Optimal Data Shaping Code Design

Yi Liu*, **Pengfei Huang***, **Alexander W. Bergman*** and **Paul H. Siegel***
*Electrical and Computer Engineering Dept., University of California, San Diego, La Jolla, CA 92093 U.S.A
{*yil333, pehuang, awbergma, psiegel*}@*ucsd.edu*

I. Introduction

Data shaping codes are designed to increase the lifetime of flash memory devices. In [1], [2] a general class of shaping codes was defined, subsuming the definitions of endurance codes and direct shaping codes [3], [4]. The minimum average cost per code symbol for a given expansion factor was characterized, and a separation theorem was proved, showing that optimal data shaping can be achieved by the concatenation of optimal lossless compression with optimal endurance coding. The expansion factor that minimizes the total cost , i.e., the average cost per input symbol, was also analyzed. Lifetime gains from the application of an MLC direct shaping code were demonstrated experimentally. In this paper, we address the construction of optimal shaping codes. Experimental results show a significant increase in device lifetime when the resulting codes are applied to English-language and Chinese-language works, providing total data capacity greater than that achieved by data compression alone.

II. Optimal Shaping Codes

In a flash memory device, programming causes cell wear, resulting in performance degradation. Wear is data dependent, with higher voltage levels causing more wear [1] , [4]. These observations suggest a costly-channel model for the flash memory, as follows. Let $\mathbf{X} = X_1X_2\ldots$ be an i.i.d source over a finite alphabet $\mathcal{X}$, where $X_i \sim X$ for all i,. Let $\mathcal{X}^*$ be the set of all finite sequences over $\mathcal{X}$, and denote the probability of any finite sequence $x^* \in \mathcal{X}^*$ by $P(x^*)$. Let $\mathcal{Y} = \{\beta_1, \ldots, \beta_v\}$ be a finite code alphabet. Each code symbol β_i is associated with a wear cost c_i. Without loss of generality, we assume that $c_1 \leqslant c_2 \leqslant \ldots \leqslant c_v$. We use a cost vector $\mathcal{C} = [c_1, c_2, \ldots, c_v]$ to represent the wear cost associated with alphabet $\mathcal{Y}$.

In [1], [2] we defined a *shaping code* as a prefix-free mapping $\phi : \mathcal{X}^q \to \mathcal{Y}^*$ which maps a length-q data sequence x_1^q to a variable length code sequence y^*. We use $\mathbf{Y}$ to denote the process $\phi(\mathbf{X}^q)$, where $\mathbf{X}^q$ is the vector process $X_1^q, X_{q+1}^{2q}, \ldots$. This subsumes the definitions of endurance codes and direct shaping codes [3], [4].

We denote the length of a codeword $\phi(x_1^q)$ by $L(\phi(x_1^q))$ and the expected length of codewords generated by length-q source sequences is given by

$$E(L) = \sum_{x_1^q \in \mathcal{X}^q} P(x_1^q) L(\phi(x_1^q)). \quad (1)$$

The *expansion factor* is defined as the ratio of the expected codeword length to the length of the input word, namely $f = E[L]/q$. Let $\hat{P} = \{\hat{p}_i\}$ be the distribution of asymptotic code symbol probabilities induced by the code. The *average cost* per code symbol is defined by $\sum_{i=1}^{v} \hat{p}_i c_i$. The *total cost*, or average cost per input symbol, is defined by $f \sum_{i=1}^{v} \hat{p}_i c_i$.

The following three theorems are from [1], [2]. The first two characterize the optimal shaping code for a given expansion factor, as well as the expansion factor that minimizes the total cost. The third presents a *separation theorem* which shows that optimal data shaping can be achieved by the concatenation of optimal lossless compression with optimal endurance coding.

Theorem 1. *Given a source distribution P and a cost vector $\mathcal{C}$, the average cost of a shaping code $\phi : \mathcal{X}^q \to \mathcal{Y}^*$ with expansion factor f is lower bounded by $\sum_i p'_i c_i$, with $p'_i = \frac{1}{N} 2^{-\mu c_i}$, where N is a normalization constant such that $\sum_i p'_i = 1$ and μ is a positive constant such that $\sum_i -p'_i \log p'_i = H(\mathbf{X})/f$.* □

Theorem 2. *Let P be a source distribution and let $\mathcal{C}$ be a cost vector. If $c_1 \neq 0$, then the minimum total cost of a shaping code $\phi : \mathcal{X}^q \to \mathcal{Y}^*$ is given by $f \sum_i p'_i c_i$, where $p'_i = 2^{-\mu c_i}$, μ is a positive constant such that $\sum_i 2^{-\mu c_i} = 1$, and the expansion factor f is*

$$f = \frac{H(\mathbf{X})}{-\sum_i p'_i \log_2 p'_i}. \quad (2)$$

If $c_1 = 0$, then the total cost is a decreasing function of f. □

Theorem 3. *For a given expansion factor f, the minimum average cost can be achieved by a concatenation of an optimal lossless compression code with an optimal endurance code.* □

The following theorem shows an equivalence between shaping codes that minimize average cost and those that minimize total cost.

Theorem 4. *Let P be the source distribution. A shaping code that minimizes total cost with cost vector $\mathcal{C}' = \{c'_i\}$ minimizes the average cost with cost vector $\mathcal{C}$ and expansion f if*

$$c'_i = -\log_2 p'_i, \quad (3)$$

where $P' = \{p'_i\}$ is the optimal probability distribution given in Theorem 1.

III. Shaping Code Design

Motivated by the separation theorem (Theorem 3) and the equivalence theorem (Theorem 4), we propose the following code design procedure in Algorithm 1 to achieve optimal data shaping for expansion factor f.

Theorem 2 implies that the minimum total cost achieved by a shaping code for a uniform random source and costs $c'_i = -\log_2 p'_i$ as defined in the algorithm above is $\log_2 |\mathcal{X}|$. For a uniform random source and given cost vector, Varn codes [5] are tree-based, fixed-to-variable length codes that minimize

Algorithm 1 Optimal Shaping Procedure

Input: Source $\mathbf{X}$, cost vector $\mathcal{C}$, expansion factor f

1: Set $f' = f \log_2 |\mathcal{X}|/H(\mathbf{X})$.
2: Calculate symbol probability distribution $P' = \{p'_i\}$ minimizing average cost for a uniform random source and expansion factor f' using Theorem 1.
3: Define a cost vector $\mathcal{C}' = \{c'_i\}$ by $c'_i = -\log_2 p'_i$.
4: Design a shaping code minimizing total cost for a uniform random source and cost vector $\mathcal{C}'$.
5: Concatenate an optimal lossless compression code with the shaping code designed in the preceding step.

total cost for a specified codebook size K. In [6], it is shown that the average *codeword* cost $A(K)$ for the Varn code with codebook size K and costs $c_i = -\log_2 p_i$, where $\sum_i p_i = 1$, is bounded by

$$\log_2 K \leqslant A(K) \leqslant \log_2 K + \max_i c_i. \tag{4}$$

Dividing by $\log_{|\mathcal{X}|} K$, we see that the total cost $T(K)$ of the Varn code with codebook size K is bounded by

$$\log_2 |\mathcal{X}| \leqslant T(K) \leqslant \log_2 |\mathcal{X}| + \frac{\max_i c_i}{\log_{|\mathcal{X}|} K} \tag{5}$$

so $\lim_{K\to\infty} T(K) = \log_2 |\mathcal{X}|$. It follows that Varn codes are asymptotically optimal when used in our construction.

IV. Shaping Code Design and Experiment Results

We considered a shaping code design motivated by the optimal shaping scheme described above. In order to compare the performance with the MLC direct shaping code presented in [4], we used a target expansion factor $f = 1$ and an empirical cost vector $\mathcal{C} = [0, 0.58, 0.87, 1.29]$. The source file was the ASCII representation of the English-language novel *The Count of Monte Cristo*. We first compressed the source using *gzip*, a standard compression tool based upon the classic LZ77 algorithm. The observed compression rate was $g = 1/2.740$. We applied Theorem 1 to compute the the target symbol probabilities of a shaping code that minimizes average cost for a uniform random source, cost vector $\mathcal{C}$, and expansion factor $f' = f/g = 2.740$. The resulting distribution is given by

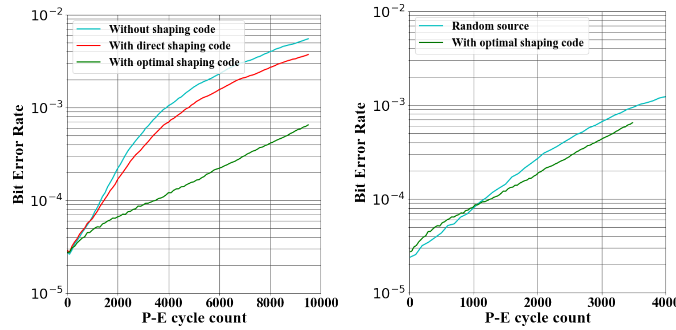
$$P' = [p'_0, p'_1, p'_2, p'_3] = [0.8604, 0.0989, 0.0335, 0.0072].$$

Using Theorem 4, we computed the costs for the equivalent code that minimizes total cost, yielding the cost vector

$$\mathcal{C}' = [c'_0, c'_1, c'_2, c'_3] = [0.2169, 3.3378, 4.8983, 7.1585].$$

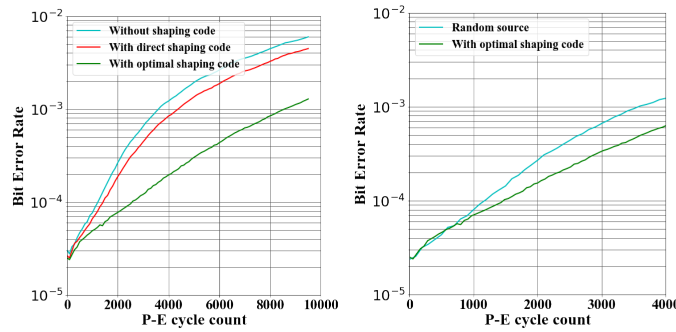
We then constructed a Varn code for the cost vector $\mathcal{C}'$ and codebook size $K = 256$.

Fig.1(a) shows the average bit error rates (BERs) for the uncoded source data, the direct shaping code [4], and the optimal shaping code. The results indicate that the optimal shaping code provides a significant increase in the memory lifetime compared to no shaping and direct shaping. As a way



(a) (b)

Fig. 1: BER Performance of English-language novel



(a) (b)

Fig. 2: BER Performance of Chinese-language novel

of comparing the performance of optimal shaping to that of data compression alone, we rescaled the P/E cycle count of the shaping code by the compression ratio 2.740 and compared the result to P/E cycling of pseudo-random data. This corresponds to a BER comparison based upon the total amount of source data stored in the memory. The results, shown in Fig. 1(b), indicate that performance of optimal shaping is superior to data compression alone as a function of total data written. Corresponding results for a Chinese-language work, *Collected Works of Lu Xun, Volumes 1–4*, represented using UTF-16LE encoding, are shown in Fig. 2(a) and Fig. 2(b).

Acknowledgment

This work was supported in part by National Science Foundation (NSF) Grant CCF-1619053 and Western Digital Corporation.

References

[1] Y. Liu and P. H. Siegel, "Shaping codes for structured data," 7th Annual Non-Volatile Memories Workshop, La Jolla, CA, Mar. 6-8, 2016.

[2] Y. Liu, P. Huang, P. H. Siegel, "Performance of optimal data shaping codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Aachen, Germany, June 25-30, 2017, pp. 1003–1007.

[3] A. Jagmohan, M. Franceschini, L. A. Lastras-Montao and J. Karidis, "Adaptive endurance coding for NAND Flash," *2010 IEEE Globecom Workshops*, Miami, FL, 2010, pp. 1841–1845.

[4] Y. Liu and P. H. Siegel, "Shaping codes for structured data," in *Proc. IEEE Globecom*, Washington, D.C., Dec. 4-8, 2016, pp. 1–5.

[5] B. Varn,"Optimal variable length codes (arbitrary symbol cost and equal code word probability)," *Inform. Contr.*, vol. 19, pp. 289–301, 1971.

[6] S. A. Savari, "Some notes on Varn coding," in *IEEE Trans. Inf. Theory*, vol. 40, no. 1, pp. 181–186, Jan. 1994.