Batch Step Time vs Number of Backward Finetuning Layers
Model: meta-llama/Llama-3.1-70B (TP=4)
Batch Size: 8 - Max Tokens per Batch: 512
BWD finetuning tokens: 1024