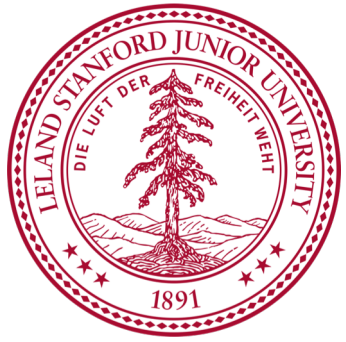


Taurus: An Intelligent Data Plane

Tushar Swamy

Alexander Rucker, Muhammad Shahbaz,

Neeraja Yadwadkar, Yaqi Zhang, and Kunle Olukotun

Managing networks is hard!

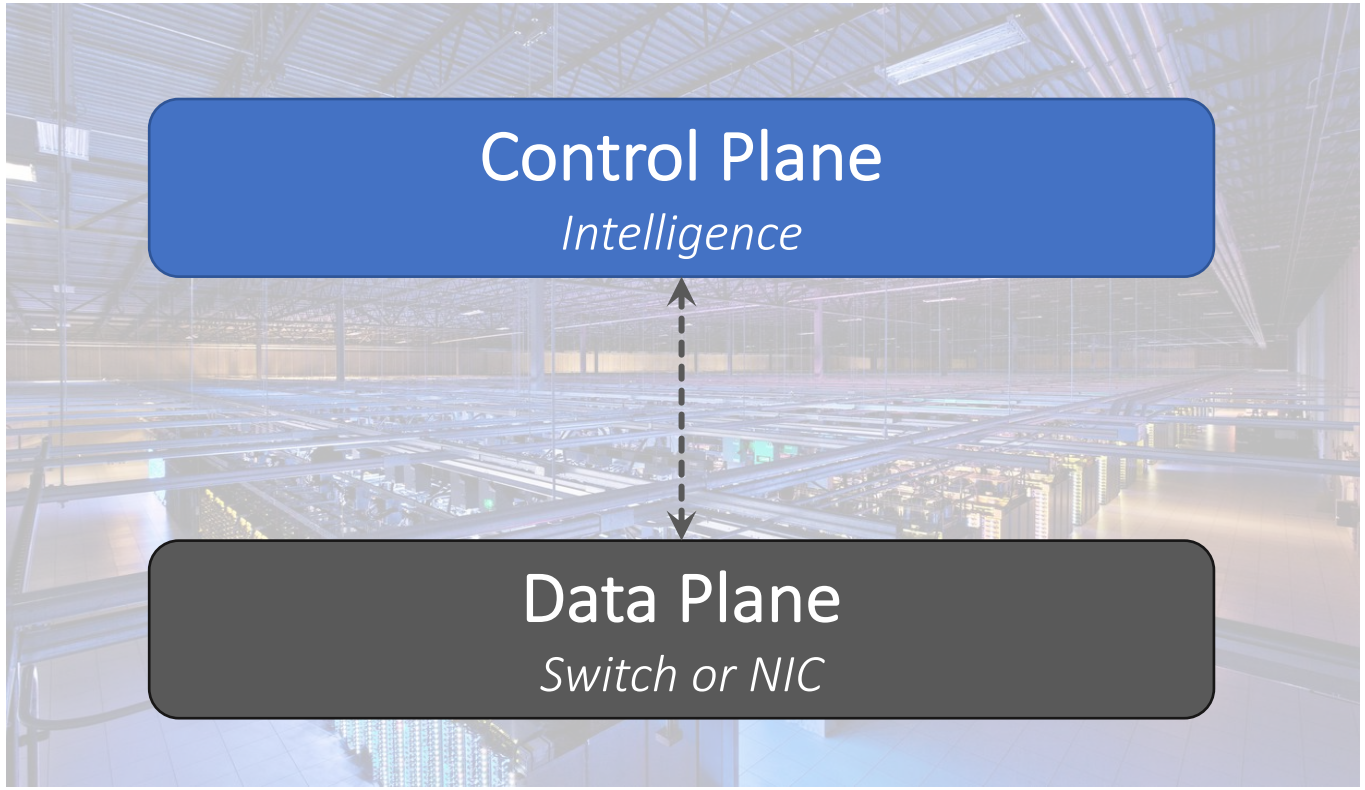
amazon Google Microsoft

Cloud Computing

Internet of Things (IoT)

Augmented and Virtual Reality (AR/VR)

Approaches to manage networks are …



amazon Google Microsoft

Slow but *intelligent*

OR

Fast yet *dumb*

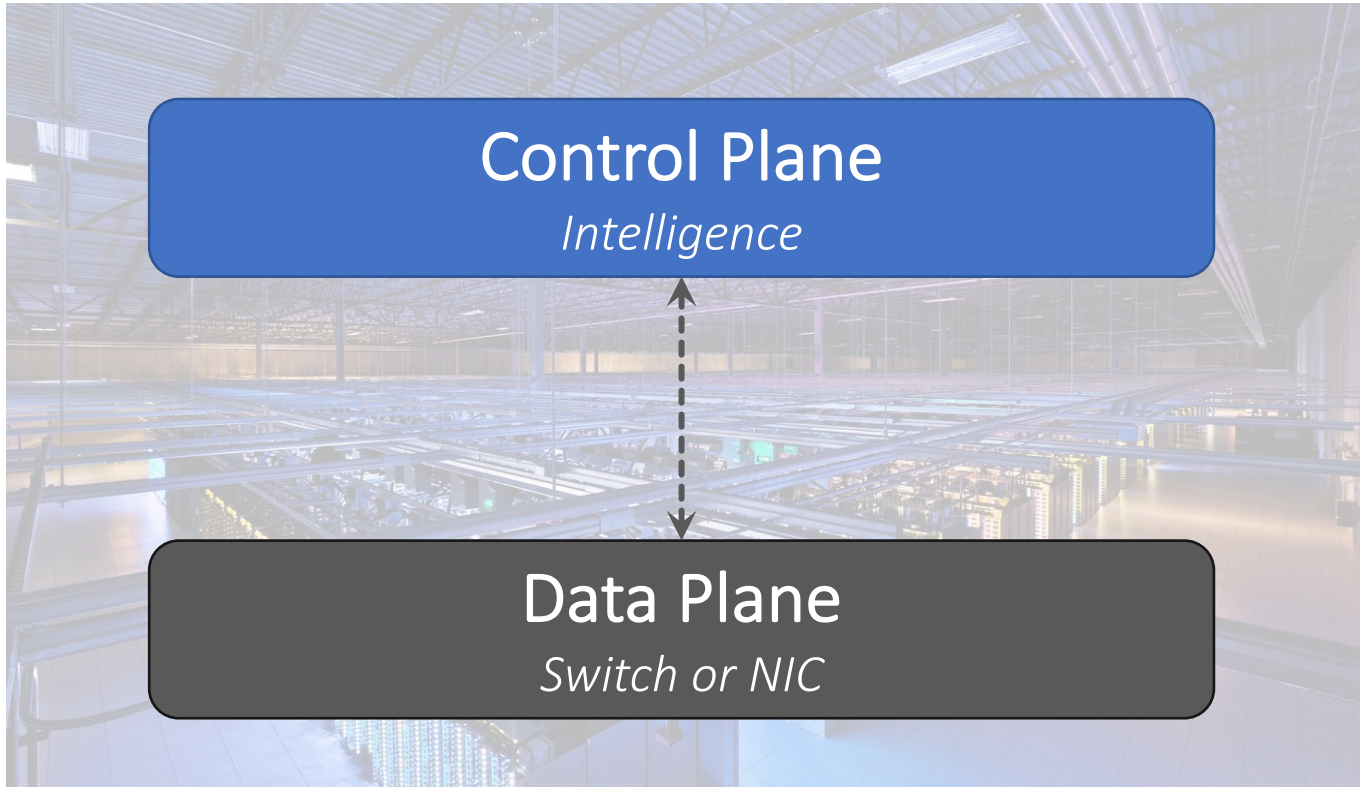
Approaches to manage networks are …

Examples:
- Congestion control
- Load balancing (ECMP, RSS)
- Queue scheduling

Characteristics:
- Operates on packet scale
- Low latency: ns
- High throughput: Tbps
- Uses heuristics: hash, …

Fast yet ***dumb***

Approaches to manage networks are …



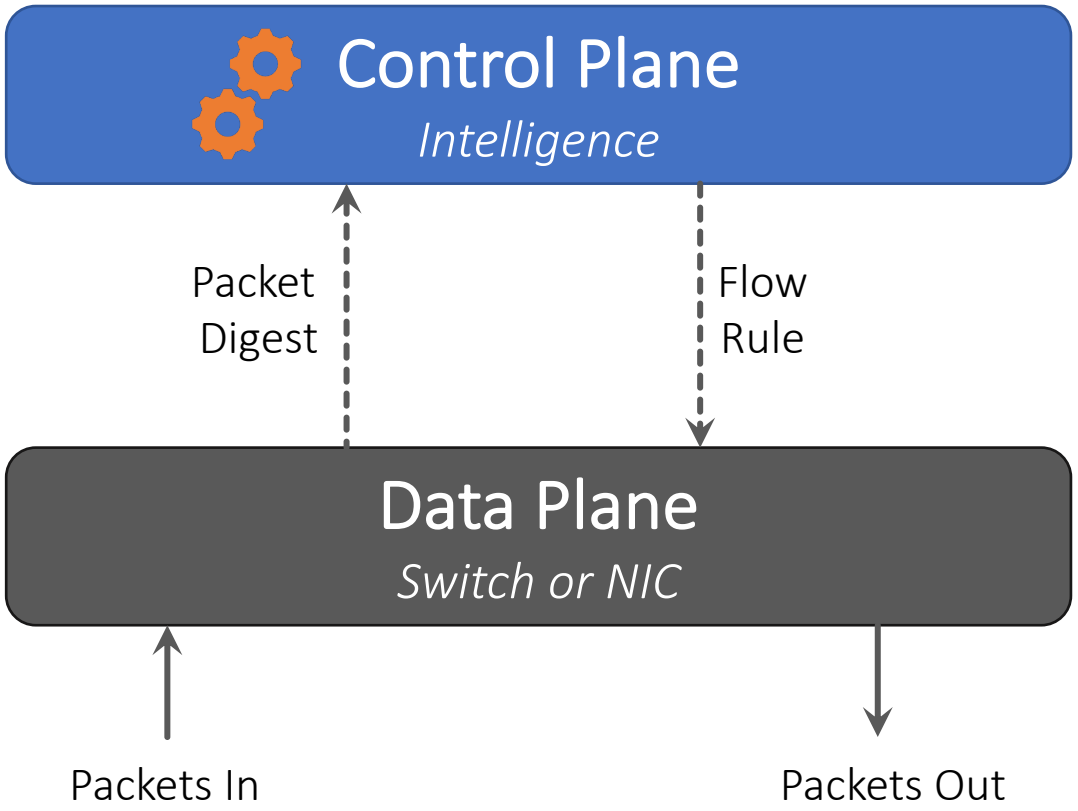
amazon Google Microsoft

Slow but *intelligent*

OR

Fast yet *dumb*

Approaches to manage networks are …



Slow but ***intelligent***

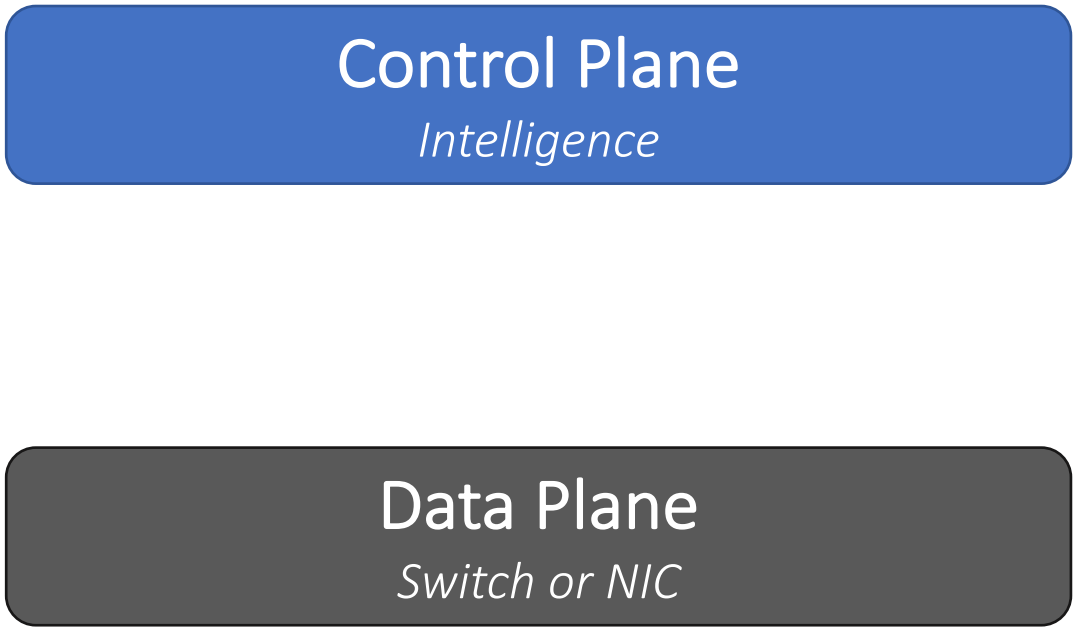
Examples:

- Anomaly detection
- Automation
- Recommendation

Characteristics:

- Can do machine learning
- Operates on flows
- Millisecond latency
- Low throughput

Approaches to manage networks are …

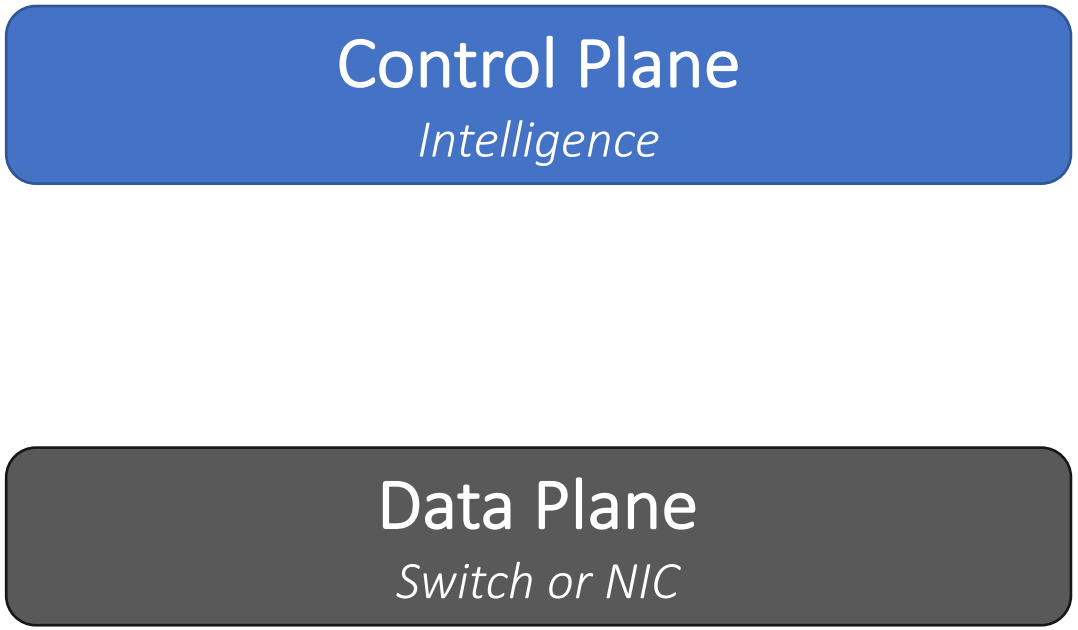


Slow but ***intelligent***

OR

Fast yet *dumb*

Approaches to manage networks are …

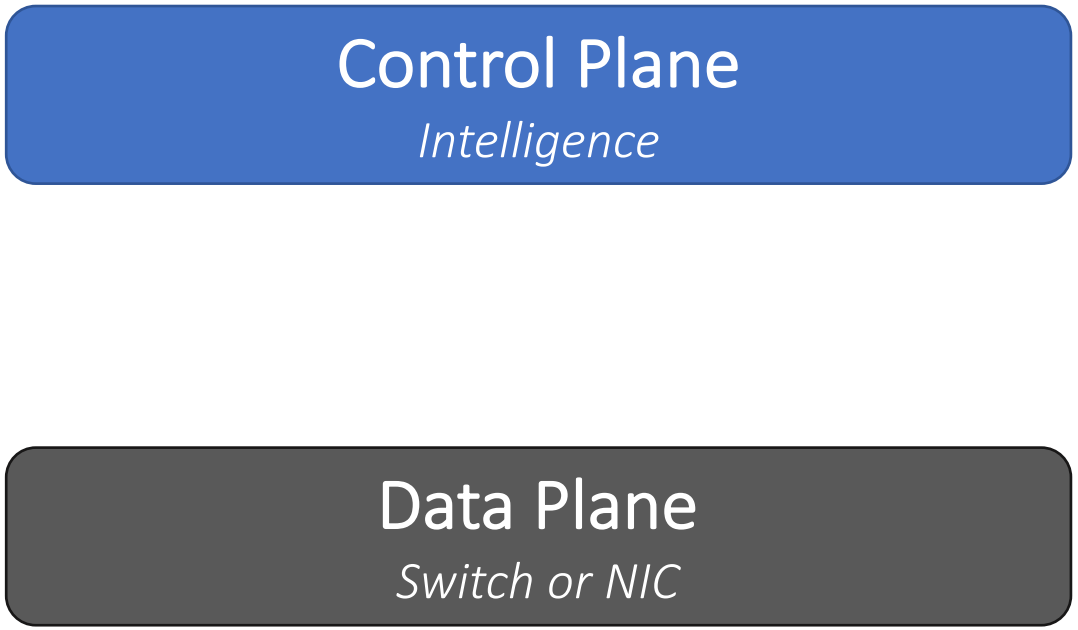


~~*Slow*~~ but ***intelligent***

OR

Fast yet ~~***dumb***~~

Network management should be …



Fast and ***intelligent***

What does “Intelligence” mean?

- Networks are becoming autonomous (***Self Driving Networks***)

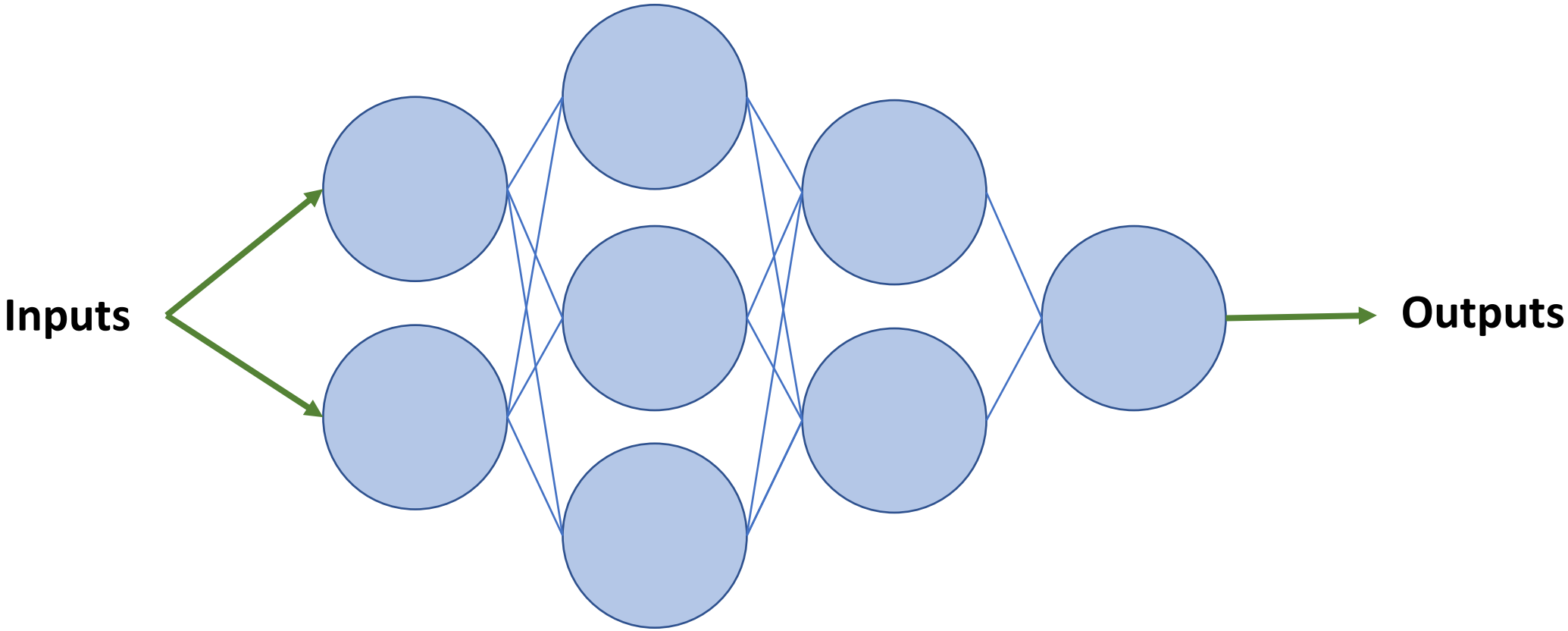
- **Machine learning (ML)** will play a key role in the future of networks [1,2,3]

Security

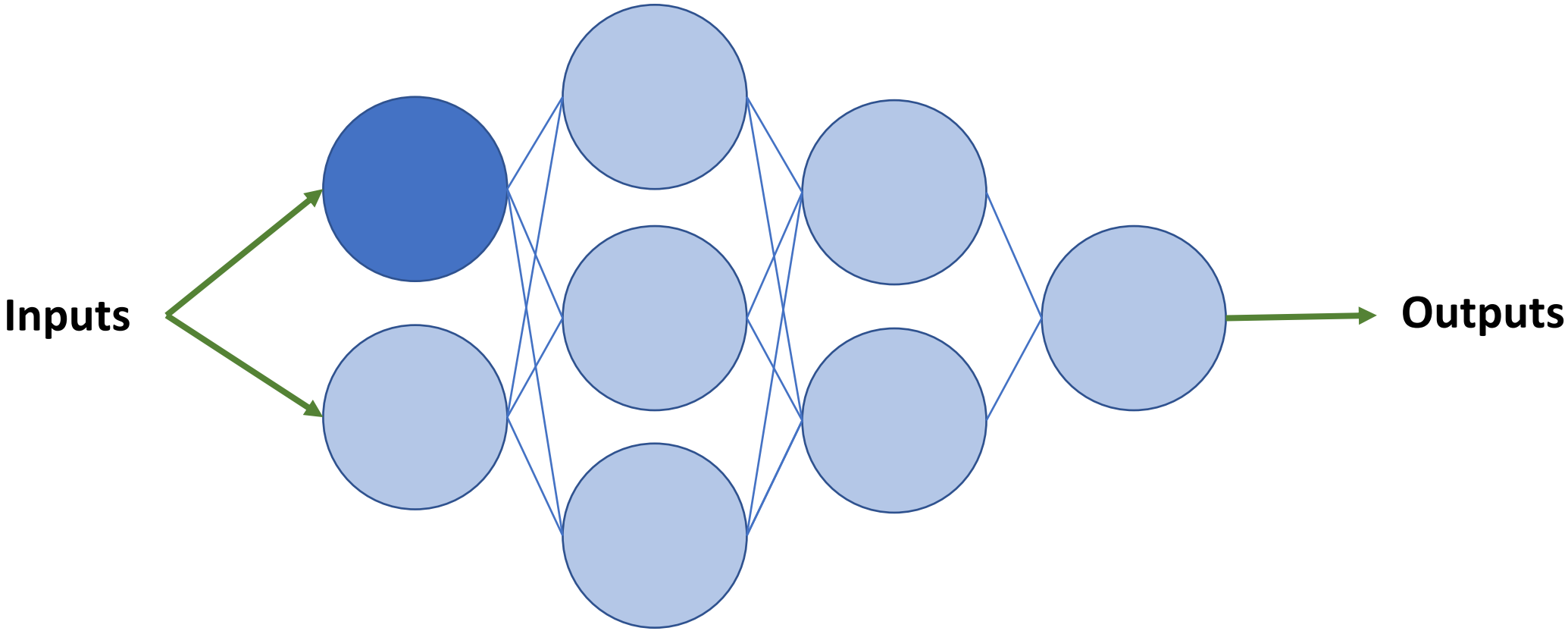
Control

Analytics

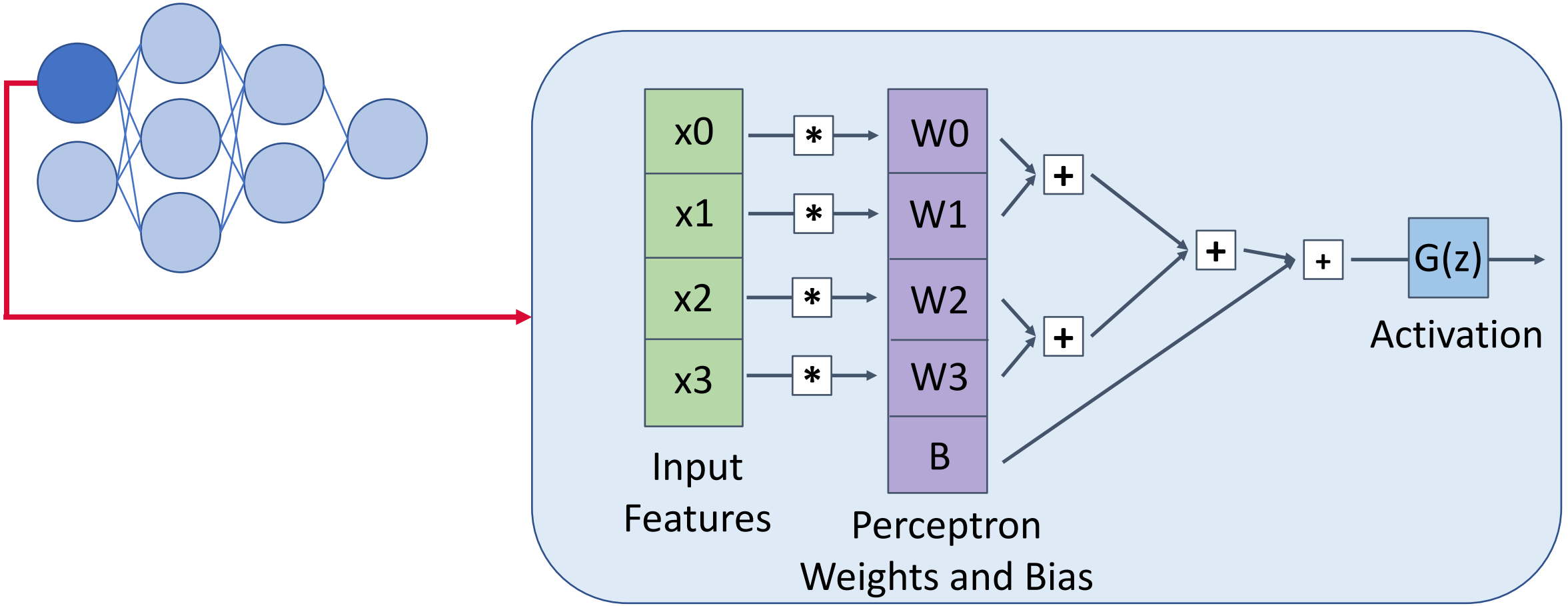
ML Inference: Neural Networks



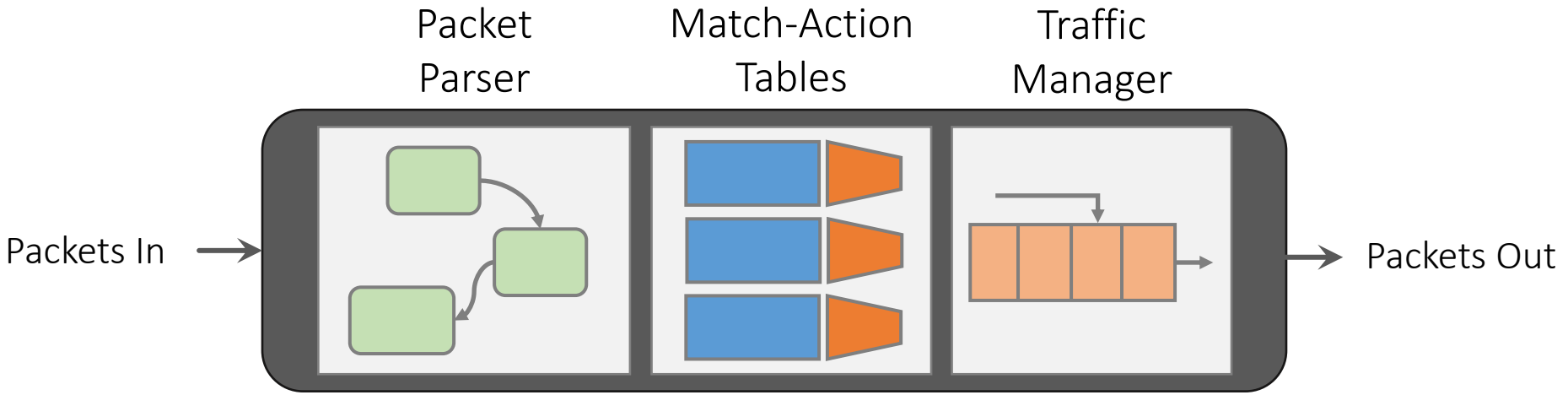
ML Inference: Neural Networks



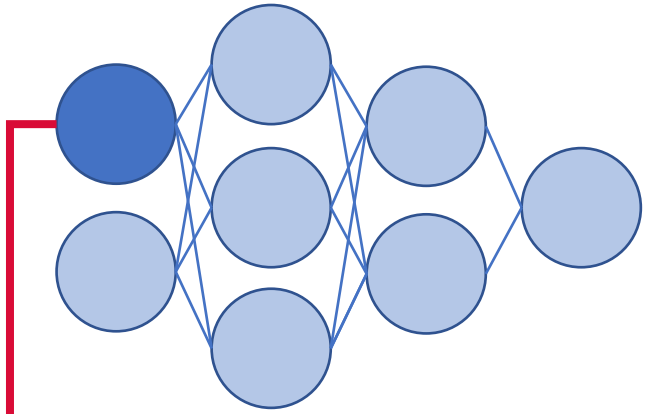
ML Inference: Single Neuron

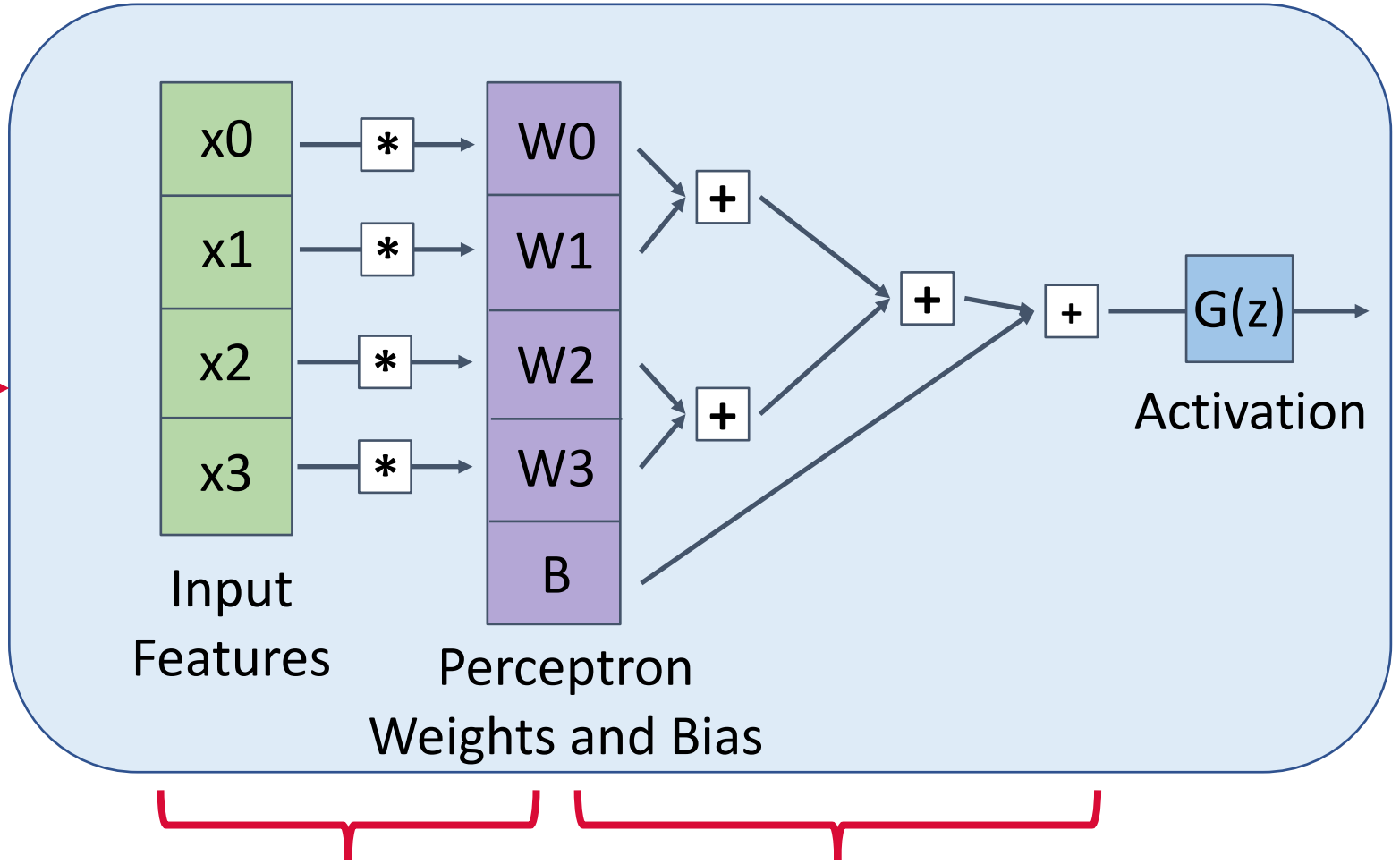


Can match-action tables perform ML inference?

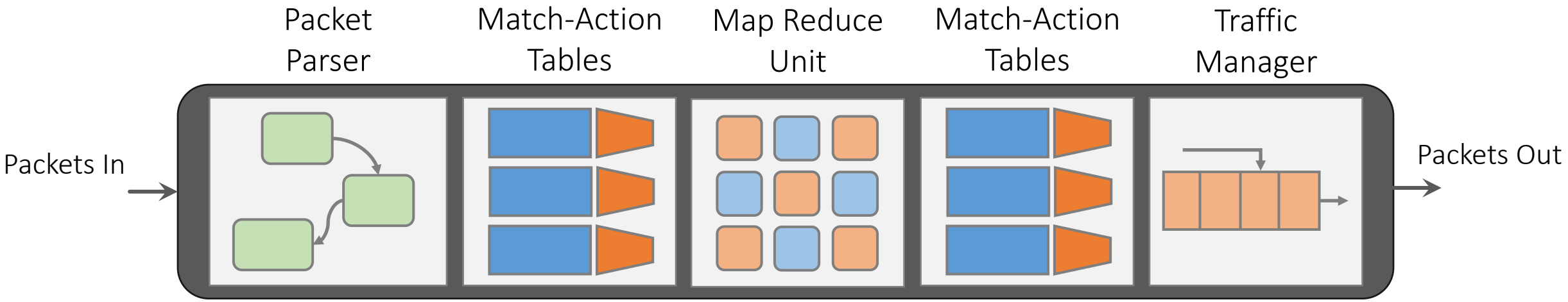


ML Inference: Single Neuron

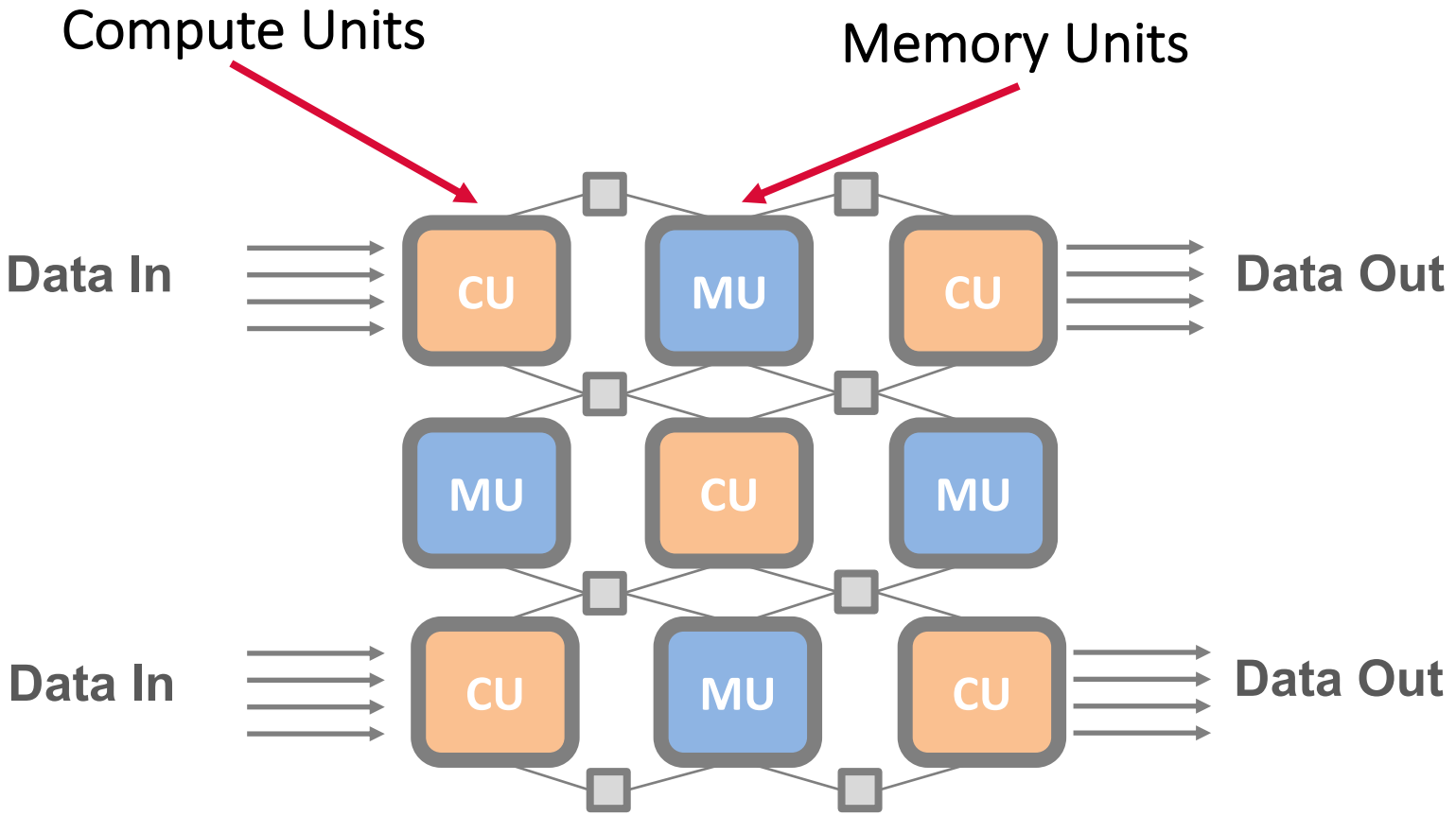




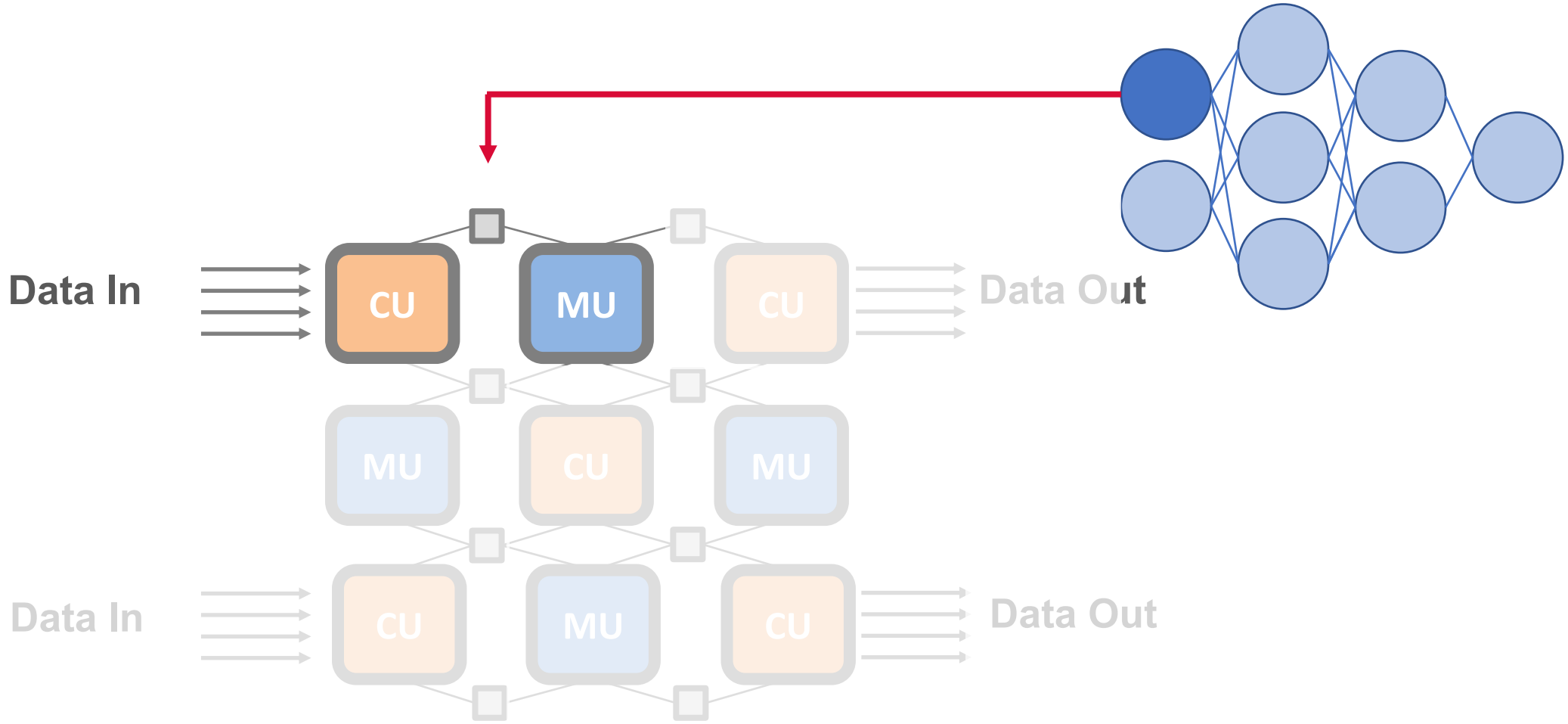
Taurus: An Intelligent Data Plane



Building blocks of Taurus: Map-Reduce

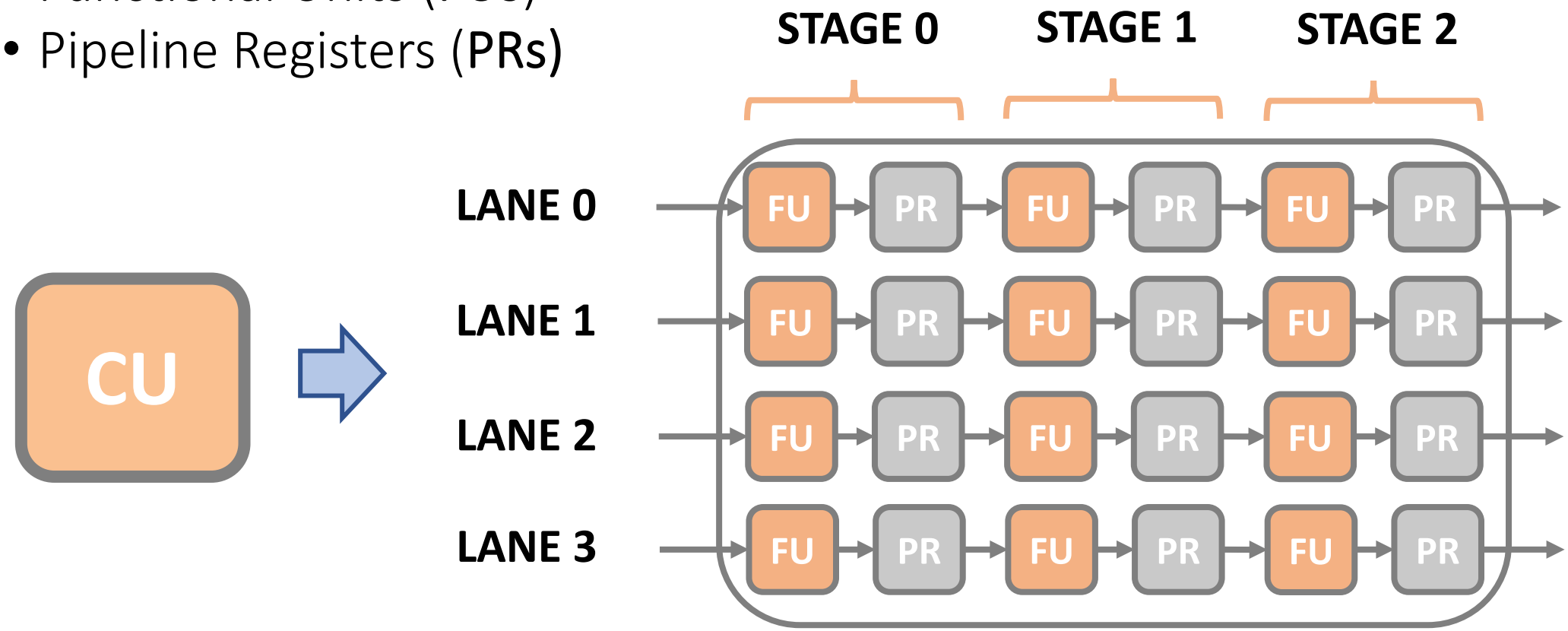


Building blocks of Taurus: Map-Reduce

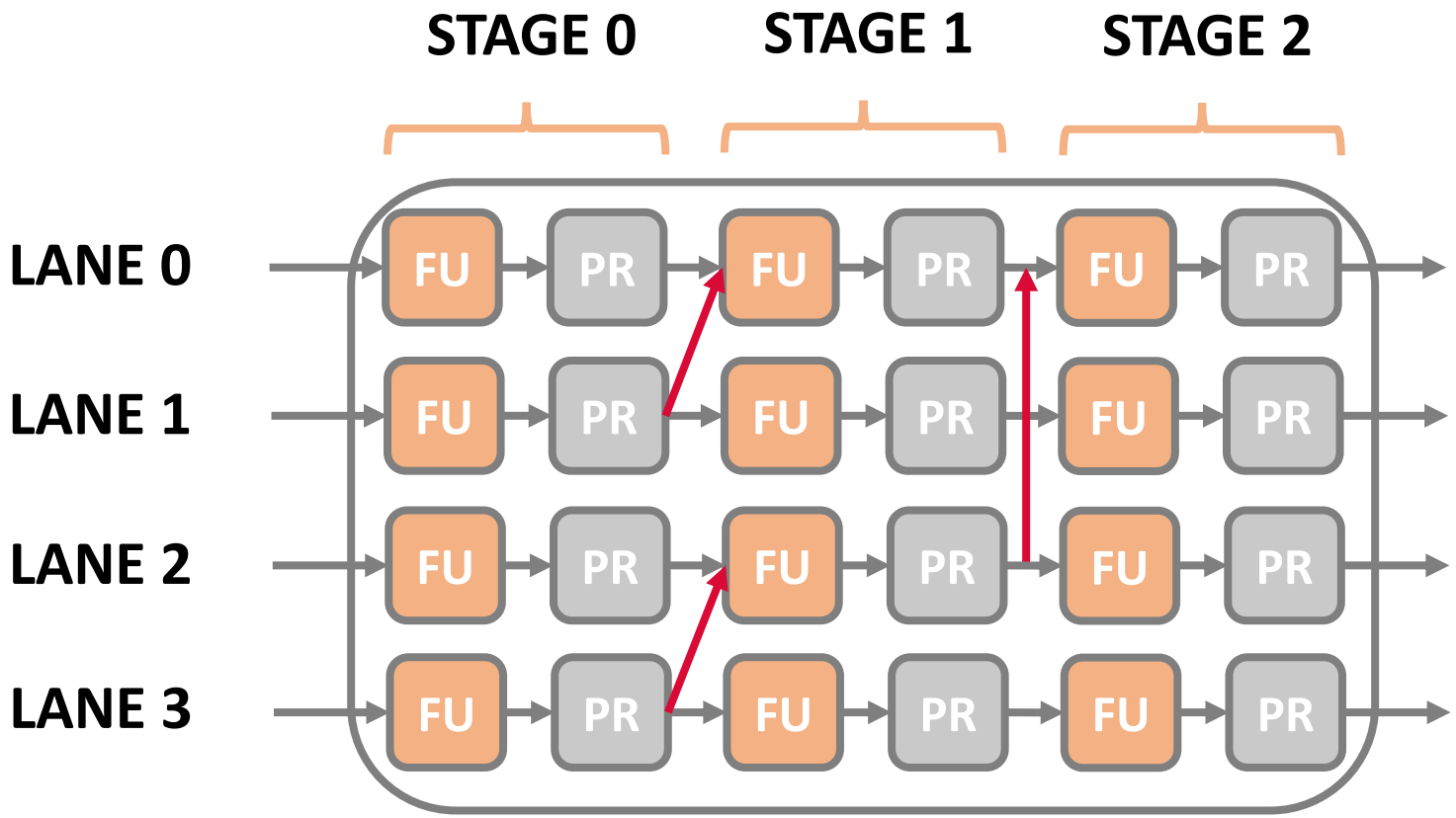


Compute unit design

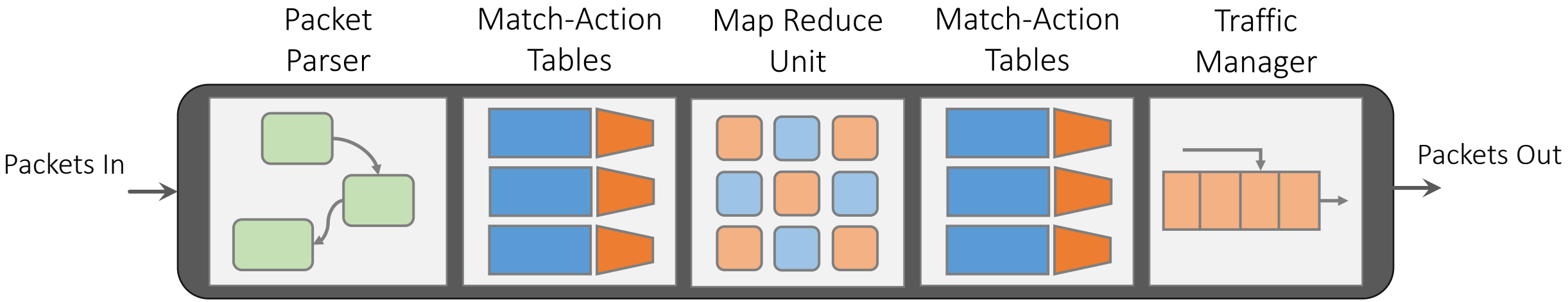
- Taurus **CUs** are array based:
  - Functional Units (**FUs**)
  - Pipeline Registers (**PRs**)



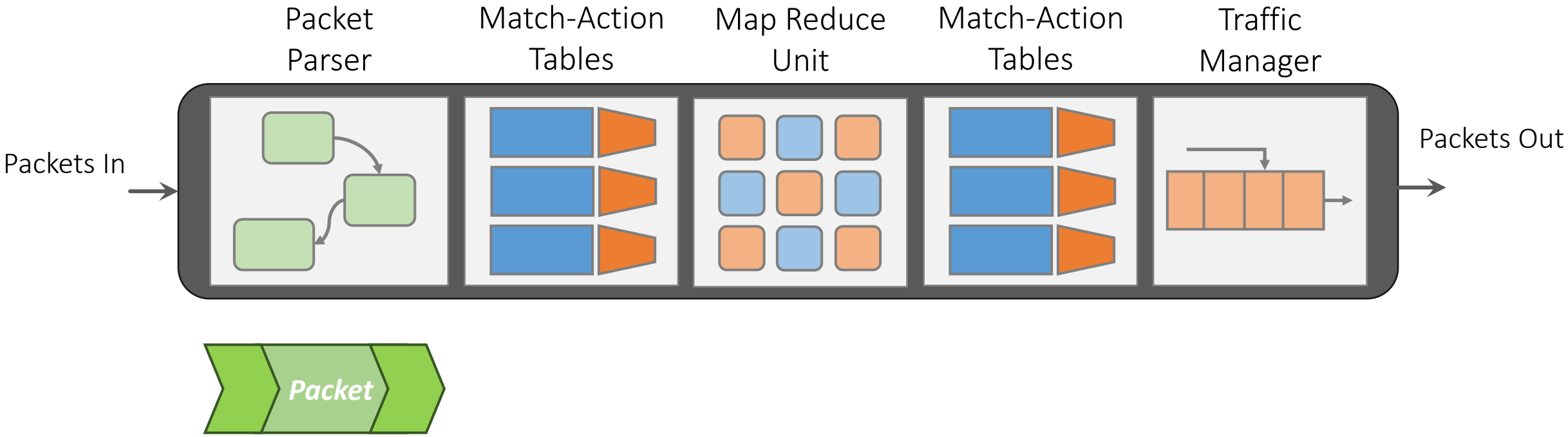
Reduction network condenses vectors to scalars



Taurus: An Intelligent Data Plane



Example: Anomaly Detection



Read local features (e.g., IP address)	Retrieve out of network events (e.g., failed logins per IP)	Apply learned anomaly detection	Select a port or action (e.g., drop if score == 1)	Send packet to destination

Evaluation: Anomaly Detection in Switches

- Taurus examines *every* packet at *every* switch

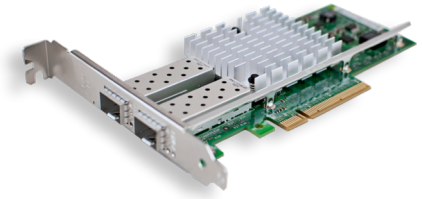
- Additional latency is less than port to port latency

			Area	Power
Model	TP (GPkt/s)	Lat (ns)	+%	+%
SVM	1	68	6.1	1.1
DNN	1	362	11.7	2.0

**Overheads are calculated relative to a 300 mm^2 chip with 4 reconfigurable pipelines, each drawing an estimated 25 W*

Evaluation: Congestion Control at the NIC

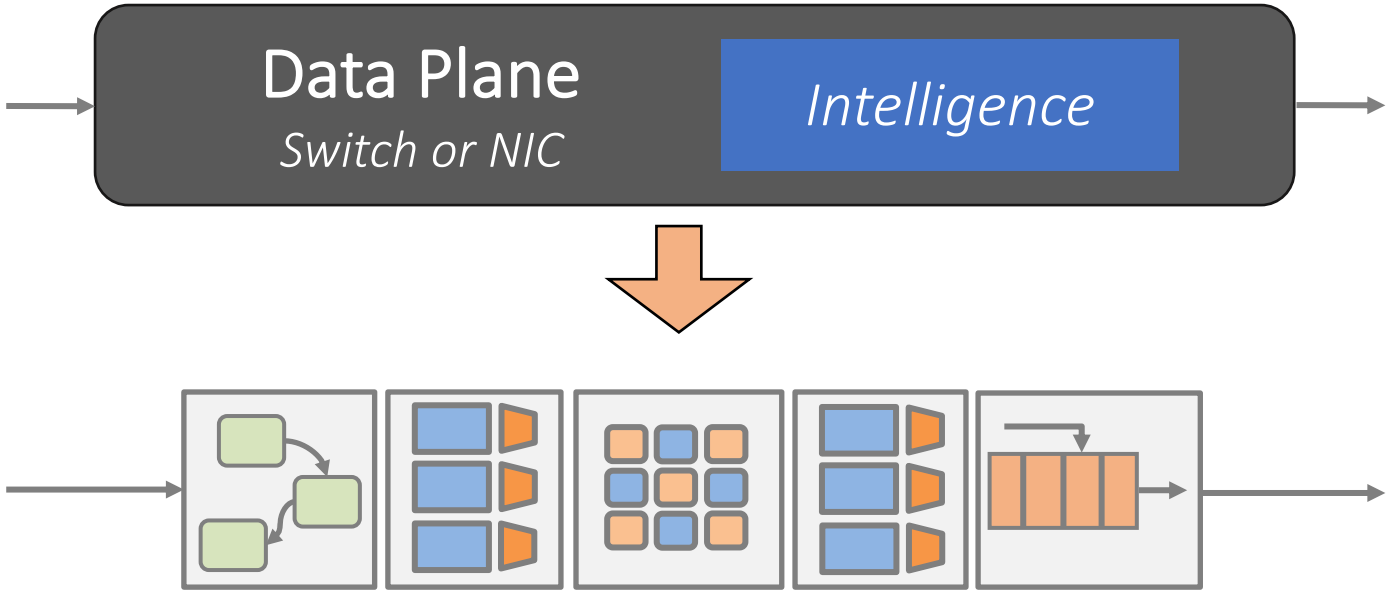
- Indigo Congestion Control LSTM Network [7]
- Taurus updates every 12.5 ns (software updates every 10 ms)



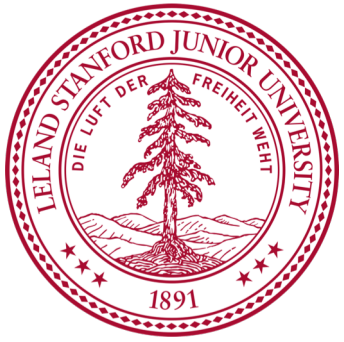
			Area	Power
Model	TP (GPkt/s)	Lat (ns)	+%	+%
LSTM	0.08	380	23.6	4.1

**Overheads are calculated relative to a 300 mm² chip with 4 reconfigurable pipelines, each drawing an estimated 25 W*

Taurus: Fast and Intelligent



- Introduces **map and reduce** to PISA
- Designed to run **ML inference** inside a data plane
- Provides **orders of magnitude improvement**



Questions?

Tushar Swamy

tswamy@stanford.edu

References

[1] Why (and How) Networks Should Run Themselves
[2] A Knowledge Plane for the Internet, SIGCOMM 2003
[3] Knowledge-Defined Networking, SIGCOMM CCR 2017
[4] Plasticine: A Reconfigurable Architecture For Parallel Patterns, ISCA 2017
[5] Deep Learning Approach for Network Intrusion Detection in Software Defined Networking, WINCOM 2016
[6] SVM for Network Anomaly Detection Using ACO Feature Subset, iSMSC 2015
[7] Pantheon: the training ground for Internet congestion-control research, ATC 2018
[8] Forwarding Metamorphosis: Fast Programmable Match-Action Processing in Hardware for SDN , SIGCOMM 2013
[9] Programmable Packet Scheduling at Line Rate , SIGCOMM 2016
[10] Design Principles for Packet Parsers, ANCS 2013