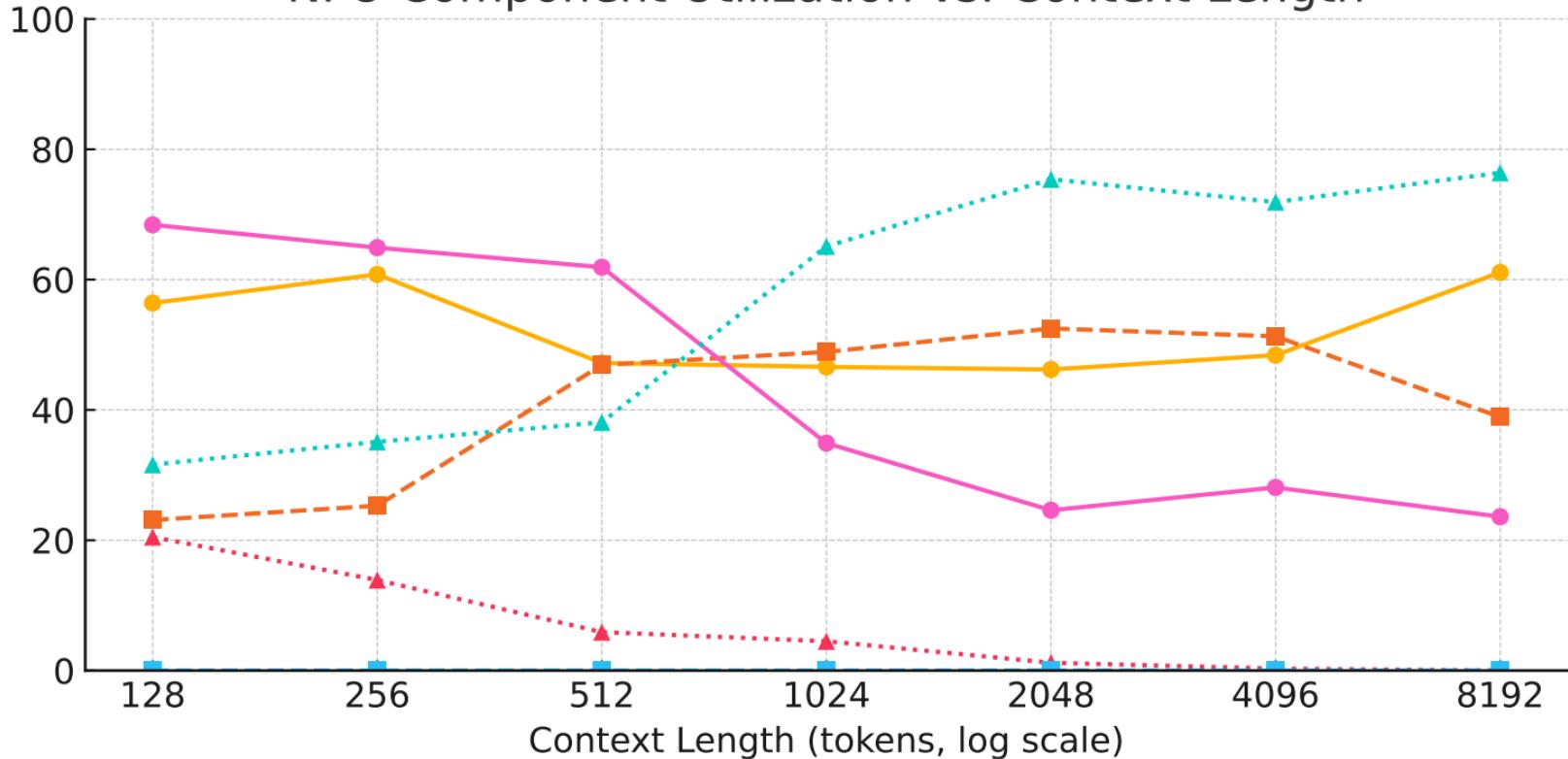


NPU Component Utilization vs. Context Length

Utilization (%)



Context Length (tokens, log scale)

- Fourier - DPU
- Fourier - DMA
- Fourier - SHAVE
- Retentive - DPU
- Retentive - DMA
- Retentive - SHAVE