
#CiscoLive

The bridge to possible

# Catalyst 9000 Switching QoS Deep Dive

## Part 2 – Silicon One ASIC

Kenny Lei – Technical Marketing Engineer

BRKENS-2096b

Content by Ninad Diwakar and Kenny Lei

Cisco Live!

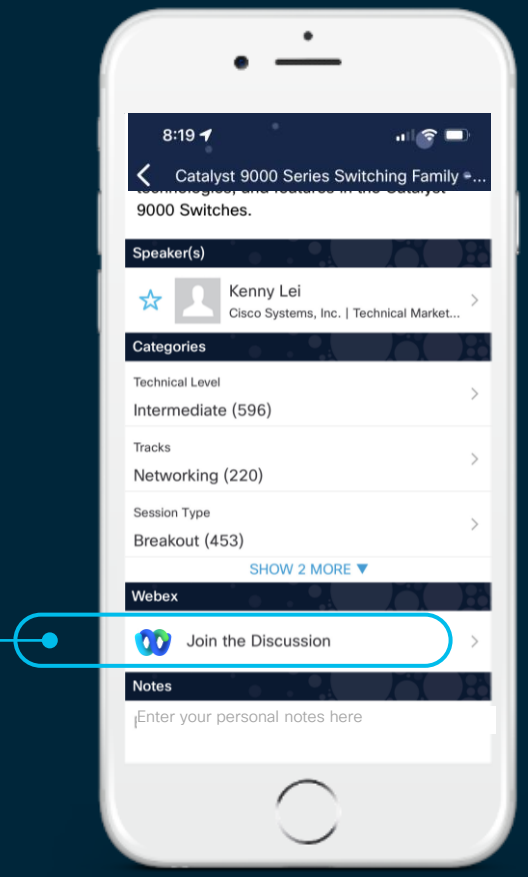
# Cisco Webex App

## Questions?

Use Cisco Webex App to chat
with the speaker after the session

## How

1. Find this session in the Cisco Live Mobile App
2. Click “Join the Discussion”
3. Install the Webex App or go directly to the Webex space
4. Enter messages/questions in the Webex space




https://ciscolive.ciscoevents.com/ciscolivebot/#BRKENS-2096b

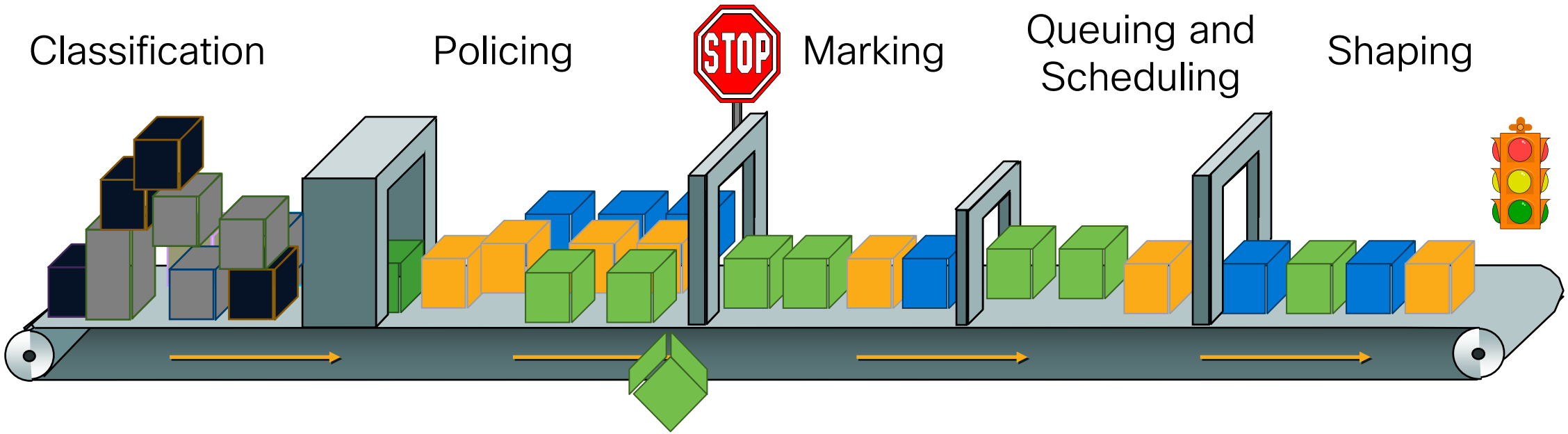
## Webex spaces will be moderated by the speaker until June 17, 2022.

# Agenda

- QoS Overview
- Silicon One QoS (VoQ) Architecture
- Classification, Marking and Policing
- Queueing, Shaping and Scheduling
- Congestion Management and Buffers
- QoS Migration to Silicon One
- Conclusion

# QoS Terminology

# The QoS Toolset




Identify and Split Traffic into Different Classes

Discard Misbehaving Traffic to Maintain Network Integrity

Mark Traffic According to Behavior and Business Policies

Prioritize, Protect and Isolate Traffic Based on Markings

Control Bursts and Conform Traffic

# QoS Terminology

| Term | Explanation |
|---|---|
| Trust | Retain the packet markings as it is |
| Classification | Identify packet priority and place it into different classes |
| Marking | Change the tags (priority) on the packets |
| Policing | Limit the traffic to specified rate. Excess traffic can either be dropped or assigned a different color. |
| Shaping | Limit the traffic to specified rate. Excess traffic will be queued and buffered. |
| Queueing | Process the packet into separate queues |
| Buffering | Storage for packets to be queued |

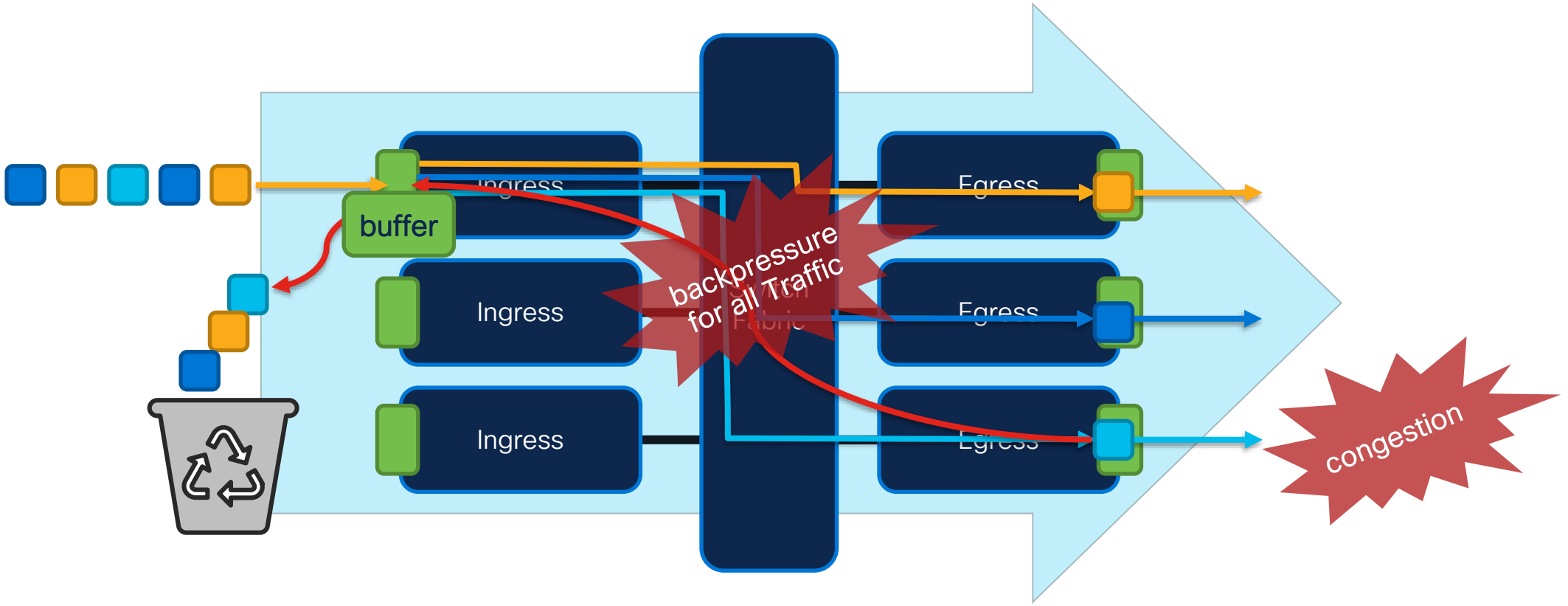
# Silicon One QoS Terminology

| Term | Explanation |
|---|---|
| VoQ | Virtual Output Queues between Ingress and Egress |
| Packet Color | Used for congestion management to prioritize packets to be dropped |
| Traffic-Class | Internal tag used by the Silicon One ASIC to differentiate packet priority |
| Traffic Manager | The block in Q200 responsible for scheduling |
| Traffic/Transmit Scheduler | When the OQ can send traffic out to the wire |
| Credit Scheduler | When the VoQ can send traffic to the Output queue |
| SMS | Shared Memory Sub-system - Primary Buffering system |
| HBM | High Bandwidth Memory - Secondary deep Buffering system used during congestion |

# VoQ and Head of Line Blocking (HoL)

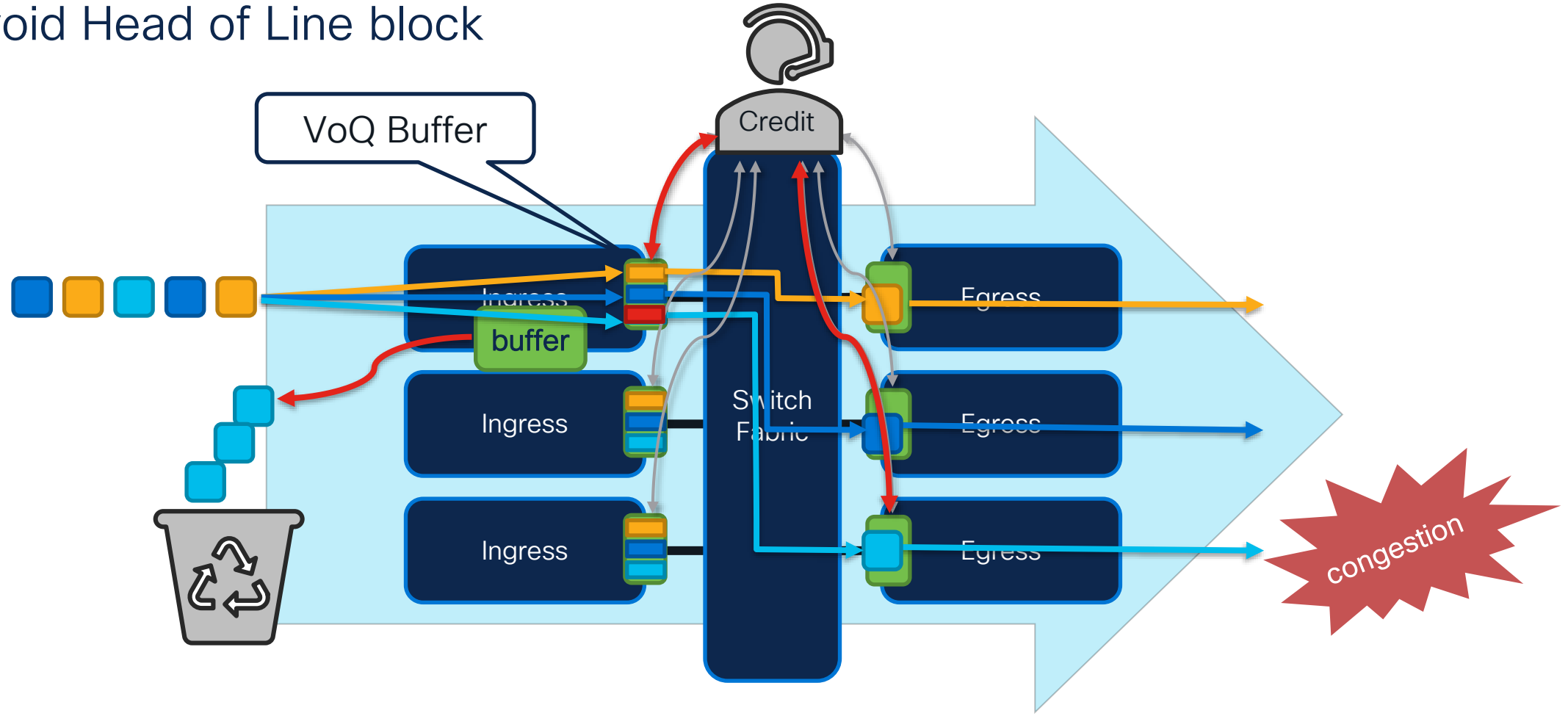
# Buffer types – Head of Line Blocking

## What is the Problem?



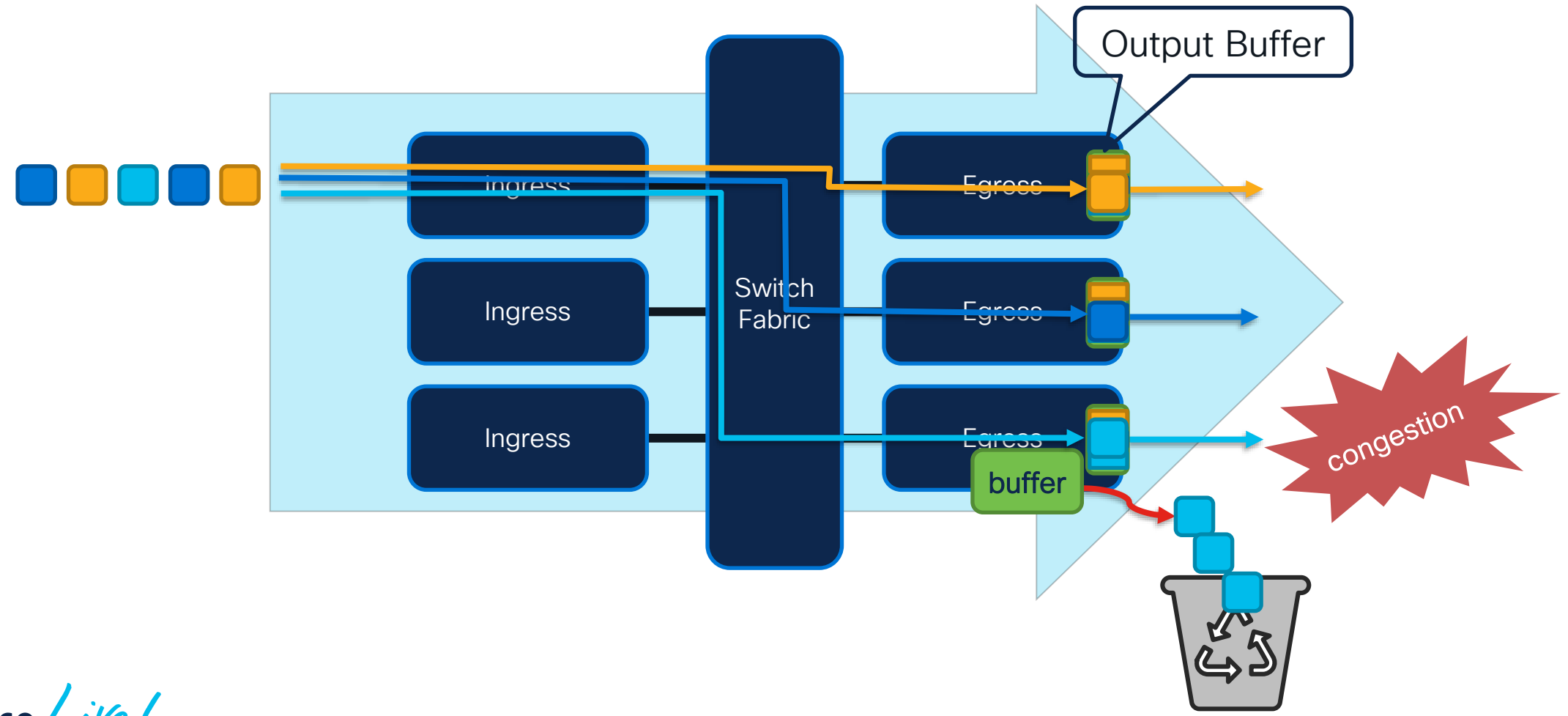
# Virtual Output Queuing – Silicon One ASIC

## Avoid Head of Line block

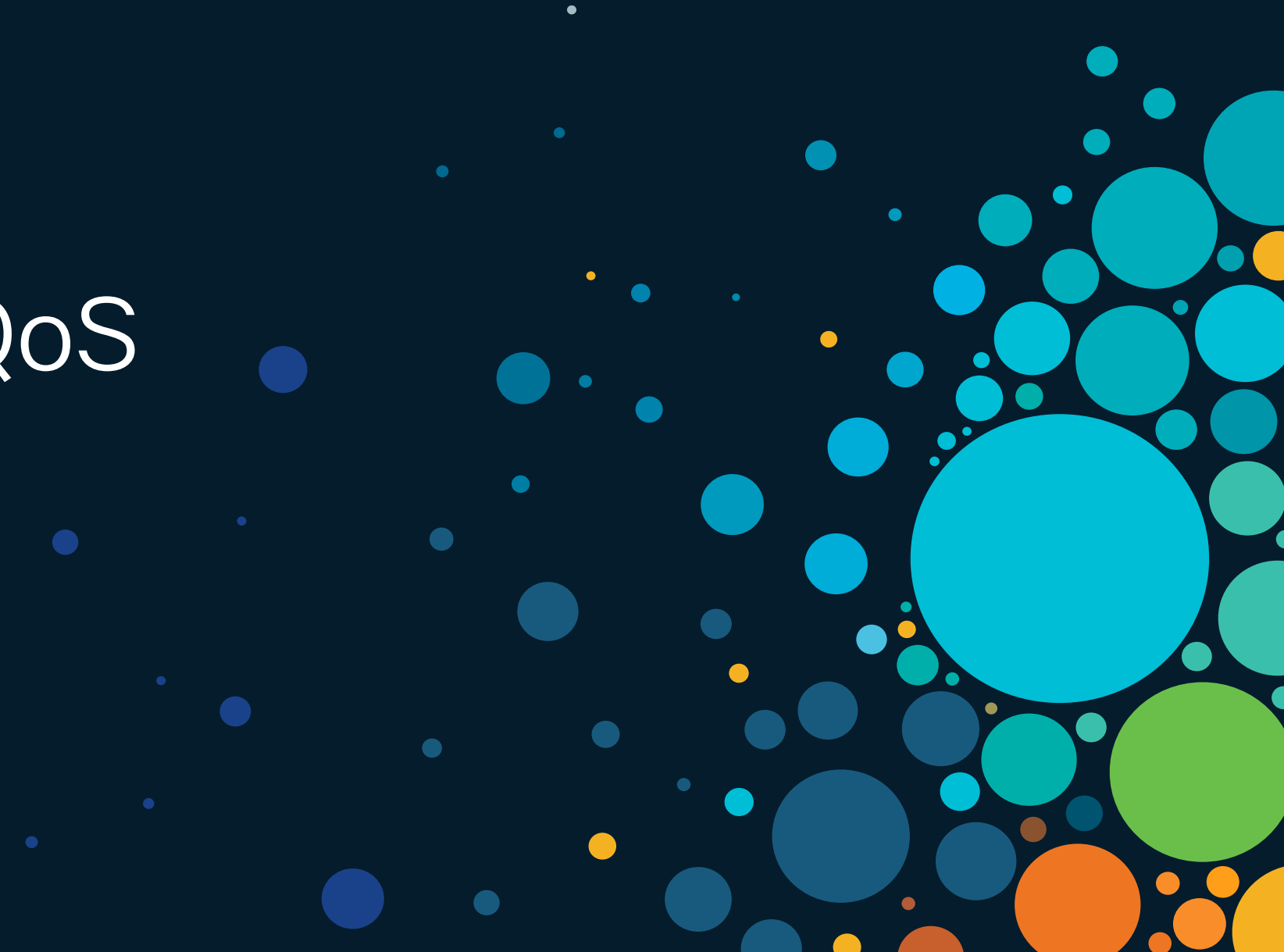


# Output Queuing - UADP

## Avoid Head of Line Blocking



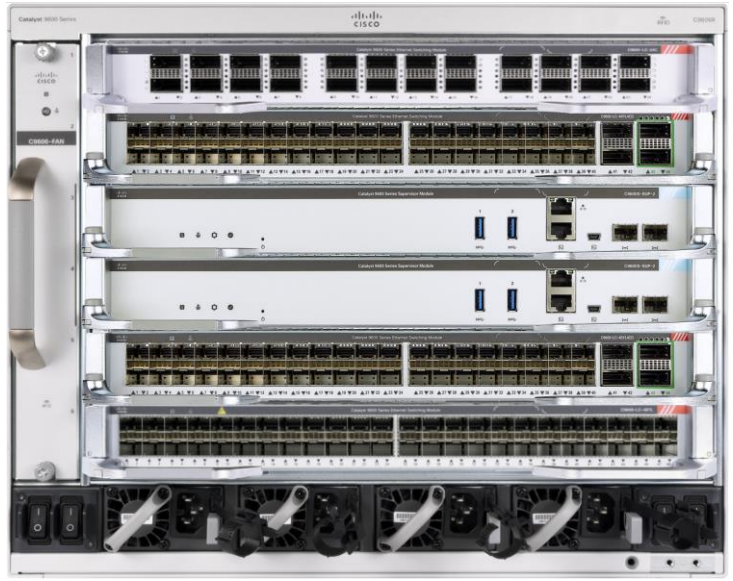
# Silicon One QoS Overview



# Catalyst 9000 Switches with Silicon One ASIC

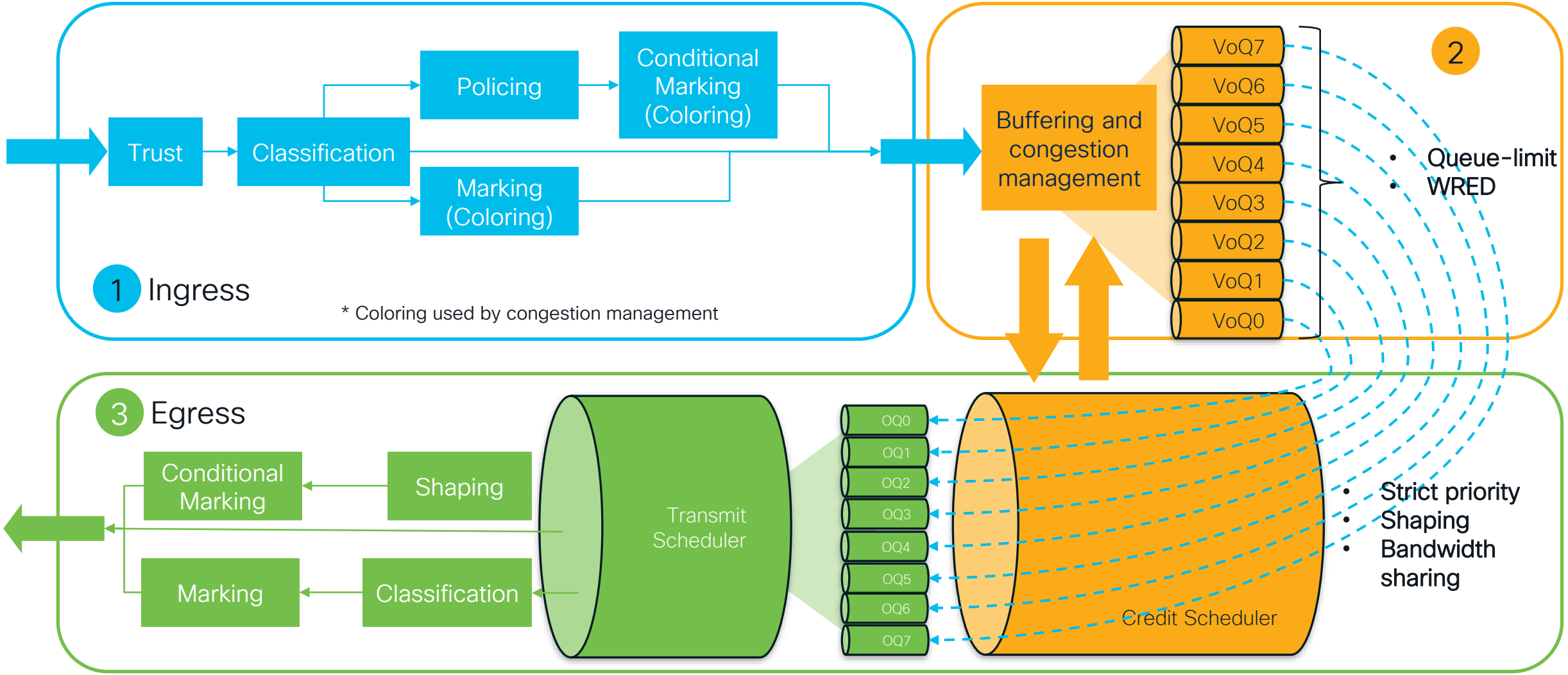
C9500X-28C8D

C9600X-Sup-2

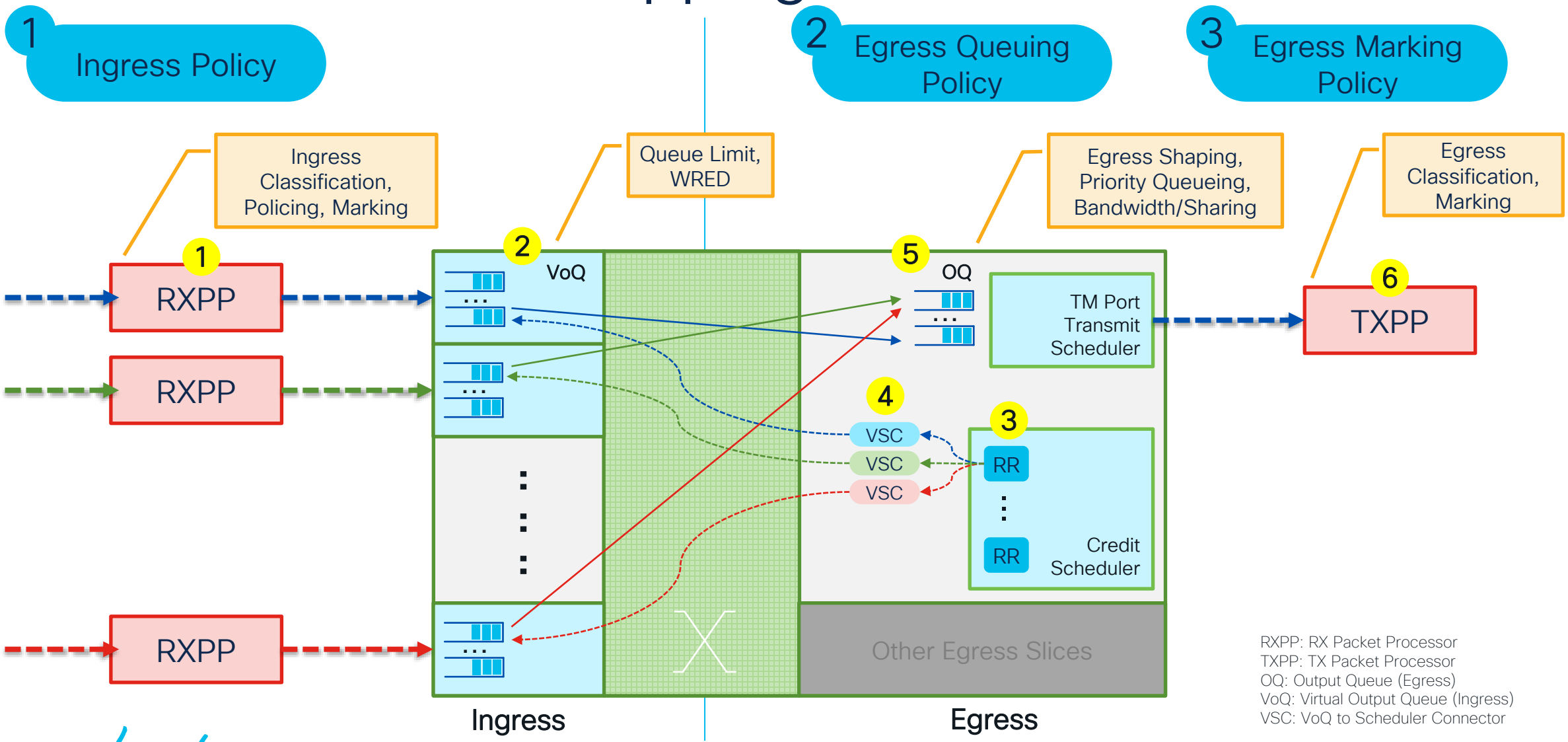


# Silicon One

## Features Mapping



# Silicon One ASIC mapping




RXPP: RX Packet Processor
TXPP: TX Packet Processor
OQ: Output Queue (Egress)
VoQ: Virtual Output Queue (Ingress)
VSC: VoQ to Scheduler Connector

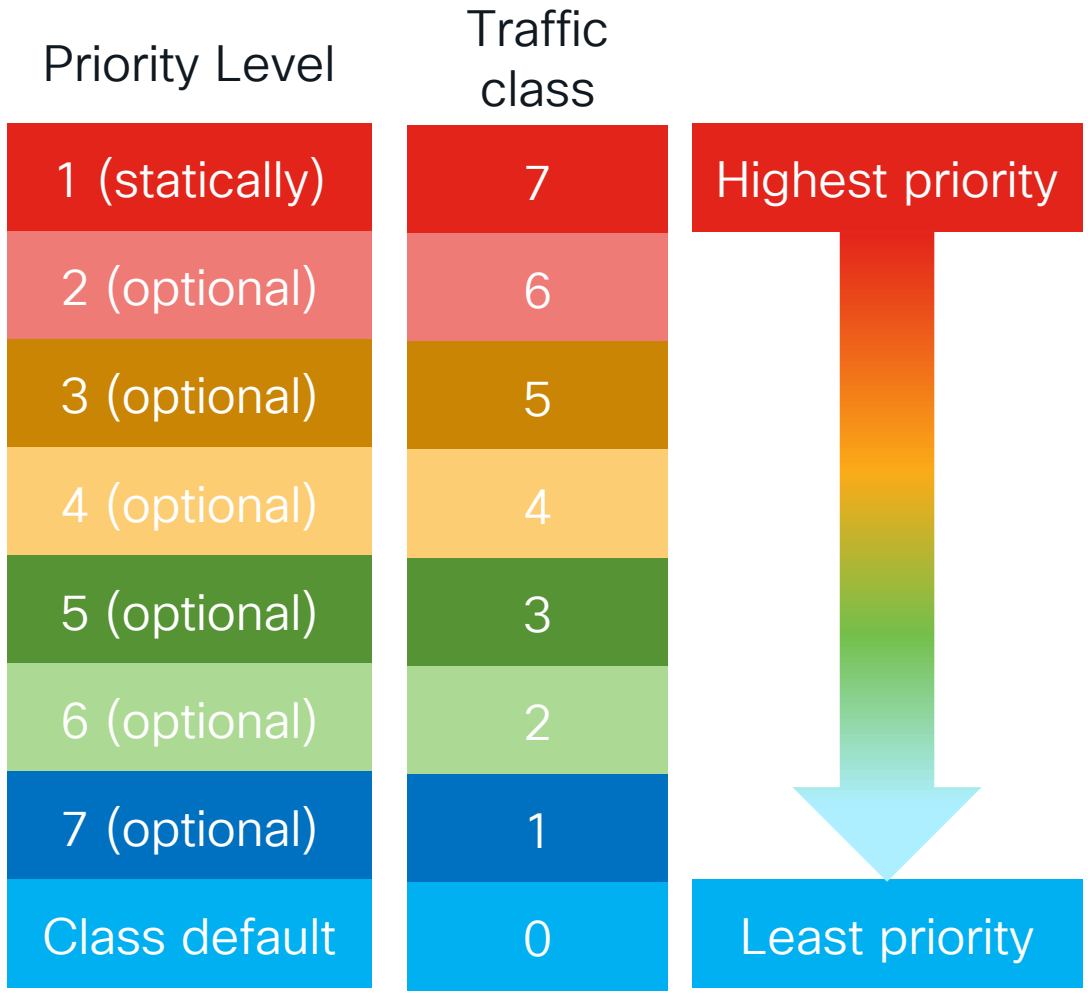
Hidden

# Silicon One ASIC mapping Hidden slide

1. Packet from the ingress interface hits the Receive Interface Group (Rx IFG). RxPP (Receive packet processor) consists of both the corresponding Interface Group as well as the Network Processing Unit (RxNPU) for the corresponding slice.
2. From the RxIFG, the packet descriptor is sent to corresponding VoQ (Virtual Output Queue) where it is queued and forwarded once it receives a credit.
3. Each VoQ connects to the scheduler via a VSC (Virtual queue scheduler) The scheduler uses a round robin (RR) algorithm to provide credits each corresponding VSC.
4. The credit is sent from the VSC to the corresponding VoQ. Once the VoQ receives the credit, it can forward the traffic to the egress.
5. The packet is sent from VoQ to corresponding OQ.
6. Once at the OQ, any marking operations, if any, are performed and the packet egresses out the TxNPU and TxIFG (collectively called TxPP) out of the switch.

# Traffic class

- S1 ASIC uses traffic classes to map traffic to different queues. "**traffic-class**" is local significant to the switch only
- 3-bit field => 8 values, traffic-class <0 - 7>
- Traffic-class 0 - lowest priority (maps to class-default); traffic-class 7 - highest priority (traffic-class 1 to 6 can be non-priority)
- **Ingress policies** classify packets to specific traffic classes
- Class-maps in **egress queuing policy** can only match traffic-class

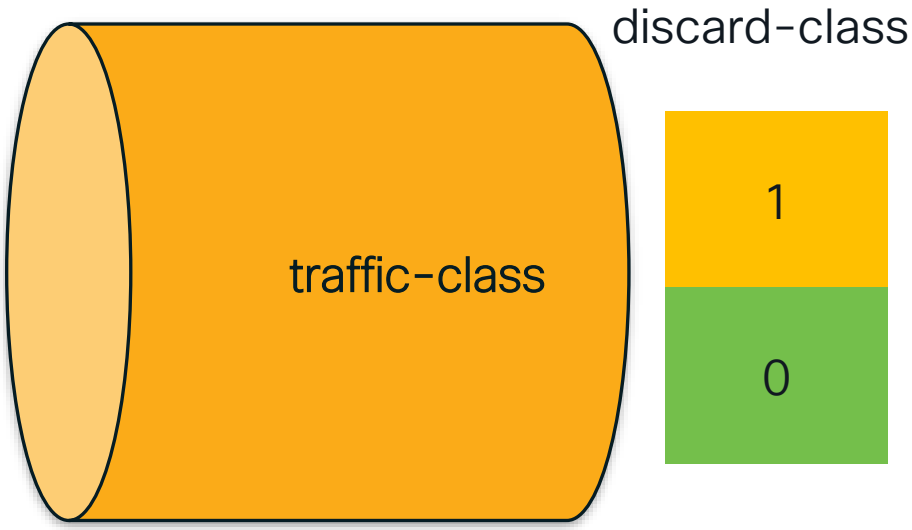


# Traffic class vs QoS-group

| | traffic-class | QoS-Group |
|---|---|---|
| = | Label for incoming packets in classification | |
| = | Local signification (switch) | |
| = | Egress class-map make use of these labels | |
| ≠ | Can be associated with **priority** in the egress | **Simple** label for use in the egress |
| ≠ | Associated with VoQ. (Multiple TCs can make to a same VoQ) | No priority or queue reference |
| ≠ | Default **mapping** (DSCP/COS to traffic-class) if not defined by ingress policy | Options for the egress |

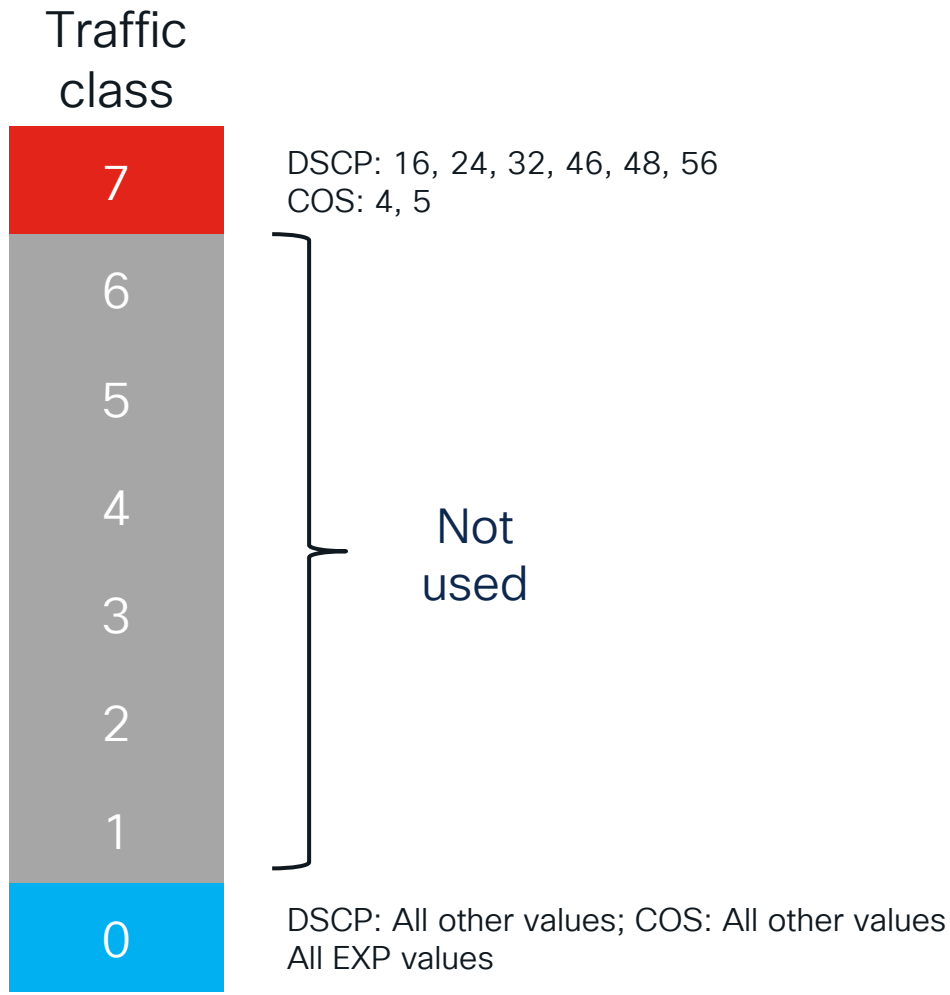
# Traffic color – discard-class

- S1 ASIC uses traffic color to assign priority for packets within a traffic-class. ”**discard-class**” is local significant to the switch only
- 1-bit field => 2 values.
- discard-class < 0- 1> (0 – green, 1 – yellow)
- **Ingress policies** can color the packet unconditionally or conditionally with a policer
- Default traffic color is 0 (green)
- Yellow (marked with 1) packet will be dropped first in event of congestion



# Silicon One Q200 QoS Default

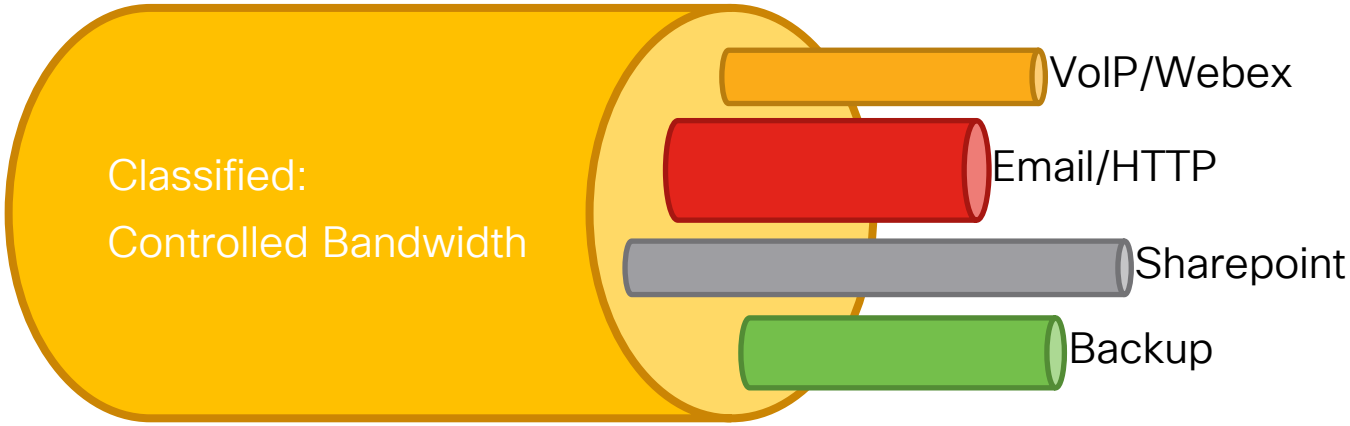
- Catalyst 9000 Switches with Silicon One Q200 ASIC
  - QoS enabled
  - All ports trust at layer2 and layer3
  - Two queues (traffic-class 7 and traffic-class 0, traffic-class7 is priority level 1)



# Classification, Marking and Policing

# Classification and Marking

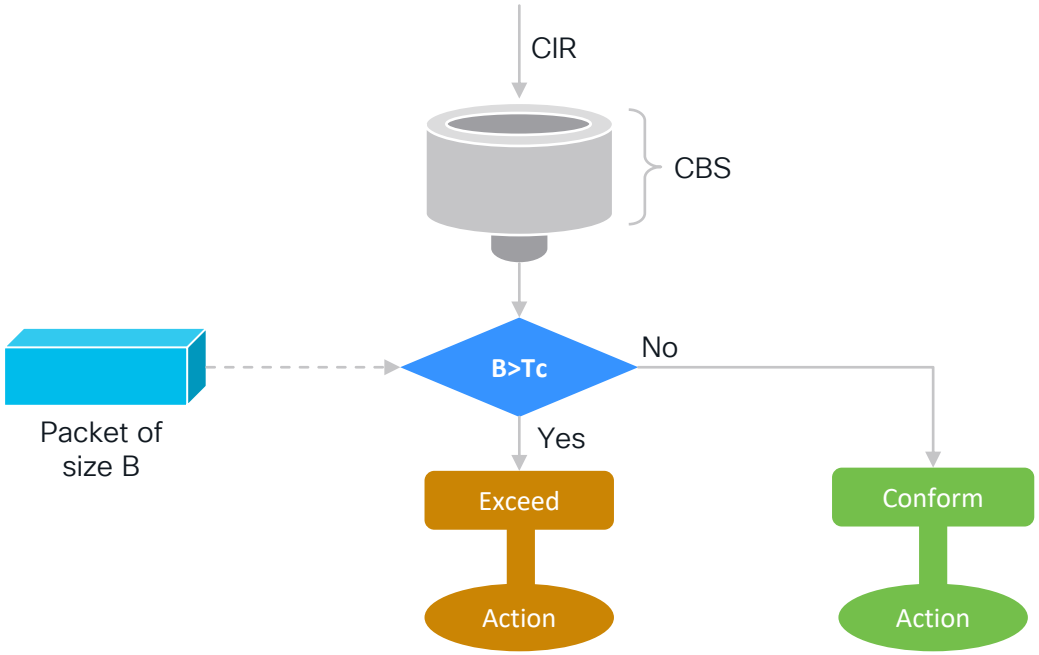
- Identify traffic
  - Access Control Lists (ACLs)
  - DSCP
  - IP precedence
  - CoS
  - QoS Group (local with the switch)
  - EXP (MPLS)
  - VLANs
- Marking (coloring)
  - Conditional or unconditional
  - Table map *
  - QoS group (local within switch)
  - Traffic-class (local within switch)
  - Traffic-color (local within switch)


* for policer markdown

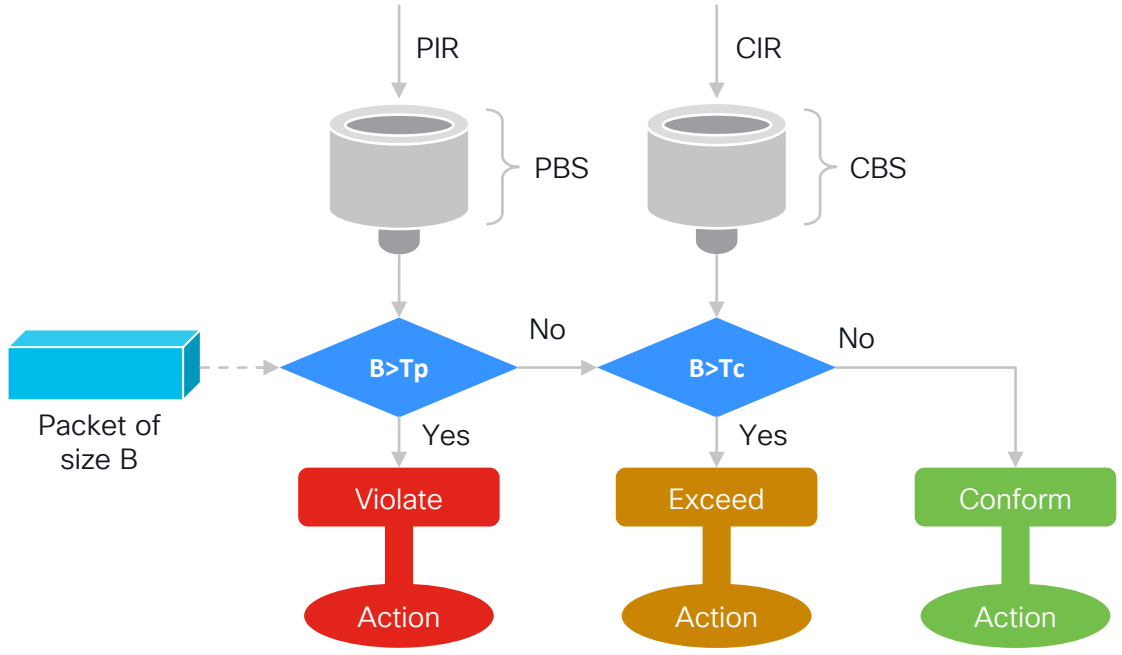
# Policing – Limit the traffic

## 1 rate 2 color




```
police cir 1g
  conform-action transmit
  exceed-action drop
```

## 2 rate 3 color




```
police cir percent 10 pir percent 50
  conform-action transmit
  exceed-action set-discard-class-transmit 1
  violate-action drop
```

Traffic color

CIR – Committed Information Rate
PIR – Peak Information Rate

PBS- Peak Burst Size
CBS – Committed Burst Size

# Policing and marking/coloring example

Unconditional Traffic Marking/Coloring

```
policy-map ingress-policy
 class class-5-green
  set traffic-class 5
 class class-5-yellow
  set traffic-class 5
  set discard-class 1
```

Different class-map
Same traffic-class

1R2C Policing:

```
policy-map test-police-1R2C
 class dscp1
  set traffic-class 3
  police rate 10g bps
     conform-action transmit
     exceed-action drop
  !
```

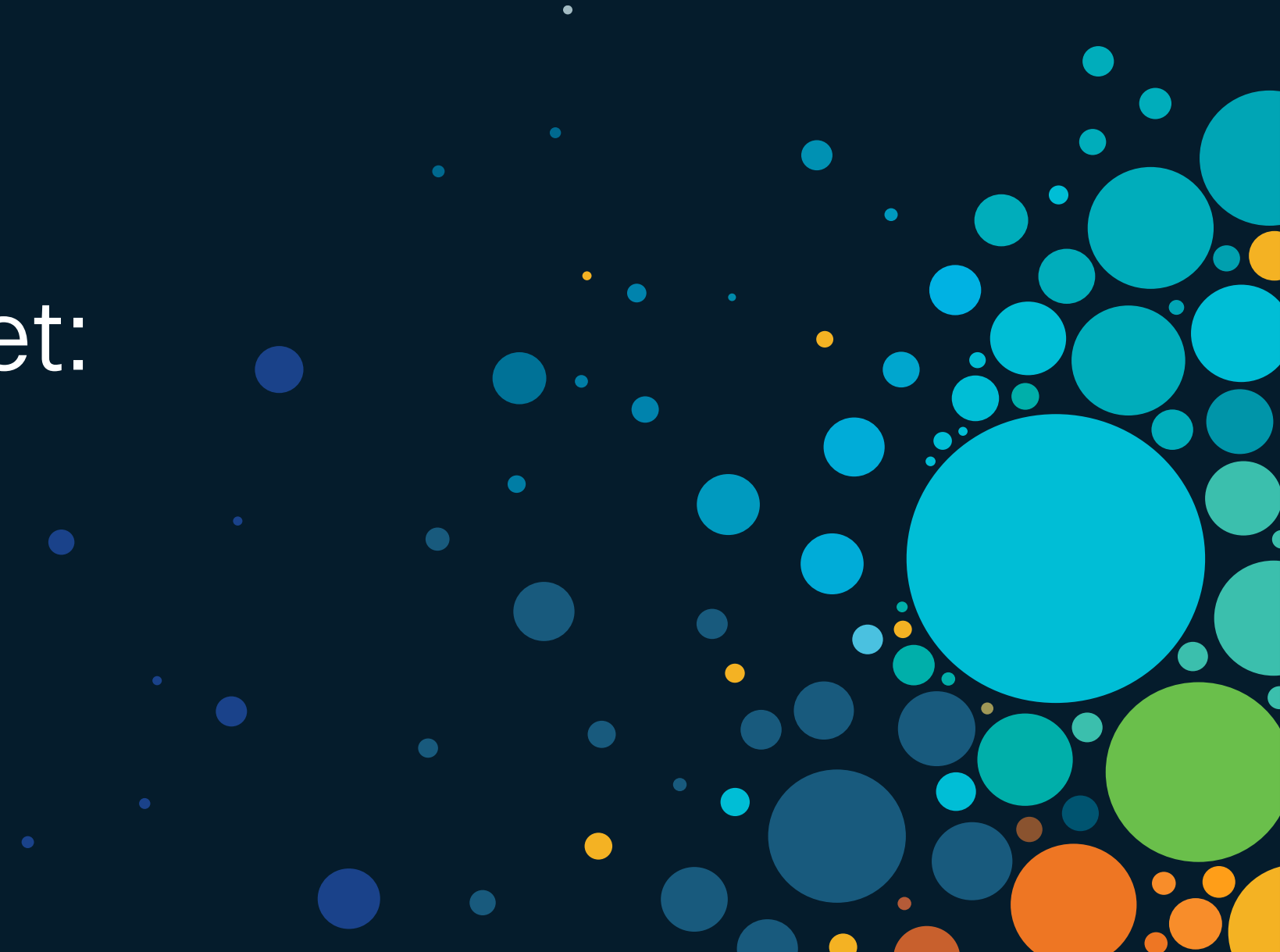
Conditional Traffic Marking/Coloring

```
policy-map ingress-policy
 class class-5
   set traffic-class 5
   police rate 5g bps peak-rate 10g bps
    exceed-action set-discard-class-transmit 1
```

2R3C Policing:

```
policy-map test-police-2R3C
 class dscp1
  set traffic-class 3
  police rate 10g bps peak-rate 20g bps
   conform-action transmit
   exceed-action set-discard-class-transmit 1
   violate-action drop
```

# Egress Toolset:
## Queueing, Shaping and Scheduling



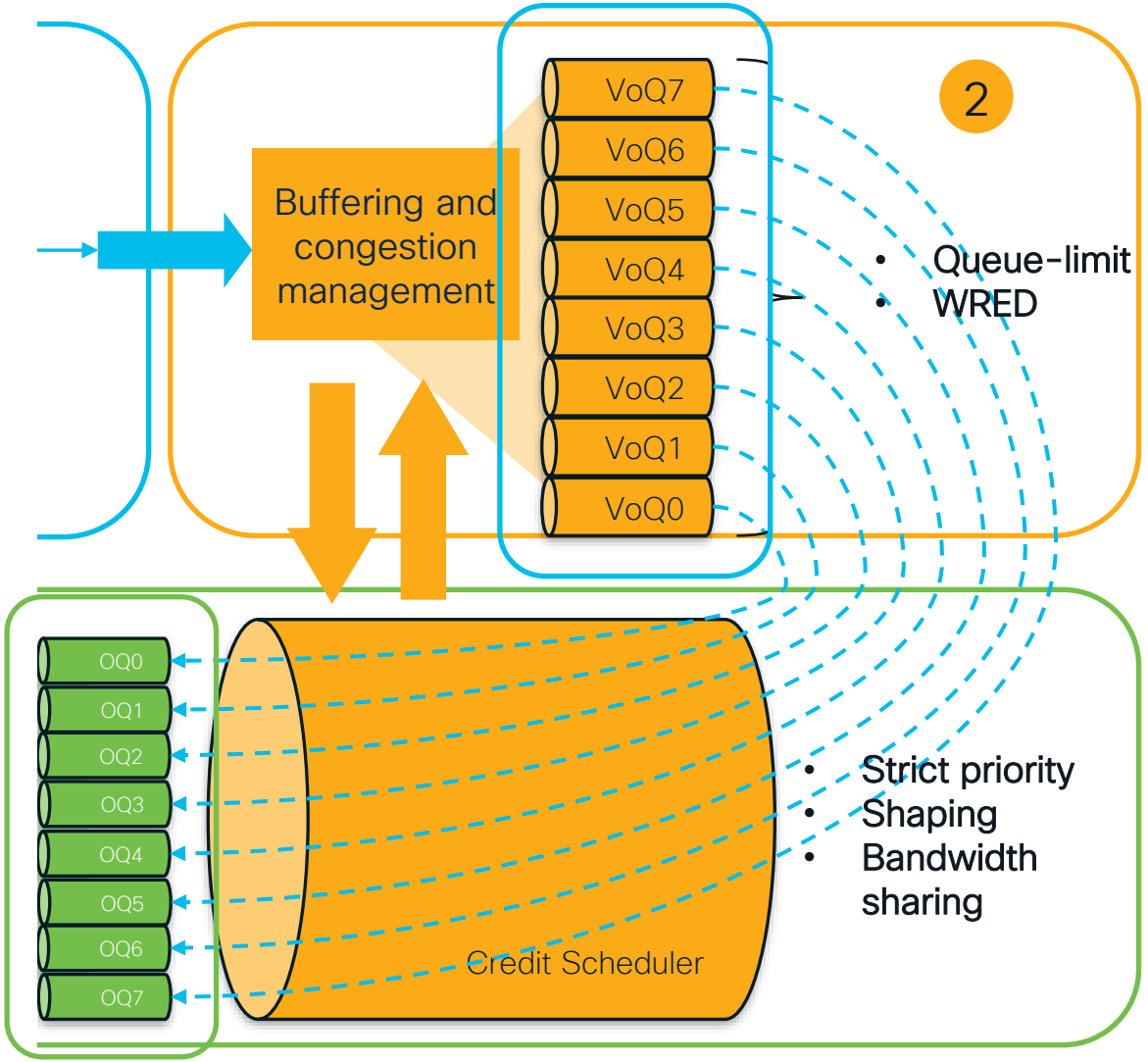
cisco Live!

# Queueing

Virtual output Queue (VoQ)
- 8 VoQ on each ingress slices for each interface
- Each traffic-class maps to a VoQ (multiple traffic-classes can map to same VoQ)

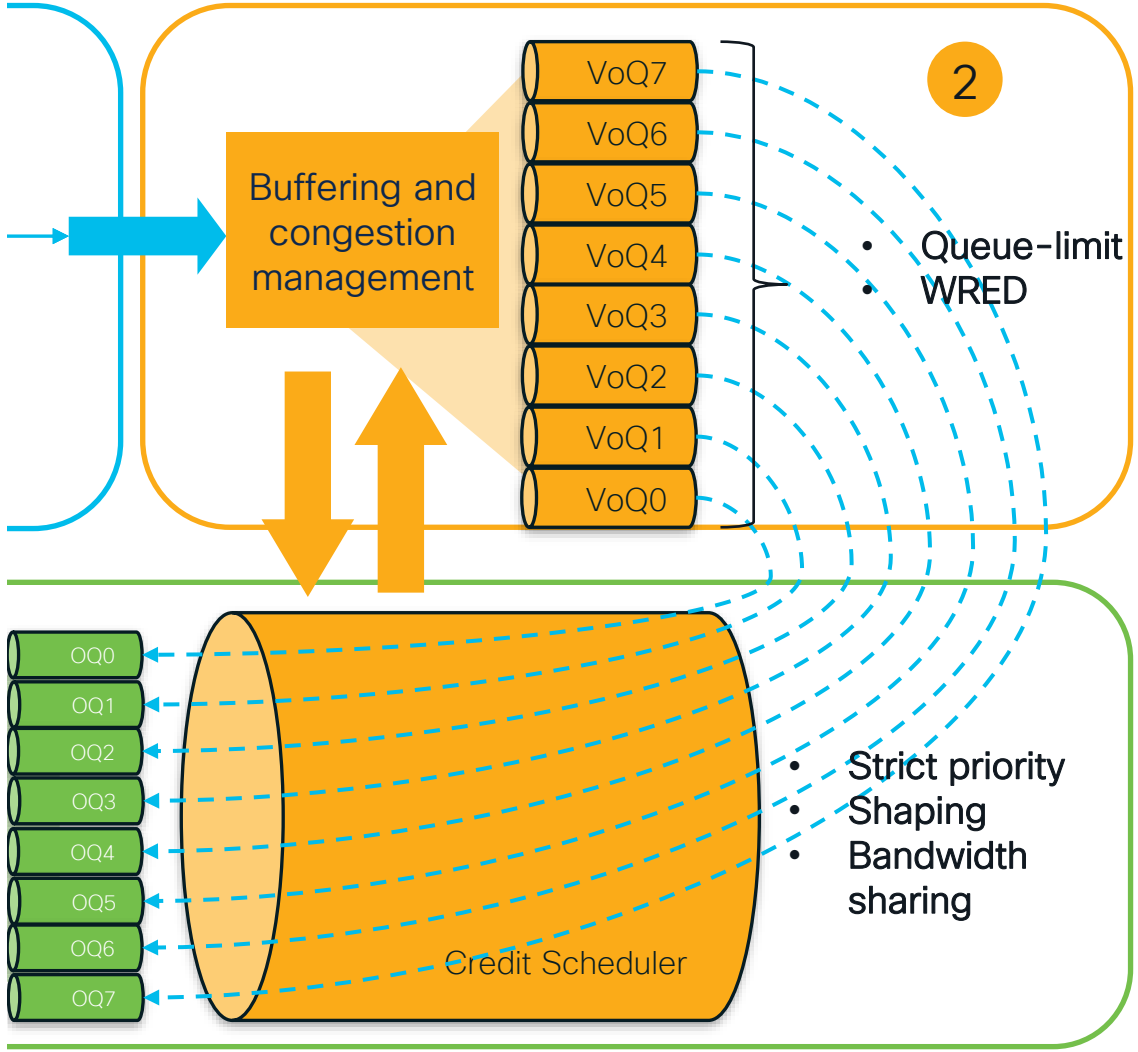
- VoQ maps to output Queue.

Output Queue
- 8 output queues (egress) for each interface
- Up to 7 strict priorities (level 1 – highest)
- Traffic-class 7 is always priority level 1, priority level is optional for other traffic-classes



# Scheduling

- Packet schedule from VoQ to OQ based on a credit scheduling system
- Packets are buffered at ingress (VoQ)
- Different type of queues are served differently
  - Strict priority queues
    - Always serviced first
    - Up to 7 PQs
  - Normal queues (without priority configured)
    - Served only after priority queues are empty
    - Use Weighted Round Robin (WRR) for scheduling



# Scheduling – Example

```
class-map match-any tc-7
 match traffic-class 7
class-map match-any tc-6
 match traffic-class 6
…
class-map match-any tc-1
 match traffic-class 1
```

Map traffic to the queues

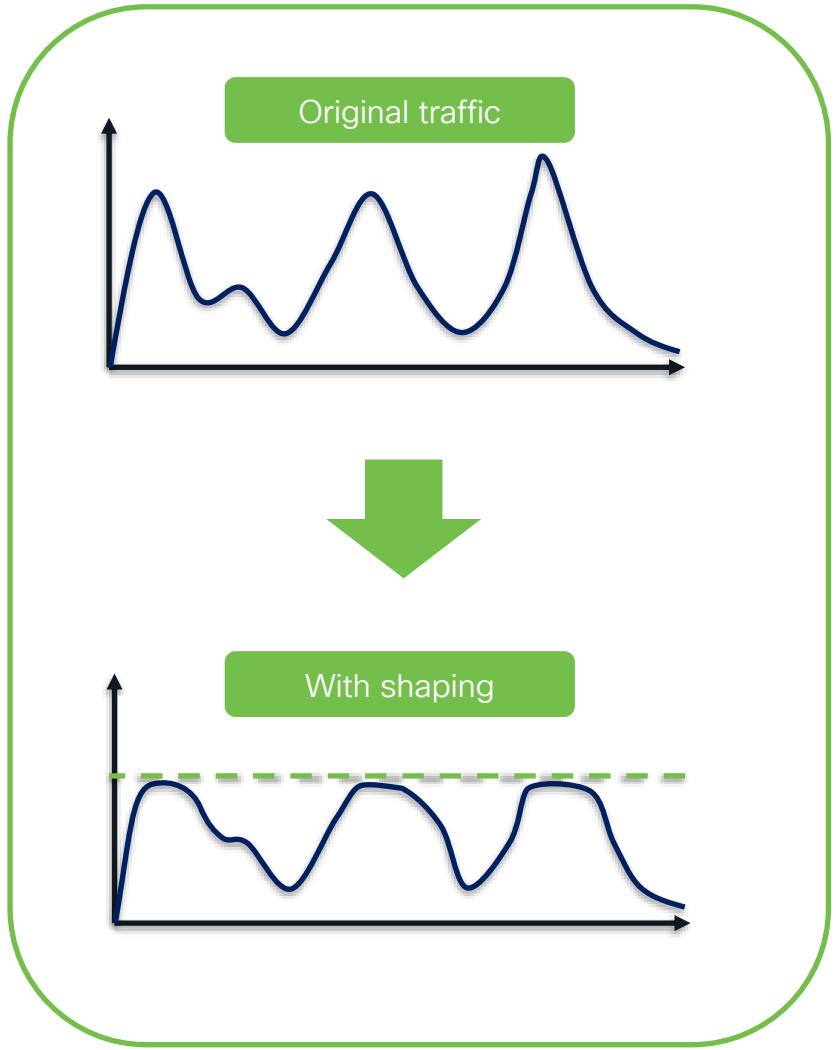
```
policy-map egress-policy
 class tc-7
  priority-level 1
 class tc-6
  Priority-level 2
 class tc-5
  bandwidth remaining ratio 1
 class t-4
  bandwidth remaining ratio 1
 class tc-3
  bandwidth remaining ratio 1
 class tc-2
  bandwidth remaining ratio 1
 class tc-1
  bandwidth remaining ratio 1
 class class-default
  bandwidth remaining ratio 1
```

- Two priority queues here
- Level 1 has the absolutely priority over level 2

- Use “bandwidth remaining ratio” to assign weight
- This example use the same weight for all the remaining 6 queues
- Served round robin around 6 queues as long as there isn’t any traffic on the two PQs

# Shaping

- Smooth out traffic peaks, microburst, with preserving traffic
- Control traffic rate to the desired value with buffering.
- Usually in the egress direction
- Can be applied on all classes, regardless of priority level.




Shaping Example:

```
policy-map type queueing egress-queueing
  class tc7
    priority level 1
    shape average 1g
   class tc6
     priority level 2
     shape average 5g
...

   class class-default
     shape average 5g
```

# Egress Marking

# Egress Marking

- Used to change packet tags of packets egressing the switch.
- A separate policy-map apart from the queueing policy-map.
- If both queueing and marking egress policy-maps are applied, marking happens after queueing actions.
- ACL matching in egress is no supported.

```
class-map match-any dscp-af41
 match dscp af41
!

policy-map egress-map
 class dscp-af41
  set dscp af31
 !

interface interface <#>
 service-policy output egress-map
```

# Congestion Management

# Weighted Tail Drop (WTD)

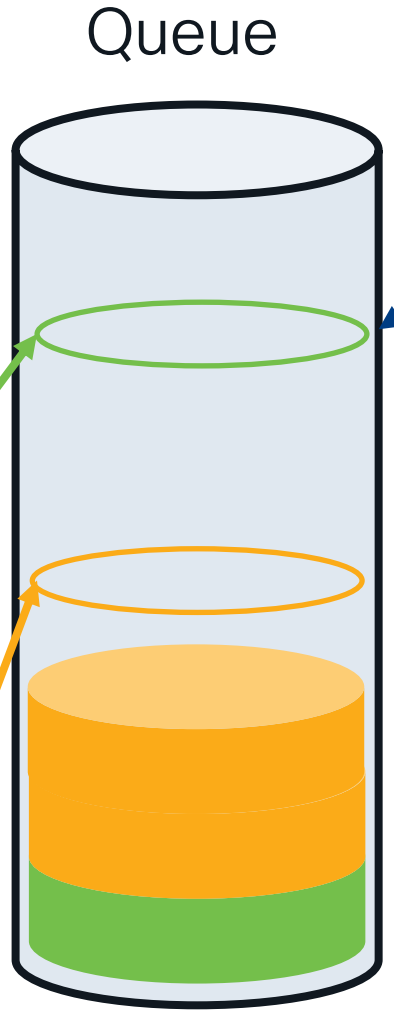
**Ingress Policy-map**

```
policy-map ingress-policy
      class class-5-green
        set traffic-class 5
      class class-5-yellow
        set traffic-class 5
        set discard-class 1
```

```
…
class-map match-any traffic-class-5
 match traffic-class 5
…
```

**Egress Policy-map**

```
policy-map type queueing queue-policy
 ...
 class traffic-class-5
  queue-limit 100000000 bytes
  queue-limit discard-class 1 50000000 bytes
```

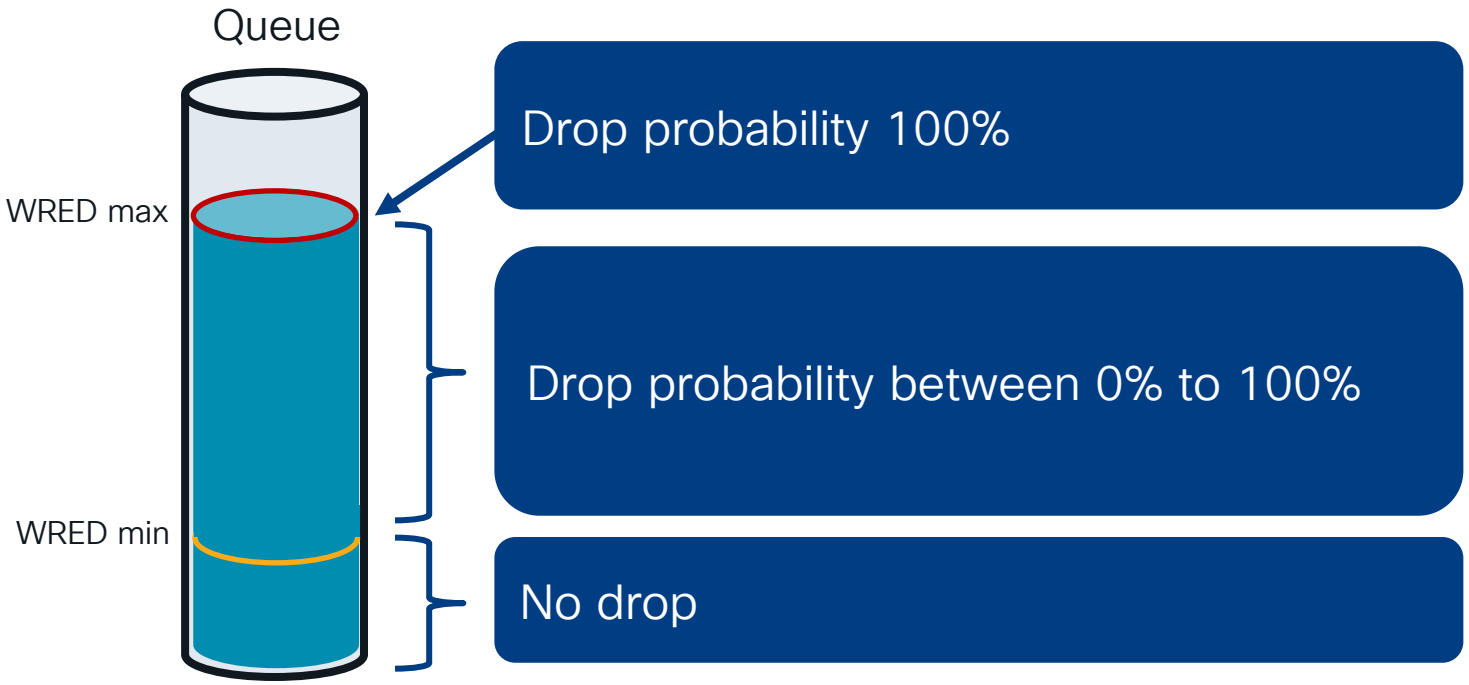



Green traffic will be dropped over this the threshold

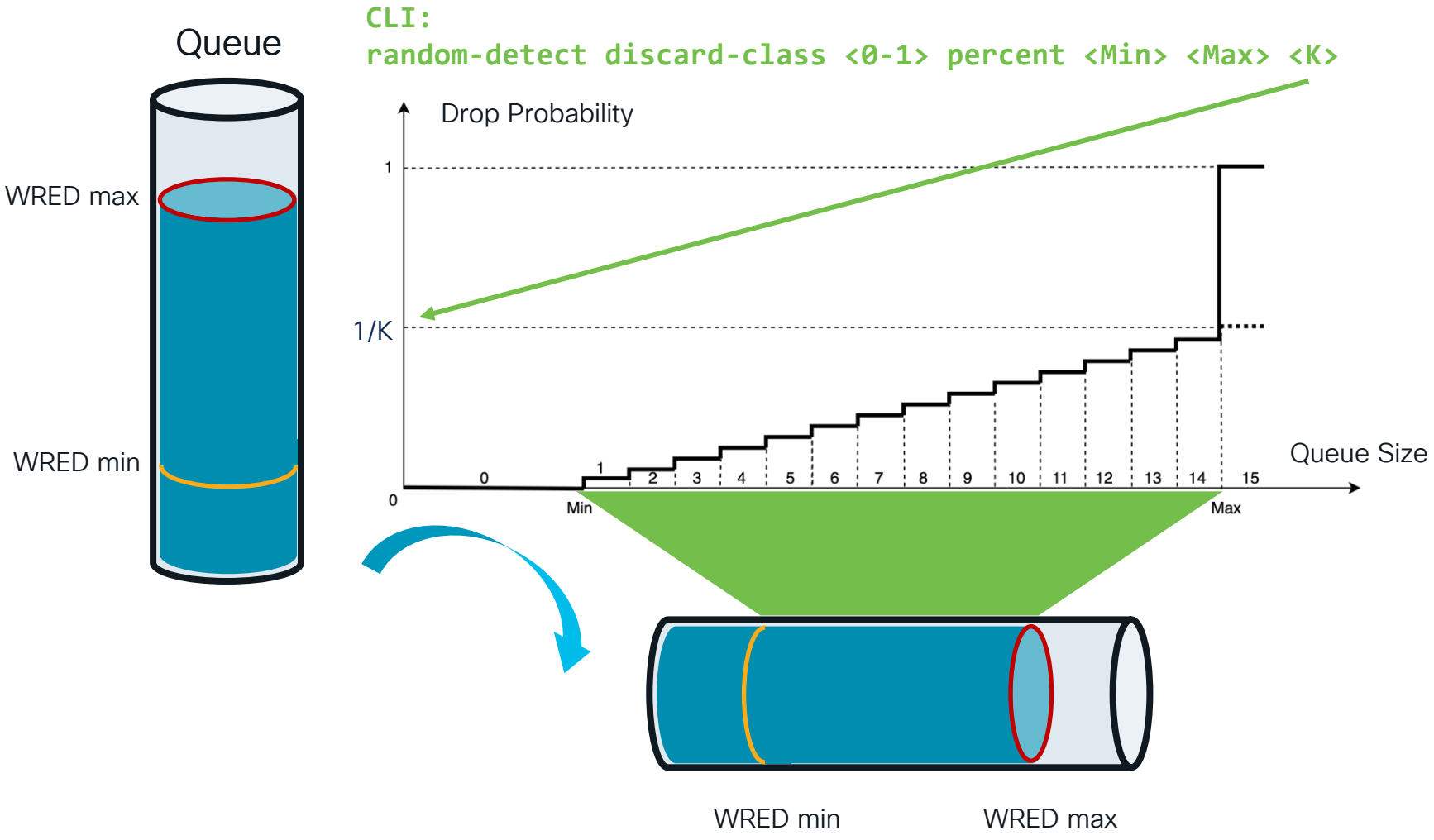
Only traffic marked with green would be here
Yellow traffic would be dropped once it is over the yellow threshold

No drop for both Green and Yellow traffic here

# Weighted Random Early Drop (WRED)



# WRED – Drop probability




- Drop probability increases as the queue utilization increases
- Silicon One ASIC provides 16 regions (drop probabilities)
- Silicon One ASIC provides a knob to influence the drop probability

# WRED – Example

**Ingress Policy-map**

```
policy-map ingress-policy
…
 class class-5-green
  set traffic-class 5
 class class-5-yellow
  set traffic-class 5
  set discard-class 1
…
```

Ingress policy with marking/coloring of packets

**class-map**

```
class tc5
  match traffic-class 5
…
```

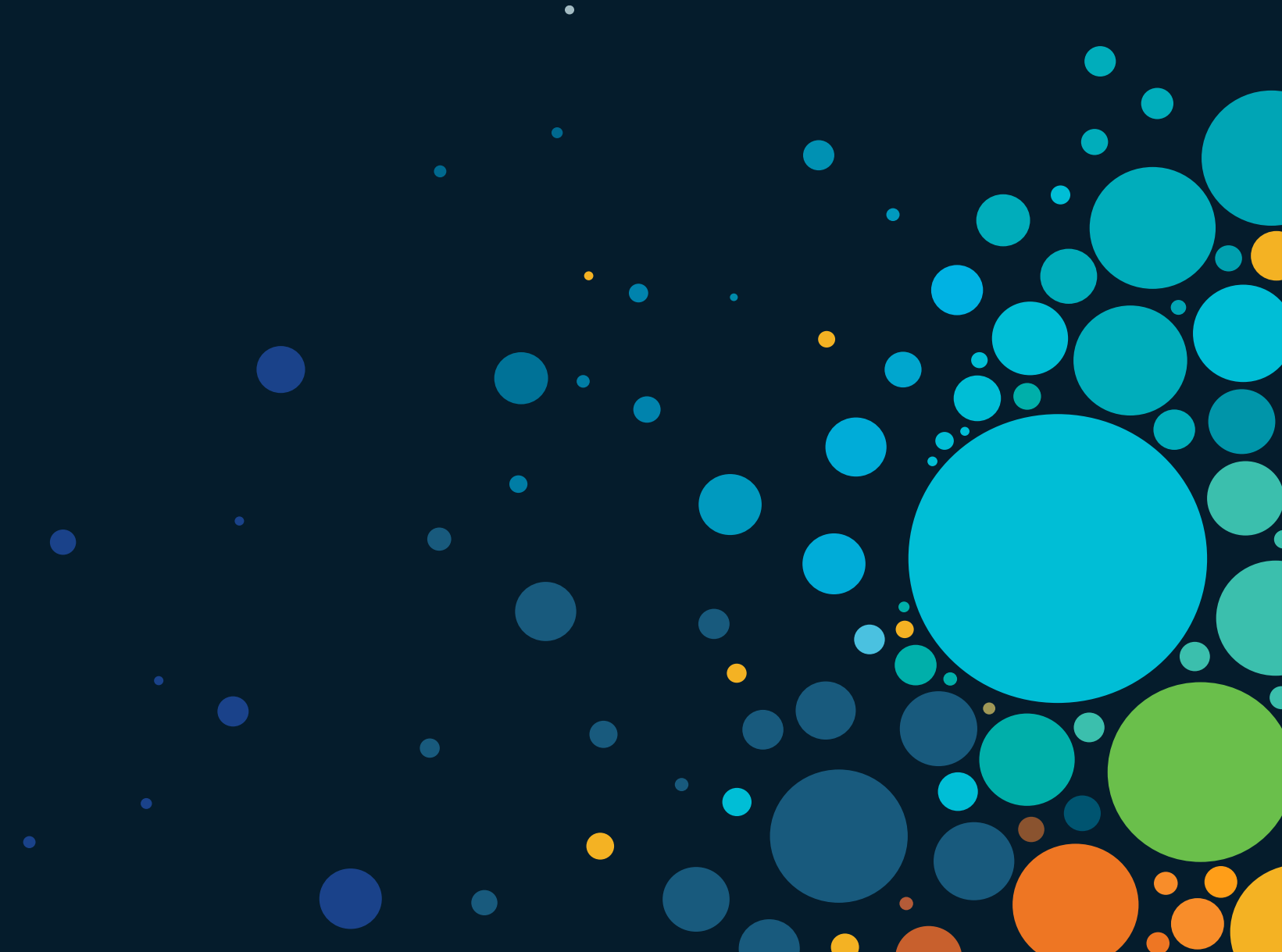
Map the ingress class to one of the traffic-class

**Egress Policy-map**

```
policy-map type queueing queue-policy
...
 class tc5
  random-detect discard-class-based
  random-detect discard-class 0 percent 80 90 5
  random-detect discard-class 1 percent 40 70 2
...
```

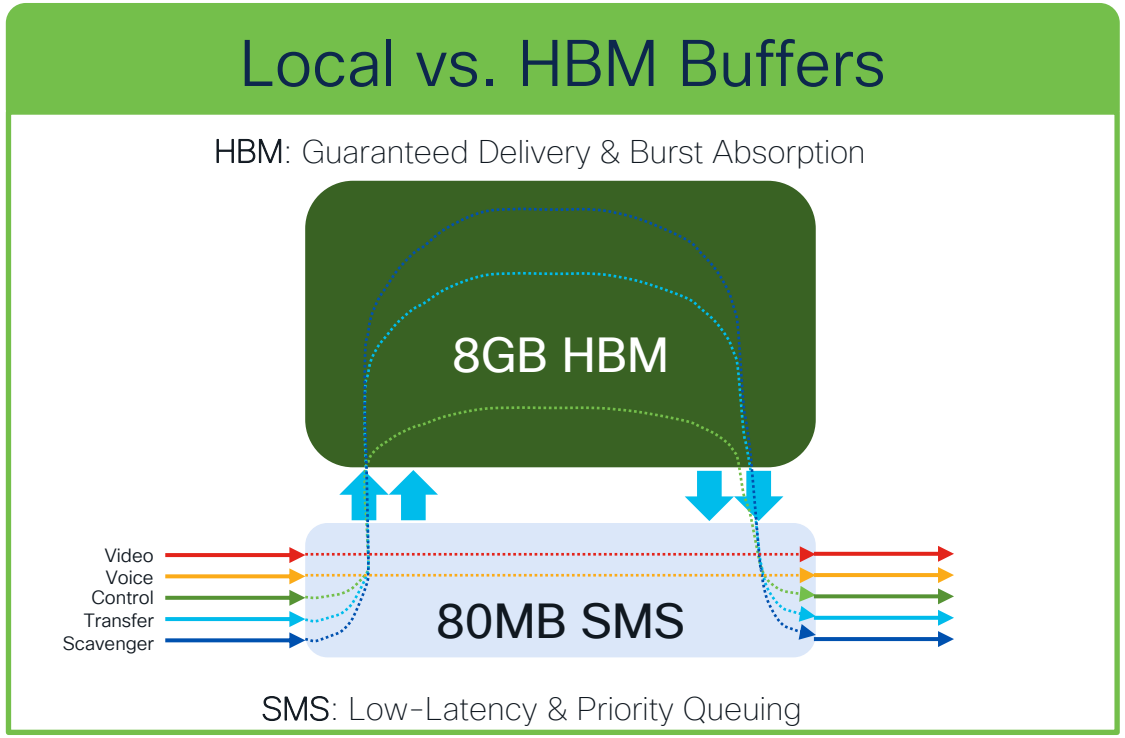
- Green traffic has higher Min and Max threshold comparing to yellow traffic
- Green traffic also has higher forwarding probability (lower drop probability) comparing to yellow traffic

# Buffers



# Silicon One Buffers

- Two different buffers to address two different requirements.
  - 80 MB of Shared Memory Sub-system (SMS) buffers:
    - Low latency packet queueing (video/voice packets)
    - Shallow specialized pool of buffers for quick queueing.
  - 8 GB of High Bandwidth Memory (HBM) buffers:
    - Deep pool of on-demand buffers for guaranteed delivery.
    - Reserve to absorb occasional bursts or address speed over-subscription between ingress and egress.

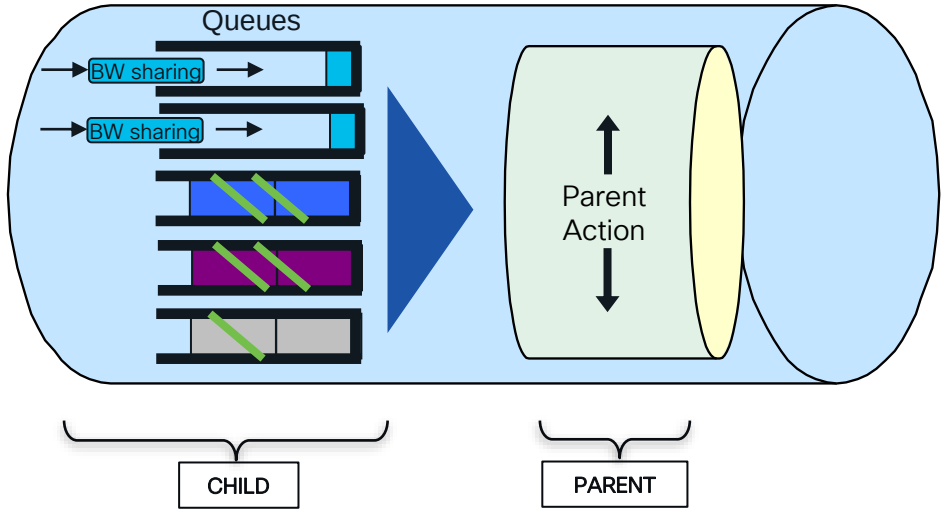



- Packet will always hit the SMS buffers first.
- SMS send the packet to HBM if additional buffers are needed.
- HBM CANNOT send the packet to the output queue, it has to be sent to the SMS again to be sent to the egress.

# HQoS

# Silicon One Hierarchical QoS (HQoS)

HQoS (two-level hierarchy) allows a parent and child policies on an interface for greater granularity. The Cisco Silicon One supports shaping as parent action.




| Child Action | Parent Action |
|---|---|
| Bandwidth sharing + Priority | Shaping |

# Silicon One vs UADP QoS

cisco Live!

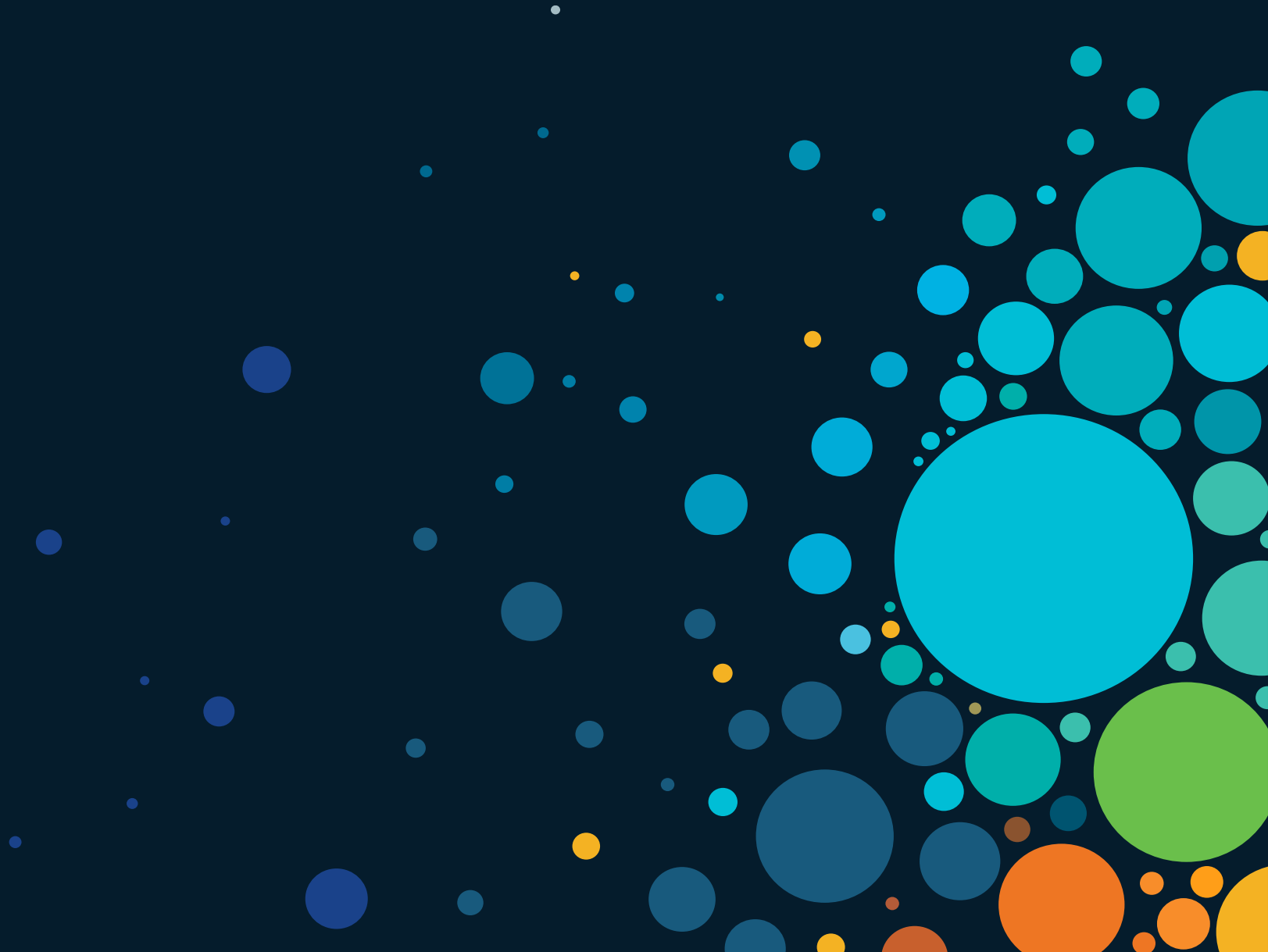
# QoS tools on UADP and Silicon One ASICs 1/2

| Features | UADP ASIC | Silicon One ASIC (Q200) |
|---|---|---|
| Trust | Trust all ports by default | Trust all ports by default |
| | | |
| Classification | Based on Packet header and ACL for both ingress and egress | Based on packet header and ACL for ingress<br>**Based on packet header for egress** |
| | | |
| Marking | Header, Table-map, QoS-Group for ingress<br>Header and table-map for egress | Header, Table-map, QoS-group, **traffic-class, discard-class** for ingress<br>Header and table-map for egress |
| | | |
| Policing | Both ingress and egress | Ingress only |

# QoS tools on UADP and Silicon One ASICs 2/2

| Features | UADP ASIC | Silicon One ASIC (Q200) |
|---|---|---|
| Queueing | Based on header or QoS group<br>Bandwidth and Bandwidth remaining | Based on **traffic-class**<br>**Bandwidth remaining** |
| | | |
| Buffering | Dedicated and shared buffer with DTS | SMS: Low-latency & priority queueing<br>HBM: Guaranteed Delivery & Burst Absorption |
| | | |
| Shaping | Egress | Egress |
| | | |
| Congestion Management | WTD: three thresholds per class<br>WRED: three thresholds' pairs per class | WTD: **two** threshold per class<br>WRED: **two** thresholds' pairs per class |

# Migration to Silicon One



# Config Migration Philosophy

1. Define the problem/behavior addressed with QoS.

   Simply copy-pasting existing configs between platform families will always throw errors due to differences in syntax and supported actions between platforms.

2. Determine the number of queues you need. Reduce if existing config has more than eight.

   Broad generalized splits often are more efficient than granular splits. Always map your qos values to corresponding traffic classes.

3. How many classes do you want to have strict priority enabled? Up to 7 supported strict priority queues.

   Know what strict means. All traffic coming into it will be serviced at the expense of other classes.

4. Define traffic shaping or sharing between queues

   Shape priority queues. Use weights to control bandwidth sharing with remaining queues

5. Do you want to modify/change WRED parameters.

   Advanced configuration options, not required for most use cases.

# Migration from Catalyst 6K to Silicon One

**Catalyst 6K Configuration**

```
class-map type lan-queuing match-all REALTIME
 match dscp ef
class-map type lan-queuing match-all NETWORK_CONTROL
 match dscp cs6  cs7
class-map type lan-queuing match-all VIDEO
 match dscp cs3  af31  af32  af33
…
policy-map type lan-queuing CAMPUS_EGRESS_6800_POLICY
 class type lan-queuing REALTIME
    priority level 1
 class type lan-queuing NETWORK_CONTROL
    bandwidth remaining percent 10
 class type lan-queuing VIDEO
    bandwidth remaining percent 20
 class class-default
     random-detect dscp-based
     random-detect dscp af11 percent 80 100
```

1. Classified Based on DSCP value
2. 4 classes (3 defined + default)

3. 4 queues
4. 1 priority queue
5. Scheduling is WRR with “bandwidth remaining”
6. Congestion management is WRED with the default class

# Config Migration from Catalyst 6K to Silicon One

**Catalyst 6K Configuration**

```
class-map type lan-queuing match-all REALTIME
 match dscp ef
class-map type lan-queuing match-all NETWORK_CONTROL
 match dscp cs6  cs7
class-map type lan-queuing match-all VIDEO
 match dscp cs3  af31  af32  af33
…
policy-map type lan-queuing CAMPUS_EGRESS_6800_POLICY
 class type lan-queuing REALTIME
    priority level 1
 class type lan-queuing NETWORK_CONTROL
    bandwidth remaining percent 10
 class type lan-queuing VIDEO
    bandwidth remaining percent 20
 class class-default
     random-detect dscp-based
     random-detect dscp af11 percent 80 100
```

1. Classified Based on DSCP value
2. 4 classes (3 defined + default)
3. 4 queues (traffic-class), traffic-7 is priority level 1

Apply policy on the ingress interface

**Catalyst 9K(Q200) Configuration**

```
class-map match-all REALTIME
 match dscp ef
class-map match-all NETWORK_CONTROL
 match dscp cs6  cs7
class-map match-all VIDEO
 match dscp cs3  af31  af32  af33
class-map match-all default-green
 match dscp af11


policy-map INGRESS
 class REALTIME
    set traffic-class 7
 class NETWORK_CONTROL
    set traffic-class 6
 class VIDEO
    set traffic-class 5
 class default-green
    set traffic-class 0
 class class-default
    set discard-class 1
```

Note: class-default is always assigned with traffic-class 0

# Config Migration from Catalyst 6K to Silicon One

Map the traffic-class marking defined on the ingress

Apply policy on the egress interface

**Catalyst 6K Configuration**

```
policy-map type lan-queuing CAMPUS_EGRESS_6800_POLICY
 class type lan-queuing REALTIME
    priority level 1
 class type lan-queuing NETWORK_CONTROL
    bandwidth remaining percent 10
 class type lan-queuing VIDEO
    bandwidth remaining percent 20
 class class-default
     random-detect dscp-based
     random-detect dscp af11 percent 80 100
```

**Catalyst 9K(Q200) Configuration**

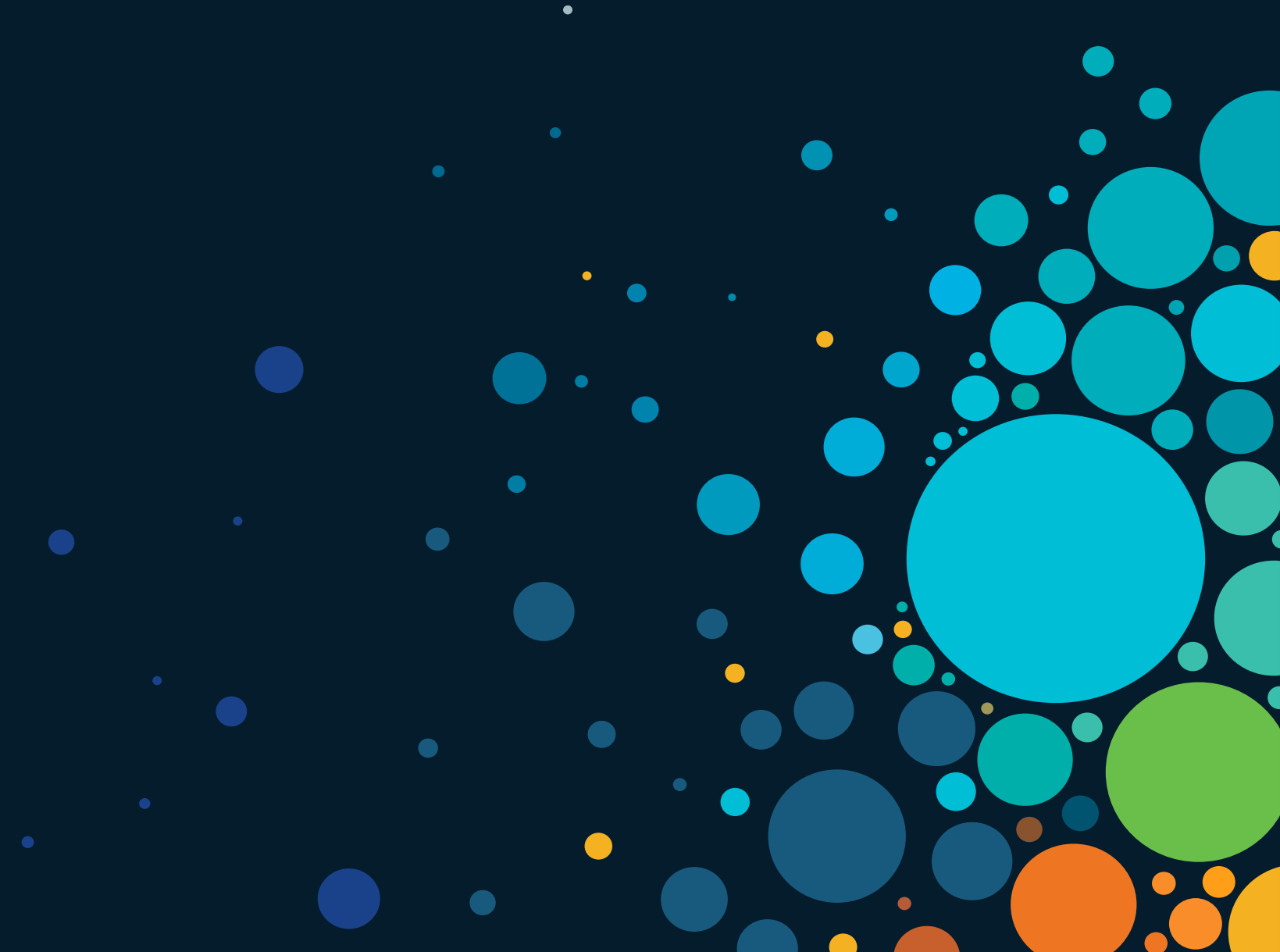
```
class-map tc7
 match traffic-class 7
class-map tc6
 match traffic-class 6
class-map tc5
 match traffic-class 5

policy-map type queuing EGRESS
 class tc7
    priority level 1
 class tc6
    bandwidth remaining ratio 1
 class tc5
    bandwidth remaining ratio 2
 class class-default
    random-detect discard-class-based
    random-detect discard-class 0 percent 80 100
    random-detect discard-class 1 percent 40 100
```

3. 1 priority queue
4. Scheduling is WRR with “bandwidth remaining”
5. Congestion management is WRED with the default class

Note: C6K WRED default min is 40, max is 100.

# Summary



# Why QoS in campus?

User Experience

Guaranteeing voice quality

Bandwidth Savvy Business Applications

protect network infrastructure to deal with abnormal events

Video Quality

de-prioritizing non-business applications

protecting the control planes

# Technical Session Surveys

- Attendees who fill out a minimum of four session surveys and the overall event survey will get Cisco Live branded socks!
- Attendees will also earn 100 points in the Cisco Live Game for every survey completed.
- These points help you get on the leaderboard and increase your chances of winning daily and grand prizes.

# Cisco Learning and Certifications

From technology training and team development to Cisco certifications and learning plans, let us help you empower your business and career. www.cisco.com/go/certs

**Pay for Learning with Cisco Learning Credits**

(CLCs) are prepaid training vouchers redeemed directly with Cisco.

## Learn

**Cisco U.**
IT learning hub that guides teams and learners toward their goals

**Cisco Digital Learning**
Subscription-based product, technology, and certification training

**Cisco Modeling Labs**
Network simulation platform for design, testing, and troubleshooting

**Cisco Learning Network**
Resource community portal for certifications and learning

## Train

**Cisco Training Bootcamps**
Intensive team & individual automation and technology training programs

**Cisco Learning Partner Program**
Authorized training partners supporting Cisco technology and career certifications

**Cisco Instructor-led and Virtual Instructor-led training**
Accelerated curriculum of product, technology, and certification courses

## Certify

**Cisco Certifications and Specialist Certifications**
Award-winning certification program empowers students and IT Professionals to advance their technical careers

**Cisco Guided Study Groups**
180-day certification prep program with learning and support

**Cisco Continuing Education Program**
Recertification training options for Cisco certified individuals

Here at the event? Visit us at **The Learning and Certifications lounge at the World of Solutions**

# Continue your education

- Visit the Cisco Showcase for related demos
- Book your one-on-one Meet the Engineer meeting
- Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs
- Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand

cisco Live!


cisco
The bridge to possible
Thank you
cisco Live!
#CiscoLive

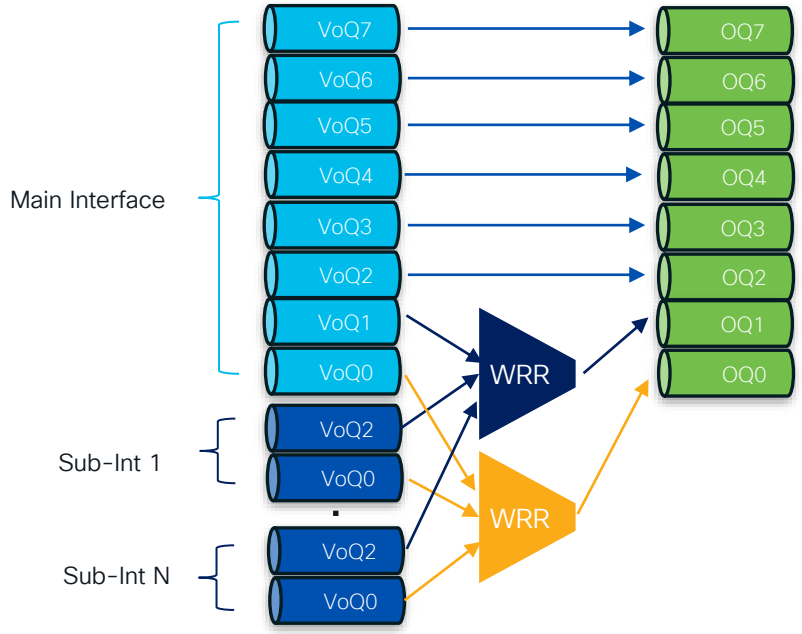

#CiscoLive

# Sub-interface queuing

Option 1

## 8 Queues main interface + 2 Queues sub-interfaces

- Main interface
  - 8 queues (up to 7 priority)
- Sub-interfaces
  - 2 queues (no priority)
- Main interface policy applies when no policy applied on sub-interface




Option 2

## 2 Queues both main and sub-interfaces

- Main interface & sub-interfaces
  - 2 queues (1 priority and 1 normal)
- CLI "**`queuing mode sub-interface priority-propagation`**" under the main interface to enable it
- Main interface policy applies when no policy applied on sub-interface
- Sub-interfaces share the BW with WRR
  - Can use CLI "*bandwidth ratio <N>*" to give additional BW to certain sub-interfaces