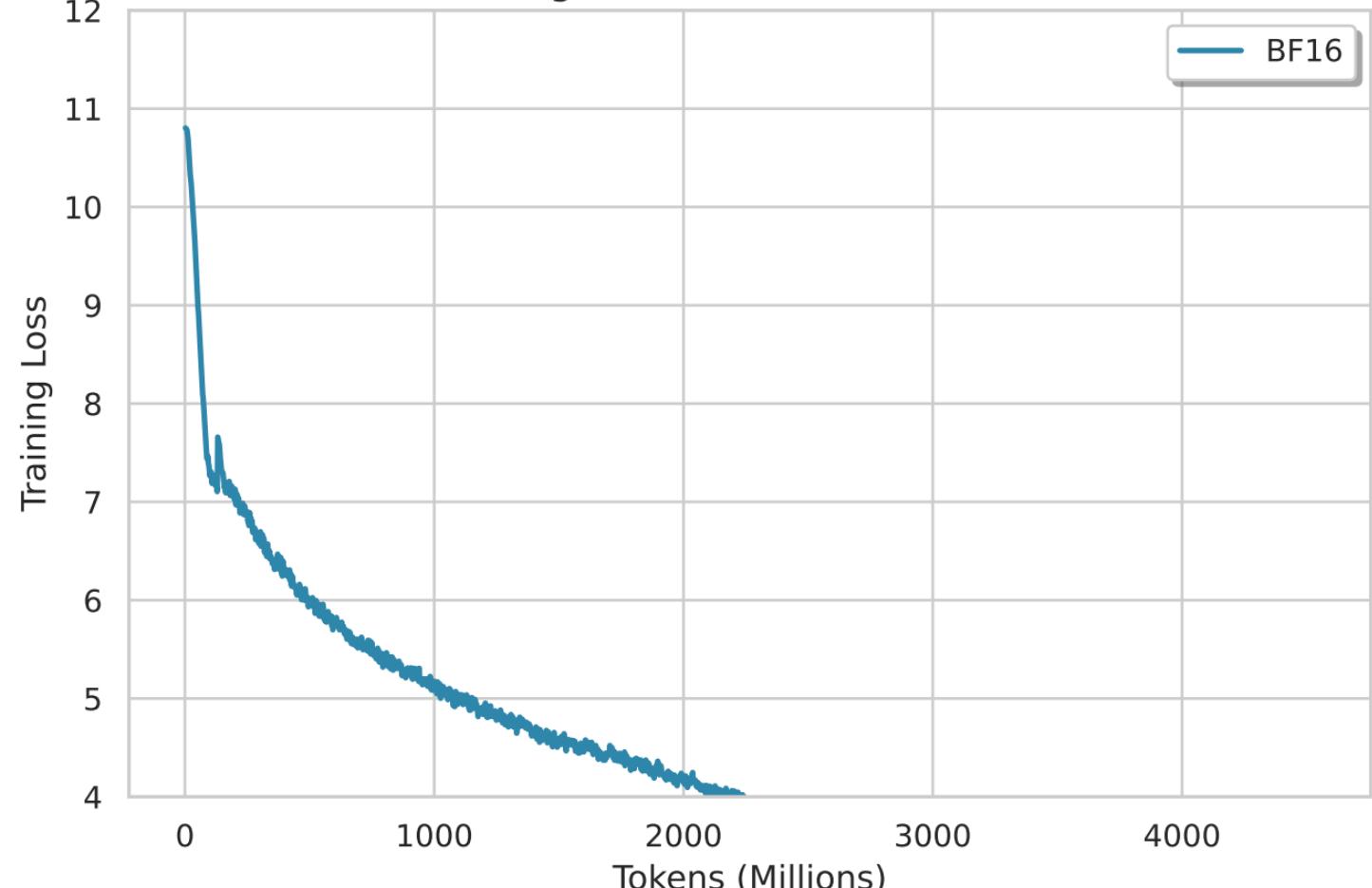


Training Loss vs Tokens Processed



Training Loss vs Wall Time

