

---

# Why Priority/Class Based PAUSE is Required?

**Asif Hazarika, Fujitsu**

**Bob Brunner, Ericsson**

ahazirik@fma.fujitsu.com

Robert.Brunner@ericsson.com

# List of Supporters

---

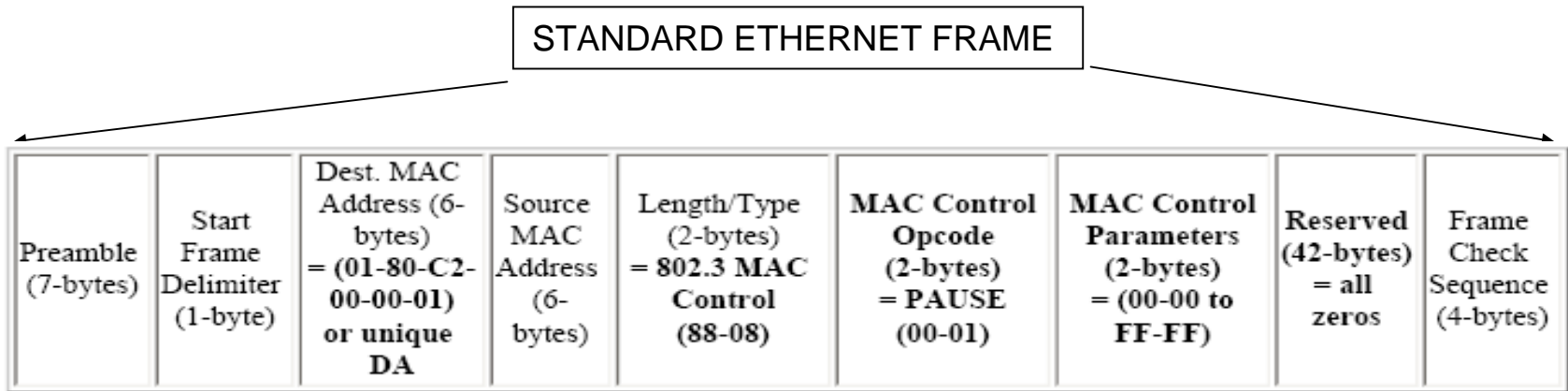
<b>Aniruddha Kundu – Intel .....</b>	<b><u><a href="mailto:Aniruddha.Kundu@intel.com">Aniruddha.Kundu@intel.com</a></u></b>
<b>Jeff Wise – Motorola ECC .....</b>	<b><u><a href="mailto:Jwise@motorola.com">Jwise@motorola.com</a></u></b>
<b>Manoj Wadekar – Intel .....</b>	<b><u><a href="mailto:Manoj.Wadekar@intel.com">Manoj.Wadekar@intel.com</a></u></b>
<b>Shashank Merchant – Nokia .....</b>	<b><u><a href="mailto:Shashank.Merchant@nokia.com">Shashank.Merchant@nokia.com</a></u></b>
<b>Uri Cummings – Fulcrum .....</b>	<b><u><a href="mailto:Uri@fulcrummicro.com">Uri@fulcrummicro.com</a></u></b>

# Agenda

---

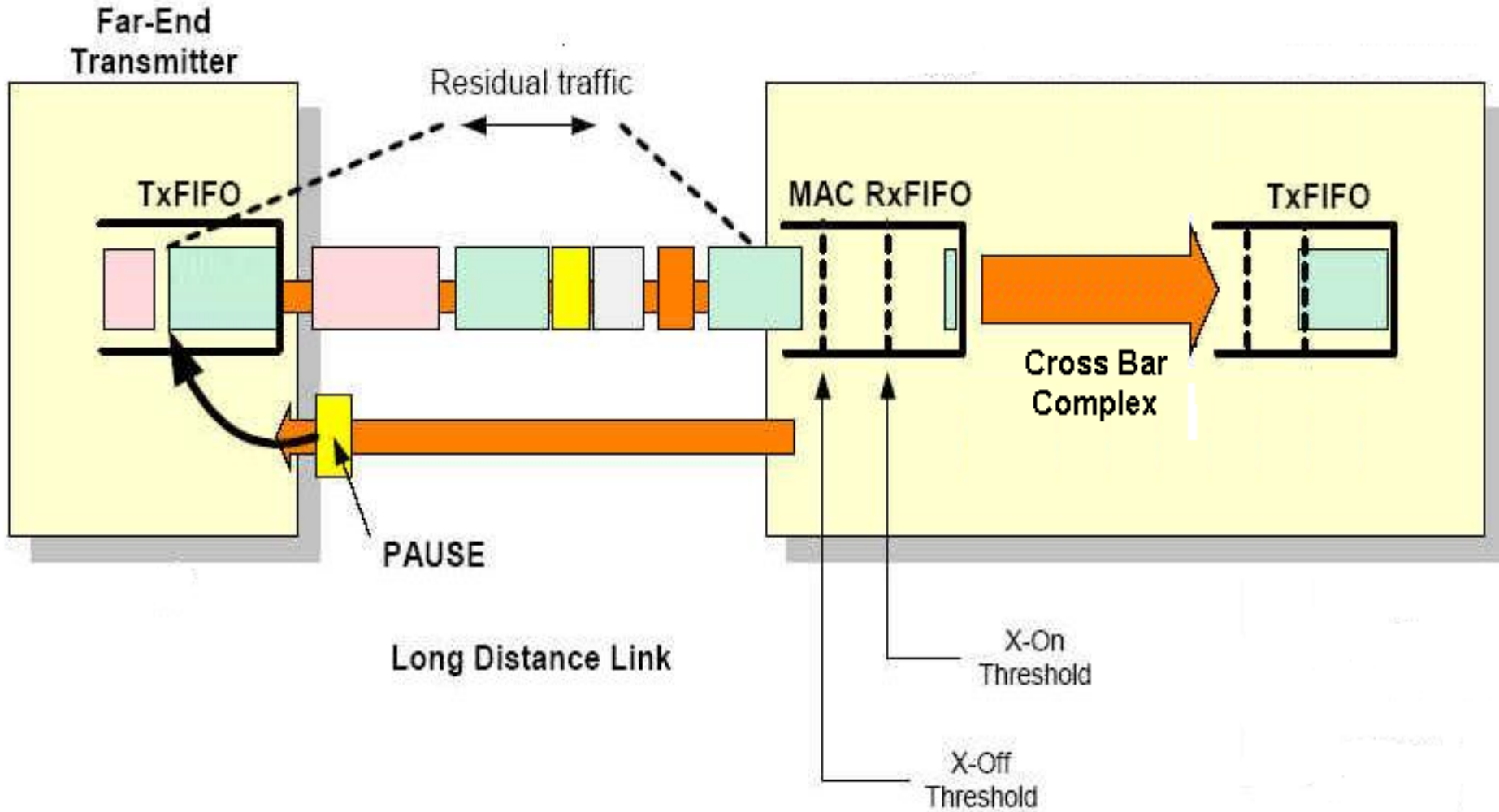
- **Current PAUSE implementation**
- **Problem with PAUSE**
- **Proposal to create a Class of Service for Pause**
- **Summary**

# PAUSE Message



- **PAUSE controls adjacent Link-Layer-Device's Transmitter**
- **PAUSE is time based (PAUSE Quantum = 512 bits of time)**
- **PAUSE time can be extended / timed-out (zero value)**

# PAUSE Mechanism exhibits coarse grained Flow Control



# PAUSE INEFFICIENCIES

---

- **Response to Pause**

- PAUSE from end-node will shutdown an adjacent LLD's transmitter
- In turn, when a switch-node's resources are exceeded, PAUSE's are propagated towards all egress ports

- **Inefficiency**

- PAUSE mechanism inevitably causes congestion spreading
- PAUSE mechanism can result in a high degree of service degradation
- shutdown of a whole link instead a particular flow causes system bandwidth to diminish
- Effect on high priority traffic
  - Higher latency and jitter for high-priority traffic
  - Delay for high priority traffic
  - No guarantees

## Solution: Selective Link Shutdown?


---

- **PAUSE for LO/HI Classes (2 levels)**
- **PAUSE for LO/MED/HI Classes (3 levels)**
- **PAUSE for each 801.1p priorities (8 levels)**
  - '7' Network Control traffic
  - '6' Delay Sensitive 10-ms Bound
  - '5' Delay Sensitive 100-ms Bound
  - '4' Delay Sensitive no-Bound
  - '3', '2' Reserved
  - '1' Reserved , Less than Best Effort
  - '0' Default, assumed BE

# How?

- **Creation of special PAUSE Frames**
  - To address COS-PAUSE which is per class of service
- **Extension to MAC Control Opcodes**
  - Use possibly non-impacting “Reserved” field.
  - New control types that can be defined.
    - PAUSE FOR DELAY SENSITIVE PRIORITY FRAMES (802.1p Class= 5 & 6)
    - PAUSE FOR BEST EFFORT PRIORITY FRAME (802.1p Class=0)
    - CONTROL FRAME WITH CREDIT/RATE/PAUSE INFORMATION FOR LOW PRIORITY CLASS
    - CONTROL FRAME WITH CREDIT/RATE/PAUSE INFORMATION FOR MEDIUM PRIORITY CLASS

Preamble (7-bytes)	Start Frame Delimiter (1-byte)	Dest. MAC Address (6- bytes) = (01-80-C2- 00-00-01) or unique DA	Source MAC Address (6- bytes)	Length/Type (2-bytes) = 802.3 MAC Control (88-08)	MAC Control Opcode (2-bytes) = PAUSE (00-01)	MAC Control Parameters (2-bytes) = (00-00 to FF-FF)	Reserved (42-bytes) = all zeros	Frame Check Sequence (4-bytes)
-----------------------	-----------------------------------------	------------------------------------------------------------------------------------	-------------------------------------------	---------------------------------------------------------------	----------------------------------------------------------	-----------------------------------------------------------------	------------------------------------------	-----------------------------------------





# Conclusions

---

- **PAUSE is very non-discriminatory, it will shut down the downlink without any concern for the priority of the traffic.**
  - This causes inefficiencies.
- **PAUSE mechanisms based on Class of Service would improve Link-Layer-Device Flow Control, allowing better management of high priority switch resources, resulting in gradual degradation of QOS**
- **An improved PAUSE mechanism would enable high priority end-to-end maintenance signaling without packet drops, facilitating a Congestion Management mechanism**
- **Implementation should result in legacy MAC's functioning as the current PAUSE mechanism.**

---

**Thank you!**