

Bit-level CCDDM 기반 확률 정형 시스템에서 심볼 라벨링 방식에 따른 성능 분석

안현대, 김인기, 박지상, 김상호*
성균관대학교 전자전기컴퓨터공학과

*iamshkim@skku.edu

On the Effect of Symbol Labeling in Bit-Level CCDDM-Based Probabilistic Shaping

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

Department of Electrical and Computer Engineering, Sungkyunkwan University

요 약

본 논문은 Bit-level CCDDM 기반 확률 정형 시스템에서 심볼 매핑 방식이 성능에 미치는 영향을 분석한다. Gray 및 Natural 라벨링에 대해 목표 Maxwell-Boltzmann 분포와 유사도를 KL divergence로 평가하고, BICM 환경에서 성능을 Generalized Mutual Information을 통해 비교하였다.

I. 서 론

고차 변조를 사용하는 현대 무선 및 유선 통신 시스템에서는 제한된 전력과 대역폭 자원 하에서 전송 효율을 극대화하는 것이 핵심 과제로 대두되고 있다. Shannon 이론에 따르면 채널 용량은 입력 신호가 가우시안 분포를 따를 때 최대가 되며, 이를 실용적인 변조 시스템에 적용하기 위한 접근법으로 확률 정형(Probabilistic Shaping, PS)이 제안되었다. 확률 정형은 심볼 성상은 고정된 채, 심볼 발생 확률을 비균등하게 설계함으로써 채널 용량에 근접하는 성능을 달성하는 기술이다[1].

확률 정형 시스템에서 분포 정렬(Distribution Matching, DM)은 균등 분포를 갖는 정보 비트를 목표 심볼 분포로 변환하는 핵심 구성 요소이다. 이 중 Constant Composition Distribution Matching(CCDDM)은 출력 심볼의 구성을 고정하여 유한 길이에서 목표 심볼 확률 분포를 보장하는 DM 기법으로, PAS(Probabilistic Amplitude Shaping) 구조에서 널리 활용되고 있다[2]. 하지만 심볼 단위 CCDDM은 심볼에 사용하는 알파벳 크기, 입력에 해당하는 정보 비트의 길이에 대한 조성 설계가 필요하여 구현 복잡도와 메모리 문제가 존재한다. 이를 완화하기 위해 심볼을 구성하는 각 비트 위치별로 이진 DM을 적용하는 bit-level CCDDM이 활용된다[3].

한편, 대부분의 통신 시스템은 Bit-Interleaved Coded Modulation(BICM) 구조를 채택하여, 수신단에서는 비트 단위 로그 우도비(LLR)를 기반으로 복호를 수행한다[4]. 이러한 구조에서 동일한 bit-level CCDDM이 적용되어도, 심볼 라벨링 방식에 따라 심볼 발생 확률과 비트 신뢰도가 달라질 수 있다. 본 논문에서는 bit-level CCDDM 기반 확률 정형 시스템에서 Gray 라벨링과 Natural 라벨링이 성능에 미치는 영향을 분석한다. 라벨링에 따라 형성되는 심볼 분포와 목표 분포 사이의 유사도를 Kullback-Leibler(KL) divergence를 통해 평가하고, BICM 시스템에서 달성 가능한 정보율을 나타내는 Generalized Mutual Information(GMI)를 이용하여 라벨링 방식에 따른 성능 차이를 비교한다.

II. 본론

2.1 Bit-level CCDDM

Bit-level CCDDM은 변조 과정에서 M -ary 심볼을 구성하는 비트 위치(bit-level)별로 0과 1의 비율을 조절하여, 심볼 출현 확률을 비균일하게 만드는 분포 정렬 기법이다. 심볼당 비트 수를 $m = \log_2 M$ 이라 하고, j 번째 bit-level에서 0이 발생할 확률을 $P_j(0)$ 로 정의한다. N 개의 심볼을 전송하는 경우, j 번째 bit-level에서는 전체 N 개의 비트 중 $NP_j(0)$ 개의 0을 포함하는 이진 시퀀스만 허용되며, 이에 따라 발생 가능한 시퀀스의 개수는 $\binom{N}{NP_j(0)}$ 로 주어진다. 각 bit-level별로 이러한 조성을 만족하는 시퀀스 집합을 이용하여 분포 정렬을 수행하면, 해당 bit-level에서 전달 가능한 정보 비트의 수는 $k_j = \left\lfloor \binom{N}{NP_j(0)} \right\rfloor$ 로 결정된다. 결과적으로 전체 bit-level CCDDM을 통해 입력되는 정보 비트의 총 개수는 $k = \sum_{j=1}^m k_j$ 가 되며, 균등 분포를 갖는 정보 비트열이 bit-level별로 비균등한 확률 분포를 갖는 출력 비트열로 변환된다[2].

2.2 심볼 라벨링: Gray vs. Natural

심볼 라벨링은 각 변조 심볼에 대응되는 비트 패턴을 정의하며, 대표적으로 Gray 라벨링과 natural 라벨링이 널리 사용된다. Gray 라벨링은 인접한 심볼 간의 비트 해밍 거리가 1이 되도록 설계하여 심볼 오류가 발생해도 비트 오류 수를 최소화하는 것을 목표로 한다. 반면 natural 라벨링은 이진수 증가 순서에 따라 심볼을 매핑하는 방식으로, 구현이 단순하고 bit-level의 구조가 직관적이라는 특징을 갖는다. 이러한 라벨링 방식의 차이는 동일한 bit-level CCDDM을 사용하더라도, 각 bit-level의 확률 변화가 심볼 분포에 반영되는 방식에 차이를 만든다. 이로 인해 bit-level별 신뢰도, 비트-심볼 간 상관 구조, 그리고 BICM 기반 복호 성능에 차이가 발생하며, 확률 정형 환경에서는 라벨링 선택이 성능에 중요한 설계 자유도로 작용한다[4]. 본 논문에서는 이러한 영향을 분석하기 위해 서로 다른 라벨링 방식에 대해 분포 및 정보율 관점에서 성능을 비교한다.

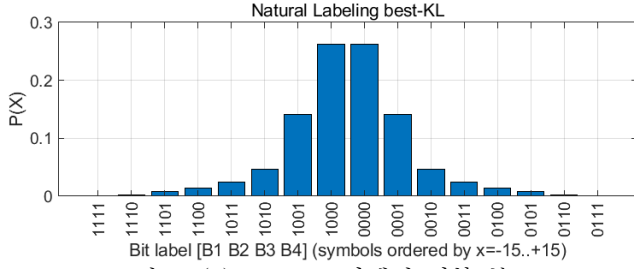


그림 1-(a). Natural 라벨링 심볼 분포

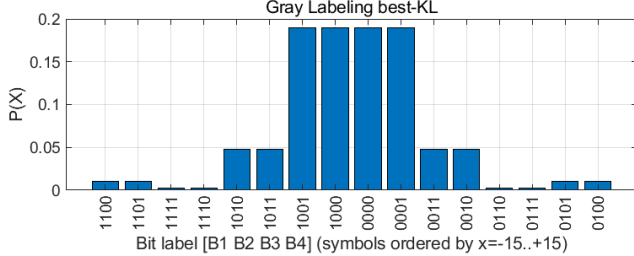


그림 1-(b). Gray 라벨링 심볼 분포

2.3 실험 설정

본 실험에서는 16-ASK 성상도 $\mathcal{X} = \{\pm 1, \pm 3, \dots, \pm 15\}$ 를 고려하고, 각 심볼을 4비트 라벨 $\mathbf{B} = (B_1, B_2, B_3, B_4)$ 로 표현하였다. 부호(sign)는 MSB인 B_1 로 결정되며, $B_1 = 0$ 일 때 양(+)의 레벨, $B_1 = 1$ 일 때 음(-)의 레벨이 되도록 정의하였다. 분포의 좌우 대칭성을 유지하기 위해 $P_1(0) = 0.5$ 로 고정하였고, 나머지 비트 위치에 대해서는 비균일 확률로 설정하여 심볼의 분포를 설정하였다. Natural 매핑은 (B_2, B_3, B_4) 를 자연 이진수로 해석하여 값이 작을수록 원점에 가까운 작은 진폭을 갖도록, 값이 클수록 원점에서 먼 큰 진폭을 갖도록 할당하였다. Gray 매핑은 (B_2, B_3, B_4) 에 표준 3-bit Gray 코드를 적용하여 인접 진폭 간 비트 차이가 1이 되도록 하였다.

Additive White Gaussian Noise(AWGN) 채널에서 Maxwell-Boltzmann(MB) 분포가 이론적으로 최적임이 알려져 있으므로[1], 고정된 MB 분포($\nu_0 = 0.05$)를 목표 분포로 설정한다. 본 실험에서는 2번째부터 4번째 bit-level의 $P_j(0)$ 을 0.05 간격으로 조절하여 다양한 후보 입력 분포를 구성하였다. 각 라벨링 방식에 대해 생성된 심볼 분포를 목표 MB 분포와 비교하였으며, 분포 간 유사도는 KL divergence를 기준으로 평가하였다. 라벨링 방식에 따른 KL divergence를 최소화하는 분포를 그림 1에 도시하였다.

2.4 실험 결과

2.4.1 KL divergence 기반 분포 근사 성능

고정된 목표 MB 분포($\nu_0 = 0.05$)에 대해 bit-level의 $P_j(0)$ 을 0.05 간격으로 탐색한 결과, natural 라벨링의 KL 최소점은 $\mathbf{P}_{NL} = [0.5 \ 0.95 \ 0.85 \ 0.65]$ 로 결정되고 gray 라벨링의 KL 최소점은 $\mathbf{P}_{GL} = [0.5 \ 0.95 \ 0.8 \ 0.5]$ 로 결정되었다. 이때 natural 라벨링으로 생성된 심볼 분포 Q_{NL} 과 MB 분포 Q_{MB} 의 KL divergence는 $D_{KL}(Q_{NL}||Q_{MB}) = 0.114356$ bits이고, Gray 매핑의 KL divergence는 $D_{KL}(Q_{GL}||Q_{MB}) = 0.400884$ bits이다. 즉, natural 라벨링이 Gray 라벨링에 비해 목표 MB 분포를 더 정확하게 근사함을 확인할 수 있으며, KL 다이버전스 기준에서 bit-level 확률 조절을 통한 분포 근사 성능이 라벨링 방식에 따라 달라질 수 있음을 확인할 수 있다.

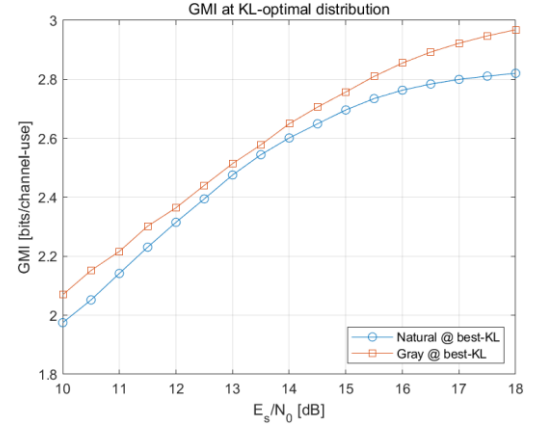


그림 2. 라벨링 방식에 따른 GMI 성능 그래프

2.4.2 BICM 성능: GMI

그림 2는 고정된 MB 분포를 기준으로 각 라벨링 방식에서 KL 다이버전스를 최소화하는 분포의 BICM GMI 성능을 비교한 결과이다. 그래프의 전 SNR 구간에서 Gray 라벨링이 natural 라벨링보다 높은 GMI를 달성하는 것을 확인할 수 있다. 이는 BICM 구조에서 Gray 라벨링이 인접한 심볼 간 해밍 거리가 1이 되도록 설계되어, 채널 잡음으로 인해 심볼이 인접 위치로 오류가 발생하더라도 대부분의 경우 단일 비트 오류로 제한되기 때문이다. 이러한 특성은 각 비트 레벨에서의 신뢰도를 보다 균등하게 유지하고, 결과적으로 bit-level mutual information을 효과적으로 증가시킨다. 따라서 KL 다이버전스 기준의 분포 근사 성능과는 무관하게, BICM-GMI 관점에서는 Gray 라벨링이 보다 효율적인 선택임을 확인할 수 있다.

III. 결론

본 논문에서는 bit-level CCDD 기반 확률 정형 시스템에서 심볼 라벨링 방식이 분포 근사 성능과 BICM 성능에 미치는 영향을 분석하였다. 실험 결과, natural 라벨링은 KL 다이버전스 기준에서 목표 MB 분포를 더 정확하게 근사하였으나, BICM-GMI 관점에서는 Gray 라벨링이 전 SNR 구간에서 일관되게 더 높은 성능을 달성하였다. 이는 분포 근사 성능과 BICM 정보율 성능이 반드시 일치하지 않으며, bit-level CCDD 기반 BICM 시스템에서 라벨링 선택이 핵심적인 설계 요소임을 보여준다.

ACKNOWLEDGMENT

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

참고 문헌

- [1] G. Böcherer, F. Steiner, and P. Schulte, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," *IEEE Transactions on Communications*, vol. 63, no. 12, pp. 4651-4665, Dec. 2015.
- [2] P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 430-434, Jan. 2016.
- [3] M. Piskun and W. Xu, "Bit-level probabilistically shaped coded modulation," *IEEE Communications Letters*, vol. 21, no. 9, pp. 1929-1932, Sept. 2017.
- [4] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 927-946, May 1998.