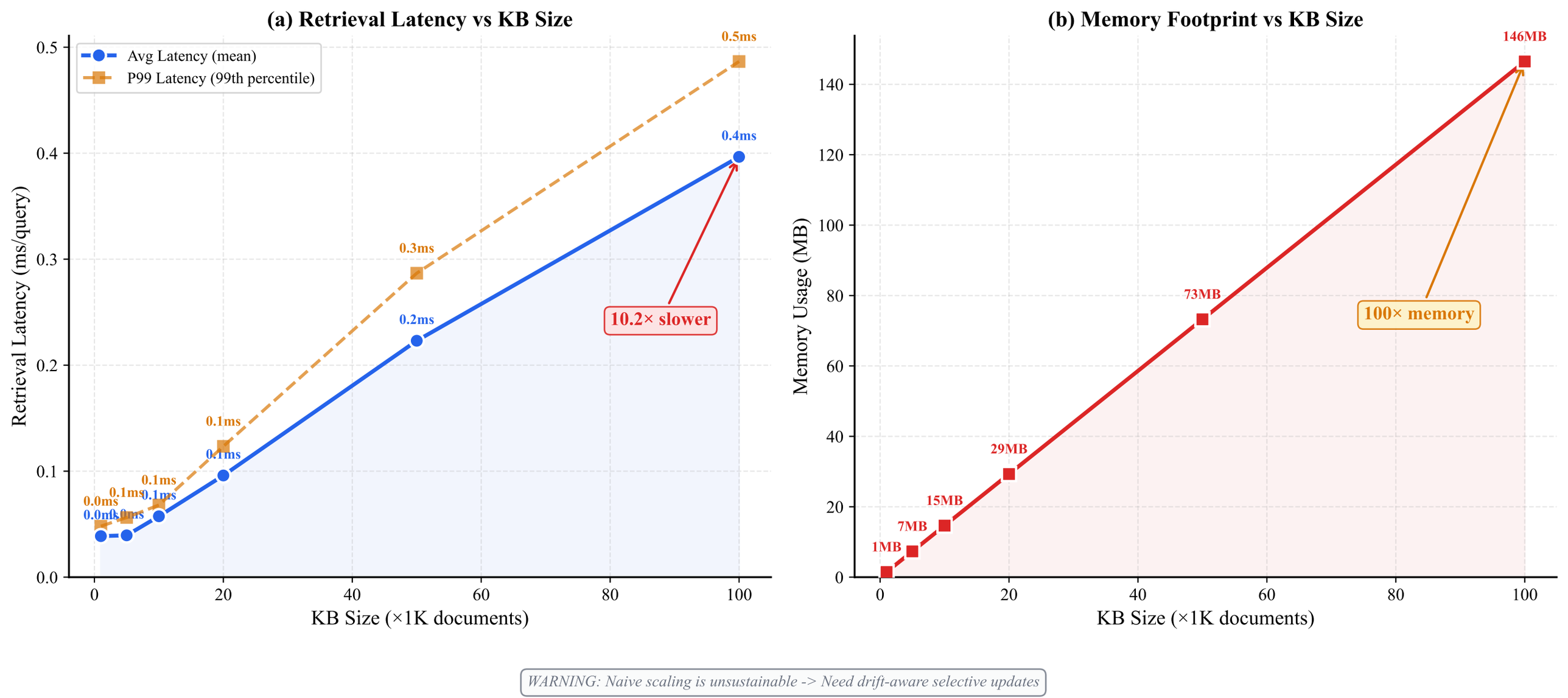
(a) Retrieval Latency vs KB Size
Avg Latency (mean)
P99 Latency (99th percentile)
Retrieval Latency (ms/query)
KB Size (×1K documents)
0.5ms
0.4ms
0.3ms
0.2ms
0.1ms
0.1ms
0.1ms
0.1ms
0.0ms
0.0ms
0.0ms
10.2× slower
(b) Memory Footprint vs KB Size
Memory Usage (MB)
KB Size (×1K documents)
146MB
73MB
29MB
15MB
7MB
1MB
100× memory
WARNING: Naive scaling is unsustainable -> Need drift-aware selective updates