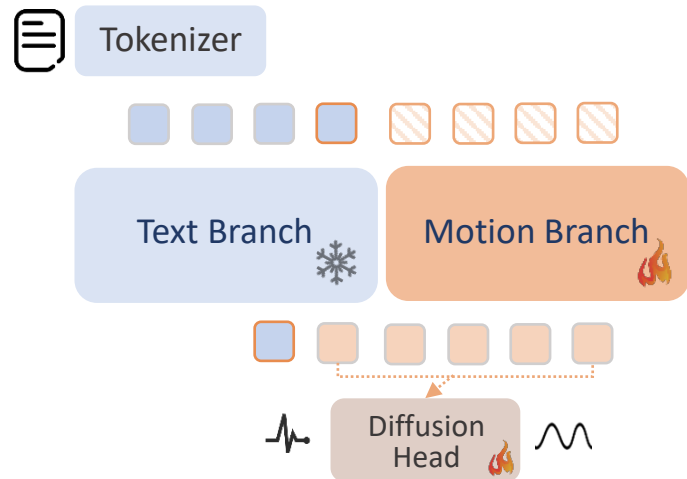


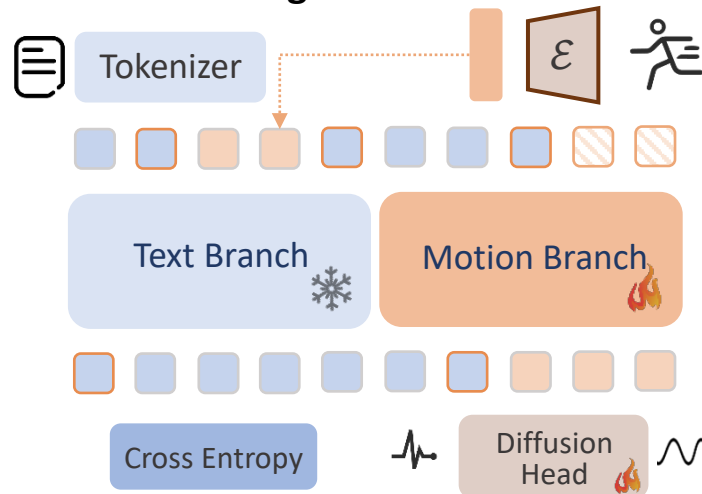
## Stage 1

## Text to Motion Pre-training



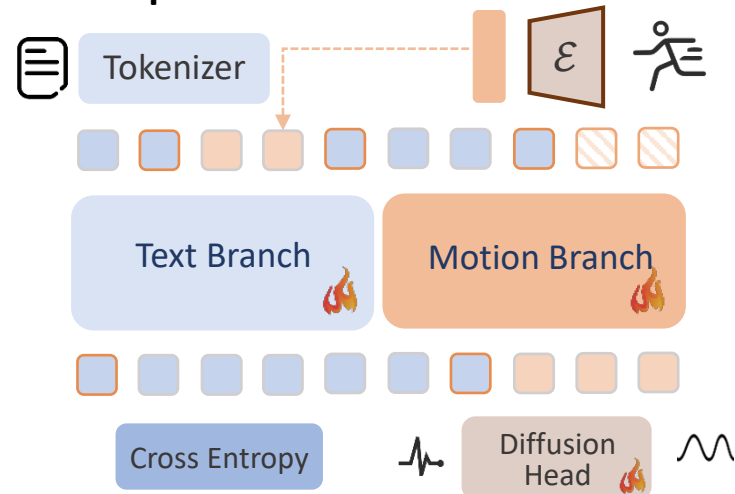
## Stage 2

## Cross-modal Alignment

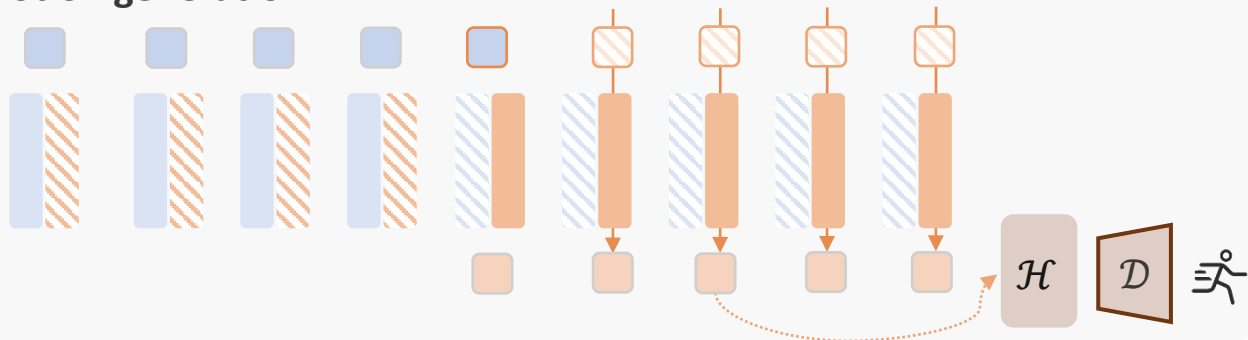


### Stage 3

## Joint Optimization



## Motion generation



## Motion understanding

