AVB - Generation 2
Latency Improvement Options

v1 – March, 2011

802.1 AVB Face to Face – Singapore

Don Pannell - dpannell@marvell.com

Need/Desire/Goal

- **Lowest Latency Physically Possible – Any Way Possible**
  - Want <5 uSec or <15 uSec/hop with <300 byte frames
  - Many, ~32 hop, Daisy Chains
  - Small Bursts of frames at known regular intervals (e.g., a 40 uSec long burst of data every 125 uSec)
- **Willing to Engineer Network Segments to meet this goal**
  - That's part of the 'Any Way Possible' statement
  - Non-Engineered (i.e., Consumer) Networks will not be able to depend on this very low latency as it can't be guaranteed in their Networks
- **The Network Structure and Usage will have to be Engineered, Managed and Controlled**
  - All links will need to be Gigabit Ethernet or Faster

Physical Realities - GE

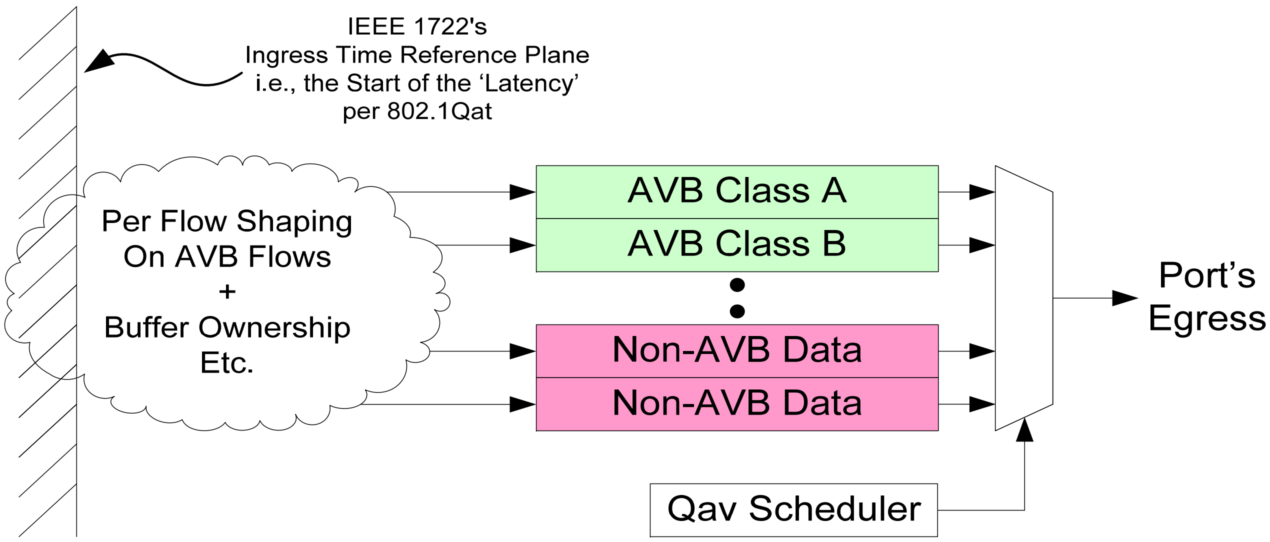
- All Latencies are Measured as Last bit in to Last bit out
- At any point on the wire, a Min size GE frame's Transmission Time is 0.512 uSec (for 64 bytes)
- Its 2.560 uSec for a 300+20 Byte frame (the 20 bytes is for the Preamble & IFG)
- CAT 5e's Propagation Delay Spec is ~538 ns/100 m

Talker: How Fast Can We Go?

Fastest? Talker's Assumption & Model

Assumptions:

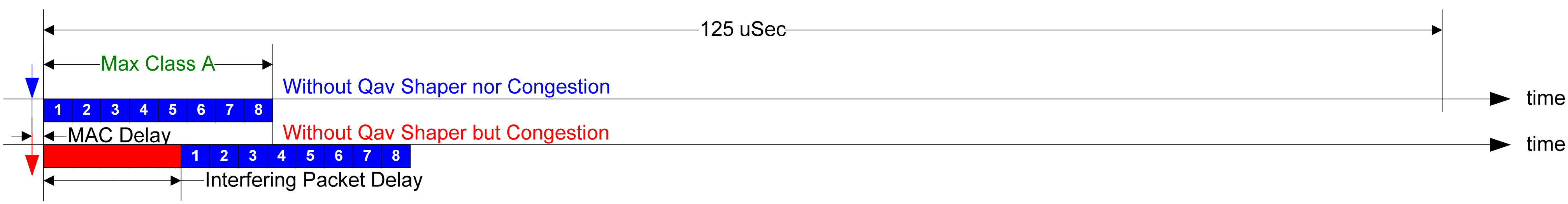
- Qav Shaper is disabled so bursting can occur – the Strict Priority Scheduler is used instead
- No other modifications are done
- At t_0 all Class A flows need to be transmitted
- 1 clk before the first Class A frame is allowed to Tx a Max size Non-AVB frame starts out the port
- This Max size frame goes 1^{st} then the Class A frames can go



Example of GE Talker Lowest? Latency

The below example is GE (300 Byte frames):

- Top line is eight 300 bytes AVB Flows with no Qav and no Congestion
  - Each blue box is 320 bytes for the 20 byte Preamble+IFG time
- Next line shows a Max size (1522 bytes) Interfering Frame
  - The red box is 1542 byes in size for the 20 byte Preamble+IFG
- An Engineered Talker can know the total number of Class A frames it will ever be bursting during any given interval
  - So it knows the Max Class A which is eight 320 byte frames in this example
  - The smaller this number is the lower the worst case latency for this talker is
- Late Interfering frames are not a issue due to disabling the Qav shaper
  - The frames are not spaced out either which is very bad for an arbitrary network so the impact of disabling the Qav shaper on the Bridges needs to be examined



Gen 2 Talker Class A Equation is:

$$\text{Gen2 T1: Max Latency}_{\text{Class A Talker}} = t_{\text{MAC Delay}} + t_{\text{Tx Max Frame}} + t_{\text{Class A Data}} + t_{\text{Tx Max Cable}}$$

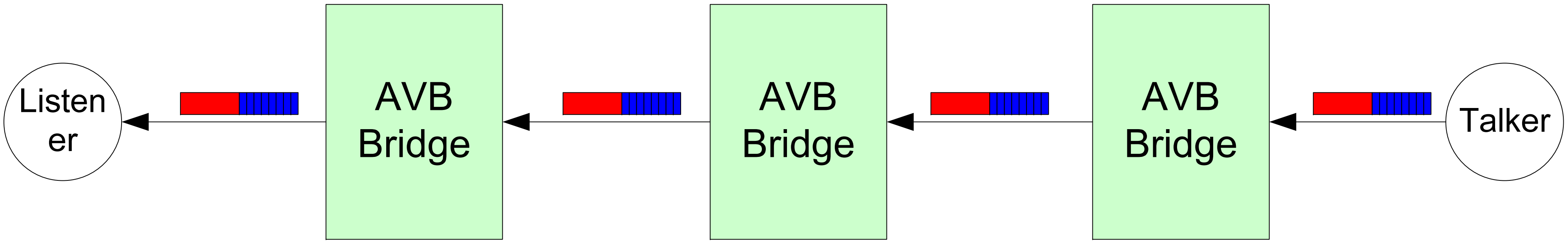
- **Total time is:**
  - Internal delay of the MAC
  - Plus the time to transmit a possible max size interfering frame
  - Plus the time to transmit the data
  - Plus the time to get the bits down a 100 meter cable
- **For Eight 300+20 Byte Class A Frames on GE:**
  - Assuming MAC Delay = 1 slot time & Max Non-AVB = 1522 bytes
  - Max Latency = 0.512 uSec + 12.336 uSec + 20.480 uSec + 0.538 uSec = **33.866 uSec**
  - This is the total time it takes to get the last bit of the last frame to the next device

Bridge: How Fast Can We Go?

Best Case Network Assumption & Model

Need to Define the Simplest Network

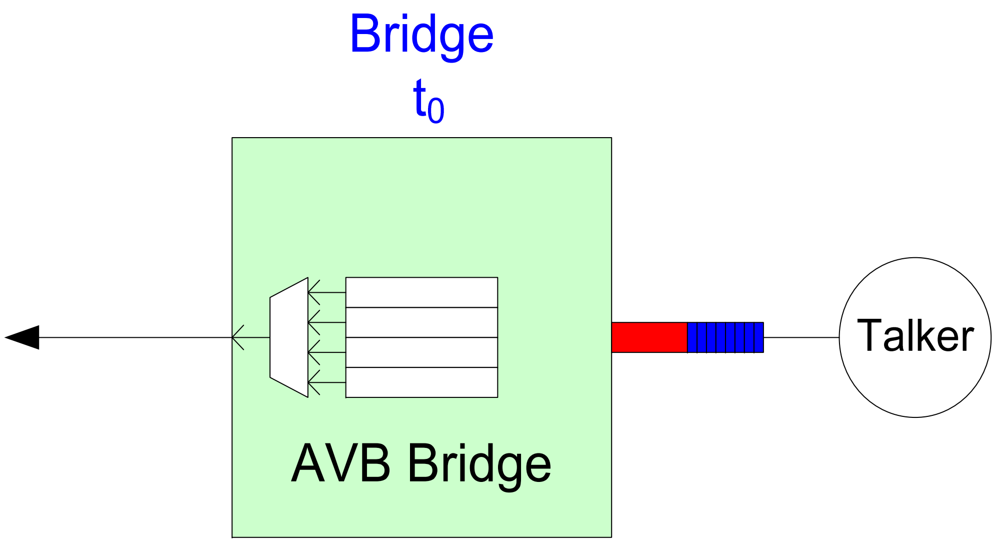
- N number of 2 Port Bridges (a 3rd port connects to internal CPU)
- How long will it take to get the previous Talker's data to the Listener
- Lets look at what goes on in the 1st AVB Bridge



Inside the 1st AVB Bridge

Time Progression – Fig 1

- At Bridge t_0 the 1st bit of the interfering frame enters the bridge

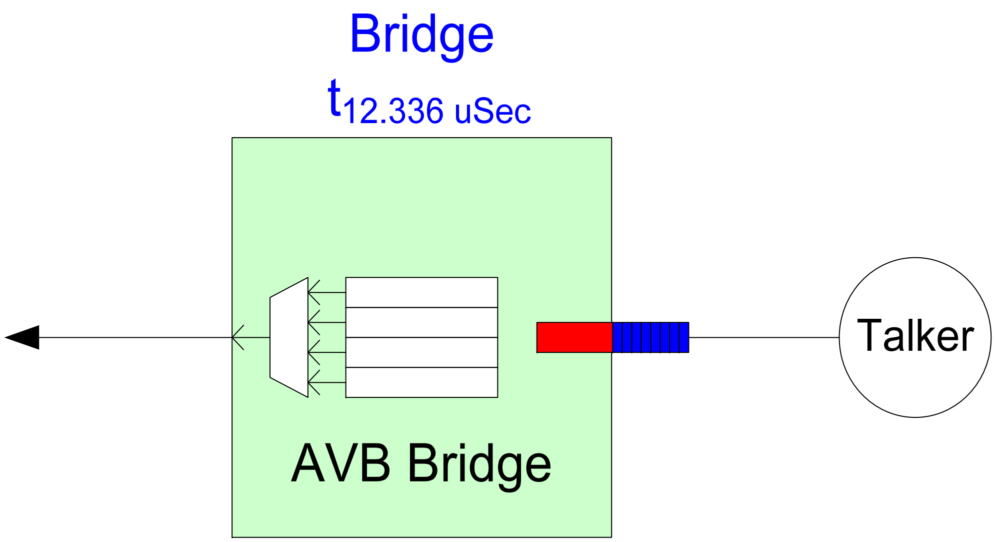


Inside the 1st AVB Bridge

Time Progression – Fig 2

- At Bridge t_0 the 1st bit of the interfering frame enters the bridge
- 1542 Bytes later the 1st bit of the 1st Blue frame enters the bridge

- **1522+20*8=**

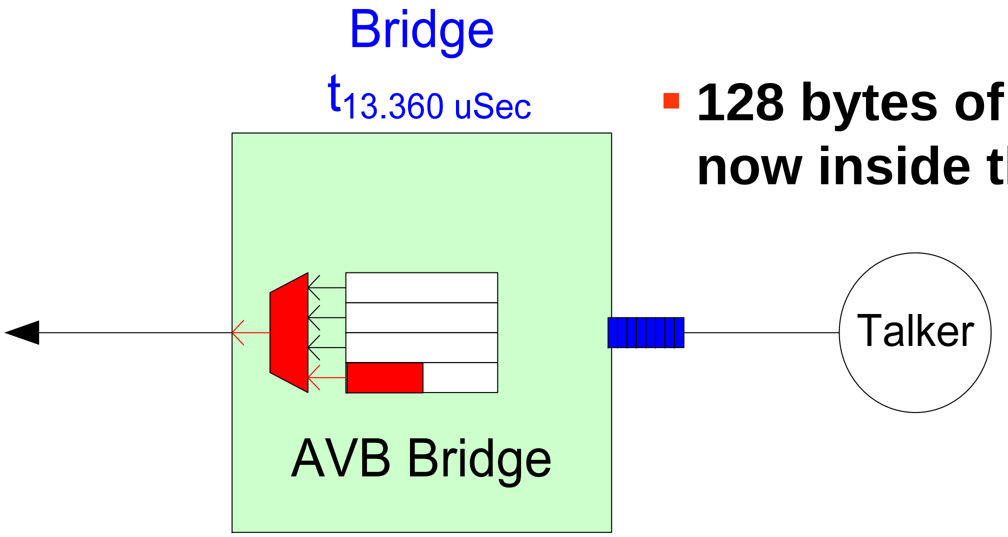


Inside the 1st AVB Bridge

Time Progression – Fig 3

- At Bridge t_0 the 1st bit of the interfering frame enters the bridge
- 1542 Bytes later the 1st bit of the 1st Blue frame enters the bridge
- Two Slot Times later the 1st bit of the interfering frame leaves the bridge

- **12.336+1.024=**



- **128 bytes of 1st Blue frame now inside the bridge**

Inside the 1st AVB Bridge

Time Progression – Fig 4

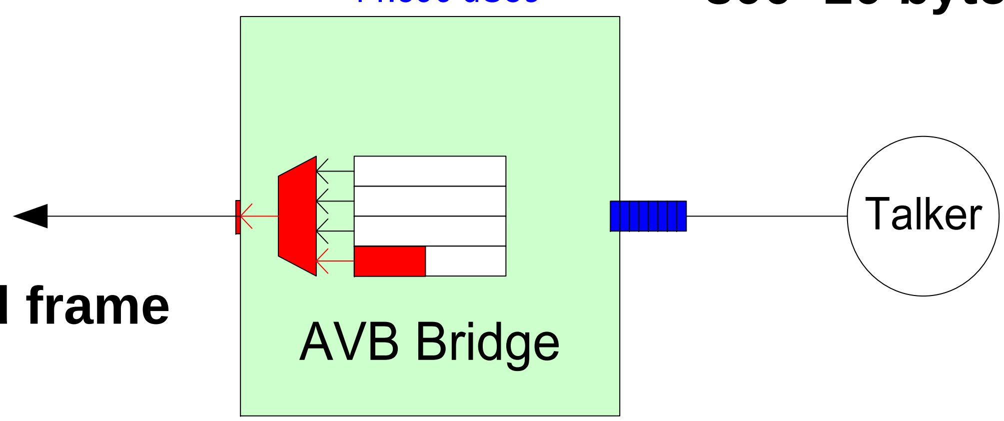
- At Bridge t_0 the 1st bit of the interfering frame enters the bridge
- 1542 Bytes later the 1st bit of the 1st Blue frame enters the bridge
- Two Slot Times later the 1st bit of the interfering frame leaves the bridge
- The 1st bit of the 2nd Blue frame enters the Bridge

- **12.336+2.560=**

Bridge

$t_{14.896\ uSec}$

- **300+20 bytes = 2.560**



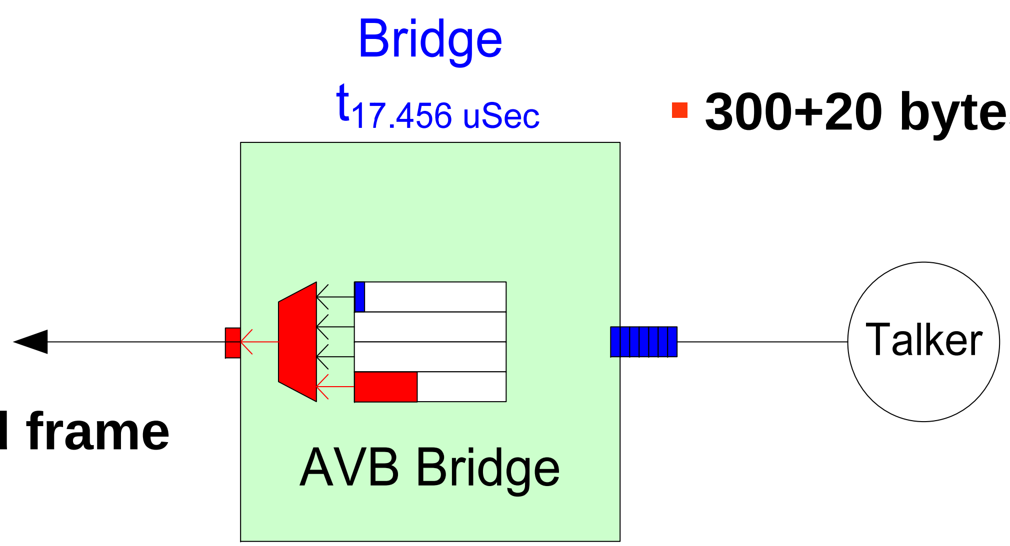
- **192 bytes of Red frame Transmitted**

Inside the 1st AVB Bridge

Time Progression – Fig 5

- At Bridge t_0 the 1st bit of the interfering frame enters the bridge
- 1542 Bytes later the 1st bit of the 1st Blue frame enters the bridge
- Two Slot Times later the 1st bit of the interfering frame leaves the bridge
- The 1st bit of the 2nd Blue frame enters the Bridge
- Then the 3rd Blue frame

- **12.336+5.120=**
- **300+20 bytes = 2.560**



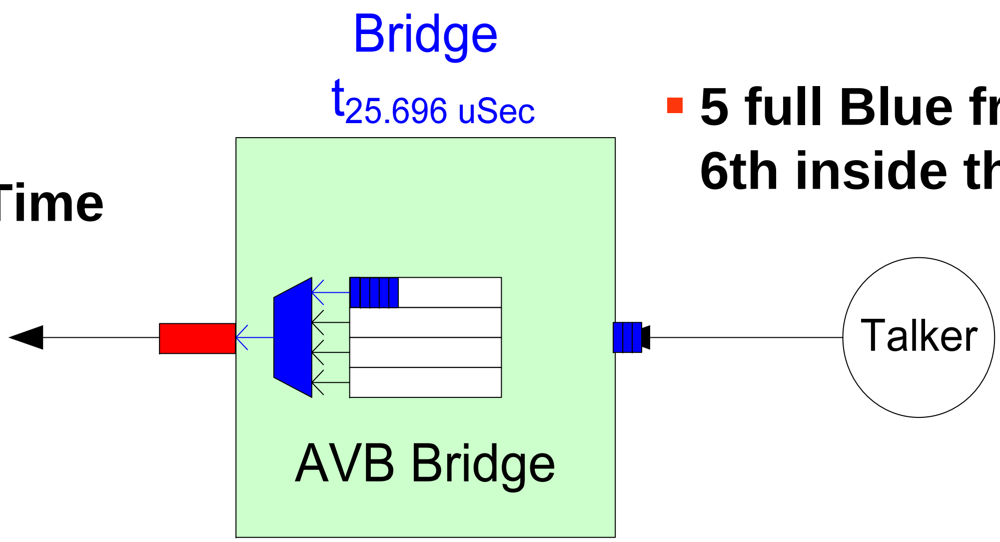
- **512 bytes of Red frame Transmitted**

Inside the 1st AVB Bridge

Time Progression – Fig 6

- 1542 Bytes later the 1st bit of the 1st Blue frame enters the bridge
- Two Slot Times later the 1st bit of the interfering frame leaves the bridge
- The 1st bit of the 2nd Blue frame enters the Bridge
- Then the 3rd Blue frame
- The 1st bit if the 1st Blue frame leaves the bridge

- **13.360+12.336=**
- **Start Time + Tx Time**



- **5 full Blue frames + 70 of 6th inside the bridge**

Inside the 1st AVB Bridge

Time Progression – Fig 7

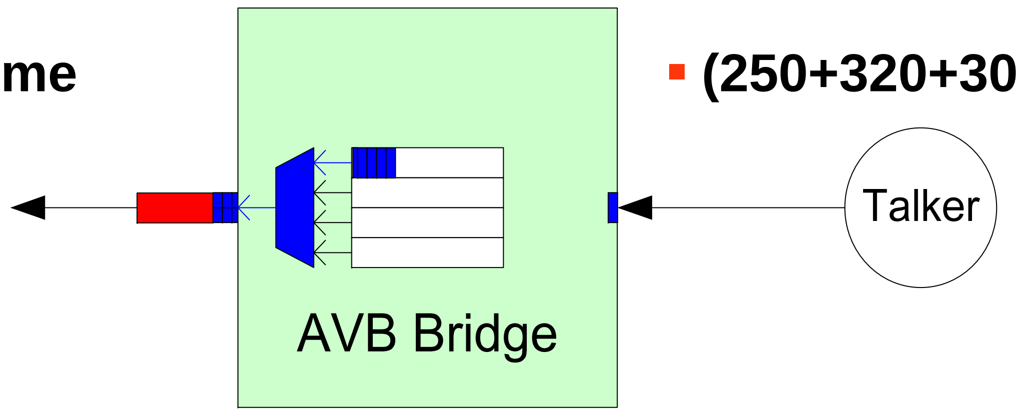
- Two Slot Times later the 1st bit of the interfering frame leaves the bridge
- The 1st bit of the 2nd Blue frame enters the Bridge
- Then the 3rd Blue frame
- The 1st bit if the 1st Blue frame leaves the bridge
- The last bit of the last Blue frame enters the bridge

Bridge

$t_{32.656\ uSec}$

- **25.696+6.960=**
- **Previous + Rx Time**

- **Rest of 6th + 7th & 8th frames**
- **(250+320+300)*8 = 6.960**

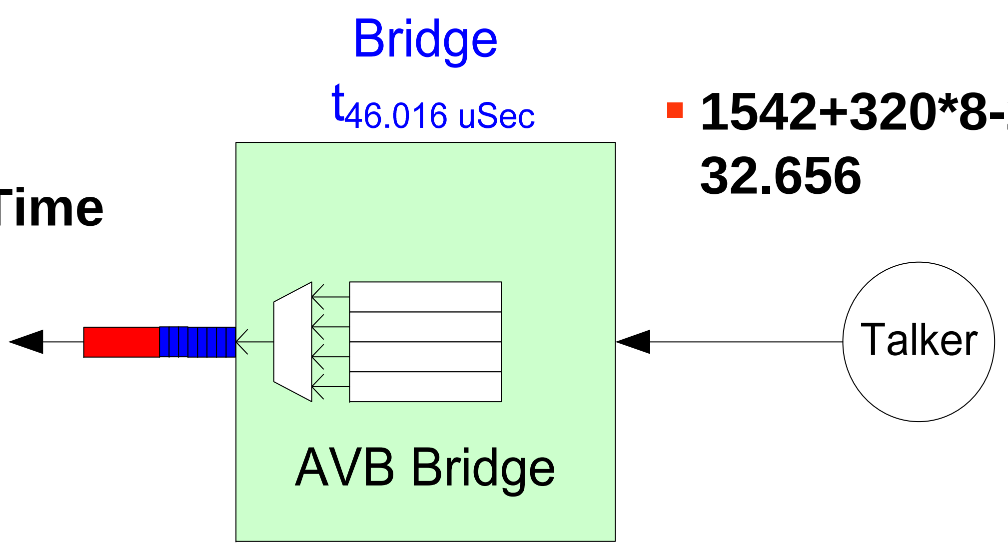


Inside the 1st AVB Bridge

Time Progression – Fig 8

- Two Slot Times later the 1st bit of the interfering frame leaves the bridge
- The 1st Blue frame enters the Bridge – then the 2nd Blue frame
- The 1st bit if the 1st Blue frame leaves the bridge
- The last bit of the last Blue frame enters the bridge
- The last bit of the last Blue frame leaves the bridge

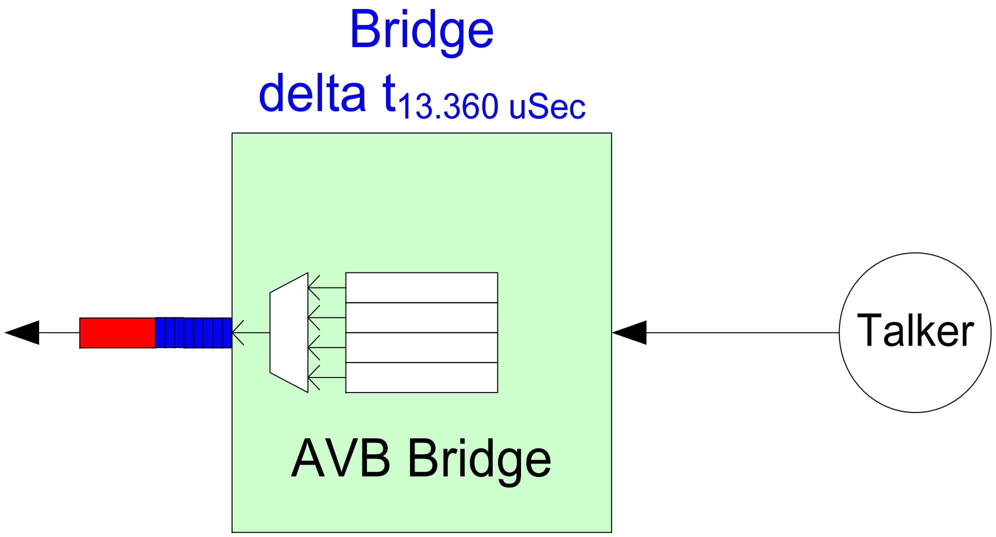
- **13.360+32.656=**
- **Start Time + Tx Time**



- **1542+320*8-20=4082*8 = 32.656**

Inside the 1st AVB Bridge

- **The 46.016 uSec number is First Bit In to Last Bit Out**
- **It needs to be Last Bit In to Last Bit Out!**
  - Subtract 12.336 uSec for the Tx time of the Red Frame
  - Subtract 20.480 uSec for the Tx time of the eight Blue frames
  - 46.016 uSec – 12.336 uSec – 20.480 uSec + 20 bytes = 13.360 uSec
  - Or its $t_{\text{Fig 8}}$ (last Blue bit in) – $t_{\text{Fig 7}}$ (last Blue bit out) = 13.360 uSec



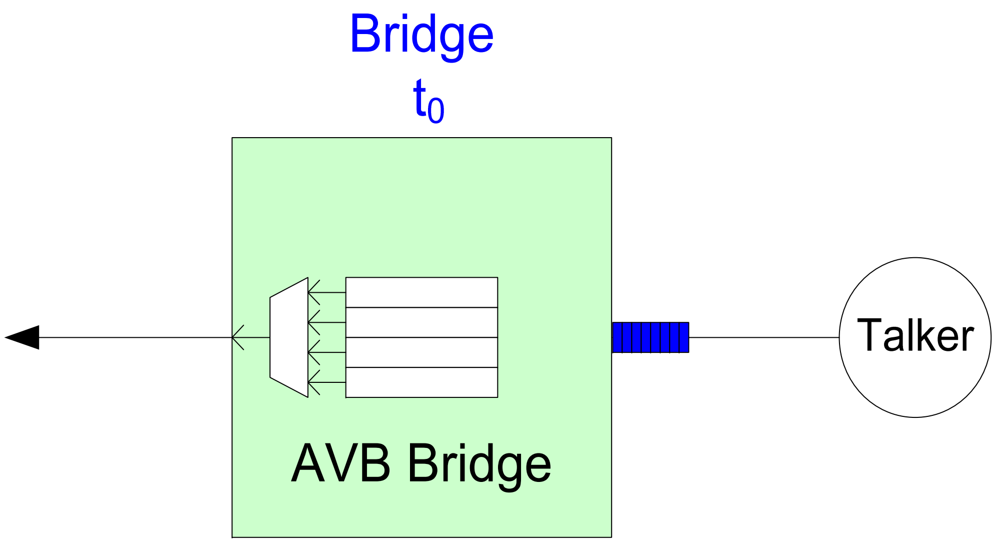
Gen 2 Bridge Type 1 Class A Equation is:

Gen2 B Ty1: Max Latency$_{\text{Class A Bridge}}$ = $t_{\text{Bridge Delay}} + t_{\text{Max Frame}} + t_{\text{Tx Max Cable}}$

- **Total time is:**
  - Internal delay of the Bridge
  - Plus the time to receive the max size interfering frame
  - Plus the time to get the bits down a 100 meter cable
- **To Daisy Chain the Eight 300+20 Byte Class A Frames on GE:**
  - Assuming Bridge Delay = 2 slot times & Max Non-AVB = 1522 bytes
  - Max Latency = 1.024 uSec + 12.336 uSec + 0.538 uSec = **13.898 uSec**
  - This is the additional time it takes to get the last bit of the last frame to the next device
  - This matches the previous number of 13.360 uSec which didn't include the cable time
- **This number does not change if the Interfering frame comes from another port**
- **How does this improve if the Interfering Frame is removed?**

Inside the 1st AVB Bridge

- **Without Interfering Frame - Time Progression – Fig 1**
  - At Bridge t_0 the 1st bit of the 1st Blue frame enters the bridge

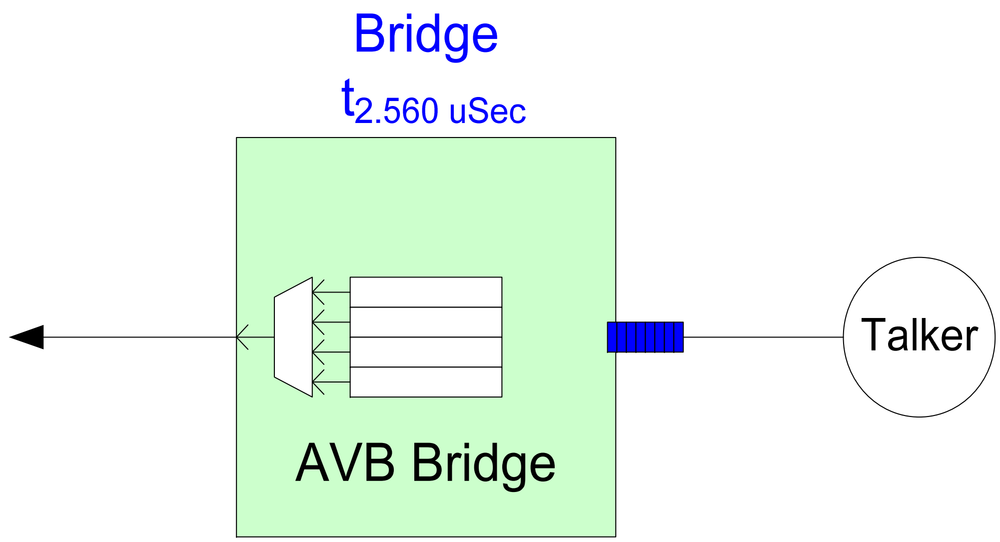


Inside the 1st AVB Bridge

Without Interfering Frame - Time Progression – Fig 2

- At Bridge t_0 the 1st bit of the interfering frame enters the bridge
- 320 Bytes later the 1st bit of the 2nd Blue frame enters the bridge

- **300+20*8=**

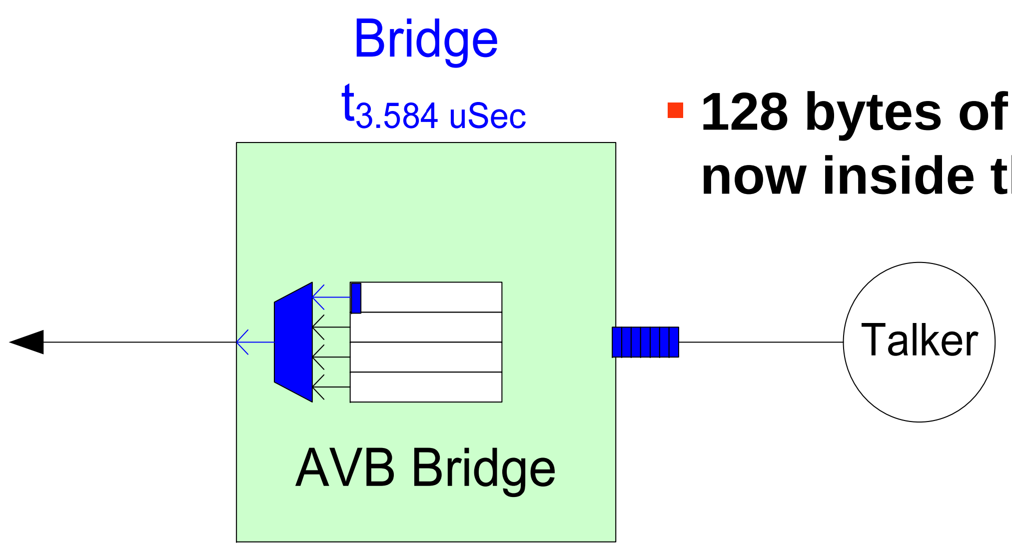


Inside the 1st AVB Bridge

Without Interfering Frame - Time Progression – Fig 3

- At Bridge t_0 the 1st bit of the interfering frame enters the bridge
- 320 Bytes later the 1st bit of the 2ndt Blue frame enters the bridge
- Two Slot Times later the 1st bit of the 1st Blue frame leaves the bridge

- **2.560+1.024=**

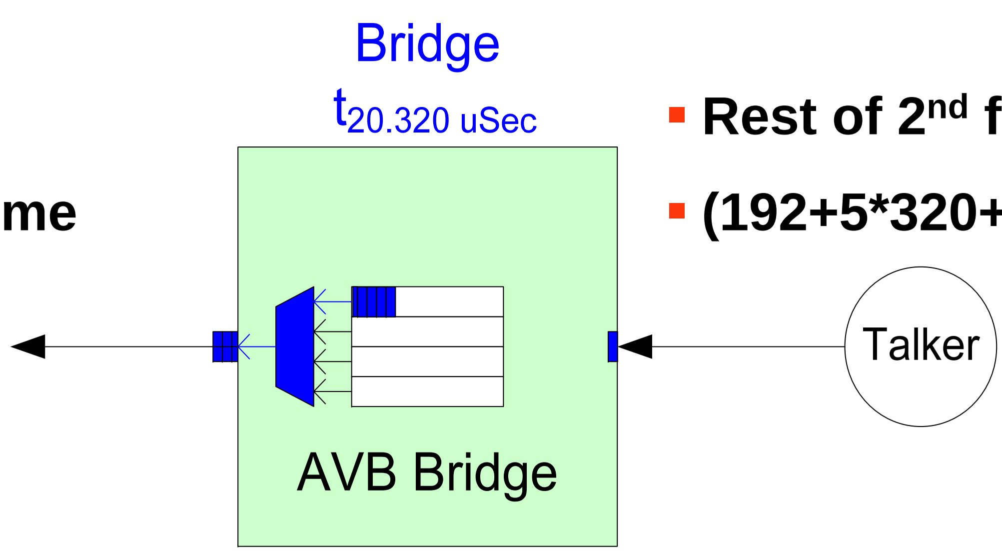


- **128 bytes of 2nd Blue frame now inside the bridge**

Inside the 1st AVB Bridge

- **Without Interfering Frame - Time Progression – Fig 4**
  - At Bridge t_0 the 1st bit of the interfering frame enters the bridge
  - 320 Bytes later the 1st bit of the 2ndt Blue frame enters the bridge
  - Two Slot Times later the 1st bit of the 1st Blue frame leaves the bridge
  - The last bit of the last Blue frame enters the bridge

- **3.584+16.736=**
- **Previous + Rx Time**



- **Rest of 2nd frame + 3rd to 8th**
- **(192+5*320+300)*8 = 16.736**

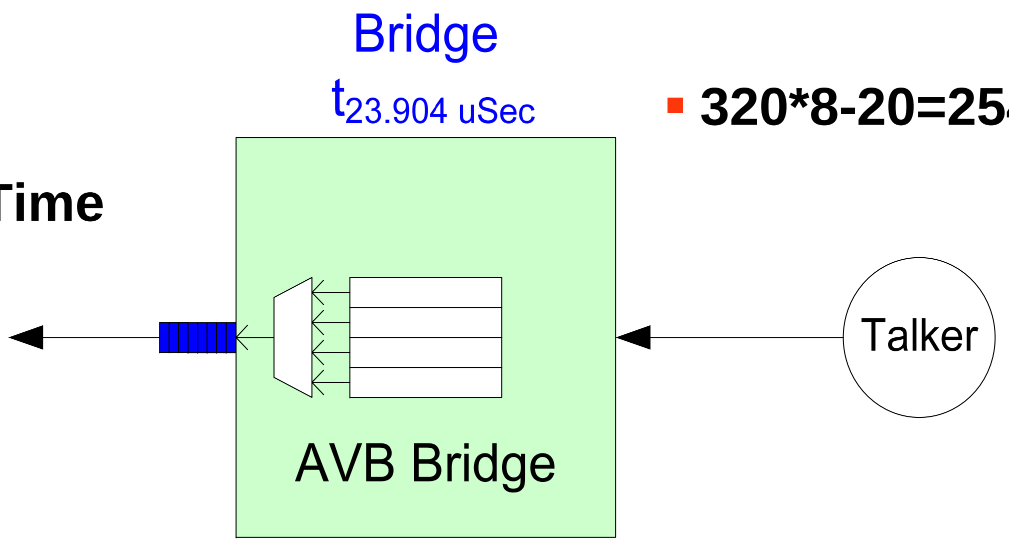
Inside the 1st AVB Bridge

Without Interfering Frame - Time Progression – Fig 5

- At Bridge t_0 the 1st bit of the interfering frame enters the bridge
- 320 Bytes later the 1st bit of the 2ndt Blue frame enters the bridge
- Two Slot Times later the 1st bit of the 1st Blue frame leaves the bridge
- The last bit of the last Blue frame enters the bridge
- The last bit of the last Blue frame leaves the bridge

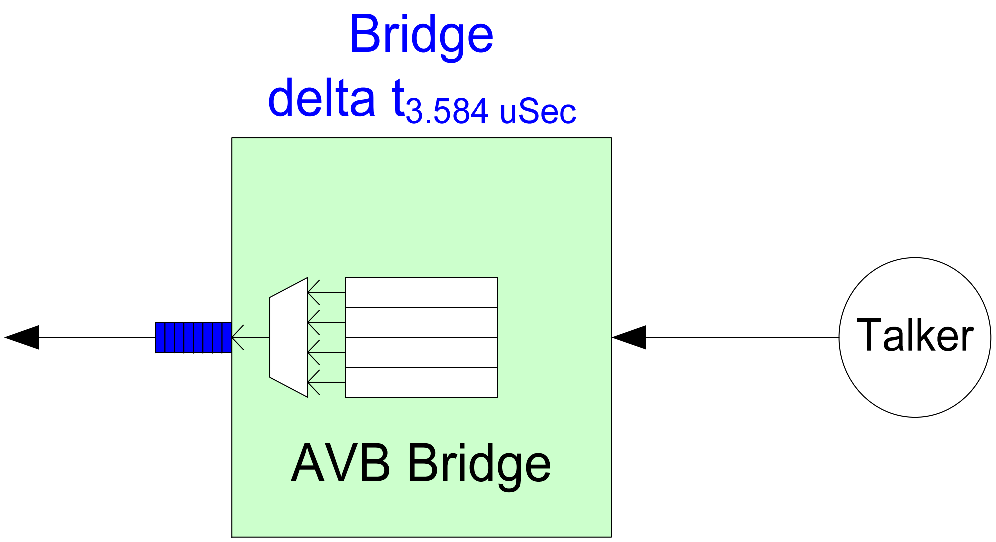
- **3.584+20.320=**
- **Start Time + Tx Time**

- **320*8-20=2540*8 = 20.320**



Inside the 1st AVB Bridge – No Interference

- **The 23.904 uSec number is First Bit In to Last Bit Out**
- **It needs to be Last Bit In to Last Bit Out!**
  - Subtract 20.480 uSec for the Tx time of the eight Blue frames
  - 23.904 uSec – 20.480 uSec + 20 bytes = 3.584 uSec
  - Or its $t_{\text{Fig 5}}$ (last Blue bit in) – $t_{\text{Fig 4}}$ (last Blue bit out) = 3.584 uSec



Gen 2 Bridge Type 2 Class A Equation is:

Gen2 B Ty2: $\text{Max Latency}_{\text{Class A Bridge}} = t_{\text{Bridge Delay}} + t_{\text{Max Stream Frame}} + t_{\text{Tx Max Cable}}$

- **Total time is:**
  - Internal delay of the Bridge
  - Plus the time to receive the largest frame size of the stream
  - Plus the time to get the bits down a 100 meter cable
- **To Daisy Chain the Eight 300+20 Byte Class A Frames on GE:**
  - Assuming Bridge Delay = 2 slots times typical & Max Steam Frame = 300 bytes
  - Max Latency = 1.024 uSec + 2.560 uSec + 0.538 uSec = **4.122 uSec**
  - This is the additional time it takes to get the last bit of the last frame to the next device
  - This matches the previous number of 3.584 uSec which didn't include the cable time

Interfering Frames are the Problem

- **The Bridge Latency with Interfering Frame is:**
  - Equal to the Size of Interfering frame + Bridge Delay + Cable Delay
  - With Max Size interfering frame this is 13.898 uSec
- **The Bridge Latency without an Interfering Frame is:**
  - Equal to the Size of AVB frame + Bridge Delay + Cable Delay
  - With a 300 byte AVB frame this is 4.122 uSec
- **Can we get rid of the Interfering Frames to get the better latency?**
- **Some Proposals are (see presentations from Nov 2010 Plenary):**
  - Interrupt the Non-AVB Interfering Frame – Preempt it & Reassemble it
  - Interrupt the Non-AVB Interfering Frame – Cause a CRC Error & Re-transmit it
  - New Ideas?
- **Lets look at the Pro's and Con's for Each of These**

Interrupt via Preemption & Reassembly

- **Pro's:**
  - Reduces the Interfering frame latency
  - Will get a the Fragmented frame out eventually
- **Con's:**
  - Non Standard Transmitter
    - Once preempted, the Transmitter has to remember where it left of
    - Some new mechanism is needed to signal from Transmitter to Receiver that this occurred
    - New Symbols are needed at the Interrupt and at the Restart of the Non-AVB frame
    - Multiple Interrupts could occur on a single Non-AVB frame
    - Requires much more complex transmit buffer controller
    - Latency of a Min Size (or largest BDPU) frame per hop is required for worst case
  - Non Standard Receiver
    - Once preempted, the Receiver has to remember where the interrupt occurred in the Non-AVB frame – it has to hold it in memory as well, as its not fully received yet
    - Need to detect the end of the fragmented frame
    - Need to detect the restart of the fragmented frame
    - Requires much more complex receive buffer controller

Interrupt via CRC Error & Retransmit

- **Pro's:**
  - Removes the Interfering frame latency
  - Standard way to indicate the Interrupt of the Non-AVB frame (it's a CRC error)
  - Uses Standard Receiver
- **Con's:**
  - Non Standard Transmitter
    - Requires a new mode for Full Duplex Transmitters
    - Requires a more complex transmit buffer controller
    - May need to re-transmit 99% of a Non-AVB frame – not bandwidth efficient
    - May never be able to get the Non-AVB frame out due to constant interruptions
  - Standard Receiver, but
    - MIB Counters indicating line quality (CRC counter) becomes useless
    - Don't know the side effects of this – Spanning Tree timeouts? Some other protocol broken?

A New Idea – TABS
Time Aware Blocking Shaper

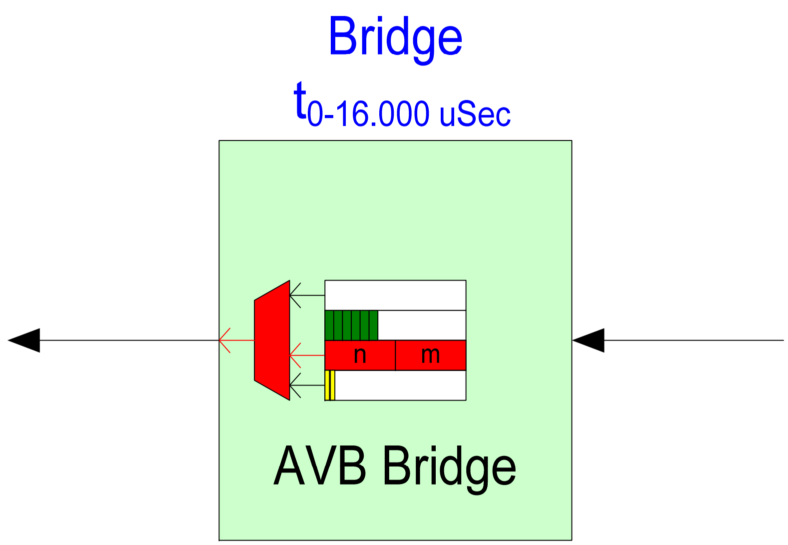
A New Idea

- **Take advantage of the target low latency data pattern**
  - i.e., That they are Small Bursts of frames at known regular intervals (for example: a 40 uSec long burst of data every 125 uSec)
- **Use this information to delay the start of non-Class A frames just before the start of the Burst Window**
  - This insures the egress port is idle so the Class A burst is not interfered
  - This is done by creating a Time Aware Blocking Shaper (TABS)
  - The Shaper can be smart and let frames out based on their size
    - Queue 1 may have a 1522 byte frame ready to go, but if it is started it would interfere with the start of the Burst Window so it is Blocked
    - While at the same time a lower priority queue 0 may have a 64 byte frame ready to go, which is allowed to start as it will finish before the start of the Burst Window

Inside an AVB Bridge w/TABS (example)

TABS Time Progression – Fig 1

- At Bridge $t_{0\text{-}16.000\ \text{uSec before the start of the Burst Window}}$ the Green Class B frames are being Shaped (gated) by Qav and can't Transmit
- So the Red Max size non-AVB High Priority frame 'n' can start

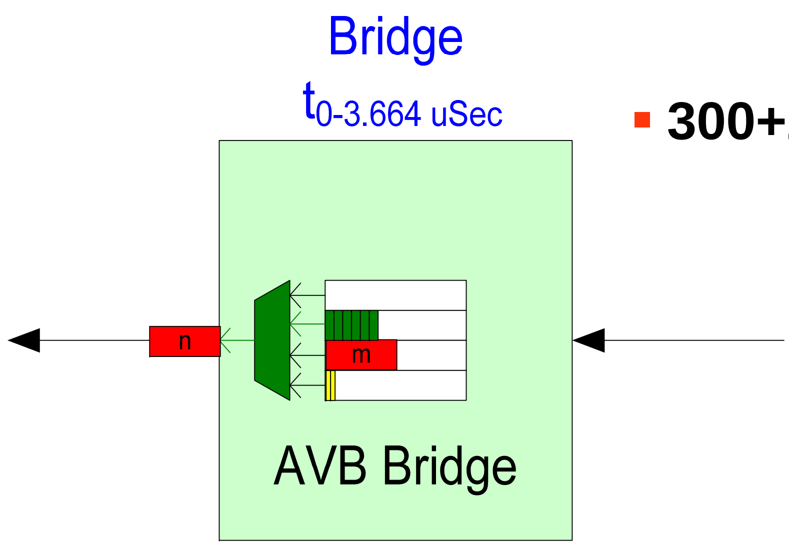


Inside an AVB Bridge w/TABS (example)

TABS Time Progression – Fig 2

- At Bridge $t_{0\text{-}3.664\text{ uSec before the start of the Burst Window}}$ the interfering Red Non-AVB frame is done
- Now the Green Class B frames are available for transmit with enough credits to burst two frames

- **16.000-12.336=**



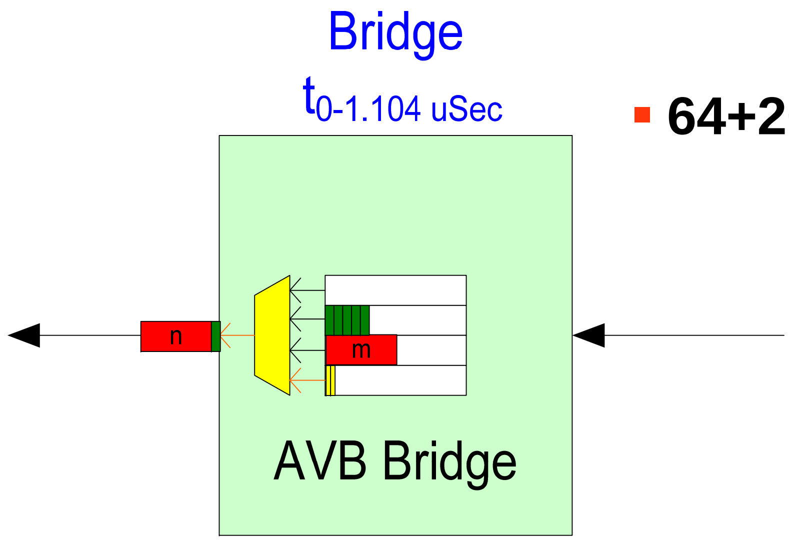
- **300+20 bytes = 2.560**

Inside an AVB Bridge w/TABS (example)

TABS Time Progression – Fig 3

- At Bridge $t_{0\text{-}1.104\ \text{uSec before the start of the Burst Window}}$ the 1st Green Class B frame is done
- Now the next Green Class B frame has credit to go, but it can't because there is not enough time before t_0 - the start of the Burst Window
- The higher priority Red 'm' frame can't go for the same reason
- But the 64 byte low priority Yellow non-AVB frame can go and does

- **3.664-2.560=**



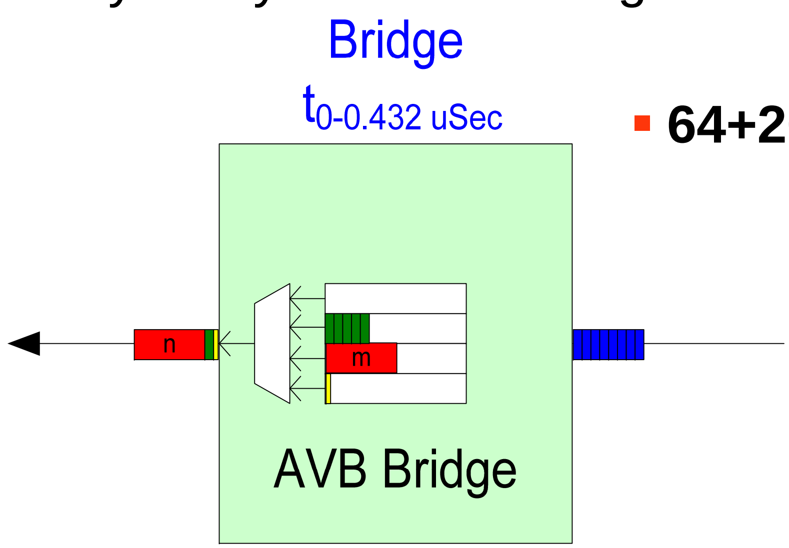
- **64+20 bytes = 672**

Inside an AVB Bridge w/TABS (example)

TABS Time Progression – Fig 4

- At Bridge $t_{0\text{-}0.432\text{ uSec before the start of the Burst Window}}$ the 64 byte Yellow frame is done
- The next Green Class B frame has credit to go, but it still can't because there is not enough time before t_0 (its credits are actually increasing)
- Same issue for the high priority non-AVB Red frame 'm'
- The next low priority 64 byte frame can't go either – not enough time

- **1.104-0.672=**

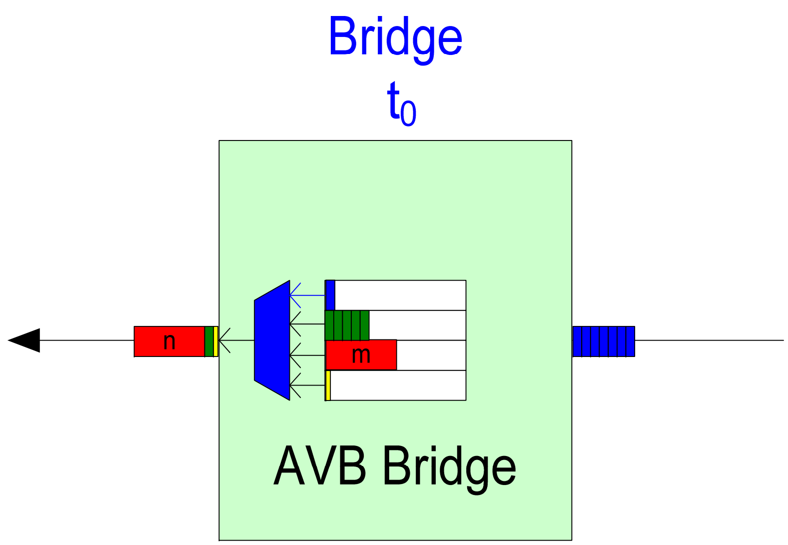


- **64+20 bytes = 672**

Inside an AVB Bridge w/TABS (example)

TABS Time Progression – Fig 5

- At Bridge t_{0} - the start of the Burst Window the port is idle so the newly arrived Blue Class A frames are allowed to egress without any interference!
- A small period after Bridge t_0 TABS release the gating on all the non-Class A queues
- The burst of Blue frames will continue since they are the top priority, but as soon as the burst is done the next higher priority frames will go



The New Idea – Pro's & Con's

Pro's:

- Modifying a port's Scheduler via a Shaper is consistent with AVB Gen 1
  - Egress ports that support 802.1Qav already gate the transmission of data out a queue – so the gating logic is already supported in the designs
  - This just adds a new time aware gate to the non-Class A queues
- Requires no other changes – the MACs, Transmit & Receive mechanisms and the Buffer controllers are not modified
- No Fragmenting and no CRC's
- Very high % usage of the line is possible

Con's:

- Only works with small bursts that occur at regular intervals
  - Although with GE ports even 75% Class A utilization is still possible
- Need a new protocol to communicate the interval period and when it will start (i.e., the exact time relative to the Grand Master)

The New Idea – in Talkers

- TABS is needed in Talkers too to prevent interference
- It requires one addition however
- Where Time Aware Blocking Shaper is needed on the non-Class A queues, The Class A queue needs a Time Aware De-blocking Shaper (TADS)
- Software in a Talker will build up the burst of frames to send out and load them up in the Class A queue
- TADS makes sure the start of the Class A burst from the Talker occurs at exactly the correct time, when the other Talker and Bridge queues in the network are all idle – it defines when to open the flood (or burst) gate

Where Do We Stand?

Bridge Latency Equation Choices

- **AVB Gen 1: @GE w/8 300 byte frames = 137.35 uSec**

Gen 1: $\text{Max Latency}_{\text{Class A Bridge}} = t_{\text{Bridge Delay}} + t_{\text{Class A Interval}} - t_{\text{Tx (Stream's Frame + 20) * 1.333}} + t_{\text{Tx Max Frame}} + t_{\text{Tx (Stream's Frame)}}$

- **Gen 2 Ty1 – No Shaper w/Max Size @1522 = 13.898 uSec**

Gen2 B Ty1: $\text{Max Latency}_{\text{Class A Bridge}} = t_{\text{Bridge Delay}} + t_{\text{Max Frame}} + t_{\text{Tx Max Cable}}$

- **Gen 2 Ty2 – No Interfere w/Max Stream @300 = 4.122 uSec**

Gen2 B Ty2: $\text{Max Latency}_{\text{Class A Bridge}} = t_{\text{Bridge Delay}} + t_{\text{Max Stream Frame}} + t_{\text{Tx Max Cable}}$

  - This value will increase as the stream size increases up to Gen 2 Ty1!

Can this be improved more?

- Yes, by supporting a Time Aware Cut-through Shaper in the Bridge (TACS)
- Cut-through bridges generally don't help normal network performance due to the low percentage of improved latency and that this improvement cannot be guaranteed
- With TACS the improved latency CAN be guaranteed
- Cut-through only works when the target ports are idle and TABS does exactly that – thus the guarantee
- And TADS makes sure the burst that needs to be cut-through shows up at the correct time

Gen 2 Bridge Type 3 Class A Equation is:

Gen2 B Ty3: Max $Latency_{Class\ A\ Bridge} = t_{Bridge\ Delay} + t_{Cut\ Through\ Point} + t_{Tx\ Max\ Cable}$

- **Total time is:**
  - Internal delay of the Bridge
  - Plus the time to receive enough of the stream to map it
  - Plus the time to get the bits down a 100 meter cable
- **To Daisy Chain the Eight 300+20 Byte Class A Frames on GE:**
  - Assuming Bridge Delay = 2 slots times typical & Cut Through time of 1 slot time
  - Max Latency = 1.024 uSec + 0.512 uSec + 0.538 uSec = **2.074 uSec**
  - This is the additional time it takes to get the last bit of the last frame to the next device
- **This number will NOT grow**
- **It is constant regardless of stream size!**

Bridge Latency Gen 2 Choices

- **Gen 2 Ty1 – No Shaper w/Max Size @1522 = 13.898 uSec**

Gen2 B Ty1: Max $Latency_{Class\ A\ Bridge} = t_{Bridge\ Delay} + t_{Max\ Frame} + t_{Tx\ Max\ Cable}$

- **Gen 2 Ty2 – No Interfere w/Max Stream @300 = 4.122 uSec**

Gen2 B Ty2: Max $Latency_{Class\ A\ Bridge} = t_{Bridge\ Delay} + t_{Max\ Stream\ Frame} + t_{Tx\ Max\ Cable}$

  - This value will increase as the stream size increases up to Gen 2 Ty1!

- **Gen 2 Ty3 – Time Aware Cut-through Shaper = 2.074 uSec**

Gen2 B Ty3: Max $Latency_{Class\ A\ Bridge} = t_{Bridge\ Delay} + t_{Cut\ Through\ Point} + t_{Tx\ Max\ Cable}$

  - This value will NOT increase as the stream size increases!

Bridge Latency Gen 2 Choices w/32 Hops

- **Gen 2 Ty1 – No Shaper w/Max Size @1522 = 444.736 uSec**

Gen2 B Ty1: Max Latency$_{\text{Class A Bridge}}$ = $t_{\text{Bridge Delay}} + t_{\text{Max Frame}} + t_{\text{Tx Max Cable}}$

- **Gen 2 Ty2 – No Interfere w/Max Stream @300 = 131.904 uSec**

Gen2 B Ty2: Max Latency$_{\text{Class A Bridge}}$ = $t_{\text{Bridge Delay}} + t_{\text{Max Stream Frame}} + t_{\text{Tx Max Cable}}$

  - This value will increase as the stream size increases up to Gen 2 Ty1!

- **Gen 2 Ty3 – Time Aware Cut-through Shaper = 66.368 uSec**

Gen2 B Ty3: Max Latency$_{\text{Class A Bridge}}$ = $t_{\text{Bridge Delay}} + t_{\text{Cut Through Point}} + t_{\text{Tx Max Cable}}$

  - This value will NOT increase as the stream size increases!

- **Just need to add in the Talker Latency to each of these**

MARVELL

We just need to decide what we want

- **Gen 2 Ty1: Do we re-define the shaper only?**
- **Gen 2 Ty2: Do we remove the interfering frames?**
  - If so by which method?
- **Gne 2 Ty3: Do we go for the stream size independent lowest latency with TACS?**

- **Note: All the presented numbers are applicable for the target use case of AVB streams that are in a daisy chain only**
- **Fan-in of AVB streams in a bridge will greatly change these numbers and will likely break Cut-through**

We can get very low latency if we Engineer it right

Thank You