Tutorial : Network ASIC

Nancy Shaw

NANOG 93

February 5, 2025

JUNIPER NETWORKS | Driven by Experience™

- Network ASIC Basics (slide2-5)
- Packet Processing Architectures (slide7-10)
- Buffer Architectures (slide12-24)
- ASIC Development Challenges (slide26-30)

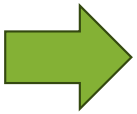
Terminology

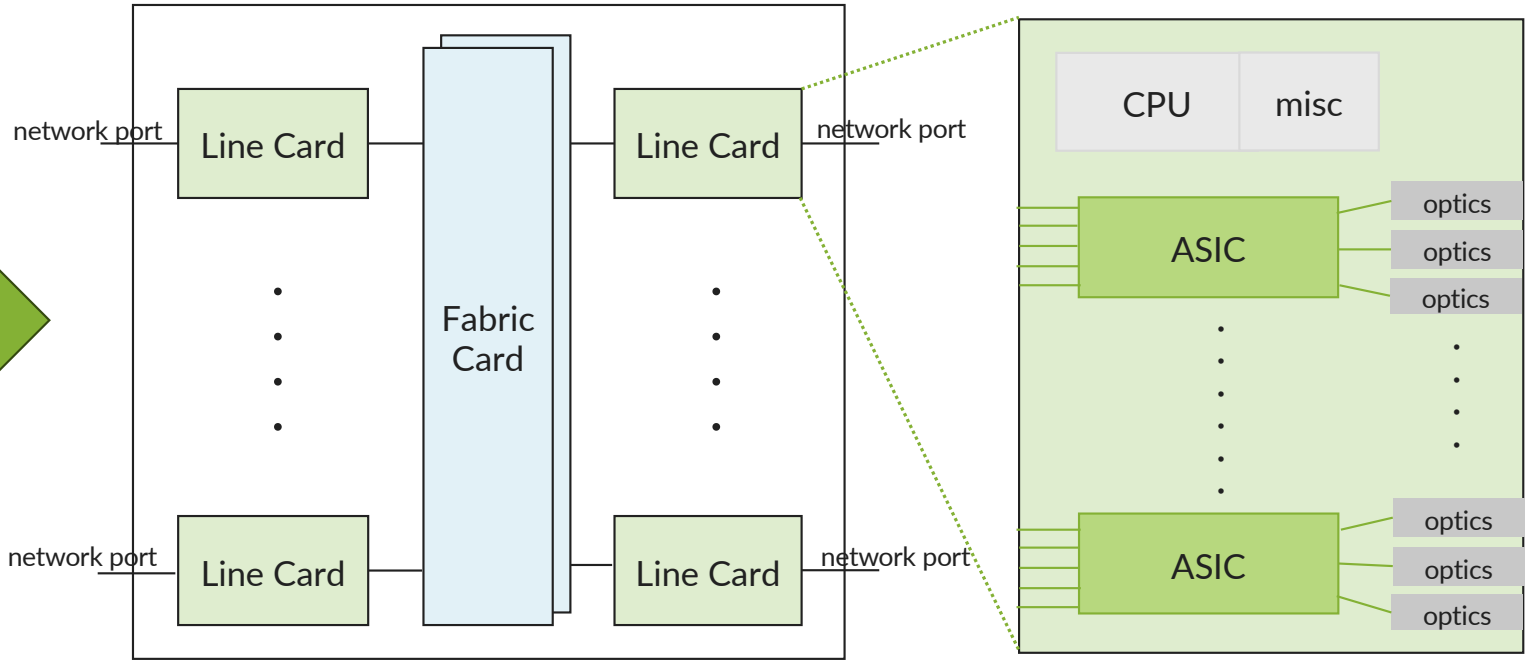
- ***ASIC*** - **A**pplication **S**pecific **I**ntegrated **Ci**rcuit

- ***chip*** – same as ASIC

- ***Network ASIC*** – custom designed chip used in networking applications; packet forwarding in a network

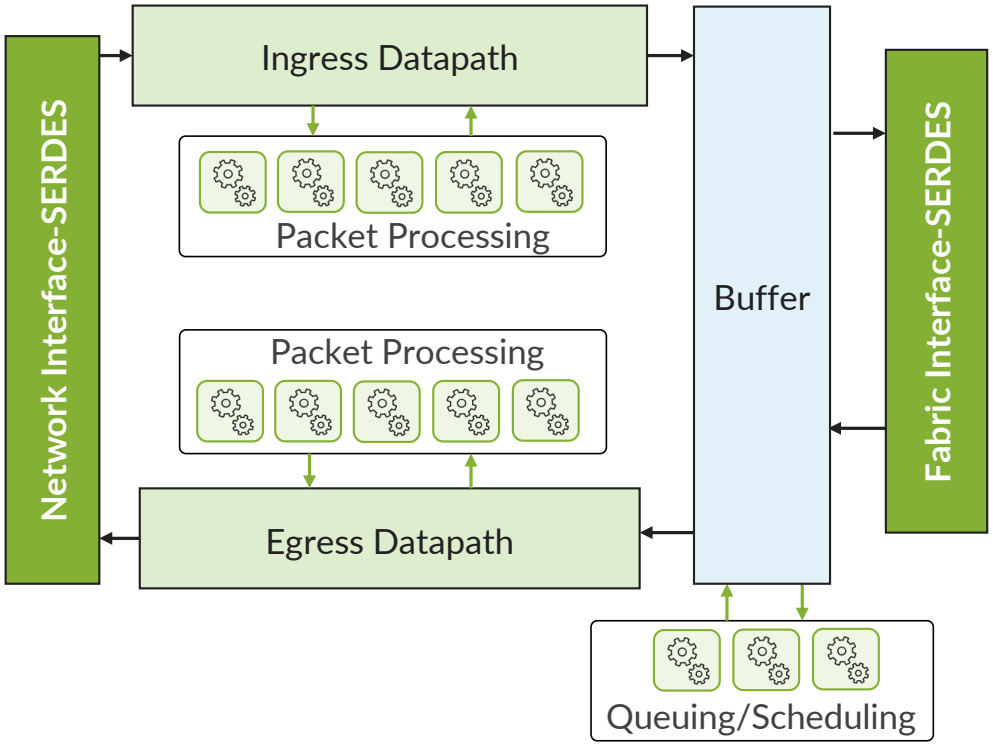
Network ASIC In The Router





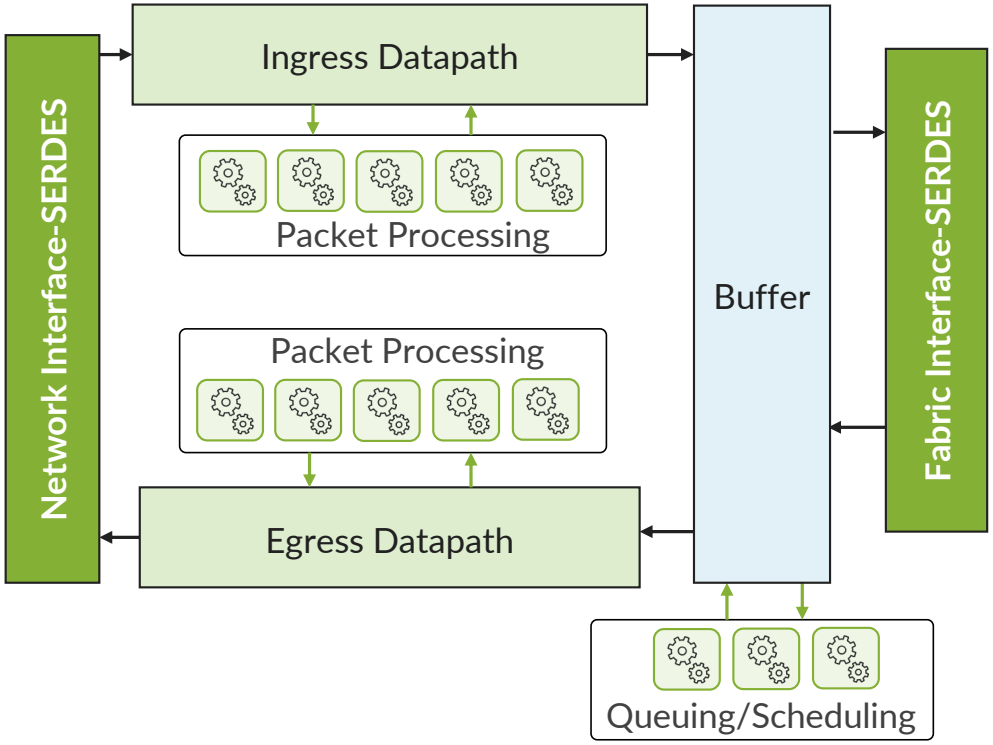
Components of a Network ASIC

- Serializer/Deserializer (SERDES)
  - Converts serial data to parallel data and vice versa
- Network Interface
  - Manages framing, error correction, alignment, etc according to Ethernet standards
- DataPath - formats data for efficient transfer



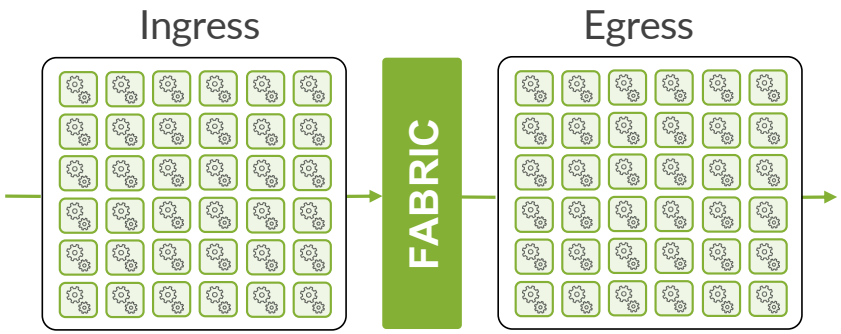
Components of a Network ASIC

- Packet Processing
  - Parsing, Classification, Tunnel, Filtering, Route Lookup, Rewrites, Encapsulations
- Buffer - provides packet storage during periods of transient congestion
- Queueing/Scheduling - ensures fair treatment of packets
- Fabric Interface – formats data, manages protocols for transfers over fabric

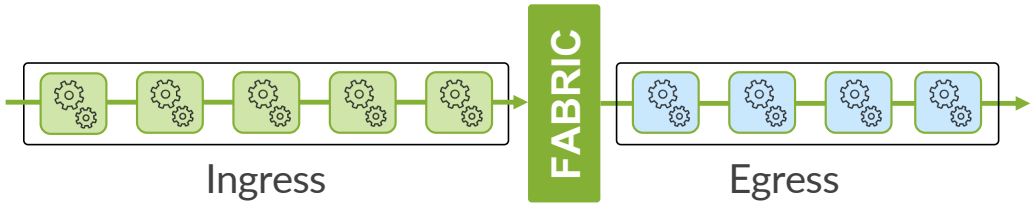


Packet Processing Architectures

Packet Processing Architectures



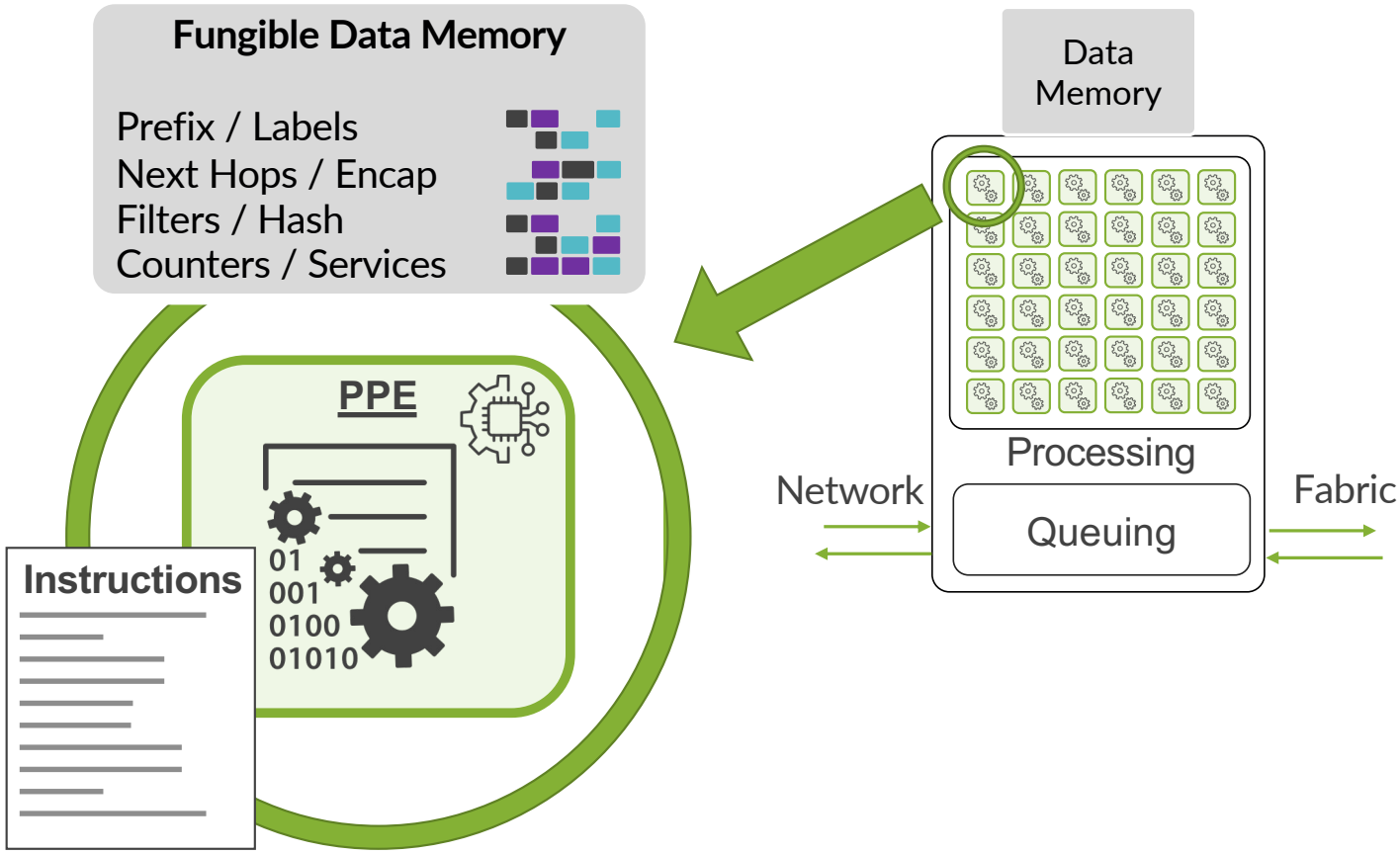
Multiple Packet Processing Engines (PPE) Architectures



Pipeline Architectures

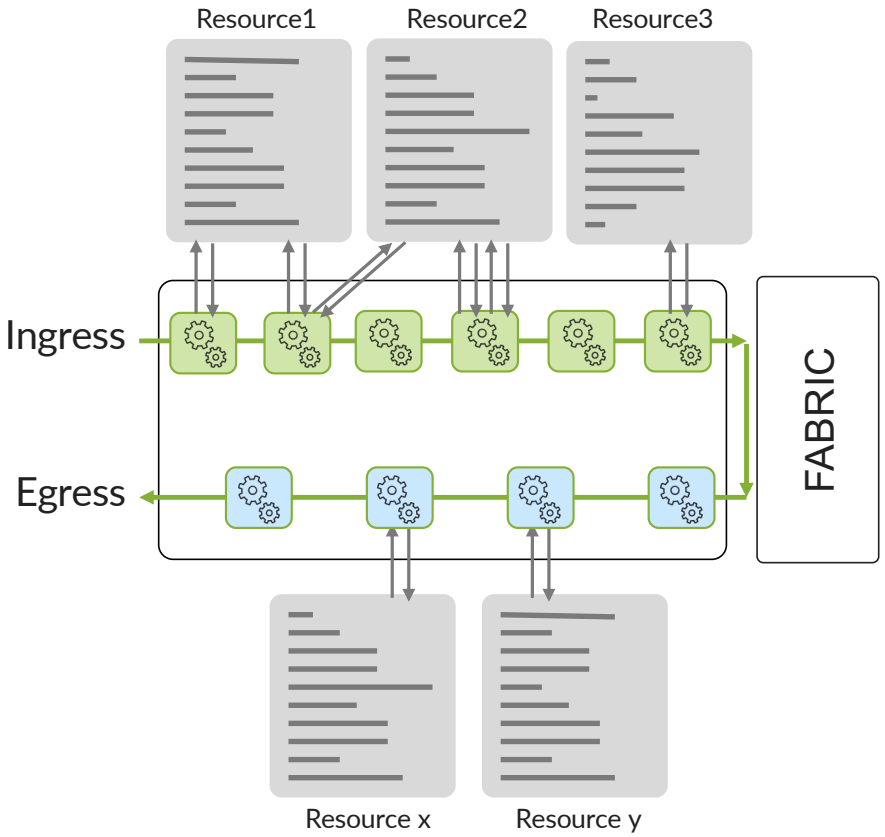
Multiple PPEs Architecture

- ASIC composed of many **Packet Processing Engines (PPE)**
- Each engine executes instructions to treat a packet header -> Programmable
- **Packet / Header stays in PPE until all instructions are executed**
  → Run-to-Completion
- Performance and Latency depends on the instructions



Pipeline Architecture

- Pipelines are made of a series of **subsequent blocks**
- Specific role for each block
  - Hardcoded functions → newly designed ASIC required for new packet types
  - Finite budget to access resources
- Predictable performance
  - Loopback / Recycle if a second pass is required



Packet Processing Architectures Pros and Cons

Multiple PPEs

- ✅ Infinite programmability
  - **All features** can be supported
- ✅ Highest level of flexibility
  - Each PPE can access a fungible memory
- ❗ Can't get the Power/Speed that pipelines offer

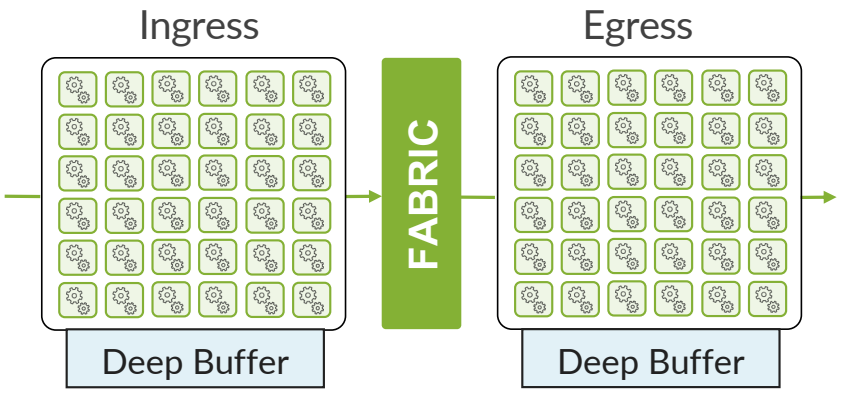
Pipeline

- ✅ Higher performance
  - Higher Speed / Lower power
- ✅ Lower latency
- ✅ Covers "most" of the major use-cases
- ✅ Can't reach the flexibility a multiple PPE architecture offers

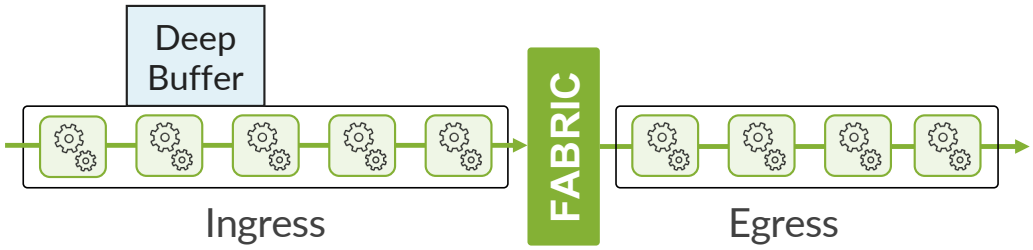
Buffer Architectures

Buffering and Queues

- Buffer provides packet storage during periods of transient congestion
- Queues store packets according to priority or Quality of Service (QoS)



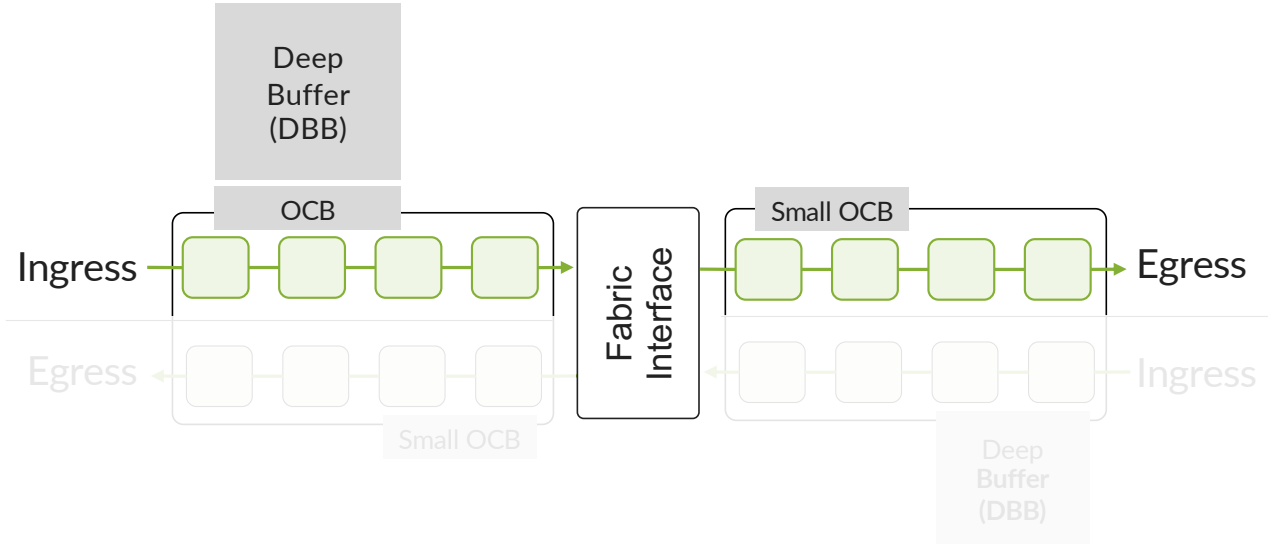
Two-Stage



Ingress-Only Single-Stage

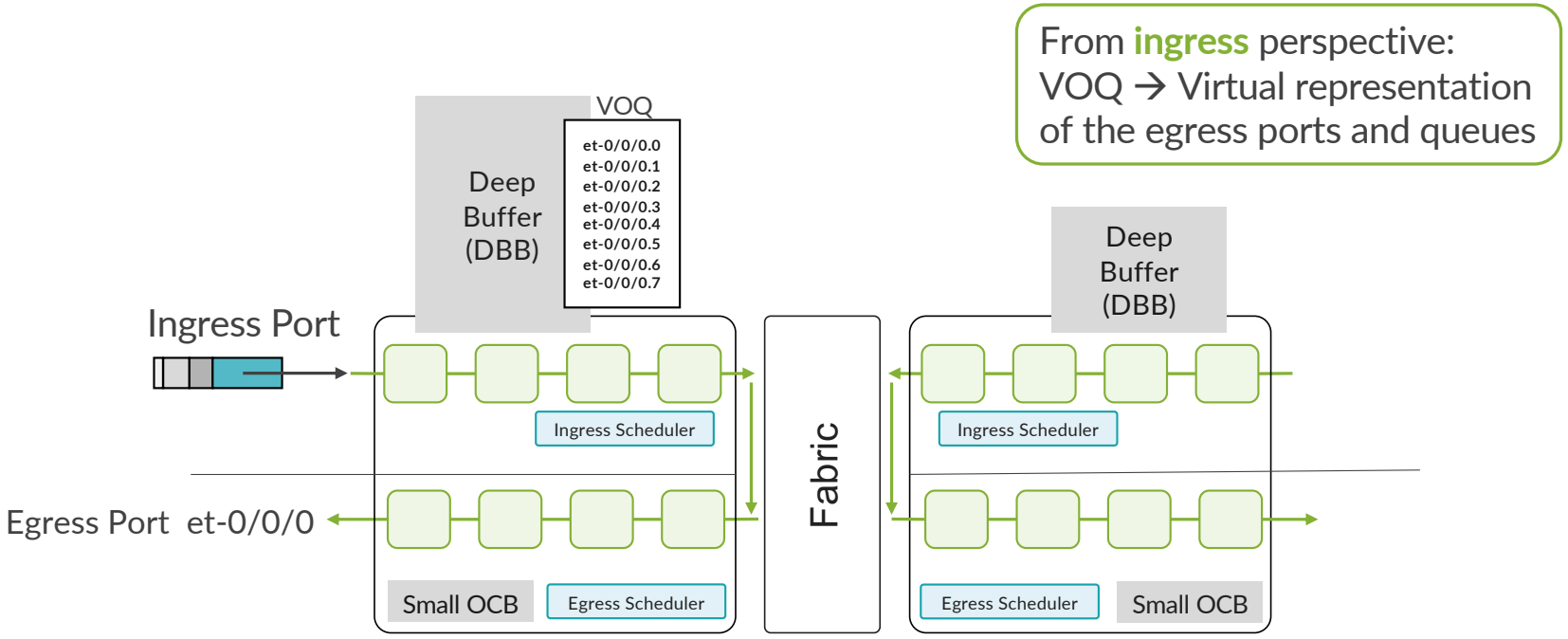
Ingress-Only Single-Stage Buffering

- Packet buffering is performed in ingress only
  - with a large Delay Bandwidth Buffer (DBB)
- Very small On-Chip Buffer (OCB) at Egress Port for packet serialization



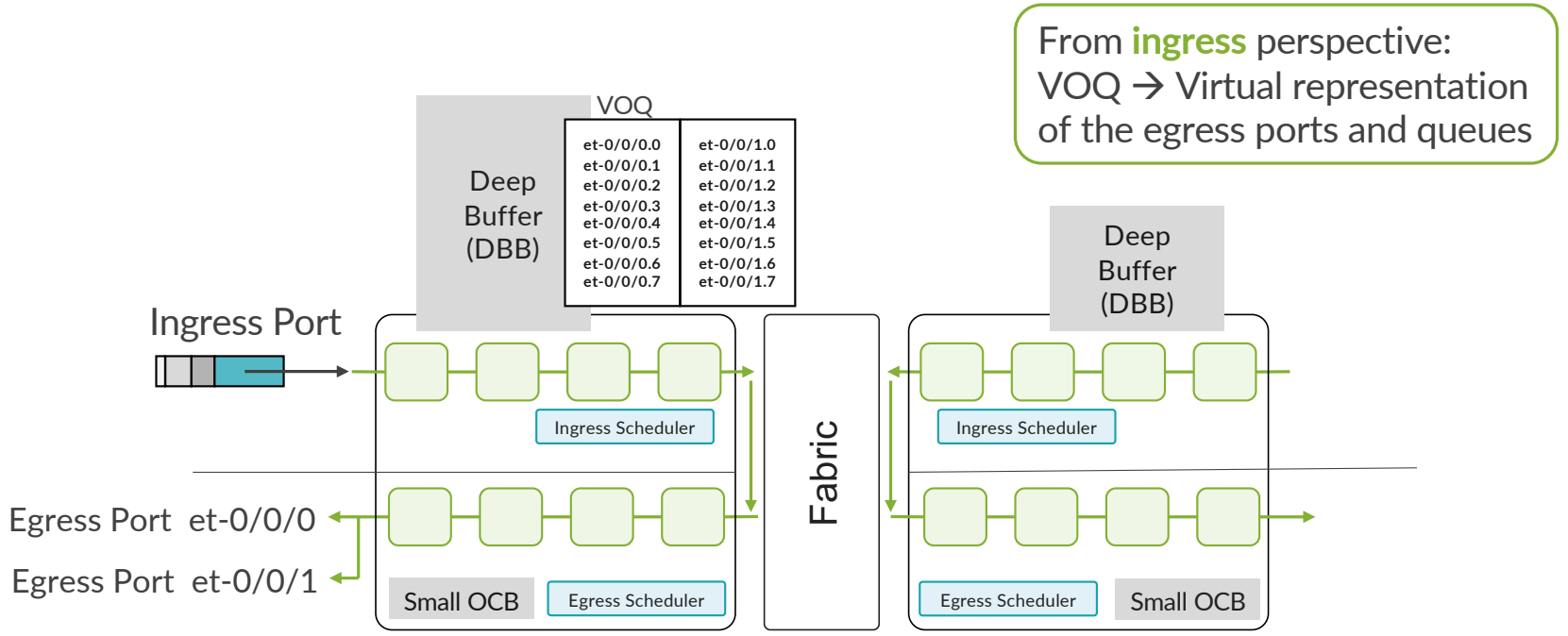
Ingress-Only Buffering: VOQ Principles (1/4)

- VOQ = Virtual Output Queue



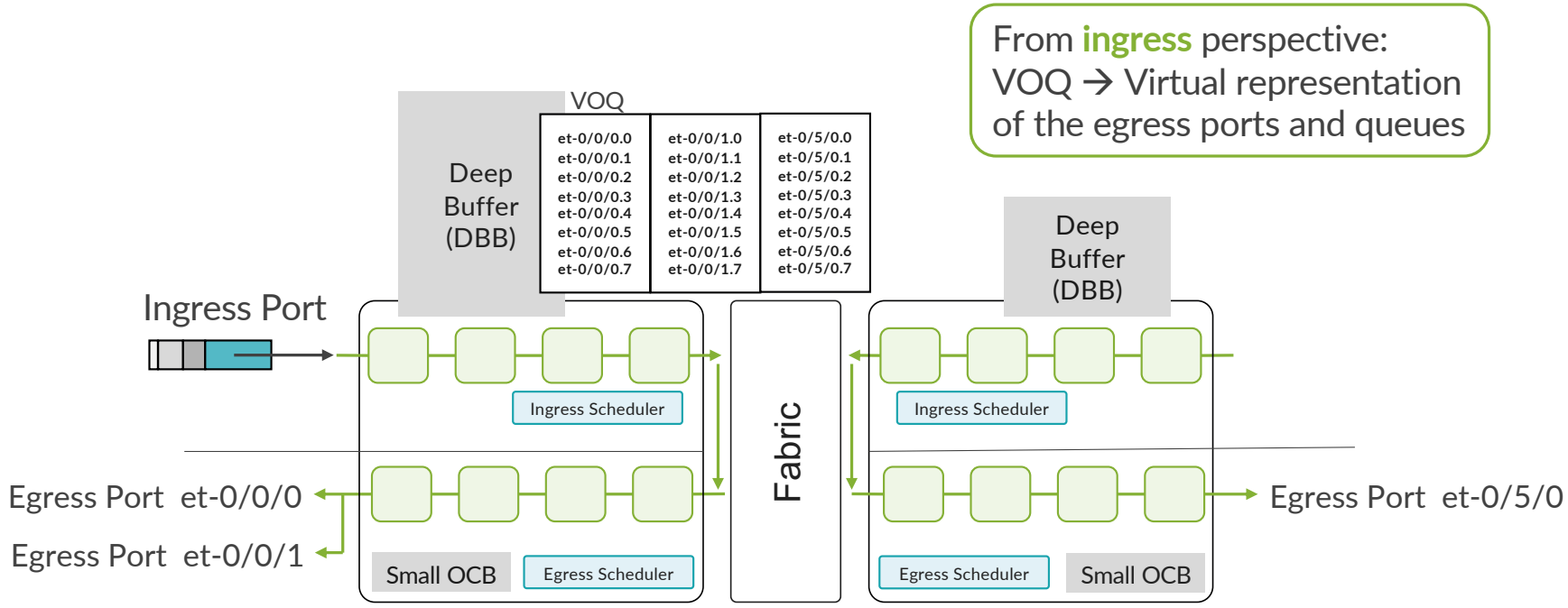
Ingress-Only Buffering: VOQ Principles (2/4)

- VOQ = Virtual Output Queue



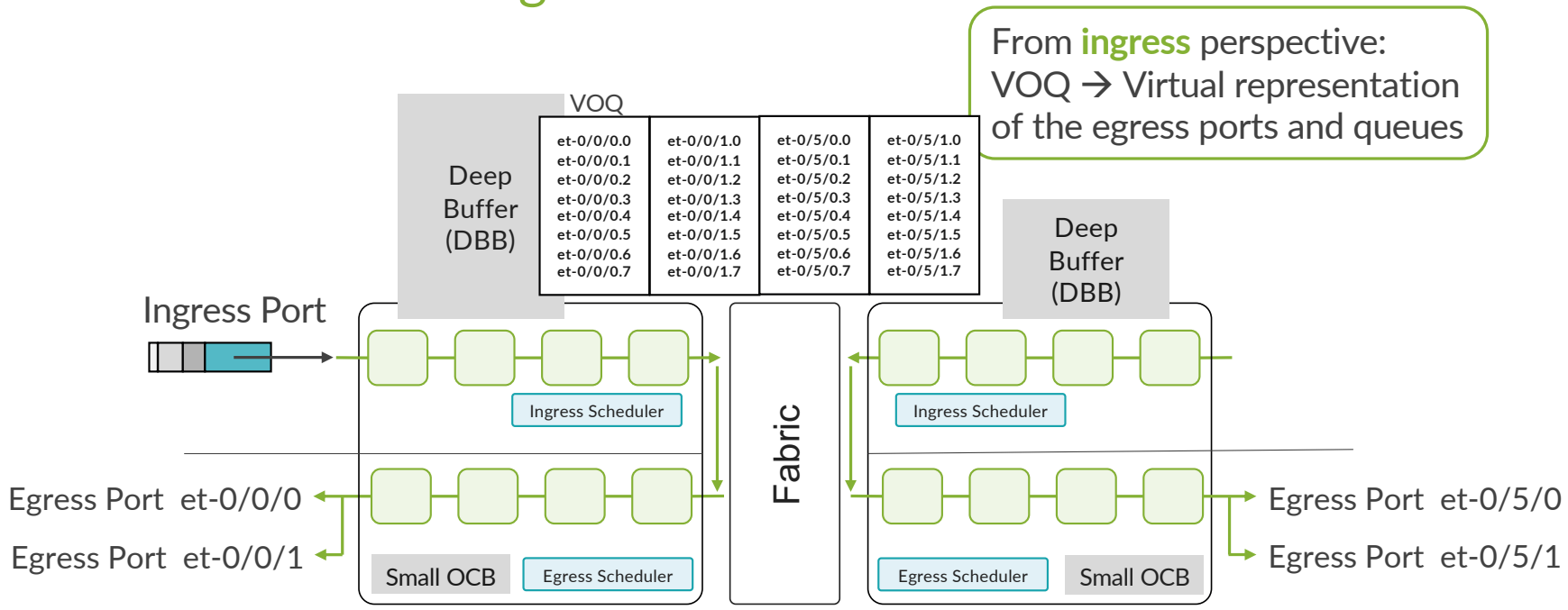
Ingress-Only Buffering: VOQ Principles (3/4)

- VOQ = Virtual Output Queue

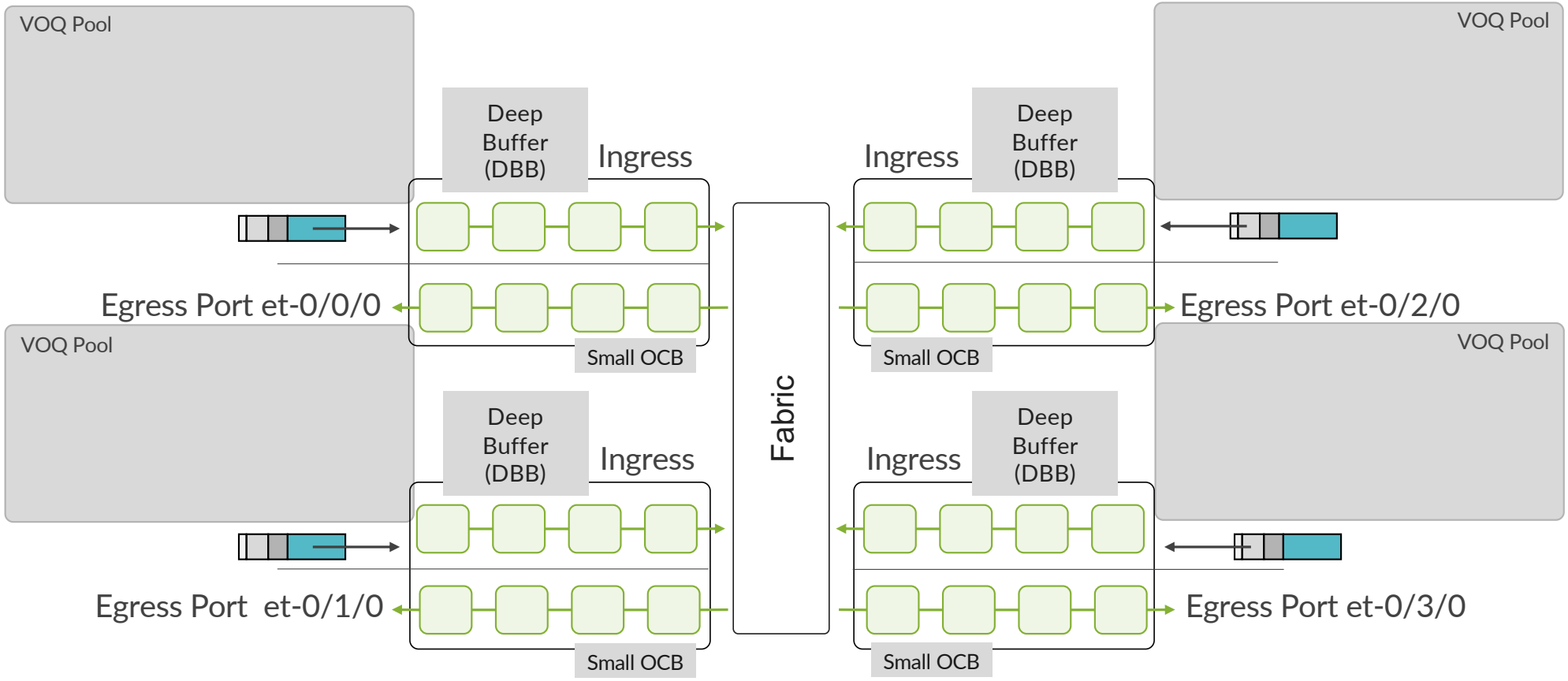


Ingress-Only Buffering: VOQ Principles (4/4)

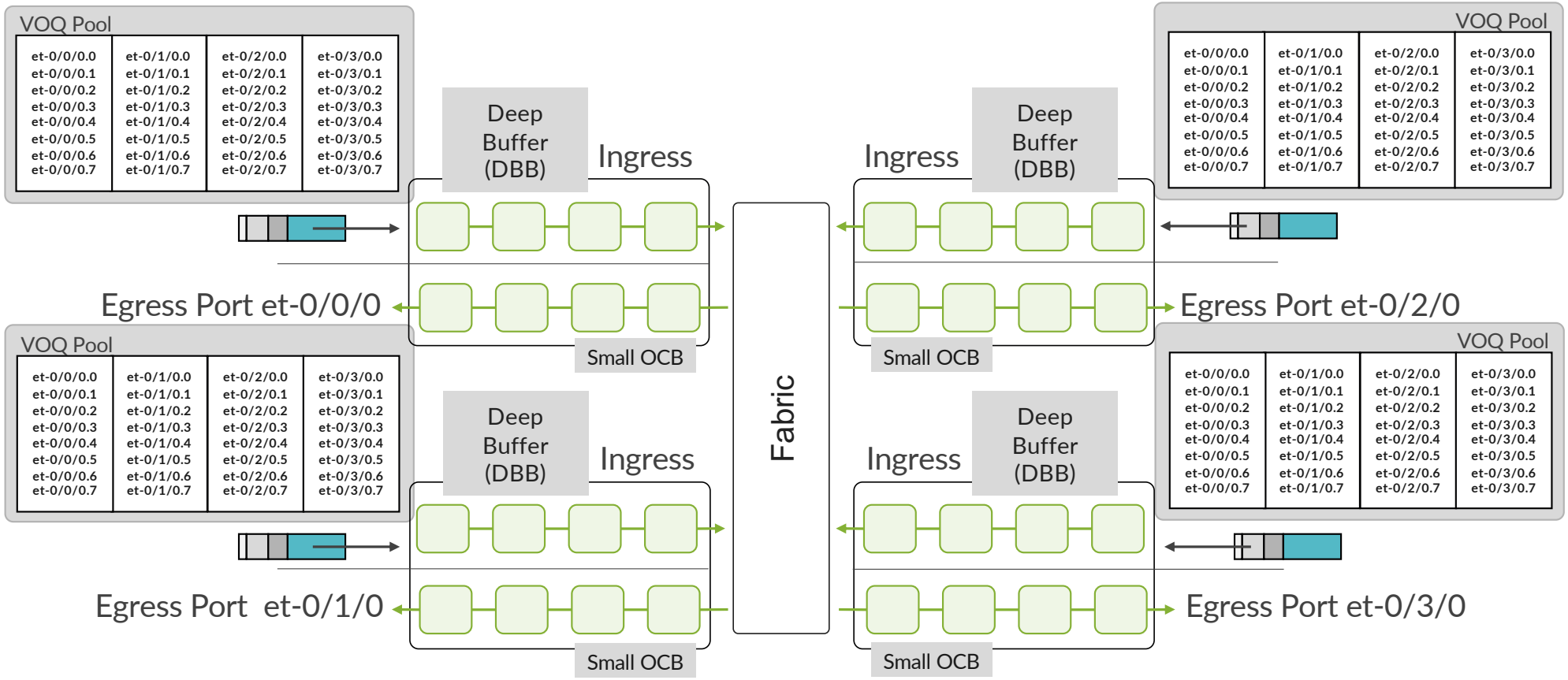
- VOQ = Virtual Output Queue
- No head of line blocking



VOQ Scale is Global / Per System (1/2)

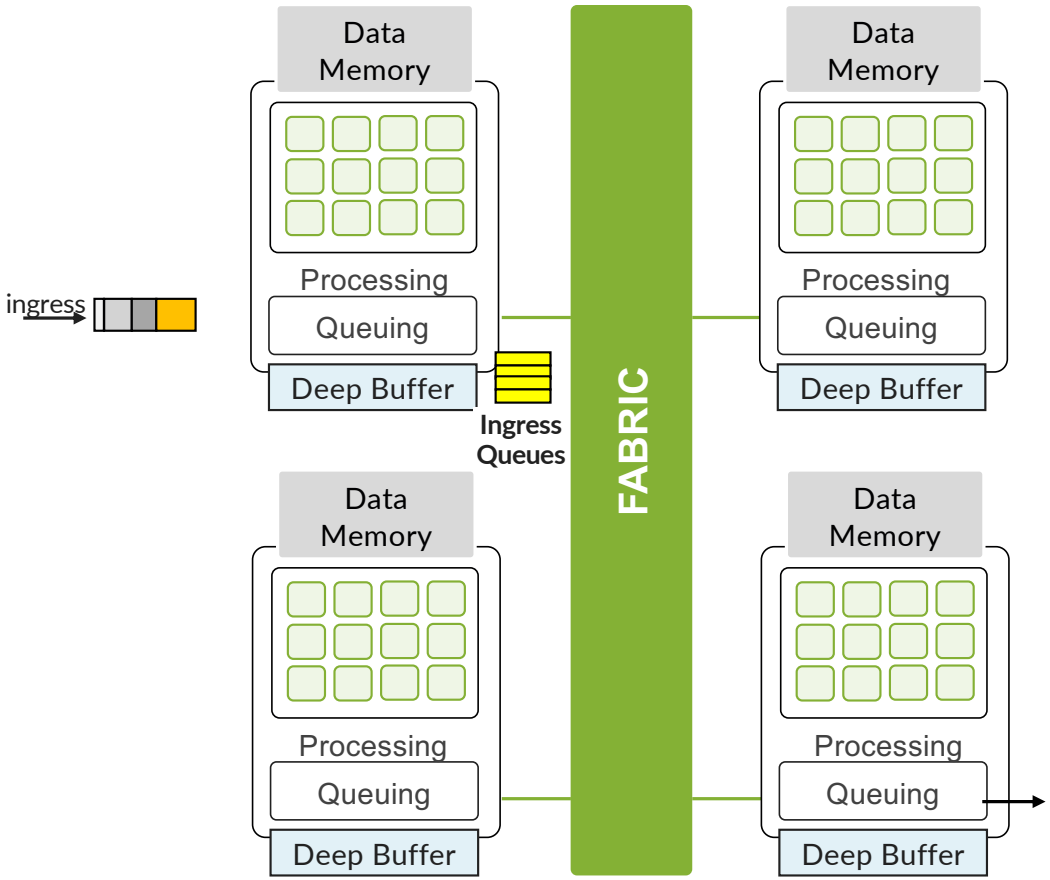


VOQ Scale is Global / Per System (2/2)



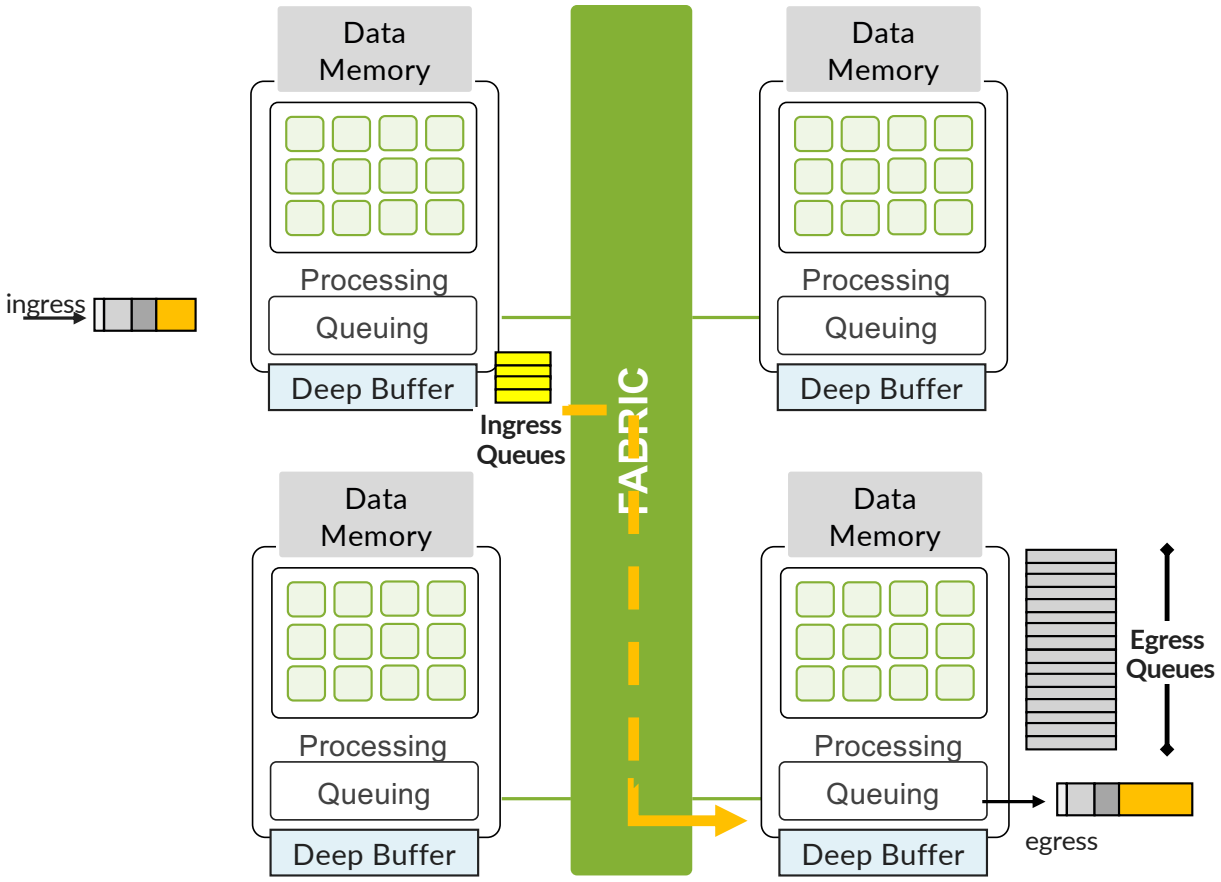
Cannot predict where the packet will be received. Need to create same structure on all ingress paths. VOQ pool resource is shared across the entire system.

Two-Stage Buffering Architecture (1/3)



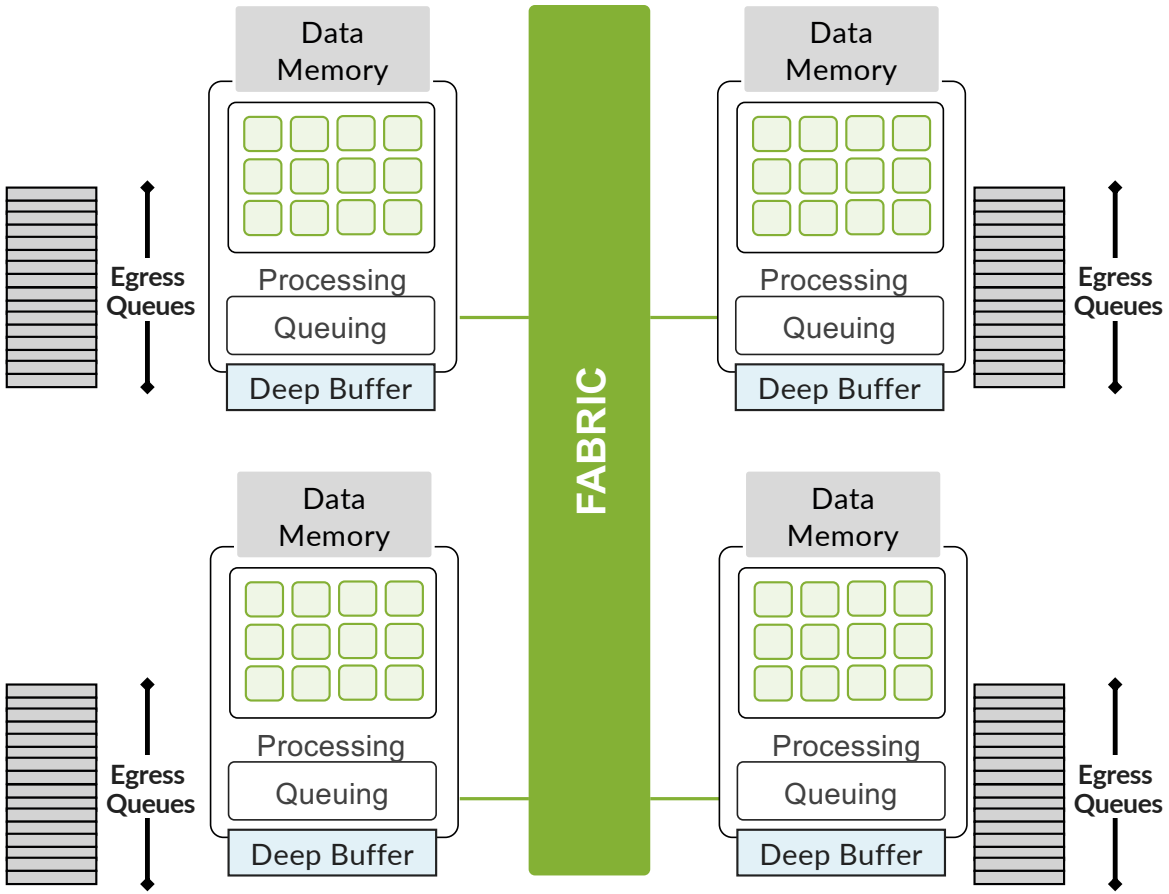
- Packet gets buffered first on ingress to be scheduled across fabric
  - Low number of queues

Two-Stage Buffering Architecture (2/3)



- Packet gets buffered first on ingress to be scheduled across fabric
  - Low number of queues
- Packet gets buffered a second time on egress to be scheduled out the local port
  - Large number of queues
  - All queues are only for “local” egress ports

Two-Stage Buffering Architecture (3/3)



- Packet gets buffered first on ingress to be scheduled across fabric
  - Low number of queues
- Packet gets buffered a second time on egress to be scheduled out the local port
  - Large number of queues
  - All queues are only for “local” egress ports
- **More line cards, more ASICs, more queues!**
  - Total queues = #cards x #ASICs x #queues/ASIC

Buffering Architectures Pros and Cons

Two-Stage Buffering	Ingress-Only Single-Stage Buffering
✅ Many more queues per egress port	✅ No Head of line blocking
✅ Supports complex Subscriber and QoS management	❗ Low Scale Queueing
❗ Head of Line Blocking	

How to Choose Architecture?

1. Design Requirements
   - Scale: route table, number of queues
   - Performance: total chip and per port
   - Feature Set: protocols supported

2. ASIC Development Limitations
   - Area
   - Power
   - Clock Speed
   - Costs

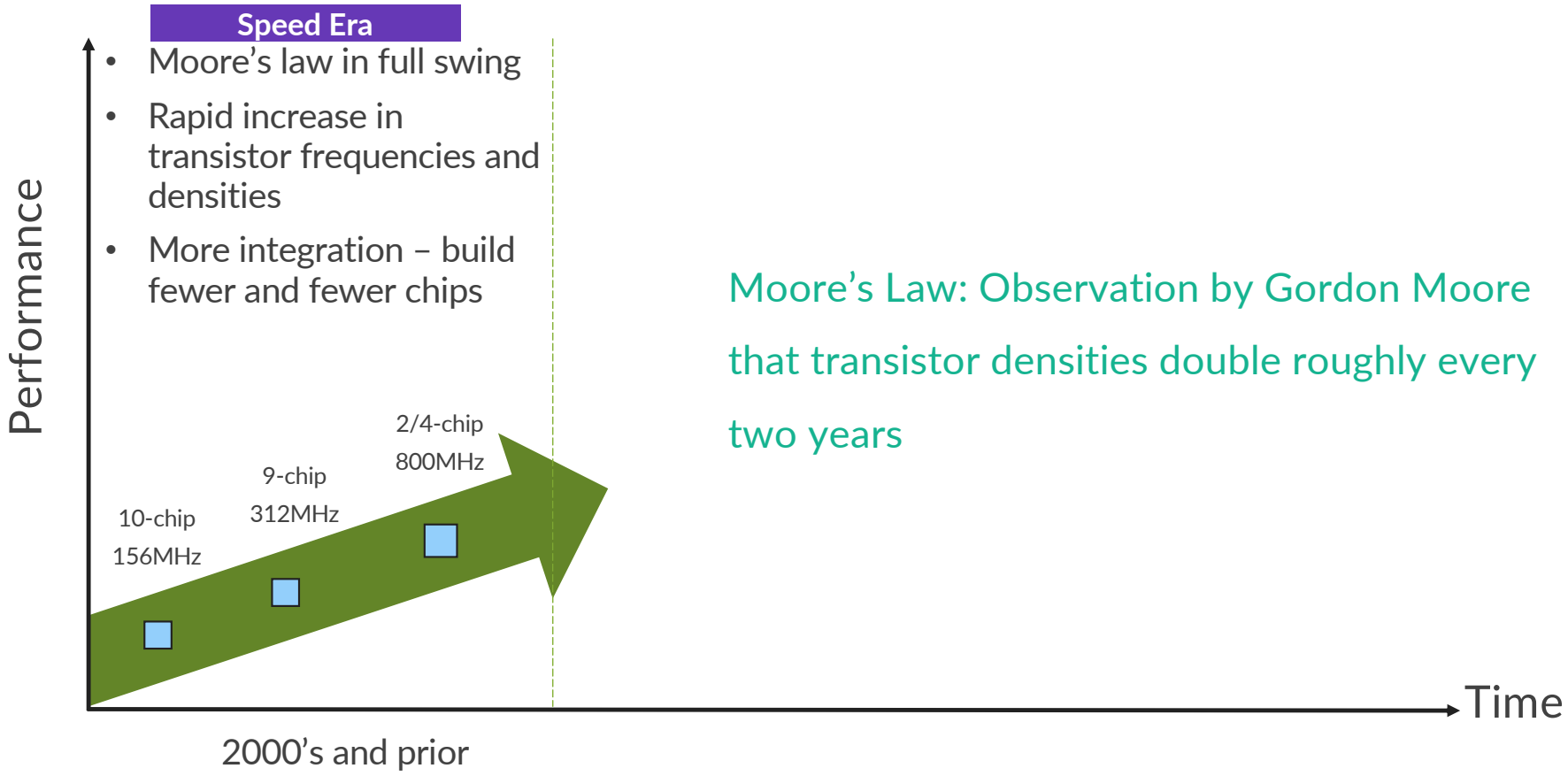
ASIC Development Challenges

More Terminology

- ***die*** – piece of integrated circuit cut from silicon wafer
- ***package*** – container that holds the die
  - Protects the die and allows connectivity on the board
- ***chip or ASIC*** - packaged die

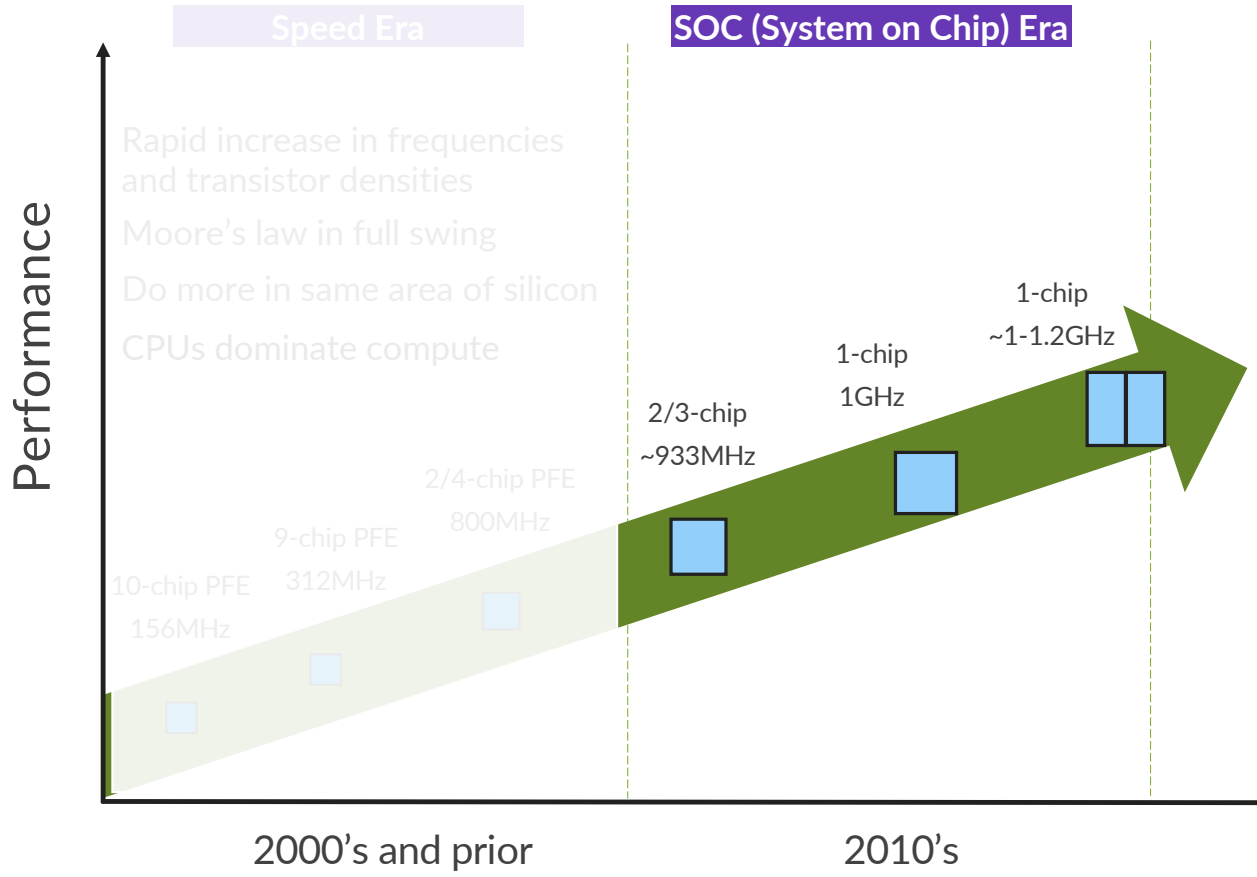
- ***process node*** - refers to the manufacturing technology used to make the die
  - newer node gives smaller, faster, lower power transistors

Network ASIC Trends



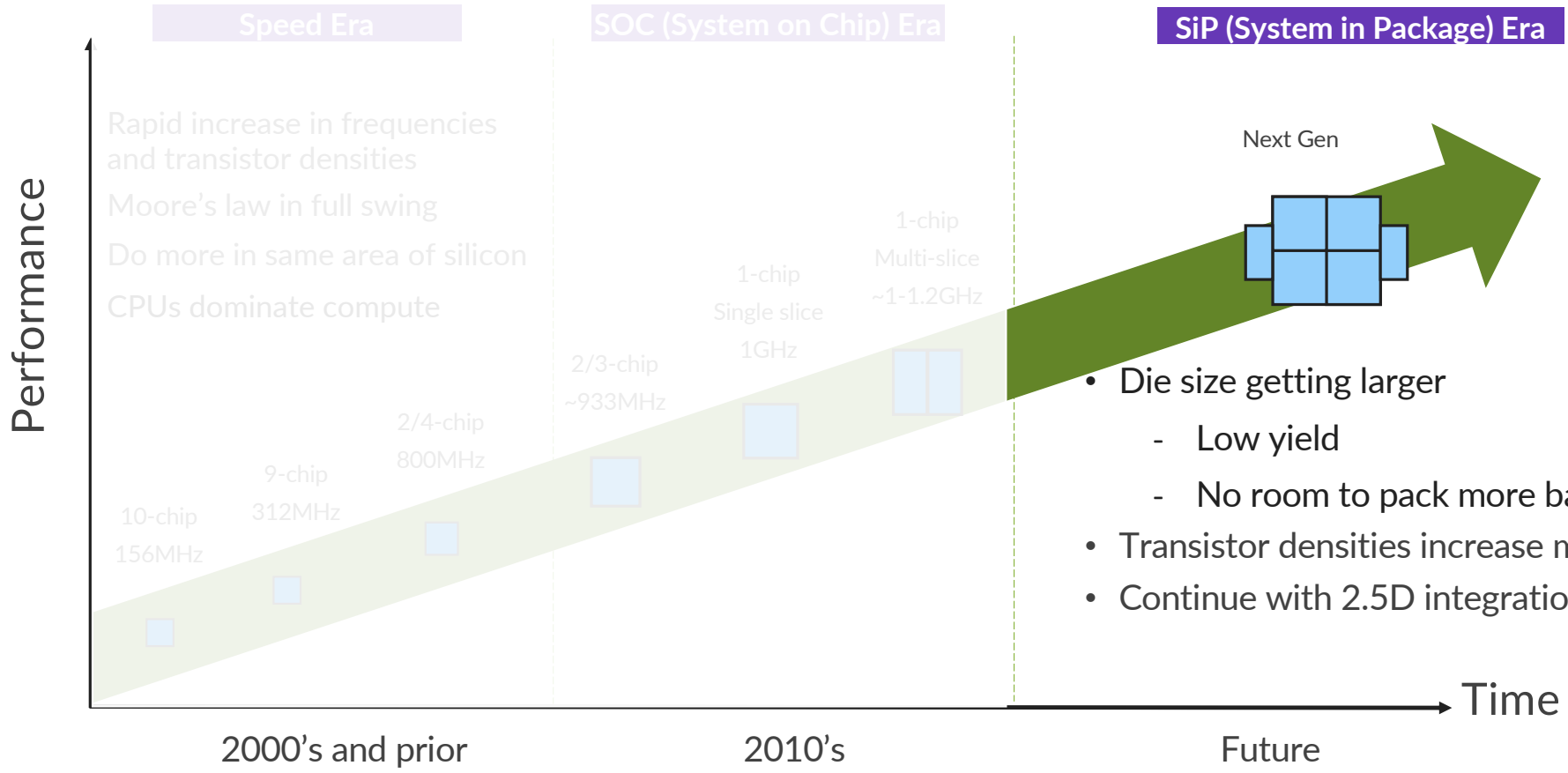
Moore's Law: Observation by Gordon Moore that transistor densities double roughly every two years

Network ASIC Trends



- Transistor densities continued to increase
- Transistor performance flat lining
- Memory wall
- Networking ASICs
  - New architectures
- Introduction of 2.5D integration with High Bandwidth Memory (HBM)

Network ASIC Trends



- Die size getting larger
  - Low yield
  - No room to pack more bandwidth
- Transistor densities increase moderately
- Continue with 2.5D integration

Future Challenges

- Process Node Limitations– how small can it go?
- How to achieve higher capacity – new architectures?
- Handling increasing ASIC power - new power saving techniques?

Glossary and Terminology

- ASIC – Application Specific Integrated Circuit
- SoC – System On Chip integrates all or most components of a system onto single die
- SiP – System in Package integrates multiple dies in single package
- HBM – High Bandwidth Memory using 3D stacking technology
- 2.5D Integration - Packaging technology combining multiple die, placed side-by-side in a package
- VOQ – Virtual Output Queue
- QoS – Quality of Service

Thank you