**Abstract**—Many of the classic problem of coding theory are highly symmetric, which makes it easy to derive sphere-packing upper bounds and sphere-covering lower bounds on the size of codes. We discuss the generalizations of sphere-packing and sphere-covering bounds to arbitrary error models. These generalizations become especially important when the sizes of the error spheres are nonuniform. The best possible sphere-packing and sphere-covering bounds are solutions to linear programs. We derive a series of bounds from approximations to packing and covering problems and study the relationships and trade-offs between them. We compare sphere-covering lower bounds with other graph theoretic lower bounds such as Turán’s theorem. We show how to obtain upper bounds by optimizing across a family of channels that admit the same codes. We present a generalization of the local degree bound of Kulkarni and Kiyavash and use it to improve the best known upper bounds on the sizes of single deletion correcting codes and single grain error correcting codes.

I. INTRODUCTION

The classic problem of coding theory, correcting substitution errors in a vector of $q$-ary symbols, is highly symmetric. First, if $s$ errors are required to change a vector $x$ into another vector $y$, then $s$ errors are also required to change $y$ into $x$. Second, the number of vectors that can be produced from $x$ by making up to $s$ substitutions, the size of the sphere around $x$, does not depend on $x$. The sizes of these spheres play a crucial role in both upper and lower bounds on the size of the largest $s$-substitution-error-correcting codes. The Hamming bound is a sphere-packing upper bound and the Gilbert-Varshamov lower bound is a sphere-covering lower bound. The two symmetries that we have described makes the proofs of the Hamming and Gilbert-Varshamov bounds extremely simple.

Many other interesting error models do not have this degree of symmetry. Substitution errors with a restricted set of allowed substitutions are sometimes of interest. The simplest example are the binary asymmetric errors, which can replace a one with a zero but cannot replace a zero with a one. Binary asymmetric errors have neither of the two symmetries that we have described. Erasure and deletion errors differ from substitution errors in a more fundamental way: the error operation takes an input from one set and produces an output from another.

In this paper, we will discuss the generalizations of sphere-packing and sphere-covering bounds to arbitrary error models. These generalizations become especially important when the sizes of the error spheres are nonuniform. Sphere-packing and sphere-covering bounds are fundamentally related to linear programming and the best possible versions of the bounds are solutions to linear programs. In highly symmetric cases, including many classical error models, it is often possible to get the best possible sphere-packing bound without directly considering any linear programs. For less symmetric channels, the linear programming perspective becomes essential.

In fact, recently a new bound, explicitly derived via linear programming, was applied by Kulkarni and Kiyavash to find an upper bound on the size of deletion-correcting codes [2]. It was subsequently applied to grain errors [3], [4] and multipermutation errors [5]. We will refer to this as the local degree bound. The local degree bound constructs a dual feasible point for the sphere-packing linear program because computation of the exact solution is intractable. Deletion errors, like most interesting error models, act on an exponentially large input space. Because computation of the best possible packing and covering bounds is often intractable, simplified bounds such as the local degree bound are useful.

Sphere-packing and sphere-covering arguments have been applied in an ad hoc fashion throughout the coding theory literature. This work aims at presenting a unifying framework that permits such arguments in their most general form applicable to both uniform and nonuniform error sphere sizes. More precisely, we derive a series of bounds from approximations to packing and covering problems. The local degree bound of [2] is one of the bounds in the series. We characterize each bound as the solution to a linear program and study the relationships between them. We use the concept of a combinatorial channel to represent an error model in a fashion that makes the connection to linear programming natural. Each approximation technique yields a sphere-packing upper bound and a sphere-covering lower bound. These bounds use varying levels of information about structure of the error model and consequently make trade-offs between performance and complexity. For example, one bound uses the distribution of the sizes of spheres in the space while another uses only the size of the smallest sphere.
Sphere-packing upper bounds and sphere-covering lower bounds are not completely symmetric. We explore the relationship between sphere-covering lower bounds and other graph theoretic lower bounds such as Turán’s theorem. In general, there are many different combinatorial channels that admit the same codes. However, each channel gives a different sphere-packing upper bound. We show that the Hamming bound, which can be derived from a substitution error channel, the Singleton bound, which can be derived from an erasure channel, and a family of intermediate bounds provide an example of this phenomenon.

Finally, we present a generalization of the local degree bound and use it to improve the best known upper bounds on the sizes of single deletion correcting codes and single grain error correcting codes.

In Section II we discuss the linear programs associated with sphere-packing bounds. In Section III we discuss various techniques for obtaining nonuniform sphere-packing bounds. These include the degree sequence, degree threshold, and local degree bounds. In Section IV we discuss families of channels that have the same codes but give different sphere-packing bounds. In Section V we discuss alternate lower bounds and their relationship with the sphere-covering lower bound. In Section VI we discuss bounds that use only the average size of spheres. In Section VII we present a generalization of the local degree bound that is related to an iterative procedure. We use this to improve the best known upper bounds on the sizes of single deletion correcting code and single grain error correcting codes.

II. SPHERE-PACKING BOUNDS, SPHERE-COVERING BOUNDS, AND LINEAR PROGRAMS

A. Notation

Let $X$ and $Y$ be a finite or countable sets. Let $R^X$ denote the set of $|X|$-dimensional column vectors of elements of $R$ indexed by $X$. Let $R^{X \times Y}$ denote the set of $|X|$ by $|Y|$ matrices of elements of $R$ with the rows indexed by $X$ and the columns indexed by $Y$. Let $2^X$ denote the power set of $X$. Let $\mathbb{N}$ denote the set of nonnegative integers and let $[n]$ denote the set of nonnegative integers less than $n$: $\{0, 1, \ldots, n-1\}$.

B. Combinatorial channels

We use the concept of a combinatorial channel to formalize a set of possible errors.

**Definition 1.** A combinatorial channel is a matrix $A \in \{0, 1\}^{X \times Y}$, where $X$ is the set of channel inputs and $Y$ is the set of channel outputs. An output $y$ can be produced from an input $x$ by the channel if $A_{x,y} = 1$. Each row or column of $A$ must contain at least one one, so each input can produce some output and each output can be produced from some input.

We will often think of a channel as a bipartite graph. In this case, the left vertex set is $X$, the right vertex set is $Y$, and $A$ is the bipartite adjacency matrix. We will refer to this bipartite graph as the channel graph. For $x \in X$, let $N_A(x) \subseteq Y$ be the neighborhood of $x$ in the channel graph (the set of outputs that can be produced from $x$). The degree of $x$ is $|N_A(x)|$.

For $y \in Y$, let $N_A(y) \subseteq X$ be the neighborhood of $y$ in the channel graph (the set of inputs that can produce $y$). In most cases, the channel involved will be evident and we will drop the subscript on $N$.

Let $1$ be the column vector of all ones. For a set $S \subseteq X$, let $1_S$ be the indicator column vector for the set $S$. Note that $A1_{\{y\}} = 1_{N(y)}$ and $1^T_s A = 1^T_{N(s)}$. Thus $A1$ is the vector of input degrees of the channel graph, $A^T1$ is the vector of output degrees, and $1^T A1$ is the number of edges.

We are interested in the problem of transmitting information through a combinatorial channel with no possibility of error. To do this, the transmitter only uses a subset of the possible channel inputs in such a way that the receiver can always determine which input was transmitted.

**Definition 2.** A code for a combinatorial channel $A \in \{0, 1\}^{X \times Y}$ is a set $C \subseteq X$ such that for all $y \in Y$, $|N(y) \cap C| \leq 1$.

This condition ensures that decoding is always possible: if $y$ is received, the transmitted symbol must have been the unique element of $N(y) \cap C$.

C. Sphere-packing

We would like to find the largest code. If $C$ is a maximum code for $A \in \{0, 1\}^{X \times Y}$, then the simplest sphere-packing upper bound on the size of a code is

$$|C| \leq \frac{|Y|}{\min_{x \in X} |N_A(x)|}.$$ 

A code is a packing of the neighborhoods of the inputs into the output space. The neighborhoods of the codewords must be disjoint and each neighborhood contains at least $\min_{x \in X} |N(x)|$ outputs.

Maximum set packing is naturally expressed as an integer linear program. Each channel output provides a constraint.

Traditionally, set packing problems have been described in the language of hypergraphs. A hypergraph consists of a vertex set and a family of hyperedges. Each hyperedge is a nonempty subset of the vertices. A hypergraph $(V, E)$ can be described by a vertex-hyperedge incidence matrix $A \in \{0, 1\}^{V \times E}$. If $A$ is the incidence matrix of a hypergraph, then $A^T$ is the incidence matrix of its dual. We will identify hypergraphs with their incidence matrices. Thus any channel can be considered as a hypergraph. Throughout this paper, we use the language of channels and bipartite channel graphs rather than that of hypergraphs. The only exceptions are the following two definitions, which are standard in the hypergraph literature [2].

**Definition 3.** Let $A \in \{0, 1\}^{X \times Y}$ be a hypergraph. A matching in $A$ is pairwise-nonintersecting subset of the hyperedges, $S \subseteq Y$. Each vertex is contained in at most one hyperedge from $S$. The maximum size of a matching in $A$ is denoted by $\nu(A)$.

A vertex packing in a hypergraph is a subset of the vertices, $S \subseteq X$, such that each hyperedge contains at most one vertex from $S$. The maximum size of a vertex packing in $A$ is denoted by $p(A)$.
A matching in $A^T$ corresponds to a vertex packing in $A$, so
$p(A) \triangleq \nu(A^T) \triangleq \max\{1^T w : w \in \{0,1\}^{|X|}, A^T w \leq 1\}.$

Consequently $p(A)$ the size of the largest input packing, or code, for a channel $A$.

We will also define the linear programming duals of these problems. At the end of Section II-E we will see an application of them.

**Definition 4.** Let $A \in \{0,1\}^{X \times Y}$ be a hypergraph. A transversal in $A$ is a subset of the vertices, $S \subseteq X$, such that every hyperedge contains at least one vertex from $S$. The minimum size of a transversal in $A$ is denoted by $\tau(A)$.

A hyperedge covering is a subset of the hyperedges, $S \subseteq Y$, such that each vertex is included in at least one hyperedge from $S$. The minimum size of a hyperedge covering in $A$ is denoted by $\kappa(A)$.

A transversal in $A^T$ corresponds to a hyperedge covering in $A$, so
$\kappa(A) \triangleq \tau(A^T) \triangleq \min\{1^T z : z \in \{0,1\}^{|Y|}, Az \geq 1\}$

For a channel $A$, $\kappa(A)$ is the size of the smallest output covering.

**D. Confusion graphs and independent sets**

**Definition 5.** Let $A \in \{0,1\}^{X \times Y}$ and $B \in \{0,1\}^{Y \times Z}$ be channels. Then define $A \circ B \in \{0,1\}^{X \times Z}$, the composition of $A$ and $B$, such that
$N_{A \circ B}(x) = \bigcup_{y \in N_A(x)} N_B(y)$

We can characterize $A \circ B$ in two other ways:

\[
(A \circ B)_{x,z} = \begin{cases} 1 & N_A(x) \cap N_B(z) \neq \emptyset \\ 0 & N_A(x) \cap N_B(z) = \emptyset \end{cases}
\]

\[
\min_{y \in Y} \max_{x \in X} (A \circ B)_{x,y} = \min_{y \in Y} \max_{z \in Z} (A \circ B)_{y,z} \tag{2}
\]

The second characterization states that $A \circ B$ is the matrix product of $A$ and $B$ in the Boolean semiring.

Let $I$ denote the identity matrix.

**Definition 6.** For a channel $A \in \{0,1\}^{X \times Y}$, define the confusion graph of $A$ to be the graph with vertex set $X$ and adjacency matrix $A \circ A^T - I$.

Because $A \circ A^T - I$ is a zero-one symmetric matrix with zeros on the diagonal, the confusion graph is simple and undirected. From (1), vertices $u$ and $v$ are adjacent in the confusion graph of $A$ if and only if $N(u)$ and $N(v)$ intersect. Figure 1 shows an example of a channel, its composition with its reverse, and its confusion graph.

**Definition 7.** Let $G$ be an undirected simple graph with vertex set $X$. A set $S \subseteq X$ is independent in $G$ if and only if for all $u,v \in S$, $u$ and $v$ are not adjacent. The maximum size of an independent set in $G$ is denoted by $\alpha(G)$.

Now we have a second important characterization of codes.

**Lemma 1.** Let $G$ be the confusion graph for a channel $A \in \{0,1\}^{X \times Y}$. Then a set $C \subseteq X$ is code for a $A$ if and only if it is an independent set in $G$. Thus $\alpha(G) = p(A)$.

**Proof:** A set $C$ is not a code if and only if there is some $y$ such that $N(y)$ contains distinct codewords $u$ and $v$, or equivalently $y \in N(u) \cap N(v)$. This means $(A \circ A^T)_{u,v} = 1$, $u$ and $v$ are adjacent in the confusion graph, and $C$ is not independent.

**E. Sphere-covering and dominating sets**

Let $G$ be the confusion graph for a channel $A \in \{0,1\}^{X \times Y}$ and let $d_G(x)$ be the degree of a vertex $x$ in $G$. If $C$ is a maximum code for $A$, then the basic most basic sphere-covering lower bound is
$|C| \geq \frac{|X|}{1 + \max_{x \in X} d_G(x)}$.

Because $C$ is maximal, each vertex in $X$ is either in $C$ or adjacent in $G$ to a vertex in $C$. Each codeword prevents at most $1 + \max_{x \in X} d_G(x)$ vertices from being added to the code.

Now we will show the relationship between this argument and the sphere-packing argument.

**Definition 8.** Let $G$ be an undirected simple graph with vertex set $X$. A set $S \subseteq X$ is dominating in $G$ if and only if for all $x \in X \setminus S$, there is some $u \in S$, such that $x$ and $U$ are adjacent. The minimum size of a dominating set in $G$ is denoted by $\gamma(G)$.

**Lemma 2.** For any graph $G$, $\gamma(G) \leq \alpha(G)$.

**Proof:** If no additional vertices can be added to an independent set, each vertex of $G$ is either in the independent set
or adjacent to a vertex in the independent set. Consequently, any maximal independent set is dominating. ■

Domingating set is a covering problem. A vertex \( u \in S \) covers itself and all adjacent vertices.

**Lemma 3.** Let \( G \) be a simple graph with vertex set \( X \) and adjacency matrix \( B - I \). We can consider \( B \) to be a channel. Then \( S \subseteq X \) is a dominating set in \( G \) if and only if \( S \) is an output covering for \( B \). Thus \( \gamma(G) = \kappa(B) \).

**Proof:** This follows immediately from the definitions. ■

**Definition 9.** Let \( G \) be a simple graph with vertex set \( X \). Then let \( G^k \), the \( k \)th distance power of \( G \), be another simple graph with vertex set \( X \). Distinct vertices are adjacent in \( G^k \) if they are connected by a path of length at most \( k \) in \( G \).

**Lemma 4.** Let \( G \) be a simple graph with vertex set \( X \) and adjacency matrix \( B - I \). Then \( C \subseteq X \) is an independent set in \( G^2 \) if and only if \( C \) is an input packing for \( B \). Thus \( \alpha(G^2) = p(B) \).

**Proof:** The confusion graph of the channel \( B \) is \( G^2 \). ■

### F. Fractional relaxations

Let \( C \) be a maximum code for a channel \( A \) and let \( G \) be the confusion graph of \( A \). Together, Lemma 1 Lemma 2 and Lemma 3 establish that

\[
\kappa(A \circ A^T) = \gamma(G) \leq |C| = \alpha(A) = p(A).
\]

However, the maximum independent set and minimum dominating set problems over general graphs are NP-Hard [6]. The approximate versions of these problems are also hard. The maximum independent set of an \( n \)-vertex graph cannot be approximated within a factor of \( n^{1-\varepsilon} \) for any \( \varepsilon \) unless \( \text{P} = \text{NP} \) [7]. We seek efficiently computable bounds. These bounds cannot be good for all graphs, but they will perform reasonably well for many of the graphs that we are interested in.

The relaxed problem, maximum fractional set packing, provides an upper bound on the original packing problem.

**Definition 10.** Let \( A \in \{0,1\}^{X \times Y} \) be a channel. The size of the maximum fractional input packing in \( A \) is

\[
p^*(A) \triangleq \max \{1^T w : w \in \mathbb{R}^X, \ 0 \leq w \leq 1, \ A^T w \leq 1 \}.
\]

The size of the minimum fractional output covering is

\[
\kappa^*(A) \triangleq \min \{1^T z : z \in \mathbb{R}^Y, \ 0 \leq z \leq 1, \ Az \geq 1 \}.
\]

The fractional programs have larger feasible spaces, so \( p(A) \leq p^*(A) \) and \( \kappa^*(A) \leq \kappa(A) \). By strong linear programming duality, \( p^*(A) = \kappa^*(A) \).

Recall that for each \( x \in X \), there is some \( y \in N(x) \) (\( N(x) \) is nonempty). Then the constraint \( \sum_{u \in N(y)} w_u \leq 1 \) appears in the program for \( p(A) \), so the constraint \( w_x \leq 1 \) is redundant. For each \( y \in Y \), \( N(y) \) is nonempty, so the vector \( z = 1 \) is feasible in the program for \( \kappa(A) \) and the constraint \( z \leq 1 \) is redundant. Dropping the redundant constraints gives

\[
p^*(A) = \max \{1^T w : w \in \mathbb{R}^X, \ 0 \leq w \leq 1, \ A^T w \leq 1 \}
\]

\[
\kappa^*(A) = \min \{1^T z : z \in \mathbb{R}^Y, \ 0 \leq z \leq 1, \ Az \geq 1 \}.
\]

Now our bounds on the maximum code \( C \) are

\[
\kappa^*(A \circ A^T) \leq \kappa(A \circ A^T) \leq |C| = p(A) \leq p^*(A).
\]

Unlike the integer programs, the values of the fractional linear programs can be computed in polynomial time. However, we are usually in sequences of channels with exponentially large input and output spaces. In these cases, finding exact solutions to the linear programs is intractable but we would still like to know as much as possible about the behavior of the solutions. We now discuss some simpler bounds that have been useful in practice.

### III. Four bounds for fractional packing and covering

In this section we consider four simple pairs of upper and lower bounds on the maximum fractional set packing number, or equivalently the minimum fractional set cover number. Each of these bounds is the value of some simplified linear program. The four upper bounds are derived either by relaxing the constraints of the primal maximization program or by tightening the constraints of the dual minimization program.

#### A. Minimum and maximum degree bounds

**Definition 11.** For a channel \( A \in \{0,1\}^{X \times Y} \), define the minimum degree upper bounds

\[
p_{\text{Min-D}}^*(A) \triangleq \max \{1^T w : w \in \mathbb{R}^X, \ w \geq 0, \ A^T w \leq |Y| \},
\]

and the maximum degree lower bounds

\[
p_{\text{Max-D}}^*(A) \triangleq \min \{1^T z : z \geq 0, \ 1^T Az \leq |X| \},
\]

The two upper bounds are equal:

\[
p_{\text{Min-D}}^*(A) = \kappa_{\text{Min-D}}^*(A) = \frac{|Y|}{\min_{x \in X} |N(x)|},
\]

and the two lower bounds are equal:

\[
p_{\text{Max-D}}^*(A) = \kappa_{\text{Max-D}}^*(A) = \frac{|X|}{\max_{y \in Y} |N(y)|}.
\]

In the next section, we will see that the programs for \( p_{\text{Min-D}}^*(A) \) and \( \kappa_{\text{Max-D}}^*(A) \) are closely related to the degree sequence bounds. The programs for \( p_{\text{Min-D}}^*(A) \) and \( \kappa_{\text{Max-D}}^*(A) \) are related to the local degree bounds.

The linear program for \( p^*(A) \) contains a constraint for each \( y \in Y \) : \( 1_n(y)w \leq 1 \). In the linear program for \( p_{\text{Min-D}}^* \) we have replaced these constraints with their sum, \( \sum_{x \in X} |N(x)|w_x \leq |X| \). Thus the feasible space has been strictly increased. This optimal \( x \) in the new program for \( p_{\text{Min-D}}^* \) assigns all weight to the input with the smallest degree.

By mechanically taking the dual of the program for \( p_{\text{Min-D}}^* \), we obtain

\[
\min \{|Y|t : t \geq 0, \ A1t \geq 1 \},
\]

which is easily rearranged into the program for \( \kappa_{\text{Min-D}}^* \). The program for \( \kappa_{\text{Min-D}}^* \) is a restriction of the program for \( \kappa^* \): the same weight must be assigned to each output.
Observe that the trivial sphere-packing bound from Section II-C is \( p_{\min D}^* (A) \) and that the trivial sphere-covering bound from Section II-E is \( \kappa_{\max D}^* (A \circ A^T) \).

### B. The degree sequence and degree threshold bounds

If the minimum degree is far from the average degree, \( p_{\min D}^* \) is likely to be a bad approximation of \( p^* \). A better bound comes from considering all of the input degrees.

**Definition 12.** For a channel \( A \in \{0, 1\}^{X \times Y} \), define the degree sequence bounds

\[
p_{DS}^* (A) \triangleq \max \{ |T w : w \in \mathbb{R}^X, \; 0 \leq w \leq 1, \; 1^T A T w \leq |Y| \}
\]

\[
\kappa_{DS}^* (A) \triangleq \min \{ |T z : z \in \mathbb{R}^Y, \; 0 \leq z \leq 1, \; 1^T A z \geq |X| \}.
\]

Recall that \( A1 \) is the vector of input degrees of the channel graph of \( A \). Note that the program for \( p_{DS}^* (A) \) is the programs for \( p_{\min D}^* (A) \) with the constraint \( w \leq 1 \) added. Consequently, \( p^*(A) \leq p_{DS}^* (A) \leq p_{\min D}^*(A) \). Similarly, \( \kappa_{\max D}^* (A) \leq \kappa_{DS}^*(A) \).

**Lemma 5.** For a channel \( A \in \{0, 1\}^{X \times Y} \) and a degree threshold \( t \in \mathbb{R} \), let

\[
X_0 = \{ x \in X : |N(x)| = t \},
\]

and let

\[
c_0 = \sum_{x \in X_0} |N(x)| = 1^T X_0 A1.
\]

\[
c_\infty = |X_\infty| = \sum_{x \in X_\infty} |N(x)| = 1^T X_\infty A1
\]

If \( c_\infty \leq |Y| \leq c_\infty + c_0 \) then

\[
p_{DS}^* (A) = |X_\infty| + \frac{|Y| - c_\infty}{t}.
\]

**Proof:** To construct a feasible point for \( p_{DS}^* (A) \), we put full weight on all of the inputs with degree below the threshold and fractional weight on inputs with degree equal to the threshold: the point \( w = 1 \chi_\infty + \frac{1-c_\infty}{|X_0|} 1 \chi_0 \) is feasible and has value \( |X_\infty| + \frac{1-c_\infty}{t} \).

The dual program is

\[
\min \{ |Y| z_0 + 1^T z : (z_0, z) \in \mathbb{R}^1 + |X|, \; A1 z_0 + z \geq 1 \}
\]

The point \( z_0 = \frac{|Y|}{t}, \; z_x = \max(0, 1 - \frac{|N(x)|}{t}) \) is feasible in the dual program. Note that \( z_x > 0 \) only for \( x \in X_\infty \). The value of this point is

\[
\frac{|Y|}{t} + \sum_{x \in X_\infty} \left( 1 - \frac{|N(x)|}{t} \right) = |X_\infty| + \frac{|Y| - c_\infty}{t}.
\]

For a given input degree distribution and output space size, there is some channel where the neighborhoods of the small degree inputs are disjoint. For this channel, the degree sequence upper bound is tight. Analogous tightness examples exist for the lower bounds. Thus the degree sequence bounds cannot be improved with incorporating more information about the structure of the channel.

The next bound, the degree threshold bound, is simpler to compute than degree sequence bound, but is often almost as good.

**Definition 13.** For a channel \( A \in \{0, 1\}^{X \times Y} \), define the degree threshold bounds to be

\[
p_{DT}^* (A) \triangleq \min_{t \in \mathbb{N}} p_{DT}^* (A, t)
\]

\[
\kappa_{DT}^* (A) \triangleq \max_{t \in \mathbb{N}} \kappa_{DT}^* (A, t)
\]

where

\[
p_{DT}^* (A, t) \triangleq \max \{ |T w : w \in \mathbb{R}^X, \; 0 \leq w \leq 1, \; c^T w \leq 1 \}
\]

\[
\kappa_{DT}^* (A, t) \triangleq \min \{ |T z : z \in \mathbb{R}^Y, \; 0 \leq z \leq 1, \; d^T z \geq 1 \}.
\]

and where \( c \in \mathbb{R}^X \) and \( d \in \mathbb{R}^Y \) such that

\[
c_x = \begin{cases} t & |N(x)| \geq t, \\
\min_{u \in X} |N(u)| & |N(x)| < t. \end{cases}
\]

\[
d_y = \begin{cases} t & |N(y)| \geq t, \\
\max_{v \in Y} |N(v)| & |N(y)| < t. \end{cases}
\]

These are equivalent to applying the degree sequence bound to a modified degree sequence. From Lemma 5, \( p_{DT}^* (A, t) \) equals

\[
|S| + \frac{|Y| - |S| d_{\min}}{t} = \frac{|Y|}{t} + |S| \left( 1 - \frac{d_{\min}}{t} \right) \leq \frac{|Y|}{t} + |S|,
\]

where \( d_{\min} = \min_{x \in X} |N(x)| \) and \( S = \{ x \in X : |N(x)| < t \} \), the members of \( X \) with small degree. If we let \( t = d_{\min} \), then \( S \) is empty and the bound reduces to the minimum degree bound.

Similarly, \( \kappa_{DT}^* (A, t) \) equals

\[
|R| + \frac{|X| - |R| d'_{\max}}{t} = \frac{|Y|}{t} + |R| \left( \frac{d'_{\max}}{t} - 1 \right).
\]

where \( d'_{\max} = \max_{y \in Y} |N(y)| \) and \( R = \{ y \in Y : |N(y)| > t \} \), the members of \( Y \) with large degree. To eliminate the dependence on \( d'_{\max} \), we can replace it with \( |X| \).

The degree threshold bounds are relatively easy to apply. Levenshtein applied both the upper and lower degree threshold bounds to the deletion channel [8]. Cullina and Kiyavash applied the upper bound to channels performing both deletions and insertions [11]. Mazumdar et al. applied the degree threshold bound to the grain error channel [9].

### C. The local degree bound

**Definition 14.** Let \( A \in \{0, 1\}^{X \times Y} \) be a channel and let \( E \) be the edge set of the channel graph for \( A \). Define the local degree bound

\[
\kappa_{LD}^* (A) \triangleq \min \{ |T z : z \in \mathbb{R}^Y, \; z \geq 0, \; C(z) \geq 1 \},
\]

\[
p_{LD}^* (A) \triangleq \max \{ |T w : w \in \mathbb{R}^X, \; w \geq 0, \; D w \leq 1 \},
\]

where \( C \in \mathbb{R}^{E \times Y} \), \( D \in \mathbb{R}^{E \times Y} \), and

\[
C_{(x,y),w} = \begin{cases} |N(x)| & y = w, \\
0 & y \neq w, \end{cases}
\]

\[
D_{(x,y),w} = \begin{cases} |N(y)| & x = u, \\
0 & x \neq u. \end{cases}
\]
To create the program for $\kappa^*_{LD}$, we have replaced each constraint $1_{N(x)}z \geq 1$ with $|N(x)|$ constraints: for each $y \in N(x)$, we require $1_{\{y\}}z \geq \frac{1}{|N(y)|}$. The old constraint is the sum of the new constraints, so the new constraints are more restrictive. This results in the program in the above definition.

Now each $z_y$ is subject to a constraint for each $x \in N(y)$. These can be combined as $z_y \min_{x \in N(y)} |N(x)| \geq 1$ or $z_y \geq 1/\min_{x \in N(y)} |N(x)|$. Thus, the optimal assignment is

$$z_y = \frac{1}{\min_{x \in N(y)} |N(x)|},$$

$$\kappa^*_{LD}(A) = \sum_{y \in Y} \frac{1}{\min_{x \in N(y)} |N(x)|}.$$  

Similarly, the optimal assignment for $p^*_{LD}(A)$ is

$$w_x = \max_{y \in N(x)} |N(y)|,$$

$$p^*_{LD}(A) = \sum_{x \in X} \frac{1}{\max_{y \in N(x)} |N(y)|}.$$  

Because we created the program for $\kappa^*_{LD}$ by restricting the program for $\kappa^*$, $\kappa^*_{LD} \geq \kappa^*$. We can also show that the local degree bounds are always at least as good as the degree sequence bounds.

**Lemma 6.** For a channel $A$, $\kappa^*_{LD}(A) \leq p^*_{DS}(A)$ and $\kappa^*_{DS}(A) \leq p^*_{LD}(A)$.

**Proof:** We will only prove $\kappa^*_{LD}(A) \leq p^*_{DS}(A)$. The proof for the lower bounds is completely analogous. We construct a point $x$ that is feasible in the primal linear program for $p^*_{DS}$ with value $\kappa^*_{LD}$.

Let $E$ be the edge set of the channel graph for $A$. The dual program for $\kappa^*_{LD}(A)$ is

$$\max \{ 1^T z : z \in \mathbb{R}^E, z \geq 0, C^T z \leq 1 \}$$

$$C_{(x,y),u} = \begin{cases} |N(x)| & y = w \\ 0 & y \neq w \end{cases}$$

We can map the parameter space for this program into the parameter space of the program for $p^*_{DS}$ in a weight preserving way: let $w_x = \sum_{y \in N(x)} z_{(x,y)}$. Now we just need to show that this map sends feasible points in the dual program for $\kappa^*_{LD}$ to feasible points in the program for $p^*_{DS}$.

In the dual program for $\kappa^*_{LD}$, $z_{(u,y)}$ is part of one constraint:

$$\sum_{x \in N(y)} |N(x)|z_{(x,y)} \leq 1.$$  

If we sum all the constraints and apply the mapping, we get

$$\sum_{y \in Y} \sum_{x \in N(y)} |N(x)|z_{(x,y)} \leq |Y|$$

$$\sum_{x \in X} |N(x)| \sum_{y \in N(x)} z_{(x,y)} \leq |Y|$$

$$\sum_{x \in X} |N(x)|w_x \leq |Y|$$

which is exactly the global constraint in the program for $p^*_{DS}$.

If we sum only the constraints involving $x$, we get

$$\sum_{y \in N(x)} |N(t)|z_{(u,y)} \leq |N(x)|$$

$$\sum_{y \in N(x)} |N(x)|z_{(x,y)} + \sum_{u \in N(y) \setminus x} |N(u)|z_{(u,y)} \leq |N(x)|$$

Thus $w_x \leq 1$, which is the local constraint on $w_x$ in the program for $p^*_{DS}$.

**Theorem 1.** For a channel $A \in \{0,1\}^{X \times Y}$,

$$p^*_{MaxD}(A) = \kappa^*_{MaxD}(A) \leq \kappa^*_{DT}(A) \leq \kappa^*_{DS}(A) \leq p^*_{LD}(A) = \kappa^*(A) \leq \kappa^*_{LD}(A)$$

$$p^*_{DS}(A) \leq p^*_{DT}(A) \leq p^*_{MinD}(A) = \kappa^*_{MinD}(A)$$

**Proof:** The program for $p^*$ is a maximization and $p^*_{DS}$, $p^*_{DT}$, and $p^*_{MinD}$ form a sequence of relaxations of that program, so $p^* = p^*_{DS} \leq p^*_{DT} \leq p^*_{MinD}$. By Lemma 6, $\kappa^*_{LD} \leq p^*_{DS}$. The program for $\kappa^*$ is a minimization and the program for $\kappa^*_{LD}$ is a restriction of it, so $\kappa^* \leq \kappa^*_{LD}$. The sequence of lower bounds on $p^*$ is analogous.

**D. Symmetric channel graphs**

**Lemma 7.** Let $A \in \{0,1\}^{X \times Y}$ be a channel and let $e = 1^TA1$, the number of edges in the channel graph. If $A1 = d1$, then $A$ is input regular and

$$\kappa_{LD} = p^*_{DS} = p^*_{DT} = p^*_{MinD} = \frac{|Y|}{d} = \frac{|X||Y|}{e}.$$  

If $A1 = d1^T$, then $A$ is output regular and

$$p^*_{LD} = \kappa^*_{DS} = \kappa^*_{DT} = \kappa^*_{MaxD} = \frac{|X|}{d^*} = \frac{|X||Y|}{e}.$$  

If $A$ is both input and output regular, then

$$p^* = \kappa^* = \frac{|X||Y|}{e}.$$  

**Proof:** This follows immediately from the definitions of the bounds.

If the input degrees are all equal to $d$ but the output degrees vary, the four upper bounds are $\kappa^*$ equal each other but are not necessarily equal to $\kappa^*$ itself. The length-one binary erasure channel,

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

demonstrates this. The erasure output covers both inputs, so $p(A) = \kappa^*(A) = \kappa(A) = 1$. Both inputs have degree 2, so all four of the upper bounds equal 3/2.
E. Example: Single binary asymmetric error channel

Consider the single-asymmetric-error channel. The input and output of this channel are binary vectors of length $n$. The channel acts separately on each entry of the vector. A zero input produces a zero output, but a one input can produce either a one or a zero (an error).

Each input with $k$ ones can produce $k + 1$ outputs. The all zero input has degree one, so $p_{\text{MinD}}^*(A) = 2^n$. There are $\sum_{i=0}^{k-1} \binom{n}{i}$ inputs with degree strictly less than $k + 1$. Thus

$$p^*_{\text{DT}}(A) = \min_{k} \sum_{i=0}^{k-1} \binom{n}{i} + \frac{2^n}{k + 1}.$$  

Each output $y$ with $k$ ones is adjacent to the input with $k$ ones and (for $\kappa$) either a one or a zero (an error).

To verify that this is a good bound on $p^*$, we compute the value of the local degree lower bound on $p^*$. Let $j = n - k$. For $k \geq 1$, each input $x$ with $k$ ones is adjacent to an output with $k - 1$ ones. That output has $n - k + 1 = j + 1$ zeros, so it has degree $j + 2$. The output with zero ones is adjacent only to the output with zero ones, which has degree $n + 1$. Thus the value of the local degree lower bound is

$$\frac{1}{n + 1} + \sum_{j=0}^{n-1} \binom{n}{j} \frac{1}{j + 2}.$$  

In this example, the input degrees are concentrated around the average degree so the degree threshold bound performs reasonably well. There is little variation in input degree within the neighborhood of a single output, so the local degree bound performs well.

F. Example: single q-ary asymmetric error channel

Now we give an example where the bounds do not perform as well. Consider the channel with input and output sets $[q] = \{1, 2, \ldots, q\}$. For each input $i$, let the possible outputs be all $j \leq i$.

For this channel, $\kappa^*$, $\kappa^*_{LD}$, $p^*_{DS}$, $p^*_{DT}$, and $p^*_{\text{MinD}}$ are all distinct. The output one can be produced by any input, so $\kappa(A) = p(A) = 1$. The input one has degree one, so $\kappa^*_{\text{MinD}}(A) = q$. If we choose $d$ as the degree threshold, then $\kappa^*_{\text{DT}}(A, d) = d + q/d$. The best choice is $\sqrt{q}$, so $\kappa^*_{\text{DT}}(A, d) = 2\sqrt{q}$. The sum of the smallest $k$ degrees is $(k + 1)/2$, so $\kappa^*_{\text{DS}}(A)$ is the largest $k$ such that $(k + 1)/2 \leq q$. This is approximately $\sqrt{2q}$. Finally, each output $j$ can be produced from each input $i \geq j$ and input $i$ has degree $j$. Thus $y_j = \min_{i=j} \frac{1}{i}$ and $\kappa^*_{\text{LD}} = \sum_{j=1}^{\sqrt{2q}} j$, which is approximately $\log q$.

In this example, the average input degree is $\frac{2q}{\sqrt{q}}$, so we might hope to get an upper bound on $\kappa^*$ of about 2. However, the input degrees are not concentrated around the average, so none of our four approximations are particularly good.

IV. FAMILIES OF CHANNELS WITH THE SAME CODES

In Section II-D we defined the confusion graph for a channel and established that a code is an independent set in the confusion graph. The confusion graph does not contain enough information to recover the original channel graph, but it contains enough information to determine whether a set is a code for the original channel.

A clique in a graph $G$ is a set of vertices $S$ such that for all distinct $u, v \in S$, $\{u, v\} \in E(G)$. If $G$ is the confusion graph for a channel $A \in \{0, 1\}^{X \times Y}$, then for each $y \in Y$, $N(y)$ is a clique in $G$.

There are many different channels that have $G$ as a confusion graph. Let $\Omega \subseteq 2^X$ be a family of cliques that covers every edge in $G$. This means that for all $\{u, v\} \in E(G)$, there is some $S \in \Omega$ such that $u, v \in S$. Let $H \in \{0, 1\}^{X \times \Omega}$ be the vertex-clique incidence matrix: $H_{x,S} = 1$ is $x \in S$ and $H_{x,S} = 0$ otherwise. Then $\alpha(G) = p(H)$.

Thus each family of cliques that covers every edge gives us an integer linear program that expresses the maximum independent set problem for $G$. These programs all contain the same integer points, the indicators of the independent sets of $G$. However, their polytopes are significantly different so the fractional relaxations of these programs give widely varying upper bounds on $\alpha(G)$.

Each edge in $G$ is a clique, so $E(G)$ is one natural choice for $\Omega$. Then $\alpha(G) = p(H_E)$, where $H_E \in \{0, 1\}^{X \times E(G)}$ is the vertex edge incidence matrix for $G$. However, relaxing the integrality constraint for this program gives a useless upper bound. The vector $w = \frac{1}{2}1$ is feasible, so $p^*(H_E) \geq \frac{|X|}{2}$ regardless of the structure of $G$.

Lemma 8. Let $G$ be a graph with vertex set $X$ and let $\Omega_1, \Omega_2 \subseteq 2^X$ be families of cliques that cover every edge in $G$. Let $H_1, H_2$ be the vertex-clique incidence matrices for $\Omega_1$ and $\Omega_2$ respectively. If for all $R \in \Omega_1$ there is some $S \in \Omega_2$ such that $R \subseteq S$, then $p^*(H_2) \leq p^*(H_1)$.

Proof: A clique $S$ gives the constraint $\sum_{x \in R} w_x \leq 1$ in $p$. If $R \in \Omega_1, S \in \Omega_2$, and $R \subseteq S$, then the constraint from $R$ is implied by the constraint for $S$. Any additional cliques in $\Omega_2$ can only reduce the feasible space for $p(H_2)$. Thus the feasible space for $p(H_2)$ is contained in the feasible space for $p(H_1)$.

Definition 15. Let $\Omega$ be the set of maximal cliques in $G$ and let $H_\Omega \in \{0, 1\}^{X \times \Omega}$ be the vertex-clique incidence matrix. Then $\alpha(G) = p(H_\Omega)$. Define the minimum clique cover of $G$, $\theta(G) \triangleq \kappa(H_\Omega)$ and the minimum fraction clique cover $\theta^*(G) \triangleq \kappa^*(H_\Omega)$.
Unlike the program derived from the edge set, $\theta^*(G)$ gives a nontrivial upper bound on $\alpha(G)$. In fact, $\theta^*(G)$ is the best sphere packing bound for any channel that has $G$ as its confusion graph.

**Corollary 1.** Let $A \in \{0,1\}^{X \times Y}$ be a channel and let $G$ be the confusion graph for $A$. Then $\theta^*(G) \leq \kappa^*(A)$.

**Proof:** For each output $y \in Y$, $N(y)$ is a clique in $G$ and these clique cover every edge of $G$. Each clique in $G$ is contained in a maximal clique, so the claim follows immediately from Lemma 8. \[\square\]

The fractional clique cover number has been considered in the coding theory literature in connection with the Shannon capacity of a graph, $\Theta(G)$. The Shannon capacity of a graph is at least as large as the maximum independent set and is extremely difficult to compute. Shannon used something equivalent to a clique cover as an upper bound for Shannon capacity $\Theta(G)$; Rosenfeld showed the connection between Shannon’s bound and linear programming [11]. Lovasz introduced the Lovasz theta function of a graph, $\vartheta(G)$, and showed that it was always between the Shannon capacity and the fractional clique cover number $\frac{\alpha(G)}{\vartheta(G) \leq \vartheta(G) \leq \theta^*(G)}$. All together, we have $\alpha(G) \leq \Theta(G) \leq \vartheta(G) \leq \theta^*(G)$.

The Lovasz theta function is derived via semidefinite programming and consequently is not a sphere-packing bound.

Corollary 1 suggests that we should ignore the structure of our original channel $A$ and try to compute $\theta^*(G)$ instead of $\kappa^*(A)$. However, there is no guarantee that we can efficiently construct the linear program for $\theta^*(G)$ by starting with $G$ and searching for all of the maximal cliques. We are often interested in graphs with an exponential number of vertices. Even worse, the number of maximal cliques in $G$ can exponentially in the number of vertices. To demonstrate this, consider a complete $k$-partite graph with 2 vertices in each part. If we select one vertex from each part, we obtain a maximal (and also maximum) clique. The graph has 2k vertices, but there are $2^k$ maximal cliques.

**A. Obtaining a bound from a family of channels**

For a given graph, we cannot necessarily find the channel that gives the best possible sphere-packing bound. However, for some graphs, we can find a small family of relatively well-behaved channels. Each channel in the family give us some insight into the structure of the confusion graph. Now we have to decide how to use this information to get the best possible bound. In some cases, it is more effective to bound $\kappa(G)$ for each channel in a family rather than creating a single channel that expresses every known constraint.

Suppose that we have a family of channels $A_i \in \{0,1\}^{X \times Y_i}$, $i \in [k]$, that all have the same confusion graph $G$. Each channel in the family identifies some set of cliques in $G$, contributes some set of constraints on the independent set, and gives an upper bound $\kappa^*(A_i)$. The simplest way to combine these bounds is to take the minimum. Alternatively, we can define a new channel that includes all of the constraints: $A = [A_0|A_1|\ldots|A_{k-1}] \in \{0,1\}^{X \times Y}$ where $Y = \bigcup_{i \in [k]} Y_i$. Adding an additional constraint to maximization linear program can only reduce the value of the program, so $\alpha(G) \leq \kappa^*(A) \leq \min_{i \in [k]} \kappa^*(A_i)$.

However, none of the approximations in from Section III have this monotonicity property. This is demonstrated by the following example. Consider the channels $A = \left[ \begin{array}{ccc} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{array} \right]$, $A' = \left[ \begin{array}{ccc} 1 & 0 & 1 \\ 0 & 1 & 1 \end{array} \right]$.

The channel $A'$ contains a subset of the constraints of $A$, but $\kappa_{MinD}(A') = 1$ while $\kappa_{MinD}(A) = 3/2$. In practice, thus the best strategy is not to apply these approximations to the channel the includes every known constraint.

**Lemma 9.** Consider a family of channels $A_i \in \{0,1\}^{X \times Y_i}$ for $i \in [k]$. Let $A = [A_0|A_1|\ldots|A_{k-1}]$. If all $A_i$ are both input and output regular, then $\kappa^*(A) = \min_{i \in [k]} \kappa^*(A_i)$.

Furthermore, unless all $k$ original channels have the same output degree, $\min_{i \in [k]} \kappa^*_L(A_i) < \kappa^*_L(A)$.

**Proof:** Let $e_i = 1^T A_i 1$, the number of edges in $A_i$, and let $d'_i = \frac{\min_{i \in [k]} \kappa^*(A_i)}{e_i}$, the output degree of $A_i$. By Lemma 7, $\kappa^*(A_i) = \kappa^*_L(A_i) = \frac{|X||Y_i|}{e_i}$, so $\alpha(G) \leq \theta^*(G)$.

The new channel $A_i$ is input regular but is not output regular. If the $d'_i$ is not the same for all $i \in [k]$, then $\kappa^*_L(A) = \frac{|X|\sum_i |Y_i|}{\sum_i e_i} = \frac{|X|\min_i 1}{\sum_i d'_i} > \frac{|X|}{\sum_i d'_i}$ and $\kappa^*_L(A) = \kappa^*_L(A_i)$, proving the second claim. \[\square\]

Note that we did not need to assume that the channels have the same confusion graph. The optimal feasible point for $\kappa^*(A)$ assigns zero weight to unhelpful constraints, but all of our approximations attempt to use every constraint regardless of quality.

The technique of optimizing over a family of channels has been successfully applied to deletion-insertions channels by Cullina and Kiyavash [11]. Any code capable of correcting $s$ deletions can also correct any combination of $s$ total insertions and deletions. Two input strings can appear in an $s$-deletion-correcting code if and only if the deletion distance between them is more than $s$. In the asymptotic regime with $n$ going to infinity and $s$ fixed, each channel in the family becomes approximately regular. Thus the degree threshold bound gives a good approximation to the exact sphere-packing bound for these channels. The best bound comes from a channel that performs approximately $\frac{\min_i 1}{q}$ deletions and $\frac{1}{q+1}$ insertions, where $q$ is the alphabet size.

**B. Hamming and Singleton Bounds**

Consider the channel that takes an $q$-ary vector of length $n$ as its input, erases $a$ symbols, and substitutes up to $b$
symbols. Thus there are \( q^n \) channel inputs, \( \binom{n}{a} q^{n-a} \) outputs, and each input can produce \( \binom{n}{a} \sum_{i=0}^{b} \binom{n-a}{i} (q-1)^i \) possible outputs. Two inputs share a common output if and only if their Hamming distance is at most \( s = a + 2b \). For each choice of \( n \) and \( s \), we have a family of channels with identical confusion graphs. Call the \( q \)-ary \( n \)-symbol \( a \)-erasure \( b \)-substitution channel \( A_{q,n,a,b} \). These channels are all input and output regular, so by Lemma 7

\[
\kappa^*(A_{q,n,a,b}) = \frac{\binom{n}{a} q^{n-a}}{\sum_{i=0}^{n} \binom{n-a}{i} (q-1)^i}.
\]

Two special cases give familiar bounds. For even \( s \), setting \( a = 0 \) and \( b = s/2 \) produces the Hamming bound:

\[
\kappa^*(A_{q,n,0,s/2}) = \frac{q^n}{\sum_{i=0}^{s/2} \binom{n}{i} (q-1)^i}.
\]

Setting \( a = s \) and \( b = 0 \) produces the Singleton bound:

\[
\kappa^*(A_{q,n,s,0}) = q^{n-s}.
\]

For \( q = 2 \), the Hamming bound is always the best bound in this family. When \( q \) is at least 3, each bound in the family is the best for some region of the parameter space.

**Lemma 10.** \( \kappa^*(A_{q,n,a,b}) \leq \kappa^*(A_{q,n,a+2,b-1}) \) when \( a + qb \leq n - 1 \).

The proof of Lemma 10 can be found in Appendix A.

**Theorem 2.** Let \( q,n,s \in \mathbb{N} \) such that \( q \geq 3 \), \( 0 \leq s \leq n - 1 \), and \( s \) even. Then

\[
\arg \min_{0 \leq b \leq s/2} \kappa^*(A_{q,n,s-2b,b}) = \left\{ \begin{array}{l}
s/2 & s \leq \frac{2}{q}(n-1) \\
\frac{n-1-s}{q} & \frac{2}{q}(n-1) < s \leq \frac{n}{q}
\end{array} \right.
\]

For fixed \( \delta \frac{2}{q} \leq \delta \leq 1 \), and \( s = \delta n \)

\[
\lim_{n \to \infty} \frac{1}{n} \log \min_{0 \leq b \leq s/2} \kappa^*(A_{q,n,s-2b,b}) = (1 - \delta) \log(q-1).
\]

**Proof:** Let \( a + 2b = s \), so \( a + qb = s + (q-2)b \). From Lemma 10 \( \kappa^*(A_{q,n,0,s/2}) \) is the smallest in the family when \( s + (q-2)b \leq n - 1 \) or equivalently \( s \leq \frac{2}{q}(n-1) \).

For \( b \geq 1 \) the following are equivalent:

\[
\kappa^*(A_{q,n,a+2,b-1}) \geq \kappa^*(A_{q,n,a,b}) \leq \kappa^*(A_{q,n,a-2,b+1})
\]

\[
s + (q-2)b \leq n - 1 \leq s + (q-2)(b+1) \]

\[
b \leq \frac{n-1-s}{q-2} \leq b + 1.
\]

Let \( b^* \) be the optimal choice of \( b \). Then \( \lim_{n \to \infty} \frac{b^*}{n} = \frac{1-\delta}{q-2} \), \( \lim_{n \to \infty} \frac{a^* + 2b^*}{n} = 1 - \delta + \frac{2(1-\delta)}{q-2} = \frac{q(1-\delta)}{q-2} \), and

\[
\lim_{n \to \infty} \frac{1}{n} \log \kappa^* = 1 - \delta \frac{q}{q-2}.
\]
that for a graph $G$, 
\[ \alpha(G) \geq \alpha_{CW}(G) \triangleq \sum_{x \in X} \frac{1}{1 + d_G(x)}. \]
Call $\alpha_{CW}(G)$ the Caro-Wei number of $G$ [13].

Let $d(G) = \frac{1}{|X|} \sum_{x \in X} d_G(x)$. Then Turán’s theorem is 
\[ \alpha(G) \geq \alpha_T(G) = \frac{|X|}{1 + d(G)}. \]

Turán’s theorem is an immediate consequence of the Caro-Wei theorem. The function $f(x) = \frac{1}{1 + x}$ is convex, so by Jensen’s inequality
\[ |X| \sum_{x \in X} \frac{1}{1 + d(x)} \geq |X| \sum_{x \in X} \frac{1}{1 + \frac{1}{|X|} \sum_{x \in X} d(x)}. \]

A. Relationships with sphere-covering bounds

The trivial sphere-covering bound from Section II-E is 
\[ \alpha(G) \geq \frac{|X|}{1 + \max_{x \in X} d_G(x)}. \]

Turán’s theorem replaces this maximum with an average. As we mentioned at the end of Section III-A, the trivial bound on $\alpha(G)$ is equal to $\kappa_{MaxD}^*(B)$. Thus $p_{MaxD}^*(B) \leq \alpha_T(G)$. The bound from Turán’s theorem is also better than the degree sequence lower bound for $B$. The vector $z = \frac{1}{1 + d(B)} = \frac{|X|}{|T_B|} 1$ is feasible in the program for $\kappa_{DS}^*(B)$, so $\kappa_{DS}^*(B) \leq \alpha_T(G)$.

Interestingly, the Caro-Wei number of $G$ is always between the local degree lower and upper bounds on $\kappa^*(B)$. For any $x \in X$, 
\[ \min_{y \in N(x)} |N(y)| \leq |N(x)| \leq \max_{y \in N(x)} |N(y)| \]
so
\[ p_{LD}^*(B) = \sum_{x \in X} \frac{1}{\max_{y \in N(x)} |N(y)|} \]
\[ \leq \alpha_{CW}(G) = \sum_{x \in X} \frac{1}{|N(x)|} \]
\[ \leq \kappa_{LD}^*(B) = \sum_{x \in X} \frac{1}{\min_{y \in N(x)} |N(y)|} \]

There are graphs for which the Caro-Wei number is strictly larger than fractional sphere-covering bound and graphs for which it is strictly smaller. Consider the $n$-vertex path graph $P_n$. Note that $\alpha_{CW}(P_{3k}) = \frac{3k - 2}{4} + \frac{2}{3} = k + \frac{1}{3}$, but $\alpha(P_{3k}) = \gamma(P_{3k}) = k$. On the other hand, $\alpha_{CW}(P_{3k+1}) = \frac{3k - 1}{4} + \frac{2}{3} = k + \frac{4}{3}$, but $\alpha(P_{3k+1}) = \gamma(P_{3k+1}) = k + 1$. These examples and the strong graph product can be used to construct graphs with arbitrarily large gaps between the two bounds.

One final example is the star graph $K_{1,k}$. For this example we have $\alpha_{CW}(K_{1,k}) = \frac{k}{k + \frac{1}{k}}$ and $\alpha(K_{1,k}) = \gamma(K_{1,k}) = 1$. The inequalities among all of these lower bounds on $\alpha(G)$ are summarized in Figure 3.

\[ \alpha(G^2) = p(B) \quad \kappa(B) = \gamma(G) \]
\[ p_{\alpha}^*(B) = \kappa^*(B) \]
\[ \kappa_{DS}^*(B) \quad \alpha_{CW}(G) \quad \alpha_T(G) \]

Fig. 3. Lower bounds on $\alpha(G)$, where $B - I$ is the adjacency matrix of $G$. If $G$ is regular, then $p_{MaxD}^*(B) = \kappa_{LD}^*(B)$ and the seven efficiently computable bounds are all equal.

VI. BOUNDS THAT USE ONLY THE NUMBER OF EDGES

The bounds of Section III use progressively more information about the structure of the channel graph. The minimum and maximum degree bounds use a single extremal degree, the degree sequence bounds use the full degree distribution of one side of the channel graph, and the local degree bounds use the degrees of the endpoints of each edge. Suppose that we only know the number of inputs, output, and edges in the channel graph. This means that we know the average input degree and the average output degree but nothing else about the degree distributions.

In Section V we noted that the Caro-Wei theorem, which uses the full degree distribution of the confusion graph, implies Turán’s theorem, which uses only the average degree. We would like to do something similar with the degree sequence bounds.

Definition 16. Define the functions $f : \mathbb{R}^X \to \mathbb{R}$ and $g : \mathbb{R}^Y \to \mathbb{R}$
\[ f(a) \triangleq \max \{1^T w : w \in \mathbb{R}^X, 0 \leq w \leq 1, a^T w \leq 1\} \]
\[ g(a) \triangleq \min \{1^T z : z \in \mathbb{R}^Y, 0 \leq z \leq 1, a^T z \geq 1\}. \]

For a channel $A \in \{0,1\}^{X \times Y}$, the degree sequence bounds can be written in terms of these functions:
\[ p_{DS}^*(A) = f \left( \frac{1}{|Y|} A 1 \right) \]
\[ \kappa_{DS}^*(A) = g \left( \frac{1}{|X|} A^T 1 \right). \]

Lemma 11. Let $a \in \mathbb{R}^X$, and let $M \in \mathbb{R}^{X \times X}$ be a doubly stochastic matrix. Then $f(Ma) \leq f(a)$ and $g(Ma) \geq g(a)$.

Proof: Suppose $(z_0, z) \in \mathbb{R}^{[1]^+}$ is the optimal feasible point is the dual to the program for $f(a)$. This means that $a_{z_0} + z \geq 1$. Multiplying both sides of this inequality by $M$ gives $M a_{z_0} + M z \geq M 1 = 1$, so $(z_0, M z)$ is feasible for the dual to the program for $f(Ma)$. Thus $f(Ma) \leq z_0 + 1^T M z = z_0 + 1^T z = f(a)$. The inequality for $g$ follows analogously.

The inequality of Lemma 11 runs in the wrong direction, so we cannot use the degree sequence upper bound to derive an upper bound that only depends on the average degree.
It turns out that the number of edges in a bipartite graph gives us weak bounds on the packing and covering numbers for the graph.

**Lemma 12.** Let $A \in \{0, 1\}^{X \times Y}$ be a channel and let $E$ be the edge set of the channel graph. Then

$$|X| + |Y| - |E| \leq p(A).$$

For any $X$, $Y$, and $R \subseteq Y$ such that $|R| \leq |X|$, there is a channel $A$ such that $|E| = |X| + |Y| - |R|$ and $R$ is an output covering in $A$.

**Proof:** For each $y \in Y$, we select $|N(y)| - 1$ inputs to forbid from the code. We forbid at most $|E| - |Y|$ total inputs, so our code contains at least $|X| + |Y| - |E|$ inputs.

We construct $A$ as follows. Choose the neighborhoods of the outputs in $R$ so that each is nonempty, they are disjoint, and $\bigcup_{y \in D} N(y) = X$. Meeting the first two conditions is possible because $|R| \leq |X|$. Because the union of neighborhoods cover all of $X$, $R$ is a covering. We have included $|X|$ edges so far. For each $y \in Y \setminus R$, let $|N(y)| = 1$ and choose the neighbor arbitrarily. Thus $|E| = |X| + |Y| - |R|$.

**Lemma 13.** Let $A \in \{0, 1\}^{X \times Y}$ be a channel and let $E$ be the edge set of the channel graph. Then

$$\kappa(A) \leq |X| - \frac{|E|}{|Y|} + 1.$$

For any $X$, $Y$, and $S \subseteq X$ such that $|S| \leq |Y|$, there is a channel $A$ such that $|E| = |Y|(|X| - |S| + 1)$ and $S$ is an input packing in $A$.

**Proof:** For any output $y \in Y$, we can construct a cover using $y$ together with $|X| - |N(y)|$ other outputs; for each $x \in X \setminus N(y)$, we add an arbitrary member of $N(x)$ to our cover. Because $\sum_{y \in Y} |N(y)| = |E|$, there is some $y$ with $|N(y)| \geq \frac{|E|}{|Y|}$.

We construct $A$ as follows. Choose the neighborhoods of the inputs in $S$ so that each is nonempty, they are disjoint, and $\bigcup_{x \subseteq S} N(x) = Y$. Meeting the first two conditions is possible because $|S| \leq |Y|$. Because the neighborhoods are disjoint, $S$ is a packing. For each $x \in X \setminus S$, let $N(x) = Y$. Then all output degrees are equal to $|X| - |S| + 1$.

Only a few edges are needed to create a single output vertex with large degree and a large number of edges are necessary to rule out the existence of a set of input vertices with small degree.

**VII. ITERATIVE ALGORITHM**

One way to look at the local degree bound is as a distributed algorithm to find a fractional covering. Each input needs coverage totaling at least one and it requests an equal amount of coverage from each output. Each output receives a list of requests and must honor the largest.

More generally, we can view this as a single step in an iterative procedure. Suppose that we have a fractional covering. Then at each input $u$, the total coverage, $(Ay)_u$, is at least one. The input $u$ informs each output in $N(x)$ that it can reduce its value by a factor of $(Ay)_u$. Each output $y$ receives such a message for each input in $N(y)$, then makes the largest reduction consistent with the messages.

This iteration and an analogous iteration for fractional packings are formalized in the following lemma.

**Definition 17.** For $z \in \mathbb{R}^Y$ such that $Az > 0$, define

$$\varphi(z)_y \triangleq \frac{z_y}{\min_{x \in N(y)}(Az)_x}.$$

For $w \in \mathbb{R}^Y$ such that $ATw > 0$, define

$$\psi(w)_x \triangleq \frac{w_x}{\max_{y \in N(y)}(ATw)_y}.$$

**Lemma 14.** For $z \in \mathbb{R}^{|Y|}$ such that $z \geq 0$ and $Az > 0$, $\varphi(z)$ is feasible in the program for $\kappa^*(A)$. If $z$ is feasible for $\kappa^*(A)$, then $\varphi(z) \leq z$.

For $w \in \mathbb{R}^T$ such that $w \geq 0$ and $ATw > 0$, $\psi(w)$ is feasible in the program for $p^*(A)$. If $w$ is feasible for $p^*(A)$, then $\psi(w) \geq w$.

**Proof:** To demonstrate feasibility of $\varphi(z)$, we need $\varphi(z) \geq 0$ and $A\varphi(z) \geq 1$. The first condition is trivially met. For $x \in X$ and $y \in N(x)$, we have

$$\varphi(z)_y = \frac{z_y}{\min_{x \in N(y)}(Az)_x} \geq \frac{z_y}{(Az)_x} = \frac{z_y}{1 + (Az)_x},$$

so $1_{T_x} A\varphi(z) \geq 1$ and $\varphi(y)$ is feasible.

If $z$ is feasible, then $Az \geq 1$. For all $y \in Y$ we have

$$\varphi(z)_y = \frac{z_y}{\min_{x \in N(y)}(Az)_x} \leq z_y.$$

The claims about $\psi(w)$ follow analogously.

For both $\varphi$ and $\psi$, scaling the input by a positive constant does not affect the output: for $c \in \mathbb{R}$, $c > 0$, $\varphi(cz)$ and $\psi(cw) = \psi(w)$.

For any channel $A$, $1$ is a feasible vector in the program for $\kappa^*(A)$ and $\frac{1}{|X|}$ is a feasible vector in the program for $p^*(A)$. The optimum of the program for $\kappa^*_L(A)$ is $\varphi(1)$ and the optimum of the program for $p^*_L(A)$ is $\psi(1)$.

We can iterate this optimization step. An iteration fails to make progress under the following condition. From the definition $\varphi(z)_y = z_y$ if and only if $\min_{x \in N(y)}(Az)_x = 1$. Thus $\varphi(z) = z$ if for all $y \in Y$ there is some $x \in N(y)$ such that $(Az)_x = 1$. This algorithm is monotonic in each entry of the feasible vector, so it cannot make progress if its input is at the frontier of the feasible space.

**A. Application to the single deletion channel**

Let $A_n$ be the $n$-bit 1-deletion channel. The input to the binary single deletion channel is a string $x \in \{0, 1\}^n$ and the output is a substring of $x$, $y \in \{0, 1\}^{n-1}$. Each output vertex in $A_n$ has degree $n + 1$. Thus $p^*(A_n) \geq p^*_{\min D}(A_n) = \frac{2^n}{n+1}$.

Levenshtein [8] showed that

$$\kappa^*(A_n) \leq \frac{2^n}{n+1} \left(1 + o(1)\right).$$

Kulkarni and Kiyavash computed the local degree upper bound, or equivalently $\varphi(1)$ [2]. This shows that $\kappa^*(A_n)$ is at most

$$\frac{2^n}{n-1} = \frac{2^n}{n+1} \left(1 + \frac{2}{n-1}\right) = \frac{2^n}{n+1} \left(1 + O(n^{-1})\right).$$
Recently, Fazeli et al. found a fractional covering for $A_n$ that provides a better upper bound \cite{14}. In this section, we compute $\varphi \circ \varphi(1)$ for these channels and analyze the values of these points. We show that Fazeli’s improved covering is related to the covering $\varphi \circ \varphi(1)$, but $\varphi \circ \varphi(1)$ provides a better bound asymptotically.

More precisely, the upper bound from $\varphi \circ \varphi(1)$, given in Theorem 4, shows that asymptotically.

Theorem 3. For

$$\sum f(r_x, u_x, b_x) = 2f(n, n, 2) + 2f(1, 0, 0)$$

and

$$+2\sum_{r=2}^{n-1}u=0b=0 \sum_{r-u-1}(n-r-1)^2 r^2 f(r, u, b)$$

Analysis of the local degree bound relies on the following identity and inequality:

$$\sum_{r=1}^{n} \sum_{r+c-1}^{n} \sum_{r+c-1}^{n} < 2n+c-1$$

We will need this along with an analogue for unit runs:

Lemma 16. For $2 \le r \le n-1$,

$$\sum_{u=0}^{r} \sum_{r-u-1}^{n} \sum_{r-u-1}^{n} \sum_{r-u-1}^{n} \sum_{r-u-1}^{n} 2(r-1)^2 n - 1 r - 1$$

Thus we can nicely sum factors of $\frac{2u-b}{(r-1)^2}$. Now we will adjust $f$ until it is in a form that we can sum.

Lemma 17. Define $f'(r, u, b)$ to be equal to

$$\begin{cases}
\frac{1}{2} \left(1 - \frac{2u-b}{(r-1)^2} \sum_{r=2}^{n} \sum_{r=2}^{n} \sum_{r=2}^{n} \sum_{r=2}^{n} 1 \right) & u > 1 \\
1 & r = 1
\end{cases}$$

Then $z_y = f'(r_y, u_y, b_y)$ is feasible for $\kappa^*(A_n)$, so $\kappa^*(A_n) \le 1Tz$.

Theorem 4. For $n \ge 3$,

$$\kappa^*(A_n) \le \frac{2n}{n+1} \left(1 + \frac{30n+12}{n(n-1)(n-2)} \right)$$

Now we will compare this bound to the bound corresponding to the cover of Fazeli et al. Let

$$f''(r, u, b) \triangleq \begin{cases}
\frac{1}{r} \left(1 - \frac{u-b}{r^2} \right) & u-b \ge 2 \\
\frac{1}{r} & u-b \le 1
\end{cases}$$

Fazeli et al. establish that $z_y = f''(r_y, u_y, b_y)$ is feasible for $\kappa^*(A_n)$. This is easy to compare with the cover given by $f'$. Note that the coefficient on $u$ is 1 in $f''$ and 2 in $f'$.

Lemma 18. Let $z_y = f''(r_y, u_y, b_y)$. Then

$$1Tz \ge \frac{2n-2}{n+1} \left(1 + \frac{3}{n-1} - \frac{3}{(n-1)(n-2)} \right)$$

This shows that the bound of Theorem 4 is asymptotically better than the bound corresponding to the cover of Fazeli et al. We could continue to iterate $\varphi$ to produce even better bounds. The fractional covers produced would depend on more statistics of the strings. For example, the value at a particular output of the cover produced by the third iteration of $\varphi$ would depend on the number of runs of length two in that output string, in addition to the total number of runs and the number of runs of length one.

The largest known single deletion correcting codes are the Varshamov-Tenengolts (VT) codes. The length-$n$ VT code
contains at least \( \frac{2^n}{n+1} \) codewords, so these codes are asymptotically optimal. The VT codes are known to be maximum independent sets for \( n \leq 10 \), but this question is open for larger \( n \) [15]. Kulkarni and Kiyavash computed the exact value of \( \kappa^*(A_n) \) for \( n \leq 14 \) [2]. For \( 7 \leq n \leq 14 \), the gap between \( \kappa^*(A_n) \) and the size of the VT codes was at least one, so it is unlikely that sphere-packing bounds will resolve the optimality of the VT codes for larger \( n \). Despite this, it would be interesting to know whether \( \kappa^*(A_n) \leq \frac{2^n}{n+1} + O(2^n) \) for some constant \( c < 1 \).

**B. Application to the single grain error channel**

Recently, there has been a great deal of interest in grain error channels, which are related to high-density encoding on magnetic media. A grain in a magnetic medium has a single polarization. If an encoder attempts to write two symbols to a single grain, only one of them will be retained. Because the locations grain boundaries are generally unknown to the encoder, this situation can be modeled by a channel.

Mazumdar et al. applied the degree threshold bound to non-overlapping grain error channels [9]. Sharov and Roth applied the degree sequence bound to both non-overlapping and overlapping grain error channels [16]. Kashyap and Zémor applied the local degree bound to improve on Mazumdar et al. for the 1,2, or 3 error cases [3]. They conjectured an extension for larger numbers of errors. Gabrys et al. applied the local degree bound to improve on Sharov and Roth [4].

The input and output of this channel are strings \( x, y \in \{0,1\}^n \). To produce an output from an input, choose a grain pattern with at most one grain of length two and no larger grains. The grain of length two, if it exists, bridges indices \( j \) and \( j+1 \) for some \( 0 \leq j \leq n-2 \). Then the channel output is

\[
y_i = \begin{cases} 
    x_i & i \neq j \\
    x_{i+1} & i = j 
\end{cases}
\]

If \( u_j = u_{j+1} \) or if there is no grain of length two, then \( y = x \).

The degree of an input string is equal to the number of runs \( r \); each of the \( r-1 \) run boundaries could be bridged by a grain or there could be no error. A grain error reduces the number of runs by 0,1, or 2. The number of runs is reduced by 1 if \( j = 0 \) and \( x_0 \neq x_1 \), by 2 if \( j \geq 1, x_j \neq x_{j+1} \), and by 0 otherwise. Equivalently, the number of runs is reduced by 1 if \( x \) has a length-1 run at index 0 is eliminated and by 2 if a length-1 run elsewhere is eliminated.

In the previous section, we let \( u_j \) be the number of length-1 runs in \( x \) and \( b_j \) be the number of length-1 runs appearing at the start or end of \( x \). For the grain channel, we need to distinguish between length-1 runs at the start and at the end, so let \( b_l^x \) and \( b_r^x \) count these.

The bipartite graph for this channel, \( B \), has a useful symmetry. Define \( z \in \{0,1\}^n \) to be the alternating string that starts with zero, so \( z_j = i \mod 2 \). Then \( (x, y) \in E(B) \) if and only if \( (y+z, x+z) \in E(B) \). If \( x = y \), this is trivially true. If \( x \neq y \), there is some \( j \) such that \( y_j = y_{j+1} = x_{j+1}, x_j \neq x_{j+1} \), and \( x_i = y_i \) for \( i \neq j, j+1 \). Then \( (y+z)_{j+1} \neq (y+z)_j+1 \) and \( (x+z)_{j+1} = (x+z)_j+1 \) for all \( j \), so \( y+z, x+z \) is an edge.

Thus, the degree of an output string \( y+z \) is equal to the degree of the input string \( y+z \), which is \( r_y+z \). Because of this, it is useful to define \( t_v = r_v+z, v_y = u_y+z, c_y^t = b_l^y+z, \) and \( c_y^r = b_r^y+z \).

**Theorem 5.** Let \( A_n \) be the primal hypergraph for the \( n \)-bit 1-grain-error channel. The vector

\[
z_y = \frac{1}{r_y} \left( 1 + \frac{2u_y - 2b_y^t - b_y^r - 2}{(r_y + 2)(r_y + 1)} \right)^{-1}
\]

is feasible for \( \kappa^*(A_n) \) and

\[
w_x = \frac{1}{t_x} \left( 1 + \frac{2u_x - 2c_x^t - c_x^r + 12}{(t_x + 1)t_x} \right)^{-1}
\]

is feasible for \( p^*(A_n) \).

By applying the techniques of Section VII-A it can be shown that Theorem 5 implies that \( \kappa^*(A_n) = \frac{2^n}{n+1}(1 + O(n^{-2})) \).

**References**

[1] D. Cullina and N. Kiyavash, “An improvement to Levenshtein’s upper bound on the cardinality of deletion correcting codes,” in *IEEE International Symposium on Information Theory* Proceedings, Jul. 2013.

[2] A. A. Kulkarni and N. Kiyavash, “Non-asymptotic upper bounds for deletion correcting codes,” *IEEE Transactions on Information Theory*, 2012. [Online]. Available: [http://arxiv.org/abs/1211.3128](http://arxiv.org/abs/1211.3128)

[3] N. Kashyap and G. Zmor, “Upper bounds on the size of grain-correcting codes,” *arXiv preprint arXiv:1302.6154*, 2013. [Online]. Available: [http://arxiv.org/abs/1302.6154](http://arxiv.org/abs/1302.6154)

[4] R. Gabrys, E. Yaakov, and D. Londeck, “Correcting grain-errors in magnetic media,” in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, 2013. p. 689693.

[5] S. Buzaglo, E. Yaakov, T. Etzion, and J. Bruck, “Error-correcting codes for multipermutations,” 2013. [Online]. Available: [http://authors.library.caltech.edu/56940](http://authors.library.caltech.edu/56940)

[6] D. B. West, *Introduction to graph theory*. Prentice Hall Upper Saddle River, 2001, vol. 2.

[7] J. Hastad, “Clique is hard to approximate within n 1-\epsiv,” in *Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on.* IEEE, 1996, p. 627636.

[8] V. I. Levenshtein, “Binary codes capable of correcting deletions, insertions, and reversals,” in *Soviet physics doklady*, vol. 10, 1966, p. 707710.

[9] A. Mazumdar, A. Barg, and N. Kashyap, “Coding for high-density recording on a 1-d granular magnetic medium,” *IEEE Transactions on Information Theory*, vol. 57, no. 11, p. 74037417, 2011.

[10] C. Shannon, “The zero error capacity of a noisy channel,” *IRE Transactions on Information Theory*, vol. 2, no. 3, pp. 8–19, 1956.

[11] M. Rosenfeld, “On a problem of C. E. Shannon in graph theory,” *Proceedings of the American Mathematical Society*, vol. 18, no. 2, pp. 315–319, Apr. 1967.

[12] L. Lovasz, “On the Shannon capacity of a graph,” *IEEE Transactions on Information Theory*, vol. 25, no. 1, pp. 1–7, Jan. 1979.

[13] N. Alon and J. H. Spencer, “Turán’s theorem,” in *The probabilistic method*. John Wiley & Sons, 2004, pp. 95–96.

[14] A. Fazeli, A. Vardy, and E. Yaakov, “Generalized sphere packing bound,” *arXiv:1401.6496 [cs, math]*, Jan. 2014. [Online]. Available: [http://arxiv.org/abs/1401.6496](http://arxiv.org/abs/1401.6496)

[15] N. J. A. Sloane, *Challenge Problems: Independent Sets in Graphs.* [Online]. Available: [http://neilsloane.com/doc/graphs.html](http://neilsloane.com/doc/graphs.html)

[16] A. Sharov and R. M. Roth, “Bounds and constructions for granular media coding,” in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, 2011. p. 23432347.

**APPENDIX A**

**Proofs**

**Lemma 10**: \( \kappa^*(A_{q,n,a,b}) \leq \kappa^*(A_{q,n,a+2,b-1}) \) when \( a + gb \leq n - 1 \).
Proof: We can rewrite the initial inequality as
\[
\kappa^* (A_{q,n,a+2,b-1}) \geq \kappa^* (A_{q,n,a,b})
\]
\[
\sum_{i=0}^{b-1} \binom{b-a-2}{i} (q-1)^i \geq \sum_{i=0}^{b} \binom{b-a}{i} (q-1)^i
\]
\[
\sum_{i=0}^{b} \binom{n-a-2}{i} (q-1)^i \geq \sum_{i=0}^{b-1} \binom{n-a-2}{i} (q-1)^i \quad (4)
\]

To simplify (4), we use the following identity:
\[
\sum_{i=0}^{b} \binom{n-c+2}{i} (q-1)^i
\]
\[
= \sum_{i=0}^{b} \left( \binom{n-c+2}{i-2} + 2 \binom{n-c}{i-1} + \binom{n-c}{i} \right) (q-1)^i
\]
\[
= \sum_{i=0}^{b-2} \binom{n-c}{i} (q-1)^i + 2 \sum_{i=0}^{b-1} \binom{n-c}{i} (q-1)^i + \sum_{i=0}^{b} \binom{n-c}{i} (q-1)^i
\]
\[
= \binom{n-c}{b} (q-1)^b - \binom{n-c}{b-1} (q-1)^{b+1} + \binom{n-c}{i} (q-1)^i
\]
\[
= \binom{n-c}{b} (q-1)^b \left( \frac{n-c-b+1}{b} - q + 1 \right) + q^b \sum_{i=0}^{b-1} \binom{n-c}{i} (q-1)^i.
\]

By setting \( c = a + 2 \), we can use this to rewrite the left side of (4). Eliminating the common term from both sides of the inequality gives
\[
\binom{n-a-2}{b-1} (q-1)^b \left( \frac{n-a-b-1}{b} - q + 1 \right) + q^b \sum_{i=0}^{b-1} \binom{n-c}{i} (q-1)^i \geq 0
\]
\[
\frac{n-a-b-1}{b} - q + 1 \geq 0
\]
\[
\frac{n-a-1}{b} - q + 1 \geq 0
\]
which proves the claim.

Theorem 3 Let
\[
f(r,u,b) \triangleq \frac{1}{r} \left( 1 + \frac{2u - b - 2}{r + (2b + r + 1)} \right)^{-1}.
\]
Then the vector \( z_y = f(r_y, u_y, b_y) \) is feasible for \( \kappa^*(A_n) \), so \( \kappa^*(A_n) \leq 17^2 z \).

Proof: By Lemma 14 \( \varphi \circ \varphi(1) \) is feasible for \( \kappa^*(A_n) \).

From the definition of \( \varphi \),
\[
\varphi(z)_y = \min_{x \in N(y)} (A_n)_x
\]
Each \( x \in [2]^n \) has \( r_x \) total substrings, so \( (A_n z''')_x = r_x \),
\[
\frac{1}{\varphi(1)_y} = \min_{x \in N(y)} (A_n)_x = \min_{x \in N(y)} r_x = r_y,
\]
and \( \varphi(1)_y = 1/r_y \).

Of the substrings of \( x \), \( u_x - b_x \) have \( r_x - 2 \) runs, \( b_x \) have \( r_x - 1 \) runs, and \( r_x - u_x \) have \( r_x \) runs, so
\[
(A_n \varphi(1))_x = \sum_{y \in N(x)} \frac{1}{d_y}
\]
\[
= u_x - b_x + \frac{b_x}{r_x - 1} + \frac{r_x - u_x}{r_x - 2}
\]
\[
= 1 + \frac{u_x}{r_x - 1} + \frac{b_x}{r_x - 2}
\]
\[
= 1 + \frac{2u_x (r_x - 1) - b_x r_x}{r_x (r_x - 1)(r_x - 2)}
\]
\[
= \frac{2u_x - b_x}{r_x (r_x - 1)(r_x - 2)}
\]
The inequality follows from \( u_x - b_x \geq 0 \).

Let \( y \in [2]^n \) be a string and let \( x \in [2]^n \) be a substring of \( y \). It is possible to create a substring by extending an existing run, adding a new run at an end of the string, or by splitting an existing run into three new runs, so \( r_x \leq r_y + 2 \) The only way to destroy a unit run in \( y \) is to extend it into a run of length two, so \( u_x \geq u_y - 1 \). Similarly, \( u_x - b_x \geq u_y - b_y - 1 \), so \( 2u_x - b_x \geq 2u_y - b_y - 2 \). Applying these inequalities to \( (A_n \varphi(1))_x \), we conclude that
\[
\frac{\varphi(1)_y}{(\varphi \circ \varphi(1))_y} = \min_{x \in N(y)} (A_n \varphi(1))_x \geq 1 + \frac{2u_y - b_y - 2}{(r_y + 2)(r_y + 1)}
\]
\[
(\varphi \circ \varphi(1))_y \leq \frac{1}{r_y} \left( 1 + \frac{2u_y - b_y - 2}{(r_y + 2)(r_y + 1)} \right)^{-1}.
\]

Lemma 16 For \( 2 \leq r \leq n - 1 \),
\[
\sum_{u=0}^{r-1} \binom{n-r-1}{r-u-1} \sum_{b=0}^{2} \binom{r-2}{b} (2u - b)
\]
\[
= \sum_{u=0}^{r-1} \binom{n-r-1}{r-u-1} \binom{r}{u} 2u - \sum_{b=0}^{2} \binom{r-2}{b} \binom{1}{b-1}
\]
\[
= \sum_{u=0}^{r-1} \binom{n-r-1}{r-u-1} \binom{r-1}{u-1} - \sum_{u=0}^{r-1} \binom{n-r-1}{r-u-1} \binom{r-1}{u-1}
\]
\[
= \frac{2(r-1)}{n-1} \left( \begin{array}{c} n-1 \\ r-1 \end{array} \right)
\]

Proof:
\[
\sum_{u=0}^{r-1} \binom{n-r-1}{r-u-1} \binom{r-2}{b} (2u - b)
\]
\[
= \sum_{u=0}^{r-1} \binom{n-r-1}{r-u-1} \binom{r}{u} 2u - \sum_{b=0}^{2} \binom{r-2}{b} \binom{1}{b-1}
\]
\[
= \sum_{u=0}^{r-1} \binom{n-r-1}{r-u-1} \binom{r-1}{u-1} - \sum_{u=0}^{r-1} \binom{n-r-1}{r-u-1} \binom{r-1}{u-1}
\]
\[
= \frac{2(r-1)}{n-1} \left( \begin{array}{c} n-2 \\ r-2 \end{array} \right)
\]
\[
= \frac{2(r-1)^2}{n-1} \left( \begin{array}{c} n-1 \\ r-1 \end{array} \right)
\]

■
Lemma 17 Define $f'(r, u, b)$ to be equal to

\[
\begin{cases}
\frac{1}{r} \left( 1 - \frac{2u - b}{(r+1)^2} \left( 1 - \frac{7}{r+1} + \frac{2}{(r+2)(r+1)} \right) \right) r > 1 \\
1 \quad r = 1
\end{cases}
\]

Then $z_y = f'(r_y, u_y, b_y)$ is feasible for $\kappa^*(A_n)$, so $\kappa^*(A_n) \leq 1Tz$. 

Proof: Recall from Theorem 3 that

\[
\kappa^*(A_{n+1}) \leq 2f'(n, n, 2) + 2f'(1, 0, 0) + 2 \sum_{r=2}^{n-1} g(n, r).
\]

From Lemma 16 $g(n, r)$ equals

\[
\frac{1}{n} \left( \frac{2}{n-1} \left( 1 - \frac{7}{n+1} \right) + \frac{2}{(r+2)(r+1)} \right) (n-1)
\]

Extend the definition of $g(n, r)$ for $r = 1$ and $r = n$ using this rational function. Note that $f'(1, 0, 0) \leq g(n, 1)$ because

\[
f'(1, 0, 0) = 1
\]

and that $f'(n, n, 2) = g(n, n)$ because both equal

\[
1 - \frac{2}{n-1} + \frac{14}{(n+1)(n-1)} + \frac{2}{(n+2)(n+1)}.
\]

Thus

\[
2f(n, n, 2) + 2f'(1, 0, 0) + 2 \sum_{r=2}^{n-1} g(n, r)
\]

\[
\leq 2 \sum_{r=1}^{n} g(n, r)
\]

\[
\leq 2 \frac{1}{n} \left( \frac{(n-3)2^n}{n-1} + \frac{14(2^{n+1})}{(n+1)(n-1)} + \frac{2(2^{n+2})}{n+2} \right)
\]

\[
= \frac{2n+1}{n} \left( \frac{(n-3)2^n}{n-1} + \frac{28}{(n+1)(n-1)} + \frac{8}{(n+2)(n+1)} \right)
\]

\[
= \frac{2n+1}{n} \left( \frac{n^2 - n - 6}{(n+2)(n-1)} + \frac{28(n+2) + 8(n-1)}{(n+2)(n+1)(n-1)} \right)
\]

\[
= \frac{2n+1}{n} \left( \frac{n + 2}{(n+2)(n+1)(n-1)} \right)
\]

\[
1Tz \geq \frac{2n+1}{n} \left( \frac{30n + 12}{n(n-1)(n-2)} \right)
\]

Lemma 18 Let $z_y = f''(r_y, u_y, b_y)$. Then

\[
1Tz \leq \frac{2n+1}{n} \left( \frac{30n + 12}{n(n-1)(n-2)} \right)
\]

Proof:

\[
f''(r, u, b) \geq \frac{1}{r} \left( 1 - \frac{u - b}{r^2} \right)
\]

A variant of (5) is

\[
\sum_{u=0}^{r} \left( \frac{n-r-1}{r-u-1} \right) \sum_{b=0}^{2} \left( \frac{r-2}{u-b} \right) \frac{2}{b} \frac{u - b}{(r+1)(r-2)}
\]

\[
= \frac{1}{n-1} \left( \frac{n-1}{r-1} \right)
\]
which we apply to show that
\[ 2 \sum_{r=1}^{n} \sum_{u=0}^{n-r} \binom{n-r-1}{r-u-1} \sum_{b=0}^{2} \binom{r-2}{b} f''(r, u, b) \]
\[ = 2 \sum_{r=1}^{n} \frac{1}{r} \left( 1 - \frac{1}{n-1} \right) \binom{n-1}{r-1} \]
\[ = \frac{2}{n} \sum_{r=1}^{n} \left( 1 - \frac{1}{n-1} \right) \binom{n}{r} \]
\[ = \frac{2}{n} \left( \frac{n-2}{n-1} \right) (2^n - 1) \]
\[ = \frac{2^{n+1} - 2}{n + 2} \left( \frac{n^2 - 4}{n(n-1)} \right) \]
\[ = \frac{2^{n+1} - 2}{n + 2} \left( 1 + \frac{1}{n} - \frac{3}{n(n-1)} \right) \]

\[ \varphi(1)_y = \min_{x \in \mathcal{N}(y)} (A(1)x) = \min_{x \in \mathcal{N}(y)} \varphi(1)_y = \min_{x \in \mathcal{N}(y)} \frac{1}{t_y} \left( 1 + \frac{2u_y - 2b^R_y - b^L_y}{(r_y + 2)(r_y + 1)} \right) \]

\[ (\varphi \circ \varphi(1))_y = \min_{x \in \mathcal{N}(y)} (A(\varphi(1))x) \geq 1 + \frac{2u_y - 2b^R_y - b^L_y}{(r_y + 2)(r_y + 1)} - 1 \]

By Lemma \[ \psi \circ \varphi(1) \] is feasible for \( p^*(A) \). From the definition of \( \psi \),

\[ w_x = \max_{y \in \mathcal{N}(x)} (A^T_y w)_y \]

Each \( y \in [2]^n \) has \( t_y \) total neighbors, so \( (A_y z''')_x = r_x \),

\[ \frac{1}{t_y} = \max_{y \in \mathcal{N}(x)} (A^T_y 1)_y = \max_{y \in \mathcal{N}(x)} t_y = \min(t_x + 2, n), \]

and \( \psi(1)_x \geq 1/(t_x + 2) \).

Then \( (A^T_y \psi(1))_y \) equals

\[ \sum_{x \in \mathcal{N}(y)} \frac{1}{t_y} = \frac{v_y - c^R_y - c^L_y}{t_y} + \frac{c^L_y}{t_y + 1} + \frac{c^R_y}{t_y + 2} \]

\[ = \frac{t_y}{t_y + 2} + \frac{(v_y - c^R_y)}{(t_y + 1)} + \frac{t_y}{t_y + 2} \]

\[ = \frac{t_y}{t_y + 2} + \frac{2(v_y - c^R_y)}{(t_y + 1)} + \frac{t_y}{t_y + 2} \]

\[ \leq 1 - \frac{2}{t_y + 2} + \frac{2v_y - 2c^R_y - c^L_y}{(t_y + 2)(t_y + 1)} \]

The inequality follows from \( v_y - c^R_y - c^L_y \leq v_y \leq t_y \). Let \( y \in [2]^n \) be an output and let \( x \in \mathcal{N}(y) \). A grain error cannot increase the number of runs, so \( r_y \geq v_y \) and \( t_y \geq t_x \). A grain error can reduce the number of unit runs by at most 3, so \( u_x \leq u_y + 3 \) and \( v_x \leq v_y + 3 \). Similarly, \( 2v_y - 2c^R_y - c^L_y \leq 2v_x - 2c^R_y - c^L_x + 6 \). Applying these inequalities to (6), we conclude that \( (A^T_y \psi(1))_y \) is at most

\[ \frac{1}{t_y + 4} + \frac{2v_y - 2c^R_y - c^L_y + 8}{(t_y + 2)(t_y + 1)} \]

\[ = \frac{t_y}{t_y + 2} + \frac{(t_y + 2)^2}{(t_y + 4)} + \frac{2v_x - 2c^R_y - c^L_y + 8}{(t_y + 1)} \]

\[ \leq \frac{t_y}{t_x + 2} + \frac{4}{(t_x + 4)t_x} + \frac{2v_x - 2c^R_y - c^L_y + 8}{(t_x + 1)} \]

\[ \leq \frac{t_x}{t_x + 2} + \frac{4}{(t_x + 4)t_x} + \frac{2v_x - 2c^R_y - c^L_y + 12}{(t_x + 1)t_x} \]
and that

$$\psi(1)_x \leq \max_{y \in N(x)} (A \phi(1))_y \geq 1 + \frac{2v_x - 2e^R_x - e^L_x + 12}{(t_x + 1)t_x},$$

$$\varphi(1)_y \leq \frac{1}{t_x} \left( 1 + \frac{2v_x - 2e^R_x - e^L_x + 12}{(t_x + 1)t_x} \right)^{-1}.$$