

# 트랜스포머 기반 FDD 대규모 MIMO 시스템에서의 효율적인 CSI 압축 및 복원 기법

허성욱, 고영재\*

고려대학교

huhbaksa@korea.ac.kr, \*koyc@korea.ac.kr

## Transformer-Based CSI Compression and Recovery Techniques for FDD Massive MIMO Systems

Sunguk Huh, Young-chai Ko  
Korea Univ.

### 요약

본 논문에서는 FDD 기반 대규모 MIMO 시스템에서 발생하는 과도한 CSI 피드백 오버헤드를 효과적으로 해결하기 위해, 커스터마이즈드 트랜스포머 인코더 구조를 도입한 TransInDecNet+를 제안한다. 제안하는 모델은 COST2100 데이터셋을 활용한 다양한 시나리오에서 실험을 수행하였으며, 기존 SOTA 모델 대비 약 35% 수준의 연산량(FLOPs)으로 다양한 환경 및 압축 비율에서 경쟁력 있는 NMSE 성능을 유지함을 확인하였다. 본 연구 결과는 TransInDecNet+의 실용적 활용 가능성을 제시한다.

### I. 서론

FDD 기반의 대규모 MIMO 시스템에서는, 기지국(BS)에서 정확한 채널 상태 정보(CSI)를 획득하는 것이 효율적인 빔포밍, 간섭 억제, 그리고 스펙트럼 효율 향상에 필수적이다. TDD는 업링크와 다운링크 채널 간의 상호성을 기반으로 기지국에서 직접 CSI를 추정할 수 있지만, FDD 시스템은 업링크와 다운링크 전송이 서로 다른 주파수 대역에서 이루어지기 때문에 직접적인 채널 상호성이 성립하지 않는다. 이로 인해 사용자 단말(UE)은 다운링크 CSI를 추정한 후 이를 피드백 링크를 통해 기지국으로 전송해야 한다. 그러나 대규모 MIMO 시스템에서 안테나 수가 증가함에 따라, 요구되는 CSI 피드백 양도 비례적으로 증가하게 되어 과도한 피드백 오버헤드가 발생한다. 따라서 높은 통신 성능을 유지하면서도 실용적인 FDD 기반 대규모 MIMO 시스템을 구현하기 위해서는 효율적인 CSI 피드백 메커니즘의 개발이 필수적이다. CsiNet[1]은 딥러닝을 활용한 첫번째 접근법이고 합성곱 신경망(CNN)을 기반으로 CSI 피드백 성능을 크게 향상시켰다. TransNet[4]은 [2]에서 소개된 트랜스포머 모델을 활용하여 CSI 피드백 품질을 향상시켜 SOTA 성능을 달성하였다. 본 연구에서는 트랜스포머 모델을 활용하여 FDD 기반 대규모 MIMO 시스템에서 효율적인 CSI 압축 및 복원 기법을 제안한다.

### II. 본론

#### A. 시스템 모델 및 문제 정의

본 논문에서는 [1]의 방법을 따라, 다수의 송신 안테나( $N_t \gg 1$ )를 갖는 단일 셀 다운링크 대규모 MIMO 시스템을 고려한다. 본 시스템은 OFDM을 기반으로 하며, 각 부반송파의 수신 신호는 다음과 같이 표현된다:

$$y_n = \tilde{h}_n^H v_n x_n + z_n \quad (1)$$

여기서  $\tilde{h}_n$ ,  $v_n$ ,  $x_n$ ,  $z_n$ 는 각각 채널 벡터, 프리코딩 벡터, 송신 심볼, 잡음을 나타낸다. 전체 CSI 행렬  $\tilde{H}$ 는 모든 부반송파의 채널 벡터를 쌓아 생성된다. UE는 파일럿 신호를 통해  $\tilde{H}$ 를 추정한 후, 이를 피드백 링크를 통해 BS로 전송한다. 하지만 송신 안테나 수나 부반송파 수가 증가할수록 CSI의 크기도 커져 피드백 오버헤드가 급격히 증가한다. 이를 해결하기 위해, 2 차원 DFT를 적용하여 각도-지연(angular-delay) 도메인에서 회소성을 이용하고, 의미 있는  $N_a$ 개의 행만 남겨 CSI를 압축한다. 압축된 CSI 행렬  $H_a$ 는 인코더에 입력되고, 인코더  $f_{en}$ 과 디코더  $f_{de}$ 를 설계하여 CSI를 압축·복원한다.

$$s = f_{en}(H_a, \theta_{en}) \quad (2)$$

$$\hat{H}_a = f_{de}(s, \theta_{de}) \quad (3)$$

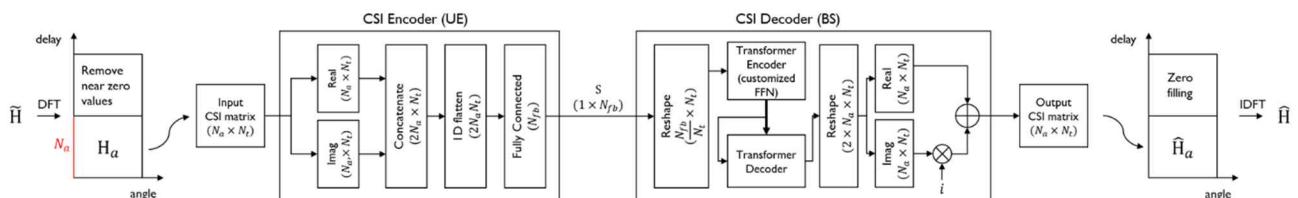


그림 1. TransInDecNet+의 구조

표 1. 제안하는 모델과 TransNet의 NMSE(dB) 및 FLOPs 성능 비교

Compression ratio ( $\eta$ )	1/4			1/8			1/16			1/32			1/64		
	Methods	NMSE		NMSE		NMSE		NMSE		NMSE		NMSE		NMSE	
		FLOPS	Indoor	Outdoor	FLOPS	indoor	Outdoor	FLOPS	indoor	Outdoor	FLOPS	indoor	Outdoor	FLOPS	indoor
(proposed)-400ep	12.54M	-27.4	-10.40	10.92M	-19.84	-7.21	10.11M	-13.92	-5.38	9.71M	-9.62	-4.69	9.51M	-5.86	-5.02
TransNet-400ep	35.72M	-29.22	-13.99	34.70M	-21.62	-9.57	34.14M	-14.98	-6.90	33.88M	-9.83	-3.77	33.75M	-5.77	-2.20
CsiNet	5.41M	-17.36	-8.75	4.37M	-12.70	-7.61	3.84M	-8.65	-4.51	3.58M	-6.24	-2.81	3.45M	-5.84	-1.93

전체 프로세스는 평균제곱오차(MSE)를 최소화하는 최적화 문제로 표현할 수 있으며, 최종 목표는  $H_a$  와 복원된  $\hat{H}_a$  간의 오차를 최소화하는 것이다.

$$(\hat{\theta}_{en}, \hat{\theta}_{dn}) = \underset{\theta_{en}, \theta_{de}}{\operatorname{argm}} \hat{n} \| H_a - \hat{H}_a \|^2 \quad (4)$$

### B. 제안하는 모델 TransInDecNet+

그림 1은 단일 배치에 대해 제안하는 TransInDecNet+의 아키텍처를 보여준다. CSI 인코더는 복소수 행렬인  $\bar{H}$ 에서 상위  $N_a$  개의 행만 선택하며, 입력 CSI 행렬  $H_a$ 는  $N_a \times N_t$  크기의 복소수 행렬이다. 이후, 잘라낸 행렬의 실수부와 허수부를 각각 이어붙여  $2N_a \times N_t$  크기의 실수 행렬  $H$ 를 구성한다. 그 다음,  $H$ 를 1 차원 벡터로 평탄화한 뒤, 이를 완전 연결 계층(Fully-connected layer, FC) 블록에 통과시켜 압축한다. 압축된 코드워드  $s$ 의 크기는  $1 \times N_{fb}$ 이며, 압축 비율 ( $\eta$ )는  $\frac{N_{fb}}{2N_a N_t}$ 로 정의된다.

기지국에서  $s$ 를 수신한 후  $\frac{N_{fb}}{N_t} \times N_t$  크기로 재구성하여 커스터마이즈드 트랜스포머 인코더의 입력으로 넣어준다. 커스터마이즈드 트랜스포머 인코더의 출력은 이후 트랜스포머 디코더로 전달된다. 트랜스포머 디코더의 출력은  $2 \times N_a \times N_t$ 로 재구성된 뒤, 실수부와 허수부로 분리된다. 복소수 행렬 복원을 위해, 허수부는 원소별로  $i$ 를 곱해 실수부와 합쳐 최종적으로 복원된 CSI 행렬  $\hat{H}_a$ 를 얻는다.  $\hat{H}_a$ 의 최종 크기는  $N_a \times N_t$ 이다.

커스터마이즈드 트랜스포머 인코더는 [2]에서 제안된 트랜스포머 인코더 구조를 기반으로 하되, 피드포워드 신경망(FFN)을 커스터마이즈하여 구현하였다. 본 연구의 FFN 블록은 CSI 인코더의 FC에서 수행된 압축을 효과적으로 복원하는 역할도 함께 수행한다. 트랜스포머 디코더는 [2]에서 제안된 원래 구조와 동일하게 사용하였다.

### C. 데이터셋, 학습 방식 및 평가 지표

[1]의 실험 환경을 따라, 본 연구에서는 COST2100 데이터셋[3]을 사용하여 실험을 진행하였다. 5.3GHz 대역의 실내 피코셀 시나리오와 300MHz 대역의 육외 농촌 시나리오의 두 가지 환경을 고려하였다. BS는  $N_t = 32$  개의 안테나로 구성하였고, FDD 시스템에서의 서브캐리어 수는  $N_c = 1024$ 로 설정하였다. 각 도도메인에서는  $N_a = 32$ 를 사용하였다. 트랜스포머 인코더와 디코더의 차원은  $d_{model} = 32$ 로 정의하였다. 데이터셋은 총 150,000 개의 샘플로 구성되며, 이 중 100,000 개는 학습, 30,000 개는 검증, 20,000 개는 테스트에 사용하였다. 학습에는 배치 크기 200과 학습률  $1 \times 10^{-4}$ 를 적용하였으며, 드롭 아웃은 사용하지 않았다. 다양한 압축 비율 1/4, 1/8, 1/16, 1/32, 1/64에서 epoch 400에 대해 NMSE를 평가하였다.

평가 지표로는 원본  $H$ 와 복원된  $\hat{H}$  간의 차이를 측정하는 정규화 평균제곱오차(NMSE)로 수행하였으며, NMSE는 다음과 같이 정의된다:

$$\text{NMSE} = E \left\{ \frac{\|H - \hat{H}\|_F^2}{\|H\|_F^2} \right\} \quad (5)$$

### D. 실험 결과

표 1은 제안한 TransInDecNet+과 기존 기법들 간의 NMSE 및 FLOPs 성능을 비교한 결과를 제시한다. 각 압축 비율( $\eta$ )별 NMSE 값은 실내 및 실외 시나리오에서 측정되었다. TransInDecNet+는 기존 SOTA 모델인 TransNet[4]에 비해 약 35% 수준의 FLOPs 만으로 작동하면서도, NMSE 성능은 충분히 수용 가능한 수준을 유지한다. 특히,  $\eta = 32$ 의 실외 시나리오,  $\eta = 64$ 의 실내 및 실외 시나리오에서는 TransInDecNet+가 더 낮은 연산 복잡도(FLOPs)로 더 우수한 NMSE 성능을 달성하여 주목할 만하다. 또한, CNN 기반의 경량 모델인 CsiNet[1]과 비교했을 때, TransInDecNet+는 전반적으로 우수한 NMSE 성능 향상을 보여준다. 다만, 일부 시나리오에서는 FLOPs 차이에 비해 NMSE 성능의 향상이 뚜렷하지 않은 경우도 있어, 이 점은 한계로 작용할 수 있다. TransInDecNet+는 CSI 인코더에서 FC 블록 외에는 연산량이 많은 구성 요소가 없기 때문에, UE 측 연산 복잡도를 크게 줄여 실제 시스템에서의 실용적인 적용 가능성을 제시한다.

### III. 결론

본 논문에서는 FDD 기반 대규모 MIMO 시스템에서 효율적인 CSI 압축 및 복원을 위해 커스터마이즈드 트랜스포머 인코더 구조를 적용한 TransInDecNet+를 제안하였다. 실험 결과, 본 모델은 기존 SOTA인 TransNet 대비 약 35% 수준의 FLOPs로 작동하면서도 다양한 환경과 압축 비율에서 경쟁력 있는 NMSE 성능을 유지하였다. 일부 시나리오에서는 더 낮은 연산 복잡도에도 오히려 복원 성능이 향상되는 결과를 보였다. 이는 TransInDecNet+가 실용적인 CSI 피드백을 위한 유망한 솔루션임을 보여준다.

### ACKNOWLEDGMENT

이 논문은 2025년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (2022-0-00704, 초고속 이동체 지원을 위한 3D-NET 핵심 기술 개발)

### 참 고 문 헌

- [1] C. -K. Wen, W. -T. Shih and S. Jin, "Deep Learning for Massive MIMO CSI Feedback," in IEEE Wireless Communications Letters, vol. 7, no. 5, pp. 748–751, Oct. 2018.
- [2] A. Vaswani et al., "Attention is all you need," in Proc. 31st Int. Conf. Neural Inf. Process. Syst. (ICONIP), vol. 30, 2017, pp. 5998–6008.
- [3] L. Liu et al., "The COST 2100 MIMO channel model," in IEEE Wireless Communications, vol. 19, no. 6, pp. 92–99, December 2012.
- [4] Y. Cui, A. Guo and C. Song, "TransNet: Full Attention Network for CSI Feedback in FDD Massive MIMO System," in IEEE Wireless Communications Letters, vol. 11, no. 5, pp. 903–907, May 2022.