Compute Express Link^{TM} (CXL^{TM})

Engineering Change Notice to the Specification

May 2021

CEDT CFMWS & QTG_DSM

LEGAL NOTICE FOR THIS PUBLICLY-AVAILABLE SPECIFICATION FROM COMPUTE EXPRESS LINK CONSORTIUM, INC.

This CXL **Specification** (this "**CXL Specification**" or this "**document**") is owned by and is proprietary to Compute Express Link Consortium, Inc., a Delaware nonprofit corporation (sometimes referred to as "**CXL**" or the "**CXL Consortium**" or the "**Company**") and/or its successors and assigns.

NOTICE TO USERS WHO ARE MEMBERS OF THE CXL CONSORTIUM:

If you are a Member of the CXL Consortium (sometimes referred to as a "**CXL Member**"), and even if you have received this publicly-available version of this CXL Specification after agreeing to CXL Consortium's Evaluation Copy Agreement (a copy of which is available https://www.computeexpresslink.org/download-the-specification, each such CXL Member must also be in compliance with all of the following CXL Consortium documents, policies and/or procedures (collectively, the "**CXL Governing Documents**") in order for such CXL Member's use and/or implementation of this CXL Specification to receive and enjoy all of the rights, benefits, privileges and protections of CXL Consortium membership: (i) CXL Consortium's Intellectual Property Policy; (ii) CXL Consortium's Bylaws; (iii) any and all other CXL Consortium policies and procedures; and (iv) the CXL Member's Participation Agreement.

NOTICE TO NON-MEMBERS OF THE CXL CONSORTIUM:

If you are **not** a CXL Member and have received this publicly-available version of this CXL Specification, your use of this document is subject to your compliance with, and is limited by, all of the terms and conditions of the CXL Consortium's Evaluation Copy Agreement (a copy of which is available at https://www.computeexpresslink.org/download-the-specification**).**

In addition to the restrictions set forth in the CXL Consortium's Evaluation Copy Agreement, any references or citations to this document must acknowledge the Compute Express Link Consortium, Inc.'s sole and exclusive copyright ownership of this CXL Specification. The proper copyright citation or reference is as follows: **"© 2019-2021 COMPUTE EXPRESS LINK CONSORTIUM, INC. ALL RIGHTS RESERVED."** When making any such citation or reference to this document you are not permitted to revise, alter, modify, make any derivatives of, or otherwise amend the referenced portion of this document in any way without the prior express written permission of the Compute Express Link Consortium, Inc.

Except for the limited rights explicitly given to a non-CXL Member pursuant to the explicit provisions of the CXL Consortium's Evaluation Copy Agreement which governs the publicly-available version of this CXL Specification, nothing contained in this CXL Specification shall be deemed as granting (either expressly or impliedly) to any party that is not a CXL Member: (ii) any kind of license to implement or use this CXL Specification or any portion or content described or contained therein, or any kind of license in or to any other intellectual property owned or controlled by the CXL Consortium, including without limitation any trademarks of the CXL Consortium.; or (ii) any benefits and/or rights as a CXL Member under any CXL Governing Documents.

LEGAL DISCLAIMERS FOR ALL PARTIES:

THIS DOCUMENT AND ALL SPECIFICATIONS AND/OR OTHER CONTENT PROVIDED HEREIN IS PROVIDED ON AN "**AS IS**" BASIS. TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, COMPUTE EXPRESS LINK CONSORTIUM, INC. (ALONG WITH THE CONTRIBUTORS TO THIS DOCUMENT) HEREBY DISCLAIM ALL REPRESENTATIONS, WARRANTIES AND/OR COVENANTS, EITHER EXPRESS OR IMPLIED, STATUTORY OR AT COMMON LAW, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, TITLE, VALIDITY, AND/OR NON-INFRINGEMENT.
In the event this CXL Specification makes any references (including without limitation any incorporation by reference) to another standard's setting organization's or any other party's ("**Third Party**") content or work, including without limitation any specifications or standards of any such Third Party (**"Third Party Specification**"), you are hereby notified that your use or implementation of any Third Party Specification: (i) is not governed by any of the CXL Governing Documents; (ii) may require your use of a Third Party's patents, copyrights or other intellectual property rights, which in turn may require you to independently obtain a license or other consent from that Third Party in order to have full rights to implement or use that Third Party Specification; and/or (iii) may be governed by the intellectual property policy or other policies or procedures of the Third Party which owns the Third Party Specification. Any trademarks or service marks of any Third Party which may be referenced in this CXL Specification is owned by the respective owner of such marks.

NOTICE TO ALL PARTIES REGARDING THE PCI-SIG UNIQUE VALUE PROVIDED IN THIS CXL SPECIFICATION:

NOTICE TO USERS: THE UNIQUE VALUE THAT IS PROVIDED IN THIS CXL SPECIFICATION IS FOR USE IN VENDOR DEFINED MESSAGE FIELDS, DESIGNATED VENDOR SPECIFIC EXTENDED CAPABILITIES, AND ALTERNATE PROTOCOL NEGOTIATION ONLY AND MAY NOT BE USED IN ANY OTHER MANNER, AND A USER OF THE UNIQUE VALUE MAY NOT USE THE UNIQUE VALUE IN A MANNER THAT (A) ALTERS, MODIFIES, HARMS OR DAMAGES THE TECHNICAL FUNCTIONING, SAFETY OR SECURITY OF THE PCI-SIG ECOSYSTEM OR ANY PORTION THEREOF, OR (B) COULD OR WOULD REASONABLY BE DETERMINED TO ALTER, MODIFY, HARM OR DAMAGE THE TECHNICAL FUNCTIONING, SAFETY OR SECURITY OF THE PCI-SIG ECOSYSTEM OR ANY PORTION THEREOF (FOR PURPOSES OF THIS NOTICE, "**PCI-SIG ECOSYSTEM**" MEANS THE PCI-SIG SPECIFICATIONS, MEMBERS OF PCI-SIG AND THEIR ASSOCIATED PRODUCTS AND SERVICES THAT INCORPORATE ALL OR A PORTION OF A PCI-SIG SPECIFICATION AND EXTENDS TO THOSE PRODUCTS AND SERVICES INTERFACING WITH PCI-SIG MEMBER PRODUCTS AND SERVICES).

CXL ENGINEERING CHANGE NOTICE

TITLE:	CEDT CFMWS & QTG _DSM
DATE:	Introduced date (11/00/2020) Updated date (05/18/2021) - Incorporated Erratum F28.
AFFECTED DOCUMENT:	CXL 2.0
SPONSOR:	Chet Douglas, Intel

Part I

1. Summary of Functional Changes

This ECN adds a new structure to the CEDT – CXL Early Discovery Table. Each instance of the structure, CXL Fixed Memory Window Structure (CFMWS), describes a range of coherent memory in terms of Host Physical Addresses (HPA) and associated restrictions that govern how these addresses may be assigned to CXL.mem capable devices.

These memory windows generally represent the host specific decoders that are set up by the System Firmware prior to the OS hand-off and cannot be modified by the OS. The OS utilizes the available fixed memory windows to assign HPA ranges to CXL memory devices it discovers during its CXL enumeration, or to handle CXL topology changes or any reassignments.

The restrictions describe the CXL Host Bridge, the type of CXL device (Type 2 vs. Type 3), the type of memory (persistent vs. volatile) and the QoS Throttling Group (QTG) that the fixed range can be assigned to. The host is expected to throttle all HPA ranges that are mapped to a single QTG as one unit. Therefore, it is important for the OS to take QTG into account when assigning the HPA. This ECN also introduces a new Device Specific Method (_DSM) so that the OS can query platform about the QTG that a given device HDM range should be mapped to.

2. Benefits as a Result of the Changes

OS utilizes the available fixed memory windows to assign HPA ranges to CXL memory devices it discovers during its CXL enumeration, or to handle CXL topology changes or any reassignments during runtime. In absence of this feature, System Firmware is required to assign HPA addresses to all CXL device and this may disallow certain usage models such as native OS hot-add or late address assignment.

The _DSM enables the OS to select the optimum QTG for a given device range without having to understand the host implementation aspects and policies regarding QTG assignment.

3. Analysis of the Hardware Implications

This ECN has no impact on CXL HW.

4. Analysis of the Software Implications

This is an optional feature that enables the OS to natively assign addresses to CXL.mem devices. Changes to System Firmware and OS are needed to enable these capabilities.

5. Analysis of the Compliance and Test Implications

This ECN has no impact on Compliance and Test.

Part II

Detailed Description of the change

Update Table 1 as follows:

Term/Acronym	Definition
..	..
QTG	QoS Throttling Group, the group of CXL.mem target resources that are throttled together in response to QoS telemetry (see section 3.3.2). Each QTG is identified using a number that is known as QTG ID.QTG ID is a positive integer.
..	..

Update 9.14.1.1 CEDT Header section as follows:

Table 215. CEDT Structure Types

Value	Description
0	CXL Host Bridge Structure
1	CXL Fixed Memory Window Structure (CFMWS)
~~1~~2-FF	Reserved

Add new section 9.14.1.3 as follows:

9.14.1.3 CXL Fixed Memory Window Structure (CFMWS)

The CFMWS structure describes zero or more Host Physical Address (HPA) windows associated with each CXL Host Bridge. Every window represents a contiguous HPA range that may be interleaved across one or more targets, some of which are CXL Host Bridges. Associated with each window, are a set of restrictions that govern its usage. It is the OSPM's responsibility to utilize each window for the stated use.

The HPA ranges described by CFMWS may include addresses that are currently assigned to CXL.mem devices. Before assigning HPAs from a fixed memory window, the OSPM must check the current assignments and avoid any conflicts.

For any given HPA address, it shall not be described by more than one CFMWS entry.

Table XXX. CFMWS Structure

Field	Byte Length	Byte Offset	Description
Type	1	0	1 = indicates this is a CFMWS entry
Reserved	1	1	Reserved
Record Length	2	2	Length of this record = 36 + 4 * NIW NIW is the raw count of Interleave ways whereas ENIW is the encoded value. If ENIW<8, NIW=2**ENIW
Reserved	4	4	Reserved
Base HPA	8	8	Base of this HPA range. This value shall be a 256 MB aligned address.
Window Size	8	16	The total number of consecutive bytes of HPA this window represents. This value shall be a multiple of NIW*256 MB.
Encoded Number of Interleave Ways (ENIW)	1	24	The encoded number of targets that this window is interleaved with. The valid encoded values are specified in Interleave Ways field of the CXL HDM Decoder Control Register (section 8.2.5.12.7). This field ~~identifies~~ determines the number of entries in the Interleave Target List starting at Offset 36.
Interleave Arithmetic	1	25	This field defines the arithmetic used for mapping HPA to an interleave target in the Interleave Target List. 0 - Standard Modulo arithmetic All other values are reserved.
Reserved	2	26	Reserved
Host Bridge Interleave Granularity (HBIG)	4	28	The number of consecutive bytes within the interleave that are decoded by each target in the Interleave Target List represented in an encoded format. ~~t~~The valid values are specified in the CXL HDM Decoder Control Register (section 8.2.5.12.7), Interleave Granularity field.

Field	Byte Length	Byte Offset	Description
Window Restrictions	2	32	A bitmap describing the restrictions being placed on the OSPM's use of the window. It is the OSPM's responsibility to adhere to these restrictions. Failure to adhere to these restrictions results in undefined behavior. More than one bit in this bit field may be set. Bit[0]: CXL Type 2 Memory: When set, the window is configured to expose host addressable memory from CXL Type 2 memory devices. Bit[1]: CXL Type 3 Memory: When set, the window is configured to expose host addressable memory from CXL Type 3 memory devices. Bit[2]: Volatile: When set, the window is configured for use with volatile memory. Bit[3]: Persistent: When set, the window is configured for use with persistent memory. Bit[4]: Fixed Device Configuration: When set, it indicates that any device ranges that have been assigned an HPA from this window must not be reassigned. Bits[15:5]: Reserved
QTG ID	2	34	The ID of the QoS Throttling Group associated with this window. The _DSM for retrieving QTG ID is utilized by the OSPM to determine which QTG a device HDM range should be assigned to. This field must not exceed the Max Supported QTG ID returned by the _DSM for retrieving QTG.

Field	Byte Length	Byte Offset	Description
Interleave Target List	4 * ~~HB~~NIW	36	A list of all the Interleave Targets. The number of entries in this list shall match the Number of Interleave Ways (NIW). The order of the targets reported in this List indicates the order in the Interleave Set. For interleave sets that only span CXL Host Bridges, this is a list of CXL Host Bridge _UIDs that are part of the Interleave Set. In this case, for each _UID value in this list, there must exist a corresponding CHBS structure. If the interleave set spans non-CXL domains, this list may contain values that do not match _UID field in any CHBS structures. These entries represent Interleave Targets that are not CXL Host Bridges. The set of HPAs decoded by Entry N in the Interleave Target List shall satisfy the following equations 1. Base HPA <= HPA < Base HPA + Windows Size 2. If (Interleave Arithmetic==0) 2.1 If ENIW=0 N=0 2.2 If ENIW=1 N= HPA[8+HBIG] 2.3 If ENIW<8 AND ENIW>1 N = HPA[7+HBIG+ENIW:8+HBIG] N is 0 based (0<= N <NIW).

Add new section 9.14.3 as follows:

9.14.3 CXL Root Device Specific Methods (_DSM)

_DSM is a control method that enables devices to provide device specific functions for the benefit of the device driver. See ACPI Specification for details. The table below lists the _DSM functions associated with the CXL Root Device (HID="ACPI0017") .

New Table XXX: _DSM Definitions for CXL Root Device

UUID	Revision	Function	Description
F365F9A6-A7DE-4071-A66A-B40C0B4F8E52	1	1	Retrieve QTG ID (See section 9.14.3.1)
	-	All others	Reserved

All other Function values are reserved. The Revision field represents the version of the individual _DSM Function. The Revision associated with a _DSM Function is incremented whenever that _DSM Function is extended to add more functionality. Backward compatibility shall be maintained during this process. Specifically, for all values of n, a _DSM Function with Revision n+1 may extend Revision ID n by assigning meaning to the fields that are marked as reserved in Revision n but must not redefine the meaning of existing fields and must not change the size or type of input/output parameters. Software that was written for a lower Revision may continue to operate on _DSM Functions with a higher Revision but will not be able to take advantage of new functionality. It is legal for software to invoke a _DSM function and pass in any non-zero Revision ID value that does not exceed the revision ID defined in this specification for that _DSM function.

For example, if the most current version of this specification defines Revision ID=4 for _DSM Function Index f, software is permitted to invoke the _DSM function with Function Index f with a Revision ID value that belongs to the set {1, 2, 2, 4}.

9.14.3.1 _DSM Function for Retrieving QTG ID

This section describes how the OSPM can request the firmware to determine the optimum QoS Throttling Group (QTG) a device HDM range should be assigned to, based on its performance characteristics. It is strongly recommended that OSPM evaluate this _DSM Function to retrieve QTG recommendations and map the device HDM range to an HPA range that is described by a CFMWS entry that follows the platform recommendations.

For each Device Scoped Memory Affinity Structure (DSMAS) in the Device CDAT, the OSPM should calculate the Read Latency, Write Latency, Read Bandwidth and Write Bandwidth from the CXL Root Port in the same VCS. The term DSMAS is defined in Coherent Device Attribute Table specification. This calculation must consider the latency and bandwidth contribution of any intermediate switches. The OSPM should call this _DSM with the performance characteristics for the Device HDM range thus calculated, utilize the return ID value(s) to pick an appropriate CFMWS, and map the DSMAS DPA range to HPA addresses that are covered by that CFMWS. This process may be repeated for each DSMAS memory range the OSPM wishes to utilize from the device.

Location:
This object shall be a child of the CXL Root Device (HID=”ACPI0017”).

Arguments:
Arg0: UUID: f365f9a6-a7de-4071-a66a-b40c0b4f8e52
Arg1: Revision ID: 1
Arg2: Function Index: 01h
Arg3: A **package** of memory device performance characteristic. The package consists of 4 DWORDs.

```
      Package {
          Read Latency
          Write Latency
          Read Bandwidth
          Write Bandwidth
```

```
        }
```

Return: A **package** containing two elements – a WORD that return the maximum throttling group the platform supports and a package containing the QTG ID(s) the platform recommends

```
        Package {
                Max Supported QTG ID
                Package {QTG Recommendations}
        }
```

New Table XXX: _DSM for retrieving QTG, Inputs and outputs

Field Name	Size	Description
Input Package:		
Read Latency	DWORD	The best case read latency as measured from the CXL root port in the same VCS, expressed in picoseconds
Write Latency	DWORD	The best case write latency as measured from the CXL root port in the same VCS, expressed in picoseconds
Read Bandwidth	DWORD	The best case read bandwidth as measured from the root port in the same VCS, expressed in MB/s
Write Bandwidth	DWORD	The best case write bandwidth as measured from the root port in the same VCS, expressed in MB/s
Return Package:		
Max Supported QTG ID	WORD	The highest QTG ID supported by the platform. The platform must be capable of supporting all QTGs whose ID, Q satisfy the following equation 0 > Q >= Max Supported QTG ID, For every value of Q, there may be zero or more CFMWS entries.
QTG Recommendations	Package	A package consisting of 0 or more WORD elements. It is a prioritized list of QTG IDs that are considered acceptable by the platform for the specified performance characteristics. If the Package contains more than one element, element[n] is preferred by the platform over element[n+1]. If the Package is empty, it indicates that the platform is unable to find any suitable QTG for this set of input values. If the OSPM does not follow platform QTG recommendations, it may result in severe performance degradation. Every element in this package must be no greater than the Max Supported QTG ID above. For example, if QTG Recommendations = Package () { 2, 1}, the OSPM should attempt to

		assign from QTG ID 2 first and attempt to assign QTG ID 1 if no assignment can be found in QTG ID 2.