

광대역 초거대 안테나 다중 입출력 시스템에서 변환 부호화 기반 채널 상태 정보 피드백을 위한 레이트 할당 기법

박영목, 박범수, 이남윤
포항공과대학교

{ympark1999,bumsupark,nylee}@postech.ac.kr

Rate Allocation for Transform Coding-Based CSI Feedback in Massive MIMO Systems

Youngmok Park, Bumsu Park, and Namyoon Lee
POSTECH

요 약

본 논문에서는 6G FDD Massive MIMO 시스템을 위한 변환 부호화 기반의 최적 rate allocation 기법을 제안한다. 기지국 안테나 수의 증가로 인해 CSI 피드백 오버헤드가 급증함에 따라 다양한 압축 기법이 연구되었으나, 기존 딥러닝 기반 기법들은 단말 측에 과도한 연산 복잡도와 메모리 소모를 야기하는 실용적 한계가 있다. 이에 본 연구는 정보이론적 관점에서 채널의 공분산 고유기저와 비상관 계수를 각각 양자화하는 변환 부호화 구조에 주목하고, 제한된 전체 피드백 rate 하에서 두 성분 간의 왜곡 trade-off 를 고려한 이론적 최적 할당 조건을 규명한다. 제안하는 기법은 유도된 최적 조건을 기반으로 저복잡도 이분법을 수행하여 rate allocation 을 수행한다.

I. 서론

6G 시스템에서 Massive MIMO 는 주파수 효율 극대화 및 간섭 제어의 핵심 기술이다. Massive MIMO 의 빔포밍 및 다중화 이득을 온전히 실현하기 위해서는 기지국(Base Station, BS)에서 정확한 채널 상태 정보(Channel State Information, CSI)의 확보가 필수적이다[1]. 주파수 분할 이중화(Frequency Division Duplex, FDD) 시스템에서는 사용자 단말(User Equipment, UE)이 하향링크(Downlink, DL) CSI 를 추정 후 BS 로 피드백 해야 하는데, 안테나와 부반송파 수의 증가는 피드백 정보량을 급격히 늘려 심각한 피드백 오버헤드를 초래한다.

이를 해결하기 위해 압축 센싱과 딥러닝 기반 기법이 제안되었으나, 각각 채널 회소성 가정에 따른 모델 불일치 문제와 UE 의 연산 및 메모리 제약이라는 실용적 한계를 보인다.

이에 본 연구는 광대역 채널을 상관된 가우시안 소스로 모델링하고, 정보이론적으로 최적인 변환 부호화(Transform Coding, TC) 기법을 적용한다. 특히 BS 가 시변하는 채널 공분산을 사전에 알 수 없는 현실적 제약을 고려하여, UE 가 공분산의 고유기저 행렬과 비상관 채널 계수를 모두 양자화하여 전송하는 구조를 채택한다. 본 논문의 핵심 기여는 전체 복원 왜곡을 기저 행렬과 계수 양자화 오차의 합으로 정의하고, 제한된 피드백 rate 내에서 이들 간의 상충 관계를 고려한 최적의 rate allocation 조건을 이론적으로 규명하는 것이다.

II. 시스템 모델 및 문제 정의

본 장에서는 TC 기반 CSI 피드백 시스템 모델을 정의하고, 제한된 피드백 rate 하에서의 문제를 정식화한다. 본 논문에서는 FDD Massive MIMO DL 시스템을 고려한다. BS 는 N_t 개의 안테나를, UE 는 단일 안테나를 가지며 OFDM 방식으로 N_c 개의 부반송파를

사용한다. 또한 UE 는 DL CSI 를 정확히 추정할 수 있다고 가정한다. k 번째 부반송파에서 DL 채널 벡터를 $\mathbf{h}_k \in \mathbb{C}^{N_t}$ 라 하고, $N := N_c N_t$ 로 정의하면, 전체 광대역 CSI 는 다음과 같은 벡터화 하여 표현할 수 있다,

$$\mathbf{h} = [\mathbf{h}_1^T, \dots, \mathbf{h}_{N_c}^T]^T \in \mathbb{C}^N. \quad (1)$$

COST2100, 3GPP SCM 등 채널 모델에 근거하여[2,3], 광대역 CSI 벡터 \mathbf{h} 는 평균이 0 이고 공분산 행렬 \mathbf{R} 을 갖는 상관된 다변량 가우시안 벡터 채널 모델링과 잘 맞는다. 공분산 행렬 \mathbf{R} 은 다음과 같이 고윳값 분해된다,

$$\mathbf{R} = \mathbf{U}_e \mathbf{\Lambda} \mathbf{U}_e^H. \quad (2)$$

여기서 \mathbf{U}_e 는 \mathbf{R} 의 고유기저 행렬이고, $\mathbf{\Lambda} := \text{diag}[\lambda_1, \dots, \lambda_N]$ 는 대응하는 고윳값들을 원소로 갖는 대각 행렬이다.

CSI 피드백은 UE 측 인코더가 \mathbf{h} 를 제한된 길이의 비트열로 압축하고 BS 측 디코더가 비트열을 $\hat{\mathbf{h}}$ 로 재구성하는 손실 압축 문제이며, 이 문제의 목표는 제한된 피드백 rate 하에 평균제곱오차(MSE)인, $\frac{1}{N} \mathbb{E}[\|\mathbf{h} - \hat{\mathbf{h}}\|_2^2]$, end-to-end(E2E) 왜곡, D_{E2E} , 을 최소화하는 인코더, 디코더 설계이다. 상관된 가우시안 소스와 MSE 에 대해 정보이론적으로 rate-distortion 하한으로 압축하기 위해, 본 논문에서는 TC 기법을 적용한다. 구체적으로 UE 는 공분산 행렬의 고유기저 행렬 \mathbf{U}_e 를 이용하여 채널을 비상관화(decorrelation) 한다,

$$\hat{\mathbf{h}} = \mathbf{U}_e^H \mathbf{h}. \quad (3)$$

채널은 서로 독립적인 가우시안을 따르는 스칼라 계수들로 변환되며 인코더에서는 스칼라 계수와 기저 행렬을 각각 유한 비트로 양자화하여 BS 로 전송한다.

이로 인해 전체 CSI 재구성 과정에서 스칼라 계수 양자화로 인한 왜곡과 기저 행렬 양자화로 인한 불일치 왜곡이 동시에 발생한다. 따라서 제한된 총 피드백 비트율 R_{tot} 은 기저 행렬 양자화를 위한 R_0 와 스칼라 계수 양자화를 위한 R_q 의 합으로 구성된다.

$$\min_{R_0, R_q} D_{E2E}(R_0, R_q) \text{ s.t. } R_0 + R_q = R_{tot}. \quad (4)$$

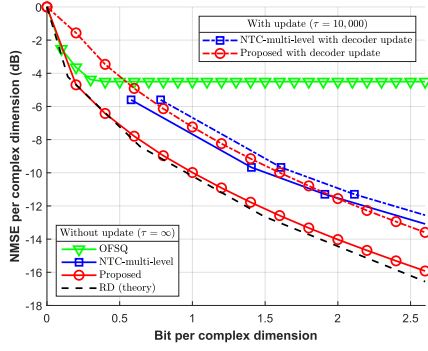


그림 1. 상관 가우시안 채널 압축 성능.

본 논문에서는 변환 부호화 기반 CSI 피드백 구조에서, 설계 변수인 R_0 와 R_q 를 조절하여 D_{E2E} 를 최소화하는 최적 rate allocation 문제를 다룬다.

III. 제안 기법

본 장에서는 앞 장에서 정의한 D_{E2E} 를 기저 행렬과 계수 양자화 오차의 합으로 분리하고 주어진 R_{tot} 하에서 R_0 와 R_q 의 최적의 rate allocation 기법을 제안한다. 분석의 편의를 위해 스칼라 계수 양자화기를 high-rate ECSQ(Entropy-Constrained-Scalar-Quantization)를 가정하면, D_{E2E} 는 각 양자화로 인한 왜곡의 독립적인 합으로 표현된다. 기저 행렬 양자화로 생기는 왜곡을 $D_0(R_0)$, 스칼라 계수 양자화로 인한 왜곡을 $D_q(R_q)$ 라 하면, D_{E2E} 는 다음과 같이 분리된다,

$$D_{E2E}(R_0, R_q) = D_0(R_0) + D_q(R_q). \quad (5)$$

기저 행렬의 양자화 방식을 RVQ(Random Vector Quantization), 스칼라 계수의 압축률을 RWF(Reverse Water-Filling)로 최적 분배했을 때, D_{E2E} 는 최종적으로 다음과 같다,

$$D_{E2E}(R_0, R_q) = \frac{1}{N} \sum_{i=1}^N \min\{\lambda_i, \mu(R_q)\} + \left(\frac{N-1}{N^2}\right) \left(\sum_{i=1}^p \lambda_i\right) 2^{-\frac{R_0 N \tau}{p(N-1)}}. \quad (6)$$

이때 $\mu(R_q)$ 는 RWF의 water level이며, τ 는 coherence block length 이고, p 는 \mathbf{R} 의 rank 이다.

$D_0(R_0)$ 와 $D_q(R_q)$ 는 동일한 rate 예산을 공유하므로, rate 할당에 따른 왜곡 변화량이 같아질 때 가장 작은 $D_{E2E}(R_0, R_q)$ 를 달성할 수 있다. 식 전개를 통해 다음 최적 rate allocation 조건을 유도할 수 있다,

$$\mu(R_q) = \alpha_0 D_0(R_0), \text{ where } \alpha_0 = N\tau/p(N-1). \quad (7)$$

식 (7)을 만족하는 최적의 R_q 는 $[0, R_{tot}]$ 구간에서 유일하게 존재하여 이분법을 활용해 저복잡도로 탐색 가능하다.

본 논문에서는 제안하는 기법을 $Nt = Nc = 32$ 인 채널에서 비교하였다. R-D 커브, NTC-multi-level[4], OFSQ[5]와 비교한다. R-D 커브는 rate matching 을 통해 채널을 가우시안으로 근사해 계산한다. NTC-multi-level, OFSQ는 딥러닝 기반 CSI 압축 기법이다. 3GPP 에 따르면[3], 딥러닝 기반 CSI 압축기는 채널의 통계적 특성이 변할 때마다 BS 의 디코더를 업데이트 해주어야 성능 손실을 피할 수 있다. 따라서 본 실험에서는 딥러닝 기반 CSI 압축 기법의 디코더 업데이트 비용을 $32 \times (\text{# of decoder parameters})$ 로 설정하였다. 이때 32 는 한

개의 실수 값 파라미터를 전송하는데 필요한 비트 수이다. 실험은 채널의 통계적 성질이 불변할 때($\tau = \infty$)와 통계적 성질이 변할 때($\tau = 10,000$)로 나누어 비교한다. 그림 1은 상관된 가우시안 채널에서 비교한

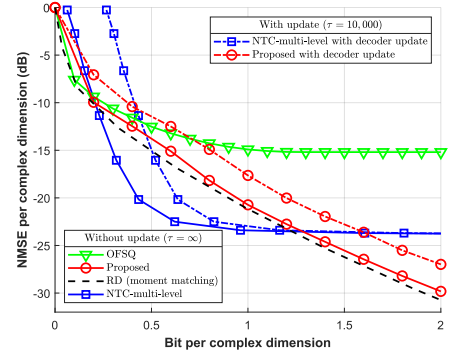


그림 2. COST2100 Indoor 5.3GHz 채널 압축 성능.

결과이며, $\tau = \infty$ 일 때, 제안하는 기법의 압축 성능은 이론적 한계에 근접하여 모든 영역에서 딥러닝 기반 압축 기법보다 3dB 이상 적은 오차를 보였다. $\tau = 10,000$ 일 때도 NTC-multi-level 보다 제안하는 기법이 high-rate 에서 1dB 가량 더 적은 압축 오차를 달성하였다. 이는 제안하는 rate 최적화 기법의 효율성을 보여준다.

그림 2 는 COST2100 실내 환경 채널에서 압축 성능을 비교한 결과이며, 중심 주파수는 5.3GHz 이다. $\tau = \infty$ 일 때, low-rate 영역에선 NTC-multi-level 이 비가우시안 분포의 특징을 잘 학습해 가장 좋은 압축 성능을 달성했지만, high-rate 영역에선 제한된 모델 용량으로 인해 제안 기법과 성능 교차가 발생했다. $\tau = 10,000$ 일 때 제안 기법은 낮은 rate(~ 0.3 bit)에서 NTC-multi-level 보다 9dB 더 낮은 오차 성능을 달성한다. 이는 기존 방법 대비 제안하는 rate 최적화 기반 디코더 업데이트 방식의 효율성을 보여준다.

ACKNOWLEDGMENT

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (2021-0-00161, Post MIMO system research for massive connectivity and new wireless spectrum). This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT).(No.RS-2023-00208552)

참 고 문 헌

- [1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590-3600, 2010.
- [2] L. Liu, C. Oestges, J. Poutanen, K. Haneda, P. Vainikainen, F. Quitin, F. Tufvesson, and P. D. Doncker, "The COST 2100 MIMO channel model," *IEEE Wireless Commun.*, vol. 19, no. 6, pp. 92-99, 2012.
- [3] "Study on channel model for frequencies from 0.5 to 100 GHz," 3rd Generation Partnership Project (3GPP), Tech. Rep. TR 38.901, 2020, release 16.
- [4] B. Park, H. Do, and N. Lee, "Transformer-based nonlinear transform coding for multi-rate CSI compression in MIMO-OFDM systems," in *IEEE Int. Conf. Commun.*, June 2025, pp. 2327-2333.
- [5] K. Liotopoulos, N. A. Mitsiou, P. G. Sarigiannidis, and G. K. Karagiannidis, "Multi-length CSI feedback with ordered finite scalar quantization," *IEEE Commun. Lett.*, vol. 29, no. 8, pp. 1973-1977, 2025.