FEABench (Finite Element
Analysis Benchmark):
Evaluating Language Models on
Multiphysics Reasoning Ability