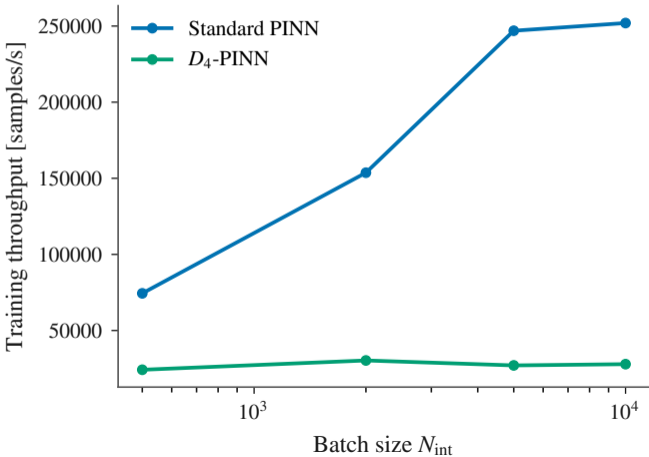


(a) Training throughput



(b) Inference latency

