# Gen-Z Data Integrity

April 2019

See https://genzconsortium.org/white-papers/ for additional technical papers on data integrity



This presentation covers Gen-Z Data Integrity techniques.

# Disclaimer

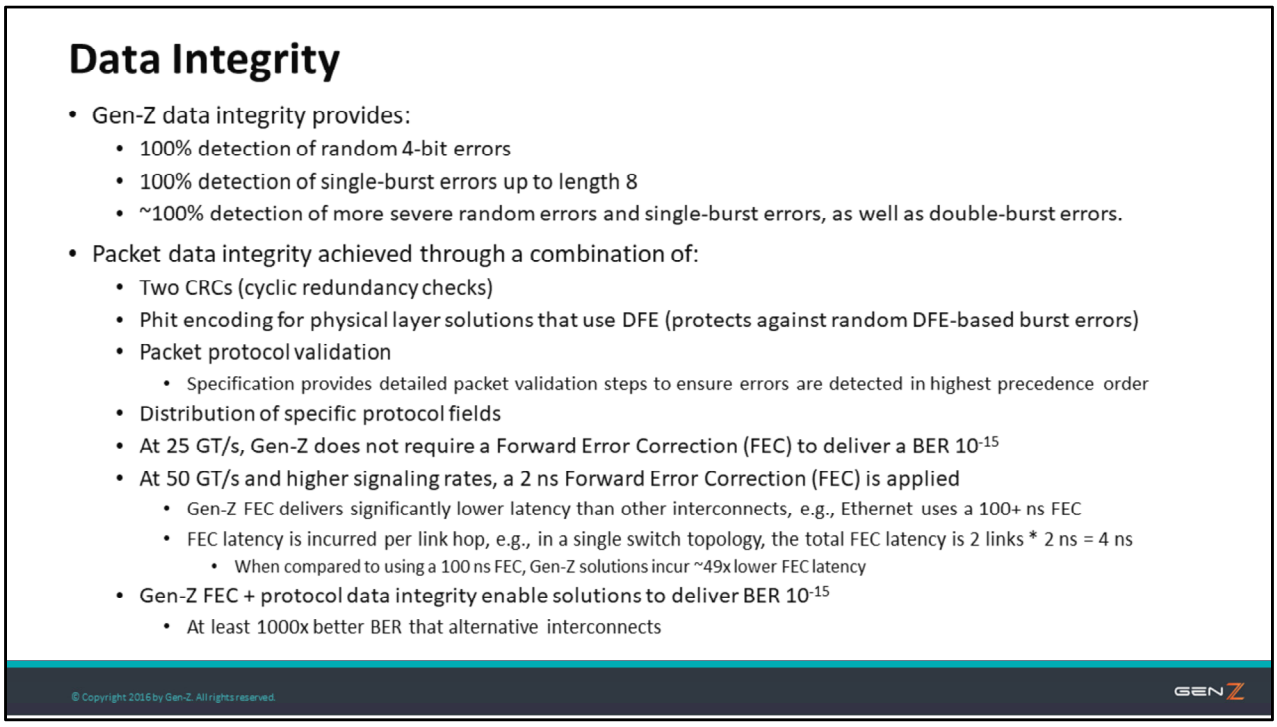
This document is provided 'as is' with no warranties whatsoever, including any warranty of merchantability, noninfringement, fitness for any particular purpose, or any warranty otherwise arising out of any proposal, specification, or sample. Gen-Z Consortium disclaims all liability for infringement of 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/



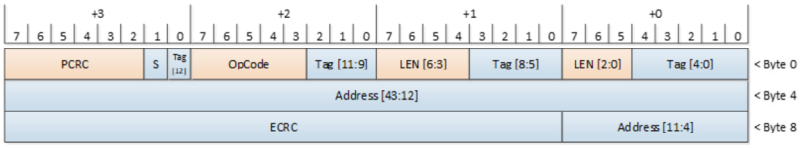

Gen-Z provides excellent packet data integrity capable of detecting random and burst errors.

All Gen-Z packets contain two CRCS. If a physical layer implementation requires DFE (decision feedback equalization), then a Phit encoding layer is transparently inserted, and this provides an additional layer of protection.

Gen-Z PHY specifies a lower raw BER (Bit Error Ratio) in order to deliver a superior solution-level BER.

Memory solutions require at least a $10^{-15}$ BER or better to ensure optimal data integrity and solution performance especially as signaling rates increase. Best-case, alternative interconnects deliver a $10^{-12}$ BER and as poor as a $10^{-9}$ BER which translates into many more transient errors and performance loss due to packet retransmission events which in turn can lead to more fabric congestion, congestion spreading, and jitter reducing overall solution performance beyond a single link transient error event.

## Prelude CRC (PCRC)




- Link-local and end-to-end packets contain a 6-bit PCRC field
- PCRC protects against errors which could cause the packet VC and length to be incorrectly interpreted
  - PCRC code word is distributed across the first three bytes in each packet format
    - Distribution enhances burst error and lane failure detection
- Whenever a packet is (re-)transmitted, the source component interface dynamically generates the PCRC
- All component ingress interfaces (including intermediate components) shall perform PCRC validation
  - As the packet is received, the interface dynamically generates the PCRC
  - If the dynamically-generated PCRC matches the packet's PCRC, then packet processing proceeds
    - If the comparison fails, then a transient error has been detected and link resynchronization is initiated
  - If a component supports remapping any field covered by the PCRC, then PCRC validation is performed prior to field remapping
  - Switches and TRs shall perform PCRC validation prior to relaying a packet



GEN-Z

The PCRC covers a subset of the packet protocol at fixed locations. This enables immediate validation of these specific fields. For example, the PCRC covers the Length field used to determine the packet's length and locate the ECRC field. If this field is corrupted, then the link needs to be resynchronized. In end-to-end packets, the PCRC also covers the VC field, which is critical to identify which receive resources to target and to prevent buffer overflow.

The originating source interface is required to dynamically generate the PCRC and ECRC fields, and all interfaces along the path to the destination are required to validate the PCRC and ECRC fields as packets are received. This ensures that transient error detection is consistently performed at every hop within a topology.

If a packet relay component supports VC remapping, then the PCRC is modified to reflect the new PCRC.

A packet cannot be transmitted to the next hop until the PCRC is validated. If an error is detected, then the packet is discarded and recovery is initiated.

## End-to-End CRC (ECRC)

- All packets contain a 24-bit ECRC field
- Whenever a packet is (re-)transmitted, the source component interface shall dynamically generate an ECRC and place it into the packet's ECRC field
- All component ingress interfaces shall perform packet ECRC validation
  - As a packet is received, the interface shall dynamically generate an ECRC
  - If the dynamically-generated ECRC matches the packet's ECRC field, then packet validation may proceed
    - A destination component shall not complete packet execution unless the ECRCs match
  - If the comparison fails, then the stomped version of the dynamically-generated ECRC is compared to the packet's ECRC
    - The stomped version is the one's complement of the dynamically-generated ECRC
    - If these do not match, then the packet is handled per transient error processing
    - If these match, then:
      - If the packet arrived at the destination component, then the packet is handled per transient error processing
      - If the packet arrived at an intermediate component and packet transmission has not begun, then the packet is silently discarded, else packet transmission continues
- If during packet transmission, an interface detects corruption, then it shall stomp the ECRC field



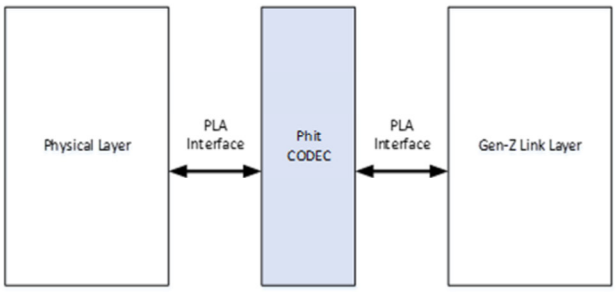
GEN-Z

All packets contain a 24-bit ECRC field. This protects the entire packet sans the ECRC field.

Unlike the PCRC validation, a packet can be relayed in spite of having a bad ECRC. For example, if cut-through packet relay is supported, then packet transmission through the egress interface can occur prior to the receipt of the ECRC field at the ingress interface. If an ECRC error is detected, then the component stomps the ECRC field to ensure the next component knows the packet is bad. All components update applicable statistics if they stomp or receive a stomped ECRC; this facilitates predictive hardware failure analysis. If transmission has not begun, then the component should silently discard the packet.

If a packet relay component supports VC remapping, then the ECRC is modified to reflect the new VC. The modification is such that if there is an error in the packet, the ECRC still detects this. Similarly, packet relay components update the Congestion field, and the ECRC is updated to reflect the new value.

## Phit Encoding



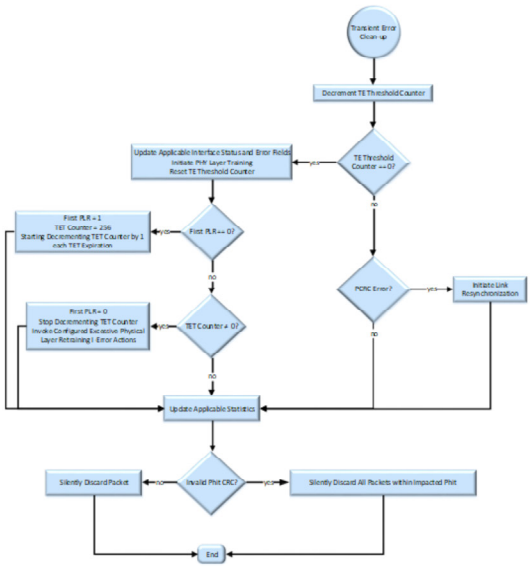

- Phit encoding layer is inserted between the physical layer and the Gen-Z link layer.
  - Applicable to physical layer implementations that use DFE.
- Packet bits segmented into 16 32-bit units of the packet stream
- Each Phit consists of a single packet, a partial packet, or multiple packets
- Each Phit is protected by a 16-bit CRC
- Phit encoding layer validates 16-bit CRC
  - If an error is detected, the corresponding packet is treated as suffering a transient CRC error
- Encoding layer contiguously places good Phits which are passed up to the Gen-Z link layer which comprehends packet boundaries



DFE is used throughout the industry across multiple technologies, primarily for longer physical channels. DFE is subject to rare, random burst errors. To detect such errors, Gen-Z specifies a Phit encoding layer that can be inserted into implementations that use DFE. Each Phit can contain a single packet, a partial packet, or multiple packets depending upon the packet sizes. Each Phit is protected by a separate 16b CRC. Should an error be detected, this is handled as a transient ECRC error.

## Transient Error Processing

- Decrement the Transient Error Threshold Counter—counter is used to determine if a link has suffered too many transient errors
- If the threshold reaches zero within the specified time period, then initiate physical layer retraining
  - Else if a PCRC error was detected, then initiate link resynchronization
- Update applicable statistics and silently discard the packet






GEN Z

It is not a question of whether transient errors will occur, but when and how often. Transient error rate depends upon the quality of the hardware implementation, environmental conditions, mechanical vibration, number of components traversed on a given path, altitude, etc. In order to facilitate predictive hardware failure report, a transient error threshold is set on each interface. If this threshold is exceeded within the specified time, then the interface automatically initiates physical layer retraining. In many cases, retraining will resolve the issue and return the physical layer to the expected BER. If physical layer retraining occurs multiple times within the specified time, then the interface can be configured to report this condition to management, and management will determine if the interface should continue to be used or shutdown.

Though Gen-Z does not mandate statistics, given the industry-wide for software-defined everything, statistics provide critical insight into operating conditions, and help software make informed decisions on actions to take, e.g., updating relay tables to bypass failing hardware, scheduling maintenance, etc.

# Thank you

This concludes this presentation.  Thank you.