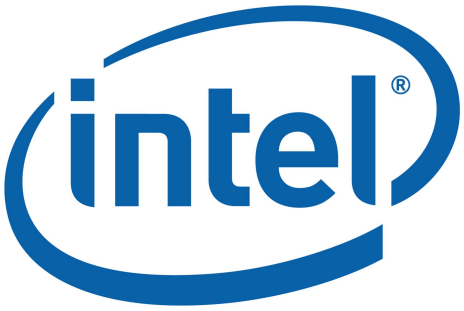


Intel® Ethernet Controller XL710

Specification Update

Networking Division (ND)

July 2014

Revision 2.0

LEGAL

By using this document, in addition to any agreements you have with Intel, you accept the terms set forth below.

You may not use or facilitate the use of this document in connection with any infringement or other legal analysis concerning Intel products described herein. You agree to grant Intel a non-exclusive, royalty-free license to any patent claim thereafter drafted which includes subject matter disclosed herein.

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

A “Mission Critical Application” is any application in which failure of the Intel Product could result, directly or indirectly, in personal injury or death. SHOULD YOU PURCHASE OR USE INTEL'S PRODUCTS FOR ANY SUCH MISSION CRITICAL APPLICATION, YOU SHALL INDEMNIFY AND HOLD INTEL AND ITS SUBSIDIARIES, SUBCONTRACTORS AND AFFILIATES, AND THE DIRECTORS, OFFICERS, AND EMPLOYEES OF EACH, HARMLESS AGAINST ALL CLAIMS COSTS, DAMAGES, AND EXPENSES AND REASONABLE ATTORNEYS' FEES ARISING OUT OF, DIRECTLY OR INDIRECTLY, ANY CLAIM OF PRODUCT LIABILITY, PERSONAL INJURY, OR DEATH ARISING IN ANY WAY OUT OF SUCH MISSION CRITICAL APPLICATION, WHETHER OR NOT INTEL OR ITS SUBCONTRACTOR WAS NEGLIGENT IN THE DESIGN, MANUFACTURE, OR WARNING OF THE INTEL PRODUCT OR ANY OF ITS PARTS.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked “reserved” or “undefined”. Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order.

Copies of documents which have an order number and are referenced in this document, or other Intel literature, may be obtained by calling 1-800-548-4725, or go to: http://www.intel.com/design/literature.htm.

Intel and Intel logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

* Other names and brands may be claimed as the property of others.

Revision History

Revision	Date	Comments
2.0[1]	July 30, 2014	Initial release (Intel public).

1. There were no previous versions of this document released.

NOTE: ***This page intentionally left blank.***

1.0 Introduction

This document applies to the Intel® Ethernet Controller XL710 (XL710).

This document is an update to a published specification, the *Intel® Ethernet Controller XL710 Datasheet*. It is intended for use by system manufacturers and software developers. All product documents are subject to frequent revision and new order numbers may apply. New documents may be added. Be sure you have the latest information before finalizing your design.

References to PCIe Express* (PCIe*) in this document refer to PCIe v3.0 (2.5GT/s, 5GT/s, and 8GT/s).

1.1 Product Code and Device Identification

Product Code: FTXL710

The following tables and drawings describe the various identifying markings on each device package:

Table 1-1 Markings

Device	Stepping	Top Marking	S-Specification[1]	Description
XL710	B0	FTXL710-AM2	S R1ZK	Ethernet controller (2x40/1x40/4x10/2x20/2x10/1x10)
			S R1ZL	
XL710	B0	FTXL710-AM1	S R1ZM	Ethernet controller (1x40/4x10/2x10/1x10)
			S R1ZN	
XL710	B0	FTX710-AM2	S R1ZP	Ethernet controller (2x10/1x10)
			S R1ZQ	

1. For Tray, Tape, Reel data, see Table 1-3.

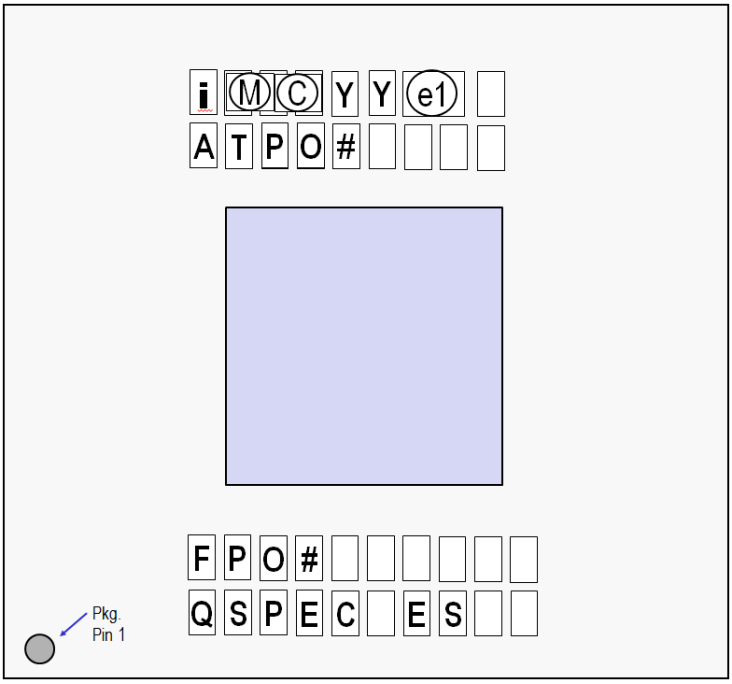
Table 1-2 Device ID

Device ID Code	Device ID	Vendor ID	Revision ID
XL710 2x40 Backplane (KR4)	0x1580	0x8086	0x01
XL710 4x10 Backplane (KR)	0x1581	0x8086	0x01
XL710 2x40 QSFP+ (XLPPI/CR4)	0x1583	0x8086	0x01
XL710 1x40 QSFP+ (XLPPI/CR4)	0x1584	0x8086	0x01
XL710 4x10 SFP+ (SFI)	0x1572	0x8086	0x01
XL710 Blank Flash (default)	0x154B	0x8086	0x01

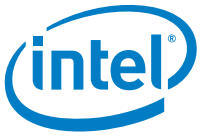
Table 1-3 MM Numbers

Product	Tray MM#	Tape and Reel MM#	Reserved
FTXL710-AM2	936550	936549	
FTXL710-AM1	936552	936551	
FTX710-AM2	936554	936553	

1.2 Marking Diagrams



- LINE1: I, Maskwork, Copyright, YY, Pb-free
- LINE2: ATPO#
- LINE3: FPO#
- LINE4: Q-spec#, ES



1.3 Nomenclature Used in This Document

This document uses specific terms, codes, and abbreviations to describe changes, errata, sightings and/or clarifications that apply to silicon/steppings. See Table 1-4 for a description.

Table 1-4 Nomenclature

Name	Description
Specification Clarifications	Greater detail or further highlights concerning a specification's impact to a complex design situation. These clarifications will be incorporated in the next release of the specifications.
Specification Changes	Modifications to the current published specifications. These changes will be incorporated in the next release of the specifications.
Errata	Design defects or errors. Errata may cause device behavior to deviate from published specifications. Hardware and software designed to be used with any given stepping must assume that all errata documented for that stepping are present on all devices.
Software Clarifications	Applies to Intel drivers, EEPROM loads.
Documentation Changes	Typos, errors, or omissions from the current published specifications. These changes will be incorporated in the next release of the specifications.
A0, B0, etc.	Stepping to which the status applies.
Doc	Document change or update that will be implemented.
Fix	This erratum is intended to be fixed in a future stepping of the component.
Fixed	This erratum has been previously fixed.
NoFix	There are no plans to fix this erratum.
Eval	Plans to fix this erratum are under evaluation.

2.0 Hardware Clarifications, Changes, Updates and Errata

See Section 1.3 for an explanation of terms, codes, and abbreviations.

Table 2-1 Summary of Specification Clarifications

Specification Clarification	Status
None	N/A

Table 2-2 Summary of Specification Changes

Specification Change	Status
1. Ingress Mirroring Cannot Be Changed on the Fly	N/A
2. RSS Field Selection is Globally Defined	N/A

Table 2-3 Summary of Documentation Updates

Specification Change	Status
None	N/A

Table 2-4 Summary of Errata; Errata Include Steppings

Erratum	Status
1. TX Performance Degradation for Small Cloud Packets	B0=Yes; NoFix
2. PCIe Subsystem ID Incorrectly Reported for All PCI Functions Except Function 0	B0=Yes; NoFix
3. Illegal Byte Error Statistical Counter Inaccuracy	B0=Yes; NoFix
4. Receive Performance Degradation with Specific Cloud Header	B0=Yes; NoFix
5. MCTP Discovery Error when Replacing Active PF	B0=Yes; NoFix
6. RX Queue Disable is Reported Done Before It is Disabled	B0=Yes; NoFix
7. TX Descriptor Might be Read Twice	B0=Yes; NoFix

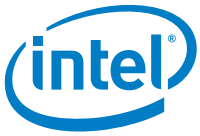


Table 2-4 Summary of Errata; Errata Include Steppings (Continued)

Erratum	Status
8. Immediate Interrupts are Delayed in Very Loaded System	B0=Yes; NoFix
9. ECRC Bits are Not RO when ECRC is Disabled	B0=Yes; NoFix
10. NC-SI I/Os Output Rise Slew Rate is Higher Than Specification	B0=Yes; NoFix
11. TC Strict Priority Does Not Work as Expected	B0=Yes; NoFix
12. Management-Only Packets Cannot Be Ignored for Wake-Up	B0=Yes; NoFix
13. Common Clock Configuration Bit Specification Compliance	B0=Yes; NoFix
14. Low Latency TC Might Be Momentarily Starved	B0=Yes; NoFix
15. Round Robin (RR) Bandwidth Distribution is Traffic Dependent	B0=Yes; NoFix
16. L2 Tag Stripped into the Wrong RX Descriptor Field	B0=Yes; NoFix
17. Internal VLAN is Not Reflected in RX Descriptor	B0=Yes; NoFix
18. Transmit Queue Group with Single Queue Enabled Performance	B0=Yes; NoFix
19. Switching Table Full Might Reduce Performance	B0=Yes; NoFix
20. Incorrect Optic Module Presence Reported to BMC	B0=Yes; NoFix
21. Set Binding Command is Not Functional for IPv4	B0=Yes; NoFix
22. Get PHY Abilities Does Not Return 20GBASE-KR2	B0=Yes; NoFix
23. “Multi-speed module timeout” NVM Parameter Has No Effect	B0=Yes; NoFix
24. Cloud Traffic Over VEB is Transmitted to LAN	B0=Yes; NoFix
25. VLAN Prune is Not Functional	B0=Yes; NoFix
26. VLAN Insertion Limitation with Specific Tunneling Packet	B0=Yes; NoFix

2.1 Specification Clarifications

None.

2.2 Specification Changes

1. Ingress Mirroring Cannot Be Changed on the Fly

Changing of Ingress Mirroring configuration during traffic might cause a bad configuration.

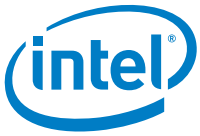
2. RSS Field Selection is Globally Defined

RSS field selection is done globally and cannot be configured differently per PF or VF.

- Functions that require the Hash (RSS) filters on IPv4 packets should set all IPv4 PCTYPEs in the PFQF_HENA / VFQF_HENA (PCTYPEs 31, 33…36).
- Functions that require the Hash filters on IPv6 packets should set all IPv6 PCTYPEs in the PFQF_HENA / VFQF_HENA (PCTYPEs 41, 43…46).
- Functions that require the Hash filters on FCoE packets should set all FCoE PCTYPEs in the PFQF_HENA / VFQF_HENA (PCTYPEs 48…50).

2.3 Documentation Updates

None.



2.4 Errata

1. TX Performance Degradation for Small Cloud Packets

Problem:

Happening for GRE+IPv6+TCP without payload. Degradation is expected to give 33 Gb/s instead of 34 Gb/s.

Implication:

This is seen if GRE+IPV6+TCP Packet is transmitted with no payload. This not typical packet format, and is not expected is most use cases.

Workaround:

None.

Status: B0=Yes; NoFix

2. PCIe Subsystem ID Incorrectly Reported for All PCI Functions Except Function 0

Problem:

All PCIe functions except Function 0 report a Subsystem ID of 0x0000 in the configuration space regardless of the value programmed in the NVM.

Implication

No functional impact to the device or drivers. However, this may impact the branding of the device if the Subsystem ID is used to select the device branding string.

Workaround:

None.

Status: B0=Yes; NoFix

3. Illegal Byte Error Statistical Counter Inaccuracy

Problem:

Short packets with bad symbols that arrive back-to-back might not be counted by GLPRT_ILLERRC.

Implication

GLPRT_ILLERRC is inaccurate.

Workaround:

None.

Status: B0=Yes; NoFix

4. Receive Performance Degradation with Specific Cloud Header

Problem:

A small performance degradation is expected when receiving back-to-back GRE+IPv6+TCP cloud frames with 128-byte Header and almost no payload.

Implication

Expected 33 Gb/s instead of 34 Gb/s.

Workaround:

None.

Status: B0=Yes; NoFix

5. MCTP Discovery Error when Replacing Active PF

Problem:

MCTP Discovery might respond with a wrong PF ID when BMC is replacing the active PF. Expected to be a rare scenario on specific machines.

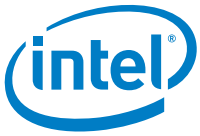
Implication

PF replacement may not work for MCTP.

Workaround:

None.

Status: B0=Yes; NoFix



6. RX Queue Disable is Reported Done Before It is Disabled

Problem:

RX Queue disable is reported done before it is disabled.

Implication

An RX Hang could result if the software re-enables the queue too early.

Workaround:

An RX Queue should be reused only after a minimum delay of 50 ms. This workaround is implemented in Intel SW Release 19.3.

Status: B0=Yes; NoFix

7. TX Descriptor Might be Read Twice

Problem:

A TX Descriptor might be read more than once in corner case conditions.

Implication

Negligible.

Workaround:

None.

Status: B0=Yes; NoFix

8. Immediate Interrupts are Delayed in Very Loaded System

Problem:

In a case where there are ten or more active queues in the system, and some of the queues are assigned with immediate interrupts, the interrupt delay may exceed the value specified in the Datasheet (”ITR and immediate interrupts jitter” table).

Implication

Low performance impact

Workaround:

None.

Status: B0=Yes; NoFix

9. ECRC Bits are Not RO when ECRC is Disabled

Problem:

ECRC bits in PCIe AER registers are writable even when ECRC is disabled.

Implication

Specification compliance issue.

Workaround:

None.

Status: B0=Yes; NoFix

10. NC-SI I/Os Output Rise Slew Rate is Higher Than Specification

Problem:

NC-SI I/Os output rise time might be as low as 400 ps, while the NC-SI Specification requires a minimum of 500 ps.

Implication

Specification compliance issue.

Workaround:

Place a 30 Ω resistor in serial to the pad.

Status: B0=Yes; NoFix

11. TC Strict Priority Does Not Work as Expected

Problem:

An UP might not get exclusive priority if PCIe bandwidth is insufficient (although gets higher priority).

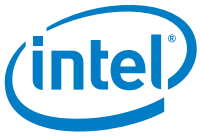
Implication

RX TC strict priority limitation.

Workaround:

None.

Status: B0=Yes; NoFix



12. Management-Only Packets Cannot Be Ignored for Wake-Up

Problem:

Due to a “NoTCO” wake-up capability malfunction, a wake event may be issued for packets that are expected to be routed to the BMC exclusively.

Implication:

Management-only packets cannot be ignored for wake-up.

Workaround:

None.

Status: B0=Yes; NoFix

13. Common Clock Configuration Bit Specification Compliance

Problem:

Common clock configuration bit should be writable for all PFs, but it is not always writable for a PF > 0.

Implication:

Specification compliance issue.

Workaround:

None.

Status: B0=Yes; NoFix

14. Low Latency TC Might Be Momentarily Starved

Problem:

Low Latency TC might be momentarily starved under TPB Non-Strict Priority (RR) policy when both Bulk and Low Latency traffic are pending.

Implication:

Low Latency TC impact.

Workaround:

None.

Status: B0=Yes; NoFix

15. Round Robin (RR) Bandwidth Distribution is Traffic Dependent

Problem:

Under RR RX Policy, RX bandwidth may be distributed unevenly among ports and TCs if PCIe bandwidth is smaller than incoming traffic, or traffic is a stream of small packets (smaller than 128 bytes).

Implication:

Uneven traffic distribution under RR.

Workaround:

Use Strict Priority policy instead of Round Robin.

Status: B0=Yes; NoFix

16. L2 Tag Stripped into the Wrong RX Descriptor Field

Problem:

If two L2 tags (for example: VLAN and S-TAG) are programmed to be extracted to the receive descriptor, and the receive packet includes only a single L2 tag, the extracted L2 tag is always posted in the L2TAG1 field regardless of the setting.

Implication:

In the above case, the software could not identify which of the two enabled L2 tags were extracted to the receive descriptor.

Workaround:

Software should not enable more than a single L2 tag to be extracted to the receive descriptor.

Status: B0=Yes; NoFix

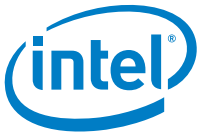
17. Internal VLAN is Not Reflected in RX Descriptor

Problem:

When SHOWIV field is set in the receive queue context, the internal VLAN is stripped, but it is not inserted in the RX descriptor as expected.

Implication:

Internal VLAN is not reflected in RX descriptor.



Workaround:

None.

Status: B0=Yes; NoFix

18. Transmit Queue Group with Single Queue Enabled Performance

Problem:

A transmit queue Group with single Queue enabled might have performance limitations when scheduling consecutive packets.

Implication:

TX Performance issue.

Workaround:

None.

Status: B0=Yes; NoFix

19. Switching Table Full Might Reduce Performance

Problem:

If switching table is relatively full, it might reduce performance in cases of packets smaller than 160 bytes.

Implication:

20% performance reduction for small packets.

Workaround:

Avoid switching table fullness.

Status: B0=Yes; NoFix

20. Incorrect Optic Module Presence Reported to BMC

Problem:

When optic module (SFP+/QSFP+) is removed, NC-SI Get Link Status reports Media Available.

Implication:

Removing the optical module is not reported in Get Link Status.

Workaround:

None.

Status: B0=Yes; NoFix

21. Set Binding Command is Not Functional for IPv4

Problem:

Set Binding command is not functional for IPv4.

Implication:

No manageability traffic after command.

Workaround:

None.

Status: B0=Yes; NoFix

22. Get PHY Abilities Does Not Return 20GBASE-KR2

Problem:

Get PHY Abilities command response does not set 20GBASE-KR2 PHY type bit (bit 30), even if set in NVM.

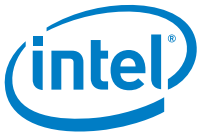
Implication:

Backplane setups with 20 GbE KR2 settings lack the report in Get PHY Abilities.

Workaround:

None.

Status: B0=Yes; NoFix



23. "Multi-speed module timeout" NVM Parameter Has No Effect

Problem:

During multi-speed module link establishment flow, the "Multi-speed module timeout" defined in NVM is not used.

Implication:

When using multi-speed module, TTL when moving from one speed to another speed, is at least 5 seconds.

Workaround:

Force a desired speed by disabling the other speed using the "set phy config" admin command.

Status: B0=Yes; NoFix

24. Cloud Traffic Over VEB is Transmitted to LAN

Problem:

Cloud traffic over VEB is transmitted to LAN.

Implication:

Cloud traffic over VEB is transmitted to LAN.

Workaround:

None.

Status: B0=Yes; NoFix

25. VLAN Prune is Not Functional

Problem:

Default action for VLAN Prune table is not set after "Add VLAN AQ" command.

Implication:

VLAN Prune is not functional.

Workaround:

Additional VLAN Prune configuration should be done by SW.

Status: B0=Yes; NoFix

26. VLAN Insertion Limitation with Specific Tunneling Packet

Problem:

VLAN cannot be inserted in the Inner L2 Header of the following tunneled packet:

L2 Header	IPV4 Header - No Checksum	UDP Header	Inner L2 Header	Inner IPv4 Header - No Checksum	Inner Payload or Encapsulation

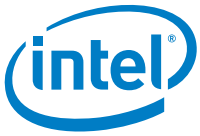
Implication:

VLAN Tag cannot be inserted.

Workaround:

None.

Status: B0=Yes; NoFix



3.0 Software Clarifications

Table 3-1 Summary of Software Clarifications

Software Clarification	Status
None	N/A

NOTE: ***This page intentionally left blank.***