Cutsize Distributions of Balanced Hypergraph Bipartitions for Random Hypergraphs

Takayuki Nozaki

Yamaguchi University

ISIT2016
2016/7/15

Outline

Previous Work (ISIT2015)

We presented a *parallel* encoding algorithm for LDPC codes

- The processing time of encoding depends on parallel degree K
- Maximum parallel degree $K_{\max}$ depends on parity check matrix $\mathbf{H}$

Aim of Research

Analyze the processing time of this encoding algorithm
$\Rightarrow$ Analyze the parallel degree $K_{\max}$ of this encoding algorithm

Main result of this work (1)

The parallel degree $K_{\max}$ depends on the minimum cutsize in balanced K-way partition for hypergraph representation of $\mathbf{H}$

However, balanced hypergraph partitioning problem is NP-hard...

Outline (2)

Solution: We take *coding theoretic approach*
(Similar technique to derive minimum distance for the LDPC ensemble)

1. considering a *random hypergraph ensemble*
2. deriving the *ensemble average* of cutsize distribution (balanced partitions with a given cutsize)
3. analyzing the growth rate for the cutsize distribution
4. clarifying the typical minimum cutsize for the hypergraph ensemble

Main result of this work (2)

Deriving the typical minimum cutsize of balanced bipartitions ($K = 2$) for random hypergraph ensemble defined from regular LDPC ensemble

Related works

- Analysis of random graphs by using coding theoritic approaches [Fujii-Wadayama2012], [Yano-Wadayama2012], [Fujii-Wadayama2013]
- Analysis of cutsize in random graph bisection [Dembo et al.2015]

Preliminaries (1: Hypergraph Representation of code)

Hypergraph $\mathcal{H} = (\mathcal{U}, \mathcal{E})$

- $\mathcal{U} := \{u_1, u_2, \dots, u_m\}$: Set of vertices
- $\mathcal{E} := \{e_1, e_2, \dots, e_n\}$: Set of nets (hyperedges)

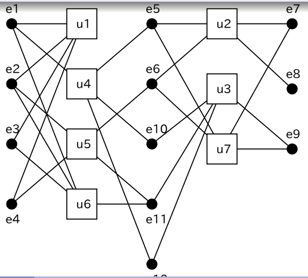
Each net connects to at least 1 vertices.

Hypergraph representation of LDPC code

i-th column $\Rightarrow$ i-th net e_i j-th row $\Rightarrow$ j-th vertex u_j

If $h_{i,j} = 1$, j-th net connects to i-th vertex

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$



Preliminaries (2: Balanced Hypergraph Partitioning)

K-way partition $\Pi_K = \{\mathcal{U}_1, \mathcal{U}_2, \ldots, \mathcal{U}_k\}$

(1) $\emptyset \neq \mathcal{U}_i \subseteq \mathcal{U}$, (2) $\mathcal{U}_i \cap \mathcal{U}_j = \emptyset$ (for $i \neq j$), (3) $\bigcup_{i=1}^{K} \mathcal{U}_i = \mathcal{U}$

A K-way partition is ϵ-balanced if

$$\max_{i=1,2,\ldots,K} |\mathcal{U}_i| \leq \frac{|\mathcal{U}|}{K}(1+\epsilon)$$

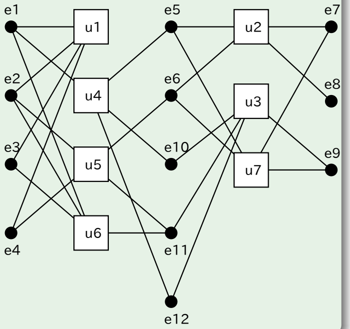
(Note) If $\epsilon = 0$, then all the parts are same size

- *Cut set* $\mathcal{X}(\Pi_K)$ is the set of vertices connecting to at least 2 parts for a partition Π_K
- *Cutsize* is the number of elements in $\mathcal{X}(\Pi_K)$

Preliminaries (3: Example of Hypergraph Partitioning)

$K = 2$
$\mathcal{U}_1 = \{u_1, u_4, u_5, u_6\}$
$\mathcal{U}_2 = \{u_2, u_3, u_7\}$

Cut set
$\mathcal{X}(\Pi_2) = \{e_5, e_6, e_{10}, e_{11}, e_{12}\}$
Cutsize : $|\mathcal{X}(\Pi_2)| = 5$



Condition for Parallel Encodable (1)

(Definition) K parallel encodable by block-diagonalization

For a given $\mathbf{H}$, an LDPC code is K *parallel encodable* if there exists a pair of permutation matrices $\mathbf{P}, \mathbf{Q}$ such that

$$\mathbf{PHQ} = \begin{pmatrix} \mathbf{H}_P & \mathbf{H}_I \end{pmatrix} = \begin{pmatrix} \mathbf{H}_{P,1} & & \mathbf{O} & \mathbf{H}_{I,1} \\ & \ddots & & \vdots \\ \mathbf{O} & & \mathbf{H}_{P,K} & \mathbf{H}_{I,K} \end{pmatrix},$$

and $\mathbf{H}_{P,i}$ is a non-singular $m_i \times m_i$ matrix for $i = 1, 2, \ldots, K$, where m_i is almost equal size ($\sum_i m_i = m$ and $\max_i m_i \leq (1+\epsilon)m/K$)

If $\mathbf{H}$ is K-parallel encodable, the parity part of codeword $\boldsymbol{p} = (\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_K)$ is parallelly solved from

$$\mathbf{H}_{P,1}\boldsymbol{p}_1^T = -\mathbf{H}_{I,1}\boldsymbol{i}^T, \qquad \cdots \qquad \mathbf{H}_{P,K}\boldsymbol{p}_K^T = -\mathbf{H}_{I,K}\boldsymbol{i}^T$$

($\boldsymbol{i}$: information part of codeword)

Condition for Parallel Encodable (2)

[Proposition 1] Necessary condition of K parallel encodable

If an LDPC code defined by $\mathbf{H}$ is K parallel encodable by block-diagonalization, the following condition holds:

$$n - m \geq \min_{\Pi_K^{(\epsilon)}} |\mathcal{X}(\Pi_K^{(\epsilon)})|.$$

- There exists the maximum parallel degree

$$K_{\max} := \max \left\{ K \mid n - m \geq \min_{\Pi_K^{(\epsilon)}} |\mathcal{X}(\Pi_K^{(\epsilon)})| \right\}$$

- Hence, processing time of encoding algorithm depends on $\min_{\Pi_K^{(\epsilon)}} |\mathcal{X}(\Pi_K^{(\epsilon)})|$
- However, It is difficult to calculate $\min_{\Pi_K^{(\epsilon)}} |\mathcal{X}(\Pi_K^{(\epsilon)})|$ (since balanced hypergraph partition problem is NP-hard)

Cutsize distribution (1: Hypergraph ensemble)

Hypergraph ensemble derived from $\mathrm{E}(n, \gamma, \delta)$

1. Define regular LDPC ensemble $\mathrm{E}(n, \gamma, \delta)$
   - n: codelength
   - γ: degree of variable node
   - δ: degree of check node
2. Convert Tanner graph to Hypergraph
   - variable node $\rightarrow$ net
   - check node $\rightarrow$ vertex

Cutsize distribution (2: Definition)

(Definition) Cutsize distribution

- $A_{\mathcal{H}}(s, m_1)$: the number of bipartitions s.t. $|\mathcal{X}(\Pi_2)| = s$, $|\mathcal{U}_1| = m_1$ and $|\mathcal{U}_2| = m_2 = m - m_1$ for a hypergraph $\mathcal{H}$
- $A(s, m_1)$: ensemble average of $A_{\mathcal{H}}(s, m_1)$

$$A(s, m_1) := \mathbb{E}_{\mathcal{H} \in \mathrm{E}(n,\gamma,\delta)}[A_{\mathcal{H}}(s, m_1)] = \frac{1}{\xi!} \sum_{\mathcal{H} \in \mathrm{E}(n,\gamma,\delta)} A_{\mathcal{H}}(s, m_1).$$

- $B_{\mathcal{H}}(s, \epsilon)$: the number of ϵ-balanced bipartitions with cutsize s for a hypergraph $\mathcal{H}$
- $B(s, \epsilon)$: ensemble average of $B_{\mathcal{H}}(s, \epsilon)$

$$B(s, \epsilon) := \mathbb{E}_{\mathcal{H} \in \mathrm{E}(n,\gamma,\delta)}[B_{\mathcal{H}}(s, \epsilon)] = \sum_{m_1 \in M_\epsilon} A(s, m_1),$$

where $M_\epsilon := [\![m(1-\epsilon)/2, m(1+\epsilon)/2]\!]$.

Cutsize distribution (3: Theorem)

(Theorem 1) Cutsize distribution

For an ensemble $\mathrm{E}(n, \gamma, \delta)$, the cutsize distribution $A(s, m_1)$ is given as follows:

$$A(s, m_1) = \frac{\binom{m}{m_1}\binom{n}{s}}{\binom{\delta m}{\delta m_1}} \mathrm{Coef}(f(u)^n, u^{\delta m_1}) \times \mathbb{I}[s \le \delta m_1]\mathbb{I}[s \le \delta(m - m_1)],$$

$$f(u) := p(u)^{s/n} q(u)^{1-s/n},$$

$$p(u) := (1+u)^\gamma - 1 - u^\gamma, \quad q(u) := 1 + u^\gamma.$$

where $\mathrm{Coef}(f(x), x^i)$ is the coefficient of x^i in the polynomial $f(x)$

Typical Minimum Cutsize (1: Definitions)

(Definition) Growth rate

Define the growth rate $g(\sigma, \mu_1)$ and $h(\sigma, \epsilon)$ for the cutsize distributions $A(\sigma n, \mu_1 m)$ and $B(\sigma n, \epsilon)$ as

$$g(\sigma, \mu_1) = \lim_{n \to \infty} \frac{1}{n} \log A(\sigma n, \mu_1 m),$$
$$h(\sigma, \epsilon) = \lim_{n \to \infty} \frac{1}{n} \log B(\sigma n, \epsilon),$$

Remark

If $h(\sigma, \epsilon) < 0$, then $B(\sigma n, \epsilon)$ is exponentially decreasing
If $h(\sigma, \epsilon) > 0$, then $B(\sigma n, \epsilon)$ is exponentially increasing

Typical Minimum Cutsize (2: Definition)

(Definition) Typical minimum cutsize

Define

$$\alpha^*(\mu_1) := \inf\{\sigma > 0 \mid g(\sigma, \mu_1) > 0\},$$
$$\beta^*(\epsilon) := \inf\{\sigma > 0 \mid h(\sigma, \epsilon) > 0\}.$$

We refer the value $\alpha^*(\mu_1)$ and $\beta^*(\epsilon)$ as the *relative typical minimum cutsizes* for $\mathrm{E}(n, \gamma, \delta)$.

[Proposition2] Necessary condition of 2 parallel encodable

If a code $\mathbf{H} \in \mathrm{E}(n, \gamma, \delta)$ is 2 parallel encodable by the block-diagonalization with high probability, the following condition holds:

$$1 - \frac{\gamma}{\delta} \geq \beta^*(\epsilon).$$

Typical Minimum Cutsize (3: Growth rate)

[Theorem 2] Growth rate

$$g(\sigma, \mu_1) = H_2(\sigma) - \gamma \frac{\delta - 1}{\delta} H_2(\mu_1) + \inf_{u>0} \{\sigma \log p(u) + (1-\sigma) \log q(u) - \mu_1 \gamma \log u\}.$$

A point u achieving the infimum satisfies

$$\sigma u p'(u) q(u) + (1-\sigma) u p(u) q'(u) = \mu_1 \gamma p(u) q(u),$$

where $p'(u) := \frac{dp}{du}$.

$$h(\sigma, \epsilon) = \max_{\mu_1 \in \bar{M}_\epsilon} g(\sigma, \mu_1).$$

Typical Minimum Cutsize (3: Property of growth rate)

[Proposition 3] Existence of typical minimum cutsize

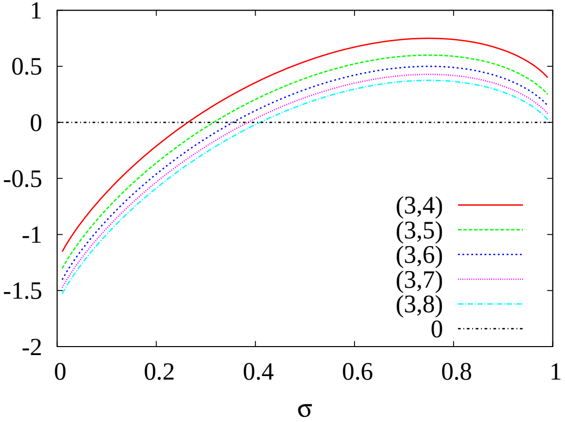
- For a fixed μ_1, there exist σ_0 such that $g(\sigma_0, \mu_1) = 0$
- For a fixed ϵ, there exist σ_0 such that $h(\sigma_0, \epsilon) = 0$

[Lemma 1] Closed form lower bound

$$h(\sigma, \epsilon) > h(\sigma, 0) = g(\sigma, 1/2).$$

$$g(\sigma, 1/2) = H_2(\sigma) + \sigma \log(2^{\gamma-1} - 1) - \gamma \frac{\delta - 1}{\delta} + 1.$$

Typical Minimum Cutsize (4: Numerical Example)



Growth rate $h(\sigma, 0)$ for hypergraph ensemble derived from $(3, \delta)$-regular LDPC ensemble.

Condition for Parallel Encodable

(Recall) Necessary condition of 2 parallel encodable

If a code $\mathbf{H} \in \mathrm{E}(n, \gamma, \delta)$ is $K = 2$ parallel encodable by the block-diagonalization with high probability, the following condition holds:

$$1 - \frac{\gamma}{\delta} \geq \beta^*(\epsilon). \tag{1}$$

Table: The left and right hand sides of (1) for $\gamma = 3$

δ	4	5	6	7	8	9
$1 - \gamma/\delta$	0.2500	0.4000	0.5000	0.5714	0.6250	0.6667
$\beta^*(0)$	0.2636	0.3157	0.3545	0.3849	0.4094	0.4297

The ensemble $\mathrm{E}(n, 3, \delta)$ for $\delta \geq 5$ satisfies the necessary condition of parallel encodable.

Conclusion and Future Works

Conclusion

- We give a necessary condition for K parallel encodable
- We derive the cutsize distribution for hypergraph ensemble
- We give the growth rate of hypergraph ensemble
- We give the typical minimum cutsize for hypergraph ensemble

Future works

- Consider $K \geq 3$
- Irregular LDPC ensemble case