Computing Largest Correcting Codes and Their Estimates Using Optimization on Specially Constructed Graphs

Sergiy Butenko

Department of Industrial Engineering

Texas A&M University

College Station, TX 77843

Joint work with P. Pardalos, I. Sergienko, V. Shylo, and P. Stetsyuk

Outline

- Introduction
- Maximum clique/independent set problems
- Error-correcting codes
- Lower bound for codes correcting one error on the Z-channel
- Conclusion

Introduction

Definitions:

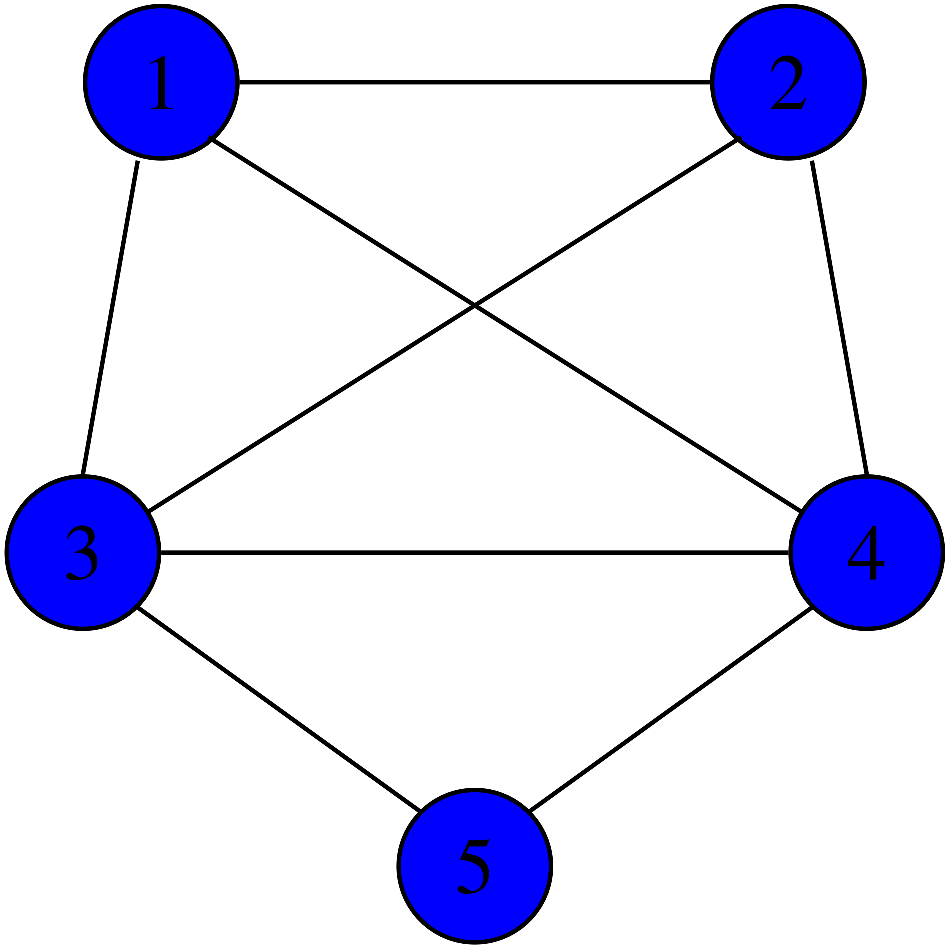
$G = (V, E)$ is a simple undirected graph,
$V = \{1, 2, \ldots, n\}$.

$\overline{G} = (V, \overline{E})$, is the *complement graph of* $G = (V, E)$, where $\overline{E} = \{(i, j) \mid i, j \in V,\ i \neq j$ and $(i, j) \notin E\}$.

For $S \subseteq V$, $G(S) = (S, E \cap S \times S)$ *the subgraph induced by* S.

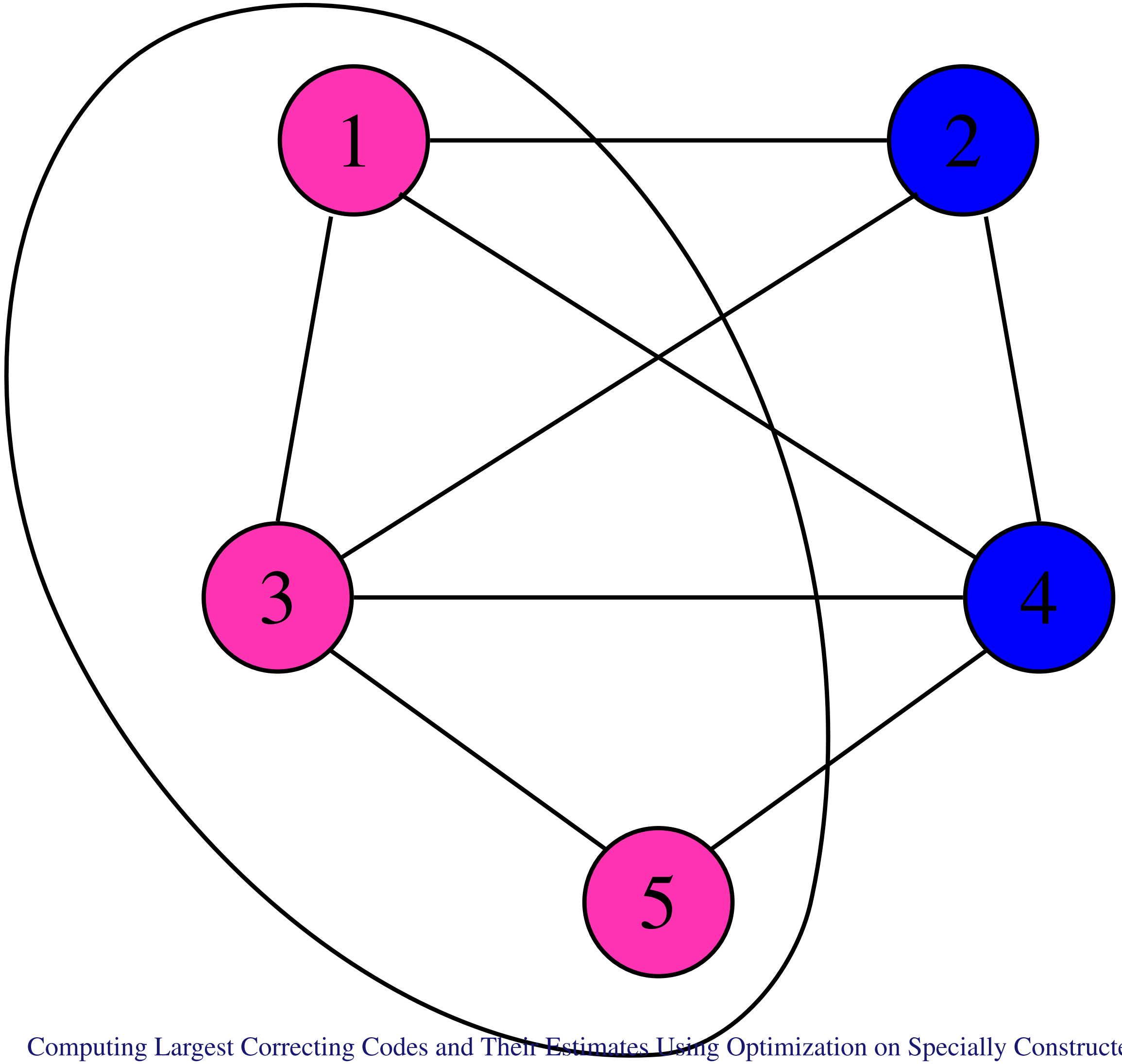
Introduction

Example:
$V =$
$\{1, 2, 3, 4, 5\}$
$E =$
$\{(1,2), (1,3),$
$(1,4), (2,3),$
$(2,4), (3,4),$
$(3,5), (4,5)\}$



Introduction

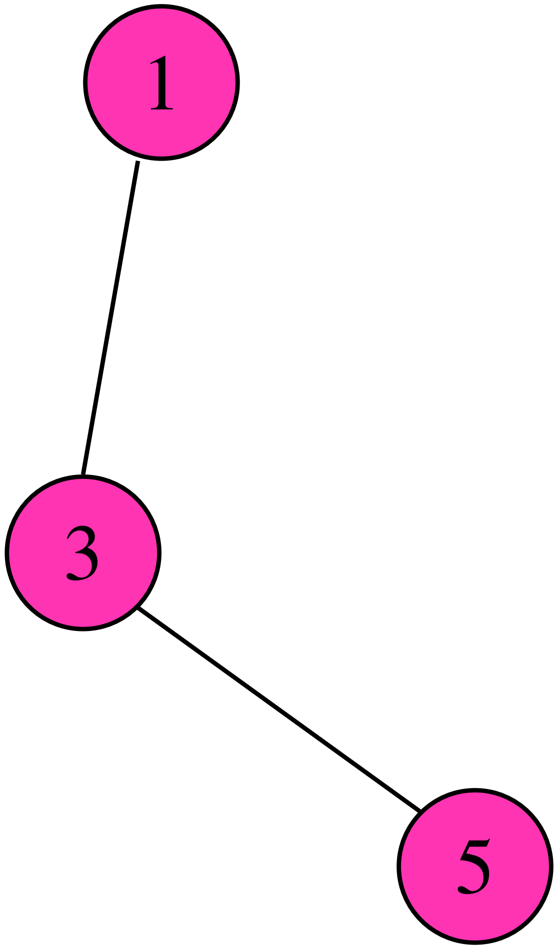
$S = \{1, 3, 5\}$



Introduction

$S = \{1, 3, 5\}$

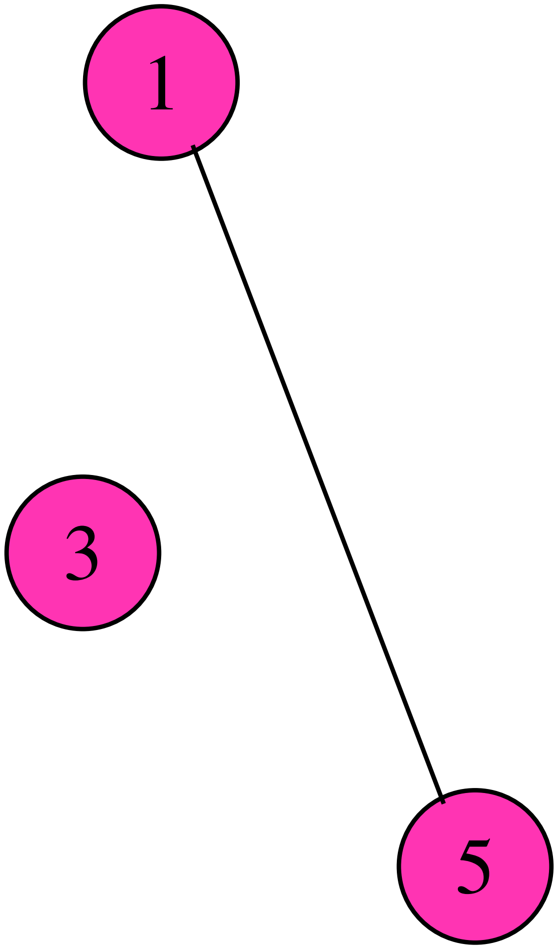
$G(S)$:



Introduction

$S = \{1, 3, 5\}$

$\overline{G(S)}$:



Introduction

A subset $I \subseteq V$ is called an *independent set* (stable set, vertex packing) if $G(I)$ has no edges.

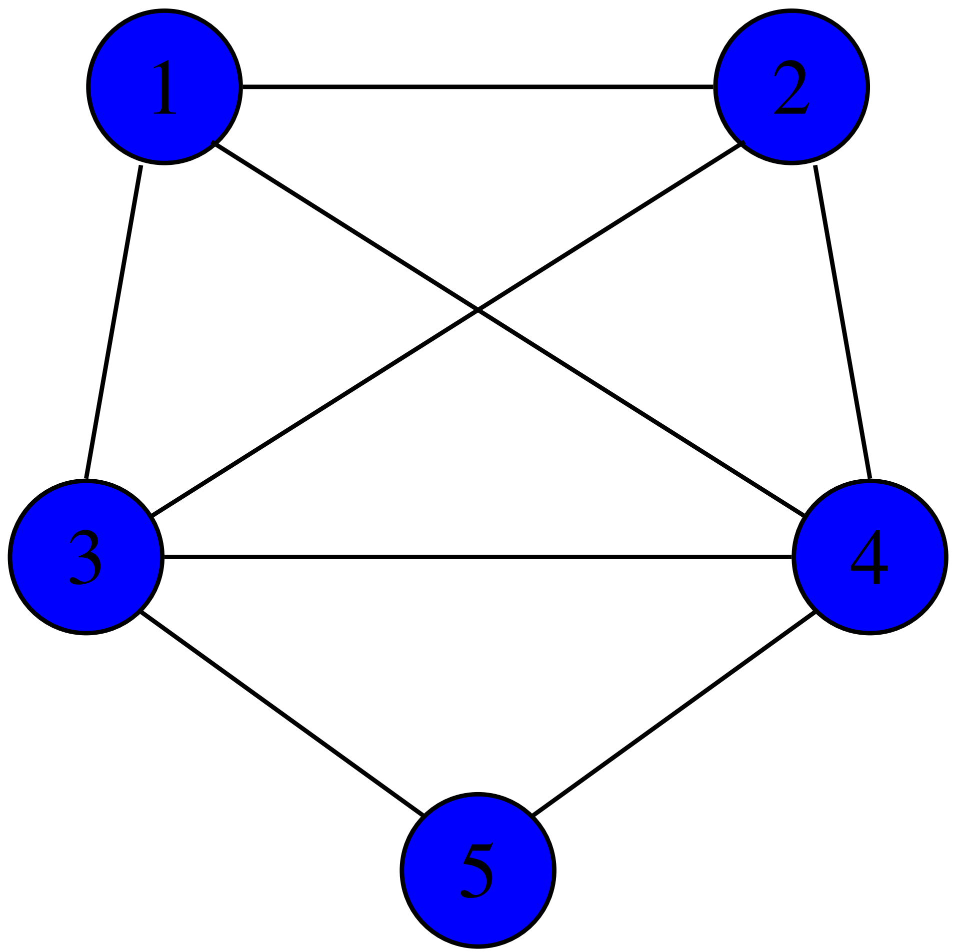
A subset $C \subseteq V$ is called a *clique* if $G(C)$ is complete, i.e. it has all possible edges.

An independent set (clique) is said to be

- *maximal*, if it is not a subset of any larger independent set (clique);
- *maximum*, if there is no larger independent set (clique) in the graph.

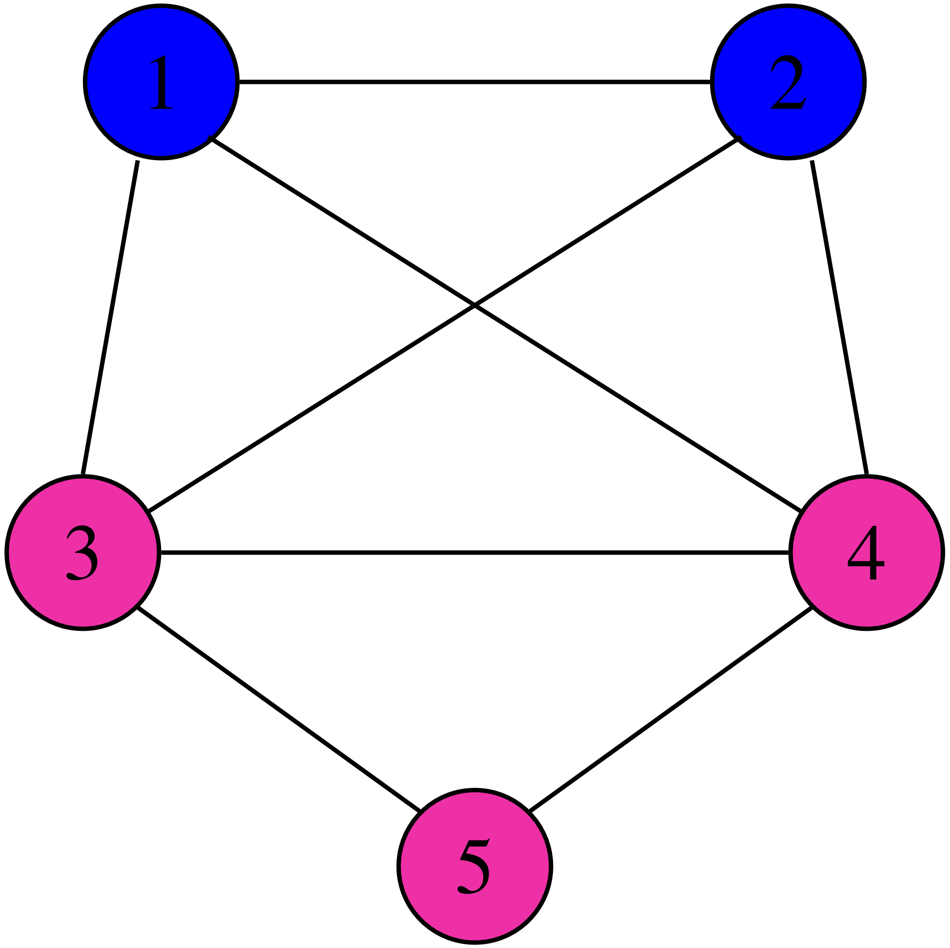
Introduction

Example:



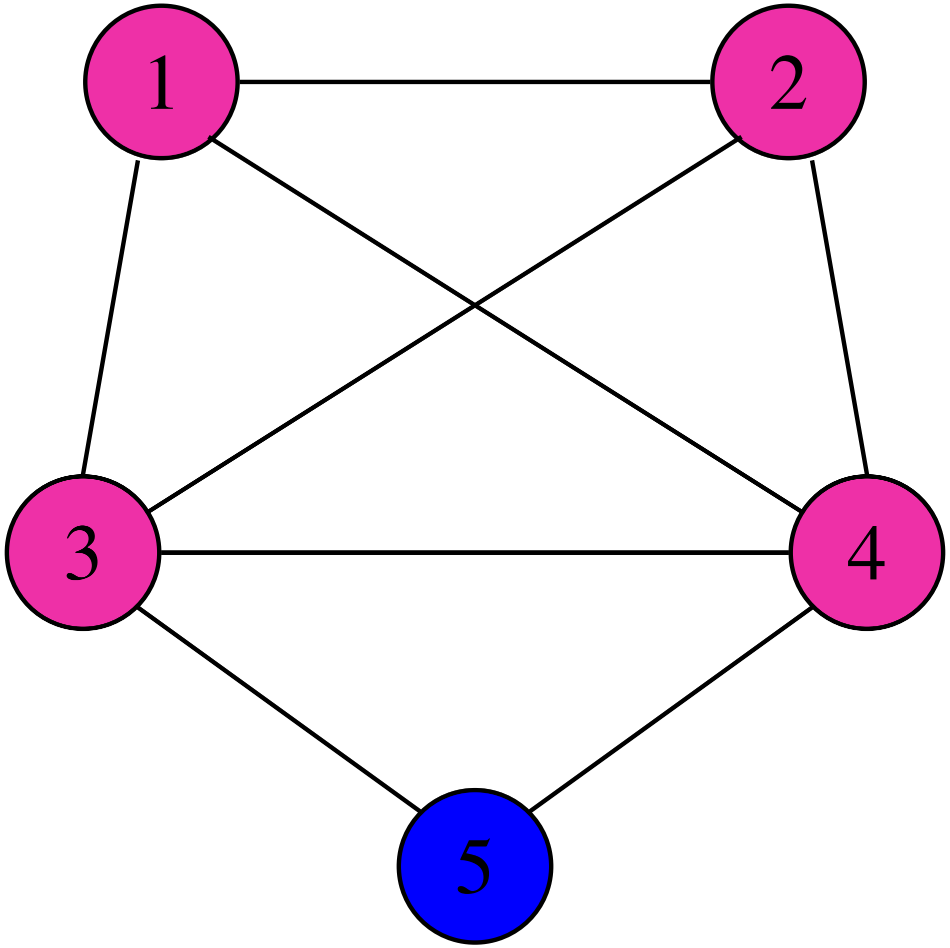
Introduction

A maximal clique: $\{3, 4, 5\}$



Introduction

The maximum clique: $\{1, 2, 3, 4\}$



Introduction

$\alpha(G)$ – the *independence (stability) number* of G.
$\omega(G)$ – the *clique number* of G.

$VC \subseteq V$ is a *vertex cover* if every edge has at least one endpoint in VC.

$$I \text{ is a maximum independent set of } G$$
$$\Updownarrow$$
$$I \text{ is a maximum clique of } \bar{G}$$
$$\Updownarrow$$
$$V \setminus I \text{ is a minimum vertex cover of } G.$$

MC, MIS and MVC problems are NP-hard

Error-correcting Codes

Given:

Set B^n of all binary vectors of length n;

For $u \in B^n$ denote by

$$F_e(u) = \left\{ v : u \xrightarrow{\text{error } e} v \right\}$$

A subset $C \subseteq B^n$ is said to be an e-correcting code if $F_e(u) \bigcap F_e(v) = \emptyset$ for all $u, v \in C,\ u \neq v$.

Find:

The largest correcting code.

Error-correcting Codes

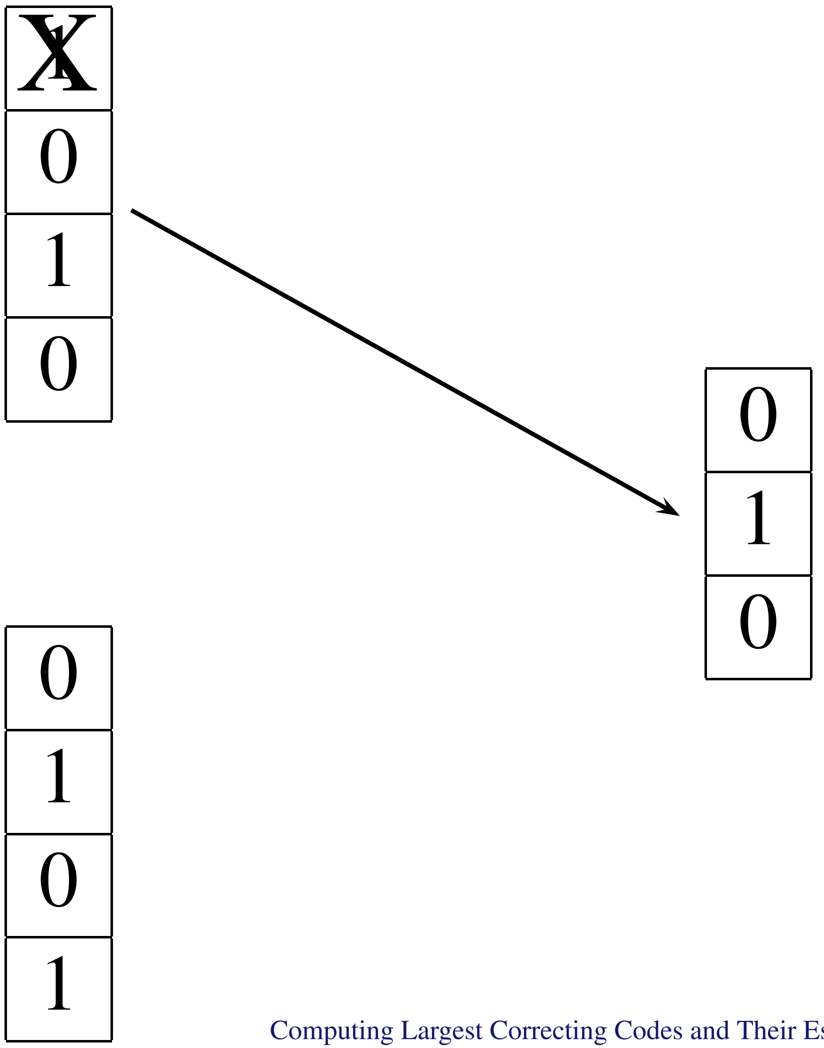
Example: Single Deletion

1
0
1
0

0
1
0
1

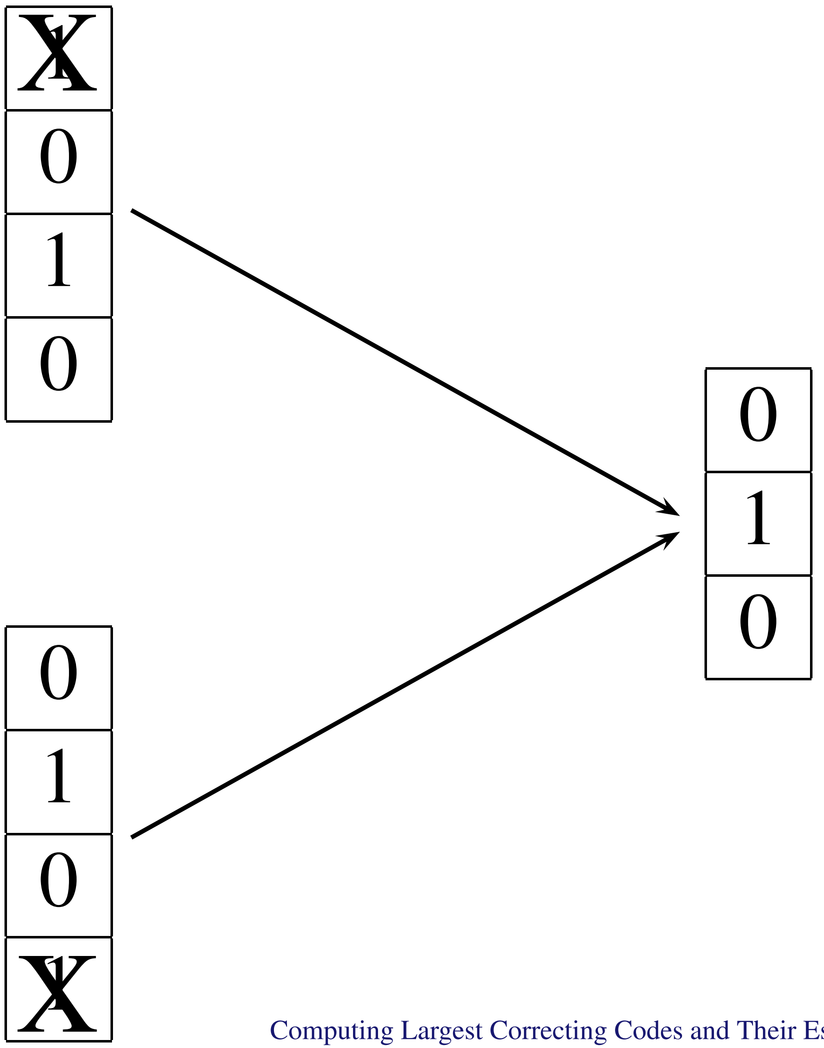
Error-correcting Codes

Example: Single Deletion



Error-correcting Codes

Example: Single Deletion



Error-correcting Codes

We construct the following graph $G_n = (V_n, E_n^{(e)})$:

- $V_n = B^n$;
- $(u, v) \in E_n^{(e)}$ if and only if $u \neq v$ and

$$F_e(u) \bigcap F_e(v) \neq \emptyset.$$

Then a correcting code corresponds to an independent set in G_n. Hence, the largest e-correcting code can be found by solving the maximum independent set problem in the considered graph.

Error-correcting Codes

- Single-Deletion-Correcting Codes (1dc);
- Two-Deletion-Correcting Codes (2dc);
- Codes For Correcting a Single Transposition, Excluding the End-Around Transposition (1tc);
- Codes For Correcting a Single Transposition, Including the End-Around Transposition (1et);
- Codes For Correcting One Error on the Z-Channel (1zc).

http://www.research.att.com/~njas/doc/graphs.html

(Neil Sloane's webpage)

Error-correcting Codes

- Preprocessing: Simplicial vertices are removed and connected components are considered separately.
- Clique Partitioning: We partition the set of vertices V of G as follows:

$$V = \bigcup_{i=1}^{k} C_i,$$

where C_i - cliques such that $C_i \cap C_j = \emptyset,\ i \neq j$.

Error-correcting Codes

- An upper bound:

$$O_{\mathcal{C}}(G) = \max \sum_{i=1}^{n} x_i$$

$$\text{s. t.} \quad \sum_{i \in C_j} x_i \le 1, j = 1, \ldots, m$$

$$x \ge 0.$$

where $C_j \in \mathcal{C}$ is a maximal clique, $\mathcal{C}$- a set of maximal cliques, $|\mathcal{C}| = m$.

Error-correcting Codes

Branch-and-Bound algorithm

- Branching: Based on the fact that the number of vertices from a clique that can be included in an independent set is always equal to 0 or 1.

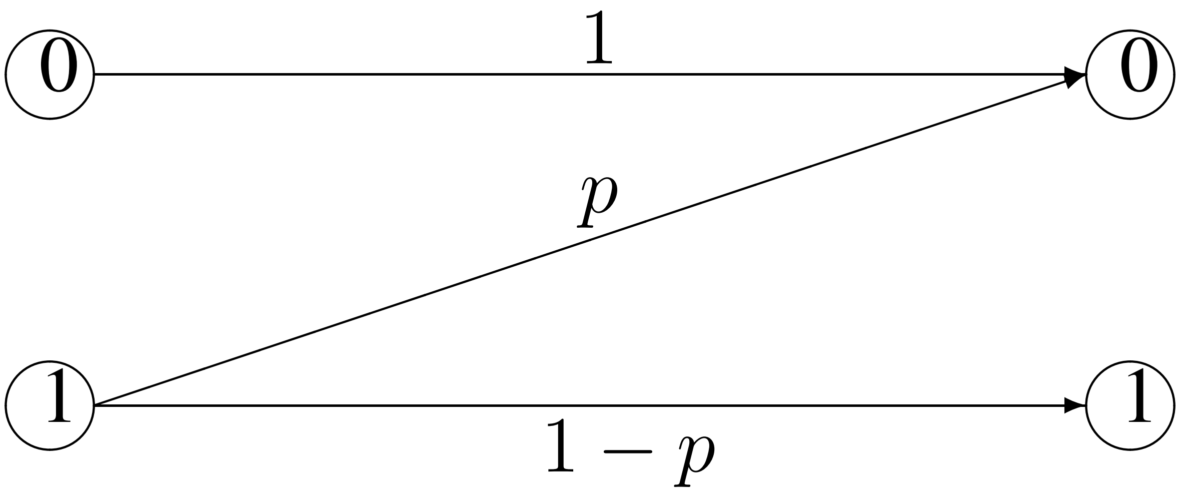
- Bounding: We use a heuristic solution as a lower bound and $O_{\mathcal{C}}(G)$ as an upper bound.

Error-correcting Codes

Exact Solutions Found.

Graph	$\|V\|$	$\|E\|$	$\alpha(G)$
1dc512	512	9727	52
2dc512	512	54895	11
1tc128	128	512	38
1tc256	256	1312	63
1tc512	512	3264	110
1et128	128	672	28
1et256	256	1664	50
1et512	512	4032	100

One Error on the Z-Channel



A scheme of the Z-channel

One Error on the Z-Channel

n	Lower bound	Upper bound
4	4	4
5	6	6
6	12	12
7	18	18
8	36	36
9	62	62
10	112	117
11	198	210
12	379*	410

Varshamov (1973), Constantin and Rao (1979), Delsarte and Piret (1981), Etzion and Östergard (1998)

One Error on the Z-Channel

The asymmetric distance $d_A(x, y)$ between vectors $x, y \in B^n$ is defined as follows:

$$d_A(x, y) = \max\{N(x, y), N(y, x)\},$$

where $N(x, y) = |\{i : (x_i = 0) \wedge (y_i = 1)\}|$. It is related to the Hamming distance $d_H(x, y) = \sum_{i=1}^{n} |x_i - y_i| = N(x, y) + N(y, x)$ by the expression

$$2d_A(x, y) = d_H(x, y) + |w(x) - w(y)|$$

One Error on the Z-Channel

The minimum asymmetric distance Δ for a code $C \subset B^n$ is defined as

$$\Delta = \min \{d_A(x, y) | x, y \in C, x \neq y\}.$$

Rao and Chawla (1975): A code C with the minimum asymmetric distance Δ can correct at most $(\Delta - 1)$ asymmetric errors.

We consider $\Delta = 2$.

One Error on the Z-Channel

The partitioning method (Van Pul and Etzion, 1989)

$$V(n) = \bigcup_{i=1}^{m} I_i, \; I_i \text{ is an independent set}, \quad I_i \bigcap I_j = \emptyset, \; i \neq j.$$

$$\Pi(n) = (I_1, I_2, \ldots, I_m).$$

The *index vector* of partition $\Pi(n)$:

$$\pi(n) = (|I_1|, |I_2|, \ldots, |I_m|),$$

We assume that $|I_1| \geq |I_2| \geq \ldots \geq |I_m|$.

One Error on the Z-Channel

Constant weight codes of weight w

Construct a graph $G(n, w)$

- $\binom{n}{w}$ vertices
- x and y are adjacent *iff* $d_H(x, y) < 4$
- an independent set partition

$$\Pi(n, w) = (I_1^w, I_2^w, \ldots, I_m^w)$$

(each ind. set is a subcode with minimum Hamming distance 4)

One Error on the Z-Channel

The *direct product* $\Pi(n_1) \times \Pi(n_2, w)$ of a partition of asymmetric codes $\Pi(n_1) = (I_1, I_2, \ldots, I_{m_1})$ and a partition of constant weight codes $\Pi(n_2, w) = (I_1^w, I_2^w, \ldots, I_{m_2}^w)$ is the set of vectors

$$C = \{(u, v) : \; u \in I_i, \, v \in I_i^w, \; 1 \leq i \leq m\},$$

where $m = \min\{m_1, m_2\}$.

Etzion and Östergard (1998): C is a code of length $n = n_1 + n_2$ with minimum asymmetric distance 2, *i.e.* a code correcting one error on the Z-channel of length $n = n_1 + n_2$.

One Error on the Z-Channel

A procedure for finding a code C of length n and minimum asymmetric distance 2:

1. Choose n_1 and n_2 such that $n_1 + n_2 = n$.
2. Choose $\epsilon = 0$ or 1.
3. Compute $\Pi(n_1)$ and $\Pi(n_2, 2i + \epsilon), i = 0, \ldots, \lfloor n_2/2 \rfloor$.
4. Set

$$C = \bigcup_{i=0}^{\lfloor n_2/2 \rfloor} \left(\Pi(n_1) \times \Pi(n_2, 2i + \epsilon) \right).$$

One Error on the Z-Channel

INPUT: $G = (V, E)$;
OUTPUT: $I_1, I_2, \ldots, I_m$.

0. i=0;
1. **while** $G \neq \emptyset$

    for $j = 1$ **to** k

        Find a maximal independent set IS_j;

        if $|IS_j| < |IS_{j-1}|$ **break**

    end

    Construct graph $\mathcal{G}$;

    Find a maximal independent set $MIS = \{IS_{i_1}, \ldots, IS_{i_p}\}$ of $\mathcal{G}$;

    $I_{i+q} = IS_{i_q},\ q = 1, \ldots, p$;

    $$G = G - \bigcup_{q=1}^{p} G(I_{i+q});\ \ i = i + p;$$

   end

One Error on the Z-Channel

- $\Pi(n, 0)$ consists of one (zero) codeword,
- $\Pi(n, 1)$ consists of n codes of size 1,
- $\Pi(n, 2)$ consists of $n - 1$ codes of size $n/2$ for even n,
- Index vectors of $\Pi(n, w)$ and $\Pi(n, n - w)$ are equal;

One Error on the Z-Channel

Partitions of asymmetric codes found.

n	#	Partition index vector	Norm	m
8	1	36,34, 34, 33, 30, 29, 26, 25, 9	7820	9
9	1	62, 62, 62, 61, 58, 56, 53, 46, 29, 18, 5	27868	11
	2	62, 62, 62, 62, 58, 56, 53, 43, 32, 16, 6	27850	11
	3	62, 62, 62, 61, 58, 56, 52, 46, 31, 17, 5	27848	11
	4	62, 62, 62, 62, 58, 56, 52, 43, 33, 17, 5	27832	11
	5	62, 62, 62, 62, 58, 56, 54, 42, 31, 15, 8	27806	11
10	1	112, 110, 110, 109, 105, 100, 99, 88, 75, 59, 37, 16, 4	97942	13
	2	112, 110, 110, 109, 105, 101, 96, 87, 77, 60, 38, 15, 4	97850	13
	3	112, 110, 110, 108, 106, 99, 95, 89, 76, 60, 43, 15, 1	97842	13
	4	112, 110, 110, 108, 105, 100, 96, 88, 74, 65, 38, 17, 1	97828	13

One Error on the Z-Channel

Partitions of constant weight codes obtained

k	w	Partition index-vector	Norm	m
10	4	30, 30, 30, 30, 26, 25, 22, 15, 2	5614	9
12	4	51, 51, 51, 51, 49, 48, 48, 42, 42, 37, 23, 2	22843	12
12	4	51, 51, 51, 51, 49, 48, 48, 45, 39, 36, 22, 4	22755	12
12	4	51, 51, 51, 51, 49, 48, 48, 45, 41, 32, 22, 6	22663	12
12	6	132, 132, 120, 120, 110, 94, 90, 76, 36, 14	99952	10
14	4	91, 91, 88, 87, 84, 82, 81, 79, 76, 73, 66, 54, 38, 11	78399	14
14	4	91, 90, 88, 85, 84, 83, 81, 79, 76, 72, 67, 59, 34, 11, 1	78305	15
14	6	278, 273, 265, 257, 250, 231, 229, 219, 211, 203, 184, 156, 127, 81, 35, 4	672203	16

One Error on the Z-Channel

Example: $n = 18,\ n_1 = 8,\ n_2 = 10.$

$\Pi(8) = \{36, 34, 34, 33, 30, 29, 26, 25, 9\};$
$\Pi(10, 4) = \{30, 30, 30, 30, 26, 25, 22, 15, 2\}.$

- $|\Pi(8) \times \Pi(10, 0)| = |\Pi(8) \times \Pi(10, 10)| = 36 \cdot 1 = 36;$
- $|\Pi(8) \times \Pi(10, 2)| = |\Pi(8) \times \Pi(10, 8)| = 256 \cdot 5 = 1280;$
- $|\Pi(8) \times \Pi(10, 4)| = |\Pi(8) \times \Pi(10, 6)| = 36 \cdot 30 + 34 \cdot 30 + 34 \cdot 30 + 33 \cdot 30 + 30 \cdot 26 + 29 \cdot 25 + 26 \cdot 22 + 25 \cdot 15 + 9 \cdot 2 = 6580;$
- The total is $2(36 + 1280 + 6580) = 15792$ codewords.

One Error on the Z-Channel

Improved lower bounds. Previous results by:
(a)-Etzion (1991); (b)- Etzion and Östergard (1998)

	Lower bound	
n	new	previous
18	15792	15762(a)
19	29478	29334(b)
20	56196	56144(b)
21	107862	107648(b)
22	202130	201508(b)
24	678860	678098(b)

Conclusion

- Improved lower bounds and exact solutions for the size of largest error-correcting codes were obtained.
- Structural properties (automorphisms, ...) of the considered graphs can be utilized more efficiently to reduce problem size.
- We used computational approach. Can the problem be solved analytically?