On the Capacity of Flash Memory Channels with Inter-Cell Interference

Yonglong Li[1] Paul H. Siegel[2] Guangyue Han[3]

ECE, NUS[1]

ECE and CMRR, UCSD[2]

Department of Mathematics, HKU[3]

July 9, 2019

Outline

- Introduction
- Channel Model and Markov Approximation
- Mutual Information Rate
- Asymptotics of Channel Capacity
- Conclusion

Introduction

- NAND flash memories have beed used widely in real-life applications such as storage devices for computers and cellphones.
- Flash memories have been more vulnerable to various device or circuit level noises due to the rapidly growing density. Various fault-tolerance techniques such as error correction coding and constrained coding have been proposed to overcome the inter-cell interference (ICI).
- The information-theoretic capacity limits of NAND flash memory have been estimated by analysis of communication channel models; Representative work includes Dong et al. 2011 and 2012, Cai et al. 2013, Li et al. 2014, Taranalli et al. 2015.

Introduction

Motivation

- It has been observed by Siegel and his group members that a "101" in the programming sequence of NAND flash memory is more prone to become a "111" due to inter-cell interference while other patterns are more likely to be programmed correctly.
- In 2014 Qin and Siegel proposed constrained codes to mitigate ICI.
- To increase the rate of constrained codes, in 2018 Buzaglo and Siegel proposed the weakly constrained codes to mitigate ICI.

In this work, we model this phenomenon by a communication channel and derive bounds on the channel capacity.

Channel Model

Consider a flash memory channel with inter-cell interference suggested by Buzaglo and Siegel, whose channel law is given by

$$P(Y_1^n = y_1^n | X_0^{n+1} = x_0^{n+1}) = \prod_{i=1}^{n} P(Y_i = y_i | X_{i-1}^{i+1} = x_{i-1}^{i+1}),$$

where $P(Y_i = 0 | X_{i-1}^{i+1} = x_{i-1}^{i+1}) = \begin{cases} 1-\varepsilon & x_{i-1}^{i+1} = 101 \\ 0 & x_i = 1 \\ 1 & \text{otherwise} \end{cases}$.

Channel Model

Remarks

- When $\varepsilon = 1$, any "101" in the input sequence will be converted to "111" in the output sequence and "101" is not decodable. We can regard this channel as an encoder for the constrained system $\mathcal{S}$ with 101 as the forbidden word. The capacity of the channel model is 0.8114.
- When $\varepsilon = 0$, Y_i is always equal to X_i. The channel is noiseless with capacity 1.
- "0" is an unambiguous symbol, that is, $Y_i = 0$ implies that $X_i = 0$ and $X_{i-1} = y_{i-1}$ and $X_{i+1} = y_{i+1}$.

Markov Approximation

For channels with memory or states, there is no closed-form characterization of channel capacity. A natural idea is using the so-called *Markov approximation* scheme to compute $C_{Markov}^{(m)}$ and its capacity achieving distribution.

Markov Approximation

For channels with memory or states, there is no closed-form characterization of channel capacity. A natural idea is using the so-called *Markov approximation* scheme to compute $C^{(m)}_{Markov}$ and its capacity achieving distribution.

Theorem

For an indecomposable finite-state channel,

$$C = C_{Shannon} = \lim_{m \to \infty} C^{(m)}_{Markov},$$

where

$$C^{(m)}_{Markov} = \sup I(X; Y),$$

and the supremum is taken over all m-th order stationary Markov chains X.

Mutual Information Rate

Theorem

For $m \geq 2$, let X be an m-th order stationary and ergodic Markov chain and Y be the corresponding output. Then

$$I(X;Y) = -p_{X_{-2}^{0}}(101)h(\varepsilon)$$

$$+\sum_{i=-\infty}^{0}\sum_{y_i^{-1}\in\mathcal{A}_i} p_{Y_{i-m}^{-1}}(0\cdots 0y_i^{-1})h(p_{Y_0|Y_i^{-1},X_{i-m}^{i-1}}(0|0\cdots 0y_i^{-1})), \qquad (1)$$

where $\mathcal{A}_i$ is the set of y_i^{-1} such that $y_i = 1$ and y_i^{-1} contains no m-consecutive 0's.

Mutual Information Rate

Proof

Lemma

For $m \geq 2$, let X be an m-th order stationary Markov chain and Y be the corresponding output. For $-n+m \leq i \leq 0$, let $y_{i-m}^{i-1} = 0 \cdots 0$, then $p(y_0|y_{-n}^{-1}) = p_{Y_0|X_{i-m}^{i-1} Y_i^{-1}}(y_0|0 \cdots 0 y_i^{-1})$.

Then

$$H(Y_0|Y_{-n}^{-1}) = -\sum_{y_{-n}^{0}} p(y_{-n}^{0}) \log p(y_0|y_{-n}^{-1}) = T1 + T2,$$

where $$T1 = -\sum_{i=-n+m}^{0} \sum_{y_{-n}^{0} \in \mathcal{A}_{i,n}} p(y_{-n}^{0}) \log p_{Y_0|Y_i^{-1}, X_{i-m}^{i-1}}(y_0|0 \cdots 0 y_i^{-1}),$$

$$T2 = -\sum_{y_{-n}^{0} \in \mathcal{B}_n} p(y_{-n}^{0}) \log p(y_0|y_{-n}^{-1})$$

and $\mathcal{B}_n$ is the set of y_{-n}^{0} such that y_{-n}^{-1} contains no m-consecutive 0's.

Mutual Information Rate

Proof

To complete the proof, it suffices to show that $T2 \to 0$ as $n \to \infty$.

$$
\begin{aligned}
T2 = & - \sum_{y_{-n}^{0} \in \mathcal{B}_n} p(y_{-n}^{0}) \log p(y_0 | y_{-n}^{-1}) \\
= & \sum_{y_{-n}^{-1} \in \mathcal{B}_n} p(y_{-n}^{-1}) h(p_{Y_0 | Y_{-n}^{-1}}(0 | y_{-n}^{-1})) \\
\leq & \sum_{y_{-n}^{0} \in \mathcal{B}_n} p(y_{-n}^{-1}) \\
= & P(\mathcal{B}_n) \stackrel{(a)}{\to} 0, \qquad (2)
\end{aligned}
$$

where (a) follows from the Poincare's recurrence theorem.

Asymptotics of Channel Capacity

Theorem

When ε is sufficiently small, we have that

$$C(\varepsilon) = 1 + \frac{\varepsilon \log \varepsilon}{8} + O(\varepsilon). \tag{3}$$

Proof

Lower bound: Using the previous theorem with X being i.i.d. with $P(X=0)=1/2$, we have that

$$C(\varepsilon) \geq 1 + \frac{\varepsilon \log \varepsilon}{8} + (\frac{3}{4} - \frac{\log 2e}{8})\varepsilon + O(\varepsilon^2).$$

Asymptotics of Channel Capacity

Proof

Upper bound:

$$
\begin{aligned}
C(\varepsilon) &= \sup_{X:\ \text{stationary}} I(X;Y) \\
&= \sup_{X} H(Y) - p_{X_{-2}^{0}}(101)h(\varepsilon) \\
&\leq \sup_{X} H(Y_0|Y_{-n}^{-1}) - p_{X_{-2}^{0}}(101)h(\varepsilon) \\
&\leq \sup_{X_{-n-1}^{1}} H(Y_0|Y_{-n}^{-1}) - p_{X_{-2}^{0}}(101)h(\varepsilon) \triangleq C_n(\varepsilon),
\end{aligned}
$$

where the supremum is taken over probability mass functions in a small neighbourhood of $p_{X_{-n-1}^{1}}(x_{-n-1}^{1}) = 1/2^{n+3}$ for small ε.

Asymptotics of Channel Capacity

Proof

We now expand $H(Y_0|Y_{-n}^{-1}) - p_{X_{-2}^0}(101)h(\varepsilon)$ around $p_{X_{-n-1}^1}(x_{-n-1}^1) = 1/2^{n+3}$ to obtain

$$
\begin{aligned}
& H(Y_0|Y_{-n}^{-1}) - p_{X_{-2}^0}(101)h(\varepsilon) \\
& = H(X_0|X_{-n}^{-1}) + p_{X_{-2}^0}(101)\varepsilon\log\varepsilon + O(\varepsilon) \\
& = 1 + \frac{1}{8}\varepsilon\log\varepsilon + \frac{1}{2}\mathbf{q}_n^T\mathbf{H}\mathbf{q}_n + (p_{X_{-2}^0}(101) - \frac{1}{8})\varepsilon\log\varepsilon + o(|\mathbf{q}_n|^2),
\end{aligned}
$$

where $\mathbf{H}$ is the Hessian of $H(X_0|X_{-n}^{-1})$ and $\mathbf{q}$ is the difference of probability mass function of X_{-n-1}^1 and $n+3$ dimensional vector $1/2^{n+3}\mathbf{1}$. Together with the concavity of $H(X_0|X_{-n}^{-1})$ with respect to $p_{X_{-n-1}^1}$, we can show that $C_n(\varepsilon) \leq 1 + \frac{\varepsilon\log\varepsilon}{8} + O(\varepsilon)$.

Numerical Simulation

Upper Bound

The capacity is upper bounded by

$$C \leq C_{stationary} \leq \sup_{p_{X_{-3}^{1}}(\cdot)} H(Y_0 | Y_{-2}^{-1}) - p_{X_{-2}^{0}}(101) h(\varepsilon),$$

where the supremum is taken over all probability mass functions $p_{X_{-3}^{1}}(\cdot)$ such that $p_{X_{-3}^{0}}(\cdot) = p_{X_{-2}^{1}}(\cdot)$.

Numerical Simulation

Upper Bound

The capacity is upper bounded by

$$C \leq C_{stationary} \leq \sup_{p_{X_{-3}^{1}}(\cdot)} H(Y_0|Y_{-2}^{-1}) - p_{X_{-2}^{0}}(101)h(\varepsilon),$$

where the supremum is taken over all probability mass functions $p_{X_{-3}^{1}}(\cdot)$ such that $p_{X_{-3}^{0}}(\cdot) = p_{X_{-2}^{1}}(\cdot)$.

Remark

Without the stationarity condition this upper bound is equal to 1 *achieved by* $p_{X_{-3}^{1}}(00000) = p_{X_{-3}^{1}}(00010) = 1/2$.

Numerical Simulation

Upper Bound

The capacity is upper bounded by

$$C \le C_{stationary} \le \sup_{p_{X_{-3}^{1}}(\cdot)} H(Y_0|Y_{-2}^{-1}) - p_{X_{-2}^{0}}(101)h(\varepsilon),$$

where the supremum is taken over all probability mass functions $p_{X_{-3}^{1}}(\cdot)$ such that $p_{X_{-3}^{0}}(\cdot) = p_{X_{-2}^{1}}(\cdot)$.

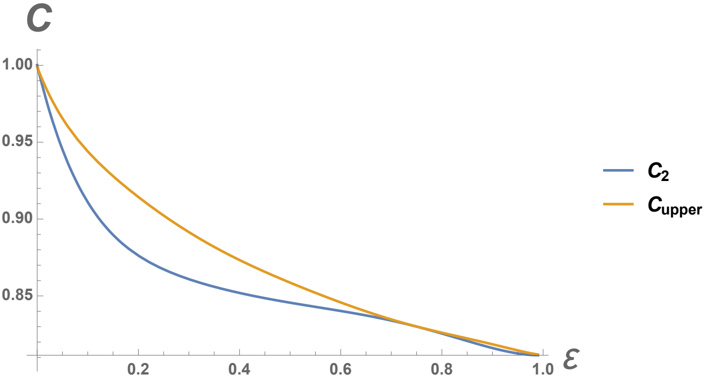
Remark

Without the stationarity condition this upper bound is equal to 1 *achieved by* $p_{X_{-3}^{1}}(00000) = p_{X_{-3}^{1}}(00010) = 1/2$.

Lower Bound

The lower bound of the channel capacity is derived by optimizing (1) for $m = 2$.

Numerical Simulation



Conclusion

Conclusion

For a flash memory channel with inter-cell interference with Markovian inputs, we derive a rather 'explicit' formula for the mutual information rate and then apply this formula to derive the asymptotics of channel capacity in the high SNR regime. We also numerically compute the lower and upper bounds on the channel capacity.

Thanks for Your Attention!