Squeezing Out the Last Shaping Gain with Optimum Enumerative Sphere Shaping for Short Block Lengths

Yizhao Chen[1], Junda Chen[2], Weihao Li[2], Mingming Zhang[2], Deming liu[1,2], Ming Tang[2]*

[1] School of Cyber Science and Engineering, Huazhong University of Science and Technology, Wuhan, 430074, China, *tangming@mail.hust.edu.cn
[2] Wuhan National Lab for Optoelectronics (WNLO) & National Engineering Laboratory for Next Generation Internet Access System, School of Optical and Electronic Information, Huazhong University of Science and Technology, Wuhan, 430074, China

Abstract *We propose a practical solution of enumerative sphere shaping to minimize the average symbol energy of the output sequences, which permits lower rate loss and a Gaussian-like distribution, squeezing out the last shaping gain at short block lengths.*

Introduction

Recent several years, considerable attention has been paid to probabilistic shaping (PS) due to its ability to achieve the shaping gain of up to 1.53 dB compared to traditional uniform modulation formats[1], as well as fine rate adaptions[2]. Probabilistic amplitude shaping (PAS) with constant composition distribution matching (CCDM) is the most popular solution for PS implementation[3,4]. However, there exist some stringent real-world constraints for efficient hardware implementation of CCDM such as the latency, throughput, implementation complexity, and power consumption. The CCDM requires output block lengths beyond approximately 500 symbols to achieve negligibly small rate loss[4]. And large block lengths lead to the drawbacks of arithmetic complexity and throughput limitation. Thus, efficient shaping techniques have recently been studied extensively for short block lengths[5-7]. Among them, enumerative sphere shaping (ESS), which map a k-bit string to an N-tuple of amplitudes satisfying a maximum-energy constraint, achieves lower rate loss than CCDM at the same shaping rate. However, ESS cannot achieve the minimum average sequence energy, because it chooses 2^k amplitude sequences lexicographically for binary transmission. To handle this problem, ESS was optimized by modifying the ESS trellis in a rather heuristic manner, resulting in an average energy very close to the minimum, but still not exactly minimum[8].

In this work, we propose a novel way for ESS to squeeze out the last shaping gain for short block lengths, called optimum ESS (OESS). By employing two ESS trellises, the sequences with low energy are preferentially used for binary transmission, permitting a minimum average energy. The energy efficiency, amplitude distribution and rate loss of OESS are theoretically analyzed and compared with ESS and CCDM. Finally, we perform a Monte Carlo simulation to investigate the rate performance of OESS in additive white Gaussian noise (AWGN) channel.

ESS and OESS

The implementation of ESS relies on the bounded-energy enumerative trellis[7]. We consider to map a k-bit string to an N-tuple of amplitudes taking value from the shaping set $A^{\bullet}$ of length N with energy no larger than $E^{\bullet}$, where $A^{\bullet}=\left\{a^N=(a_1,a_1,\cdots,a_N):e(a^N)<E^{\bullet}\right\}$ and $a_n \in A=\left\{1,3,\cdots,2^m-1\right\}$. The trellis for N = 4, m = 3 and $E^{\bullet}$ = 60 is shown in Fig. 1.

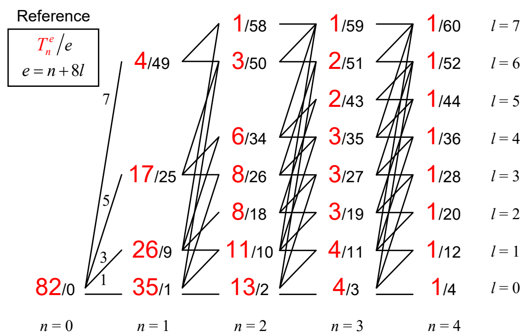

Fig. 1: Enumerative trellis for N = 4, m = 3 and $E^{\bullet}$ = 60.

The node in the trellis are addressed with (n, e), where e represents the accumulated energy of amplitude sequences over the first n dimensions for n = 0, 1, · · · ,N. These energy values are indicated in black. Branches connecting a node in column n − 1 with a node in column n represent the n^{th} components of s^N, $s^N \in \mathcal{A}$. The number written in bold&red is the number of possible ways to reach a final node starting from (n, e) which is denoted by T_n^e, which can be computed by

$$T_n^e \triangleq \sum_{a\in\mathcal{A}} T_{n+1}^{e+a^2} \tag{1}$$

978-1-6654-3868-1/21/$31.00 ©2021 IEEE

where $T_N^e = 1$ for $e \le E^\bullet$. The details about ESS can refer to Ref. [7]. The number of sequences that can be represented by the trellis is given by T_0^0, which is 79 for the trellis in Fig. 1. For binary transmission, only 2^k sequences are actually employed, which means that 64 sequences are chosen from 79 sequences. However, as ESS orders sequences lexicographically, it cannot guarantee that low-energy sequences are preferentially used[8]. As shown in Fig. 2, ESS employs 64 sequences from 79 sequences bounded by the energy of 60. Some sequences with lower energy of 36, 44 and 52 are abandoned, while 15 sequences with energy of 60 are employed, degrading the energy-efficiency of ESS.

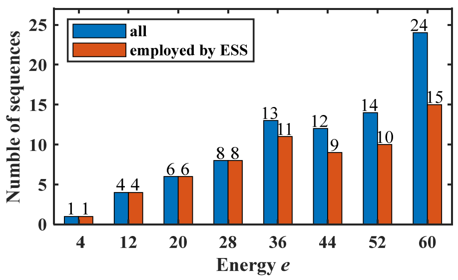

Fig. 2: The distribution of sequences employed by ESS.

To further promote the performance of ESS, we propose OESS which can achieve the minimum average symbol energy. First, we choose the minimum $E^\bullet$ for which the trellis has at least 2^k sequences, i.e., $\mathcal{A}^\bullet\left(E^\bullet\right) \ge 2^k \ge \mathcal{A}^\bullet\left(E^\bullet - 8\right)$. Here, we consider the situation of N = 4, m = 3 and $E^\bullet$ = 60, which corresponds to the Fig.1. Second, two ESS trellises are used for shaping mapping, as shown in Fig. 3. The first trellis is the normal ESS trellis but with bounded energy $E^\bullet - 8$, which has $T_0^0 < 2^k$. Thus, there are 58 sequences represented in the first trellis. The second trellis shown in Fig. 3(b) is different from normal ESS trellis, which is represented by $\hat{T}$. This trellis is established by setting $\hat{T}_N^e = 0$ for $e < E^\bullet$ and $\hat{T}_N^{E^\bullet} = 1$. In other words, the paths to the node of energy $E^\bullet$ are closed for the first trellis, and these paths consist of the second trellis. Finally, consider an input sequences, which can be converted to a decimal number I. If $I \le T_0^0$, it will be sent to first trellis for coding. Otherwise, we get $I' = I - T_0^0$, and it is sent to the second trellis for coding. In this way, the sequences with energy lower than $E^\bullet$ are fully used, permitting the minimum average symbol energy. As for the decoding processing, the energy of the input

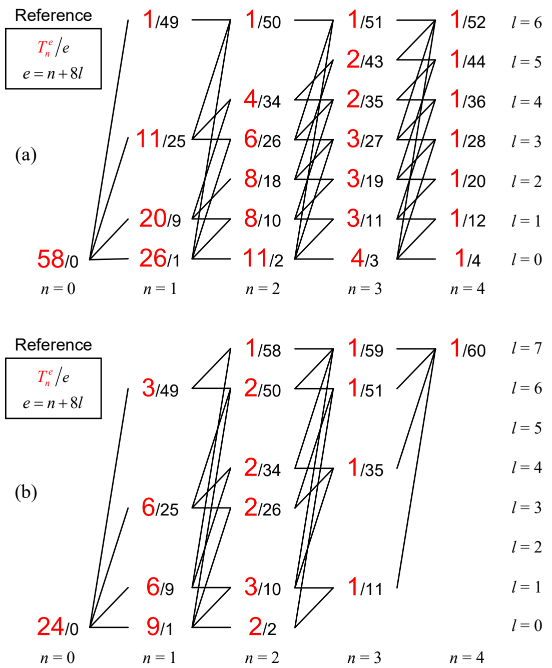

Fig. 3: Two trellises for OESS. (a) The normal ESS trellis with N = 4, m = 3 and $E^\bullet$ = 52. (b) The specific trellis for sequences with energy $E^\bullet$ = 60.

amplitude sequences is first calculated, then the corresponding trellis is chosen for decoding. The energy distribution of sequences employed by OESS is shown in Fig. 4. Consequently, OESS achieves the minimum average symbol energy of 9.6875, while ESS can only achieve the average symbol energy of 10.1875.

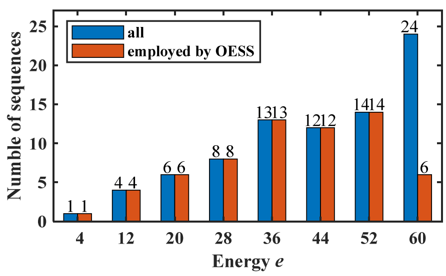

Fig. 4: The distribution of sequences employed by OESS.

Analysis

The performance of OESS is investigated in terms of operational amplitude distribution, energy efficiency and rate loss. The amplitude distribution of the sequences represented by the first trellis can be calculated by

$$p_1^\bullet(a) = [T_1^1, T_1^9, \cdots, T_1^{(2^m-1)^2}]. \qquad (2)$$

Because all sequences in the first trellis are used, the set of sequences is permutation invariant. The amplitude distribution of the sequences represented by the second trellis is more complex. The calculation processing is introduced in Ref.

[8], and the distribution is represented by $p_2^{\bullet}(a)$. Then the total distribution can be calculated by

$$p^{\bullet}(a) = p_1^{\bullet}(a) * T_0^0 + p_2^{\bullet}(a) * (2^k - T_0^0)\,. \qquad (3)$$

The operational amplitude distribution of OESS with $N = 20$ is shown in Fig. 5. We use $\Delta P = P^{\text{OESS}} - P^{\text{MB}}$ to represent the probability difference between the distribution produced by OESS and corresponding Maxwell-Boltzmann (MB) with the same average energy. The absolute values of ΔP fluctuate within 0.015 over the shaping rate range from 0.2 to 2 bits/amp, where the shaping rate is defined as k/N.

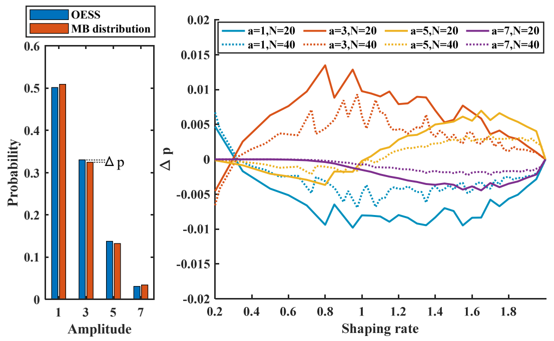

Fig. 5: Probability distribution (left) of ESS and MB distribution and their difference (right).

The energy efficiency of OESS, ESS and CCDM is shown in Fig. 6. At the shaping rate of 1.5 bits/amp, OESS achieves the average symbol energy of 8.416, which is 0.236 and 4.584 less than that of ESS and CCDM. The numerical rate loss comparison with N of 20 and 40 are shown in Fig. 7. The rate loss R_{loss} with a finite block length is defined as $R_{\text{loss}} \triangleq \mathbb{H}(X_{\text{MB}}) - R_{\text{S}}$, where X_{MB} is a MB-distributed random variable with the same average energy as the output sequence[7]. In general, over the shaping rate range from 0 to 2 bits/amp, OESS results in lower rate loss compared to ESS and CCDM. The maximum shaping gain improvement is estimated to be 0.031 and 0.271 bits/amp compared with ESS and CCDM with $N = 20$.

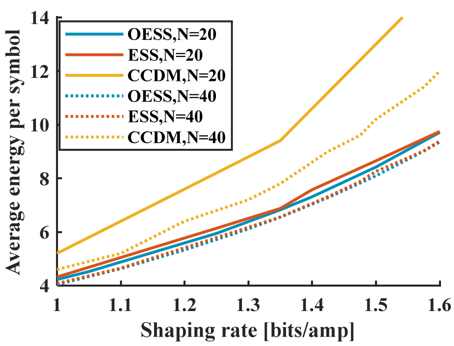

Fig. 6: Energy efficiency of OESS, ESS and CCDM.

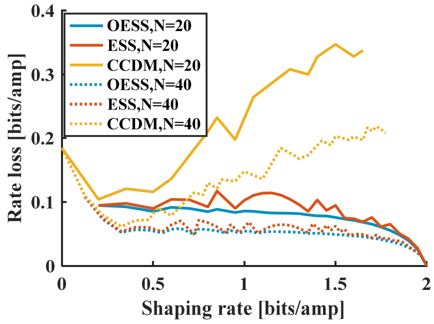

Fig. 7: Investigation on the rate loss.

Simulation

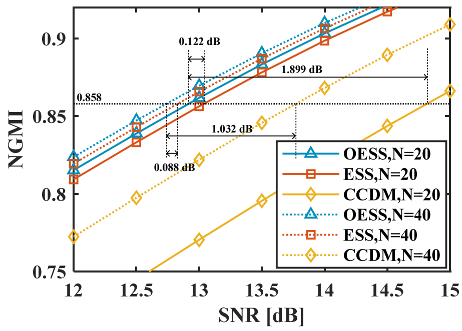

Fig. 8: NGMI versus SNR in AWGN channel.

Finally, we perform a 64-QAM Monte Carlo simulation to evaluate the actual performance of OESS in AWGN channel. The PAS architecture is utilized and the shaping rate is chosen to be 1.5 bits/amp. Error-free decoding can be verified by the normalized generalized mutual information (NGMI)[9]. At the NGMI threshold of 0.858, compared with ESS and CCDM, OESS achieves receiver sensitivity gains of 0.122 dB and 1.899 dB with N = 20, and 0.088 dB and 1.032 dB with N = 40, respectively.

Conclusions

We propose OESS which can always achieve the minimum average sequence energy, squeezing out the last shaping gain with short block lengths. By using two ESS trellises for the shaping encoding and decoding, the sequences with low energy are preferentially used for binary transmission. The operational amplitude distribution of OESS is very close to MB distribution, permitting remarkable shaping gain. Compared with ESS and CCDM, OESS achieves average symbol energy reduction of 0.236 and 4.584, and shaping gain improvement of up to 0.031 and 0.271 bits/amp. In the simulation, OESS achieves receiver sensitivity gains of 0.122 dB and 1.899 dB compared the other two methods.

Acknowledgements

National Key R&D Program of China 2018YFB1801205; NSFC Grants 61931010.

References

[1] U. Wachsmann, R. F. H. Fisher, and J. B. Huber, “Multilevel codes: Theoretical concepts and practical design rules,” IEEE Transactions on Information Theory, vol. 45, no. 5, pp. 1361–1391, Jul. 1999.

[2] F. Buchali, F. Steiner, G. Böcherer, L. Schmalen, P. Schulte, and W. Idler, “Rate adaptation and reach increase by probabilistically shaped 64-QAM: An experimental demonstration,” J. Light. Technol. 34, 1599– 1609 (2016).

[3] G. Böcherer, F. Steiner, and P. Schulte, “Bandwidth efficient and ratematched low-density parity-check coded modulation,” IEEE Trans. Commun., vol. 63, no. 12, pp. 4651–4665, Dec. 2015.

[4] P. Schulte and G. Böcherer, “Constant composition distribution matching,” IEEE Transactions on Inf. Theory 62, 430–434 (2016).

[5] T. Fehenberger, D. S. Millar, T. Koike-Akino, K. Kojima, and K. Parsons, “Multiset-partition distribution matching,” IEEE Trans. Commun., vol. 67, no. 3, pp. 1885–1893, Mar. 2019.

[6] P. Schulte and F. Steiner, “Divergence-optimal fixed-to-fixed length distribution matching with shell mapping,” IEEE Wireless Commun. Lett., vol. 8, no. 2, pp. 620–623, Apr. 2019.

[7] Y. C. Gültekin, W. J. van Houtum, A. G. C. Koppelaar, and F. M. J. Willems, “Enumerative sphere shaping for wireless communications with short packets,” IEEE Trans. Wireless Commun., vol. 19, no. 2, pp. 1098–1112, Feb. 2020.

[8] Y. C. Gültekin, W. J. van Houtum, A. G. C. Koppelaar and F. M. J. Willems, "Comparison and Optimization of Enumerative Coding Techniques for Amplitude Shaping," in IEEE Communications Letters, vol. 25, no. 4, pp. 1231-1235, April 2021, doi: 10.1109/LCOMM.2020.3044701.

[9] A. Alvarado, E. Agrell, D. Lavery, R. Maher, and P. Bayvel, “Replacing the Soft-Decision FEC Limit Paradigm in the Design of Optical Communication Systems,” J. Lightwave Technol. 33(20), 4338–4352 (2015).