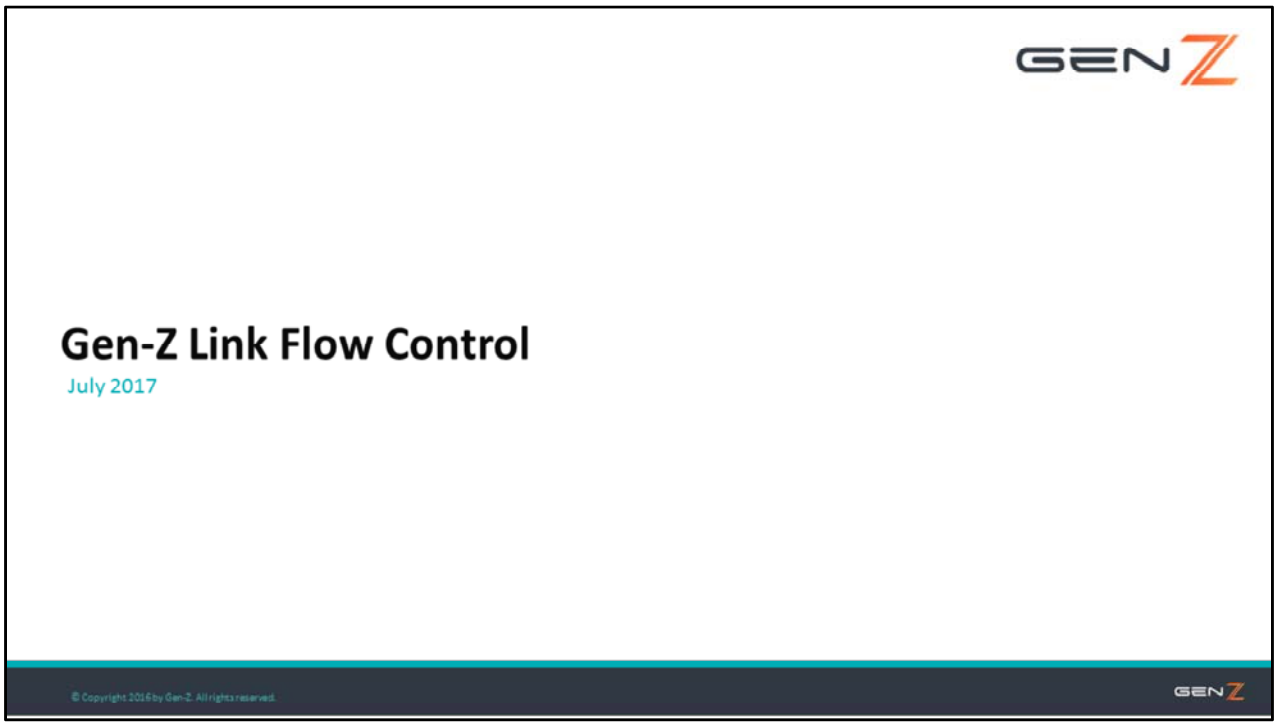


This presentation covers Gen-Z link flow control. Link flow control uses link-local packets to exchange flow-control credits.

Disclaimer

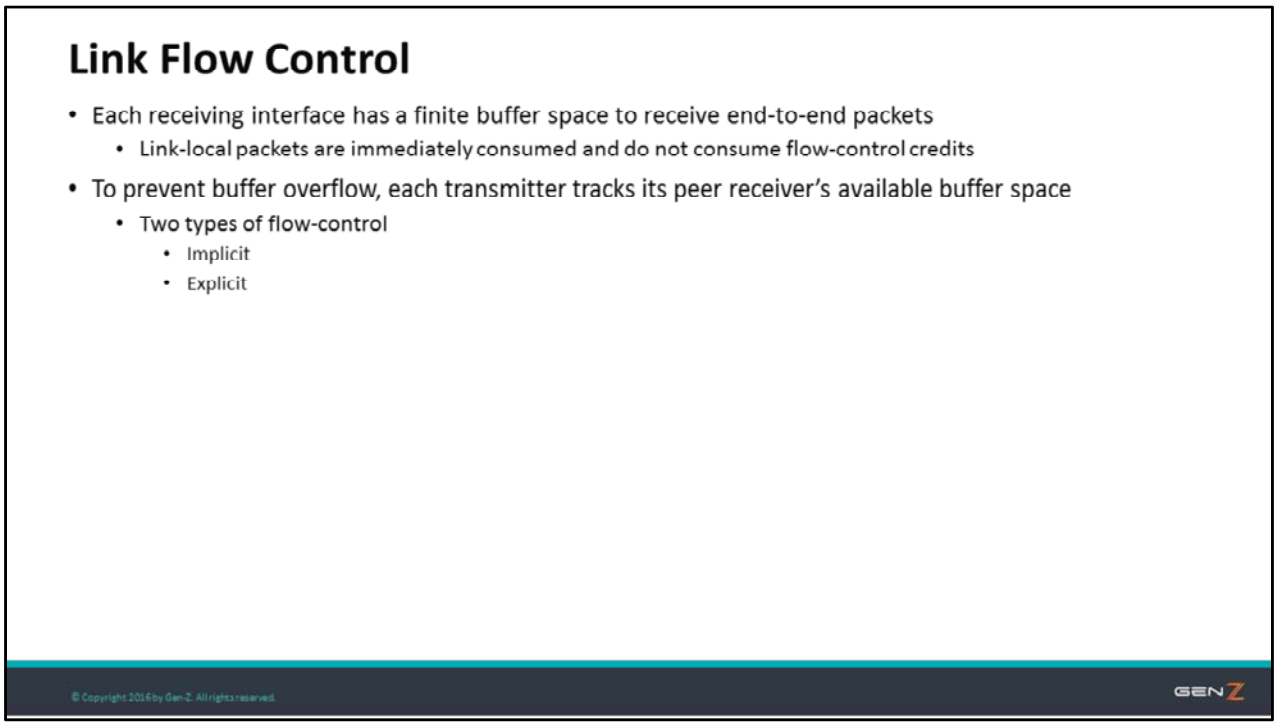
This document is provided 'as is' with no warranties whatsoever, including any warranty of merchantability, noninfringement, fitness for any particular purpose, or any warranty otherwise arising out of any proposal, specification, or sample. Gen-Z Consortium disclaims all liability for infringement of proprietary rights, relating to use of information in this document. No license, express or implied, by estoppel or otherwise, to any intellectual property rights is granted herein.

Gen-Z is a trademark or registered trademark of the Gen-Z Consortium.

All other product names are trademarks, registered trademarks, or servicemarks of their respective owners.

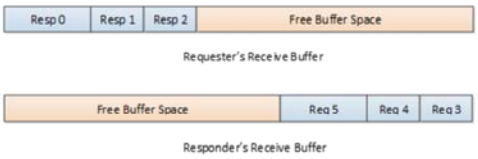
All material is subject to change at any time at the discretion of the Gen-Z Consortium

http://genzconsortium.org/



Link flow control is used to prevent buffer overflow. Gen-Z supports two types of link flow control—Implicit and Explicit.

Implicit Link Flow-Control



- Implicit flow control does not exchange link-local flow-control packets
- Implicit flow control is limited to a single point-to-point topology
  - Applicable to interfaces that support the P2P-Core, P2P-Coherency, and P2P-Vendor-defined OpClasses
  - Configured on a per component interface basis through the Interface Structure
    - Both interfaces need to support implicit flow-control
  - Requester tracks each Responder interface's available receive buffer space
    - Requester only transmits a request packet if the Responder interface has receive buffer space that can hold the request packet and the Requester has receive buffer space that can hold the corresponding response packet
      - Responder does not track Requester available buffer—assumes response buffer space is always available for each request
    - Responder provisions buffer space to hold (Max RSP Supported Requests * sizeof (max request packet size)
      - For example, if Max RSP Supported Requests = 64 and maximum P2P-Core packet size is 86 bytes (64B write + 12B protocol), then
        - Provisioned buffer = 64 * 86 bytes

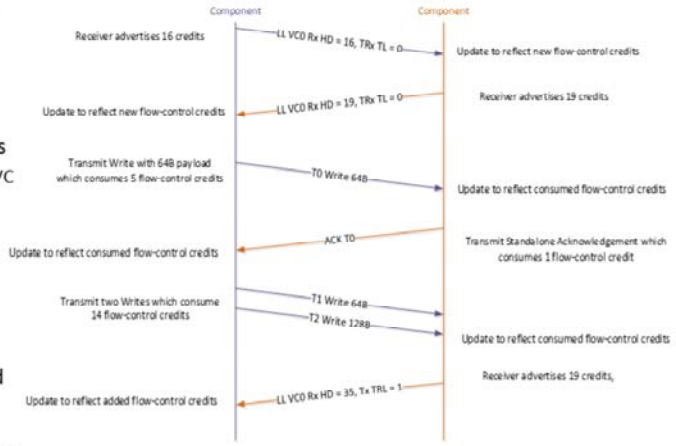
GEN Z

Implicit flow control is applicable to component interfaces that support the P2P-Core, P2P-Coherency, or P2P-Vendor-defined OpClass. Implicit flow control improves wire protocol efficiency by eliminating the exchange of explicit flow-control credit packets.

To support implicit flow control, the Requester tracks each Responder interface's available receive buffer space. To transmit a request packet, the Requester needs to ensure that it has sufficient receive buffer space to contain the corresponding response packet, and it needs to ensure that the Responder has sufficient buffer space to receive the request packet.

Explicit Flow Control

- Explicit flow control may be used in any topology
- Explicit flow control uses link-local packets to exchange each receiver's present flow-control credit state
  - Maximum flow-control credits per VC is 2^{16} credits
    - A credit represents 16 bytes for a total of 1 MiB / VC
      - Reserved bits to enable future growth
    - Sufficient to drive multi-TB/s bandwidth
    - Sufficient to drive relatively long physical distances
- Flow-control credits per VC may vary
- Optional support for adaptive flow control
  - Enables credits from under utilized VCs to be used for over-subscribed VCs to alleviate congestion
  - Eliminates the need to overprovision credits for worst-case operating conditions—reduces cost and complexity when supporting large number of VCs

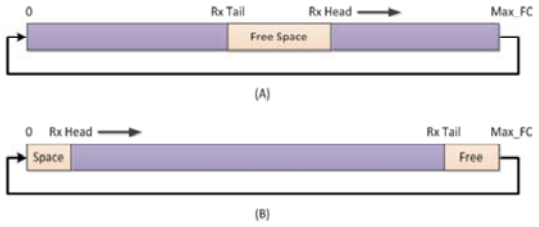


GEN Z

Explicit flow control uses link-local packets to exchange explicit flow-control credits for each enabled VC. In order to support multi-TB/s bandwidths as well as cable solutions, a link can provision up to 1 MiB of receive buffer space per VC. The actual amount of buffer space provisioned per VC may vary.

Further, an implementation can support adaptive flow control. Adaptive flow control enables a link to dynamically optimize flow-control credits by shifting credits from under utilized VCs to over subscribed VCs. Adaptive flow-control credits can reduce the probability of congestion

Logical Buffer



- Available flow-control credits are calculated using a sliding window algorithm
  - Receiver maintains a logical circular number space of size Max_FC (2^{16})
  - Receiver slides a logical window across this number space
  - The window's leading edge is referred to as the head (Rx Head) and the trailing edge as its tail (Rx Tail)
    - As buffer space becomes available, the receiver increments Rx Head
    - As buffer space is consumed, the receiver increments Rx Tail
  - Available buffer space = Rx Head – Rx Tail
  - Transmitter maintains an analog of the receiver's logical circular number space with a Tx Head and Tx Tail
    - Upon receipt of a flow-control packet, the transmitter sets Tx Head = Rx Head
    - As packets are transmitted, the transmitter sets Tx Tail = (Tx Tail + Y credits) modulo Max_FC

GEN Z

Conceptually, the receive space associated with each VC is a circular buffer containing Max_FC credits. The available receive space is calculated as the difference between the head and tail pointers (modulo Max_FC).

For each supported VC, the implementation needs to provision at least sufficient flow-control credits to support the maximum packet size (sum of the maximum number of protocol bytes for the supported OpClasses plus the maximum number of payload bytes).

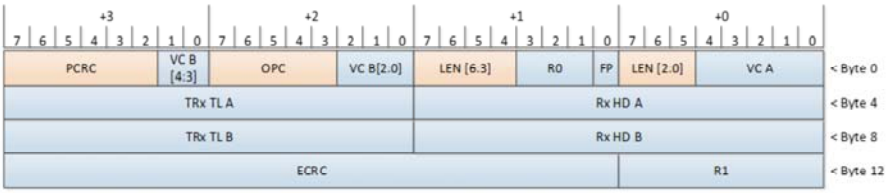
Periodic Flow-Control Credit Updates

- Each interface should transmit flow-control credit updates whenever credits are available and no higher precedence packets are awaiting transmission
- Each receiver maintains a single, interface-local FCTT (Flow Control Transmission Timer)
  - The FCTT period is 2^{24} UI
  - While the link state is L-Up or L-Up-LP, the FCTT runs continuously
    - For each enabled VC, an interface shall transmit at least one flow-control credit packet every FCTT period
    - For each enabled VC, an interface should transmit one link-local flow-control packet at every (Max_FC * Credit Size * 8 / Max_Lanes) UI, where Max_Lanes = maximum number of enabled receiver lanes
  - If an interface fails to receive at least one flow-control packet per enabled VC for two consecutive FCTT periods, then the interface shall initiate physical layer retraining
  - FCTT is suspended whenever the link state is L-LP or L-Down and during physical layer retraining

GEN Z

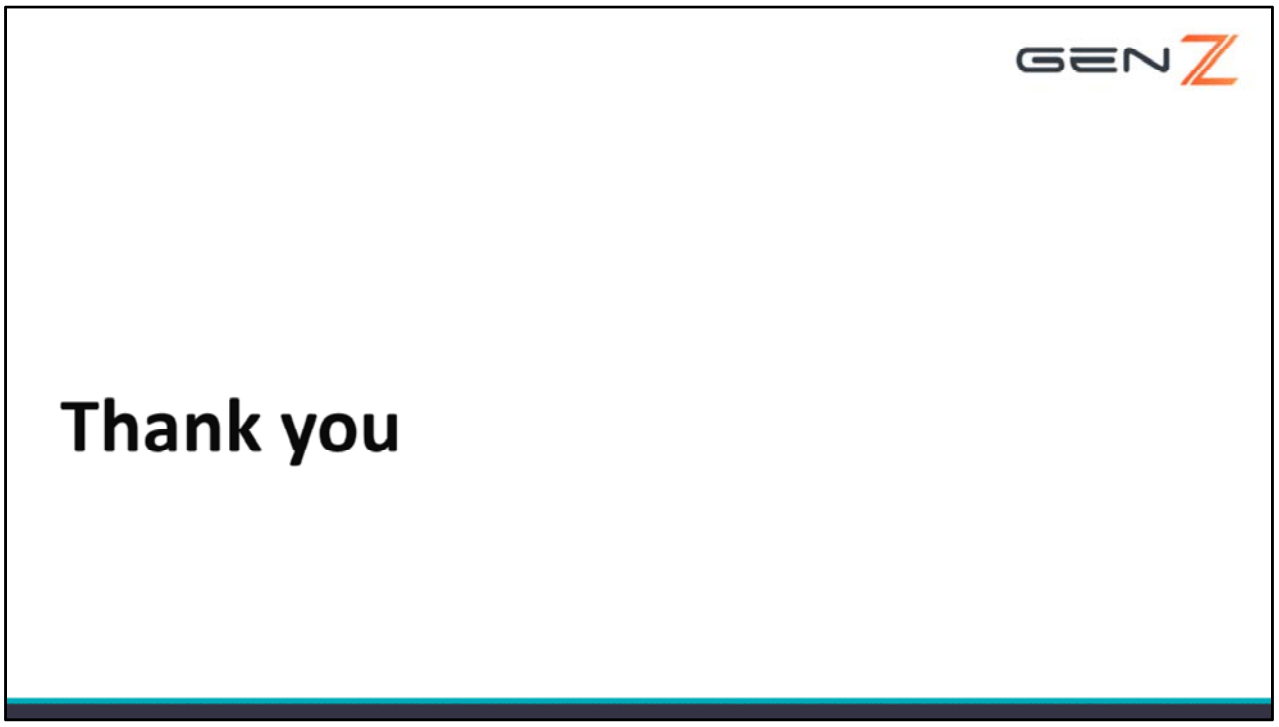
Since link-local packets are transmitted as unreliable datagrams, each interface is required to periodically transmit flow-control credit packets for each enabled VC. If an interface fails to receive flow-control credits for each enabled VC for two consecutive timer expirations, then the interface automatically initiates physical layer retraining.

Link-Local Dual-VC Flow-Control Packet Format



- Single-VC + LLR and Dual-VC (shown) flow-control packet formats
- VC A and VC B identify the virtual channels
- Rx HD A and Rx HD B
  - Advance the receiver's understanding of the head pointer associated with each VC
- TRx TL A and TRx TL B
  - Transmitter's understanding of the peer receiver's tail associated with the indicated VC
  - Used to ensure interfaces remain synchronized in the event of transient errors

There are two types of link flow-control packets—a single VC-LLR (link-level reliability) combination packet and a dual-VC packet. The single VC / LLR combination can be used to improve protocol efficiency by eliminating the need for a separate LLR acknowledgment packet.



This concludes this presentation. Thank you.