

$$\text{thrpt}_{\text{app}} = \min \left( \begin{array}{l} \frac{\text{op}}{\text{sec}} \\ \frac{\text{op}}{\text{byte}_{\text{off}}} \text{BW}_{\text{off}} \\ \frac{\text{op}}{\text{byte}_{\text{on}}} \text{BW}_{\text{on}} \\ \frac{\text{op}}{\text{byte}_{\text{net}}} \text{BW}_{\text{net}} \end{array} \right)$$

**Throughput**

**Proportional To**

Compute

$$\text{op} \sim P, D$$

Off-chip  
Memory

$$\text{byte}_{\text{off}} \sim P, D$$

On-chip  
Memory

$$\text{byte}_{\text{on}} \sim P, D$$

On-chip  
Network

$$\text{byte}_{\text{net}} \sim P^2, D$$

Application-specific

Hardware-specific

P: Parallelization factor

D: Pipelining depth