SEMTECH

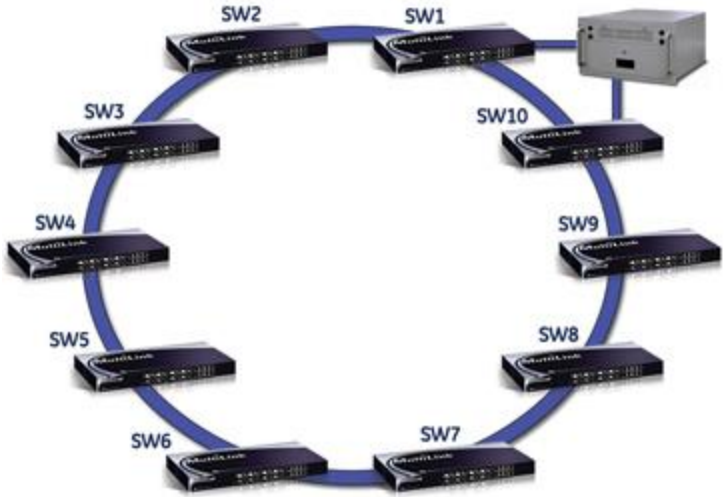
www.semtech.com

Robust Time Delivery – Turning a Wish Into a Reality (How To Use PTP With A Ring Architecture)

David Spencer – WSTS 2014

Ring Architectures

- **Ring architectures are already common for data transport**
  - Traffic can pass from any node to any other node through one of two routes
    - Traditionally called East and West (viewed from bottom of ring)
  - Provides alternate route in event of single failure
    - As long as ring is geographically separated
- **For communication between any pair of nodes one path is likely preferred**
  - Normally the shortest one
  - But other path provides a reasonable backup
  - Different for different nodes
    - For example SW9 would talk to SW1 anticlockwise while SW5 would talk to SW1 clockwise
- **But how does this fit in with PTP?**
  - Does a ring of boundary clocks work?



SEMTECH

www.semtech.com

Simple BC Ring

Boundary Clock Ring: Node Architectures
A 'normal' Boundary clock

SEMTECH

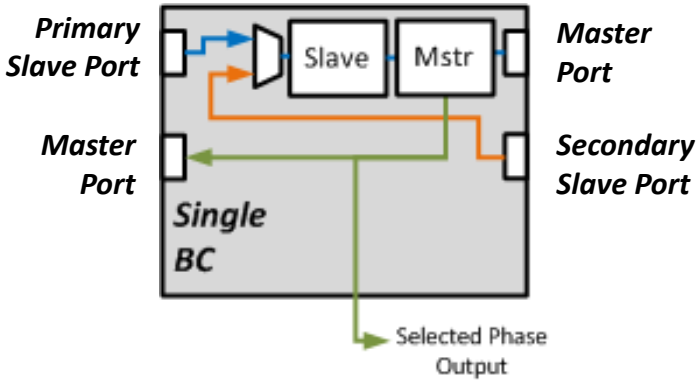
Standard BC:

- A PTP slave port locks to a selected master via a selected port creating a timebase.
- The Master, all egress ports *and* the selected output are tied to the slave timebase.
- The slave port may or may not be able to hitlessly switch on failure.

Ports can be physical or logical

Let's consider a choice of two ports for use as sources of PTP (slave ports)

PTP master function will egress through at least two ports.

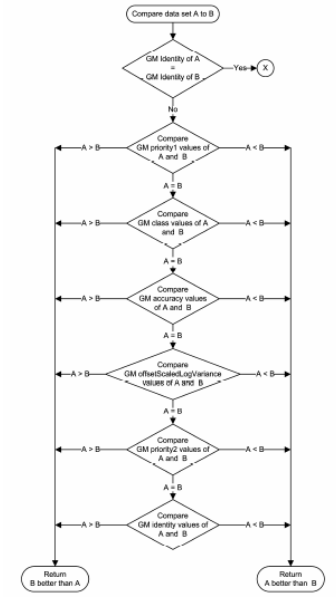


Let's call this a 'Single BC'

The BMCA – A Quick Refresher

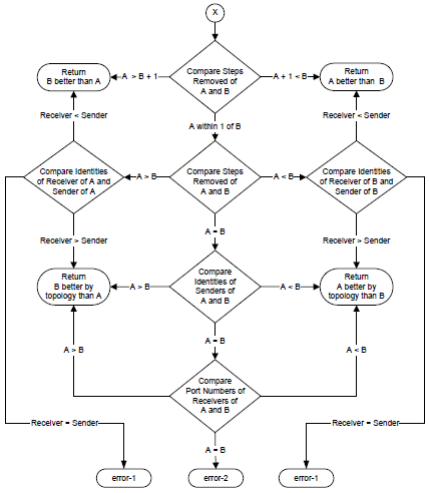
❑ Mechanism to select best master

- When multiple masters are available selects the best one based on number of factors
  - Programmed priorities
  - Characteristics of clock source

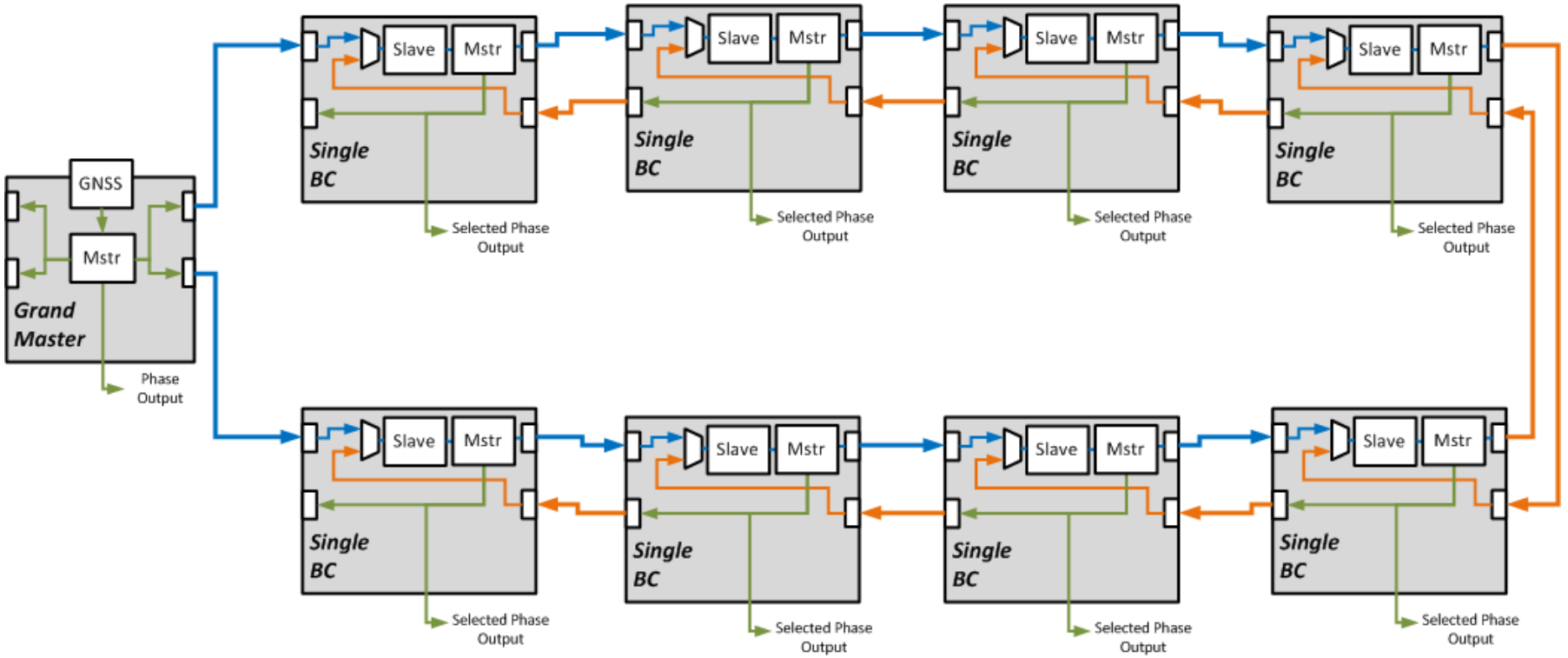


❑ What about a ring

- Same master will be seen from both sides
- Second stage of BMCA handles this
  - Looks at Steps Removed field
  - Prefers links through least number of BCs
- Path closest to GM will be selected
  - Either east or west depending on where in ring

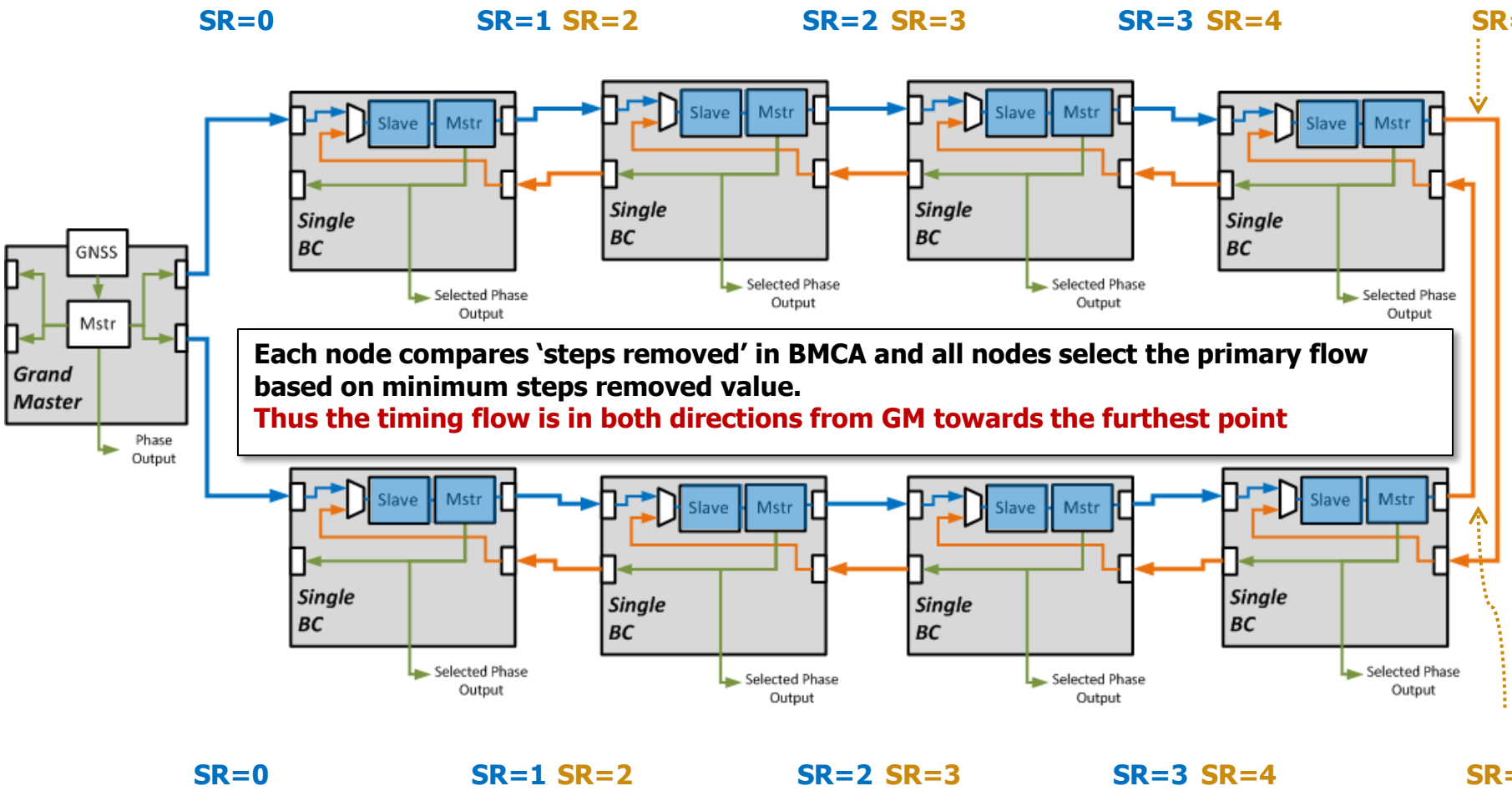


Boundary Clock Architecture:
Conceptual- Single GM, 8-BC node ring



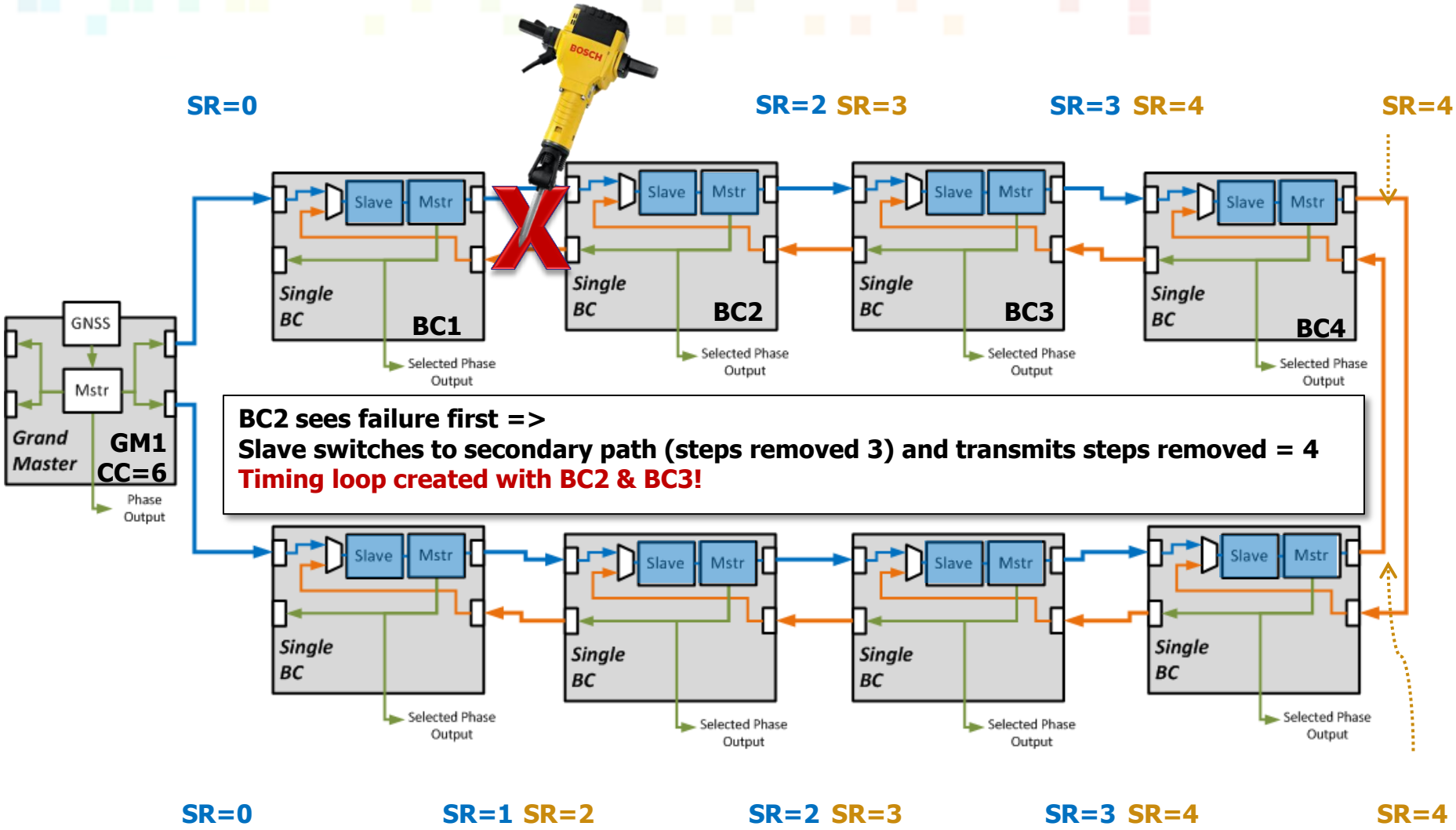
Each node has two potential masters- selected in Slave by BMCA
The Slave's timebase is used for all the egress ports and the output

Boundary Clock Architecture:
Single GM, BC nodes, with transmitted 'steps removed'

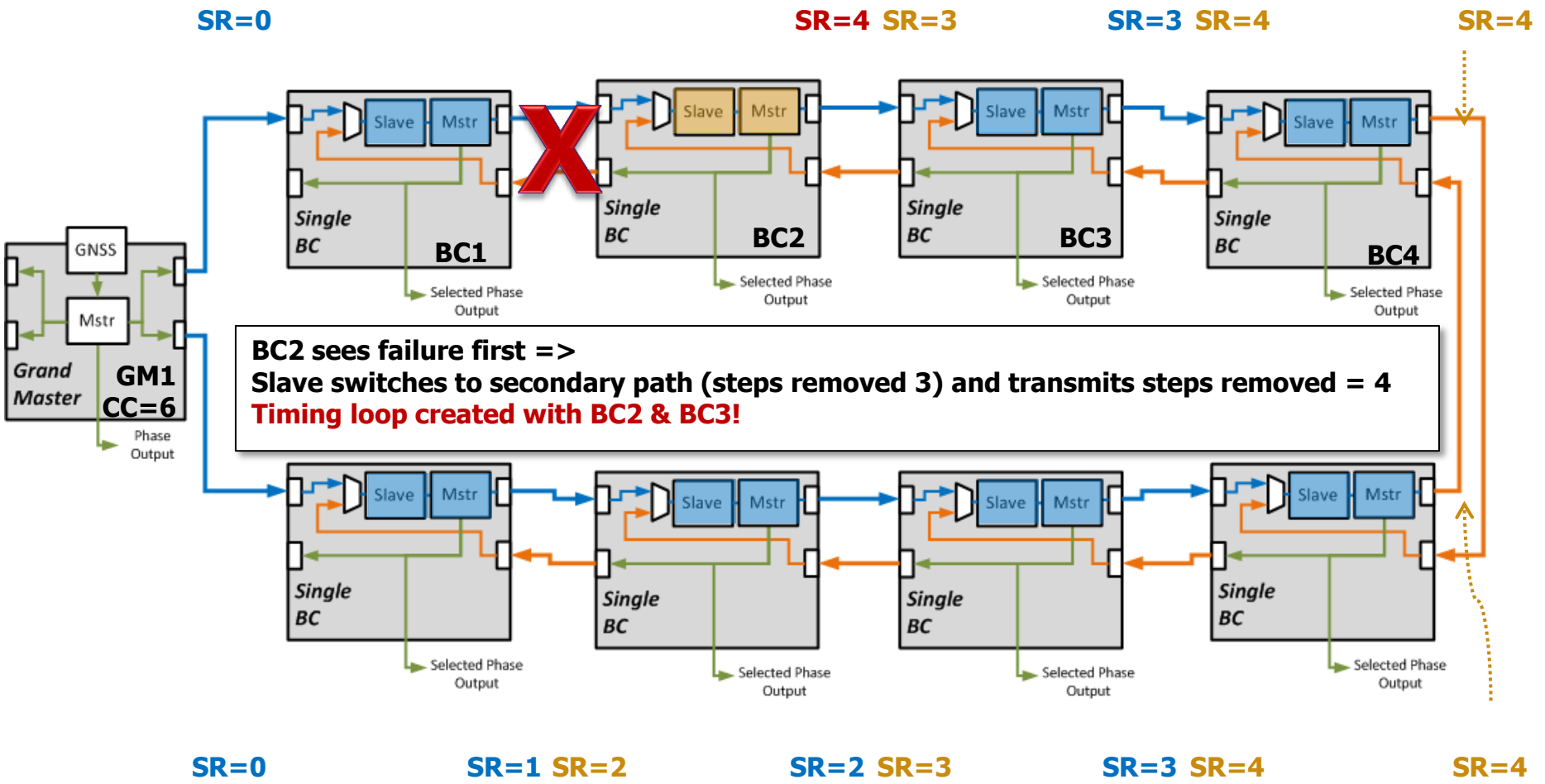


Boundary Clock Architecture:

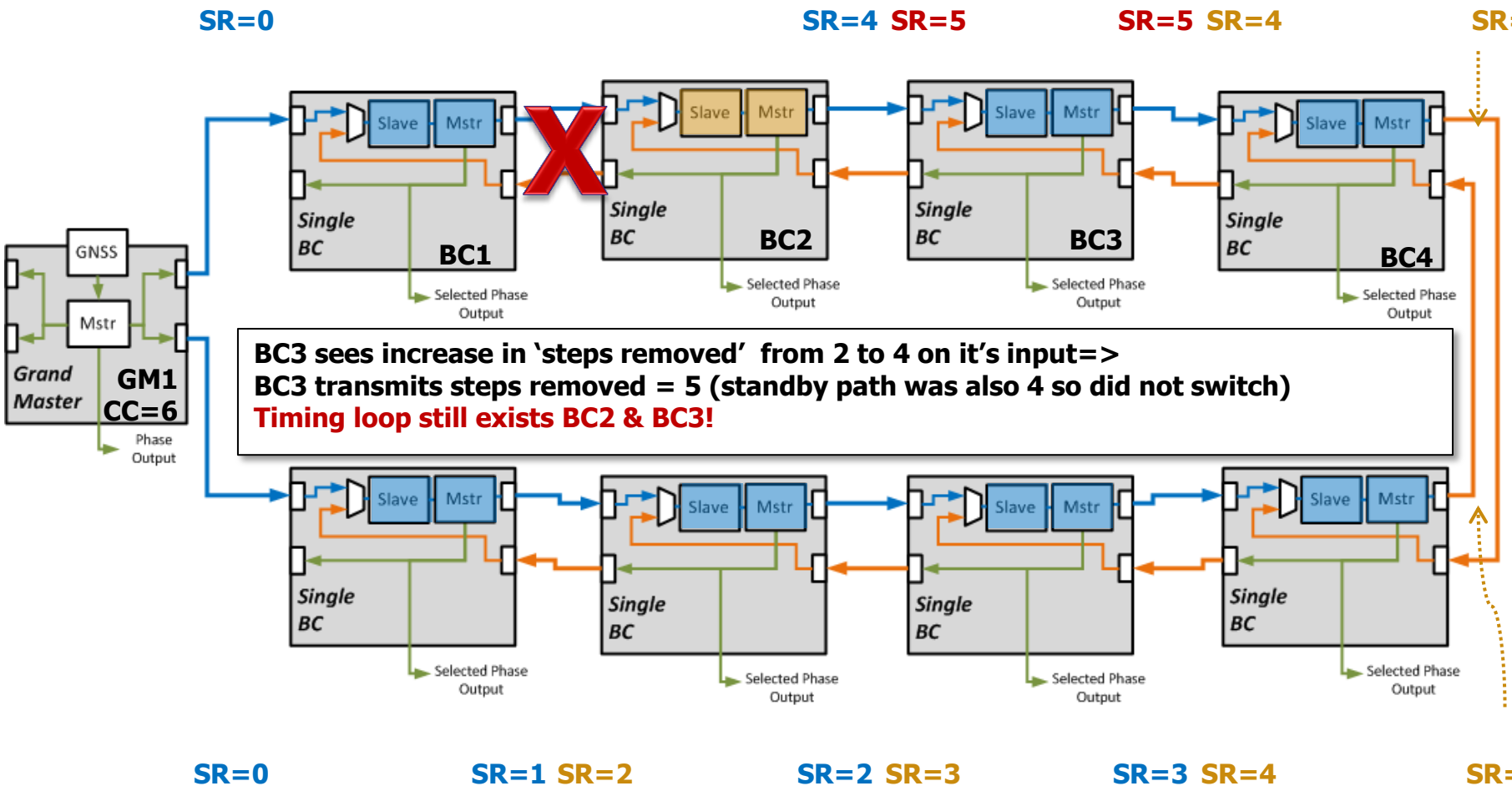
Failure: first action, everything controlled by SR



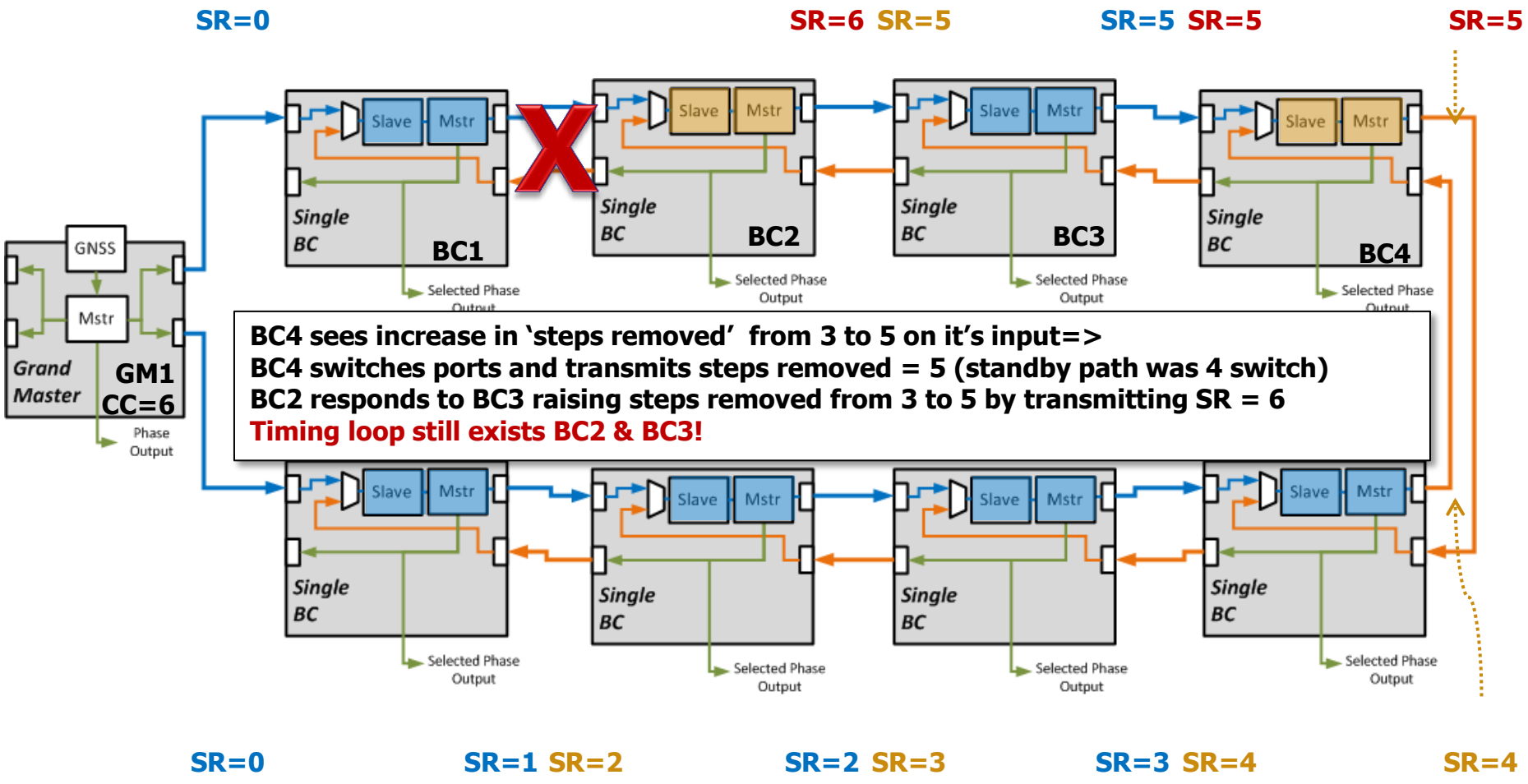
Boundary Clock Architecture:
Failure: first action, everything controlled by SR



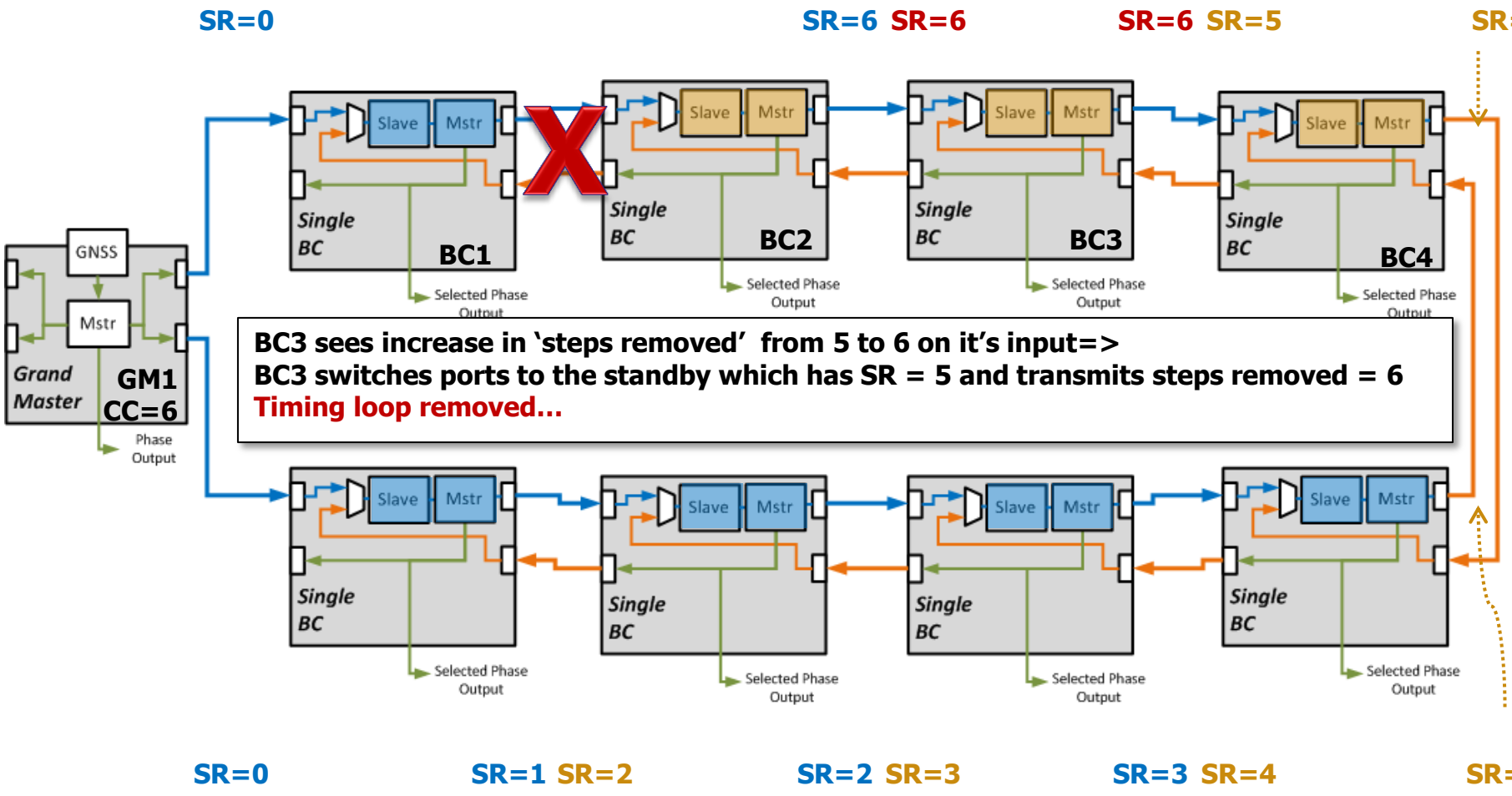
Boundary Clock Architecture:
Failure: second action



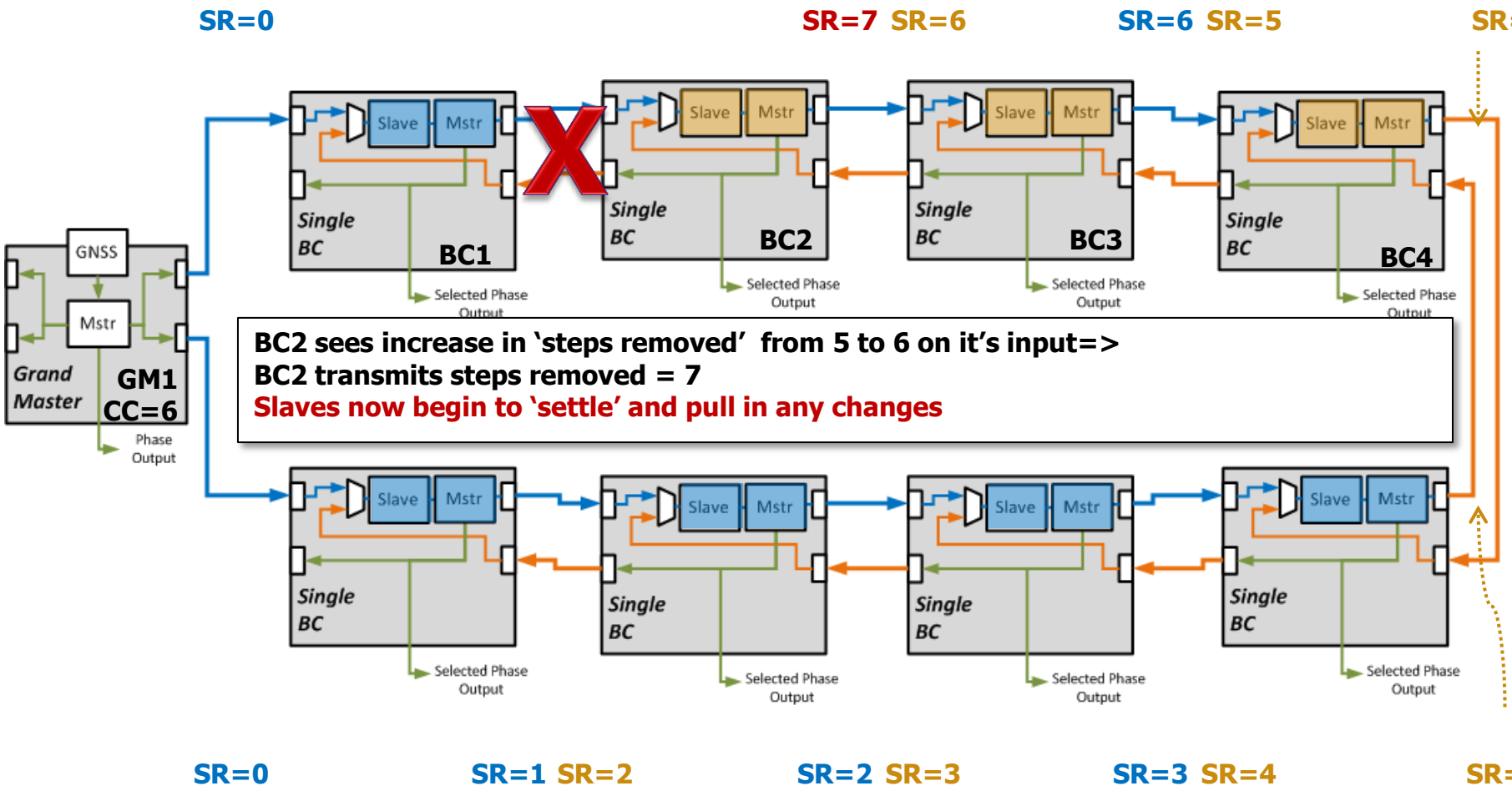
Boundary Clock Architecture:
Failure occurs: third action



Boundary Clock Architecture:
Failure occurs: Fourth action



Boundary Clock Architecture:
Failure occurs: Final action

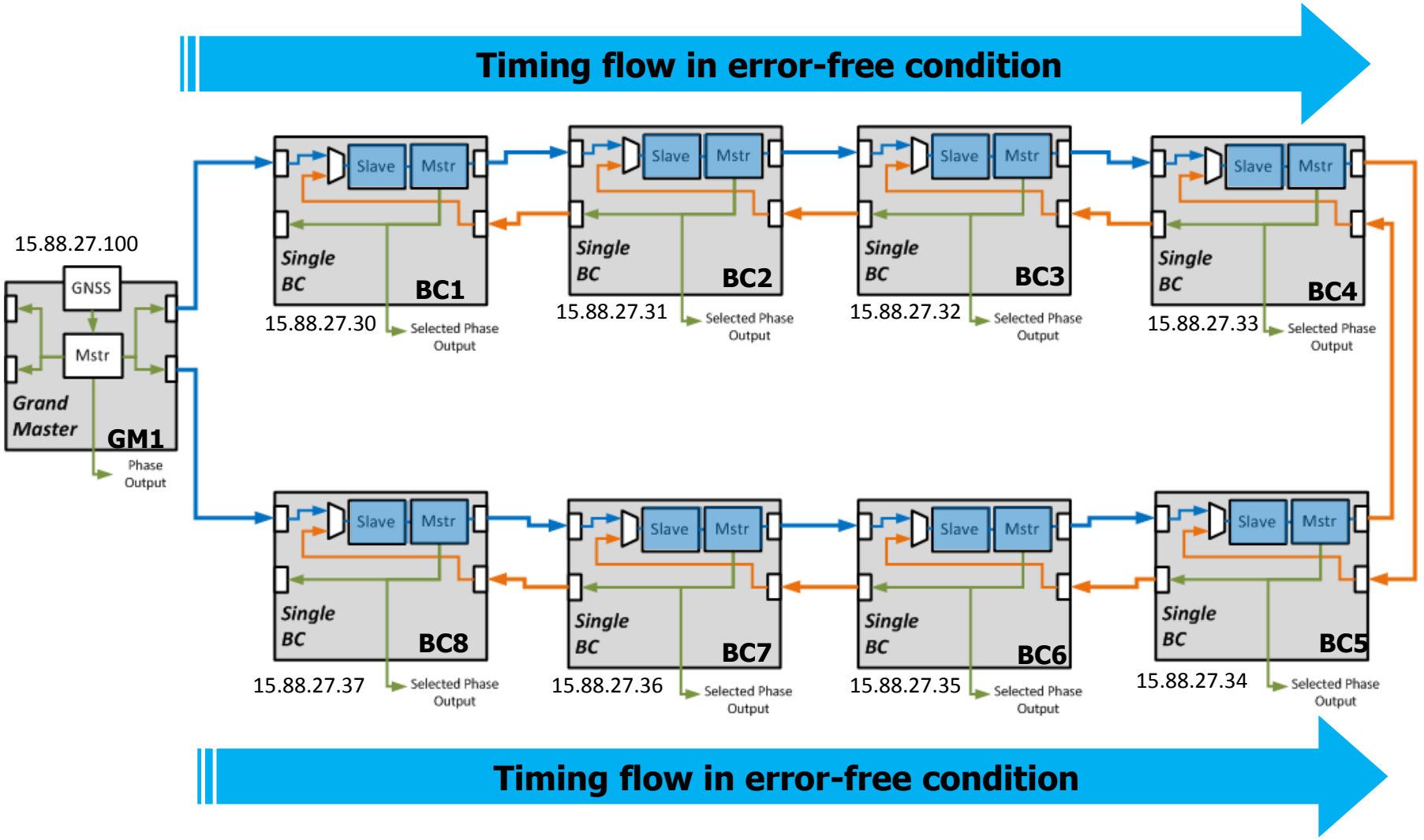


SEMTECH

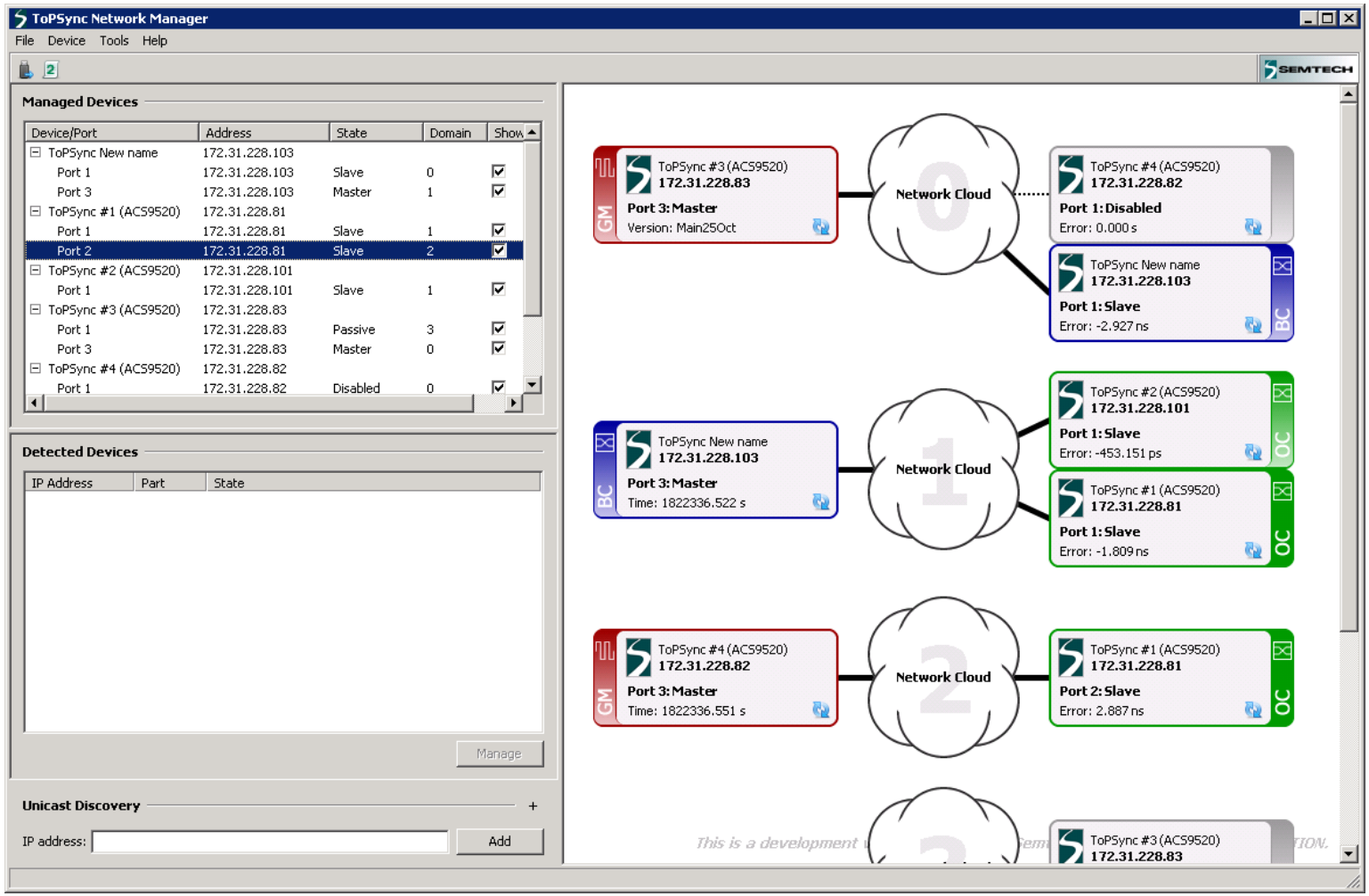
www.semtech.com

Single BC ring Lab test

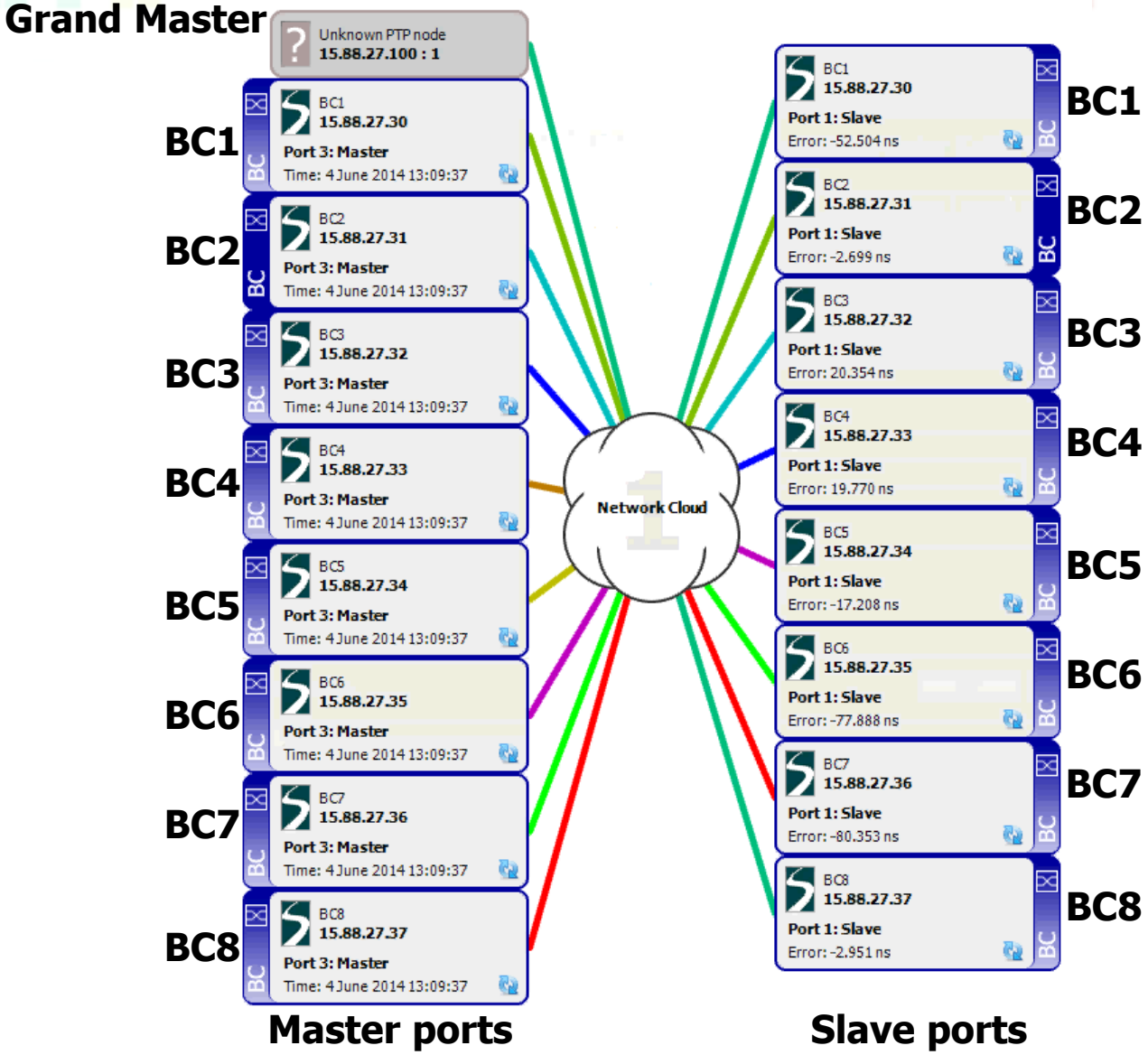
Boundary Clock Architecture:
Single GM, Lab Setup Initial state



Network Manager Tool:
A Display of Networked PTP elements



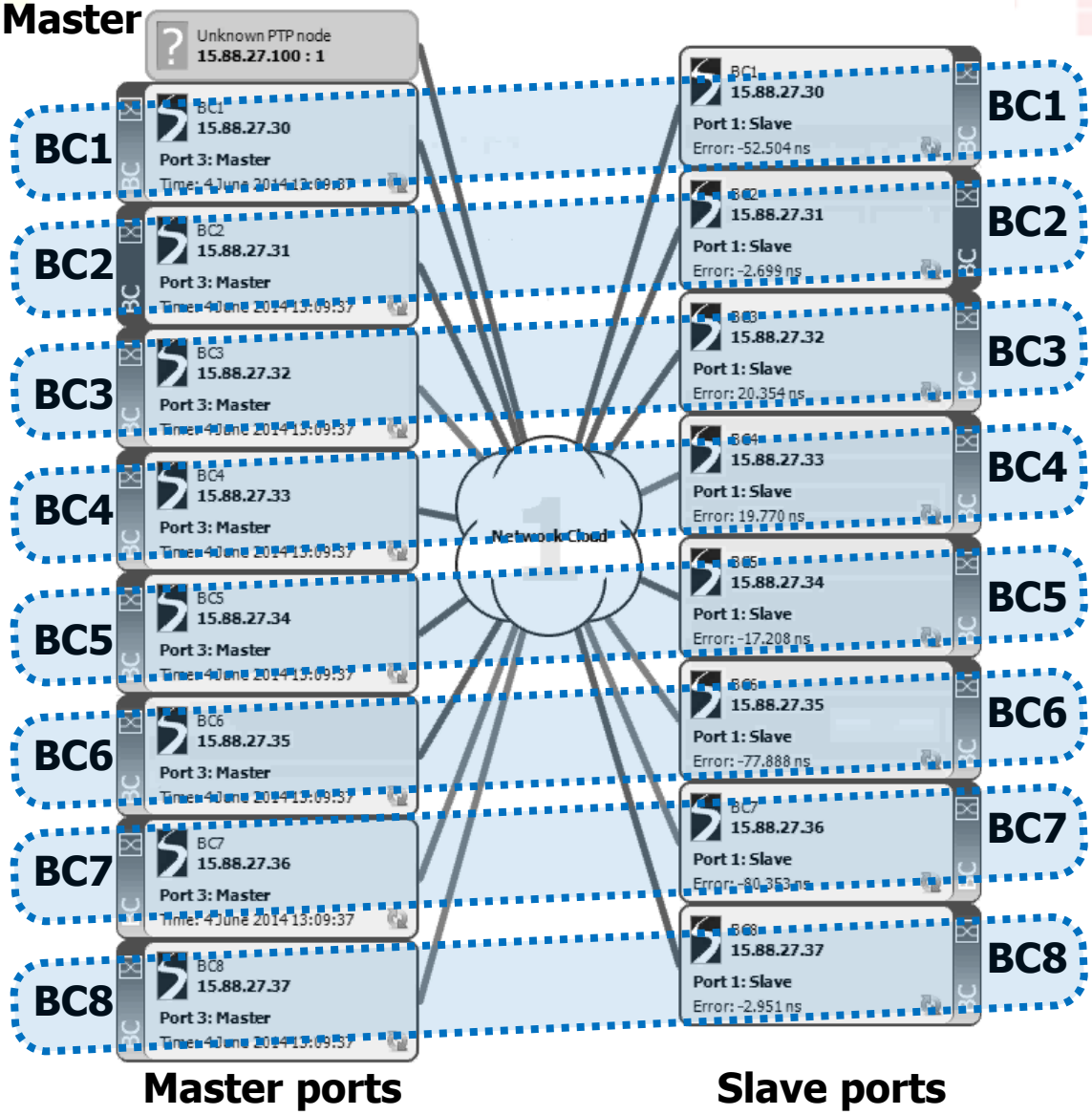
Initial Lab Setup: GUI screenshot



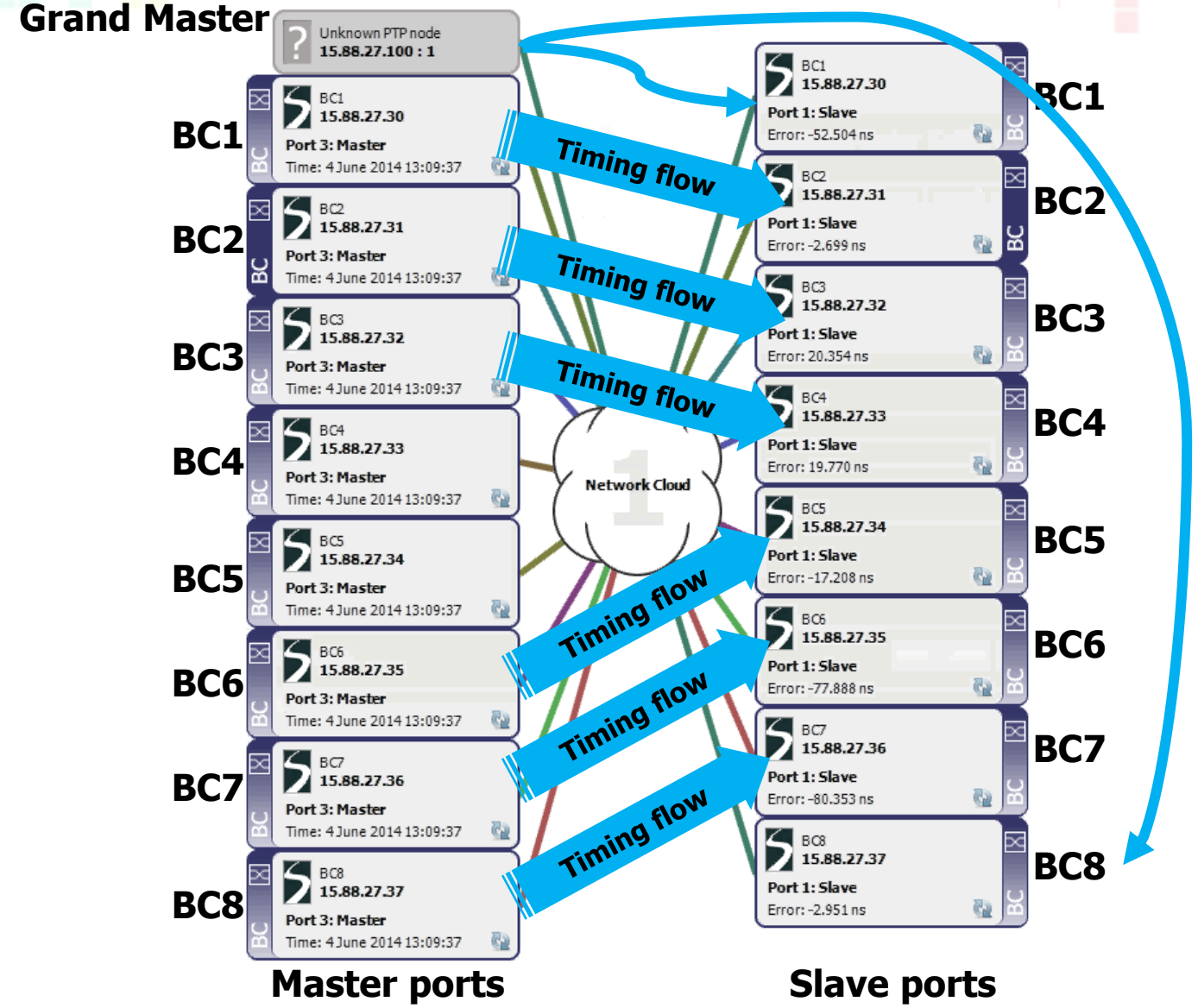
Each BOX is a PTP *PORT*

Each BC has two PORTS- one is Master & one is Slave

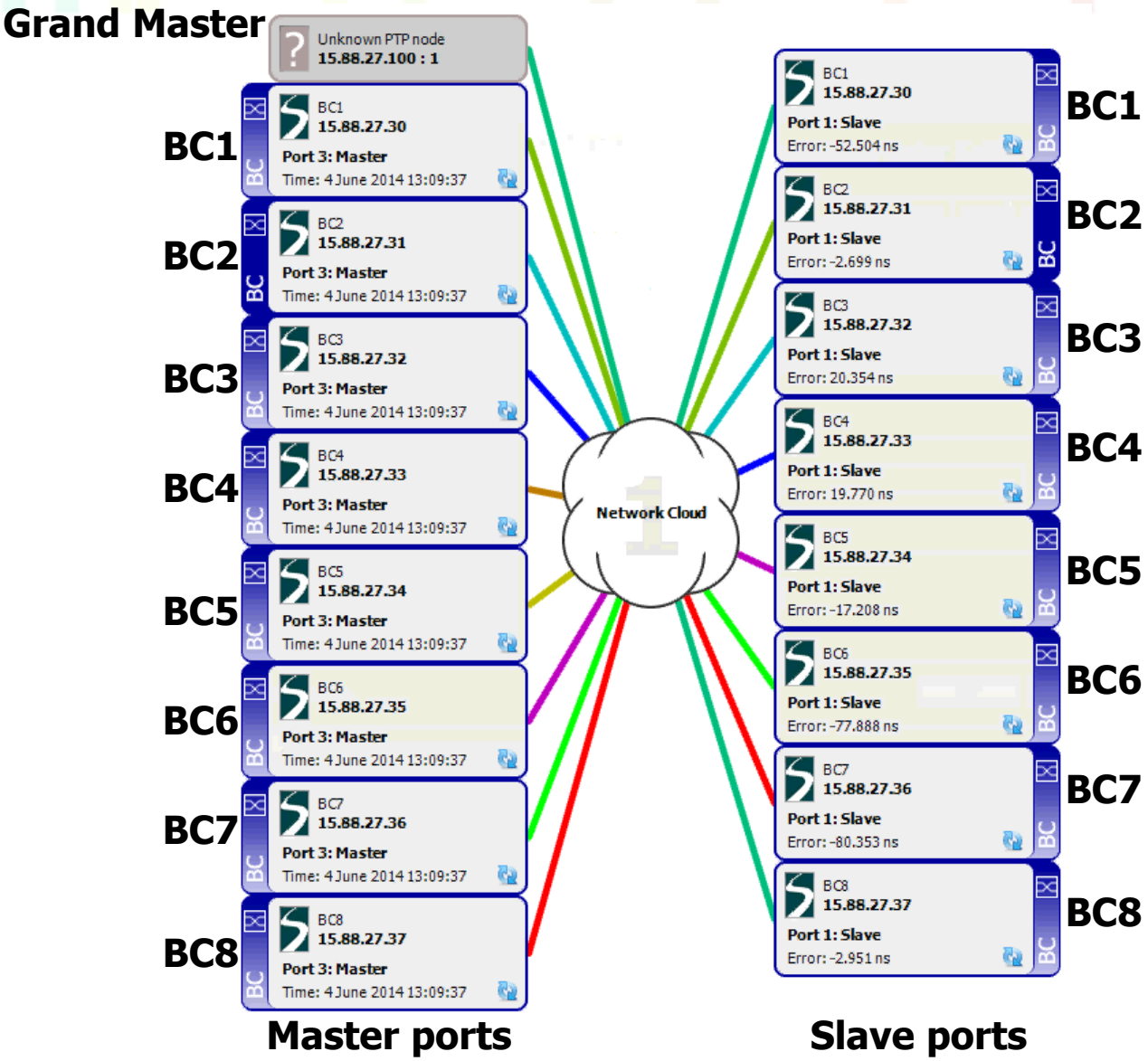
Single BC loop: Initial Lab Setup
Semtech GUI screenshot



Single BC loop: Initial Lab Setup
Semtech GUI screenshot



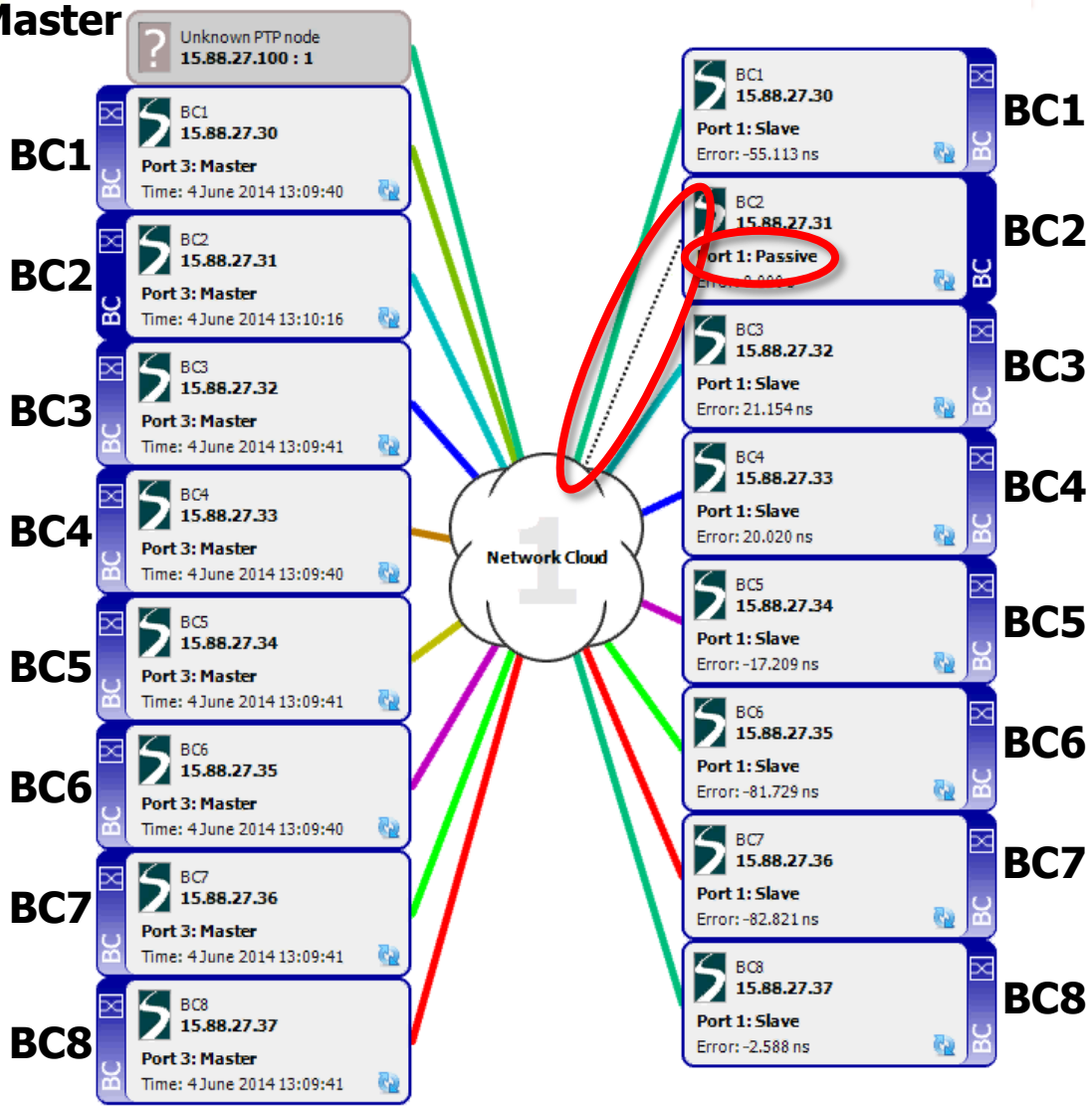
Initial Lab Setup: GUI screenshot



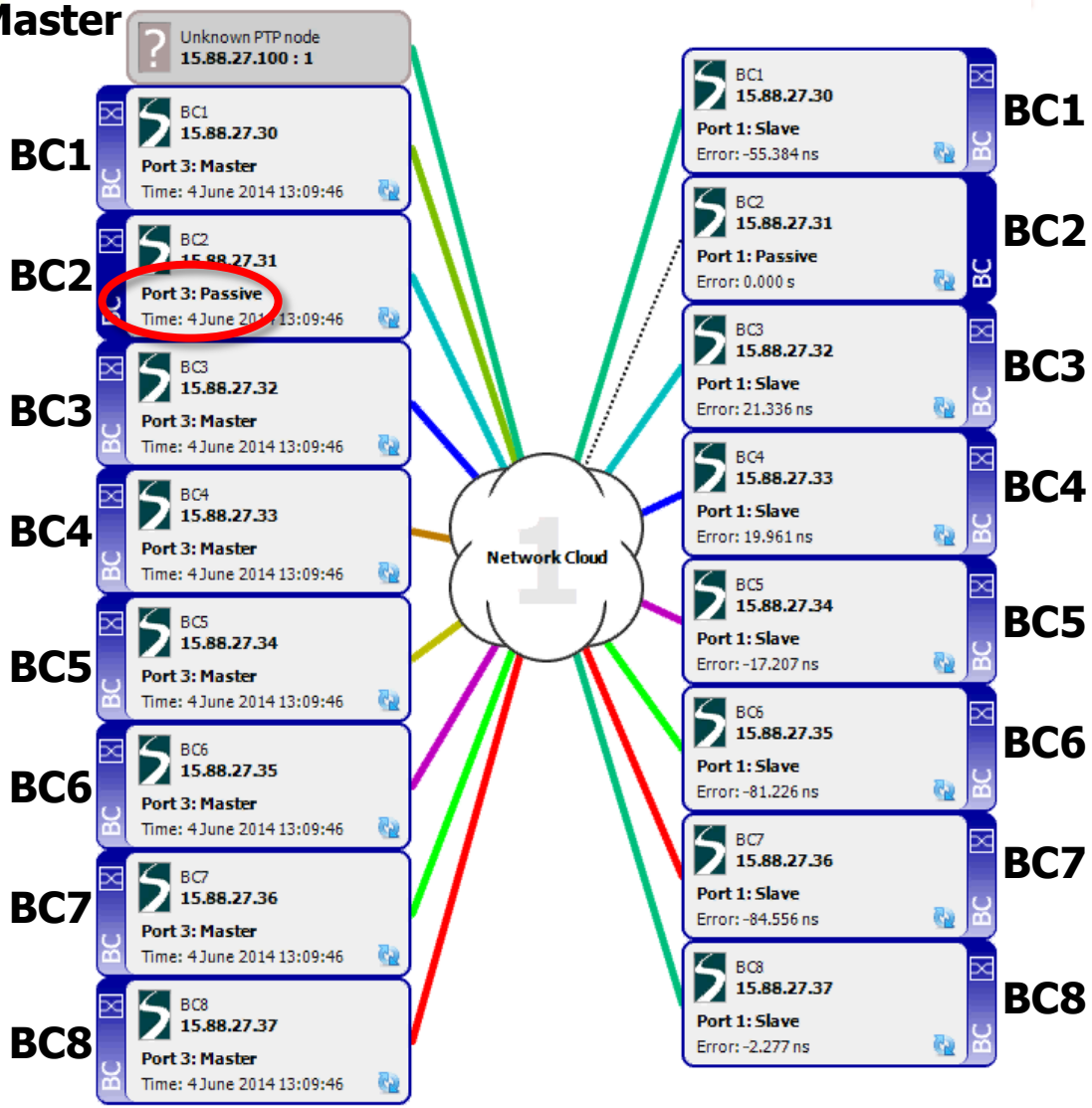
Line color denotes selected master/slave link

Failure + 2s

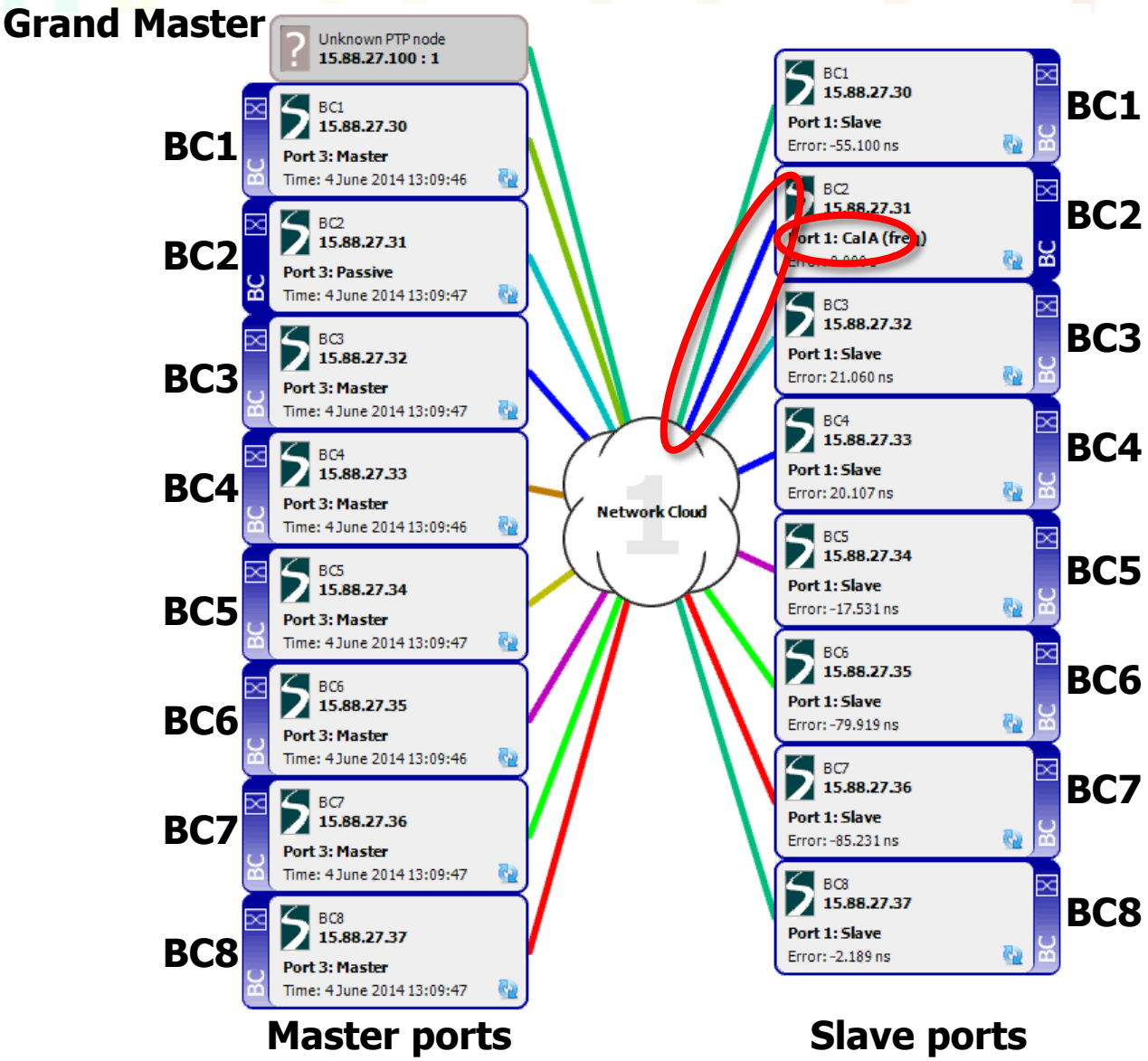
Initial Failure: BC1 to BC2
BC2 slave port goes passive



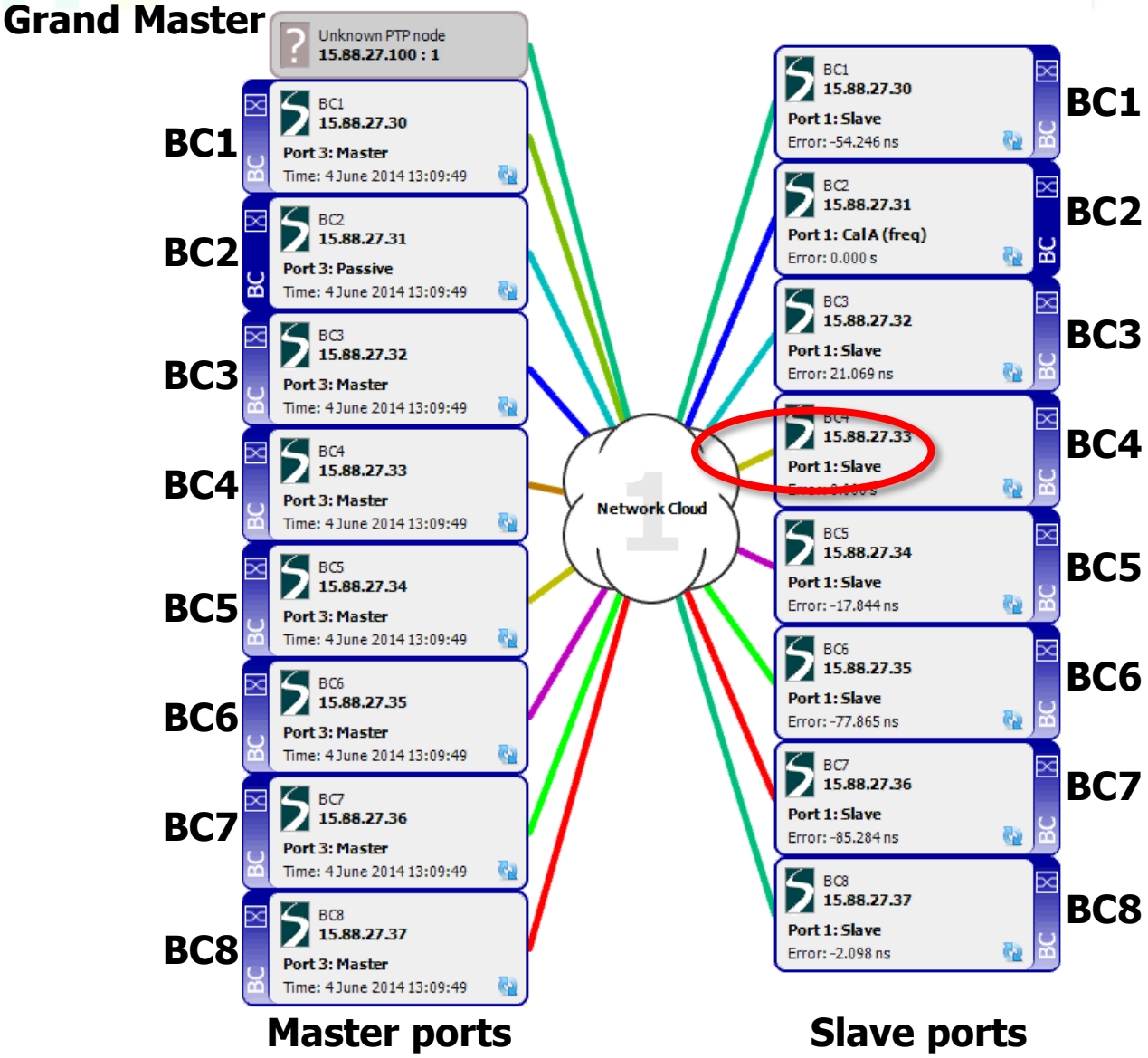
Initial Failure: BC1 to BC2
BC2 Master port goes passive



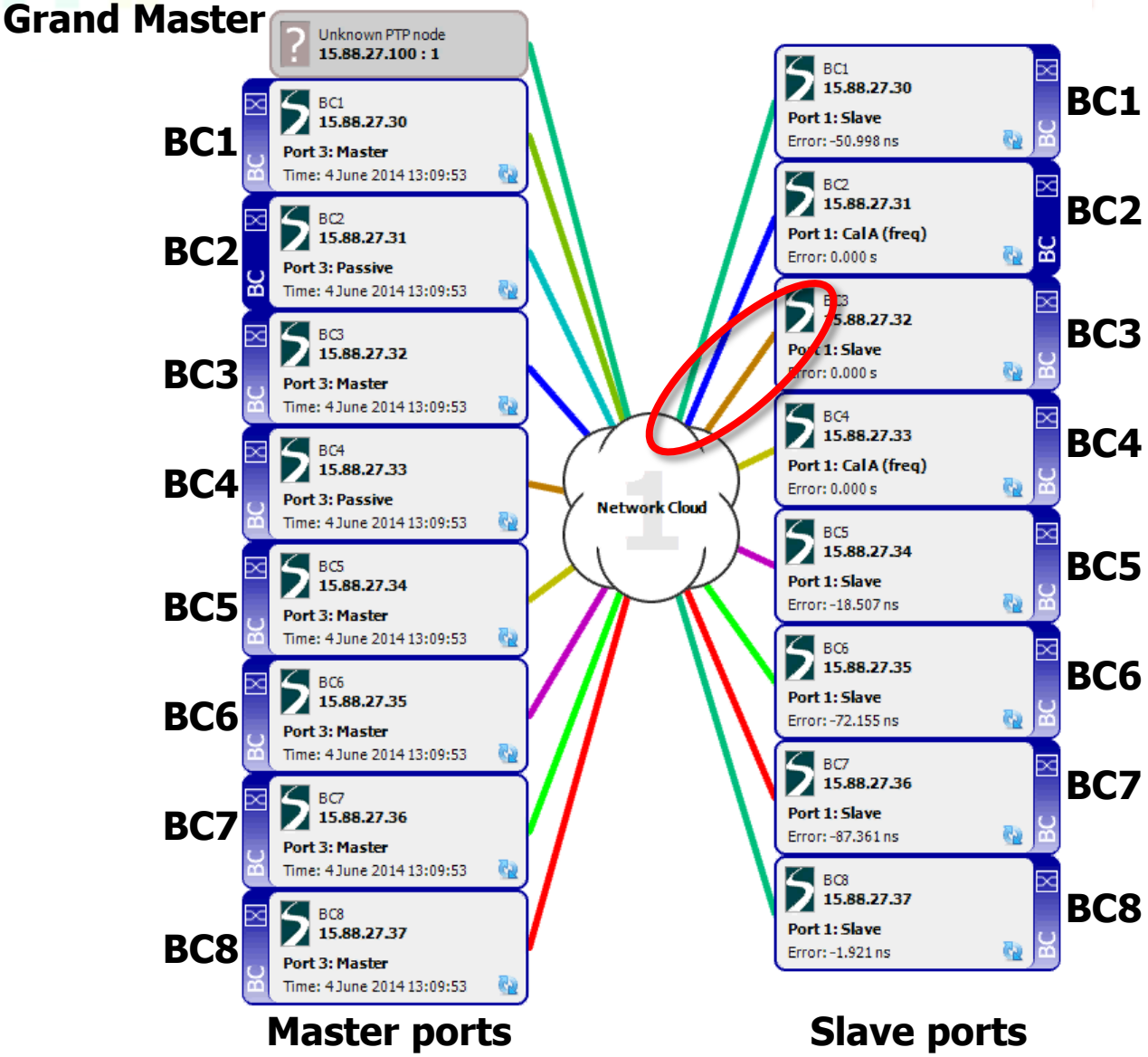
Initial Failure: BC1 to BC2
BC2 Slave port switches from BC1 to BC3



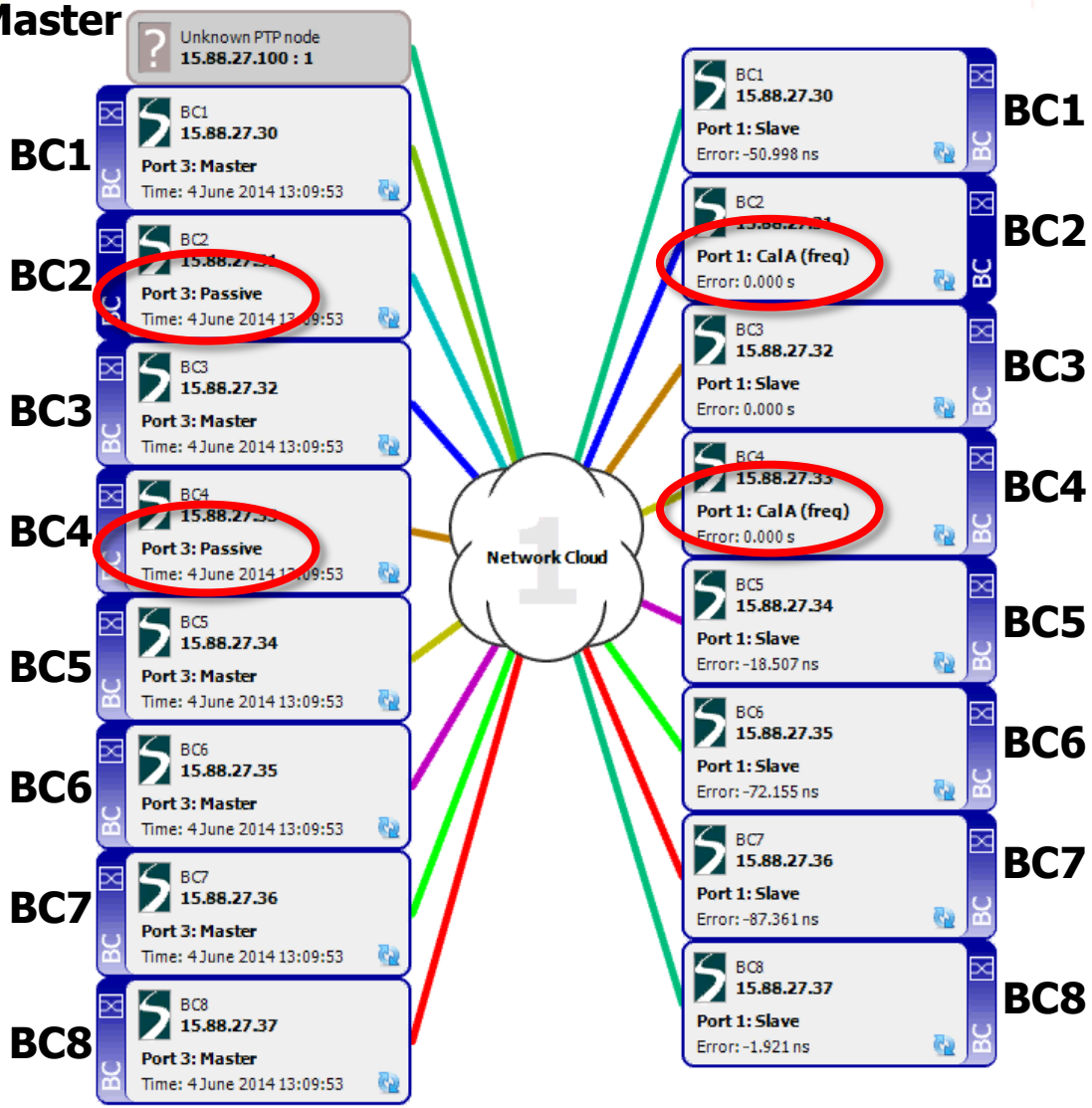
Initial Failure: BC1 to BC2
BC4 Slave port switches from BC3 to BC5



Initial Failure: BC1 to BC2
BC3 Slave port switches from BC2 to BC4

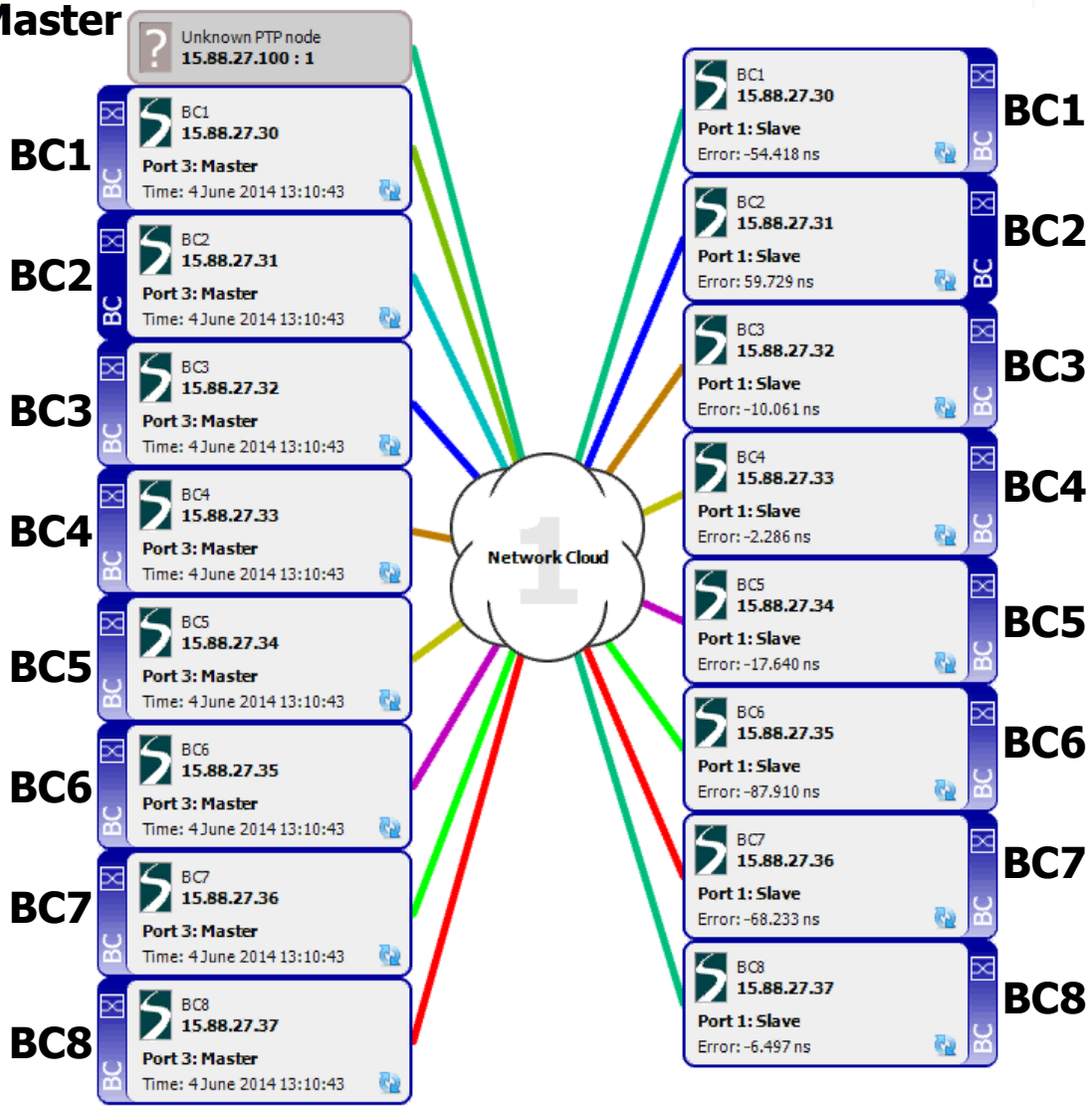


Initial Failure: BC1 to BC2
Slave ports calibrating, Masters passive



Failure + 14s

Final Post Failure: BC1 to BC2
Everything settled 24 seconds after failure



Single BC Ring- Summary

- PTP Rings can be created
- ‘Steps removed’ is the primary arbitration in each node
- Timing passes in two directions from GM to furthest point
- Temporary timing loops are created
- A failure triggers a chain of events that take time to settle
- Every node that ‘switches’ may have a ‘hit’ whilst slave settles
- What to do on recovery?

SEMTECH

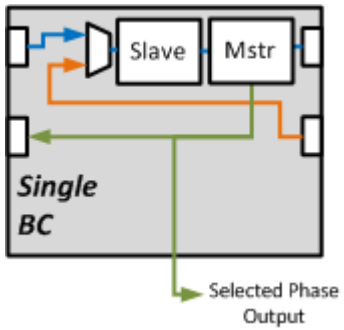
www.semtech.com

Dual BC Ring

Let's consider other Node Architectures

- Can we make fault tolerance hitless?
- Can we remove transient timing loops?
- Can we improve failover and recovery times?

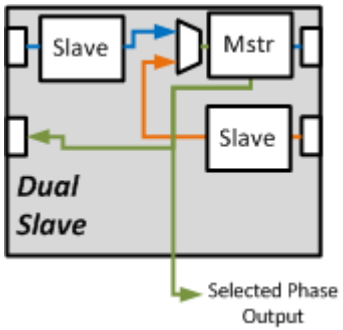
Boundary Clock Ring: Node Architectures
A 'normal' Boundary clock



Single BC:
A single slave port locks to a selected master creating a single timebase.

The Master, all egress ports and the selected output are tied to the slave timebase.

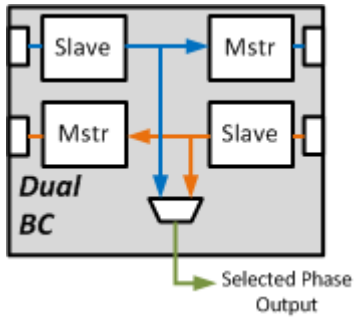
The slave port may or may not be able to hitlessly switch on failure.



Dual Slave:
Two independent slave ports create two internal timebases.

A single Master selects a slave timebase and uses it for all egress ports and the selected output.

The master can hitlessly switch between slaves on failure.



Dual BC:
Two completely independent boundary clocks.

Two slaves create a timebase for two dedicated masters.

One of the two timebases is selected as the output.

Master ports never switch on failure, merely enter holdover.

All switching is performed hitlessly in the output selection.

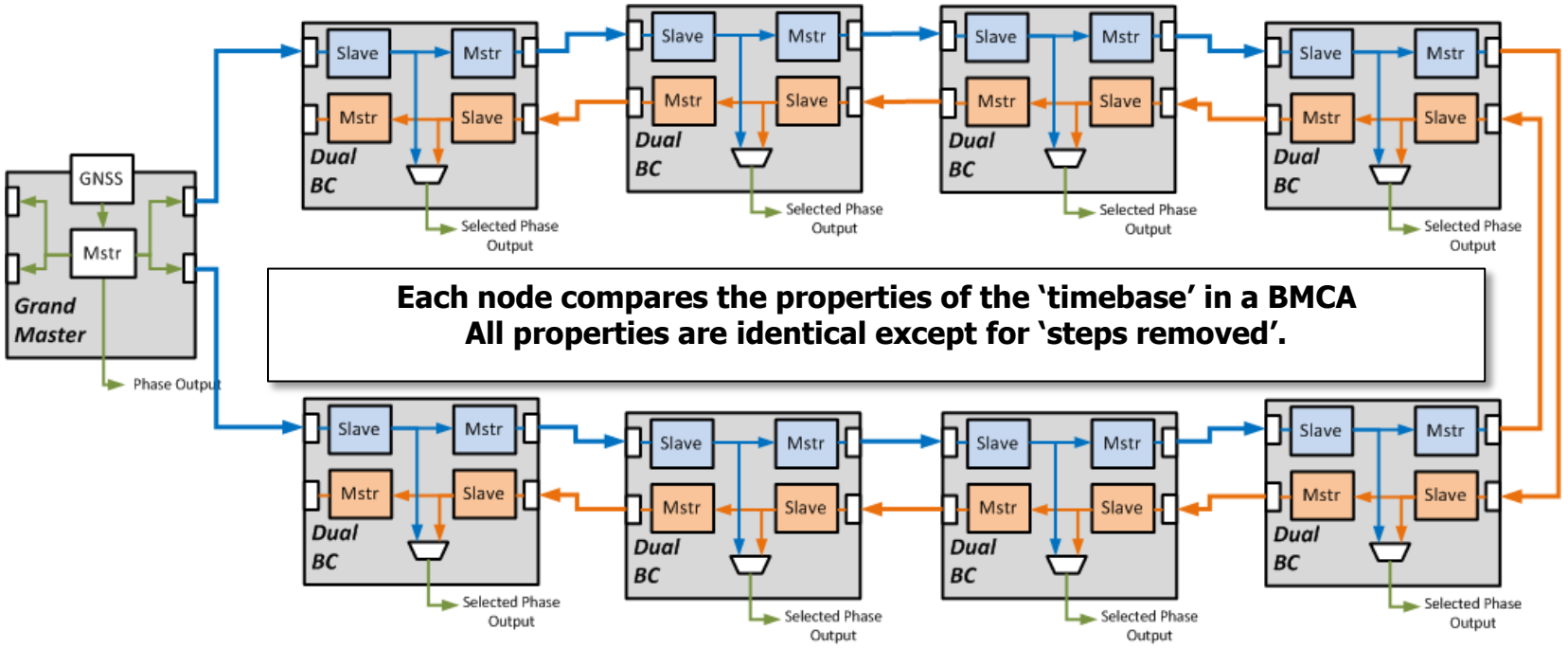
Node Architectures Summary

- **‘Single BC’**
  - Traditional in the sense that it selects a timing ingress port and locks to it
  - ‘Crosses that bridge…’
    ie. When a failure occurs, then press the mechanisms into action, not before

- **‘Dual Slave’**
  - A BC with two independent slave ports.
  - Each slave is permanently locked to the selected master
  - Switching performance is hitless, as slaves are pre-locked but fault management is the same as the ‘Single BC’

- **‘Dual BC’**
  - Two completely independent boundary clocks
  - Can maintain a timing flow in both directions
  - Permanently ready for an instant and hitless response to a failure

Dual Boundary Clock Architecture:
Conceptual- Single GM, 2 contra-rotating 'sync rings'

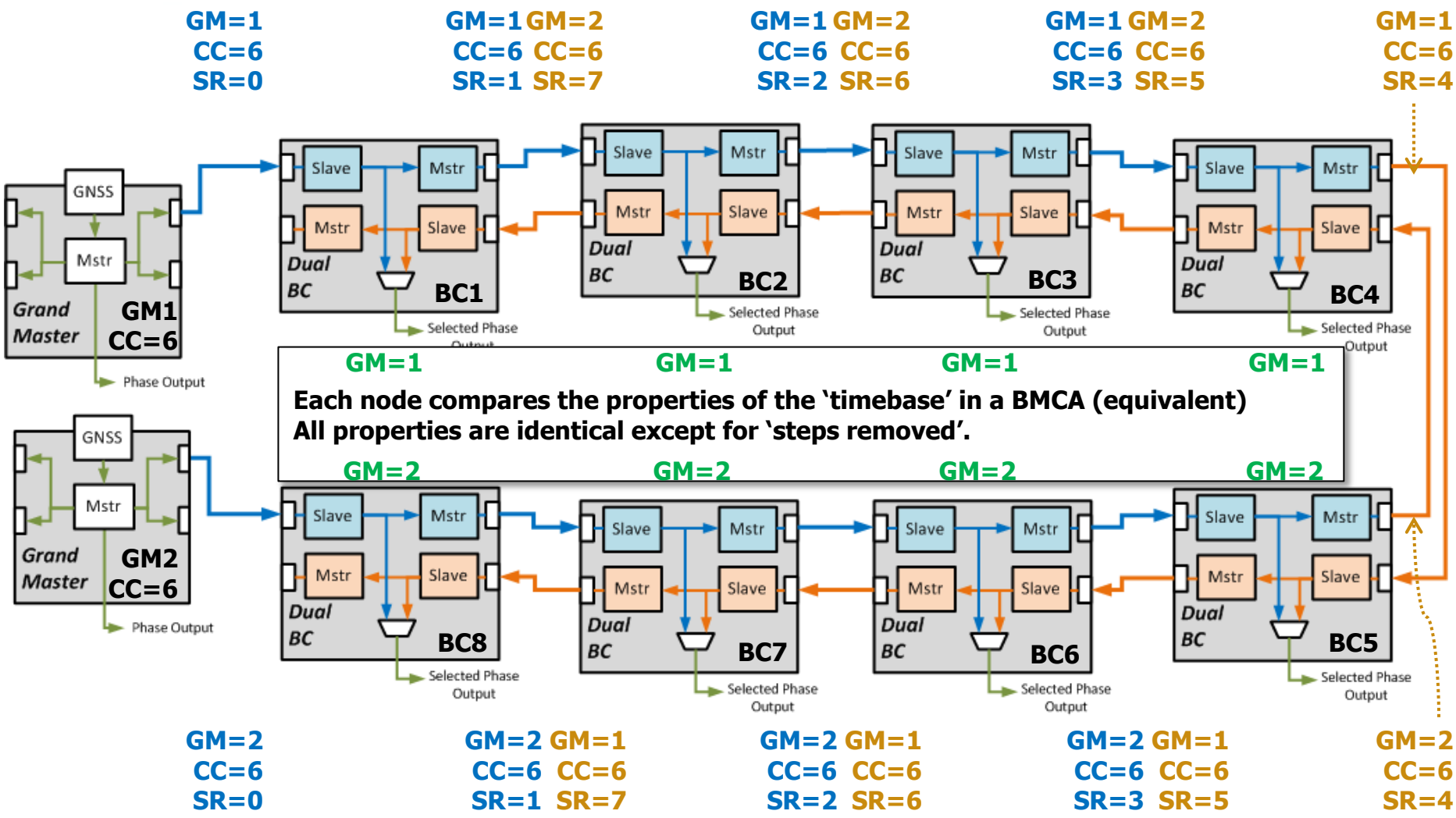


Each node has two independent boundary clocks
Each slave has only ONE acceptable master- the upstream node
Each slave is dedicated to a master
Node phase output can switch hitlessly between timebases based on BMCA

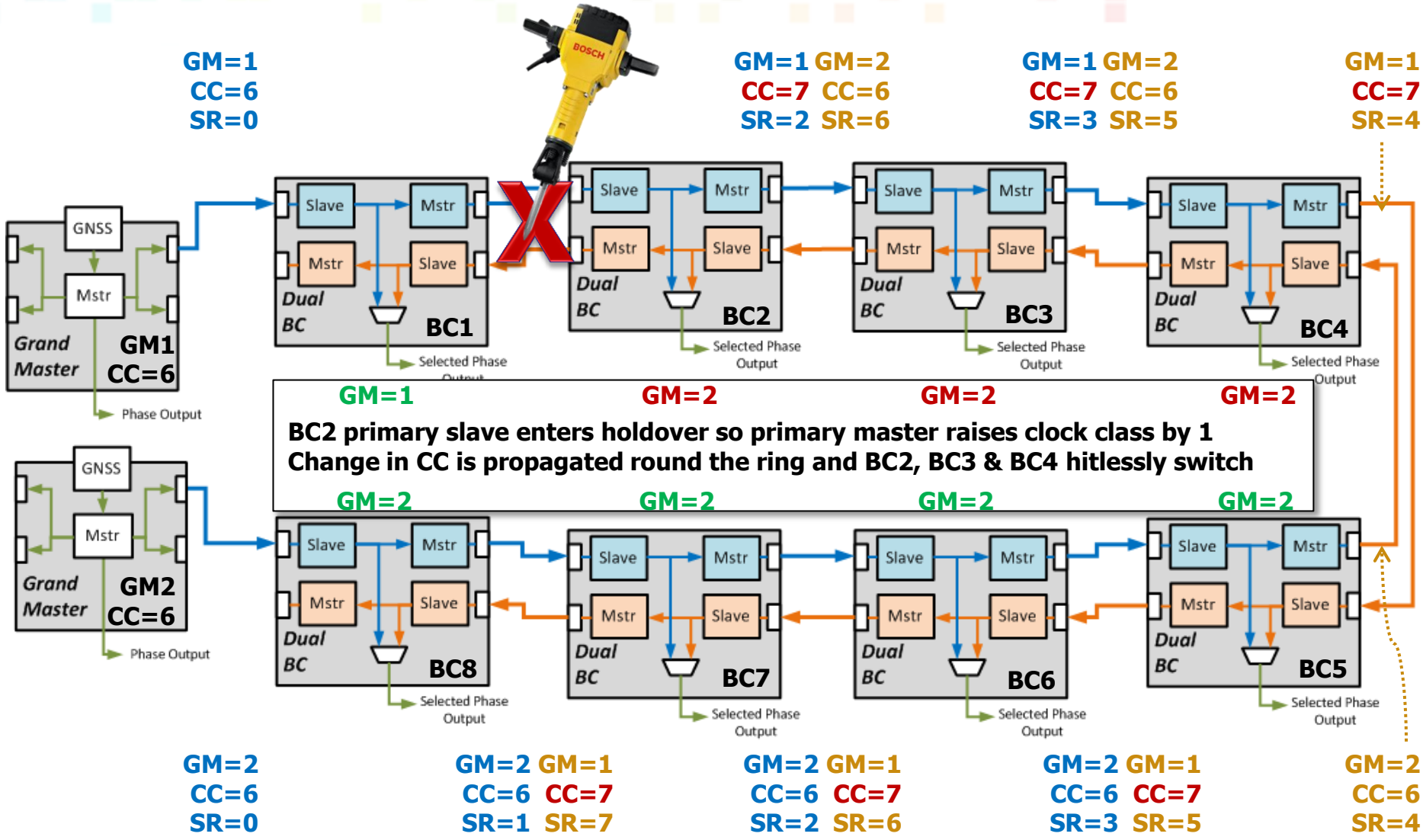
Primary Flow
Secondary Flow

Dual BC Test setup 1:
Lab setup (add a second GM for clarity)

CC=Clock Class
SR=Steps Removed



Dual BC Test setup 1: Failure between BC1 & BC2

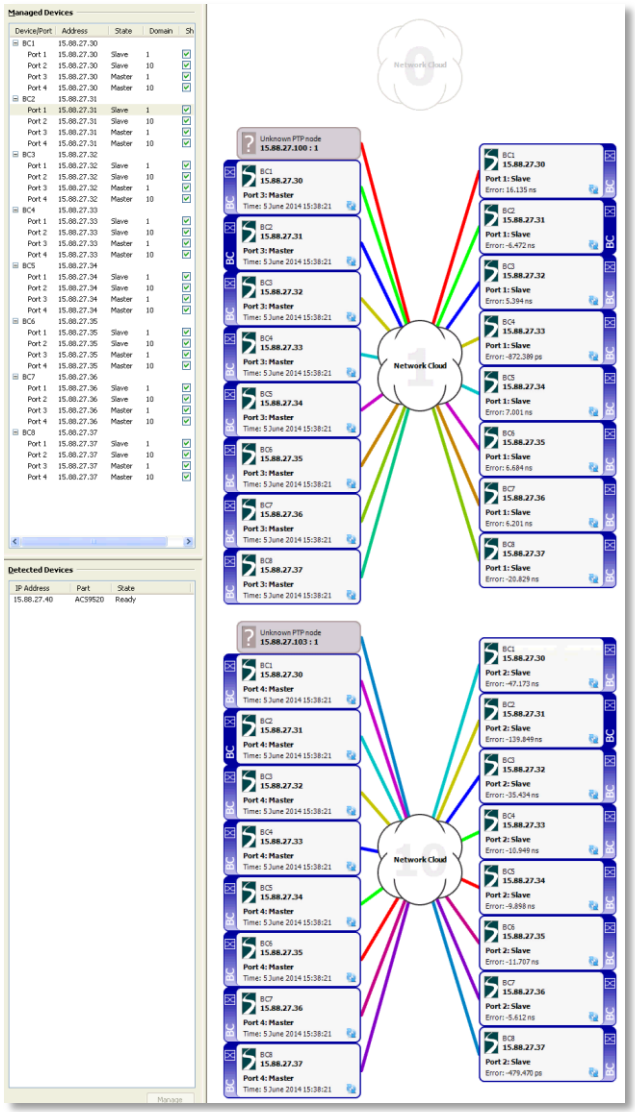


SEMTECH

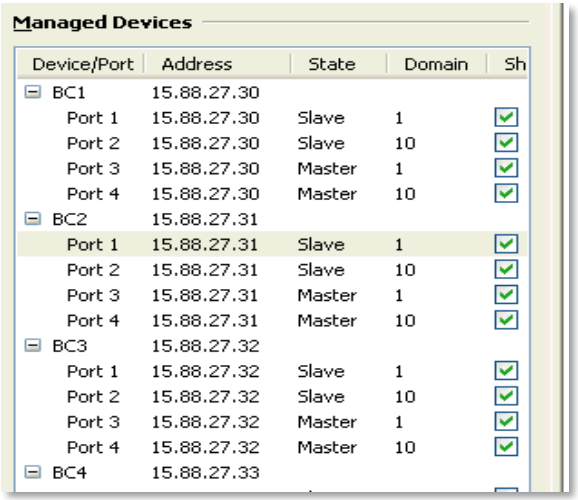
www.semtech.com

Dual BC ring Lab test

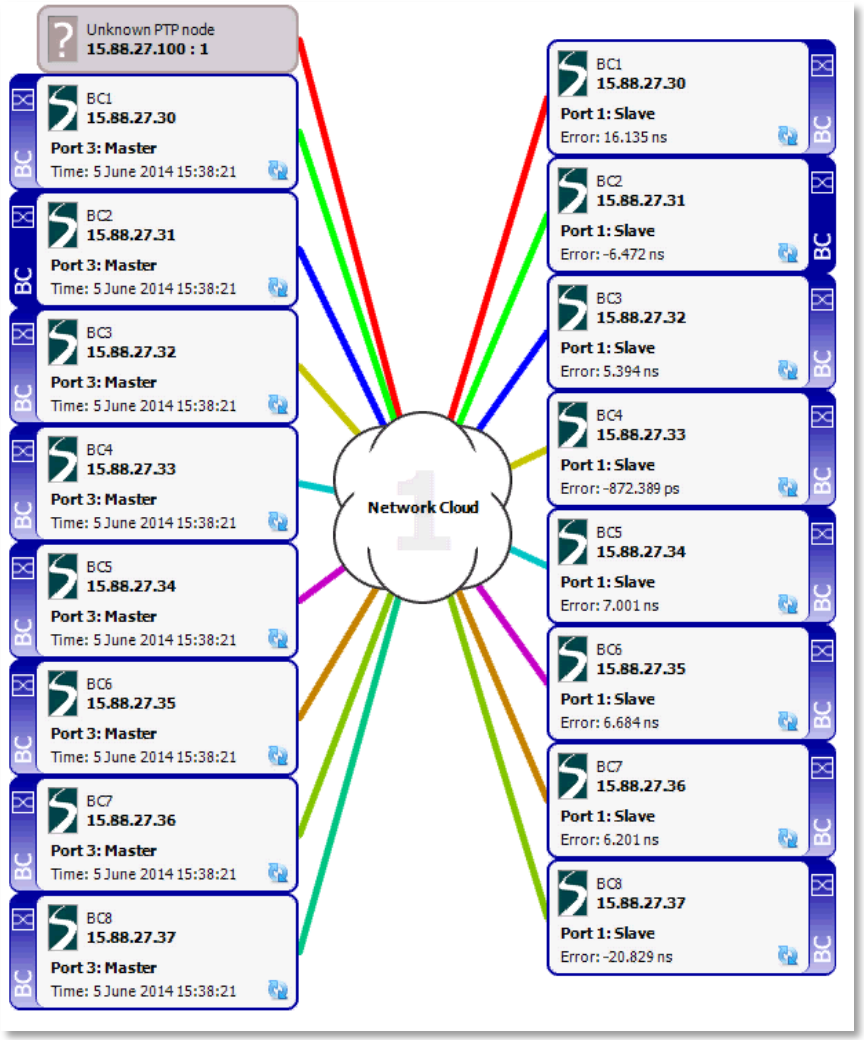
How the dual BC-Ring looks in the GUI

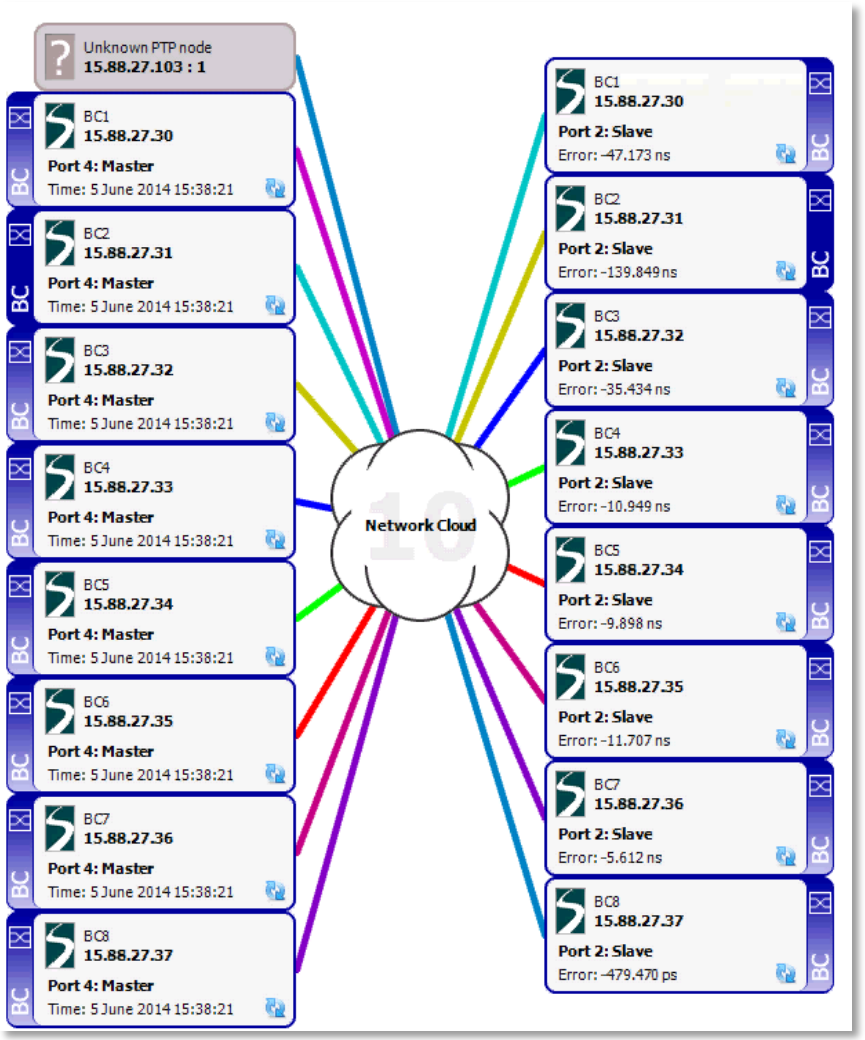


- **Each cloud represents a domain**
  - Clockwise ring – domain 1
  - Anticlockwise ring – domain 10
- **Two grandmasters**
  - One per domain
- **Four PTP ports per node**
  - Two slaves
  - Two masters

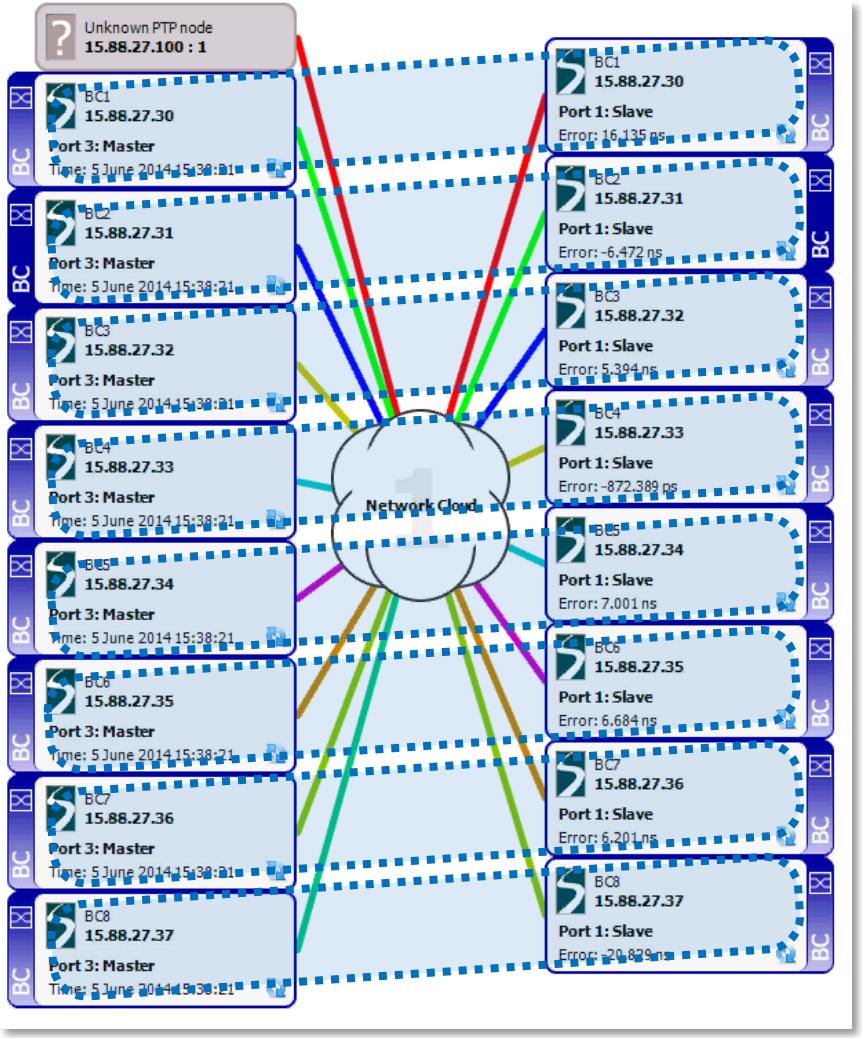


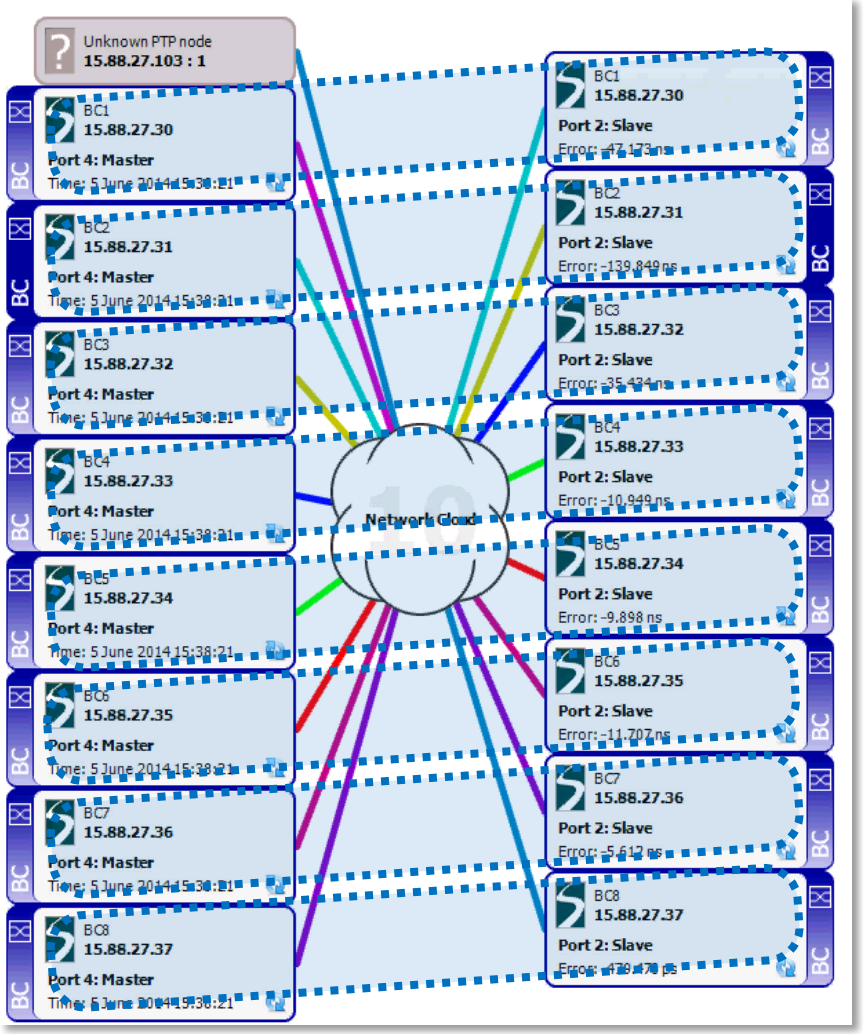
Prior to the failure:



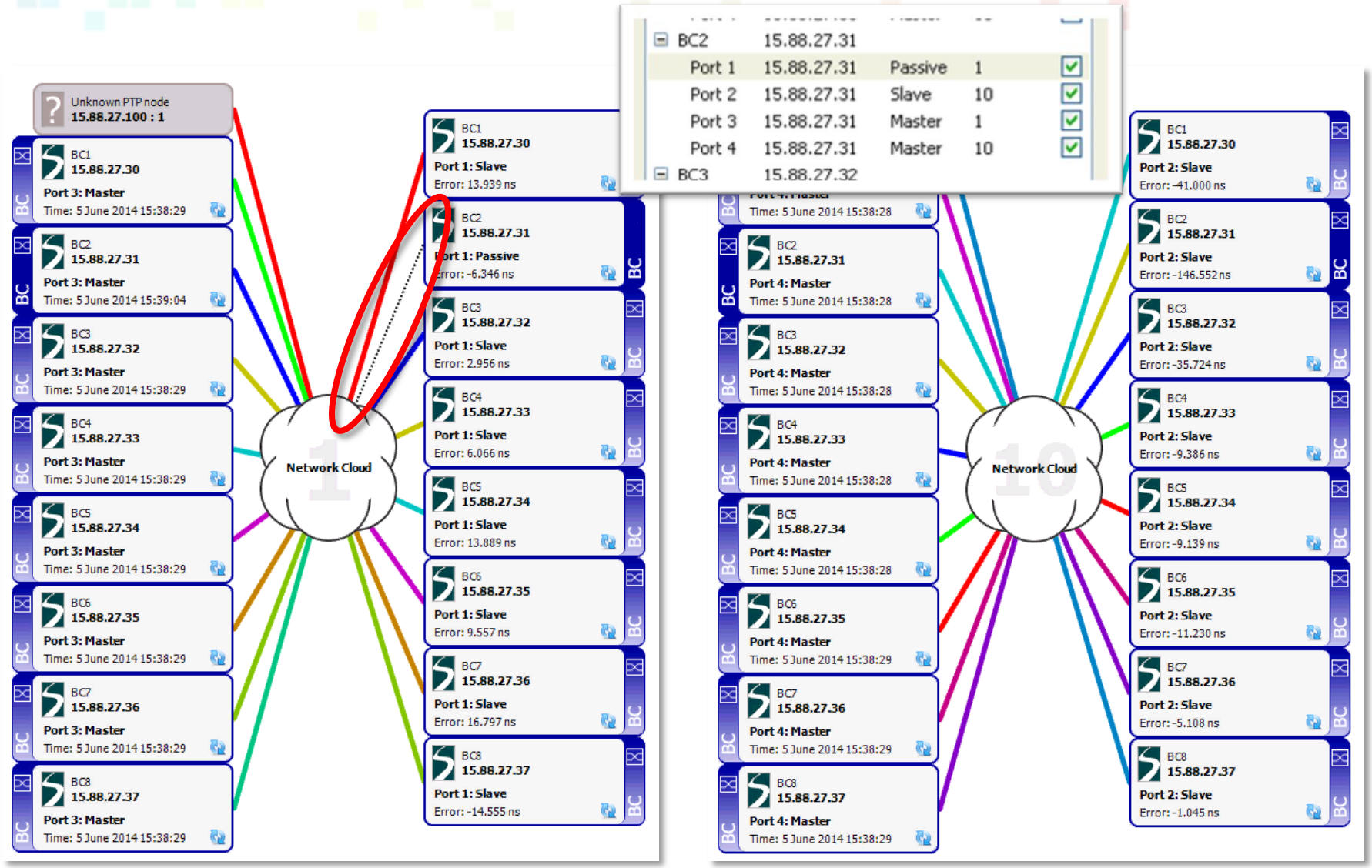


Prior to the failure:

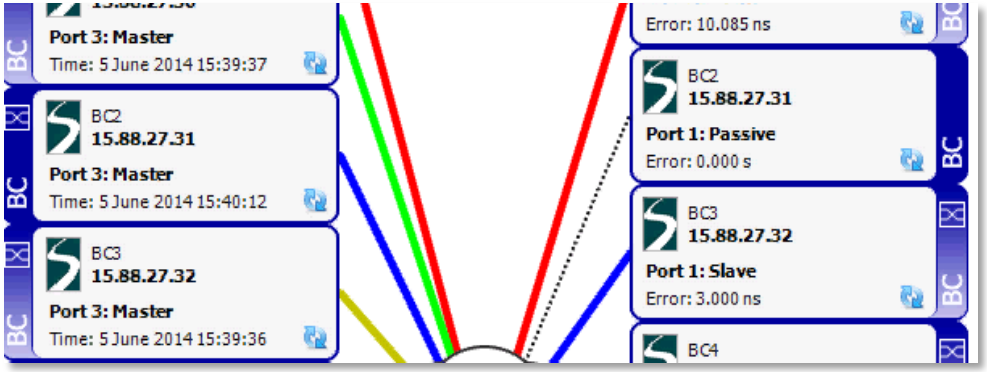




Failure +4 seconds: only change is BC2 port 1 enters holdover



No further change:
as long as the failure persists



Managed Devices

Device/Port	Address	State	Domain	Sh
BC1	15.88.27.30			
Port 1	15.88.27.30	Slave	1	☑
Port 2	15.88.27.30	Slave	10	☑
Port 3	15.88.27.30	Master	1	☑
Port 4	15.88.27.30	Master	10	☑
BC2	15.88.27.31			
Port 1	15.88.27.31	Passive	1	☑
Port 2	15.88.27.31	Slave	10	☑
Port 3	15.88.27.31	Master	1	☑
Port 4	15.88.27.31	Master	10	☑
BC3	15.88.27.32			
Port 1	15.88.27.32	Slave	1	☑
Port 2	15.88.27.32	Slave	10	☑
Port 3	15.88.27.32	Master	1	☑
Port 4	15.88.27.32	Master	10	☑
BC4	15.88.27.33			
Port 1	15.88.27.33	Slave	1	☑
Port 2	15.88.27.33	Slave	10	☑
Port 3	15.88.27.33	Master	1	☑
Port 4	15.88.27.33	Master	10	☑
BC5	15.88.27.34			
Port 1	15.88.27.34	Slave	1	☑
Port 2	15.88.27.34	Slave	10	☑
Port 3	15.88.27.34	Master	1	☑
Port 4	15.88.27.34	Master	10	☑
BC6	15.88.27.35			
Port 1	15.88.27.35	Slave	1	☑
Port 2	15.88.27.35	Slave	10	☑
Port 3	15.88.27.35	Master	1	☑
Port 4	15.88.27.35	Master	10	☑
BC7	15.88.27.36			
Port 1	15.88.27.36	Slave	1	☑
Port 2	15.88.27.36	Slave	10	☑
Port 3	15.88.27.36	Master	1	☑
Port 4	15.88.27.36	Master	10	☑
BC8	15.88.27.37			
Port 1	15.88.27.37	Slave	1	☑
Port 2	15.88.27.37	Slave	10	☑
Port 3	15.88.27.37	Master	1	☑
Port 4	15.88.27.37	Master	10	☑

Conclusion

- **Rings of boundary clocks work**

- **Simple boundary clock**
  - Single slave instance
  - Has to relock on network break
  - Will recover in time

- **Dual-slave boundary clocks**
  - Two slave instances so backup path is pre-locked
  - Can failover immediately on network break
  - Much shorter recovery time

- **Dual-path boundary clock**
  - Two boundary clocks in one
  - Dual redundant flows at all times
  - No rearrangement needed on network break

And If It Goes Wrong … ?

- We all have to work to an impossible hour to fix it

- Thank you to Rich Lansdowne and Malcolm Green for planning and running these tests