

Step Time vs Number of BWD Finetuning Layers  
Max Tokens per Batch: 256

