of every cycle in G is in both $S(K)$ and $S(\bar{K})$. By Lemma 9, there are an irreducible component G_1 of K and an irreducible component G_2 of $\bar{K}$ such that $S(G) \subseteq S(G_1)$ and $S(G) \subseteq S(G_2)$. Certainly, G_1 can reach H by the definition of K, and G_2 can be reached from H by the definition of $\bar{K}$. ∎

Define G' to be the irreducible component of $\hat{G}$ that contains H. The following proposition shows the importance of this component when the constraint is finite-type.

Proposition 11: If S is finite-type, then G' is the only non-trivial irreducible component of $\hat{G}$.

Proof: Let m be the memory of S. By Proposition 4, $\hat{S}$ and $\hat{G}$ have memory m. Let G_1 be a nontrivial irreducible component of $\hat{G}$. There is a cycle γ in G_1 of length at least m with some label w. Let u be the initial state of γ. Then any $x \in \Phi(w)$ is a homing word for a state in H and $x \in \mathsf{F}(u)$, so H is accessible from G_1. Conversely, by Lemma 6, there is a path η starting in H with label w. But $\hat{G}$ has memory m, so η terminates at u. Therefore, G_1 is reachable from H, which shows that G_1 is the irreducible component containing H. ∎

Note that Proposition 11 does not hold in general for non finite-type constraints. See Examples 14 and 28 in Sections IV and VI.

IV. TRADEOFF FUNCTIONS

Let q be an integer, and $I \subseteq \{1, \ldots, q\}$. Let S be a constrained system. Define $M(q, I)$ to be the number of words $w \in \hat{S}$ such that

- $|w| = q$,
- $w_i = \square$ if and only if $i \in I$.

Hence, the set I represents the positions of $\square$ in w; we say that w *satisfies* I.

For $0 \leq \rho \leq 1$, define $\mathcal{I}(\rho)$ to be the set of all sequences (I_q) such that $I_q \subseteq \{1, 2, \ldots, q\}$ and $\lim_{q \to \infty} |I_q|/q = \rho$. We define the overall optimum code rate for S at insertion rate ρ to be

$$f(\rho) = f_S(\rho) = \sup_{(I_q) \in \mathcal{I}(\rho)} \limsup_{q \to \infty} \frac{\log M(q, I_q)}{q}. \quad (3)$$

We call this overall rate the insertion rate–code rate tradeoff function for S, or tradeoff function for short. Thus, the tradeoff function is the frontier of all achievable rate regions for this scheme using unconstrained positions. Note that $f(0) = \mathrm{cap}(S)$ by taking each I_q to be empty and $f(\rho)$ can be ngative. With this notation, we define the maximum insertion rate to be

$$\mu = \mu(S) = \sup_{f(\rho) > 0} \rho.$$

We remark that the $\sup$ in the definition of μ can in fact be taken over $f(\rho) \geq 0$ (see the second half of proof of Theorem 18).

The following proposition states that the supremum in the definition of the tradeoff function is achieved by some sequence (I_q).

Proposition 12: Let $0 \leq \rho \leq 1$. There is a sequence $(I_q) \in \mathcal{I}(\rho)$ such that

$$f(\rho) = \limsup_{q \to \infty} \frac{\log M(q, I_q)}{q}.$$

Proof: From the definition of f, there exist $q_1 < q_2 < \cdots$ and $I_{q_n}^{(n)} \subseteq \{1, \ldots, q_n\}$ such that

$$\left| \frac{|I_{q_n}^{(n)}|}{q_n} - \rho \right| \leq \frac{1}{n}$$

$$\frac{\log M\left(q_n, I_{q_n}^{(n)}\right)}{q_n} \geq f(\rho) - \frac{1}{n}.$$

Fix $(I_q^0) \in \mathcal{I}(\rho)$ arbitrarily. For each $q \in \mathbb{N}$, define I_q to be $I_{q_n}^{(n)}$ if $q = q_n$ for some n, and I_q^0 otherwise. It is easy to check that $(I_q) \in \mathcal{I}(\rho)$. Moreover

$$\limsup_{q \to \infty} \frac{\log M(q, I_q)}{q} \geq \limsup_{n \to \infty} \frac{\log M(q_n, I_{q_n}^{(n)})}{q_n} \geq f(\rho).$$

Hence (I_q) achieves the supremum in definition of $f(\rho)$. ∎

The following example shows why we use $\limsup$ in the definition of f, namely, the limit may not exist for some choice of (I_q).

Example 13: Let S be the RLL$(0,1)$ constraint. Its Shannon cover G and $\hat{G}$ are shown in Fig. 3. We will give an example of $(I_q) \in \mathcal{I}(1/3)$ such that $\lim_{q\to\infty}(\log M(q, I_q))/q$ does not exist. To simplify the construction of I_q, we represent a set $J \subseteq \{1, 2, \ldots, q\}$ by a binary word x, where $x_j = 1$ if $j \in J$.

We construct a right-infinite sequence x recursively as follows. Let $y = 010010$ and $z = 100010$. Set $k \leftarrow 2$ and $x \leftarrow yz$. Then recursively compute $x \leftarrow xy^k z^k$ and $k \leftarrow 3k$. We define $I_q = \{i \in \{1, \ldots, q\} : x_i = 1\}$. It is clear that $(I_q) \in \mathcal{I}(1/3)$.

Let

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

The matrix A is the part of the adjacency matrix $A_{\hat{G}}$ of $\hat{G}$ that corresponds to edges with label 0 and 1, while the matrix B is the part of $A_{\hat{G}}$ that corresponds to edges with label $\square$ (see Fig. 3). Let $C(q) = \prod_{i=1}^{q} C_i$, where $C_i = A$ if $i \notin I_q$ and $C_i = B$ if $i \in I_q$. It can be checked that $M(q, I_q) = c_{11} + c_{12}$, where c_{11} and c_{12} are the $(1, 1)$ and $(1, 2)$ entries of $C(q)$, respectively. The basic building blocks that correspond to y^k and z^k are

$$(ABAABA)^k = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \quad (BAAABA)^k = \begin{bmatrix} 2^k & 0 \\ 0 & 0 \end{bmatrix}.$$

It follows that for any $n \in \mathbb{N}$, if $q = 12 \times 3^n$, then

$$M(q, I_q) = 2^{3^n} = M(2q, I_{2q}).$$

Thus, for $q = 12 \times 3^n$, $(\log M(q, I_q))/q = 1/12$. For $q = 24 \times 3^n$, $(\log M(q, I_q))/q = 1/24$. These are two subsequences of $(\log M(q, I_q))/q$ that converge to different values and hence the limit does not exist.

The following two examples show that $f(\mu)$ need not be zero.

Example 14: The Shannon cover G and $\hat{G}$ of the 2-charge constraint is shown in Fig. 5. The capacity for this constraint is 0.5, and so $f(0) = 0.5$. However, there is no word in $\hat{S}$ that has more than two $\square$. Therefore, $f(\rho) = -\infty$ when $\rho > 0$.

Example 15: Let G be the graph in Fig. 6 without the dashed edges and let $S = S(G)$. The constraint S is primitive since G is irreducible and aperiodic. The graph G has memory 7 and hence S is finite-type. The irreducible component G' of $\hat{G}$ that contains H is obtained by adding the dashed edges. It can be shown that $\mu = 1/8$ and $f(\mu) = 1/8$. This example shows that $f(\mu)$ need not be zero.

In the spirit of Campello *et al.* [1], which examined constructions with periodic insertions of $\square$, we can also define the notion of *periodicity* here. A set $I \subseteq \mathbb{N}$ is said to be *periodic* if there exists a period $N \in \mathbb{N}$ and $U \subseteq \{0, \ldots, N-1\}$ such that $i \in I$ if and only if $i \equiv i' \mod N$ for some $i' \in U$. A sequence $(I_q) \in \mathcal{I}(\rho)$ is *periodic* if

- $I_q \subseteq I_{q+1}$ for all $q \in \mathbb{N}$, i.e., the sequence is increasing,
- $I = \bigcup_{q=1}^{\infty} I_q$ is a periodic set.
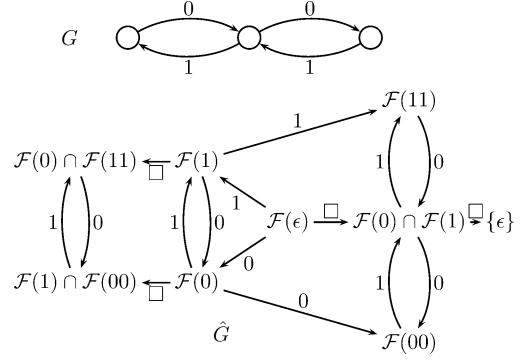
In this case, the sequence of (I_q) is determined by I.

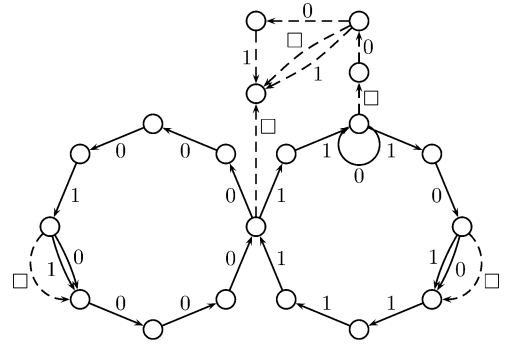Fig. 5. The Shannon cover G and $\hat{G}$ for the 2-charge constraint.

Fig. 6. An example showing that $f(\mu)$ need not be zero, even when S is finite-type and primitive.

Let $\mathcal{I}_p(\rho) = \{(I_q) \in \mathcal{I}(\rho) : (I_q) \text{ is periodic}\}$. We define the maximum achievable code rate for periodic unconstrained positions as

$$f^{(p)}(\rho) = \sup_{(I_q)\in\mathcal{I}_p(\rho)} \limsup_{q\to\infty} \frac{\log M(q, I_q)}{q}.$$

This function $f^{(p)}$ is defined only on the rationals $0 \le \rho \le 1$. Clearly, $f^{(p)}(\rho) \le f(\rho)$.

In contrast to Example 13, if the sequence (I_q) is periodic, the $\limsup$ in the definitions of f and $f^{(p)}$ can always be replaced by the limit.

Proposition 16: Let (I_q) be periodic with period N and let $I = \bigcup_{q=1}^{\infty} I_q$. Let A and B be the parts of $A_{\hat{G}}$ that correspond to symbols 0, 1, and symbol $\square$, respectively. Let $C = \prod_{i=1}^{N} C_i$ where $C_i = A$ if $i \notin I$ and $C_i = B$ if $i \in I$. Then

$$\lim_{q\to\infty} \frac{\log M(q, I_q)}{q} = \frac{1}{N} \log \lambda(C).$$

Proof: For simplicity, define $L_q = (\log M(q, I_q))/q$. Since $\hat{G}$ is deterministic

$$\|C^n\|_\infty \le M(nN, I_{nN}) \le |V_{\hat{G}}| \|C^n\|_\infty$$

where $\|C^n\|_\infty$ is the maximum row sum of C^n. Therefore,

$$\lim_{n\to\infty} L_{nN} = \lim_{n\to\infty} \frac{\log \|C^n\|_\infty}{nN}.$$

It is well known, e.g., [7, Corollary 5.6.14], that $\lambda(C) = \lim_{n\to\infty} \|C^n\|_\infty^{1/n}$. Since log is a continuous function

$$\lim_{n\to\infty} L_{nN} = \frac{1}{N} \log\left(\lim_{n\to\infty} \|C^n\|_\infty^{1/n}\right) = \frac{1}{N} \log\lambda(C).$$

Since $I_q \subseteq I_{q+1}$, the number of words that satisfy I_{q+1} is at most twice the number of words that satisfy I_q. Therefore, if n is the integer such that $nN \leq q < (n+1)N$

$$-\frac{1}{n+1} + L_{(n+1)N} \leq L_q \leq \frac{1}{n} + L_{nN}.$$

Hence (L_q) also converges to the same limit as its subsequence (L_{nN}). ∎

Let $f^{(o)}$ denote the definition of code rate in [1]. We repeat $f^{(o)}$ in our notation for clarity of presentation as follows:

$$f^{(o)}(\rho) = \limsup_{q\to\infty} \max_{\frac{|I_q|}{q}=\rho} \frac{\log|M(q, I_q)|}{q}.$$

It can be shown from the definitions of $f^{(p)}$ and $f^{(o)}$ and Proposition 16 that $f^{(p)}(\rho) \leq f^{(o)}(\rho) \leq f(\rho)$. We will show in Section VII that $f^{(o)}(\rho) = f(\rho)$ on the rationals $0 \leq \rho \leq \mu$, and for finite-type constraints $f(\rho) = f^{(p)}(\rho)$.

To prove further results, we need to consider the tradeoff function in a more general setting. We define the tradeoff function for any labeled graph G with alphabet $\mathcal{A}_3$ by simply letting $M(q, I)$ in (3) to be the number of words $w \in S(G)$ such that $|w| = q$ and $w_i = \square$ if and only if $i \in I$. The tradeoff function for G is denoted by f_G. Similarly, we define $\mu_G = \sup_{f_G(\rho)>0} \rho$. In this setting, the tradeoff function f is equal to $f_{\hat{G}}$. Similar extensions can be applied to $f^{(p)}$ and $f^{(o)}$.

V. COMPUTING MAXIMUM INSERTION RATES

Let π be a path in $\hat{G}$. Define $\nu(\pi)$ to be the ratio of the number of $\square$ in the label of π to its length.

Proposition 17: For any cycle γ in $\hat{G}$, there exists a simple cycle α such that $\nu(\alpha) \geq \nu(\gamma)$.

Proof: Any cycle γ can be written as $\gamma_1\alpha_1\gamma_2$, where α_1 is the first simple cycle (from left to right). Extract out α_1 and this results in a new path $\gamma_1\gamma_2$. Repeating this process until no simple cycles remain, we obtain a collection of simple cycles $\alpha_1, \alpha_2, \ldots, \alpha_k$ whose total length equals $|\gamma|$ and whose total number of $\square$ equals the number of $\square$ of γ. Then $\nu(\gamma)$ is the weighted average of $\nu(\alpha_i)$, $1 \leq i \leq k$. Hence, there is a simple cycle α_i such that $\nu(\alpha_i) \geq \nu(\gamma)$. ∎

We define

$$\nu^* = \max_{\text{all cycles } \gamma \text{ in } \hat{G}} \nu(\gamma).$$

A cycle that achieves ν^* is called a *max-insertion-rate cycle*. Proposition 17 implies that ν^* exists and we only need to consider simple max-insertion-rate cycles. Also, note that ν^* corresponds to the maximum cycle mean when a weight of 1 is assigned to the edges labeled $\square$ and a weight of 0 is assigned to the other edges.

Theorem 18: The maximum insertion rate μ is equal to ν^*.

Proof: Let γ be a max-insertion-rate cycle, so $\nu(\gamma) = \nu^*$. First, we show that $\mu \leq \nu(\gamma)$. Suppose that $\rho > \nu(\gamma)$ and $f(\rho) \geq 0$. Then there exist q and $I_q \subseteq \{1, \ldots, q\}$ such that

$$|I_q| > q\nu(\gamma) + |V_{\hat{G}}| \tag{4}$$

and $M(q, I_q) \geq 1$. Hence, there exists a path π with label w that satisfies I_q.

We claim that any path β can be written as

$$\alpha_1 e_1 \alpha_2 e_2 \ldots e_{m-1}\alpha_m e_m$$

where the α_j are cycles (possibly of length 0), the e_j are edges, where m is at most the number of distinct states visited by β. To see this, use induction on the number of distinct states visited by β. Let u_i be the initial state of the edge β_i (so u_{i+1} is the terminal state of β_i). The claim is trivial if only one state is visited (in which case β must consist entirely of loops at that state). Assume the claim is true when the number of distinct states visited is less than k (where $k \geq 2$), and assume β visits k distinct states. We then take α_1 to be the subpath $\beta_1 \ldots \beta_{k_1-1}$ where k_1 is the time of the last visit to u_1, i.e., $k_1 = \max\{j : u_j = u_1\}$ (if $k_1 = 1$, take the path to be empty). Write β as $\alpha_1 e_1 \beta'$. By the inductive hypothesis, β' can be decomposed as $\alpha_2 e_2 \alpha_3 e_3 \ldots e_{m-1}\alpha_m e_m$ with $m \leq k$, which completes the proof of the claim.

Applying the above claim to the path π, the number of $\square$ in w is at most

$$|V_{\hat{G}}| + \sum_{i=1}^{m} |\alpha_i|\nu(\alpha_i) \leq |V_{\hat{G}}| + q\nu(\gamma)$$

which contradicts (4). Hence, $f(\rho) = -\infty$ when $\rho > \nu(\gamma)$ and so $\mu \leq \nu(\gamma)$.

Next, we show that $\mu \geq \nu(\gamma)$. Let w be the label of γ. Let k be the number of $\square$ in w and let $N = |\gamma|$. For any $\epsilon > 0$, there is an integer ℓ such that

$$\nu(\gamma) - \epsilon < \frac{\ell k - 1}{\ell N} < \frac{k}{N} = \nu(\gamma).$$

Let i_1 be the first position of $\square$ in w. Let

$$I'_q = \Big\{i \in \{1, \ldots, q\} : w_{(i \bmod \ell N)+1} = \square,$$
$$\text{and } (i \bmod \ell N) + 1 \neq i_1\Big\}.$$

It is clear that $(I'_q) \in \mathcal{I}((\ell k - 1)/(\ell N))$. Let n be an integer. Consider the word $(w^\ell)^n$. By viewing this word as having period ℓN, we replace the first $\square$ in each period by 0 and 1 to obtain 2^n words of length $n\ell N$ that satisfy $I'_{n\ell N}$. It follows that $M(q, I'_q) \geq 2^{q/\ell N}$ when q is a multiple of ℓN. Thus,

$$f\left(\frac{\ell k - 1}{\ell N}\right) \geq \frac{1}{\ell N} > 0.$$

Hence, $\mu \geq (\ell k - 1)/(\ell N) > \nu(\gamma) - \epsilon$. Since this is true for any $\epsilon > 0$, $\mu \geq \nu(\gamma)$. ∎

From Proposition 17 and Theorem 18, we can compute the maximum insertion rate by finding a simple max-insertion-rate cycle. This is accomplished by applying Karp's algorithm [9] to $\hat{G}$. Karp's algorithm is a polynomial time algorithm that finds the maximum cycle mean of a weighted directed graph by searching over simple cycles.

A. Periodic Insertions at the Maximum Insertion Rate

In this subsection, we will show that $f(\mu) = f^{(p)}(\mu) = f^{(o)}(\mu)$. This result will be used later in Section VII, where we prove that $f^{(o)}(\rho) = f(\rho)$ in general, and $f^{(p)}(\rho) = f(\rho)$ if the constraint is finite-type.

Throughout this subsection, let T be the subgraph of $\hat{G}$ consisting of all edges that belong to a max-insertion-rate cycle.

Lemma 19: A cycle in $\hat{G}$ is contained in T if and only if it is a max-insertion-rate cycle.

Proof: Let γ be a max-insertion-rate cycle. Then γ is contained in T by definition.

For the other direction, suppose that γ is a cycle of length n in T. For each edge γ_i, there is a path γ_i' such that $\gamma_i \gamma_i'$ is a max-insertion-rate cycle. Note that $\gamma_1 \ldots \gamma_n \gamma_n' \ldots \gamma_1'$ is a cycle in T, and so is $\gamma_n' \ldots \gamma_1'$. Then its insertion rate is

$$\nu(\gamma_1 \ldots \gamma_n \gamma_n' \ldots \gamma_1') = \sum_{i=1}^{n} \theta_i \nu(\gamma_i \gamma_i') = \sum_{i=1}^{n} \theta_i \mu = \mu$$

where $\theta_i = |\gamma_i \gamma_i'|/|\gamma_1 \ldots \gamma_n \gamma_n' \ldots \gamma_1'|$ and $\sum_{i=1}^{n} \theta_i = 1$.

However,

$$\mu = \nu(\gamma_1 \ldots \gamma_n \gamma_n' \ldots \gamma_1') = \theta \nu(\gamma) + (1 - \theta) \nu(\gamma_n' \ldots \gamma_1')$$

for $0 < \theta < 1$. Note that if either $\nu(\gamma) < \mu$ or $\nu(\gamma_n' \ldots \gamma_1') < \mu$, then this would imply $\nu(\gamma_1 \ldots \gamma_n \gamma_n' \ldots \gamma_1') < \mu$, which is a contradiction. So $\nu(\gamma) = \nu(\gamma_n' \ldots \gamma_1') = \mu$. ∎

Lemma 20: For any $\epsilon > 0$, there exists $\delta > 0$ such that if π is a path in $\hat{G}$ with $|\nu(\pi) - \mu| < \delta$ and $|\pi| > 1/\delta$, then

$$\frac{\text{number of edges in } \pi \text{ that are not in } T}{|\pi|} < \epsilon.$$

Proof: By the same path decomposition argument as in the proof of Proposition 17, we can find a collection of simple cycles $\alpha_1, \ldots, \alpha_k$ and a simple path β such that

$$|\pi| = \sum_{i=1}^{k} |\alpha_i| + |\beta|$$

$$|\beta| \leq |V_{\hat{G}}|$$

$$\nu(\pi) = \sum_{i=1}^{k} \nu(\alpha_i) \cdot \frac{|\alpha_i|}{|\pi|} + \nu(\beta) \cdot \frac{|\beta|}{|\pi|}. \tag{5}$$

Let $J \subseteq \{1, \ldots, k\}$ be the set of indices such that $j \in J$ if and only if $\nu(\alpha_j) = \mu$ (i.e., these cycles α_j are contained in T). Let $\mu' = \max_{j \notin J} \nu(\alpha_j)$.

Fix $\epsilon > 0$, pick

$$\delta = \frac{\epsilon(\mu - \mu')}{2|V_{\hat{G}}|}.$$

From (5)

$$\nu(\pi) = \sum_{i \in J} \mu \cdot \frac{|\alpha_i|}{|\pi|} + \sum_{i \notin J} \nu(\alpha_i) \cdot \frac{|\alpha_i|}{|\pi|} + \nu(\beta) \cdot \frac{|\beta|}{|\pi|}$$

$$\leq \mu \sum_{i \in J} \frac{\alpha_i}{|\pi|} + \mu' \sum_{i \notin J} \frac{\alpha_i}{|\pi|} + \frac{|V_{\hat{G}}|}{|\pi|}.$$

Let $\theta = \left(\sum_{i \notin J} |\alpha_i| + |\beta| \right) /|\pi|$. Note that θ is an upper bound on the fraction of edges in π not contained in T. By the assumptions on π in the statement of the lemma

$$\mu - \delta < \nu(\pi) < \mu(1 - \theta) + \mu'\theta + \frac{\epsilon(\mu - \mu')}{2}.$$

So

$$\mu - \frac{\epsilon(\mu - \mu')}{2} < \mu - \theta(\mu - \mu') + \frac{\epsilon(\mu - \mu')}{2}$$

which is equivalent to

$$\theta < \epsilon.$$

This completes the proof. ∎

For each pair of states (u, v) of a graph G, define $M_G(q, I_q, u, v)$ to be the number of words of length q satisfying I_q and generated by a path in G starting at u and ending at v.

Lemma 21: Let $T_1, \ldots, T_n$ be the irreducible components of T. Let t_i be the period of T_i. Let t be the least common multiple of $\{t_1, \ldots, t_n\}$. For any $\epsilon > 0$, there exists an integer L with the following property. Suppose that $\ell > L$ and ℓ is a multiple of t. Then there are a state $u \in V_T$ and a set $I_\ell' \subseteq \{1, \ldots, \ell\}$ such that $|I_\ell'|/\ell = \mu$ and

$$\max_{I_\ell \subseteq \{1, \ldots, \ell\}} \frac{\log M_T(\ell, I_\ell)}{\ell} < \epsilon + \frac{\log M_T(\ell, I_\ell', u, u)}{\ell}.$$

Proof: It is clear that T has no transitional edges, i.e., T is the union of its irreducible components. Since t_i is the period of the component T_i, by [10, Proposition 4.5.6], V_{T_i} can be partitioned into t_i period classes: $D(i, 1), \ldots, D(i, t_i)$. (Two states are in the same period class if they are connected by a path whose length is a multiple of the period of the graph. This is clearly an equivalence relation.)

For any integer ℓ that is a multiple of t, any path of length ℓ has initial and terminal states in the same period class. Thus, for any $I_\ell \subseteq \{1, \ldots, \ell\}$

$$M_T(\ell, I_\ell) = \sum_{i=1}^{n} \sum_{j=1}^{t_i} M_T^{(i,j)}(\ell, I_\ell) \tag{6}$$

where $M_T^{(i,j)}(\ell, I_\ell)$ is the number of words of length ℓ, satisfying I_ℓ and generated by a path in T beginning and ending in period class $D(i, j)$.

Let (i^*, j^*) maximize $M_T^{(i,j)}(\ell, I_\ell)$. Restricting our attention to words starting and ending in the period class $D(i^*, j^*)$, it follows from (6) that

$$M_T(\ell, I_\ell) \leq s M_T^{(i^*, j^*)}(\ell, I_\ell) \tag{7}$$

where $s = \sum_{i=1}^{n} t_i$.

Let A be the adjacency matrix of T, with the rows and columns grouped by period classes. From [10, pp. 127–128], A^t is block diagonal. Each block is primitive and corresponds to one of the period classes. It follows that for each period class $D(i, j)$, there exists an integer $k(i, j)$, a multiple of t, such that

for any states u and v in $D(i,j)$, there is a path from u to v of length k', for all $k' \geq k(i,j)$ that are multiples of t. Let

$$k = \max_{1 \leq i \leq n} \max_{1 \leq j \leq t_i} k(i,j).$$

Suppose that $\ell > k$. Let $q = \ell - k$. Define $I_q = I_\ell \cap \{1, \ldots, q\}$. Then,

$$M_T^{(i^*,j^*)}(\ell, I_\ell) \leq 2^k M_T^{(i^*,j^*)}(q, I_q). \tag{8}$$

Next, observe that there are states u and v in $D(i^*, j^*)$ such that

$$\frac{M_T^{(i^*,j^*)}(q, I_q)}{|D(i^*,j^*)|^2} \leq M_T(q, I_q, u, v). \tag{9}$$

Suppose that $M_T(q, I_q, u, v) = 0$. We pick any $I'_\ell \subseteq \{1, \ldots, \ell\}$ such that $|I'_\ell|/\ell = \mu$. Then $M_T(q, I_q, u, v) \leq M_T(\ell, I'_\ell, u, u)$ trivially holds. On the other hand, suppose that $M_T(q, I_q, u, v) \geq 1$. Let z be the label of a path from u to v of length q such that z satisfies I_q. Let z' be the label of a path from v to u of length k. Let $w = zz'$. Let $I'_\ell = \{i \in \{1, \ldots, \ell\} : w_i = \square\}$. By Lemma 19, $|I'_\ell|/\ell = \mu$ since w is the label of a cycle in T. Furthermore, the one-to-one correspondence $z \mapsto zz'$ gives

$$M_T(q, I_q, u, v) \leq M_T(\ell, I'_\ell, u, u). \tag{10}$$

From (7)–(10)

$$\begin{aligned} M_T(\ell, I_\ell) &\leq s2^k |D(i^*,j^*)|^2 M_T(\ell, I'_\ell, u, u) \\ &\leq s2^k |V_T|^2 M_T(\ell, I'_\ell, u, u). \end{aligned}$$

Choose L such that $L > k$ and

$$\frac{\log s2^k |V_T|^2}{L} < \epsilon.$$

Then the lemma follows. ∎

Proposition 22: $f_T(\mu) = f_T^{(p)}(\mu) = f_T^{(o)}(\mu).$

Proof: It suffices to show that $f_T(\mu) \leq f_T^{(p)}(\mu)$ and $f_T(\mu) \leq f_T^{(o)}(\mu)$ since the other direction is immediate.

Let $(I_m) \in \mathcal{I}(\mu)$ realize $f_T(\mu)$. Fix $\epsilon > 0$. Let t and L be as described in Lemma 21. Let q be a multiple of t such that $q > L$ and

$$f_T(\mu) < \epsilon + \frac{\log M_T(q, I_q)}{q}.$$

From Lemma 21, there exist a state $u \in V_T$ and a set $I'_q \subseteq \{1, \ldots, q\}$ such that $|I'_q|/q = \mu$ and

$$f_T(\mu) < \epsilon + \epsilon + \frac{\log M_T(q, I'_q, u, u)}{q}.$$

For each $l \in \mathbb{N}$, let

$$J_l = \{i \in \{1, \ldots, l\} : (i \bmod q) + 1 \in I'_q\}.$$

It is clear that $(J_l) \in \mathcal{I}_p(\mu)$. Now, for any $n \in \mathbb{N}$

$$\begin{aligned} f_T(\mu) &\leq 2\epsilon + \frac{\log \left(M_T(q, I'_q, u, u) \right)^n}{nq} \\ &\leq 2\epsilon + \frac{\log M_T(nq, J_{nq})}{nq}. \end{aligned}$$

Letting $n \to \infty$ yields

$$f_T(\mu) \leq 2\epsilon + f_T^{(p)}(\mu) \quad \text{and} \quad f_T(\mu) \leq 2\epsilon + f_T^{(o)}(\mu).$$

Since ϵ is arbitrary, it follows that

$$f_T(\mu) \leq f_T^{(p)}(\mu) \quad \text{and} \quad f_T(\mu) \leq f_T^{(o)}(\mu). \qquad ∎$$

Proposition 23: $f_{\hat{G}}(\mu) = f_T(\mu).$

Proof: It suffices to show $f_T(\mu) \geq f_{\hat{G}}(\mu)$, since the other direction is immediate.

Let $\epsilon > 0$. By Lemma 21, for each $n \in \mathbb{N}$, there exist an integer $\ell(n)$ and a set $I'_{\ell(n)} \subseteq \{1, \ldots, \ell(n)\}$ such that

$$\begin{aligned} |I'_{\ell(n)}|/\ell(n) &= \mu \\ \ell(n+1) &> \ell(n) \\ \max_{I \subseteq \{1,\ldots,\ell(n)\}} \frac{\log M_T(\ell(n), I)}{\ell(n)} &< \epsilon + \frac{\log M_T(\ell(n), I'_{\ell(n)})}{\ell(n)}. \end{aligned} \tag{11}$$

For each $n \in \mathbb{N}$, replace ϵ in Lemma 20 with $\epsilon/\ell(n)$ and choose an appropriate δ_n. Let $(I_m) \in \mathcal{I}(\mu)$ achieve $f_{\hat{G}}(\mu)$. Let $q(n)$ be a multiple of $\ell(n)$ such that

$$\frac{\log M_{\hat{G}}(q(n), I_{q(n)})}{q(n)} > f_{\hat{G}}(\mu) - \epsilon$$

$$\left| \frac{|I_{q(n)}|}{q(n)} - \mu \right| < \delta_n \tag{12}$$

$$q(n) > \frac{1}{\delta_n}. \tag{13}$$

Let Γ_n be a set of paths of length $q(n)$ with labels that satisfy $I_{q(n)}$ and $|\Gamma_n| = M_{\hat{G}}(q(n), I_{q(n)})$. By Lemma 20, (12) and (13) imply that any path in Γ_n has at most $\lfloor \epsilon q(n)/\ell(n) \rfloor$ edges that are not in T. Subdivide any path in Γ_n into $q(n)/\ell(n)$ nonoverlapping subpaths of length $\ell(n)$. Then there are at most $\lfloor \epsilon q(n)/\ell(n) \rfloor$ nonoverlapping subpaths that are not entirely in T.

Let $s(n) = \lfloor \epsilon q(n)/\ell(n) \rfloor$. We partition Γ_n into $s(n) + 1$ classes: $C_0, \ldots, C_{s(n)}$, where C_j is the set of paths with exactly j subpaths not entirely contained in T. We further partition each class C_j into $\binom{q(n)/\ell(n)}{j}$ subclasses corresponding to the positions of subpaths that are not entirely in T. From (11), the number of paths in each subclass of C_j is at most

$$\left(2^{\ell(n)} \right)^j \times \left(2^{\epsilon\ell(n)} M_T(\ell(n), I'_{\ell(n)}) \right)^{\frac{q(n)}{\ell(n)} - j}.$$

Since $\sum_{k=0}^{m} \binom{m}{k} = 2^m$, we have $\binom{q(n)/\ell(n)}{j} \leq 2^{q(n)/\ell(n)}$. Thus,

$$
\begin{aligned}
& M_{\hat{G}}(q(n), I_{q(n)}) \\
& \leq \sum_{j=0}^{s(n)} 2^{\frac{q(n)}{\ell(n)}} \left(2^{\ell(n)}\right)^j \left(2^{\epsilon \ell(n)} M_T(\ell(n), I'_{\ell(n)})\right)^{\frac{q(n)}{\ell(n)}-j} \\
& \leq (s(n)+1) 2^{\frac{q(n)}{\ell(n)}} \left(2^{\ell(n)}\right)^{s(n)} \left(2^{\epsilon \ell(n)} M_T(\ell(n), I'_{\ell(n)})\right)^{\frac{q(n)}{\ell(n)}} \\
& \leq (\epsilon q(n)+1) 2^{\frac{q(n)}{\ell(n)}} 2^{\epsilon q(n)} 2^{\epsilon q(n)} M_T(\ell(n), I'_{\ell(n)})^{\frac{q(n)}{\ell(n)}}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\frac{\log M_{\hat{G}}(q(n), I_{q(n)})}{q(n)} & \leq \frac{\log M_T(\ell(n), I'_{\ell(n)})}{\ell(n)} + 2\epsilon \\
& \quad + \frac{1}{\ell(n)} + \frac{\log(\epsilon q(n)+1)}{q(n)} \\
\limsup_{n\to\infty} \frac{\log M_{\hat{G}}(q(n), I_{q(n)})}{q(n)} & \leq \limsup_{n\to\infty} \frac{\log M_T(\ell(n), I'_{\ell(n)})}{\ell(n)} \\
& \quad + 2\epsilon + 0 + 0 \\
f_{\hat{G}}(\mu) - \epsilon & \leq f_T(\mu) + 2\epsilon.
\end{aligned}
$$

Since ϵ is arbitrary, it follows that $f_{\hat{G}}(\mu) \leq f_T(\mu)$. ∎

Corollary 24: $f_{\hat{G}}(\mu) = f_{\hat{G}}^{(p)}(\mu) = f_{\hat{G}}^{(o)}(\mu)$.

Proof: Since T is a subgraph of $\hat{G}$

$$
f_{\hat{G}}^{(p)}(\mu) \geq f_T^{(p)}(\mu) \quad \text{and} \quad f_{\hat{G}}^{(o)}(\mu) \geq f_T^{(o)}(\mu).
$$

By the definitions of $f^{(p)}$ and $f^{(o)}$

$$
f_{\hat{G}}(\mu) \geq f_{\hat{G}}^{(p)}(\mu) \quad \text{and} \quad f_{\hat{G}}(\mu) \geq f_{\hat{G}}^{(o)}(\mu).
$$

Then the corollary follows from Propositions 22 and 23. ∎

B. Maximum Insertion Rates for RLL and MTR

Next we present explicit formulas for the maximum insertion rates of RLL and MTR constraints. Instead of finding a max-insertion-rate cycle in $\hat{G}$, we find a word that can be freely concatenated and has the maximum ratio of $\square$ to its length. This approach is equivalent to finding a max-insertion-rate cycle by Lemma 8, as can be seen by the following argument. Suppose that $w \in \hat{S}$ can be freely concatenated. Then there is a cycle in $\hat{G}$ with label w^k for some k. Thus, the ratio of $\square$ in w is at most the maximum insertion rate. The converse is obvious; given a max-insertion-rate cycle, the label of the cycle can freely concatenated and its density of $\square$ is exactly the maximum insertion rate.

Theorem 25:

$$
\mu(\text{RLL}(d, k)) = \frac{\left\lfloor \frac{k-d}{d+1} \right\rfloor}{\left\lfloor \frac{k+1}{d+1} \right\rfloor (d+1)}.
$$

Proof: We will find a word w that can be freely concatenated and has the maximum density of $\square$. We first claim that, every run of 0 in w has length d. To see this, if there is a run of

0 with length greater than d, we replace this run with a run of length d. The new word still satisfies the constraint and induces a higher insertion rate.

Thus, it suffices to assume that w has the form

$$
a_1 \, 0^d \, a_2 \, 0^d \, \ldots \, a_\ell \, 0^d
$$

where a_i is a sequence comprising only 1 and $\square$. If $d \geq 1$, a sequence of 1 and $\square$ of length greater than one will violate the constraint; thus, each a_i is a single symbol of 1 or $\square$. If $d = 0$, we can write w as a word of length ℓ: $a_1 \ldots a_\ell$, where $a_i \in \{1, \square\}$.

Now, from the expression of w, it is clear that the maximum insertion rate is at most $1/(d+1)$. Suppose that there are at least two 1 in w. If there are $\leq k - d - 1$ symbols between two consecutive 1, we insert $0^d \square$ after the first 1 and the insertion rate does not decrease (because the ratio of $\square$ of the inserted word is $1/(d+1) \geq$ maximum possible insertion rate). Thus, any portion of w between two consecutive 1 is $1(0^d \square)^m 0^d 1$, where m is the largest integer such that $(d+1)m + d \leq k$, i.e., $m = \lfloor (k-d)/(d+1) \rfloor$. Hence, w can be assumed to have the form $(1(0^d \square)^m 0^d)^n$, and the insertion rate is

$$
\begin{aligned}
\frac{m}{(m+1)(d+1)} &= \frac{\left\lfloor \frac{k-d}{d+1} \right\rfloor}{\left(\left\lfloor \frac{k-d}{d+1} \right\rfloor + 1\right)(d+1)} \\
&= \frac{\left\lfloor \frac{k-d}{d+1} \right\rfloor}{\left\lfloor \frac{k+1}{d+1} \right\rfloor (d+1)}.
\end{aligned}
$$
∎

Theorem 26: Let S be the MTR (j, k) constraint. If $\gcd(j+1, k+1) \neq 1$

$$
\mu(S) = 1 - \frac{1}{j+1} - \frac{1}{k+1}.
$$

If $\gcd(j+1, k+1) = 1$, let m be the smallest positive integer such that $m(j+1) = k \bmod (k+1)$, and let n be the smallest positive integer such that $n(j+1) = 1 \bmod (k+1)$. Then

$$
\mu(S) = \begin{cases} L_1, & \text{if } m > n \\ \max\{L_0, L_1\}, & \text{if } m < n \end{cases}
$$

where

$$
\begin{aligned}
L_0 &= 1 - \frac{n}{n(j+1) - 1} - \frac{1}{k+1} \\
L_1 &= 1 - \frac{1}{j+1} - \frac{m(j+1) + 1}{m(j+1)(k+1)}.
\end{aligned}
$$

Proof: Let w be a word that can be freely concatenated. We first show that w can be assumed to start with 10. Let c be the smallest integer such that $1\square^c 0$ is a subword of ww. Then we can cyclically shift the positions of all 0 in w to the left by c to obtain another word with the same ratio of $\square$. This word contains 10 as a subword, and we can cyclically shift the positions of all 0 and 1 so that the word starts with 10.

Starting with the prefix 10, we will show that an optimal strategy to construct a word that has the maximum ratio of $\square$
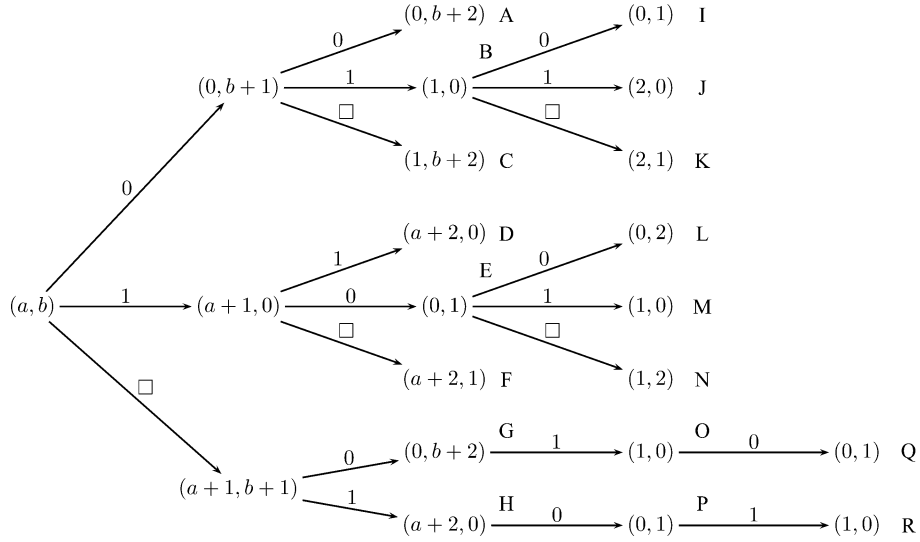
Fig. 7. Choices of appending symbols to a word with state (a, b).

is to keep appending $\square$ until the constraint is almost violated. Given a word with prefix 10, define

$$a = |w| - \max_{w_i = 0} i, \quad b = |w| - \max_{w_i = 1} i.$$

For example, if $w = 10\square\square0\square$, then $a = 1$ and $b = 5$. This pair (a, b) keeps track of the last 0 and 1 in w. We can view (a, b) as the "state" of w. In the above example, the history of (a, b) is given by

$$(0, 1) \xrightarrow{\square} (1, 2) \xrightarrow{\square} (2, 3) \xrightarrow{0} (0, 4) \xrightarrow{\square} (1, 5).$$

If S is the MTR (j, k), then $w \in \hat{S}$ only if $a \leq j$ and $b \leq k$. Moreover, it specifies what symbol can be append to w. In other words, $\mathcal{F}_{\hat{S}}(w)$ can be determined from (a, b). It is easy to check that if (a, b) is the state of a word w, (c, d) is the state of a word w', $a \leq c$, and $b \leq d$, then $\mathcal{F}_{\hat{S}}(w') \subseteq \mathcal{F}_{\hat{S}}(w)$.

Let w and w' be words in $\hat{S}$ with prefix 10. Let (a, b) and (c, d) be the states of w and w', respectively. We wish to have a way to compare which word is better as a prefix. If z is a word in $\mathcal{A}_3$, define $B(z)$ to be the number of $\square$ in z. If both of the following are true, we prefer prefix w over w' to construct a word with maximum ratio of $\square$ to its length.

a) $B(w) - B(w') \geq |w| - |w'| \geq 0$,
b) $(a, b) \leq (c, d)$ componentwise.

To see this, first note that if $w'z$ is a word that can be freely concatenated, then so is wz; this is because $\mathcal{F}_{\hat{S}}(w') \subseteq \mathcal{F}_{\hat{S}}(w)$ by b), and w and w' have prefix 10. Since $B(w'z) \leq |w'z|$ and $|w| \geq |w'|$

$$\begin{aligned} \frac{B(w'z)}{|w'z|} &\leq \frac{B(w'z) + |w| - |w'|}{|w'z| + |w| - |w'|} \\ &\leq \frac{B(w'z) + B(w) - B(w')}{|wz|} \\ &\leq \frac{B(wz)}{|wz|} \end{aligned}$$

where the second inequality follows from a). Hence, wz has a better ratio of $\square$ than $w'z$.

TABLE I
ADDING $\square$ IS BETTER WHEN (14) HOLDS

choices	better alternative	remark
A,C	G	
B	O	if $b \leq k - 2$
D,F	H	
E	P	if $a \leq j - 2$
I,K	P	if $b = k - 1$ (and so $a \leq j - 2$)
J	R	if $b = k - 1$ (and so $a \leq j - 2$)
L	Q	if $a = j - 1$ (and so $b \leq k - 2$)
M,N	O	if $a = j - 1$ (and so $b \leq k - 2$)

Let $w \in \hat{S}$ with state (a, b). Suppose that

$$a \neq j, \quad b \neq k, \quad (a, b) \neq (j - 1, k - 1). \qquad (14)$$

Then we can append $\square$ to w and still be able to continue, i.e., $\mathcal{F}_{\hat{S}}(w\square) \neq \{\epsilon\}$. We will show that it is optimal to append $\square$ when (14) holds. Fig. 7 shows some choices, denoted by A–R, of appending symbols to a word w with state (a, b). From the discussion above, it is clear that G is better than A. Table I summarizes all choices and shows that it is optimal to add $\square$ whenever (14) is true.

Now we have a strategy to construct word w. We start with $w = 10$ and add $\square$ when (14) holds; we add 0 when $a = j$ and add 1 when $b = k$. We keep adding symbols until the word ends with 10 again. But what should we add when $(a, b) = (j - 1, k - 1)$?

Case 1: $\gcd(j + 1, k + 1) \neq 1$. In this case, we will show that the state $(j - 1, k - 1)$ will never occur if we start with state $(0, 1)$. By using the above strategy, there are exactly k symbols between two consecutive 1 and exactly j symbols between two consecutive 0. If $(j - 1, k - 1)$ occurs, there is an integer x such that $x = (j - 1) \bmod (j + 1)$ and $x + 1 = (k - 1) \bmod (k + 1)$. Equivalently, there are integers y and z such that $x = y(j + 1) - 2 = z(k + 1) - 3$. But this contradicts the assumption that $\gcd(j + 1, k + 1) \neq 1$, so $(j - 1, k - 1)$ does not occur. When the state $(0, 1)$ occurs again, the word length (not including the last 10) is $\operatorname{lcm}(j + 1, k + 1)$ and there are $(\operatorname{lcm}(j + 1, k + 1))/(j + 1)$
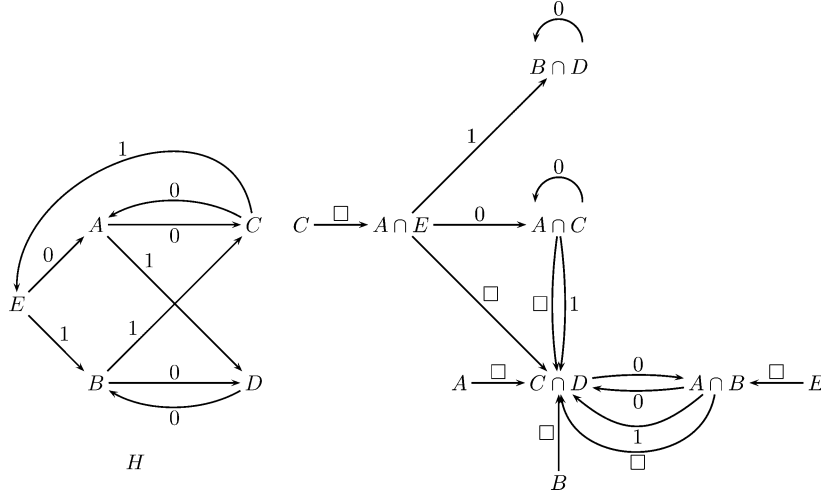
Fig. 8. Subgraphs of $\hat{G}$ that can be reached from H for Example 28.

and $(\mathrm{lcm}(j+1, k+1))/(k+1)$ numbers of 0 and 1, respectively. Thus the maximum insertion rate is

$$1 - \frac{\frac{\mathrm{lcm}(j+1,k+1)}{j+1} + \frac{\mathrm{lcm}(j+1,k+1)}{k+1}}{\mathrm{lcm}(j+1,k+1)} = 1 - \frac{1}{j+1} - \frac{1}{k+1}.$$

Case 2: $\gcd(j+1, k+1) = 1$. Without loss of generality, we assume that the state sequence

$$s_0 = (0, 1) \longrightarrow s_1 \longrightarrow \cdots \longrightarrow s_{|w|} = (0, 1)$$

does not contain the same state twice except the initial and terminal states (by the same idea as Proposition 17). Since m is the smallest integer such that $m(j + 1) = k \bmod (k + 1)$, $s_{m(j+1)-2} = (j - 1, k - 1)$. Let $p = m(j + 1) - 2$. There are two choices to continue adding symbols after this state.

i)

$$s_p \xrightarrow{\;1\;} s_{p+1} = (j, 0) \xrightarrow{\;0\;} s_{p+2} = (0, 1).$$

Now we return to the state $(0, 1)$ again and finish constructing w. We have $|w| = p+2 = m(j+1)$; the number of 0 in w is m; the number of 1 in w is $(m(j+1)+1)/(k+1)$. Thus, the maximum insertion rate in this case is

$$L_1 = 1 - \frac{1}{j+1} - \frac{m(j+1)+1}{m(j+1)(k+1)}.$$

ii)

$$s_p \xrightarrow{\;0\;} s_{p+1} = (0, k) \xrightarrow{\;1\;} s_{p+2} = (1, 0).$$

From state $(1, 0)$, we wish to determine whether we will see state $(0, 1)$ or $(j-1, k-1)$ first. If we reach $(0, 1)$ first, then $(0, 1) = s_{t+p+2}$ for some t where $t+1 = 0 \bmod (j+1)$ and $t = 1 \bmod (k+1)$. Hence, $t+p+2 = 0 \bmod (k+1)$. Moreover, there is an integer x such that $t = x(j+1)-1$ and thus,

$$t+p+2 = x(j+1)-1+m(j+1) = (x+m)(j+1)-1.$$

This implies that $x+m$ is an integer such that $(x+m) \cdot (j+1) = 1 \bmod (k+1)$. Since n is the smallest such integer, if $m > n$ then $x+m > k+1$.

If we reach $(j - 1, k - 1)$ first, then $(j - 1, k - 1) = s_{y+p+2}$ for some y where $y + 1 = j - 1 \bmod (j + 1)$

and $y = k - 1 \bmod (k + 1)$. Hence, $y + p + 2 = k - 2 \bmod (k + 1)$. Moreover, there is an integer z such that $y = z(j + 1) - 3$ and thus

$$y+p+2 = z(j+1)-3+m(j+1) = (z+m)(j+1)-3.$$

This implies that $(z + m)(j + 1) = 0 \bmod (k + 1)$. It follows that $z + m = k + 1$.

If $m > n$, then $x > z$ and hence $t > y$, which means that we will reach state $(j - 1, k - 1)$ before $(0, 1)$. This contradicts the assumption that no state is repeated twice. Hence, we must choose i) when $m > n$. If $m < n$, then $x+m = n < k+1$ and hence $t < y$. We will reach $(0, 1)$ first and w is completed. In this case, $|w| = t + p + 2 = n(j + 1) - 1$; the number of 0 is n; the number of 1 is $(n(j + 1) - 1)/(k + 1)$. Thus, the maximum insertion rate is

$$L_0 = 1 - \frac{n}{n(j + 1) - 1} - \frac{1}{k + 1}. \qquad \blacksquare$$

VI. PROPERTIES OF THE TRADEOFF FUNCTION

From Proposition 5, the Shannon cover for a constrained system S is isomorphic to a subgraph H of $\hat{G}$. Recall also that G' is the irreducible component of $\hat{G}$ that contains H. The following corollary shows that if S is finite-type, it suffices to use only G' to compute the tradeoff function. This result is immediate from Proposition 11.

Corollary 27: Let S be a finite-type constrained system. Then $f(\rho) = f_{G'}(\rho)$.

The following example shows that Corollary 27 may not hold for a constraint that is not finite-type.

Example 28: Let S be the constrained system presented by the graph H in Fig. 8. It is easy to check that H is the Shannon cover for S. The subgraph of $\hat{G}$ that can be reached from H is given in two separate components for ease of viewing. Since H has a cycle at A, C with label 00, S is not finite-type.

From the figure, if we start from a state in H and follow an edge with label $\square$, we cannot return to H. Thus, the irreducible component of $\hat{G}$ containing H is H itself. Since H does not have any edge with label $\square$, $f_H(\rho) = -\infty$ for all $\rho > 0$. Consider the component with states $A \cap B$ and $C \cap D$. This component can generate a word $0x0\square$, where x can be 0 or 1. This word can be freely concatenated and still satisfies $\hat{S}$. Therefore, $f(1/4) \geq 1/4 > -\infty = f_H(1/4)$. This implies that Corollary 27 may not hold for a non finite-type constraint.

Proposition 29: Let S be a constrained system. Let f be the tradeoff function and let μ be the maximum insertion rate for S. Then f is decreasing on $[0, \mu]$. In fact

$$f(\rho') \geq \rho - \rho' + f(\rho). \tag{15}$$

Proof: Let $0 \leq \rho' < \rho \leq \mu$ and $(I_q) \in \mathcal{I}(\rho)$. Let J_q be the set of the smallest $\min\{\lfloor(\rho-\rho')q\rfloor, |I_q|\}$ entries of I_q. Then

$$\lim_{q \to \infty} \frac{|J_q|}{q} = \min\left\{\lim_{q \to \infty} \frac{\lfloor(\rho-\rho')q\rfloor}{q}, \lim_{q \to \infty} \frac{|I_q|}{q}\right\}$$
$$= \min\{\rho - \rho', \rho\} = \rho - \rho'. \tag{16}$$

Let $I'_q = I_q \setminus J_q$. It follows from (16) that $(I'_q) \in \mathcal{I}(\rho')$. By filling in $\square$ at positions in J_q of every word of length q that satisfies I_q with 0 and 1, we obtain

$$M(q, I'_q) \geq 2^{|J_q|} M(q, I_q)$$
$$\limsup_{q \to \infty} \frac{\log M(q, I'_q)}{q} \geq \lim_{q \to \infty} \frac{|J_q|}{q} + \limsup_{q \to \infty} \frac{\log M(q, I_q)}{q}$$
$$= \rho - \rho' + \limsup_{q \to \infty} \frac{\log M(q, I_q)}{q}.$$

It follows that

$$f(\rho') \geq \rho - \rho' + f(\rho).$$

Thus f is decreasing on $[0, \mu]$. ∎

The following proposition gives a continuity property of the tradeoff function. Later, we will show that f is continuous on $[0, \mu]$.

Proposition 30: Let S be a constrained system. Let f be the tradeoff function and let μ be the maximum insertion rate for S. Then f is left-continuous on $[0, \mu]$.

Proof: Fix $\rho \in [0, \mu]$ and let $\rho_n \to \rho$ with $\rho_n < \rho$. We will show that $f(\rho_n) \to f(\rho)$ as $n \to \infty$.

Since

$$f(\rho_n) = \sup_{(I_q) \in \mathcal{I}(\rho_n)} \limsup_{q \to \infty} \frac{\log M(q, I_q)}{q}$$

for each $n \in \mathbb{N}$, we can choose $q_n \in \mathbb{N}$ and a set $I^n_{q_n} \subseteq \{1, \ldots, q_n\}$ such that

$$\frac{\log M(q_n, I^n_{q_n})}{q_n} \geq f(\rho_n) - \frac{1}{n}$$
$$\left|\frac{|I^n_{q_n}|}{q_n} - \rho_n\right| \leq \frac{1}{n}$$
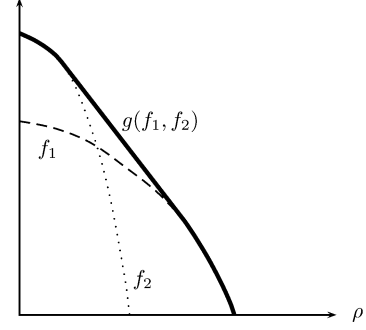$$q_{n+1} > q_n.$$

Fig. 9. An example of $g(f_1, f_2)$.

Let (I^0_q) be an arbitrary sequence in $\mathcal{I}(\rho)$. Now define a sequence (I_q) by

$$I_q = \begin{cases} I^n_{q_n}, & \text{if } q = q_n \text{ for some } n \\ I^0_q, & \text{otherwise.} \end{cases}$$

We have $|I^0_q|/q \to \rho$ and by the triangle inequality, we also have

$$\left|\frac{|I^n_{q_n}|}{q_n} - \rho\right| \leq \left|\frac{|I^n_{q_n}|}{q_n} - \rho_n\right| + |\rho_n - \rho|$$
$$\leq \frac{1}{n} + |\rho_n - \rho| \to 0$$

so $(I_q) \in \mathcal{I}(\rho)$. For the code rate, we have

$$f(\rho) \geq \limsup_{q \to \infty} \frac{\log M(q, I_q)}{q}$$
$$\geq \limsup_{n \to \infty} \frac{\log M(q_n, I^n_{q_n})}{q_n}$$
$$\geq \limsup_{n \to \infty} f(\rho_n).$$

By Proposition 29, f is decreasing, so $f(\rho_n) \geq f(\rho)$ for all n. It follows that

$$\limsup_{n \to \infty} f(\rho_n) \geq \liminf_{n \to \infty} f(\rho_n) \geq f(\rho).$$

Thus, $f(\rho_n) \to f(\rho)$. ∎

We will make use of the following lemma to prove several important properties of the tradeoff function f. The proof of the Lemma is given in the Appendix .

Lemma 31: Let G be an irreducible graph with alphabet $\mathcal{A}_3$. Let $0 \leq \rho \leq 1$ and $(I_q) \in \mathcal{I}(\rho)$. There is $(I'_q) \in \mathcal{I}(\rho)$ such that

$$\lim_{q \to \infty} \frac{\log M(q, I'_q)}{q} = \limsup_{q \to \infty} \frac{\log M(q, I_q)}{q}.$$

Let $f_i : [0, 1] \to [-\infty, \infty)$, $i \in \{1, \ldots, n\}$, be functions. We define $g(f_1, \ldots, f_n) : [0, 1] \to [-\infty, \infty]$ to be the smallest concave function such that $g(f_1, \ldots, f_n)(\rho) \geq f_i(\rho)$ for all $i \in \{1, \ldots, n\}$ and $\rho \in [0, 1]$. Fig. 9 shows an example of $g(f_1, f_2)$ to illustrate the definition.

In order to express $g(f_1, \ldots, f_n)(\rho)$ in terms of f_i, we need some results from convex analysis, in particular the Caratheodory's theorem [13, Theorem 17.1]. The following proposition, which appeared as [13, Corollary 17.1.3], is a consequence of the theorem.

Proposition 32 ([13]): Let $\{f_i : i \in I\}$ be an arbitrary collection of proper convex functions on $\mathbb{R}$, and let f be the largest convex function such that $f_i(x) \geq f(x)$ for all $x \in \mathbb{R}$. Then, for every x

$$f(x) = \inf \left\{ \sum_{i \in I} \lambda_i f_i(x_i) : \sum_{i \in I} \lambda_i x_i = x \right\}$$

where the infimum is taken over all expressions of x as a convex combination in which at most two of the coefficients λ_i are nonzero.

We remark that, according to the proofs of [13, Corollary 17.1.1] and [13, Corollary 17.1.3], each f_i need not be convex if we do not require that it appears only once in the combination. In other words, even if f_i are not convex

$$f(x) = \inf \left\{ \sum_{1 \leq k \leq 2} \lambda_k f_{i_k}(x_k) : \sum_{1 \leq k \leq 2} \lambda_k x_k = x, \ i_k \in I \right\}. \tag{17}$$

We use this result to prove the following lemma.

Lemma 33: Let $f_i : [0,1] \to [-\infty, \infty)$, $i \in \{1, \ldots, n\}$, be functions. For any $\rho \in [0,1]$

$$g(f_1, \ldots, f_n)(\rho) = \sup \ \theta f_i(x) + (1 - \theta) f_j(y)$$

where the supremum is subject to
- $\theta, x, y \in [0,1]$,
- $i, j \in \{1, \ldots, n\}$,
- $\theta x + (1 - \theta) y = \rho$.

Proof: By definition, $-g(f_1, \ldots, f_n)$ is the largest convex function such that $-g(f_1, \ldots, f_n)(\rho) \leq -f_i(\rho)$. Define $\bar{f}_i : \mathbb{R} \to [-\infty, \infty)$ by $\bar{f}_i(z) = f_i(z)$ if $z \in [0,1]$ and $\bar{f}_i(z) = -\infty$ otherwise. For $\rho \in [0,1]$, it follows from (17) that

$$-g(f_1, \ldots, f_n)(\rho) = \inf \ \theta(-\bar{f}_i(x)) + (1 - \theta)(-\bar{f}_j(y))$$
$$g(f_1, \ldots, f_n)(\rho) = \sup \ \theta \bar{f}_i(x) + (1 - \theta) \bar{f}_j(y) \tag{18}$$

where $\theta \in [0,1]$, $\theta x + (1 - \theta) y = \rho$, and $i, j \in \{1, \ldots, n\}$. If $\theta = 0$, then $y = \rho \in [0,1]$ and we can assume that $x \in [0,1]$ since $\theta \bar{f}_i(x) + (1 - \theta) \bar{f}_j(y)$ does not vary with x. Similarly, if $\theta = 1$, then $x, y \in [0,1]$. If $\theta \in (0,1)$ and either x or y is not in $[0,1]$, then $\theta \bar{f}_i(x) + (1 - \theta) \bar{f}_j(y) = -\infty$. Thus, we can limit the possible values of x and y to be in $[0,1]$ without changing the value of the right-hand side of (18), and the lemma follows. ∎

Lemma 34: Let $f_i : [0,1] \to [-\infty, \infty)$, $i \in \{1, \ldots, n\}$, be functions. For any fixed $\rho \in [0,1]$

$$g(f_1, \ldots, f_n)(\rho) = \max_{i,j \in \{1,\ldots,n\}} g(f_i, f_j)(\rho).$$

Proof: It is obvious from the definition of g that

$$g(f_1, \ldots, f_n)(\rho) \geq \max_{i,j \in \{1,\ldots,n\}} g(f_i, f_j)(\rho).$$

To show the other direction, let $\rho \in [0,1]$. From Lemma 33, there are $\theta_k, x_k, y_k \in [0,1]$, and $i_k, j_k \in \{1, \ldots, n\}$ such that $\theta_k x_k + (1 - \theta_k) y_k = \rho$ and

$$\lim_{k \to \infty} \theta_k f_{i_k}(x_k) + (1 - \theta_k) f_{j_k}(y_k) = g(f_1, \ldots, f_n)(\rho).$$

Let $(k(m))$ be a subsequence of (k) such that all $i_{k(m)}$ are equal to some i and all $j_{k(m)}$ are equal to some j. Then

$$g(f_i, f_j)(\rho) \geq \lim_{m \to \infty} \theta_{k(m)} f_i(x_{k(m)}) + (1 - \theta_{k(m)}) f_j(y_{k(m)})$$
$$= g(f_1, \ldots, f_n)(\rho). \quad \blacksquare$$

Lemma 35: Let $G_1, \ldots, G_n$ be the irreducible components of $\hat{G}$. Then

$$f(\rho) \leq g(f_{G_1}, \ldots, f_{G_n})(\rho).$$

The proof of Lemma 35 is given in the Appendix . The key idea in the proof is that every path in $\hat{G}$ must follow a chain of irreducible components, but there are finitely many such chains, so there must exist one chain, $G_1 \to G_2 \to \cdots \to G_c$, whose code rate dominates in the following sense. For a given ρ, there exists $\theta_i \geq 0$, ρ_i, for $i = 1, \ldots, c$ such that
   a) $\sum_{i=1}^{c} \theta_i = 1$,
   b) $\sum_{i=1}^{c} \theta_i \rho_i = \rho$,
   c) $\sum_{i=1}^{c} \theta_i f_{G_i}(\rho_i) \geq f(\rho)$.
Then use the inequality

$$\sum_{i=1}^{c} \theta_i f_{G_i}(\rho_i) \leq g(f_{G_1}, \ldots, f_{G_c})(\rho) \leq g(f_{G_1}, \ldots, f_{G_n})(\rho)$$

to complete the proof.

The next lemma introduces an important concept that we will use to prove subsequent results. Suppose that G_1 and G_2 are irreducible components of a graph G such that G_2 can be reached from G_1. Clearly, for any ρ_1 and ρ_2, $f_G(\rho_1) \geq f_{G_1}(\rho_1)$ and $f_G(\rho_2) \geq f_{G_2}(\rho_2)$. The lemma states that the graph of f_G lies above the straight line connecting $(\rho_1, f_{G_1}(\rho_1))$ and $(\rho_2, f_{G_2}(\rho_2))$. We prove the lemma by using the concept of "time-sharing," which can be described roughly as follows. Suppose that $\rho_1 < \rho < \rho_2$. We concatenate the sequences in G_1 with insertion rate ρ_1 and the sequences in G_2 with insertion rate ρ_2 to obtain sequences with insertion rate ρ. Then we show that $f_G(\rho)$ must be at least the weighted average of $f_{G_1}(\rho_1)$ and $f_{G_2}(\rho_2)$.

Lemma 36: Let G be a graph with alphabet $\mathcal{A}_3$. Let G_1 and G_2 be irreducible components of G such that G_2 can be reached from G_1. Then

$$g(f_{G_1}, f_{G_2})(\rho) \leq f_G(\rho).$$

Proof: Let $\rho \in [0,1]$. Let $\theta, \rho_1, \rho_2 \in [0,1]$ satisfy $\theta \rho_1 + (1 - \theta) \rho_2 = \rho$. Let $j_1, j_2 \in \{1, 2\}$. (Note that j_1 and j_2 may be the same.) For $i \in \{1, 2\}$ and $I \subseteq \mathbb{N}$, let $M_i(q, I)$ be the number of words of length q in $S(G_{j_i})$ that satisfy I. Let (I_q^i) be a sequence in $\mathcal{I}(\rho_i)$ such that $(\log M(q, I_q^i))/q$ converges. By Lemma 31, such a sequence exists and it suffices to consider only all such sequences.

Let $(a(n))$ and $(b(n))$ be strictly increasing sequences of positive integers such that $a(n)/(a(n) + b(n))$ converges to θ and $a(n+1) + b(n+1) > a(n) + b(n) + |V_G|$ for all n. For each $n \in \mathbb{N}$, we define $q(n)$ and $I'_{q(n)}$ as follows. Let $p_1 = a(n)$ and $p_2 = b(n)$. For each $i \in \{1, 2\}$, let $J_i = I_{p_i}^i$. For each word w_i of length p_i whose labels satisfy J_i, choose exactly one path that contributes to $M_i(p_i, J_i)$. Let Γ_i be the set of such paths. There exist states u_i and v_i in G_{j_i} such that the number

of paths in Γ_i with initial state u_i and terminal state v_i is at least $M_i(p_i, J_i)/|V_G|^2$. Let z_i be a word of length p_i that satisfies J_i.

Case 1: G_{j_1} can be reached from G_{j_2}. There is a path of length at most $|V_G|$ from v_2 to u_1. Let x be the label of this path. Let $w = z_2 x z_1$.

Case 2: G_{j_2} can be reached from G_{j_1}. There is a path of length at most $|V_G|$ from v_1 to u_2. Let x be the label of this path. Let $w = z_1 x z_2$.

Set $q(n) = |w|$ and

$$I'_{q(n)} = \{i \in \{1, \ldots, q(n)\} : w_i = \square\}.$$

It is clear that $q(n)$ is strictly increasing with n. Choose an arbitrary sequence $(K_p) \in \mathcal{I}(\rho)$. We extend the sequence $(I'_{q(n)})$ to (I'_p) for all positive integers p by setting $I'_p = K_p$ if there is no n such that $q(n) = p$.

Then

$$\frac{|I'_{q(n)}|}{q(n)} \geq \frac{|I^1_{a(n)}| + |I^2_{b(n)}|}{a(n) + b(n) + |V_G|}$$

$$= \frac{a(n)}{a(n) + b(n) + |V_G|} \times \frac{|I^1_{a(n)}|}{a(n)}$$

$$+ \frac{b(n)}{a(n) + b(n) + |V_G|} \times \frac{|I^2_{b(n)}|}{b(n)}$$

$$\liminf_{n \to \infty} \frac{|I'_{q(n)}|}{q(n)} \geq \theta\rho_1 + (1 - \theta)\rho_2 = \rho.$$

Moreover

$$\frac{|I'_{q(n)}|}{q(n)} \leq \frac{|I^1_{a(n)}| + |I^2_{b(n)}| + |V_G|}{a(n) + b(n)}$$

$$\limsup_{n \to \infty} \frac{|I'_{q(n)}|}{q(n)} \leq \rho.$$

It follows that $(I_p) \in \mathcal{I}(\rho)$.

Furthermore

$$M(q(n), I'_{q(n)}) \geq \frac{M_1(a(n), I^1_{a(n)})}{|V_G|^2} \times \frac{M_2(b(n), I^2_{b(n)})}{|V_G|^2}$$

$$\frac{\log M(q(n), I'_{q(n)})}{q(n)} \geq \frac{\log M_1(a(n), I^1_{a(n)})}{a(n) + b(n) + |V_G|}$$

$$+ \frac{\log M_2(b(n), I^2_{b(n)})}{a(n) + b(n) + |V_G|} - \frac{4\log|V_G|}{q(n)}.$$

Hence,

$$f(\rho) \geq \limsup_{q \to \infty} \frac{\log M(q, I'_q)}{q}$$

$$\geq \lim_{n \to \infty} \frac{\log M(q(n), I'_{q(n)})}{q(n)}$$

$$\geq \lim_{n \to \infty} \frac{\log M_1(a(n), I^1_{a(n)})}{a(n) + b(n) + |V_G|}$$

$$+ \lim_{n \to \infty} \frac{\log M_2(b(n), I^2_{b(n)})}{a(n) + b(n) + |V_G|} - \lim_{n \to \infty} \frac{4\log|V_G|}{q(n)}$$

$$= \theta \lim_{n \to \infty} \frac{\log M_1(a(n), I^1_{a(n)})}{a(n)}$$

$$+ (1 - \theta) \lim_{n \to \infty} \frac{\log M_2(b(n), I^2_{b(n)})}{b(n)}.$$

Since this is true for any sequence $(I^i_q) \in \mathcal{I}(\rho_i)$ such that $(\log M_i(q, I^i_q))/q$ converges, $f(\rho) \geq \theta f_{j_1}(\rho_1) + (1-\theta)f_{j_2}(\rho_2)$. Since this holds for any $j_1, j_2 \in \{1, 2\}$ and $\theta, \rho_1, \rho_2 \in [0, 1]$ satisfying $\theta\rho_1 + (1 - \theta)\rho_2 = \rho$, we have $f(\rho) \geq g(f_1, f_2)(\rho)$ by Lemma 33. ∎

Proposition 37: Let G be an irreducible graph. Then f_G is concave.

Proof: Letting $G = G_1 = G_2$ in Lemma 36, we obtain $f_G \geq g(f_G, f_G)$. By definition, $g(f_G, f_G) \geq f_G$ and $g(f_G, f_G)$ is concave. ∎

In the two lemmas below, we will use the following notations. Let $G_1, \ldots, G_n$ be the irreducible components of $\hat{G}$. Define

$$P = \{(i, j) \in \{1, \ldots, n\}^2 : G_i \text{ can reach } G_j\}.$$

By letting $G = \hat{G}$ in Lemma 36, we obtain the following.

Lemma 38:

$$\max_{(i,j) \in P} g(f_{G_i}, f_{G_j})(\rho) \leq f(\rho).$$

Lemma 39: Let S be an irreducible constraint. Then

$$\max_{i,j \in \{1, \ldots, n\}} g(f_{G_i}, f_{G_j})(\rho) \leq \max_{(i,j) \in P} g(f_{G_i}, f_{G_j})(\rho).$$

Proof: Let G_i and G_j be irreducible components of $\hat{G}$. By Proposition 10, there exist irreducible components $G_{i'}$ and $G_{j'}$ such that $G_{i'}$ can reach H, H can reach $G_{j'}$, $S(G_i) \subseteq S(G_{i'})$, and $S(G_j) \subseteq S(G_{j'})$. Thus, $(i', j') \in P$ and $g(f_{G_i}, f_{G_j})(\rho) \leq g(f_{G_{i'}}, f_{G_{j'}})(\rho)$. This completes the proof. ∎

Lemmas 34, 35, 38, and 39 together establish a chain of inequalities. It follows from these inequalities that

$$f(\rho) = g(f_{G_1}, \ldots, f_{G_n})(\rho)$$

where $G_1, \ldots, G_n$ are the irreducible components of $\hat{G}$. This observation along with Proposition 10 give the following theorem and corollary, which are the main properties of the tradeoff function.

Theorem 40: Let S be an irreducible constrained system. Let G' be the irreducible component of $\hat{G}$ that contains H. Let $G_1, \ldots, G_k = G'$ be the irreducible components of $\hat{G}$ that can be reached from H. Let $G' = G_k, \ldots, G_m$ be the irreducible components of $\hat{G}$ that can reach H. Then

$$f(\rho) = g(f_{G_1}, \ldots, f_{G_k})(\rho) = g(f_{G_k}, \ldots, f_{G_m})(\rho).$$

Corollary 41: The tradeoff function f for an irreducible constraint S is concave. Moreover, f is continuous on $[0, \mu]$, where μ is the maximum insertion rate for S.

Proof: Let f be the tradeoff function for S. From Theorem 40, f is concave on $[0, 1]$. From Proposition 29, $f(\rho) > -\infty$ for $\rho \in [0, \mu)$. Thus, f is continuous on $(0, \mu)$. Since f is nonincreasing and concave, it is continuous at 0. By Proposition 30, f is left continuous at μ. ∎

From Theorem 40, we can compute the tradeoff function from either the subgraph of $\hat{G}$ that can reach H or the subgraph that can be reached from H. Thus, we can avoid constructing the whole graph $\hat{G}$ but still be able to compute the tradeoff function.

Next we describe an algorithm for constructing the subgraph $\tilde{G}$ of $\hat{G}$ that can be reached from H. First we need the following notation. Let G be a graph. Let $x \in S(G)$ and let $U \subseteq V_G$. Recall that $T(x, U)$ to be the set of terminal states of all paths from U with label x.

CONSTRUCTION OF THE SUBGRAPH $\tilde{G}$ OF $\hat{G}$ REACHABLE FROM H

Input: Shannon cover G
$V_{\tilde{G}} \leftarrow \{\{v\} : v \in V_G\}$
Let $V_1, \ldots, V_{|V_{\tilde{G}}|}$ be the elements in $V_{\tilde{G}}$
$k \leftarrow 1$
while $k \leq |V_{\tilde{G}}|$
    for $b = 0, 1$
        if every state in V_k has an outgoing edge with label b
          if $T(b, V_k) \notin V_{\tilde{G}}$
            $V_{|V_{\tilde{G}}|+1} \leftarrow T(b, V_k)$
            $V_{\tilde{G}} \leftarrow V_{\tilde{G}} \cup \{T(b, V_k)\}$
          Assign the edge $V_k \xrightarrow{b} T(b, V_k)$
        if every state in V_k has two outgoing edges
      $T_\square \leftarrow T(0, V_k) \cup T(1, V_k)$
      if $T_\square \notin V_{\tilde{G}}$
        $V_{|V_{\tilde{G}}|+1} \leftarrow T_\square$
        $V_{\tilde{G}} \leftarrow V_{\tilde{G}} \cup \{T_\square\}$
      Assign the edge $V_k \xrightarrow{\square} T_\square$
    $k \leftarrow k + 1$
Reduce $\tilde{G}$ using, e.g., the proof of [10, Theorem 3.4.14]
Output: $\tilde{G}$

We remark that there is a constraint whose tradeoff function f is not equal to $f_{\tilde{G}}$.

VII. ON THE TRADEOFF FUNCTIONS $f^{(p)}$ AND $f^{(o)}$

In this section, we will explore some properties of the periodic tradeoff function $f^{(p)}$ and the tradeoff function $f^{(o)}$ given by [1]. Both functions were defined earlier in Section IV. We will also establish their relation to the tradeoff function f. Recall from the end of Section IV that the periodic tradeoff function $f^{(p)}$ is equal to $f_{\hat{G}}^{(p)}$ and similarly, $f^{(o)}$ is equal to $f_{\hat{G}}^{(o)}$. As in the previous sections, the maximum insertion rate of $\hat{G}$ is given by μ.

Proposition 42: The periodic tradeoff function $f^{(p)}(\rho)$ is nonincreasing on rationals $0 \leq \rho \leq \mu$.

Proof: Let $\rho_1 < \rho_2$ be rational numbers in $[0, \mu]$. Pick $(I_q) \in \mathcal{I}_p(\rho_2)$ with period N and periodic set $U \subseteq \{1, \ldots, N\}$. Choosing a common denominator, write $\rho_1 = a/c$, $\rho_2 = b/c$. Without loss of generality, we can assume N is divisible by c (since any multiple of N is also a period for (I_q)). Then let $U' = U \setminus J$ where J is any fixed subset of U of size $b - a$. This corresponds to filling in $b - a$ boxes in each period for the configuration given by (I_q). Let (I'_q) be the periodic sequence determined by N and U'. Then $(I'_q) \in \mathcal{I}_p(\rho_1)$ and $M(I'_q, q) \geq M(I_q, q)$. So $f^{(p)}(\rho_1) \geq f^{(p)}(\rho_2)$, showing that $f^{(p)}$ is nonincreasing. $\blacksquare$

Lemma 43: Let S be an irreducible finite-type constrained system. Let $\rho \leq \mu$ and let $\epsilon > 0$. There is a rational number $\tilde{\rho}$ with $|\tilde{\rho} - \rho| < \epsilon$ and $f^{(p)}(\tilde{\rho}) > f(\rho) - \epsilon$.

Proof: If S is finite-type, then by Corollary 27, it suffices to use only the only nontrivial irreducible component G' of $\hat{G}$ to compute the tradeoff functions, i.e., $f_{\hat{G}} = f_{G'}$.

For each pair of states (u, v) in $V_{G'}$, we will again use $M_{G'}(q, I_q, u, v)$ to denote the number of words of length q satisfying I_q and generated by a path in G' beginning at u and ending at v.

Fix $\epsilon > 0$. By the definition of $f(\rho)$, there exist a $q_0 \in \mathbb{N}$ and $I'_{q_0} \subseteq \{1, \ldots, q_0\}$ such that

$$\frac{\log M_{G'}(q_0, I'_{q_0})}{q_0} \geq f(\rho) - \frac{\epsilon}{3}$$

$$\left| \frac{|I'_{q_0}|}{q_0} - \rho \right| \leq \frac{\epsilon}{2}$$

$$\frac{q_0}{q_0 + |V_{G'}|} \geq 1 - \frac{\epsilon}{3},$$

$$\frac{2 \log |V_{G'}|}{q_0} \leq \frac{\epsilon}{3}.$$

Let (σ^*, τ^*) be a pair of states such that

$$M_{G'}(q_0, I'_{q_0}, \sigma^*, \tau^*) \geq \frac{M_{G'}(q_0, I'_{q_0})}{|V_{G'}|^2}.$$

Since $\rho \leq \mu$, $M_{G'}(q_0, I'_{q_0}) \geq 1$. Thus, there is a path from σ^* to τ^* with label, denoted by α, satisfying I'_{q_0}. Let x be the label of a path from τ^* to σ^* such that $|x| \leq |V_{G'}|$. Such a path exists since G' is irreducible. Let $c = |x|$ and let k be the number of $\square$ in x.

Now define $I \subseteq \mathbb{N}$ by putting $i \in I$ iff the ith position of the right-infinite sequence $(\alpha x)^\infty$ is $\square$. Then let $(I_q)_{q=1}^\infty$ be the periodic sequence given by

$$I_q = I \cap \{1, \ldots, q\}.$$

Let $(w_1, \ldots, w_n)$ be an n-tuple, where w_i is the label of a path beginning at σ^* and ending at τ^* and w_i satisfies I'_{q_0}. We map $(w_1, \ldots, w_n)$ to the sequence $w_1 x w_2 x \cdots w_n x$ which is generated by the cycle

$$\sigma^* \xrightarrow{w_1} \tau^* \xrightarrow{x} \sigma^* \xrightarrow{w_2} \tau^* \longrightarrow \cdots \longrightarrow \sigma^* \xrightarrow{w_n} \tau^* \xrightarrow{x} \sigma^*.$$

The mapping $(w_1, \ldots, w_n) \mapsto w_1 x w_2 x \ldots w_n x$ is one-to-one. Thus, for any n, there are at least $\left(M_{G'}(q_0, I'_{q_0}) / |V_{G'}|^2 \right)^n$ codewords of length $n(q_0 + c)$ satisfying $I_{n(q_0+c)}$.

The new insertion rate $\tilde{\rho}$ induced by (I_q) is thus

$$\frac{|I'_{q_0}| + k}{q_0 + c} = \frac{|I'_{q_0}|}{q_0} \times \frac{q_0}{q_0 + c} + \frac{k}{q_0 + c}$$
$$\leq \rho + \epsilon,$$

since $k/(q_0 + c) \leq c/(q_0 + c) \leq |V_{G'}|/(q_0 + |V_{G'}|) \leq \epsilon/3$. Also,

$$\frac{|I'_{q_0}| + k}{q_0 + c} \geq \left(\rho - \frac{\epsilon}{2} \right) \left(1 - \frac{\epsilon}{3} \right) + \frac{k}{q_0 + c}$$
$$\geq \left(\rho - \frac{\epsilon}{2} \right) \left(1 - \frac{\epsilon}{2} \right)$$
$$\geq \rho - \epsilon.$$

Thus, $|\tilde{\rho} - \rho| \leq \epsilon$.

The code rate is bounded as follows:

$$f_{\hat{G}}^{(p)}(\tilde{\rho}) \geq \limsup_{n \to \infty} \frac{\log M_{G'}(n(q_0 + c), I_{n(q_0 + c)})}{n(q_0 + c)}$$

$$\geq \limsup_{n \to \infty} \frac{\log \left(\frac{M_{G'}(q_0, I'_{q_0})}{|V_{G'}|^2} \right)^n}{n(q_0 + c)}$$

$$= \frac{\log M_{G'}(q_0, I_{q_0})}{q_0} \cdot \frac{q_0}{q_0 + c} - \frac{2 \log |V_{G'}|}{q_0 + c}$$

$$\geq \left(f_{\hat{G}}(\rho) - \frac{\epsilon}{3} \right) \left(1 - \frac{\epsilon}{3} \right) - \frac{\epsilon}{3}$$

$$\geq f_{\hat{G}}(\rho) - \epsilon. \qquad \blacksquare$$

Using the above lemma, we show in the following theorem that for finite-type constraints, periodic insertions can achieve $f(\rho)$ for rational $0 \leq \rho \leq \mu$. This justifies the intuitive usage of periodic insertions in certain (finite-type) constrained block codes [1], [14].

Theorem 44: Let S be an irreducible finite-type constrained system. Then $f^{(p)}(\rho) = f(\rho)$ on rationals $0 \leq \rho \leq \mu$.

Proof: By Corollary 24, $f^{(p)}(\mu) = f(\mu)$, so it suffices to prove the result for rational $0 \leq \rho < \mu$.

Fix a rational $\rho < \mu$. Fix $\epsilon_0 > 0$. We will show that $f^{(p)}(\rho) \geq f(\rho) - \epsilon_0$. By Corollary 41, $f(\rho)$ is continuous at ρ, so there exists $\delta > 0$ such that $|f(\rho) - f(\rho')| < \epsilon_0/2$ whenever $|\rho - \rho'| < \delta$. Pick ρ' to be a rational number in $(\rho, \rho + \min\{\delta, \mu - \rho\})$. Applying Lemma 43 with $\epsilon = \min\{\rho' - \rho, \epsilon_0/2\}$, and ρ' in place of ρ, we obtain

$$f^{(p)}(\tilde{\rho}) > f(\rho') - \frac{\epsilon_0}{2} > f(\rho) - \epsilon_0$$

where $\tilde{\rho} \geq \rho$. Since $f^{(p)}$ is nonincreasing by Proposition 42, we have

$$f^{(p)}(\rho) \geq f^{(p)}(\tilde{\rho}) > f(\rho) - \epsilon_0.$$

This holds for all $\epsilon_0 > 0$, so $f^{(p)}(\rho) \geq f(\rho)$. $\qquad \blacksquare$

We remark that $f^{(p)}$ is not equal to f in general because periodic time-sharing between any two components of $\hat{G}$ is not achievable.

Corollary 45: The tradeoff function f for a finite-type constrained system is the uniformly continuous extension of $f^{(p)}$ to the domain $[0, \mu]$.

The following proof for $f^{(o)}$ being nonincreasing on rational ρ is similar to that of Proposition 42.

Proposition 46: The tradeoff function $f^{(o)}(\rho)$ is nonincreasing on rationals $0 \leq \rho \leq \mu$.

Proof: Let $\rho_1 < \rho_2$ be rational numbers in $[0, \mu]$. In lowest terms, write $\rho_1 = a/b$ and $\rho_2 = c/d$, where a, b are coprime, and c, d are coprime. Let q_n be an increasing sequence, $I_{q_n} \subseteq \{1, \ldots, q_n\}$ with $|I_{q_n}|/q_n = \rho_2$. We can rewrite ρ_1 as r_n/s_n where $s_n = b\lceil q_n/b \rceil$, so that $q_n \leq s_n \leq q_n + b$. Now

$$r_n = a \left\lceil \frac{|I_{q_n}| d}{cb} \right\rceil$$

$$\leq \frac{ad}{cb} |I_{q_n}| + a$$

$$\leq |I_{q_n}|$$

for all sufficiently large n since $ad < cb$ by $\rho_1 < \rho_2$. Let $I'_{q_n} = I_{q_n} \setminus J_n$, where J_n is a subset of I_{q_n} of size $|I_{q_n}| - r_n$. This corresponds to filling in $|I_{q_n}| - r_n$ boxes for the configuration given by (I_{q_n}). Then extend the sequence to length s_n by a path of length $s_n - q_n$ with no boxes. This forms a new sequence (I''_{s_n}) such that $M(I''_{s_n}, s_n) \geq M(I_{q_n}, q_n)$. Since s_n and q_n differ by at most b, $f^{(o)}(\rho_1) \geq f^{(o)}(\rho_2)$, showing that $f^{(o)}$ is nonincreasing. $\qquad \blacksquare$

The following proposition reconciles our definition of the tradeoff function f with that given by [1].

Theorem 47: Let S be an irreducible constrained system. Then $f(\rho) = f^{(o)}(\rho)$ on rationals $0 \leq \rho \leq \mu$.

Proof: We first show that $f(\rho) \geq f^{(o)}(\rho)$ for rational ρ in $[0, \mu]$. From the definition of $f^{(o)}(\rho)$, there exists a subsequence of pairs (q, I_q) such that

$$\frac{\log M(I_q, q)}{q} \to f^{(o)}(\rho)$$

where $|I_q|/q = \rho$.

Using this subsequence, we can form a single $(I'_q) \in \mathcal{I}(\rho)$ such that

$$f(\rho) \geq \limsup_{q \to \infty} \frac{\log M(q, I'_q)}{q} = f^{(o)}(\rho).$$

To show the other direction, first note that $f^{(o)}(\mu) = f(\mu)$ by Corollary 24, so it suffices to prove $f^{(o)}(\rho) \geq f(\rho)$ for rational $0 \leq \rho < \mu$.

Fix $\rho < \mu$. Let $\rho = k/N$ in lowest terms, so k, N are coprime. Fix $\epsilon > 0$. By Corollary 41, $f(\rho)$ is continuous at ρ, so there exists $\delta > 0$ such that $|f(\rho) - f(\tilde{\rho})| < \epsilon$ whenever $|\rho - \tilde{\rho}| < \delta$. Fix a rational $\tilde{\rho} \in (\rho, \min\{\rho + \delta, \mu\})$.

Let $(I_q) \in \mathcal{I}(\tilde{\rho})$ achieve $f(\tilde{\rho})$. For any sequence $\tilde{\alpha}_q$ of length q satisfying I_q, we construct a new sequence α_q that has insertion rate exactly ρ (by filling in extra boxes and extending the length). The new sequence will induce the code rate equal to $f(\tilde{\rho})$.

For each q, rewrite $\rho = k/N = r_q/s_q$, where $s_q = N\lceil q/N \rceil$, so that $q \leq s_q \leq q + N$. For all sufficiently large q

$$r_q = k \left\lceil \frac{q}{N} \right\rceil$$

$$\leq \frac{kq}{N} + k = \frac{\rho q}{|I_q|} |I_q| + k$$

$$\leq |I_q|$$

since $\rho q/|I_q| \to \rho/\tilde{\rho} < 1$. Thus, we can fill in $|I_q| - r_q$ boxes within α_q and extend the length of $\tilde{\alpha}_q$ by a small amount $s_q - q$, using only a path with no boxes in its label. Denote the resulting set of box positions by I'_{s_q}. The sequence α_q obtained by this procedure has insertion rate exactly ρ, length s_q, and satisfies I'_{s_q}.

Now

$$M(q, I_q) \leq M(s_q, I'_{s_q})$$

$$\frac{\log M(q, I_q)}{q} \leq \frac{\log M(s_q, I'_{s_q})}{s_q} \times \frac{s_q}{q}$$

$$f(\tilde{\rho}) = \limsup_{q \to \infty} \frac{\log M(q, I_q)}{q} \leq \limsup_{q \to \infty} \frac{\log M(s_q, I'_{s_q})}{s_q} \times 1$$

$$\leq f^{(o)}(\rho).$$

This implies that $f^{(o)}(\rho) \geq f(\rho) - \epsilon$ for all $\epsilon > 0$. Thus, $f^{(o)}(\rho) \geq f(\rho)$. ∎

VIII. BOUNDS FOR TRADEOFF FUNCTIONS

A. General Bounds

The following proposition gives a simple, but somewhat poor, upper bound on the tradeoff function f.

Proposition 48: For $0 \leq \rho \leq \mu$,

$$f(\rho) \leq \operatorname{cap}(S) - \rho.$$

Proof: Let $\rho' = 0$ and $f(\rho') = \operatorname{cap}(S)$ in (15). ∎

Let G be a deterministic graph. We will present bounds for the tradeoff function f_G. First, we can use Proposition 16 to derive a greedy lower bound as follows. Let $A_G = A + B$ where A corresponds to the edges with label 0 and 1 in G and B corresponds to the edges with label $\square$. Let C be a square matrix. We recall that $\lambda(C)$ denotes the spectral radius of C. For a given number c, the following algorithm returns an insertion rate ρ such that $f_G(\rho) \geq c$.

GREEDY LOWER BOUND
    Inputs: A, B, a real number c, an integer N
    $C \leftarrow A$
    $b \leftarrow 0$
    for $1 \leq k \leq N - 1$
        if $((\lambda(C))^{1/k} \geq c$ and $\lambda(CBA^{N-k-1})^{1/N} \geq c)$
            then $C \leftarrow CB$
                $b \leftarrow b + 1$
        else $C \leftarrow CA$
    Output: b/N.

At the end, C is a product $\prod_{i=1}^{N} C_i$ where C_i is either A or B. The algorithm ensures that $(\log \lambda(C))/N \geq c$. By Proposition 16, c is a lower bound for $f_G(b/N)$, where b is the number of B in the product $\prod_{i=1}^{N} C_i$.

Next we offer another method to compute a lower bound and an approximate upper bound. This method is based on the concept of dynamic programming [4, Ch. 16].

DYNAMIC PROGRAMMING LOWER BOUND
    Inputs: A, B, an integer N
    $C_0 \leftarrow A$
    $C_1 \leftarrow B$
    for $2 \leq \ell \leq N$
        $C_0' = C_0 A$
        for $1 \leq b \leq \ell - 1$
            $D \leftarrow C_{b-1} B$
            $E \leftarrow C_b A$
            if $(\lambda(D) \geq \lambda(E))$
                then $C_b' \leftarrow D$
                else $C_b' \leftarrow E$
        $C_\ell' \leftarrow C_{\ell-1} B$
        $C_j \leftarrow C_j'$ for all $0 \leq j \leq \ell$
    Outputs: $(\log \lambda(C_j))/N$ for all $0 \leq j \leq N$

When the preceding algorithm terminates, C_j is a product consisting of $N - j$ copies of A and j copies of B. Hence, $(\log \lambda(C_j))/N$ is a lower bound for $f_G(j/N)$.

Next we give an algorithm to compute an approximate upper bound for f_G.

DYNAMIC PROGRAMMING APPROXIMATE UPPER BOUND
    Inputs: A, B, an integer N
    $x_0 \leftarrow$ all-one vector of dimension $|V_G|$
    for $1 \leq \ell \leq N$
        $x_0' \leftarrow A x_0$
        for $1 \leq b \leq \ell - 1$
            $y \leftarrow B x_{b-1}$
            $z \leftarrow A x_b$
            $x_b' \leftarrow \max\{y, z\}$ componentwise
        $x_\ell' \leftarrow B x_{\ell-1}$
        $x_j \leftarrow x_j'$ for all $0 \leq j \leq \ell$
    Outputs: $\dfrac{\log \sum_{1 \leq k \leq |V_G|} x_j(k)}{N}$ for all $0 \leq j \leq N$ (where $x_j(k)$ is the kth entry of vector x_j)

At the end of the algorithm, $\sum_k x_j(k)$ represents, for all $I \subseteq \{1, \ldots, N\}$ such that $|I| = j$, an upper bound on the number of words $w = w_1 \ldots w_N \in S(G)$ satisfying $w_i = \square$ if and only if $i \in I$. Note that this is an approximate upper bound as N is finite, but it converges to a true upper bound as $N \to \infty$.

To compute bounds for the tradeoff function f, we can replace G with $\hat{G}$ in the preceding algorithms. However, in practice, we do not use the whole graph $\hat{G}$. Instead, we compute bounds $h_1, \ldots, h_k$ for the irreducible components $G_1, \ldots, G_k$ that can be reached from H. Then, the convex combination $g(h_1, \ldots, h_k)$ is a bound for f by Theorem 40. In particular, by Corollary 27, we use only the irreducible component G' when the constraint is finite-type.

We further remark that for a fixed constraint, other lower bounds for f can be derived by considering specific bit-insertion schemes. This is shown in the next section for the case of MTR constraints.

B. Bit-Stuffing Lower Bounds for MTR Constraints

A lower bound for MTR constraints can be obtained by considering *bit-stuffing* schemes (see [1] and [6]). The bit-stuffing scheme for the MTR (j) constraint is given as follows. For b-bit-stuffing, begin with a string s of length N that satisfies the MTR $(j - b)$ constraint and subdivide s into intervals of size $j - b + 1$; then, in between each of these intervals, insert a string of b ones. The resulting string satisfies the MTR (j) constraint and has insertion rate $\rho = b/(j+1)$ with asymptotic optimal code rate (as $N \to \infty$), of $\frac{j-b+1}{j+1} \log \lambda_{j-b}$. Thus, Campello *et al.* [1] established that the piecewise-linear curve connecting the $j + 1$ points

- $(0, \log \lambda_j)$
- $(\frac{1}{j+1}, \frac{j}{j+1} \log \lambda_{j-1})$
- $(\frac{2}{j+1}, \frac{j-1}{j+1} \log \lambda_{j-2})$
- $\ldots$
- $(\frac{j-1}{j+1}, \frac{2}{j+1} \log \lambda_1)$
- $(\frac{j}{j+1}, \frac{1}{j+1} \log \lambda_0) = (\mu(\mathrm{MTR}\,(j)), 0)$

is a concave lower bound for $f_{\mathrm{MTR}\,(j)}$.

We remark that such a scheme can be imitated for the MTR (j, k) constraint to obtain similar lower bounds, which we describe below.

Since MTR (j, k) has a symmetry between j and k (by swapping 0 and 1), we assume without loss of generality (WLOG) that $j \leq k$. Let $b \leq \min(j, k) - 1$. We define a b-bit stuffing scheme for the MTR (j, k) constraint as follows. Begin with a string s that satisfies the MTR $(j - b, k - b)$ constraint. Subdivide s into intervals of length $k - b + 1$. In between each of these intervals, insert a string of b ones. It is easily checked that the resulting string satisfies the MTR (j, k) constraint and has parity insertion rate $\frac{b}{k+1}$. The set of all such strings of a fixed length N can be viewed as a block code, and the asymptotic optimal code rate of such codes as $N \rightarrow \infty$ is $\frac{k-b+1}{k+1} \log \lambda_{j-b,k-b}$ (where $\lambda_{j-b,k-b}$ denotes the largest eigenvalue of the standard adjacency matrix $A_{(j-b,k-b)}$ for the MTR $(j - b, k - b)$ constraint).

For $S = \text{MTR} (j, k)$ and parity insertion rate ρ, let $h_S(\rho)$ denote the code rate obtained by the weighted average of b-bit stuffing and $(b + 1)$-bit stuffing, as described above. The graph of $h_S(\rho)$ is the piecewise-linear curve that connects the $j + 1$ points

- $(0, \log \lambda_{j,k})$
- $(\frac{1}{k+1}, \frac{k}{k+1} \log \lambda_{j-1,k-1})$
- $(\frac{2}{k+1}, \frac{k-1}{k+1} \log \lambda_{j-2,k-2})$
- $\cdots$
- $(\frac{j-1}{k+1}, \frac{k-j+2}{k+1} \log \lambda_{1,k-(j-1)})$
- $(\mu, 0)$.

We remark that the value of μ can be found separately from Theorem 26.

For $j = k$, the above reduces to the piecewise-linear curve connecting the following j points:

- $(0, \log \lambda_{j,j})$
- $(\frac{1}{j+1}, \frac{j}{j+1} \log \lambda_{j-1,j-1})$
- $(\frac{2}{j+1}, \frac{j-1}{j+1} \log \lambda_{j-2,j-2})$
- $\cdots$
- $(\frac{j-2}{j+1}, \frac{3}{j+1} \log \lambda_{2,2})$
- $(\frac{j-1}{j+1}, \frac{2}{j+1} \log \lambda_{1,1}) = (\mu, 0)$.

Proposition 49: The curve $h_S(\rho)$ on $[0, \mu]$ is a lower bound to $f(\rho)$.

Note that the first j points are achieved by b-bit stuffing schemes, which induces periodic insertions. The last point $(\mu, 0)$ is also achievable by a max-insertion-rate cycle. Any point on the line connecting two adjacent points can be achieved by time-sharing.

Example 50: The MTR $(2, 3)$ has capacity 0.7947 and maximum insertion rate 0.3750 (see Theorem 26). Taking $N = 1000$, the solid line in Fig. 10 shows the greedy lower bound of f for MTR $(2, 3)$. The dotted line is the upper bound in Proposition 48. The dashed line is a lower bound from the GREEDY LOWER BOUND algorithm with the condition $\lambda(C)^{1/k} \geq c$ removed from the if statement. This dashed line is very close to the straight line between $(0, \text{cap}(S))$ and $(\mu, 0)$.

Taking $N = 1000$ again, the solid line in Fig. 11 shows the dynamic programming lower bound and the dotted line is the approximate upper bound. The dashed line is the analytical lower bound based on bit-stuffing schemes.
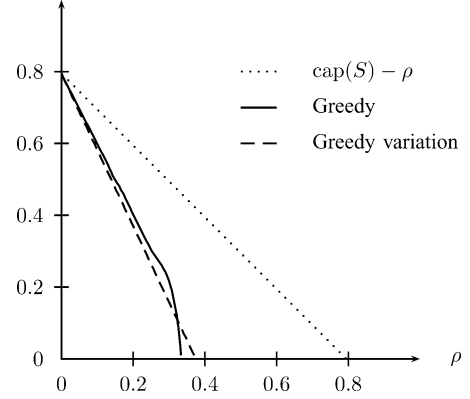

Fig. 10. Greedy lower bounds (solid and dashed) and the upper bound $\text{cap}(S) - \rho$ in Proposition 48 (dotted) of the tradeoff function for the MTR $(2, 3)$.
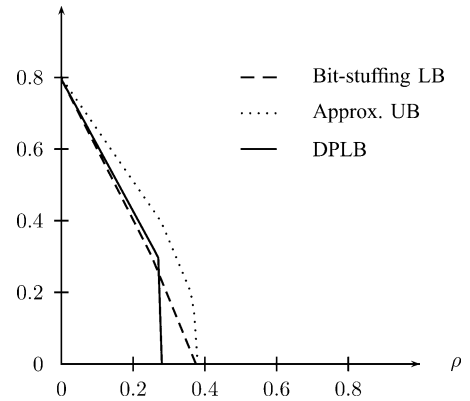

Fig. 11. Bit-stuffing lower bound (dashed), dynamic programming lower bound (solid), and approximate upper bound (dotted) of the tradeoff function for the MTR $(2, 3)$.
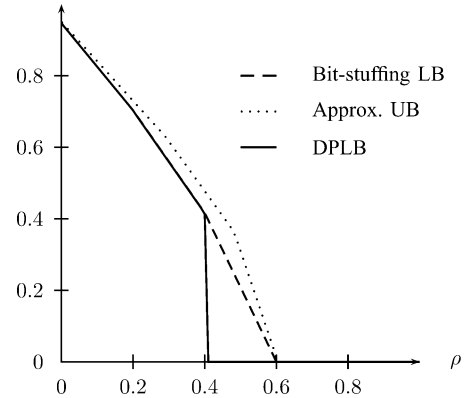

Fig. 12. Bit-stuffing lower bound (dashed), dynamic programming lower bound (solid) and approximate upper bound (dotted) of the tradeoff function for the MTR(4, 4).

By Corollary 41, the tradeoff function f is concave. Thus, if f_1 and f_2 are lower bounds for f, their convex combination $g(f_1, f_2)$ is also a lower bound for f. This technique can be used to construct a better lower bound.

Example 51: The MTR $(4, 4)$ has capacity 0.9468 and maximum insertion rate 0.6 (see Theorem 26). Taking $N = 1000$, the solid line in Fig. 12 shows the dynamic programming lower bound and the dotted line is the approximate upper bound. The dashed line is the analytical lower bound based on bit-stuffing schemes.

IX. Some Properties of the Tradeoff Functions for Certain RLL and MTR Constraints

A. NRZ Versus NRZI Tradeoff Functions for RLL $(0, k)$ and MTR (j)

In recording channels, binary data is often represented in two ways: the nonreturn to zero inverted (NRZI) and the nonreturn to zero (NRZ) representations. Let $x = x_1 \ldots x_n$ be a binary sequence in NRZI. The corresponding NRZ sequence y is given by $y_i = y_{i-1} \oplus x_i$. Thus, given a sequence in one representation, there are two corresponding sequences in the other representation, depending on the value of y_0. For example, 001011 in NRZI corresponds to 001101 and 110010 in NRZ; 001101 in NRZ corresponds to 001011 and 101011 in NRZI. Suppose that S_1 is a constrained system in NRZI and S_2 is the NRZ-precoded version of S_1. It can be shown that $\operatorname{cap}(S_1) = \operatorname{cap}(S_2)$.

Our definitions of RLL and MTR constraints use the NRZI representation. Here, we show that the tradeoff functions of RLL $(0, k)$ and MTR (j) constraints are greater than their respective NRZ counterparts.

Proposition 52: Let S_1 be the RLL $(0, k)$ constraint or the MTR (j) constraint. Let S_2 be the NRZ-precoded version of S_1. Then for $0 \le \rho \le 1$

$$f_{S_2}(\rho) \le f_{S_1}(\rho).$$

Proof: Letting $q \in \mathbb{N}$ and $I \subseteq \{1, \ldots, q\}$, define $M_i(q, I)$ to be the number of words in $\hat{S}_i$ of length q satisfying I. Let $\rho \in [0, 1]$. By Proposition 12, there is a sequence $(I_q) \in \mathcal{I}(\rho)$ such that

$$\limsup_{q \to \infty} \frac{\log M_2(q, I_q)}{q} = f_{S_2}(\rho).$$

Let W_q be the set of words in $\hat{S}_2$ of length q satisfying I_q.

For each q, we construct a set Y_q of words of length q from W_{q+1} as follows.

1) If $1 \in I_{q+1}$, replace the first position of every word in W_{q+1} with 1 to obtain W'_{q+1}. Otherwise, let $W'_{q+1} = W_{q+1}$.
2) For each $w \in W'_{q+1}$, define $\gamma(w)$ to be the word of length q as follows.
   Case 1: S_1 is the RLL $(0, k)$. If $w_{i+1} \ne \square$, let $\gamma(w)_i = w_{i+1} \oplus w_j$ where j is the largest index such that $j \le i$ and $w_j \ne \square$. If $w_{i+1} = \square$, let $\gamma(w)_i = \square$.
   Case 2: S_1 is the MTR (j). If $w_{i+1} \ne \square$, let $\gamma(w)_i = w_{i+1} \oplus w_j \oplus (i - j \bmod 2)$ where j is the largest index such that $j \le i$ and $w_j \ne \square$. If $w_{i+1} = \square$, let $\gamma(w)_i = \square$.
3) Let $Y_q = \{\gamma(w) \mid w \in W'_{q+1}\}$.

Observe the following properties of Y_q.

- $Y_q \subseteq \hat{S}_1$ and all $y \in Y_q$ satisfy

$$I'_q = \{i - 1 : i \in I_{q+1} \setminus \{1\}\}.$$

- The number $|I'_q|$ of unconstrained positions of Y_q satisfies $|I_{q+1}| - 1 \le |I'_q| \le |I_{q+1}|$. This is because we neglect at most one unconstrained position when constructing W'_{q+1}.

- $|Y_q| \ge |W_{q+1}|/2$. This follows from the fact that, given $y \in Y_q$, there are at most two words, w and w', in W'_{q+1} such that $\gamma(w) = \gamma(w') = y$. These words are the "flipped" version of each other.

From the second property, $(I'_q) \in \mathcal{I}(\rho)$. Hence,

$$f_{S_1}(\rho) \ge \limsup_{q \to \infty} \frac{\log M_1(q, I'_q)}{q} \ge \limsup_{q \to \infty} \frac{\log |Y_q|}{q}$$

$$\ge \limsup_{q \to \infty} \frac{\log \left(\frac{|W_{q+1}|}{2} \right)}{q} = f_{S_2}(\rho). \quad \blacksquare$$

B. On the RLL $(0, 1)$ and MTR $(2, 2)$ Tradeoff Functions

The RLL $(0, 1)$ constraint (see Fig. 3) is equivalent to the MTR $(j = 1)$ constraint, except that the labels are reversed $(0 \to 1, 1 \to 0)$. Thus, RLL $(0, 1)$ has the same capacity as MTR (1), which is 0.6942. From Theorem 25, the maximum insertion rate for RLL $(0, 1)$ is $1/2$.

The tradeoff functions for MTR $(j = 1)$ and MTR $(j = 2)$ are given in [1, Theorem 3]. We restate the theorem specifically for RLL $(0, 1)$ in a proposition below.

Proposition 53 ([1]): Let S denote the RLL $(0, 1)$ constraint. Then f_S is given by the straight line connecting the points $(0, 0.6942)$ and $(1/2, 0)$.

The MTR $(2, 2)$ constraint is the NRZ version of the RLL $(0, 1)$ constraint, so at zero insertion rate, the optimal code rate is also 0.6942. The Shannon cover for MTR $(2, 2)$ is given in Fig. 13. From Theorem 26, the maximum insertion rate for MTR $(2, 2)$ is $1/3$. Clearly, any point on the straight line connecting the points $(0, 0.6942)$ and $(1/3, 0)$ can be achieved by weighted bit-stuffing schemes, and the line presents a lower bound to the tradeoff function for MTR $(2, 2)$. We further show that this lower bound curve is in fact the tradeoff function itself.

Proposition 54: Let S denote the MTR $(2, 2)$ constraint. Then f_S is given by the straight line connecting the points $(0, 0.6942)$ and $(1/3, 0)$.

Proof: We will use similar proof techniques from [1] for this proof. The essence of this proof is to use a matrix inequality to imply an eigenvalue inequality: namely, if X and Y are nonnegative matrices and $X \succcurlyeq Y$, where $\succcurlyeq$ denotes entry-by-entry inequality, then $\lambda(X) \ge \lambda(Y)$.

Since MTR $(2, 2)$ is finite-type, it suffices by Corollary 27 to use only G' to compute f_S. However, by Theorem 44, $f_S^{(p)}(\rho)$ achieves $f_S(\rho)$, so it suffices to show that $f_S^{(p)}(\rho)$ is upper-bounded by the straight line.

From G' in Fig. 13, let A denote the part of $A_{G'}$ corresponding to the edges with binary labels and B denote the part of $A_{G'}$ corresponding to the edges with box labels, i.e.,

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$
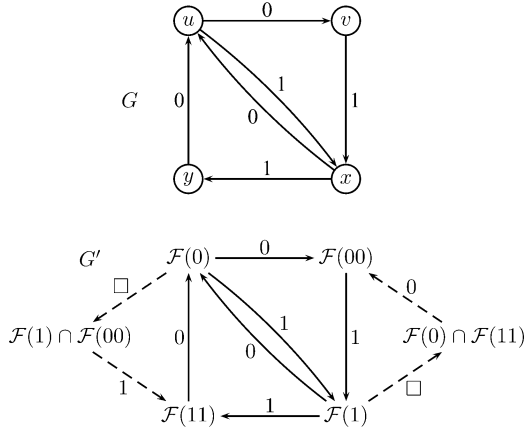
Fig. 13. The Shannon cover G and the only nontrivial component G' (of $\hat{G}$) for MTR $(2,2)$. The subgraph of G' with solid edges is H, which is isomorphic to G.

and

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

By inspecting G' in Fig. 13, it is clear that an unconstrained position with label $\square$ is always followed by two constrained positions. This can only happen in the follower set transitions:

$$\mathcal{F}(1) \xrightarrow{\square} \mathcal{F}(0) \cap \mathcal{F}(11) \xrightarrow{0} \mathcal{F}(00) \xrightarrow{1} \mathcal{F}(1)$$
$$\mathcal{F}(0) \xrightarrow{\square} \mathcal{F}(1) \cap \mathcal{F}(00) \xrightarrow{1} \mathcal{F}(11) \xrightarrow{0} \mathcal{F}(0).$$

Given a length N, a rational insertion rate $\rho \in [0, 1/3]$, and a specification of unconstrained positions $U \subseteq \{0, \ldots, N-1\}$ with $|U|/N = \rho$, define the matrix representing the periodic (U, N) configuration by

$$M = M_0 M_1 \cdots M_{N-1}$$

where $M_i = B$ if $i \in U$ and $M_i = A$ if $i \notin U$, with the indices taken $\mathrm{mod}\ N$. Each entry of M represents the number of fillings in $\hat{S}$ satisfying U, with length exactly N, that begin with a restricted set of prefixes and suffixes.

The restrictions in the follower set transitions deduced above imply that any $M_0 M_1 \cdots M_{N-1}$ can be written in terms of A and another matrix $C = BA^2$, which when evaluated has only two nonzero entries (each of value 1), in the $(1,1)$ and $(3,3)$ entries. Since the nonzero entries are situated along the main diagonal, $C^n = C$ for any $n \geq 1$. It is also straightforward to verify the matrix inequality $AC \preceq A$.

Thus, for any N and $M = A^{n_1} C^{m_1} A^{n_2} C^{m_2} \ldots A^{n_l} C^{m_k}$, where n_i, m_i, l, k are nonnegative integers that satisfy
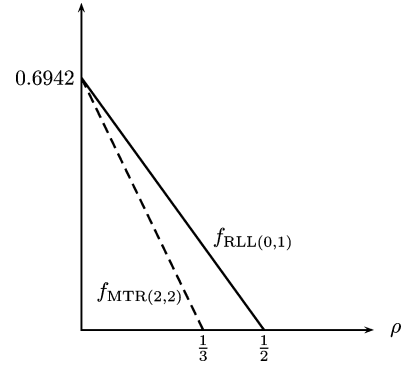
$\sum_{i=1}^{l} n_i + 3 \sum_{i=1}^{k} m_i = N$, we can use $AC \preceq A$ to upper-bound M repeatedly

$$M = A^{n_1} C^{m_1} A^{n_2} C^{m_2} \ldots A^{n_l} C^{m_k}$$
$$\preceq A^{n_1+n_2} C^{m_2} \ldots A^{n_l} C^{m_k}$$
$$\preceq A^{\sum_{i=1}^{l} n_i} C^{m_k}$$
$$= A^{\sum_{i=1}^{l} n_i} C^{\sum_{i=1}^{k} m_i}$$

where the last inequality uses the relation $C^n = C$ for any $n \geq 1$.

The expression on the right corresponds to another periodic configuration sharing the same length N but with a different set U' of unconstrained positions, and higher code rate (due to the eigenvalue inequality). However, the new (U', N) configuration is one that uses a weighted average of the 0-bit stuffing and the one-in-three bit stuffing schemes, which have code rates 0.6942 and 0, respectively. This in turn implies that the tradeoff function f_S must be a straight line connecting those two points. ∎

The tradeoff functions for RLL $(0,1)$ and MTR $(2,2)$ are given in Fig. 14. We remark that the tradeoff function of a constraint is not necessarily a straight line. An example is the tradeoff function of MTR (2), which was proven in [1] to equal the bit-stuffing lower bound.

Fig. 14. Tradeoff functions for RLL $(0,1)$ and MTR $(2,2)$.

X. Summary

We summarize the main definitions and results of this paper below.

For a binary constraint S, we let $\hat{S}$ be the set of words w over $\{0, 1, \square\}$ such that if we replace each occurrence of $\square$ in w independently by 0 or 1, we obtain a word in S. In Section III, a graph construction $\hat{G}$ is given, and $\hat{S}$ is shown to be the constrained system presented by $\hat{G}$. The graph $\hat{G}$ contains all information on the permissible sets of unconstrained positions for the constraint S. The graph $\hat{G}$ is deterministic and reduced, but not necessarily irreducible (even if S is irreducible). We show that there is a unique subgraph H of $\hat{G}$ that is isomorphic to the Shannon cover G of S. For finite-type constraints, $\hat{G}$ has only one nontrivial irreducible component G'. For irreducible constraints, a special topological feature of $\hat{G}$ exists (which we refer to as the "alias" property): Given an irreducible component G of $\hat{G}$, there are irreducible components G_1, G_2 of $\hat{G}$ such that G_1 can reach H and H can reach G_2; moreover, the constrained system S is contained within that of G_1 and G_2, respectively.

Let q be an integer. For $0 \le \rho \le 1$, $\mathcal{I}(\rho)$ refers to the set of all sequences (I_q) such that $I_q \subseteq \{1, 2, \ldots, q\}$ and $\lim_{q \to \infty} |I_q|/q = \rho$. In Section IV, we give a general definition of the tradeoff function f

$$f(\rho) = f_S(\rho) = \sup_{(I_q) \in \mathcal{I}(\rho)} \limsup_{q \to \infty} \frac{\log M(q, I_q)}{q}$$

where $M(q, I)$ is the number of words $w \in \hat{S}$ such that $|w| = q$ and $w_i = \square$ if and only if $i \in I$. The maximum insertion rate is given as

$$\mu = \mu(S) = \max_{f(\rho) \ge 0} \rho.$$

Other definitions of the tradeoff function are also examined, namely, the periodic tradeoff function $f^{(p)}$, and the definition of code rate $f^{(o)}$ that was given in [1]. Both $f^{(p)}$ and $f^{(o)}$ are defined only on the rationals $0 \le \rho \le 1$. We show in Section VII that $f^{(p)}(\rho) \le f^{(o)}(\rho) \le f(\rho)$, $f^{(o)}(\rho) = f(\rho)$ on the rationals $0 \le \rho \le \mu$, and for finite-type constraints, $f(\rho) = f^{(p)}(\rho)$.

More generally, for a graph G over $\{0, 1, \square\}$, we denote its tradeoff function by f_G, and its maximum insertion rate by μ_G.

In Section V, we show that the maximum insertion rate μ of a constraint is rational, and can be computed directly from its $\hat{G}$ presentation by finding a max-insertion-rate cycle that yields the largest ratio of box-labeled edges to cycle length. For any constraint, the maximum insertion rate can be computed numerically using Karp's algorithm on each irreducible component of $\hat{G}$ (if more than one) and then taking the maximum. We also present explicit closed-form expressions for the maximum insertion rates of the RLL (d, k) and MTR (j, k) constraints. The maximum insertion rate corresponds to the highest redundancy (or lowest rate) ECC that is compatible with the constraint in this scheme of unconstrained positions. The fact that it is easily computable for any constraint S enables it to be used as a first check for code compatibility when attempting to combine any ECC with S for this scheme.

In Section VI, properties of the tradeoff function are presented. We show that f is decreasing and left-continuous for any constraint. For an irreducible, finite-type constraint S, its tradeoff function f_S is shown to equal $f_{G'}$, where G' is the single nontrivial irreducible component in $\hat{G}$. An irreducible constraint S may have more than one irreducible component in its $\hat{G}$ presentation. We show that timesharing between two components G_1, G_2, where one can be reached from the other, can be achieved by proportionate concatenation of long sequences from one component with those obtained from the other. Using timesharing and the alias property of $\hat{G}$ for S, we establish that f_S is concave and hence continuous on the domain $[0, \mu_S]$. Furthermore, f_S can be determined from the nontrivial irreducible components that can be reached from H.

However, we remark that it is still an open problem to determine the tradeoff function exactly from an irreducible component. The existence of such an algorithm would not only yield the tradeoff function exactly for any finite-type constraint, but would also yield the tradeoff function for any irreducible constraint. A practical method for the latter would be to apply the algorithm to each component G_i of $\hat{G}$ to determine each tradeoff function f_{G_i}, and then take the smallest concave function dominating each f_{G_i}.

APPENDIX
PROOFS OF LEMMAS 31 AND 35

A. *Proof of Lemma 31*

For each I_q, define J_q as follows. Let Γ be a set of paths of length q with labels that satisfy I_q and $|\Gamma| = M(q, I_q)$. There exist states u and v in V_G such that the number of paths in Γ with initial state u and terminal state v is at least $M(q, I_q)/|V_G|^2$. Since G is irreducible, there is a path of length at most $|V_G|$ from v to u. Let z be the label of this path. Let z' be a word of length q that satisfies I_q. Let w be the right-infinite word $(z'z)^\infty$. Define $J_q = \{i \in \mathbb{N} : w_i = \square\}$. Moreover, define $T_q = |z'z|$. The set J_q is periodic with period T_q.

Let $(q(n))$ be a sequence such that

$$\lim_{n \to \infty} \frac{\log M(q(n), I_{q(n)})}{q(n)} = \limsup_{q \to \infty} \frac{\log M(q, I_q)}{q}.$$

For each $p \in \mathbb{N}$, we define I'_p as follows. If there is an integer n such that $T_{q(n)} \le \sqrt{p}$, define $I'_p = J_{q(m(p))} \cap \{1, \ldots, p\}$, where $m(p)$ is the largest such integer. (Note that $\sqrt{p}$ can be replaced by any function $\alpha(p)$ such that $\alpha(p) \to \infty$ and $\alpha(p)/p \to 0$ as $p \to \infty$.) Let p^* be the smallest integer p such that $m(p)$ exists. If $p < p^*$, define $I'_p = \emptyset$.

Observe the following.
- $m(p)$ is unbounded, i.e., $m(p) \to \infty$ as $p \to \infty$.
- Let words z and z' be defined as above. Since $q(m(p)) = |z'|$ and $T_{q(m(p))} = |z'z|$

$$T_{q(m(p))} - |V_G| \le q(m(p)) \le T_{q(m(p))} \le \sqrt{p}.$$

  Since $m(p)$ is unbounded, $q(m(p))$ and $T_{q(m(p))}$ are unbounded.

If $p \ge p^*$, I'_p contains a periodic pattern with period $T_{q(m(p))}$. In each period, the number of $\square$ is at least $|I_{q(m(p))}|$ and at most $|I_{q(m(p))}| + |V_G|$. It follows that

$$|I'_p| \ge |I_{q(m(p))}| \left(\frac{p}{T_{q(m(p))}} - 1 \right)$$

$$\frac{|I'_p|}{p} \ge \frac{|I_{q(m(p))}|}{T_{q(m(p))}} - \frac{|I_{q(m(p))}|}{p}$$

$$\ge \frac{|I_{q(m(p))}|}{T_{q(m(p))}} - \frac{q(m(p))}{p}$$

$$\ge \frac{|I_{q(m(p))}|}{q(m(p))} \times \frac{q(m(p))}{T_{q(m(p))}} - \frac{1}{\sqrt{p}}$$

$$\liminf_{p \to \infty} \frac{|I'_p|}{p} \ge \rho \times 1 - 0 = \rho.$$

Moreover

$$|I'_p| \le \left(|I_{q(m(p))}| + |V_G| \right) \left(\frac{p}{T_{q(m(p))}} + 1 \right)$$

$$\frac{|I'_p|}{p} \le \frac{|I_{q(m(p))}| + |V_G|}{T_{q(m(p))}} + \frac{|I_{q(m(p))}| + |V_G|}{p}$$

$$\le \frac{|I_{q(m(p))}|}{T_{q(m(p))}} + \frac{|V_G|}{T_{q(m(p))}} + \frac{1}{\sqrt{p}} + \frac{|V_G|}{p}$$

$$\limsup_{p \to \infty} \frac{|I'_p|}{p} \le \rho + 0 + 0 + 0 = \rho.$$

Therefore, $(I'_p) \in \mathcal{I}(\rho)$.

If $p \geq p^*$, the number of words $M(p, I_p')$ satisfies

$$M(p, I_p') \geq \left(\frac{M(q(m(p)), I_{q(m(p))})}{|V_G|^2} \right)^{\left(\frac{p}{T_{q(m(p))}} - 1 \right)}$$

$$\frac{\log M(p, I_p')}{p} \geq \left(\frac{1}{T_{q(m(p))}} - \frac{1}{p} \right)$$
$$\times \left(\log M(q(m(p)), I_{q(m(p))}) - 2 \log |V_G| \right)$$
$$\geq \frac{\log M(q(m(p)), I_{q(m(p))})}{q(m(p))} \times \frac{q(m(p))}{T_{q(m(p))}}$$
$$- \frac{q(m(p))}{p} - \left(\frac{2}{T_{q(m(p))}} - \frac{2}{p} \right) \log |V_G|.$$

Hence,

$$\liminf_{p \to \infty} \frac{\log M(p, I_p')}{p} \geq \lim_{p \to \infty} \frac{\log M(q(m(p)), I_{q(m(p))})}{q(m(p))} \times 1$$
$$- 0 - 0$$
$$= \limsup_{q \to \infty} \frac{\log M(q, I_q)}{q}.$$

Furthermore

$$M(p, I_p') \leq \left(M(q(m(p)), I_{q(m(p))}) 2^{|V_G|} \right)^{\left(\frac{p}{T_{q(m(p))}} + 1 \right)}$$

$$\frac{\log M(p, I_p')}{p} \leq \left(\frac{1}{T_{q(m(p))}} + \frac{1}{p} \right)$$
$$\times \left(\log M(q(m(p)), I_{q(m(p))}) + |V_G| \right)$$
$$\leq \frac{\log M(q(m(p)), I_{q(m(p))})}{T_{q(m(p))}} + \frac{q(m(p))}{p}$$
$$+ \frac{|V_G|}{T_{q(m(p))}} + \frac{|V_G|}{p}.$$

Hence,

$$\limsup_{p \to \infty} \frac{\log M(p, I_p')}{p} \leq \lim_{p \to \infty} \frac{\log M(q(m(p)), I_{q(m(p))})}{T_{q(m(p))}}$$
$$+ 0 + 0 + 0$$
$$= \limsup_{q \to \infty} \frac{\log M(q, I_q)}{q}. \qquad \blacksquare$$

B. Proof of Lemma 35

For simplicity, we write f_j for f_{G_j}. We will show that for any ρ and $(I_q) \in \mathcal{I}(\rho)$, there are $G_{j_1}, \ldots, G_{j_d}$ such that

$$\limsup_{q \to \infty} \frac{\log M(q, I_q)}{q} \leq g(f_{j_1}, \ldots, f_{j_d})(\rho).$$

Then the lemma will follow since

$$g(f_{j_1}, \ldots, f_{j_d})(\rho) \leq g(f_1, \ldots, f_n)(\rho).$$

If

$$\limsup_{q \to \infty} (\log M(q, I_q))/q = -\infty$$

for all $(I_q) \in \mathcal{I}(\rho)$, then

$$f(\rho) = -\infty \leq \max_{(i,j) \in P} g(f_i, f_j)(\rho).$$

Therefore, we can consider only the sequences (I_q) such that

$$0 \leq \limsup_{q \to \infty} (\log M(q, I_q))/q \leq 1.$$

It follows that there exists a subsequence $(q(m))$ such that

$$\lim_{m \to \infty} \frac{\log M(q(m), I_{q(m)})}{q(m)} = \limsup_{q \to \infty} \frac{\log M(q, I_q)}{q}.$$

Define $W(q)$ to be the set of words of length q in $\hat{S}$ that satisfy I_q. If $w \in W(q)$, let $\gamma(w)$ be the path in $\hat{G}$ from state $\mathcal{F}(\epsilon)$ with label w. We can partition $W(q)$ into classes where w and w' are in the same class if the orders of irreducible components that $\gamma(w)$ and $\gamma(w')$ traverse are the same. There are at most $(n+1)!$ classes, where n is the number of irreducible components in $\hat{G}$. At a given $q(m)$, let $C_{q(m)}$ be a class with the most number of words. Let $W(q(m), C_{q(m)})$ be the set of words of length $q(m)$ in class $C_{q(m)}$. Then

$$|W(q(m), C_{q(m)})| \geq \frac{M(q(m), I_{q(m)})}{(n+1)!}.$$

Since the number of classes is finite, there is a subsequence $(q(m,l))$ of $(q(m))$ such that $C_{q(m,1)} = C_{q(m,2)} = \cdots$; we denote this class by C. Now for any $q(m,l)$, $W(q(m,l), C)$ can be partitioned into subclasses, where w and w' are in the same subclass if whenever $(\gamma(w))_i$ is a transitional edge, $(\gamma(w))_i = (\gamma(w'))_i$. Letting E be the set of transitional edges, we can represent the subclass of a word w by

$$\{(i, e) \in \{1, \ldots, q(m,l)\} \times E : (\gamma(w))_i = e\}.$$

If $N = |E|$ is the number of transitional edges, then there are at most $(qN)^N$ subclasses. Hence, there is a subclass $D_{q(m,l)}$ with the number of words at least

$$\frac{M(q(m,l), I_{q(m,l)})}{(n+1)!(q(m,l)N)^N}. \qquad (19)$$

Observe that the class C specifies the order of the irreducible components, say $G_1, G_2, \ldots, G_c$, that all paths $\gamma(w)$ traverse, where $w \in W(q(m,l), C)$. For each $q(m,l)$, the subclass $D_{q(m,l)}$ induces c sets of unconstrained positions: $J_1(q(m,l)), \ldots, J_c(q(m,l))$, which is defined as follows. Let $D_{q(m,l)} = \{(i_1, e_1), \ldots, (i_{c-1}, e_{c-1})\}$. Then set $J_1(q(m,l)) = I_{q(m,l)} \setminus \{i_1, \ldots, q(m,l)\}$. For $2 \leq j \leq c-1$, set

$$J_j(q(m,l)) = \{i - i_{j-1} : i \in I_{q(m,l)}, \ i_{j-1} + 1 \leq i \leq i_j - 1\}.$$

Finally, set

$$J_c(q(m,l)) = \{i - i_{c-1} : i \in I_{q(m,l)}, \ i \geq i_{c-1} + 1\}.$$

Moreover, $D_{q(m,l)}$ induces the length of the subpath of $\gamma(w)$ that stay in each irreducible component. Set $b_1(q(m,l)) = i_1 - 1$. For $2 \leq j \leq c-1$, set $b_j(q(m,l)) = i_j - i_{j-1} - 1$. Finally, set $b_c(q(m,l)) = q(m,l) - i_{c-1}$.

We claim that there is a subsequence $(q(m,l,k))$ of $(q(m,l))$ such that

- $\lim_{k \to \infty} \frac{|J_j(q(m,l,k))|}{q(m,l,k)} = \rho_j'$ for some ρ_j', where $\sum_{j=1}^{c} \rho_j' = \rho$;
- $\lim_{k \to \infty} \frac{b_j(q(m,l,k))}{q(m,l,k)} = \theta_j$ for some θ_j, where $\sum_{j=1}^{c} \theta_j = 1$;
- $b_j(q(m,l,k))$ is nondecreasing with k for all $1 \leq j \leq c$;
- if $\theta_j > 0$, then $b_j(q(m,l,k)) > 0$ for all k.

To prove the claim, observe that

$$|I_{q(m,l)}| - N \leq \sum_{j=1}^{c} |J_j(q(m,l))| \leq |I_{q(m,l)}|$$

where N is the number of transitional edges of $\hat{G}$. Since $(I_q) \in \mathcal{I}(\rho)$

$$\lim_{l \to \infty} \frac{\sum_{j=1}^c |J_j(q(m,l))|}{q(m,l)} = \rho. \tag{20}$$

Next, observe that $\sum_{j=1}^c b_j(q(m,l)) = q(m,l) - c + 1$. Hence,

$$\lim_{l \to \infty} \frac{\sum_{j=1}^c b_j(q(m,l))}{q(m,l)} = 1. \tag{21}$$

Since $|J_j(q(m,l))| \geq 0$ for all j, we can find a subsequence of $(q(m,l))$ such that the corresponding subsequence of $(|J_1(q(m,l))|/q(m,l))$ converges. Similarly, we can find a subsequence of $(b_1(q(m,l))/q(m,l))$ that converges. Moreover, since $b_j(q(m,l))$ are nonnegative integers, we can find a subsequence of $(b_1(q(m,l)))$ that is nondecreasing; if the limit of the subsequence of $(b_1(q(m,l))/q(m,l))$ is strictly positive, then we can also choose a subsequence that is strictly positive and nondecreasing. It is clear that any subsequences of $(J_j(q(m,l)))$ and $(b_j(q(m,l)))$ still satisfy (20) and (21) respectively. Thus, we can take further and further subsequences (as many as $3c$ times) to obtain $(q(m,l,k))$ that satisfies the claim.

Let K be the set of indices j such that $\theta_j > 0$. We write $K = \{j_1, \ldots, j_d\}$. If $j \in K$, we define $\rho_j = \rho'_j/\theta_j$. It is clear that

$$\lim_{k \to \infty} \frac{|J_j(q(m,l,k))|}{b_j(q(m,l,k))}$$
$$= \lim_{k \to \infty} \left(\frac{|J_j(q(m,l,k))|}{q(m,l,k)} \times \frac{q(m,l,k)}{b_j(q(m,l,k))} \right) = \frac{\rho'_j}{\theta_j} = \rho_j.$$

Note that if $\theta_j = 0$, then $\rho'_j = 0$ since $|J_j(q(m,l,k))| \leq b_j(q(m,l,k))$. Therefore,

$$\sum_{j=1}^c \rho'_j = \sum_{j \in K} \rho'_j = \sum_{j \in K} \theta_j \rho_j = \rho.$$

If $1 \leq j \leq c$ and $I \subseteq \mathbb{N}$, we define $M_j(q, I)$ to be the number of words of length q in $S(G_j)$ that satisfy I. Then the number of words in the subclass $D_{q(m,l,k)}$ is at most $\prod_{j=1}^c M_j(b_j(q(m,l,k)), J_j(q(m,l,k)))$. From (19)

$$M(q(m,l,k), I_{q(m,l,k)}) \leq (n+1)!(q(m,l,k)N)^N$$
$$\times \prod_{j=1}^c M_j(b_j(q(m,l,k)), J_j(q(m,l,k))).$$

Hence,

$$\limsup_{q \to \infty} \frac{\log M(q, I_q)}{q}$$
$$= \lim_{k \to \infty} \frac{\log M(q(m,l,k), I_{q(m,l,k)})}{q(m,l,k)}$$
$$\leq \limsup_{k \to \infty} \frac{\log(n+1)!(q(m,l,k)N)^N}{q(m,l,k)}$$
$$+ \sum_{j=1}^c \limsup_{k \to \infty} \frac{\log M_j(q(m,l,k), J_j(q(m,l,k)))}{q(m,l,k)}$$

$$= 0 + \sum_{j=1, j \notin K}^c \limsup_{k \to \infty} \frac{\log M_j(q(m,l,k), J_j(q(m,l,k)))}{q(m,l,k)}$$
$$+ \sum_{j \in K} \limsup_{k \to \infty} \frac{\log M_j(q(m,l,k), J_j(q(m,l,k)))}{q(m,l,k)}$$
$$\leq \sum_{j=1, j \notin K}^c \limsup_{k \to \infty} \frac{b_j(q(m,l,k))}{q(m,l,k)}$$
$$+ \sum_{j \in K} \limsup_{k \to \infty} \left(\frac{b_j(q(m,l,k))}{q(m,l,k)} \right.$$
$$\left. \times \frac{\log M_j(q(m,l,k), J_j(q(m,l,k)))}{b_j(q(m,l,k))} \right)$$
$$\leq 0 + \sum_{j \in K} \theta_j f_j(\rho_j)$$
$$\leq g(f_{j_1}, \ldots, f_{j_d})(\rho). \qquad \blacksquare$$

Acknowledgment

References

[1] J. C. de Souza, B. H. Marcus, R. New, and B. A. Wilson, "Constrained systems with unconstrained positions," *IEEE Trans. Inf. Theory*, vol. 48, no. 4, pp. 866–879, Apr. 2002.

[2] P. Chaichanavong and B. H. Marcus, "Optimal block-type-decodable encoders for constrained systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1231–1250, May 2003.

[3] R. D. Cideciyan, E. Eleftheriou, B. H. Marcus, and D. S. Modha, "Maximum transition run codes for generalized partial response channels," *IEEE J. Select. Areas Commun.*, vol. 19, no. 4, pp. 619–634, Apr. 2001.

[4] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed. Cambridge, MA: McGraw-Hill/MIT Press, 2000.

[5] J. L. Fan and A. R. Calderbank, "A modified concatenated coding scheme, with applications to magnetic storage," *IEEE Trans. Inf. Theory*, vol. 44, no. 4, pp. 1565–1574, Jul. 1998.

[6] J. Fan and J. Cioffi, "Constrained coding techniques for soft iterative decoders," in *Proc. IEEE GLOBECOM*, Rio de Janeiro, Brazil, 1999, pp. 723–727.

[7] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[8] K. A. S. Immink, "A practical method for approaching the channel capacity of constrained channels," *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1389–1399, Sep. 1997.

[9] R. M. Karp, "A characterization of the minimum cycle mean in a digraph," *Discr. Math*, vol. 23, pp. 309–311, 1978.

[10] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding*. Cambridge, U.K.: Cambridge Univ. Press, 1995.

[11] B. H. Marcus, R. M. Roth, and P. H. Siegel, *Handbook of Coding Theory*. Amsterdam, The Netherlands: Elsevier, 1998, ch. 20.

[12] J. Moon and B. Brickner, "Maximum transition run codes for data storage systems," *IEEE Trans. Magn.*, vol. 32, no. 5, pp. 3992–3994, Sep. 1996.

[13] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.

[14] A. J. van Wijngaarden and K. A. S. Immink, "Maximum runlength-limited codes with error control capabilities," *IEEE J. Select. Areas Commun.*, vol. 19, no. 4, pp. 602–611, Apr. 2001.