Attention Mechanism Latency Scaling on Apple M4