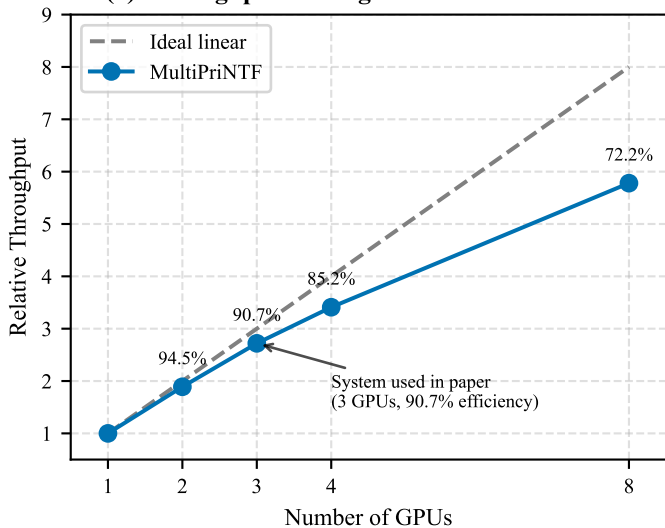
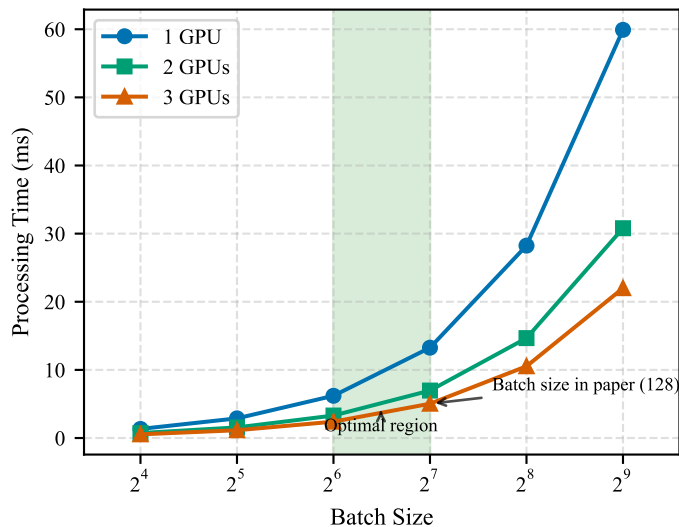


# System Scalability Analysis

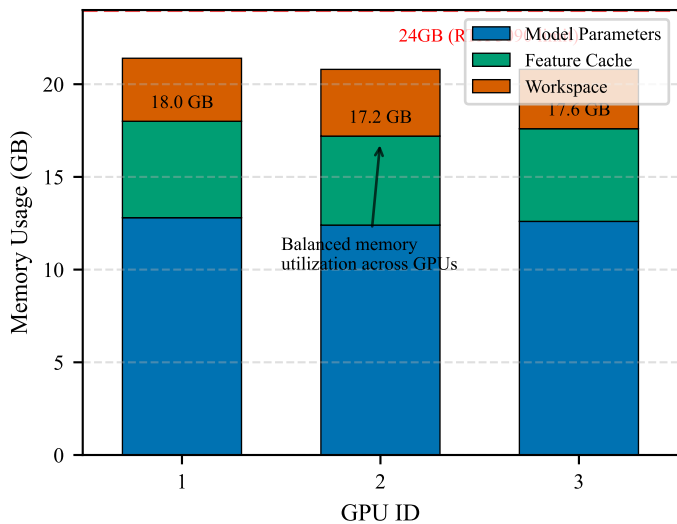
(a) Throughput Scaling with Number of GPUs



(b) Processing Time vs. Batch Size



(c) Memory Utilization Patterns Across GPUs



(d) Communication Overhead vs. Dataset Size

