



1
2
3
4

Document Identifier: DSP0238

Date: 2024-01-22

Version: 1.2.1

5
6
7

Management Component Transport Protocol (MCTP) PCIe VDM Transport Binding Specification

8
9
10
11

Supersedes: 1.2.0

Document Class: Normative

Document Status: Published

Document Language: en-US

12 Copyright Notice

13 Copyright © 2009, 2014, 2018, 2021, 2024 DMTF. All rights reserved.

14 DMTF is a not-for-profit association of industry members dedicated to promoting enterprise and systems
15 management and interoperability. Members and non-members may reproduce DMTF specifications and
16 documents, provided that correct attribution is given. As DMTF specifications may be revised from time to
17 time, the particular version and release date should always be noted.

18 Implementation of certain elements of this standard or proposed standard may be subject to third-party
19 patent rights, including provisional patent rights (herein “patent rights”). DMTF makes no representations
20 to users of the standard as to the existence of such rights and is not responsible to recognize, disclose, or
21 identify any or all such third-party patent right owners or claimants, nor for any incomplete or inaccurate
22 identification or disclosure of such rights, owners, or claimants. DMTF shall have no liability to any party,
23 in any manner or circumstance, under any legal theory whatsoever, for failure to recognize, disclose, or
24 identify any such third-party patent rights, or for such party’s reliance on the standard or incorporation
25 thereof in its products, protocols, or testing procedures. DMTF shall have no liability to any party
26 implementing such standards, whether such implementation is foreseeable or not, nor to any patent
27 owner or claimant, and shall have no liability or responsibility for costs or losses incurred if a standard is
28 withdrawn or modified after publication, and shall be indemnified and held harmless by any party
29 implementing the standard from any and all claims of infringement by a patent owner for such
30 implementations.

31 For information about patents held by third-parties which have notified DMTF that, in their opinion, such
32 patents may relate to or impact implementations of DMTF standards, visit
33 <https://www.dmtf.org/about/policies/disclosures>.

34 PCI-SIG, PCIe, and the PCI HOT PLUG design mark are registered trademarks or service marks of PCI-
35 SIG.

36 All other marks and brands are the property of their respective owners.

37 This document’s normative language is English. Translation into other languages is permitted.

38

CONTENTS

39	Foreword	4
40	Introduction.....	5
41	1 Scope	6
42	2 Normative references	6
43	3 Terms and definitions	7
44	4 Symbols and abbreviated terms.....	7
45	5 Conventions	8
46	5.1 Reserved and unassigned values.....	8
47	5.2 Byte ordering.....	8
48	6 MCTP over PCI Express VDM transport.....	8
49	6.1 Packet format.....	8
50	6.2 Supported media.....	11
51	6.3 Physical address format for MCTP control messages.....	11
52	6.4 Message routing	12
53	6.5 Bus owner address	12
54	6.6 Bus address assignment for PCIe	13
55	6.7 Host dependencies	13
56	6.8 Discovery Notify message use for PCIe	13
57	6.9 MCTP over PCIe endpoint discovery.....	14
58	6.10 MCTP messages timing requirements.....	18
59	ANNEX A (informative) Notations and conventions.....	20
60	ANNEX B (informative) Change log.....	21
61		

62 Figures

63	Figure 1 – MCTP over PCI Express Vendor Defined Message (VDM) packet format	9
64	Figure 2 – Flow of operations for full MCTP Discovery over PCIe	16
65	Figure 3 – Flow of operations for Partial Endpoint Discovery.....	17
66		

67 Tables

68	Table 1 – PCI Express medium-specific MCTP packet fields.....	9
69	Table 2 – Supported media.....	11
70	Table 3 – Physical address format.....	11
71	Table 4 – Timing specifications for MCTP Control messages on PCIe VDM.....	18
72		

73

Foreword

74 The *Management Component Transport Protocol (MCTP) PCIe VDM Transport Binding Specification*
75 (DSP0238) was prepared by the PMCI Working Group.

76 DMTF is a not-for-profit association of industry members dedicated to promoting enterprise and systems
77 management and interoperability.

78 **Acknowledgments**

79 DMTF acknowledges the following individuals for their contributions to this document:

80 **Editors:**

- 81 • Hemal Shah – Broadcom Inc.
- 82 • Tom Slaight – Intel Corporation

83 **Contributors:**

- 84 • Patrick Caporale – Lenovo
- 85 • Yuval Itkin – NVIDIA Corporation
- 86 • Eliel Louzoun – Intel Corporation
- 87 • Patrick Schoeller – Hewlett Packard Enterprise
- 88 • Bob Stevens – Dell Technologies

89

Introduction

90 The Management Component Transport Protocol (MCTP) over PCIe VDM transport binding defines a
91 transport binding for facilitating communication between platform management subsystem components
92 (e.g., management controllers, management devices) over PCIe.

93 The [MCTP Base Specification](#) describes the protocol and commands used for communication within and
94 initialization of an MCTP network. The MCTP over PCIe VDM transport binding definition in this
95 specification includes a packet format, physical address format, message routing, and discovery
96 mechanisms for MCTP over PCIe VDM communications.

97

98 1 Scope

99 This document provides the specifications for the Management Component Transport Protocol (MCTP)
100 transport binding using PCIe Vendor Defined Messages (VDMs).

101 2 Normative references

102 The following referenced documents are indispensable for the application of this document. For dated
103 references, only the edition cited applies. For undated references, the latest edition of the referenced
104 document (including any amendments) applies.

105 CXL Consortium, *Compute Express Link™ (CXL™) Specification Revision 1.0*,
106 <https://www.computeexpresslink.org>

107 CXL Consortium, *Compute Express Link™ (CXL™) Specification Revision 1.1*,
108 <https://www.computeexpresslink.org>

109 CXL Consortium, *Compute Express Link™ (CXL™) Specification Revision 2.0*,
110 <https://www.computeexpresslink.org>

111 DMTF DSP0236, *Management Component Transport Protocol (MCTP) Base Specification 1.0*
112 https://www.dmtf.org/sites/default/files/standards/documents/DSP0236_1.0.pdf

113 DMTF DSP0236, *Management Component Transport Protocol (MCTP) Base Specification 1.3*
114 https://www.dmtf.org/sites/default/files/standards/documents/DSP0236_1.3.pdf

115 DMTF DSP0239, *Management Component Transport Protocol (MCTP) IDs and Codes 1.0*
116 https://www.dmtf.org/sites/default/files/standards/documents/DSP0239_1.0.pdf

117 DMTF DSP0239, *Management Component Transport Protocol (MCTP) IDs and Codes 1.8*
118 https://www.dmtf.org/sites/default/files/standards/documents/DSP0239_1.8.pdf

119 ISO/IEC Directives, Part 2, *Principles and rules for the structure and drafting of ISO and IEC documents*
120 <https://www.iso.org/sites/directives/current/part2/index.xhtml>

121 PCI-SIG, *PCI Express® Base Specification Revision 1.1*, March 8, 2005,
122 <http://www.pcisig.com/specifications/>

123 PCI-SIG, *PCI Express® Base Specification Revision 2.0*, December 20, 2006,
124 <http://www.pcisig.com/specifications/>

125 PCI-SIG, *PCI Express® Base Specification Revision 2.1*, March 4, 2009,
126 <http://www.pcisig.com/specifications/>

127 PCI-SIG, *PCI Express® Base Specification Revision 3.0*, November 10, 2010,
128 <http://www.pcisig.com/specifications/>

129 PCI-SIG, *PCI Express® Base Specification Revision 3.1a*, December 7, 2015,
130 <http://www.pcisig.com/specifications/>

131 PCI-SIG, *PCI Express® Base Specification Revision 4.0*, October 5, 2017,
132 <http://www.pcisig.com/specifications/>

133 PCI-SIG, *PCI Express® Base Specification Revision 5.0*, May 28, 2019,
134 <http://www.pcisig.com/specifications/>

135 **3 Terms and definitions**

136 In this document, some terms have a specific meaning beyond the normal English meaning. Those terms
137 are defined in this clause.

138 The terms "shall" ("required"), "shall not", "should" ("recommended"), "should not" ("not recommended"),
139 "may", "need not" ("not required"), "can" and "cannot" in this document are to be interpreted as described
140 in [ISO/IEC Directives, Part 2](#), Clause 7. The terms in parentheses are alternatives for the preceding term,
141 for use in exceptional cases when the preceding term cannot be used for linguistic reasons. Note that
142 [ISO/IEC Directives, Part 2](#), Clause 7 specifies additional alternatives. Occurrences of such additional
143 alternatives shall be interpreted in their normal English meaning.

144 The terms "clause", "subclause", "paragraph", and "annex" in this document are to be interpreted as
145 described in [ISO/IEC Directives, Part 2](#), Clause 6.

146 The terms "normative" and "informative" in this document are to be interpreted as described in [ISO/IEC](#)
147 [Directives, Part 2](#), Clause 3. In this document, clauses, subclauses, or annexes labeled "(informative)" do
148 not contain normative content. Notes and examples are always informative elements.

149 Refer to [DSP0236](#) for terms and definitions that are used across the MCTP specifications. For the
150 purposes of this document, the following additional terms and definitions apply.

151 **3.1**

152 **MCTP PCIe Endpoint**

153 a PCIe endpoint on which MCTP PCIe VDM communication is supported

154 **4 Symbols and abbreviated terms**

155 Refer to [DSP0236](#) for symbols and abbreviated terms that are used across the MCTP specifications. The
156 following symbols and abbreviations are used in this document.

157 **4.1**

158 **PCIe®**

159 PCI Express™

160 **4.2**

161 **VDM**

162 Vendor Defined Message

163 **4.3**

164 **CXL™**

165 Compute Express Link™

166 **5 Conventions**

167 The conventions described in the following clauses apply to this specification.

168 **5.1 Reserved and unassigned values**

169 Unless otherwise specified, any reserved, unspecified, or unassigned values in enumerations or other
170 numeric ranges are reserved for future definition by DMTF.

171 Unless otherwise specified, numeric or bit fields that are designated as reserved shall be written as 0
172 (zero) and ignored when read.

173 **5.2 Byte ordering**

174 Unless otherwise specified, byte ordering of multi-byte numeric fields or bit fields is "Big Endian" (that is,
175 the lower byte offset holds the most significant byte, and higher offsets hold lesser significant bytes).

176 **6 MCTP over PCI Express VDM transport**

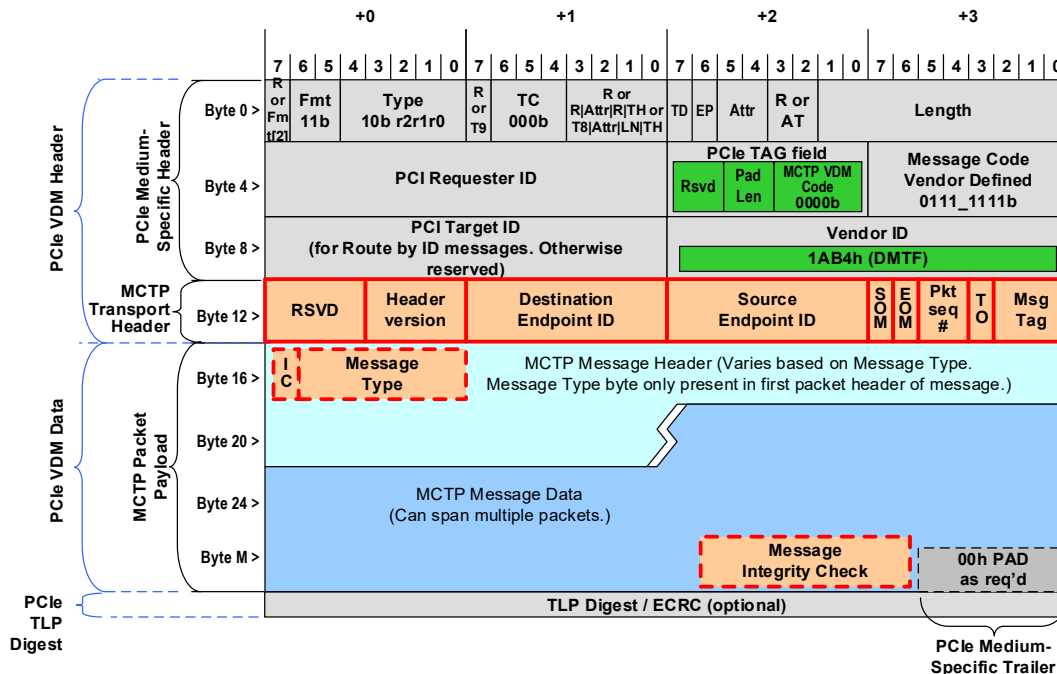
177 This document defines the medium-specific transport binding for transferring MCTP packets between
178 endpoints on PCI Express™ using PCIe Vendor Defined Messages (VDMs).

179 A MCTP over PCIe VDM compliant PCIe device shall support MCTP over PCIe VDM communications on
180 at least one PCIe Physical Function (PF) of the device. If a MCTP over PCIe VDM compliant PCI device
181 supports MCTP over PCIe VDM communications on more than one PCIe function, then MCTP over PCIe
182 VDM communication on each function shall be independent from MCTP over PCIe VDM communications
183 on other PCIe functions.

184 **6.1 Packet format**

185 The MCTP over PCI Express (PCIe) VDM transport binding transfers MCTP messages using PCIe Type
186 1 VDMs with data. MCTP messages use the MCTP VDM code value (0000b) that uniquely differentiates
187 MCTP messages from other DMTF VDMs.

188 Figure 1 shows the encapsulation of MCTP packet fields within a PCIe VDM.



189

190

Figure 1 – MCTP over PCI Express Vendor Defined Message (VDM) packet format

191 The fields labeled “PCIe Medium-Specific Header” and “PCIe Medium-Specific Trailer” are specific to
 192 carrying MCTP packets using PCIe VDMs. The fields labeled “MCTP Transport Header” and “MCTP
 193 Packet Payload” are common fields for all MCTP packets and messages and are specified in [MCTP](#). This
 194 document defines the location of those fields when they are carried in a PCIe VDM. The PCIe
 195 specification allows the last four bytes of the PCIE VDM header to be vendor defined. The MCTP over
 196 PCIe VDM transport binding specification uses these bytes for MCTP Transport header fields under the
 197 DMTF Vendor ID. This document also specifies the *medium-specific* use of the MCTP “Hdr Version” field.

198 Table 1 lists the PCIe medium-specific fields and field values that shall be used in MCTP over PCIe VDM
 199 communications. When not specified, field values shall be set according to PCIe specifications. Note that
 200 the presence of TLP prefixes in MCTP over PCIe VDM packets is implementation dependent and outside
 201 the scope of this specification.

202

Table 1 – PCI Express medium-specific MCTP packet fields

Field	Description
R or Fmt[2]	PCIe 1.1/2.0: PCIe reserved bit (1 bit). PCIe 2.1, 3.X, 4.X, 5.X: Fmt[2]. Set to 0b.
Fmt	Format (2 bits). Set to 11b to indicate 4 dword header with data.
Type	Type and Routing (5 bits). [4:3] Set to 10b to indicate a message [2:0] PCI message routing (r2r1r0) 000b : Route to Root Complex 010b : Route by ID 011b : Broadcast from Root Complex Other routing fields values are not supported for MCTP.

Field	Description
R or T9	PCIe 1.1/2.0/2.1/3.X: PCIe reserved bit (1 bit). Refer to the PCI Express™ specification (PCIe). Set to 0b. PCIe 4.X/5.X: T9 (1bit). Refer to the PCI Express™ specification (PCIe) Gen 4. Set to 0b.
TC	Traffic Class (3 bits). Set to 000b for MCTP over PCIe VDM.
R or R Attr R TH or T8 Attr LN TH	PCIe 1.1/2.0: PCIe reserved bits (4 bits). Set to 0000b PCIe 2.1/3.X: PCIe reserved bit (1 bit), Attr[2] (1 bit) – Set to 0b, reserved bit (1bit), and TH (1bit) – Set to 0b. PCIe 4.X/5.X: T8 bit (1 bit) – Set to 0b, Attr[2] (1 bit) – Set to 0b, LN (1bit) – Set to 0b, and TH (1bit) – Set to 0b
TD	TLP Digest (1 bit). 1b indicates the presence of the TLP Digest field at the end of the PCIe TLP (transaction layer packet). The TD bit should be set in accordance with the devices overall support for the TLP Digest capability, and whether that capability is enabled. See description of the TLP Digest / ECRC field, below, for additional information. Note that earlier versions of this specification erroneously required this bit to be set to 0b, which would have required devices to not support the TLP Digest capability.
EP	Error Poisoned (1 bit).
Attr[1:0]	Attributes (2 bits). Set to 00b or 01b for all MCTP over PCIe VDM.
R or AT	PCIe 1.1: PCIe reserved bits (2 bits). PCIe 2.0/2.1/3.X/4.X/5.X: Address Type (AT) field. Set to 00b.
Length	Length: Length of the PCIe VDM Data in dwords. Implementations shall support the baseline transmission unit defined in the MCTP Base Specification . For example, supporting a baseline transmission unit of 64 bytes requires supporting PCIe VDM data up to 16 dwords. An implementation may optionally support larger transfer unit sizes.
PCI Requester ID	Bus/device/function or bus/function number of the managed endpoint sending the message.
Pad Len	Pad Length (2-bits). 1-based count (0 to 3) of the number of 0x00 pad bytes that have been added to the end of the packet to make the packet dword aligned with respect to. PCIe . Because only packets with the EOM bit set to 1b are allowed to be less than the transfer unit size, packets that have the EOM bit set to 0b will already be dword aligned and will thus not require any pad bytes and will have a pad length of 00b.
MCTP VDM Code	Value that uniquely differentiates MCTP messages from other DMTF VDMs. Set to 0000b for this transport mapping as defined in this specification.
Message Code	(8 bits). Set to 0111_1111b to indicate a Type 1 VDM.
PCI Target ID	(16 bits). For Route By ID messages, this is the bus/device/function number or bus/function number that is the physical address of the target endpoint. This field is ignored for Broadcast and for Route to Root Complex messages.
Vendor ID	(16 bits). Set to 6836 (0x1AB4) for DMTF VDMs. The most significant byte is in byte 10, the least significant byte is byte 11.
RSVD	MCTP reserved (4 bits). Set these bits to 0 when generating a message. Ignore them on incoming messages.
Hdr Version	MCTP version (4 bits) 0001b : For MCTP devices that conform to the MCTP Base Specification and this version of the PCIe VDM transport binding. All other settings: Reserved to support future packet header field expansion or header version.

Field	Description
00h PAD	Pad bytes. 0 to 3 bytes of 00h as required to fill out the overall PCIe VDM data to be an integral number of dwords. Because only packets with the EOM bit set to 1b are allowed to be less than the transfer unit size, packets that have the EOM bit set to 0b will already be dword aligned, and will thus not require any pad bytes and will have a pad length of 00b.
TLP Digest / ECRC	(32 bits). TLP Digest / ECRC (End-to-end CRC). This field is defined for all PCIe TLPs (Transaction Layer Packets). Device support for this field is optional. However, per PCIe v2.1/3.X/4.X/5.X : "If a device Function is enabled to generate ECRC, it must calculate and apply ECRC for all TLPs originated by the Function. If the device supports generating this field, it must support it for all TLPs." Additionally, per PCIe v2.1/3.X/4.X/5.X , if the ultimate PCI Express Receiver of the TLP does not support ECRC checking, the receiver must ignore the TLP Digest.

203 **6.2 Supported media**

204 This physical transport binding has been designed to work with the following media as defined in
 205 [DSP0239](#) and listed in Table 2. Use of this binding with other types of physical media is not covered by
 206 this specification. Refer to DSP0239 for all supported physical media by MCTP transport bindings.

207 An implementation that is compliant with this specification shall at least support one of the PCIe media
 208 listed in Table 2. Note that the CXL is built on the [PCI Express](#) (PCIe) physical and electrical interface.

209 **Table 2 – Supported media**

Physical Media Identifier	Description
0x08	PCIe 1.1 compatible
0x09	PCIe 2.0 compatible
0x0A	PCIe 2.1 compatible
0x0B	PCIe 3.X compatible
0x0C	PCIe 4.X compatible
0x0D	PCIe 5.X compatible, CXL 1.X/2.X compatible

210 **6.3 Physical address format for MCTP control messages**

211 The address format shown in Table 3 is used for MCTP control commands that require a physical
 212 address parameter to be returned for a bus that uses this transport binding with one of the supported
 213 media types listed in 6.2. This includes commands such as the Resolve Endpoint ID, Routing Information
 214 Update, and Get Routing Table Entries commands.

215 **Table 3 – Physical address format**

Format Size	Address Type	Layout and Description	
		byte 1	byte 2
2 bytes (BDF ID)	Bus Device Function (BDF)	byte 1	[7:0] – Bus number
		byte 2	[7:3] – Device number [2:0] – Function number
2 bytes (ARI ID)	Alternate Routing Identifier (ARI)	byte 1	[7:0] – Bus number
		byte 2	[7:0] – Function number

216 6.4 Message routing

217 Physical packet routing within a PCIe bus uses routing as defined by the PCIe specification. PCIe
218 physical routing/bridging is not the same as MCTP bridging. PCIe physical routing/bridging is generally
219 transparent to MCTP. There are no MCTP-defined functions for configuring or controlling the setup of a
220 PCIe bus. The following types of PCIe addressing are used with MCTP messages:

- 221 • **Route by ID**

222 All MCTP over PCIe VDM packets between endpoints that are not the bus owner shall use
223 Route by ID for message routing.

224 The MCTP bus owner shall use Route by ID for messages to individual MCTP endpoints.

225 MCTP endpoints are required to capture the PCIe requester ID and the MCTP source EID when
226 receiving an EID assignment MCTP control request message. This is because this request can
227 only be issued by the MCTP bus owner.

- 228 • **Route to root complex**

229 MCTP endpoints shall use this routing for the Discovery Notify request message to the MCTP
230 bus owner as part of the MCTP over PCIe VDM discovery process.

231 The MCTP endpoints shall use this routing for responding to the MCTP control request
232 messages that were sent using Broadcast from Root Complex.

233 Communication of MCTP PCIe VDM packets that are destined to MCTP bus owner using
234 routed to root complex is implementation specific and is outside the scope of this specification.

- 235 • **Broadcast from root complex**

236 The MCTP bus owner should use this routing for the Prepare for Endpoint Discovery and
237 Endpoint Discovery messages as part of the MCTP over PCIe VDM discovery process.

238

239 6.4.1 Routing peer transactions on a PCIe bus

240 Because the PCIe specification does not require peer-to-peer routing support in PCIe root complexes,
241 MCTP over PCIe VDM messages are not required to be routed to peer devices directly. When peer-to-
242 peer routing is not supported by a PCIe root complex, all MCTP over PCIe VDM messages between two
243 MCTP endpoints shall be routed to or through the MCTP bus owner as an MCTP bridge. If the PCIe root
244 complex, as the MCTP bus owner, supports peer-to-peer routing, it shall use direct physical addressing to
245 support routing between two MCTP endpoints on the PCIe bus.

246 6.4.2 Routing messages between PCIe and other buses

247 All MCTP messages that span between PCIe and other buses shall be sent through the MCTP bus
248 owner. The MCTP bus owner has the destination EID routing tables necessary to route messages
249 between the two bus segments.

250 If an endpoint is aware of multiple routes to a destination over multiple bus types, a higher level
251 algorithm/protocol above MCTP shall be used to determine which bus/route to use. Typically this decision
252 can be based on things like power state and MCTP discovery state.

253 6.5 Bus owner address

254 The MCTP PCIe VDM bus owner functionality shall be accessible through "Route-to-Root Complex"
255 addressing.

256 6.6 Bus address assignment for PCIe

257 PCIe bus addresses are assigned per the mechanisms specified in [PCIe](#).

258 6.7 Host dependencies

259 MCTP over PCIe VDM, when used in a typical “PC” computer system, has a dependency on the host
260 CPU, host software, power management states, link states, and reset. Some of these dependencies are
261 described as follows:

262 • **Reset**

263 Assertion of “Fundamental Reset” on the bus causes both the host functionality as well as the
264 MCTP PCIe VDM communication on an MCTP PCIe endpoint to be reset. From the assertion
265 “Fundamental Reset” until the PCIe fabric has been configured and enumerated, no “MCTP
266 over PCI Express” messages can be sent.

267 Similarly, if MCTP PCIe VDM communication is supported on a function, a function level reset
268 (FLR) could reset MCTP PCIe VDM endpoint as well as MCTP PCIe VDM communication on
269 that function.

270 • **Configuration and enumeration**

271 Following the de-assertion “Fundamental Reset”, the software running on the host CPU
272 configures and enumerates the PCIe fabric. Failure of the host CPU or boot software to properly
273 configure and enumerate the PCIe fabric prevents it from being used for MCTP over PCIe VDM
274 messaging.

275 • **Power management states**

276 The host (as defined in the context of the [PCI Express™ specification](#)) controls PCIe bus power
277 management. The host may power down PCIe devices and links, or place them in sleep states,
278 independent of management controllers, which may cause MCTP PCIe VDM communication to
279 be unavailable. Depending on the device usage in the system, a PCIe device may retain or lose
280 states such as EID, “discovered” state, and routing information (if the device is a bridge). A
281 PCIe device that loses MCTP PCIe VDM communication state needs to be reinitialized and/or
282 rediscovered after it returns to a power state that supports MCTP over PCIe VDM
283 communication.

284 • **Link states**

285 The PCIe link states affect MCTP over PCIe VDM communications. MCTP over PCIe VDM
286 communication can be performed only when the PCIe link is in a state that allows VDM
287 communications. The mechanisms for PCIe link state transitions are outside the scope of this
288 specification.

289 • **PCIe Root Complex**

290 PCIe Root Complex (RC) is responsible for communicating route-to-root complex MCTP over
291 PCIe VDM discovery messages to the MCTP bus owner.

292 6.8 Discovery Notify message use for PCIe

293 An MCTP control Discovery Notify message shall be sent from a PCIe endpoint to the MCTP bus owner
294 whenever the physical address for the device changes (that is, the endpoint receives a Type 0
295 configuration write request and the bus number is different than the currently stored bus number). This
296 occurs on the first Type 0 configuration write following a PCIe bus reset during initial enumeration, or
297 during re-enumeration where the bus number has changed (for example, because of a hot plug event,
298 bus reset, and so on).

299 Endpoints use the Discovery Notify command to inform the MCTP bus owner that it needs to update the
300 endpoint's ID. The Discovery Notify command shall be sent with the PCIe message routing set to 000b
301 (Route-to-Root Complex), the Destination Endpoint ID for the Discovery Notify message shall be set to
302 the Null Destination EID. The Source Endpoint ID field shall be set to the Null Source EID if the device
303 has not yet been assigned an EID; otherwise, it shall contain the assigned EID value.

304 **6.9 MCTP over PCIe endpoint discovery**

305 This clause describes the steps used to support discovering MCTP endpoints on PCIe.

306 **6.9.1 Discovered flag**

307 Each endpoint (except the bus owner) on the PCIe bus maintains an internal flag called the *Discovered*
308 flag.

309 The flag is set to the *discovered* state when the Set Endpoint ID command is received.

310 The Prepare for Endpoint Discovery message causes each recipient endpoint on the PCIe bus to set their
311 respective Discovered flag to the *undiscovered* state. For the Prepare for Endpoint Discovery request
312 message, the routing in the physical transport header should be set to 011b (Broadcast from Root
313 Complex).

314 An endpoint also sets the flag to the *undiscovered* state at the following times:

- 315 • Whenever the PCI bus/device/function or bus/function number associated with the endpoint is
316 initially assigned or is changed to a different value.
- 317 • Whenever an endpoint first appears on the bus and requires an EID assignment. A device shall
318 have been enumerated on PCI and have a bus/device/function or bus/function number before it
319 can do this.
- 320 • During operation if an endpoint enters a state that causes it to lose its EID assignment.
- 321 • For endpoints that have already received an EID assignment but are in any temporary state
322 where the endpoint was unable to respond to MCTP control requests for more than $T_{RECLAIM}$
323 seconds.

324 Only endpoints that have their Discovered flag set to *undiscovered* shall respond to the Endpoint
325 Discovery message. Endpoints that have the flag set to *discovered* shall not respond to the Endpoint
326 Discovery message.

327 For PCIe endpoints, an Endpoint Discovery broadcast request message can be sent by the MCTP bus
328 owner to discover all MCTP-capable devices. MCTP-capable endpoints respond with an Endpoint
329 Discovery response message.

330 An MCTP-capable endpoint shall respond to broadcast MCTP control request messages only if a PCI bus
331 number is assigned to the associated PCIe function; otherwise, the endpoint should silently discard such
332 MCTP messages.

333 **6.9.2 PCIe endpoint announcement**

334 One or more endpoints may announce their presence and their need for an EID assignment by
335 autonomously sending a Discovery Notify message to the bus owner. This would typically trigger the
336 MCTP bus owner to perform the PCIe endpoint discovery/enumeration processes described in the
337 following subclauses.

338 6.9.3 Full endpoint Discovery/Enumeration

339 The following process is typically used when the MCTP bus owner wishes to discover and enumerate all
340 MCTP endpoints on the PCIe bus.

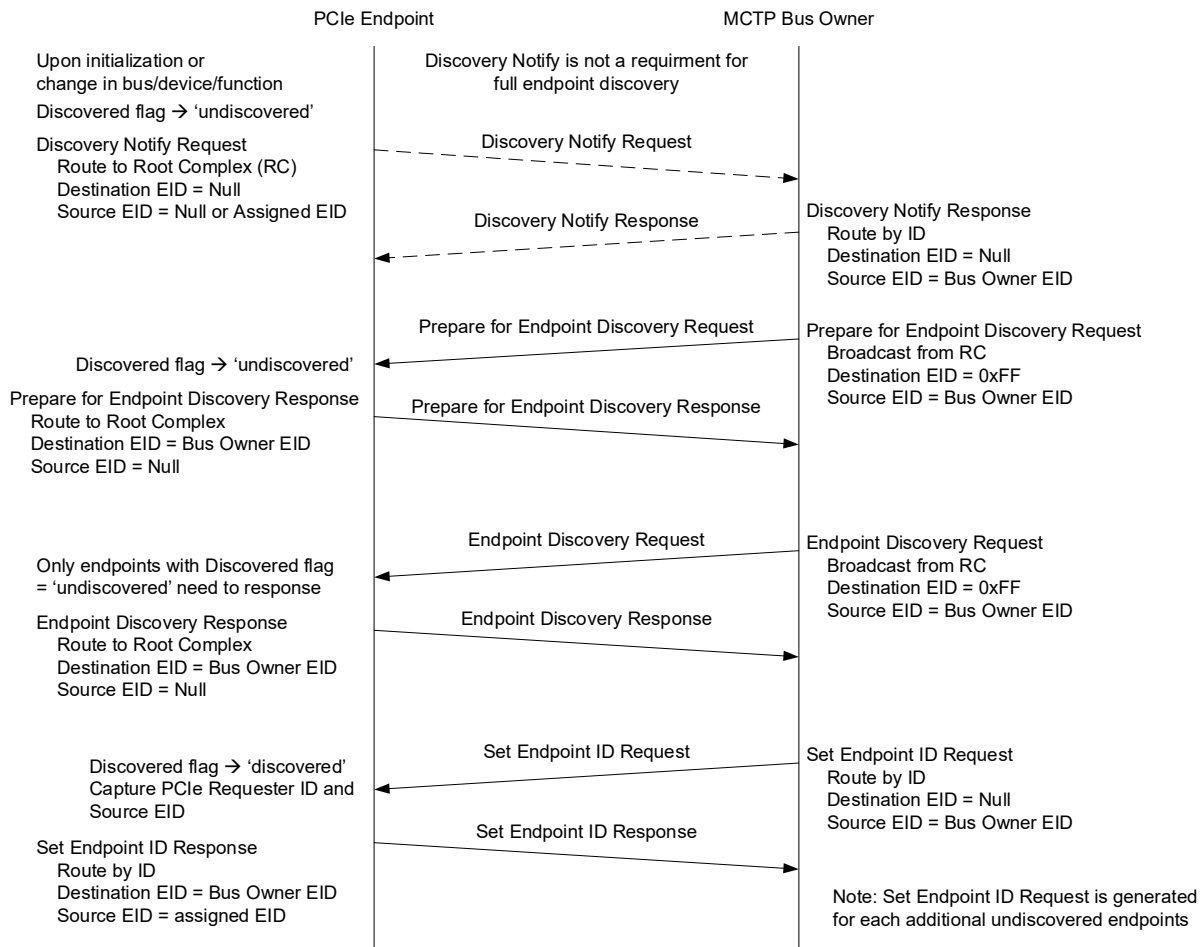
- 341 1) The MCTP bus owner issues a broadcast Prepare for Endpoint Discovery message. This
342 message causes each discoverable endpoint on the bus to set its PCIe endpoint Discovered
343 flag to undiscovered. Depending on the number of endpoints and the buffer space available in
344 the MCTP bus owner, the MCTP bus owner may not receive all of the response messages. The
345 discovery process does not require the MCTP bus owner to receive all the response messages
346 to the Prepare for Endpoint Discovery request. Because the MCTP bus owner cannot determine
347 that all endpoints have received the Prepare for Endpoint Discovery request, it is recommended
348 that Prepare for Endpoint Discovery request is retried MN1 times to help ensure that all
349 endpoints have received the request. The MCTP bus owner is not required to wait for MT2 time
350 interval between the retries.
- 351 2) The MCTP bus owner should wait for MT2 time interval to help ensure that all endpoints that
352 received the Prepare for Endpoint Discovery request have processed the request.
- 353 3) The MCTP bus owner issues a broadcast Endpoint Discovery request message. All MCTP-
354 capable devices that have their Discovered flag set to undiscovered will respond with an
355 Endpoint Discovery response message.
- 356 4) Depending on the number of endpoints and the buffer space available in the MCTP bus owner,
357 the MCTP bus owner receives some or all of these response messages. For each response
358 message received from an undiscovered MCTP-capable device PCIe bus/device/function or
359 bus/function number, the MCTP bus owner issues a Set Endpoint ID command to the physical
360 address for the endpoint. This causes the endpoint to set its Discovered flag to *discovered*.
361 From this point, the endpoint shall not respond to the Endpoint Discovery command until
362 another Prepare for Endpoint Discovery command is received or some other condition causes
363 the Discovered flag to be set back to *undiscovered*.
- 364 5) If the MCTP bus owner received any responses to the Endpoint Discovery request issued in
365 Step 3, then it shall repeat steps 3 and 4 until it no longer gets any responses to the Endpoint
366 Discovery request. In this case, then the MCTP bus owner is allowed to send the next Endpoint
367 Discovery request without waiting for MT2 time interval. If no responses were received by the
368 MCTP bus owner to the Endpoint Discovery request within the MT2 time interval, then the
369 discovery process is completed.

370 After the initial endpoint enumeration, it is recommended that the MCTP bus owner maintains a list of the
371 unique IDs for the endpoints it has discovered and reassigns the same IDs to those endpoints if a
372 bus/device/function or bus/function number changes during system operation.

373 An MCTP-capable endpoint may respond to *Route by ID* Prepare for Endpoint Discovery and Endpoint
374 Discovery request messages.

375 Figure 2 provides an example flow of operations for full endpoint discovery.

Full PCIe MCTP Endpoint Discovery



376

377

Figure 2 – Flow of operations for full MCTP Discovery over PCIe

378 **6.9.4 Partial endpoint Discovery/Enumeration**

379 This process is used when the MCTP bus owner wishes to discover endpoints that may have been added
 380 to the bus after a full enumeration has been done. This situation can occur if a device has its
 381 bus/device/function or bus/function number change after the full enumeration has been done, or when a
 382 hot-plug device is added to the system, or if a device that is already present in the system — but was in a
 383 disabled or powered-down state — comes on-line.

384 The partial discovery process is the same as the full discovery process except that the MCTP bus owner
 385 skips the step of broadcasting a Prepare for Endpoint Discovery command in order to avoid clearing the
 386 Discovered flags of already discovered endpoints.

387 The partial discovery process may be initiated when a device that is added or enabled for MCTP sends a
 388 Discovery Notify message to the MCTP bus owner. The MCTP bus owner may also elect to periodically
 389 issue a broadcast Endpoint Discovery message to test for whether any undiscovered endpoints have
 390 been missed. The Discovery Notify message provides the MCTP bus owner with the bus/device/function
 391 or bus/function number of the MCTP PCIe endpoint. The MCTP bus owner can then send a directed
 392 Endpoint Discovery message to the endpoint to confirm that the device has not been discovered. The

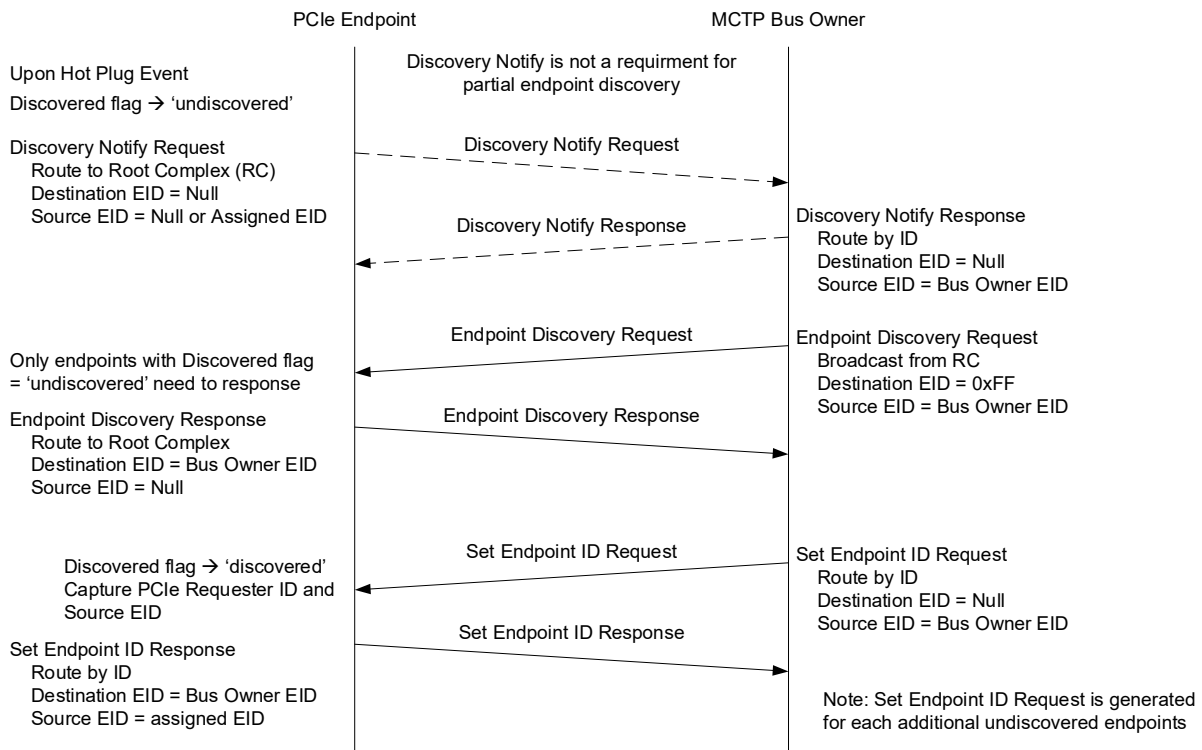
393 MCTP bus owner then issues a Set Endpoint ID command to the physical address for the endpoint which
 394 causes the endpoint to set its Discovered flag to *discovered*.

395 It is recommended that the MCTP bus owner maintains a list of the unique MCTP EIDs for the endpoints
 396 it has discovered and reassigns the same MCTP EIDs to those endpoints if a bus/device/function or
 397 bus/function number changes during system operation.

398 An MCTP-capable endpoint may respond to *Route by ID* Endpoint Discovery request messages.

399 Figure 3 provides an example flow of operations for partial endpoint discovery.

Partial PCIe MCTP Endpoint Discovery



400

401

Figure 3 – Flow of operations for Partial Endpoint Discovery

402 6.9.5 Endpoint re-enumeration

403 If the bus implementation includes hot-plug devices, the bus owner shall perform a full or partial endpoint
 404 discovery any time the MCTP bus owner goes into a temporary state where the MCTP bus owner can
 405 miss receiving a Discovery Notify message (for example, if the bus owner device is reset or receives a
 406 firmware update). Whether a full or partial endpoint discovery is required is dependent on how much
 407 information the MCTP bus owner retains from prior enumerations.

408 **6.10 MCTP messages timing requirements**

409 Table 4 lists MCTP-specific timing requirements for MCTP Control messages and operation on the PCIe
 410 VDM medium. All MCTP Control Messages over PCIe VDM shall comply to the timing specification listed
 411 in Table 4.

412 **Table 4 – Timing specifications for MCTP Control messages on PCIe VDM**

Timing Specification	Symbol	Min	Max	Description
Endpoint ID reclaim	T _{RECLAIM}	–	5 sec	<p>Maximum interval that an endpoint is allowed to be non-responsive to MCTP control messages before its EID may be reclaimed by the bus owner.</p> <p>A bus owner shall wait at least for this interval before an EID of the non-responsive endpoint is reclaimed.</p>
Number of request retries	MN1	2	See Description column	Total of three tries, minimum: the original try plus two retries. The maximum number of retries for a given request is limited by the requirement that all retries shall occur within MT4, max of the initial request.
Request-to-response time	MT1	–	120 ms	This interval is measured at the responder from the end of the reception of an MCTP control request to the beginning of the transmission of the corresponding MCTP control response. This requirement is tested under the condition where the responder can successfully transmit the response on the first try.
Time-out waiting for a response	MT2	MT1 max ^[1] + 6 ms	MT4, min ^[1]	<p>This interval at the requester sets the minimum amount of time that a requester should wait before retrying a MCTP control request. This interval is measured at the requester from the end of the successful transmission of the MCTP control request to the beginning of the reception of the corresponding MCTP control response.</p> <p>NOTE: This specification does not preclude an implementation from adjusting the minimum time-out waiting for a response to a smaller number than MT2 based on the measured response times from responders. The mechanism for doing so is outside the scope of this specification.</p>
Instance ID expiration interval	MT4	5 sec ^[2]	6 sec	Interval after which the instance ID for a given response will expire and become reusable if a response has not been received for the request. This is also the maximum time that a responder tracks an instance ID for a given request from a given requester.

Timing Specification	Symbol	Min	Max	Description
<p>NOTE 1: Unless otherwise specified, this timing applies to the mandatory and optional MCTP commands.</p> <p>NOTE 2: If a requester is reset, it may produce the same sequence number for a request as one that was previously issued. To guard against this, it is recommended that sequence number expiration be implemented. Any request from a given requester that is received more than MT4 seconds after a previous, matching request should be treated as a new request, not a retry.</p>				

413
414
415
416

ANNEX A (informative)

Notations and conventions

417 Notations

418 Examples of notations used in this document are as follows: list into text needed

- 419 • 2:N In field descriptions, this will typically be used to represent a range of byte offsets
420 starting from byte two and continuing to and including byte N. The lowest offset is on
421 the left, the highest is on the right.
- 422 • (6) Parentheses around a single number can be used in message field descriptions to
423 indicate a byte field that may be present or absent.
- 424 • (3:6) Parentheses around a field consisting of a range of bytes indicates the entire range
425 may be present or absent. The lowest offset is on the left, the highest is on the right.
- 426 • [PCIe](#) Underlined, blue text is typically used to indicate a reference to a document or
427 specification called out in Clause 2, "Normative References" or to items hyperlinked
428 within the document.
- 429 • rsvd Abbreviation for Reserved. Case insensitive.
- 430 • [4] Square brackets around a number are typically used to indicate a bit offset. Bit offsets
431 are given as 0-based values (that is, the least significant bit [LSb] offset = 0).
- 432 • [7:5] A range of bit offsets. The most significant bit is on the left, the least significant bit is
433 on the right.
- 434 • 1b The lower case "b" following a number consisting of 0s and 1s is used to indicate the
435 number is being given in binary format.
- 436 • 0x12A A leading "0x" is used to indicate a number given in hexadecimal format.

437
 438
 439
 440
 441

ANNEX B (informative)

Change log

Version	Date	Description
1.0.0	2009-07-28	
1.0.1	2009-10-30	Created erratum to clarify Length field definition of PCIe VDM header for MCTP PCIe VDM transport binding, modify introduction section, and clean up references section.
1.0.2	2014-12-07	Clarifications to TD bit usage. Added TLP Digest/ECRC to packet figure and to field descriptions table.
1.1.0	2018-10-24	Added support for PCIe Gen 3, PCIe Gen 4, and ARI. Fixed Figure 1 to cover PCIe 1.0/2.0/2.1/3.X/4.0. Clarified MCTP over PCIe VDM compliant management device requirements. Clarified Endpoint ID reclaim definition. Clarified MCTP bus owner requirements in the specification. Eliminated PCIe bus owner term and replaced it with PCIe root complex where applicable.
1.2.0	2021-03-02	Added support for PCIe Gen 5.X, CXL 1.X, and CXL 2.X.
1.2.1	2024-01-22	Added a clarification for bus number assignment before responding to broadcast MCTP control messages. Added clarifications that an MCTP-capable endpoint may respond to Route by ID discovery request messages.

442