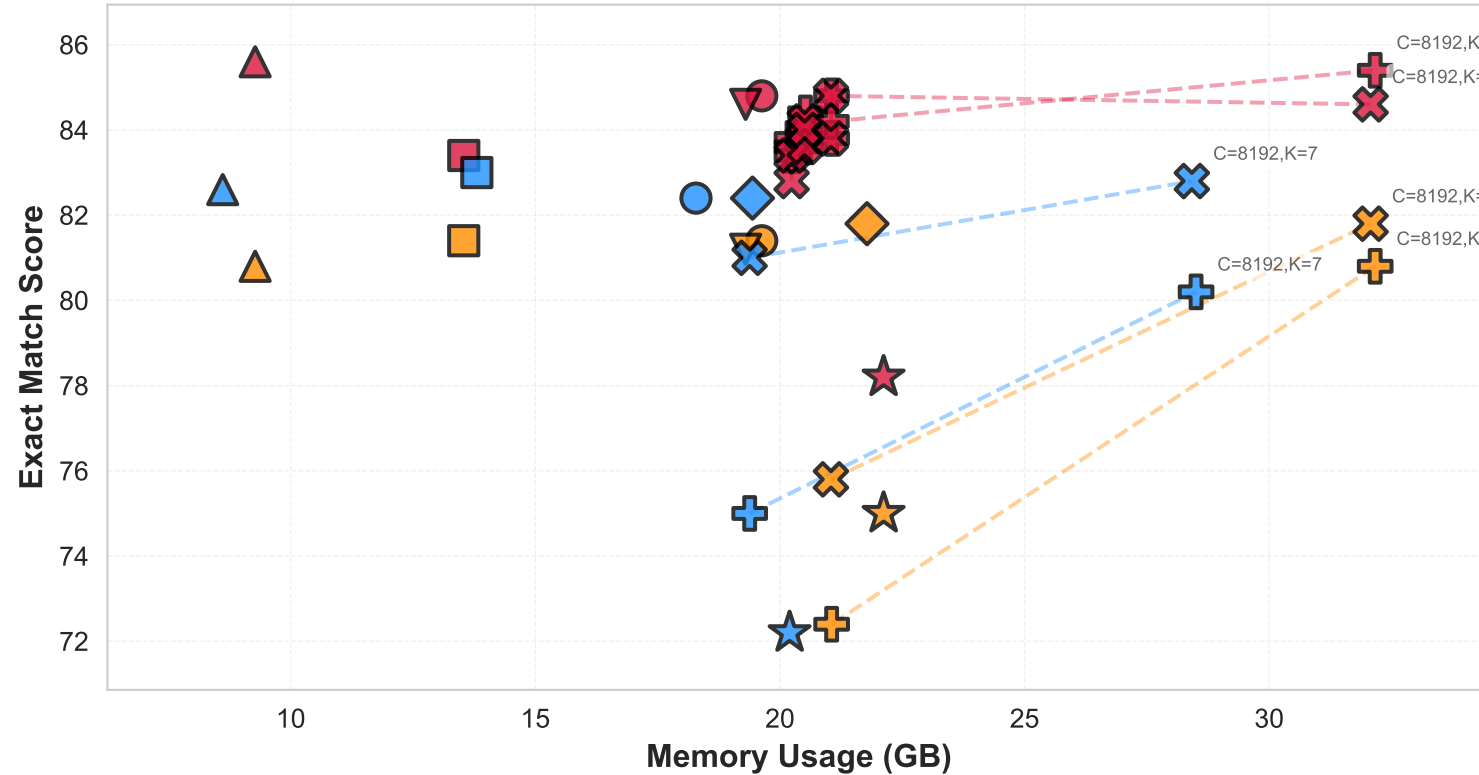


ICL Banking Task Performance Analysis (16K Context)

Memory and Latency vs Exact Match

Memory vs Performance



Throughput vs Performance

