New Upper Bounds for Grain-Correcting and Grain-Detecting Codes

Artyom Sharov, Ronny Roth

Department of Computer Science

Technion – Israel Institute of Technology

Combinatorial model

- $\langle s \rangle = \{0, 1, \ldots, s-1\}$
- Binary alphabet $\Sigma = \langle 2 \rangle$
- *Grain* (of length 2) ending at location $e \in \langle n \rangle \setminus \{0\}$ in a word $\mathbf{x} = (x_i)_{i \in [n]}$ overruns the value of cell e by that of cell $e-1$: $x_e \leftarrow x_{e-1}$
- *Grain pattern* $\mathcal{S} \subseteq \langle n \rangle \setminus \{0\}$ contains all the grain locations and inflicts errors to $\mathbf{x}$ by means of operator $\sigma_{\mathcal{S}}$

Example

- $n = 6$, $\mathbf{x} = 101011$, $\mathcal{S} = \{1, 3, 5\}$ and $\mathcal{S}' = \{1, 2\}$

$\boldsymbol{x}$:	1	0	1	0	1	1
$\mathcal{S}$:	1	0	1	0	1	1
$\sigma_{\mathcal{S}}(\boldsymbol{x})$:	1	1	1	1	1	1

$\boldsymbol{x}$:	1	0	1	0	1	1
$\mathcal{S}'$:	1	0	1	0	1	1
$\sigma_{\mathcal{S}'}(\boldsymbol{x})$:	1	1	0	0	1	1

Applications

- Conventional magnetic recording media: *grains*
- Wood *et al.* suggested a new approach for magnetizing areas as tiny as a single grain
- Mazumdar *et al.* considered a 1-D combinatorial error model with specific substitution errors and grains of length 1 and 2
- Related to shingled writing (Iyengar *et al.*)
- $\mathcal{S}$ has *overlaps* if there exist $e, e' \in \mathcal{S}$ such that $e' = e+1$

Combinatorial model (cont.)

- Words $\mathbf{x}, \mathbf{x}' \in \Sigma^n$ are *t-confusable* if there exist $\mathcal{S}$, $\mathcal{S}'$ of size t at most for which $\sigma_{\mathcal{S}}(\mathbf{x}) = \sigma_{\mathcal{S}'}(\mathbf{x}')$
- Code $\mathcal{C} \subseteq \Sigma^n$ is *t-grain-correcting* if no two distinct codewords are t-confusable
- Largest size $M(n,t)$, *rate* $R(\tau) = \overline{\lim}_{n\to\infty} \frac{1}{n} \log_2 M(n,t)$
- Code $\mathcal{C} \subseteq \Sigma^n$ is *∞-grain-detecting* if $\sigma_{\mathcal{S}}(\mathbf{x}) \notin \mathcal{C}$ for any $\mathbf{x} \in \mathcal{C}$ and any grain pattern $\mathcal{S}$
- Largest sizes $M_Z(n,t)$, $M_H(n,t)$ and rate $R_H(\tau)$
- Entropy function $\mathrm{H}(x) = -x \log_2 x - (1{-}x) \log_2(1{-}x)$

Upper bound on $M(n,t)$

Lemma 1. *Let n be a positive integer and let $t \leq \frac{n}{2}$ be an integer. Then*

$$M(n,t) \leq 2^{\lceil n/2 \rceil} \cdot M_Z(\lfloor n/2 \rfloor, t) \,.$$

Proof. Let $\mathcal{C}$ be a t-grain-correcting code of length n. For $\mathbf{x} = (x_i)_{i \in \langle \lceil n/2 \rceil \rangle}$, let

$$\mathcal{C}(\mathbf{x}) = \left\{ \boldsymbol{c} = (c_i)_{i \in \langle n \rangle} \in \mathcal{C} : \text{for all } i \in \langle \lceil n/2 \rceil \rangle \,, c_{2i} = x_i \right\} .$$

There exists $\mathbf{x} \in \Sigma^{\lceil n/2 \rceil}$ such that $|\mathcal{C}(\mathbf{x})| \geq \dfrac{|\mathcal{C}|}{2^{\lceil n/2 \rceil}}$.

Upper bound on $M(n,t)$ (cont.)

Equivalent code of length $\lfloor n/2 \rfloor$ and size $|\mathcal{C}(\mathbf{x})|$ correcting t asymmetric errors $1 \to 0$

$$\mathcal{C}^{\oplus} = \left\{ \mathbf{y} = (c_{2i} \oplus c_{2i+1})_{i \in \langle \lfloor n/2 \rfloor \rangle} : \boldsymbol{c} = (c_i)_{i \in \langle n \rangle} \in \mathcal{C}(\mathbf{x}) \right\} .$$

Therefore

$$|\mathcal{C}| \leq 2^{\lceil n/2 \rceil} |\mathcal{C}(\mathbf{x})| = 2^{\lceil n/2 \rceil} \left| \mathcal{C}^{\oplus} \right| \leq 2^{\lceil n/2 \rceil} M_Z(\lfloor n/2 \rfloor, t) .$$

$\square$

Best known bounds on $M(n,t)$

t \ n	2	3	4	5	6	7	8	9	10	11	12	13
1	*2*	*4*	*6*	*8*	*16*	*26*	*44*	88(72)	176(112)	352(210†)	682*(372)	1260*(702†)
2			*4*	*8*	*10*	*16*	*22*	*32*	**64** (44)	**128** (68$^\diamond$)	**256** (88)	**512** (136$^\diamond$)
3					*8*	*16*	*18*	*32*	**64** (38)	**128** (64)	**128** (76)	**256** (128)

t \ n	14	15	16	17	18
1	**2304**(1272)	4368*(2400†)	8190*(4522)	15420*(8428)	29126*(15348)
2	**512** (176)	**1024** (312$^\diamond$)	**1792** (418$^\diamond$)	**3584** (836$^\diamond$)	**6144** (1318$^\diamond$)
3	**256** (152)	**512** (260$^\diamond$)	**1024** (304)	**2048** (520$^\diamond$)	**2048** (608)

$\star$ — [KZ13], $\dagger$ and $\diamond$ — [GYD13]

Upper bound on $R(\tau)$

- *Asymmetric distance* between $\boldsymbol{c} = (c_i)_{i \in \langle n \rangle} \in \Sigma^n$ and $\boldsymbol{c}' = (c'_i)_{i \in \langle n \rangle} \in \Sigma^n$

$$\Delta(\boldsymbol{c}, \boldsymbol{c}') \triangleq \max \{\Delta^\star(\boldsymbol{c}, \boldsymbol{c}'), \Delta^\star(\boldsymbol{c}', \boldsymbol{c})\} \ ,$$

  where

$$\Delta^\star(\boldsymbol{c}, \boldsymbol{c}') = |\{i \in \langle n \rangle : c_i = 0, c'_i = 1\}| \ .$$

- *Minimum asymmetric distance* of a code $\mathcal{C} \subseteq \Sigma^n$ as

$$\Delta(\mathcal{C}) \triangleq \min_{\boldsymbol{c}, \boldsymbol{c}' \in \mathcal{C} : \boldsymbol{c} \neq \boldsymbol{c}'} \{\Delta(\boldsymbol{c}, \boldsymbol{c}')\} \, .$$

- Hamming distance $\mathsf{d}(\boldsymbol{c}, \boldsymbol{c}')$, minimum Hamming distance $\mathsf{d}(\mathcal{C})$

Upper bound on $R(\tau)$ (cont.)

Theorem. *Let* $\tau \in [0, \frac{1}{8}]$. *Then*

$$R(\tau) \leq \rho(\tau) \triangleq \frac{1}{2}\left(1 + \min_{0 < x \leq 1-8\tau} \{b(x)\}\right), \tag{1}$$

where

$$b(x) = 1 + h(x^2) - h(x^2 + 8\tau x + 8\tau)$$

and

$$h(x) = \mathrm{H}\left(0.5(1 - \sqrt{1-x})\right) .$$

Proof. Let $\mathcal{C}$ be a code of length n and size $M_Z(n, \lceil \tau n \rceil)$ correcting $\lceil \tau n \rceil$ asymmetric errors. Then $\Delta(\mathcal{C}) \geq \lceil \tau n \rceil + 1$. There exists a constant-weight subcode $\mathcal{C}(w) \subseteq \mathcal{C}$ of size at least $|\mathcal{C}| / (n+1)$ whose codewords are of weight $w \in \langle n \rangle \setminus \{0\}$.

Upper bound on $R(\tau)$ (cont.)

For any $\boldsymbol{c}, \boldsymbol{c}' \in \mathcal{C}(w)$, we have

$$\mathsf{d}(\boldsymbol{c}, \boldsymbol{c}') = 2\Delta(\boldsymbol{c}, \boldsymbol{c}') \geq 2(\lceil \tau n \rceil + 1) \ ,$$

Thus,

$$M_Z(n, \lceil \tau n \rceil) = |\mathcal{C}| \leq (n+1)\,|\mathcal{C}(w)| \leq (n+1) M_H(n, \lceil \tau n \rceil) \ ,$$

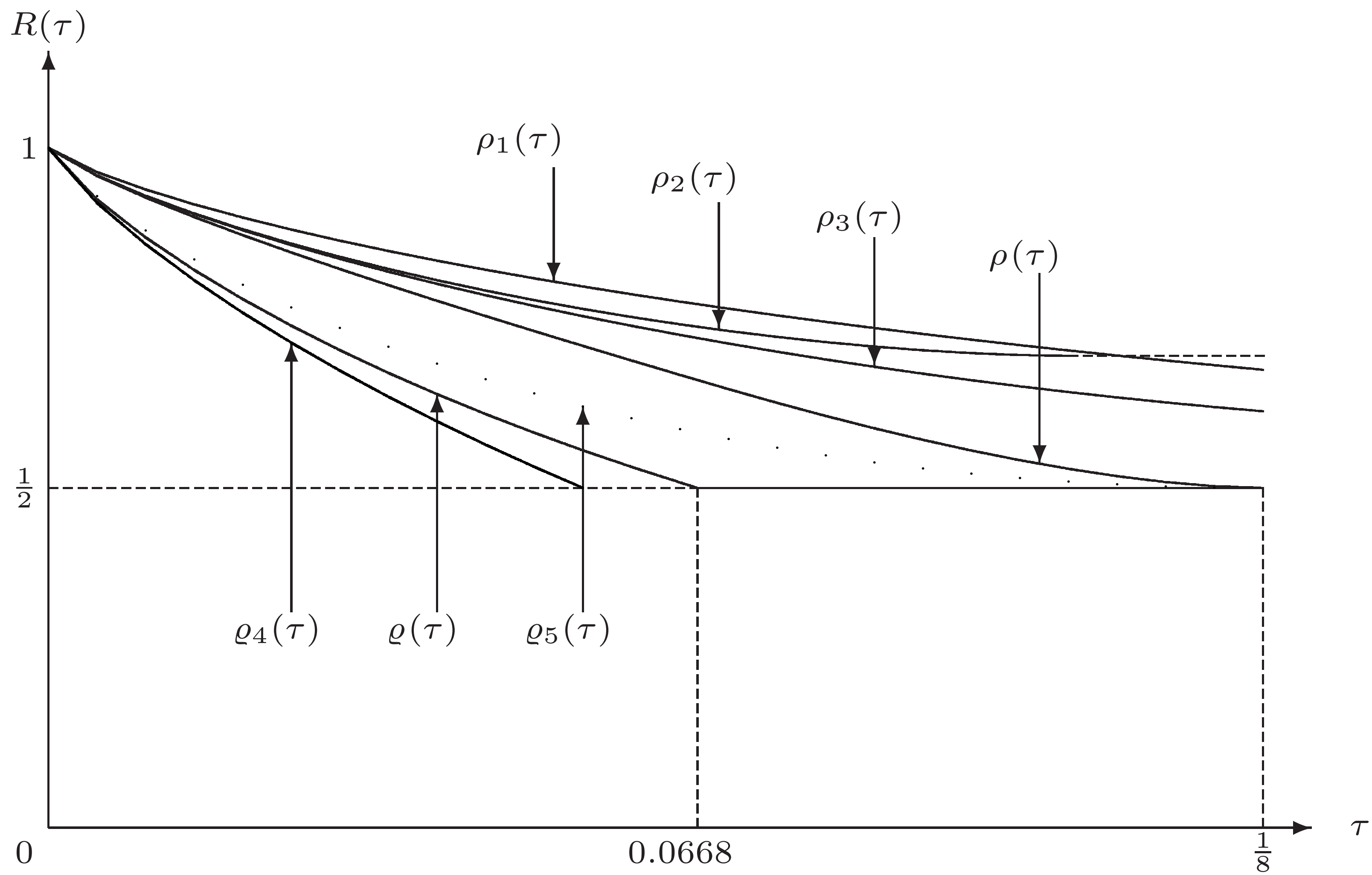
hence, by Lemma 1,

$$M(n, \lceil \tau n \rceil) \leq 2^{\lceil n/2 \rceil} \cdot \Big((\lfloor n/2 \rfloor + 1) \cdot M_H(\lfloor n/2 \rfloor, \lceil \tau n \rceil) \Big) \ ;$$

switching to asymptotics yields $R(\tau) \leq \frac{1}{2}(1+R_H(2\tau))$.

Finally, the second MRRW bound on $R_H(2\tau)$ is used. $\square$

Bounds on $R(\tau)$



Detection of all grain errors

- *Minimum redundancy* of a largest ∞-grain-detecting code of length n

$$r_n \triangleq n - \max_{\substack{\mathcal{C}\subseteq\Sigma^n \text{ is an} \\ \infty\text{-grain-detecting code}}} \{\log_2 |\mathcal{C}|\}$$

- Upper bound of $1.5\log_2 n + O\left(\frac{1}{n}\right)$ on r_n due to a construction
- Lower bound of $0.5\log_2 n + O(1)$ on r_n due to Lemma 1 (upper bound of $2^{\lceil n/2\rceil}\binom{\lfloor n/2\rfloor}{\lfloor n/4\rfloor}$ on the largest size)
- For the overlapping scenario, the upper bound on the largest size can be improved by a factor of $\sqrt{2}$

Detection of all overlapping grain errors

- Inspiration: Knuth's Christmas tree pattern
- Idea: to partition Σ^ℓ into s_ℓ totally ordered chains (subsets) $\mathsf{C}_{\ell;j}$ for $j \in \langle \mathsf{s}_\ell \rangle$ by induction on ℓ
- For $\mathbf{x}, \mathbf{y} \in \Sigma^n$, $\mathbf{x}$ is *dominated* by $\mathbf{y}$, $\mathbf{x} \preceq \mathbf{y}$, if there exists a grain pattern $\mathcal{S}$ such that $\sigma_{\mathcal{S}}(\mathbf{y}) = \mathbf{x}$
- The "biggest" and the "smallest" words in $\mathsf{C}_{\ell;j}$ denoted by $\mathsf{F}(\mathsf{C}_{\ell;j})$ and $\mathsf{f}(\mathsf{C}_{\ell;j})$
- The upper bound on the largest size will be $2\mathsf{s}_n$

Christmas tree for $\ell = 5$

10100,10101

10010,10110

10000,10001,10011,10111

11000,11001

01010,11010

01000,01001,01011,11011

01100,11100

00100,00101,01101,11101

00010,00110,01110,11110

00000,00001,00011,00111,01111,11111

Construction

Basis ($\ell = 1$). Let $\mathsf{C}_{1;0} = \{0\}$.
Step ($\ell \geq 2$). For $j \in \langle \mathsf{s}_{\ell-1} \rangle$, from a set $\mathsf{C}_{\ell-1;j}$ of size 1, we derive a new set

(C1) $\mathsf{C}_{\ell-1;j} \times \Sigma$.

From a set $\mathsf{C}_{\ell-1;j}$ of size at least 2 whose words all end with $a \in \Sigma$, we derive two new sets

(C2) $(\mathsf{C}_{\ell-1;j} \times \{\overline{a}\}) \cup \big\{\mathsf{f}(\mathsf{C}_{\ell-1;j})a\big\}$,

(C3) $(\mathsf{C}_{\ell-1;j} \times \{a\}) \setminus \big\{\mathsf{f}(\mathsf{C}_{\ell-1;j})a\big\}$,

where $\overline{a}$ denotes the *binary complement* of the symbol $a \in \Sigma$. From a set $\mathsf{C}_{\ell-1;j}$ of size at least 2 such that there exists only one word $\boldsymbol{c} \in \mathsf{C}_{\ell-1;j}$, $\boldsymbol{c} \neq \mathsf{F}(\mathsf{C}_{\ell-1;j})$, that ends with a and the rest of the words end with $\overline{a}$, we derive two new sets

(C4) $(\mathsf{C}_{\ell-1;j} \times \{a\}) \cup \{\boldsymbol{c}\overline{a}\}$,

(C5) $(\mathsf{C}_{\ell-1;j} \times \{\overline{a}\}) \setminus \{\boldsymbol{c}\overline{a}\}$.

Construction example

$$
\begin{tikzcd}[column sep=small]
\ell = 1: & & & \mathsf{C}_{1;0} = \{0\} \arrow[d, "\textcolor{red}{(C1)}"] & & \\
\ell = 2: & & & \mathsf{C}_{2;0} = \{\textcolor{red}{00}, 01\} \arrow[dl, "\textcolor{red}{(C4)}"'] \arrow[drr, "\textcolor{red}{(C5)}"] & & \\
\ell = 3: & & \mathsf{C}_{3;0} = \{000, \textcolor{red}{001}, 010\} \arrow[ddl, "\textcolor{red}{(C4)}"'] \arrow[ddrr, "\textcolor{red}{(C5)}"] & & & \mathsf{C}_{3;1} = \{011\} \arrow[d, "\textcolor{red}{(C1)}"] \\
\ell = 4: & & & & & \mathsf{C}_{4;2} = \{0111, 0110\} \\
 & \mathsf{C}_{4;0} = \{0001, 0011, \textcolor{red}{0010}, 0101\} & & & \mathsf{C}_{4;1} = \{\textcolor{red}{0000}, 0100\} \arrow[dll, "\textcolor{red}{(C2)}"'] \arrow[dr, "\textcolor{red}{(C3)}"] & \\
\ell = 5: & & \mathsf{C}_{5;2} = \{00000, 00001, 01000\} & & & \mathsf{C}_{5;3} = \{010001\}
\end{tikzcd}
$$

Total order, chains number

Lemma 2. *For any positive integer ℓ and any $j \in \langle \mathsf{s}_\ell \rangle$, the set $\mathsf{C}_{\ell;j}$ is totally ordered with respect to $\preceq$.*

- Values of s_ℓ for small values of ℓ:

ℓ	1	2	3	4	5	6	7	8	9	10
s_ℓ	1	1	2	3	6	10	20	35	70	126

- The number $\binom{\ell-1}{\lfloor(\ell-1)/2\rfloor}$ of walks of length $\ell-1$ on the square lattice from the origin $(0,0)$ by moving down or moving right, all the while staying on the points (x,y) satisfying $\mathsf{x}+\mathsf{y} \geq 0$

- Improvement of $2^{\lceil n/2\rceil}\binom{\lfloor n/2\rfloor}{\lfloor n/4\rfloor}/2\binom{n-1}{\lfloor(n-1)/2\rfloor} \approx \sqrt{2}$ over the result of Lemma 1

Largest grain-detecting codes without and with overlaps

t \ n	2	3	4	5	6	7	8
1	2	4	8	16	32	64	128
2			8	10	18	34	58
3					18	32	56
4							56

t \ n	2	3	4	5	6	7	8
1	2	4	8	16	32	64	128
2		4	6	10	18	30	52
3			6	8	12	22	42
4				8	12	20	32
5					12	20	32