

QAM 시스템의 확률 정형을 위한 산술 부호화 기반 CCDM 구현 및 성능 분석

김인기, 안현태, 박지상, 김상효*
성균관대학교

*iamshkim@skku.edu

Arithmetic Coding-Based CCDM for Probabilistic Shaping in QAM Systems: Implementation and Performance Analysis

Inki Kim, Hyun Tae Ahn, Jisang Park, Sang-Hyo Kim*
Sungkyunkwan University

Department of Electrical and Computer Engineering, Sungkyunkwan University

요 약

확률 정형은 차세대 통신 시스템의 스펙트럼 효율을 극대화하여 새넌 한계에 근접하기 위한 필수 기술이다. 본 논문에서는 기존 가변 길이 부호화의 한계를 극복하기 위해 산술 부호화 기반의 CCDM 기법을 구현하고, 블록 길이에 따른 전송 손실과 정규화된 정보 발산의 상관관계를 분석한다. 나아가, LDPC 부호를 적용한 시뮬레이션을 수행하여 고차 QAM 전송 환경에서 실질적인 성능 이득을 정량적으로 검증하였다.

I. 서 론

확률 정형(Probabilistic shaping) 기술은 균등(uniform) 확률 분포를 갖는 입력 비트열을 채널 상태에 최적화된 비균등(non-uniform) 분포를 갖는 심볼열로 변환함으로써, 새넌(Shannon) 한계에 근접하는 전송 효율을 달성하는 기법이다. 최근 5G 및 6G 통신 시스템에서는 스펙트럼 효율을 극대화하기 위해 고차 직교 진폭 변조(Quadrature Amplitude Modulation, QAM)를 필수적으로 사용하고 있으며 [2], 이러한 고차 변조 환경에서 균등 확률 분포를 사용할 경우 발생하는 셰이핑 이득(shaping gain)의 손실을 보상하기 위해 분포 매칭(distribution matching) 기술에 대한 연구가 활발히 진행되고 있다 [3-5].

전통적인 QAM 시스템은 부가 백색 가우시안 잡음(Additive White Gaussian Noise, AWGN) 채널의 용량을 달성하기 위해, 균등 분포가 아닌 AWGN 분포와 유사한 분포를 지녀야 한다 [3-5]. 하지만 데이터 전송 시 송수신 간의 엄격한 동기화가 요구되고, LDPC(Low-Density Parity-Check) 부호 [1]와 같은 오류 정정 부호와 결합되는 시스템에서는 매처(matcher)의 입출력 길이를 고정하는 것이 시스템 설계 측면에서 중요하다 [2]. 따라서 고정 길이(fixed-to-fixed length)를 가지면서도 목표 분포와의 정보 발산(informational divergence)을 최소화하는 매칭 기법이 요구된다.

Schulte 와 Böcherer 는 이러한 문제를 해결하기 위해 Constant Composition Distribution Matching (CCDM) 기법을 제안하였다 [4]. CCDM 은 산술 부호화(arithmetic coding)를 기반으로 하여 출력 시퀀스의 심볼 구성(composition)을 일정하게 유지함으로써 고정 길이 입출력을 보장하며, 거대한 코드북을 저장할 필요 없이 알고리즘적으로 매핑을 수행할 수 있는 장점을 갖는다. 특히 CCDM 은 블록 길이가 증가함에 따라 목표 분포의 엔트로피에 수렴하는 전송률을 달성하며, 출력 분포와 목표 분포 사이의 정규화된 발산(normalized divergence)이 0 으로 수렴하여 점근적으로 최적의 성능을 보장한다.

본 논문에서는 QAM 시스템의 전송 용량 증대를 위한 산술 부호화 기반의 CCDM 을 구현하고 그 성능을 분석한다.

II. 본 론

2.1. 시스템 모델 및 성능 지표 [3]

본 논문에서는 m 비트의 균등 입력 시퀀스 $B^m \in \{0,1\}^m$ 를 목표 확률 질량 함수(Probability Mass Function, PMF) P_A 를 따르는 n 심볼의 출력 시퀀스 $\bar{A}^n = (\bar{A}_1, \dots, \bar{A}_n) \in A^n$ 으로 매핑하는 시스템을 고려한다. 이때 매칭 전송률(matching rate)은 $R_{DM} = m/n$ [bits/symbol]로 정의된다. 이상적인 분포 매칭은 출력 분포 $P_{\bar{A}^n}$ 와 목표 분포 P_A^n 사이의 차이를 최소화해야 한다. 이를 측정하는 척도로 정규화된 정보 발산(normalized informational divergence)을 사용하며, 이는 (수식 1)과 같이 정의된다.

$$\frac{\mathbb{D}(P_{\bar{A}^n} \| P_A^n)}{n} = \frac{1}{n} \sum_{a^n \in \text{supp}(P_{\bar{A}^n})} P_{\bar{A}^n} \log_2 \frac{P_{\bar{A}^n}(a^n)}{P_A^n(a^n)} \quad (\text{수식 1})$$

정보이론적 관점에서, $n \rightarrow \infty$ 일 때 정규화된 발산이 0 으로 수렴하면서 가역성(invertibility)을 만족하고, 이때 매처가 달성 가능한 전송률의 상한은 (수식 2)와 같다. 이때 H_A 는 목표 분포의 엔트로피이다.

$$R_{DM} \leq H(A) \quad (\text{수식 2})$$

즉, 이상적인 매처는 블록 길이가 증가함에 따라 엔트로피에 근접하는 전송률을 만족하게 된다.

2.2. Constant Composition 부호화와 매핑 조건 [4]

고정 길이 입출력을 보장하기 위해, CCDM 은 목표 확률 분포 P_A 를 고정하고, 이에 맞춰 각 심볼의 등장 횟수를 결정하여 모든 코드워드(codeword)가 동일한 구성(constant composition)을 지니도록 하는 고정 길이 기법이다.

주어진 블록 길이 n 과 목표 확률 $P_A(a)$ 에 대해 시퀀스 내 각 심볼 a 의 등장 횟수 n_a 는 $n_a \approx n \cdot P_A(a)$ 가 되도록 결정된다. 결정된 심볼 개수 조합을 갖는 모든 시퀀스들의 집합 $T_{P_A}^n$ 을 코드북으로 사용한다. 이때 일대일 매핑을 보장하기 위한 입력 비트 길이 m 은 생성 가능한 시퀀스의 총 개수에 따라 (수식 3)과 같이 결정된다.

$$m = \lceil \log_2 |T_{P_A}^n| \rceil \quad (\text{수식 3})$$

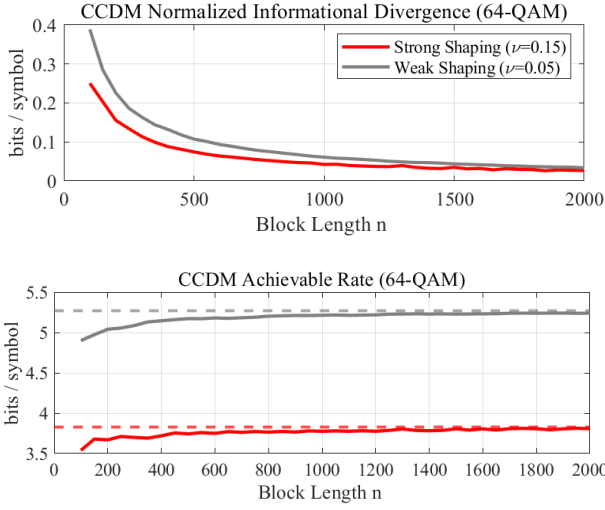


그림 1. 64-QAM 에서 CCDDM 의 블록 길이에 따른 성능 분석:
(a) 정규화 정보 발산, (b) 달성 전송률

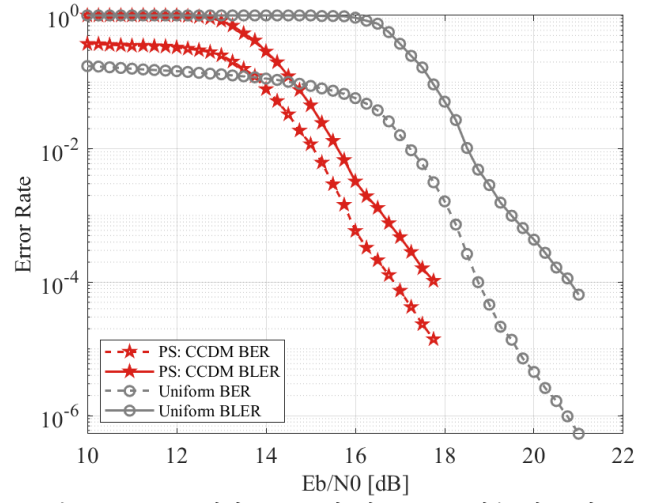


그림 2. 64-QAM 에서 CCDDM 과 기존 QAM 전송 성능 비교:
(LDPC 부호 적용)

위 조건은 입력 공간의 크기 2^m 이, 미리 결정된 심볼 개수 조합으로 만들 수 있는 순열(permutation)의 총 개수를 초과하지 않도록 제한하여 매칭 실패를 방지한다.

2.3 산술 부호화(Arithmetic Coding) 기반 구현

집합 $T_{p_A}^n$ 의 크기는 n 에 대해 지수적으로 증가하므로, 이를 직접 저장하는 대신 산술 부호화를 이용한 알고리즘 기반 매핑을 수행한다. 산술 부호화는 입력 비트열 B^m 에 대응하는 구간(interval) $[0, 1)$ 을 출력 시퀀스 집합의 누적 확률 구간으로 매핑한다. 일반적인 산술 부호화와 달리, CCDDM 은 비복원 추출(drawing without replacement) 모델을 따른다. 즉, i 번째 심볼을 생성할 때 특정 심볼 a 가 선택될 확률은 (수식 4)와 같이 동적으로 갱신된다.

$$P(\bar{A}_i = a | \bar{A}^{i-1}) = \frac{n'_a}{n'} \quad (\text{수식 4})$$

이때 n' 은 현재 남아있는 전체 심볼 수, n'_a 는 남아있는 심볼 a 의 개수를 의미한다. 이 방식은 마지막 n 번째 심볼까지 생성했을 때, 결과 시퀀스가 반드시 초기 설정된 구성 n_a 를 만족함을 보장한다.

III. 시뮬레이션 및 결과

그림 1 은 64-QAM 전송에서 CCDDM 적용 시 블록 길이에 따른 정규화 정보 발산과 달성 전송률을 나타낸다. (a)에서 블록 길이가 증가함에 따라 정규화 정보 발산은 감소하며, 강한 셰이핑($\nu=0.15$)이 약한 셰이핑($\nu=0.05$) 대비 더 작은 값을 보여 제한된 블록 길이 환경에서 우수한 분포 정합 성능을 보인다. (b)에서는 블록 길이 증가에 따라 달성 전송률이 목표 엔트로피에 근접함을 확인할 수 있으며, 강한 셰이핑의 경우 목표 엔트로피가 낮아 약한 셰이핑 대비 달성 전송률이 전반적으로 낮게 나타난다.

그림 2 는 동일한 조건에서 CCDDM 의 달성 전송률을 나타낸다. 블록 길이가 증가할수록 달성 전송률은 목표 엔트로피에 근접하며, CCDDM 이 유한 블록 길이에서도 높은 전송 효율을 달성할 수 있음을 보여준다. 강한 셰이핑의 경우 목표 분포의 엔트로피가 낮기 때문에 약한 셰이핑 대비 달성 전송률이 전반적으로 낮게 나타나며, 이는 확률 정형에 따른 자연스러운 결과이다.

IV. 결론

본 논문에서는 확률 정형을 위한 산술 부호화 기반의 CCDDM 기법을 구현하고 이를 LDPC 부호와 결합하여 그 성능을 분석하였다. 제안된 기법은 목표 확률 분포에 따라 결정된 심볼 구성을 기반으로 산술 부호화를 수행함으로써, 고정 길이 입출력을 보장하고 셰이핑 이득을 얻어 기존 QAM 전송보다 더 우수한 성능을 달성함을 확인하였다. 또한, 블록 길이가 증가함에 따라 매칭 전송률이 타겟 엔트로피에 수렴하여 전송률 손실이 최소화됨을 확인하였다. 그러나 저지연 통신을 위해 짧은 블록 길이를 사용할 경우, CCDDM 구조상 불가피한 전송률 손실이 발생하여 시스템 효율이 저하되는 트레이드오프가 존재함을 확인하였다. 향후 연구로 짧은 블록 길이에서도 전송률 손실을 최소화할 수 있는 비트단위(bit-wise) CCDDM 이나 더 큰 셰이핑 이득을 얻는 ESS(Enumerative Sphere Shaping) 기법 [5]에 대한 분석을 수행할 예정이다.

ACKNOWLEDGMENT

이 논문은 2026 년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(RS-2024-00343913) 및 정보통신기획평가원의 지원(RS-2024-00398449)을 받아 수행된 연구임.

참고 문헌

- [1] R. Gallager, "Low-density parity check codes," *IRE Trans. Inf. Theory*, vol. 8, no. 1, pp. 21-28, Jan. 1962.
- [2] 3GPP, NR; Multiplexing and channel coding, TS 38.212, v 15.2.0, Dec. 2018.
- [3] G. Böcherer, F. Steiner, and P. Schulte, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4651-4670, Dec. 2015.
- [4] P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 430-434, Jan. 2016.
- [5] W. Liu, T. Richardson, O. Shental, L. Wu, C. Xu, and H. Xu, "Energy-based arithmetic coding methods for probabilistic amplitude shaping," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2023.