Latency Analysis for Full NeRF Pass Using Memristor Crossbar

1. NeRF Architecture Configuration

The target NeRF model architecture is as follows:

- **Input Layer**: 5-dimensional input
- **Positional Encoding**: Encoded to 84 features
- **Network Layers**:
  - Layer 1: $84 \rightarrow 256$
  - Layer 2: $256 \rightarrow 256$
  - Layer 3: $256 \rightarrow 256$
  - Output Layer: $256 \rightarrow 4$

This leads to four weight matrices implemented in separate crossbar tiles:

Layer	Shape	Tile Size
Layer 1	256×84	256×84
Layer 2	256×256	256×256
Layer 3	256×256	256×256
Output Layer	4×256	4×256

2. Latency Contributors Per Layer

Each inference through a crossbar tile includes the following components:

2.1. Crossbar Computation Latency

Parallel analog computation in memristor crossbars is extremely fast due to Kirchhoff's Law. Based on LTSpice simulations: $\sim$**0.3 ns**

2.2. DAC Delay

High-speed DACs such as `TI DAC5670` and `Analog Devices AD9116` operate at 1 GSPS+. Conservative average delay: $\sim$**5 ns**
Source: TI E2E Forum — `https://e2e.ti.com/support/data-converters-group/data-converters/f/data-converters-forum/546029`

2.3. ADC Delay

High-speed ADCs like `TI ADS5400` exhibit latency of $\sim$**7 ns**
Source: TI E2E Forum (same thread)

2.4. Crossbar Reprogramming Time

Changing weights in a memristor crossbar involves applying write pulses. Programming time: **∼100 μs per array**
Source: Nature Communications: `https://www.nature.com/articles/s41467-023-44620-1`

3. Per-Layer Latency Breakdown (Excluding Programming Time)

Layer	Crossbar Time	DAC Delay	ADC Delay	Total Latency per Layer
Layer 1 (84→256)	0.3 ns	5 ns	7 ns	**12.3 ns**
Layer 2 (256→256)	0.3 ns	5 ns	7 ns	**12.3 ns**
Layer 3 (256→256)	0.3 ns	5 ns	7 ns	**12.3 ns**
Output (256→4)	0.3 ns	5 ns	7 ns	**12.3 ns**

Total Inference Latency (1 sample): ∼49.2 ns

4. Including Crossbar Programming Time

Crossbar programming latency is only relevant when weights are changed (e.g., during training or layer updates):

- 4 arrays × 100 μs = **400 μs total one-time reprogramming latency**
- Not applicable during inference (weights are static)

5. Summary Table

Component	Value
Crossbar latency (per tile)	∼0.3 ns
DAC delay (avg)	∼5 ns
ADC delay (avg)	∼7 ns
Total latency per layer	∼12.3 ns
Total latency for 4 layers	**∼49.2 ns**
Programming latency (once)	∼400 μs

6. References

- Texas Instruments Forum on Low-Latency Converters: `https://e2e.ti.com/support/data-converters-group/data-converters/f/data-converters-forum/546029`
- Memristor Programming Latency (Nature Communications, 2023): `https://www.nature.com/articles/s41467-023-44620-1`