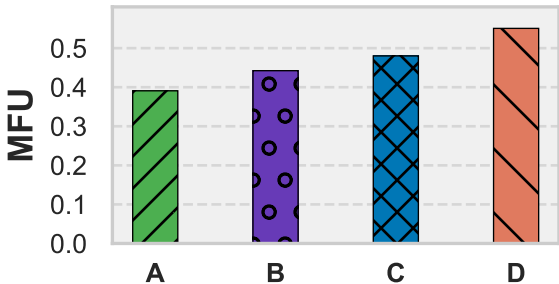**Legend:**
- (A) BurstAttention w. DoubleRing w/o Selective Checkpointing
- (B) BurstAttention w. Ulysses(intra node) w/o Selective Checkpointing
- (C) BurstAttention w. DoubleRing
- (D) BurstAttention w. Ulysses(intra node)

MFU for 7B-32xA100 1024k Training

MFU

A B C D

Tokens/s for 7B-32xA100 1024k Training

Tokens/s

A B C D