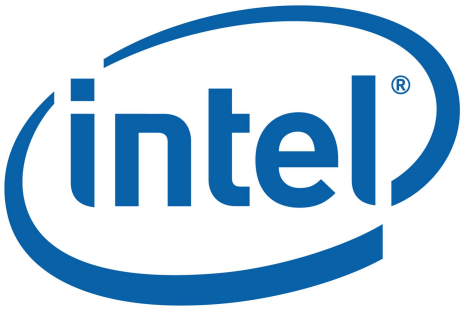


Intel® Ethernet Controller X540

Specification Update

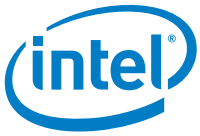
Ethernet Networking Division (ND)

January 2018

Revision 3.4
334566-003

Revision History

Revision	Date	Comments
3.4	January 2, 2018	**Miscellaneous Updates** • Updated AT2 and BT2 descriptions in Table 1-1, "Markings".
3.3	July 5, 2017	**Errata added or updated:** • 44. Internal PHY Slow Initialization (Added)
3.2	July 1, 2016	**Specification Clarifications added or updated:** • 12. FCOE_PARAM Field (Added) **Errata added or updated:** • Removed Erratum #21 • 22. LED Does Not Blink in Invert Mode (Updated) • 41. PCIe Advanced Error Reporting: First Error Pointer (Added) • 42. IPv4 Checksum Error Might be Reported for a Fragmented Packet (Added) • 43. LLC Packet without SNAP Header (Added) **Software Clarifications added or updated:** • 5. PF/VF Drivers Should Configure Registers That are Not Reset by VFLR (Updated)
3.1	July 14, 2015	**Miscellaneous Updates** • Changed Q-specification from SLJEJ to SLKTL.
3.0	February 23, 2015	**Specification Clarifications added or updated:** • 10. CRCERRS Statistic Counter (Added) • 11. Rx Statistics Counters Do Not Count Runt Frames or Fragments Smaller Than 12 Bytes (Added) **Specification Changes added or updated:** • 7. ETQF[19:16] are Reserved (Added)
2.9	March 25, 2014	**Specification Clarifications added or updated:** • 3. MCTP/DMTF Standard Compliance (Updated) • 9. SR-IOV Prefetchable Address Space (Added) **Specification Changes added or updated:** • 6. OS2BMC Requires Flow Control Rx Threshold Configuration (Updated) **Errata added or updated:** • 40. PCIe SR-IOV Reserved Bits are Writable (Added - Replaces Internal Erratum #49)
2.8	December 19, 2013	**Specification Changes added or updated:** • 6. OS2BMC Requires Flow Control Rx Threshold Configuration (Added) **Errata added or updated:** • 38. Clearing RXEN During VM-to-VM Loopback Traffic Might Cause an Rx Hang (Added) • 39. NC-SI Hardware Arbitration Issues (Added)
2.7	September 17, 2013	**Specification Clarifications added or updated:** • 3. MCTP/DMTF Standard Compliance (Updated) • 7. X540 Single Port NVM Image File (Added) • 8. VLAN Anti-Spoof Filter of an Untagged Packet (Added) **Documentation Updates added or updated:** • Removed Documentation Updates #1 and #2.



Revision	Date	Comments
2.6	May 6, 2013	**Specification Clarifications added or updated:** • 6. Selecting a Rx Pool Using VLAN Filters (Added) **Specification Changes added or updated:** • 5. The Flow Director FDIRErr(0) Bit in the Rx Descriptor is Valid Only if the FLM Bit is Set (Added) **Documentation Updates added or updated:** • Added Documentation Update #2. **Errata added or updated:** • Removed External Erratum #4 (Replaced by External Erratum #37). • 35. Flow Director: Collision Indication Can be Cleared by Adding a New Filter (Added) • 36. RXMEMWRAP Register Content is Inaccurate (Added) • 37. Flow Director Statistics Inaccuracy (Added)
2.5	February 12, 2013	**Errata added or updated:** • 33. NC-SI: Get NC-SI Pass-Through Statistics Response Might Contain Incorrect Packet Counts (Added) • 34. IPv4 Checksum Error Might be Reported for Multicast Frames Over 12 KB (Added) **Documentation Updates added or updated:** • Added Documentation Update #1. **Miscellaneous Updates** • Added X540-BT2 Product Code and Device Identification information (Table 1-1 through Table 1-3).
2.4	December 1, 2012	**Specification Clarifications added or updated:** • 5. IPv6 Extended Headers are Parsed by the X540 (Added) **Specification Changes added or updated:** • 4. RXMTRL.*UDPT* Initial Value (Added) **Errata added or updated:** • 31. PCIe Rx Termination During Power-Up (Added) • 32. EICR Bit 23 Can be Read as Set (Added) **Miscellaneous Updates** • Revised Section 1.1, "Product Code and Device Identification" (Added single-port SKU ordering information).
2.3	July 26, 2012	**Specification Clarifications added or updated:** • 4. PCIe: No Snoop is Enabled By Default (Added) **Errata added or updated:** • 29. PCIe Compliance Pattern is Not Transmitted When Connected to a x4/x2/x1 Slot (Added) • 30. PF's MSI TLP Might Contain the Wrong Requester ID When a VF Uses MSI-X (Added) **Software Clarifications added or updated:** • 5. PF/VF Drivers Should Configure Registers That are Not Reset by VFLR (Added)
2.2	March 30, 2012	**Errata added or updated:** • 28. Flow Director Filters Configuration Issue (Added) **Software Clarifications added or updated:** • 4. Identity Network Adapter Port by Blinking LED (Added)
2.1	February 21, 2012	**Errata added or updated:** • Removed External Erratum #23. • 27. Double Image Policy Flow is Not Applicable to PHY Image Module (Added)
2.0	January 29, 2012	**Miscellaneous Updates** • Revision change to reflect latest software release. No hardware updates.

Revision	Date	Comments
1.9	January 9, 2012	Initial release (Intel Public)

1. Introduction

This document applies to the Intel® Ethernet Controller X540 (X540).

This document is an update to a published specification, the *Intel® Ethernet Controller X540 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* in this document refer to PCIe v2.1 (2.5GT/s and 5GT/s).

1.1 Product Code and Device Identification

Product Codes: ELX540AT2 and ELX540BT2

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

Table 1-1 Markings

Device	Stepping	Top Marking	S-Specification	Description
X540	B0	ELX540AT2	SLKTL[1]/SLKTK[2]	10 GbE, 2-port, 12.5 W maximum device power, 25 x 25 mm - Lead (Pb) free and RoHS compliant
X540	B0	ELX540BT2	SLKTQ[3]/SLKTP[2]	10 GbE, 2-port, 14 W maximum device power, 25 x 25 mm - Lead (Pb) free and RoHS compliant

1. Tray
2. Tape and Reel
3. Did not sample externally

Table 1-2 Device IDs

X540 Device	Vendor ID	Device ID	Revision ID
Intel® Ethernet Controller X540-AT2/X540-BT2	8086	1528	0
Intel® Ethernet Controller X540-AT2/X540-BT2 (with single port NVM)	8086	1560	0
Intel® X540 Virtual Function (Mailbox Communication)	8086	1515	0
Intel® X540 Virtual Function (Microsoft* Hyper-V)	8086	1530	0

Table 1-3 MM Numbers

Product	Tray MM#	Tape and Reel MM#
ELX540AT2	940872	940875
ELX540BT2	940880	940879

1.2 Marking Diagrams

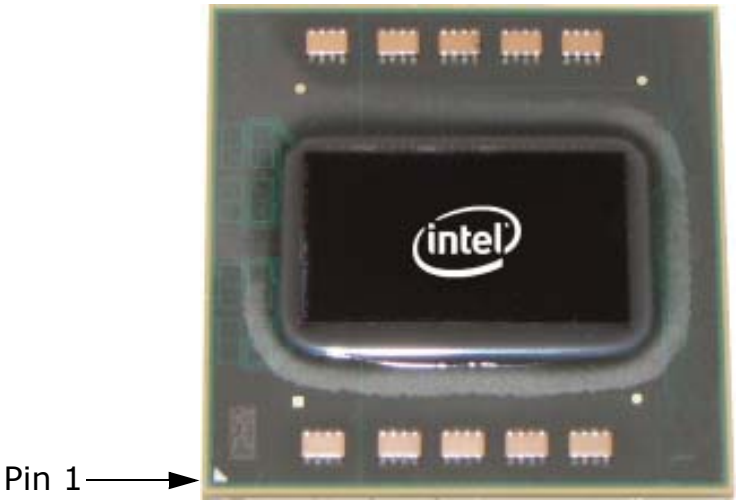
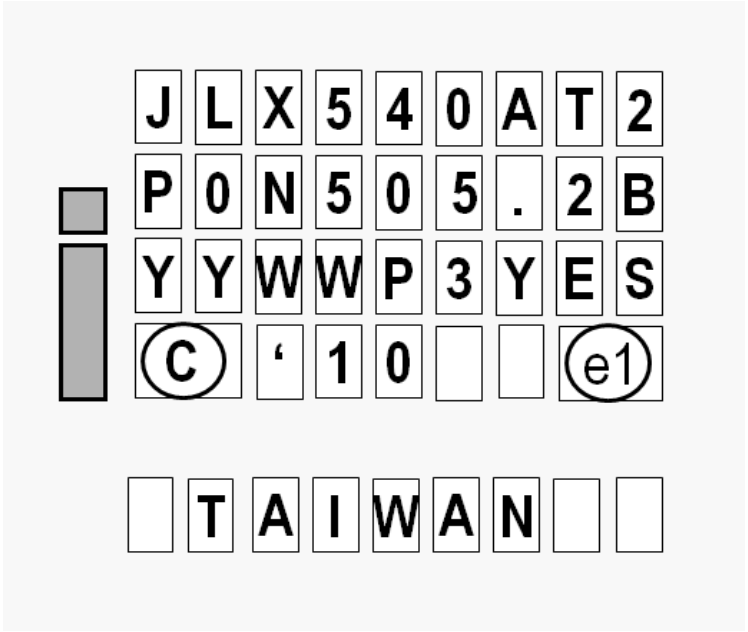
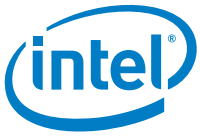


Figure 1-1 Example with Identifying Marks



- LINE1: Product code
- LINE2: Wafer lot# concatenated with Assembler vendor code
- LINE3: Assy YYWW followed by Q-spec# (no “Q”) and ES for Engineering Sample
- LINE4: Copyright, ' YY, Pb-free mark
- LINE5: Country of origin (COO)



1.3 Nomenclature Used in This Document

This document uses specific terms, codes, and abbreviations to describe changes, errata, 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.
Fixed	This erratum has been fixed.
Fix Planned	This erratum is intended to be fixed in a future stepping of the component.
NoFix	There are no plans to fix this erratum.
Fixed in NVM	This erratum has been fixed in NVM X.XX.
Fix Planned in NVM	This erratum is intended to be fixed in a future NVM version.
Eval	Plans to fix this erratum are under evaluation.

2. 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
1. PCIe Completion Timeout Value Must be Properly Set	N/A
2. Master Disable (Datasheet Section 5.2.5.3.2)	N/A
3. MCTP/DMTF Standard Compliance	N/A
4. PCIe: No Snoop is Enabled By Default	N/A
5. IPv6 Extended Headers are Parsed by the X540	N/A
6. Selecting a Rx Pool Using VLAN Filters	N/A
7. X540 Single Port NVM Image File	N/A
8. VLAN Anti-Spoof Filter of an Untagged Packet	N/A
9. SR-IOV Prefetchable Address Space	N/A
10. CRCERRS Statistic Counter	N/A
11. Rx Statistics Counters Do Not Count Runt Frames or Fragments Smaller Than 12 Bytes	N/A
12. FCOE_PARAM Field	N/A

Table 2-2 Summary of Specification Changes

Specification Change	Status
1. PBA Number Module — Word Address 0x15-0x16	N/A
2. Updates to PXE/iSCSI EEPROM Words	N/A
3. NC-SI Pull-Down Resistor Value Change	N/A
4. RXMTRL.*UDPT* Initial Value	N/A
5. The Flow Director FDIRErr(0) Bit in the Rx Descriptor is Valid Only if the FLM Bit is Set	N/A
6. OS2BMC Requires Flow Control Rx Threshold Configuration	N/A
7. ETQF[19:16] are Reserved	N/A

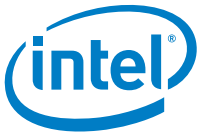


Table 2-3 Summary of Documentation Updates

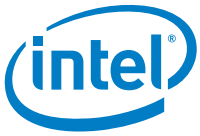
Documentation Update	Status
None.	N/A

Table 2-4 Summary of Errata; Errata Include Steppings

Erratum	Status
1. Flow Director: Length Error Bit Not Updated on a Remove Operation	B0=Yes; NoFix
2. Flow Director: Filter Might Lose the Length-Error Attribute in Perfect-Match Mode	B0=Yes; NoFix
3. Flow Director: L4 Packet Type Gives Wrong Indication	B0=Yes; NoFix
4. Replaced by Erratum #37	See Erratum #37.
5. No Length Error on VLAN Packets with Bad Type/Length Field	B0=Yes; NoFix
6. GPRC and GORCL/H Also Count Missed Packets	B0=Yes; NoFix
7. FCoE: In Order to Read DMA-Rx FCoE Context, CSRs Need to Add a Dummy Write	B0=Yes; NoFix
8. In 100 Mb/s Link Mode, CSR Access to DMA-Rx Might Reach an Internal Timeout	B0=Yes; NoFix
9. MACSec: When PN=0b, a Packet is Not Dropped	B0=Yes; NoFix
10. MACSec Statistics: LSECRXUC, LSECRXNUSA and LSECRXUNSA Statistics Counters Not implemented According to Specification	B0=Yes; NoFix
11. Cause of an Interrupt Might Never be Cleared	B0=Yes; NoFix
12. The X540 Does Not Meet the Timing Requirements for PAUSE Operation in 1 GbE Speed	B0=Yes; NoFix
13. The X540 Does Not Meet the Timing Requirements for PAUSE Operation in 100 Mb/s	B0=Yes; NoFix
14. NC-SI Additional Multicast Packets Might be Forwarded to the MC	B0=Yes; NoFix
15. SMBus: Unread Packets Received on One Port Might Cause Loss of Ability to Receive on Other Port	B0=Yes; NoFix
16. NC-SI: Packet Loss When the MC Sends Packets to Both Ports and One Port Has Link Down	B0=Yes; NoFix
17. FCoE: Exhausted Receive Context is Not Invalidated if Last Buffer Size is Equal to User Buffer Size	B0=Yes; NoFix
18. Gen1 Tx Compliance Pattern Test Mode (Wrong Disparity)	B0=Yes; NoFix
19. LEDs Cannot be Configured to Blink in LED_ON Mode	B0=Yes; NoFix
20. NVM Missing or Blank	B0=Yes; NoFix
21. Removed	N/A
22. LED Does Not Blink in Invert Mode	B0=Yes; NoFix
23. In Certain Configurations, LPLU (at S5) Can Link at 1 GbE	B0=Yes; Fixed in NVM 4.3
24. External POR Assertion	B0=Yes; NoFix
25. The Allow Link Down (ALD) Feature Does Not Work While Using Function Swap	B0=Yes; NoFix
26. PCIe Gen2 Tx Common Return Loss	B0=Yes; NoFix

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

Erratum	Status
27. Double Image Policy Flow is Not Applicable to PHY Image Module	B0=Yes; NoFix
28. Flow Director Filters Configuration Issue	B0=Yes; NoFix
29. PCIe Compliance Pattern is Not Transmitted When Connected to a x4/x2/x1 Slot	B0=Yes; NoFix
30. PF's MSI TLP Might Contain the Wrong Requester ID When a VF Uses MSI-X	B0=Yes; NoFix
31. PCIe Rx Termination During Power-Up	B0=Yes; NoFix
32. EICR Bit 23 Can be Read as Set	B0=Yes; NoFix
33. NC-SI: Get NC-SI Pass-Through Statistics Response Might Contain Incorrect Packet Counts	B0=Yes; NoFix
34. IPv4 Checksum Error Might be Reported for Multicast Frames Over 12 KB	B0=Yes; NoFix
35. Flow Director: Collision Indication Can be Cleared by Adding a New Filter	B0=Yes; NoFix
36. RXMEMWRAP Register Content is Inaccurate	B0=Yes; NoFix
37. Flow Director Statistics Inaccuracy	B0=Yes; NoFix
38. Clearing RXEN During VM-to-VM Loopback Traffic Might Cause an Rx Hang	B0=Yes; NoFix
39. NC-SI Hardware Arbitration Issues	B0=Yes; NoFix
40. PCIe SR-IOV Reserved Bits are Writable	B0=Yes; NoFix
41. PCIe Advanced Error Reporting: First Error Pointer	B0=Yes; NoFix
42. IPv4 Checksum Error Might be Reported for a Fragmented Packet	B0=Yes; NoFix
43. LLC Packet without SNAP Header	B0=Yes; NoFix
44. Internal PHY Slow Initialization	B0=Yes; NoFix



2.1 Specification Clarifications

1. PCIe Completion Timeout Value Must be Properly Set

The X540 Completion Timeout Value[3:0] must be properly set by the system BIOS in the X540 PCIe Configuration Space Device Control 2 register (0xC8; W). Failure to do so can cause unexpected completion timeouts.

The X540 complies with the PCIe 2.0 specification for the completion timeout mechanism and programmable timeout values. The PCIe 2.0 specification provides programmable timeout ranges between 50 µs to 64 s with a default time range of 50 µs - 50 ms. The X540 defaults to a range of 16 ms - 32 ms.

The completion timeout value must be programmed correctly in PCIe configuration space (in Device Control 2 register); the value must be set above the expected maximum latency for completions in the system in which the X540 is installed. This ensures that the X540 receives the completions for the requests it sends out, avoiding a completion timeout scenario. Failure to properly set the completion timeout value can result in the device timing out prior to a completion returning.

By default, the X540 does not resend the request upon a completion timeout; however, it can be programmed to do so. In this case after the completion timeout occurs, the device assumes the original completion is lost, and re-sends the original request. In this condition, if the completion for the original request arrives at the X540, this results in two completions arriving for the same request, which might cause unpredictable system behavior. NVM images provided by Intel set the resend feature to off and it is recommended to not enable it.

For details on completion timeout operation, refer to the *Intel® Ethernet Controller X540 Datasheet*.

2. Master Disable (Datasheet Section 5.2.5.3.2)

The driver might time out if the PCIe Master Enable Status bit is not cleared within a given time. Examples that delay the clearing of the PCIe Master Enable Status bit include flow control, link down, or DMA completions not making it back to the DMA block. In these instances, the driver should check that the Device Status register Transaction Pending bit (bit 5) in the PCI config space is clear before proceeding. Also, the driver should flush the transmit data path and initiate two consecutive software resets with a delay larger than 1 µs between them.

The recommended method to flush the transmit data path is as follows:

1. Inhibit data transmission by setting the HLREG0.*LPBK* bit and clearing the RXCTRL.*RXEN* bit. This configuration avoids transmission even if flow control or link down events are resumed.
2. Set the GCR_EXT.*Buffers_Clear_Func* bit for 20 µs to flush internal buffers.
3. Clear the HLREG0.*LPBK* bit and the GCR_EXT.*Buffers_Clear_Func* bit.

The *Intel® Ethernet Controller X540 Datasheet* will reflect this specification clarification in the next release.

3. MCTP/DMTF Standard Compliance

The X540 supports NC-SI over MCTP protocol over SMBus. The X540 MCTP protocol implementation is based on an early draft of the DSP0261 Standard, and it includes a Payload Type field removed in the final release of the standard.

Note: The X540 MCTP protocol implementation does not support pass-through traffic.

4. PCIe: No Snoop is Enabled By Default

The X540 enables the No Snoop feature by default after power on. No Snoop feature must be disabled during Rx flow software initialization if there is no intention to use it. To disable No Snoop, the CTRL_EXT.*NS_DIS* bit should be set to 1b.

5. IPv6 Extended Headers are Parsed by the X540

IPv6 extended headers are parsed by the X540, enabling TCP layer header recognition. As such, the IPv6 extended header fields are not taken into account for the queue classification by a flow director filter. Note that this rule does not apply for security headers and fragmentation headers. Packets with fragmentation headers miss this filter. Packets with security extended headers are parsed only up to these headers and therefore can match only filters that do not require fields from the L4 protocol.

6. Selecting a Rx Pool Using VLAN Filters

Rx Pool selection is described in the *Intel® Ethernet Controller X540 Datasheet*, Section 7.10.3.2. Note that pools are first selected by MAC Address filtering, and then by VLAN filtering. If the application is aiming to map packets to pools exclusively by their VLAN tags, it needs to replicate all incoming packets to all the different pools by their MAC Address.

To achieve the packet replication, PFVTCTL.*Rpl_En* should be set and the relevant MAC Address filtering bits should be set:

- MPSAR, PFUTA, MTA and VFTA tables.
- Relevant bits in PFVML2FLT registers – ROMPE, ROPE, BAM and MPE.

Pool selection by VLAN is then controlled by the PFVLVF and PFVLVFB registers.

7. X540 Single Port NVM Image File

When the X540 is used in a single-port configuration, the use of a specialized single-port NVM image file that saves power on the unused port is highly recommended. Please contact your Intel representative to obtain NVM images.

8. VLAN Anti-Spoof Filter of an Untagged Packet

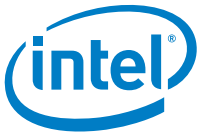
The VLAN anti-spoofing capability insures that a VM always uses a VLAN tag that is part of the set of VLAN tags defined on the Rx path. A Tx packet with a non-matching VLAN tag is dropped, preventing spoofing of the VLAN tag. Note that an untagged packet is not dropped.

9. SR-IOV Prefetchable Address Space

In SR-IOV mode, memory space should be allocated to the multiple VFs enabled. To accommodate the full extent of possible memory allocation, 64-bit addressing should be used. The PCI bridge specification requires that a 64-bit BAR be prefetchable.

The Prefetchable bit must be set in the NVM IOV Control Word and has been set in the NVM Dev Starter releases since revision 4.50.

To obtain an updated NVM image, contact your Intel representative.



10. CRCERRS Statistic Counter

A packet counted by the ILLERRC or the ERRBC statistics counter is not counted by the CRCERRS counter.

11. Rx Statistics Counters Do Not Count Runt Frames or Fragments Smaller Than 12 Bytes

TPR, RFC, and RUC statistics counters do not count runt frames or fragments smaller than 12 bytes.

12. FCOE_PARAM Field

The FCOE_PARAM field reported in the Rx descriptor indicates the size of the entire exchange only when the *Relative Offset Present* bit is set in the F_CTL field. If the bit is clear, software can conclude the size of the DDP data by reading the DDP buffer pointers FCPTRL, FCPTRH and the DDP buffer offset FCBUFF.*OFFSET*.

Intel implementation on Linux/ESX/Windows for the X540 is using DDP only for solicited data. In solicited data the *Relative Offset Present* bit is always set. As such, with Intel implementation, the FCOE_PARAM field reported in the Rx descriptor reflects the size of the DDP data.

2.2 Specification Changes

1. PBA Number Module — Word Address 0x15-0x16

The nine-digit Printed Board Assembly (PBA) number used for Intel manufactured Network Interface Cards (NICs) is stored in the EEPROM.

Note: Through the course of hardware ECOs, the suffix field is incremented. The purpose of this information is to enable customer support (or any user) to identify the revision level of a product.

Network driver software should not rely on this field to identify the product or its capabilities.

Current PBA numbers have exceeded the length that can be stored as hex values in these two words. For these PBA numbers the high word is a flag (0xFAFA) indicating that the PBA is stored in a separate PBA block. The low word is a pointer to a PBA block.

PBA Number	Word 0x15	Word 0x16
G23456-003	FAFA	Pointer to PBA Block

The PBA block is pointed to by word 0x16.

Word Offset	Description
0x0	Length in words of the PBA block (default 0x6).
0x1 ... 0x5	PBA number stored in hexadecimal ASCII values.

The PBA block contains the complete PBA number including the dash and the first digit of the 3-digit suffix. For example:

PBA Number	Word Offset 0	Word Offset 1	Word Offset 2	Word Offset 3	Word Offset 4	Word Offset 5
G23456-003	0006	4732	3334	3536	2D30	3033

Older PBA numbers starting with (A,B,C,D,E) are stored directly in words 0x15 and 0x16. The dash itself is not stored nor is the first digit of the 3-digit suffix, as it is always 0b for relevant products.

PBA Number	Byte 1	Byte 2	Byte 3	Byte 4
123456-003	12	34	56	03

2. Updates to PXE/iSCSI EEPROM Words

Words 0x30 and 0x34 (bits 2:0) are now defined as follows:

Bit(s)	Value	Port Status	CLP (Combo) Executes	iSCSI Boot Option ROM CTRL-D Menu	FCoE Boot Option ROM CTRL-D Menu
2:0	0	PXE	PXE	Displays port as PXE. Allows changing to Boot Disabled, iSCSI Primary or Secondary.	Displays port as PXE. Allows changing to Boot Disabled, FCoE enabled.
	1	Boot Disabled	NONE	Displays port as Disabled. Allows changing to iSCSI Primary/Secondary.	Displays port as Disabled. Allows changing to FCoE enabled.
	2	iSCSI Primary	iSCSI	Displays port as iSCSI Primary. Allows changing to Boot Disabled, iSCSI Secondary.	Displays port as iSCSI. Allows changing to Boot Disabled, FCoE enabled.
	3	iSCSI Secondary	iSCSI	Displays port as iSCSI Secondary. Allows changing to Boot Disabled, iSCSI Primary.	Displays port as iSCSI Allows changing to Boot Disabled, FCoE enabled.
	4	FCoE	FCOE	Displays port as FCoE. Allows changing port to Boot Disabled, iSCSI Primary or Secondary.	Displays port as FCoE Allows changing to Boot Disabled.
	7:5	Reserved	Same as disabled.	Same as disabled.	Same as disabled.
4:3	Same a before.				
5	Bit 5: formerly used to indicate iSCSI enable / disable, is no longer valid and is not checked by software.				
15:7	Same a before.				

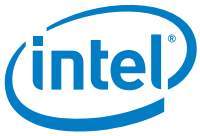
3. NC-SI Pull-Down Resistor Value Change

Previous version 1.2 documentation suggested a 100 Ω pull-down resistor value for NC-SI. It was determined that the 100 Ω value was to strong and prevented Input High Voltage (V_{IH}) from reaching 2.0 V minimum. To correct, replace all 100 Ω NC-SI pull-down resistors with 10 KΩ pull-down resistors.

4. RXMTRL.*UDPT* Initial Value

If the Time Sync (IEEE 1588) feature is used, the RXMTRL.UDPT field should be initialized to 0x13F.

This is fixed in ixgbe v3.11.20.



5. The Flow Director FDIRErr(0) Bit in the Rx Descriptor is Valid Only if the FLM Bit is Set

The FDIRErr(0) bit in Rx descriptor (length error) is valid only if the FLM bit is set (a packet matches a flow director filter) in the Extended Status of the Advanced Receive Descriptor.

6. OS2BMC Requires Flow Control Rx Threshold Configuration

If flow control is not enabled but OS2BMC is enabled, the FCRTH[n].*RTH* fields must be set as if flow control is enabled.

This information now appears in the *Intel® Ethernet Controller X540 Datasheet*, Revision 2.7.

7. ETQF[19:16] are Reserved

Bits 19:16 of the EType Queue Filter (ETQF) registers are reserved and should not be set by software.

2.3 Documentation Updates

None.

2.4 Errata

1. Flow Director: Length Error Bit Not Updated on a Remove Operation

Problem:

To avoid high latency, the length of the Flow Director (FD) filters linked list is limited. The length limit is programmable (FDIRCTRL.*MAX_LENGTH* field). If a linked list exceeds this limit, a length error is reported in the FDIRErr.*length* field in the Rx descriptor.

This erratum exists because once a filter is assigned to have the length-error attribute, it stays with this attribute even if an error condition doesn't exist anymore (such as a previous filter was removed from the list).

Implication:

When the FD table is programmed with many filters while dynamic filter removal is used, the driver might get an indication for over length lists (FDIRErr.*length*) even though the linked lists are not too long. This indication could be used by the software driver to remove filters from the table. Note that the current software driver does not use the dynamic filter removal option.

Workaround:

Software - Reset Flow Director (FD) tables when max-length indication is observed, or hold image of all the FD table and update the FD table (holding the image is less recommended).

The FD table is the hardware internal memory structure. Clearing this table means that the packet buffer memory of FD is cleared and linked to the empty link-list and head/tail CSRs are initialized. All other CSR are re-configured by software. (See the *Intel® Ethernet Controller X540 Datasheet*, Section 7.1.2.7).

Status: B0=Yes; NoFix

2. Flow Director: Filter Might Lose the Length-Error Attribute in Perfect-Match Mode

Problem:

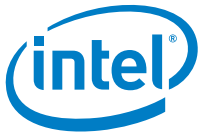
To avoid high latency, the length of the Flow Director (FD) filters linked list is limited. The length limit is programmable (FDIRCTRL.*MAX_LENGTH* field). If a linked list exceeds this limit, a length error is reported in the FDIRErr.*length* field in the Rx descriptor.

In some rare cases, a filter that has the length-error attribute might change the attribute to No-Length-Error. As a result, the FD table includes long lists, which are not reported to software. Once a packet matches these filters it causes a slightly higher latency in the device.

Implication:

There is no expected impact. In the cases where this indication is important, we expect other filters to indicate length-error.

FD tables are reset, which lowers the probability of reaching this case.There is also no impact to packet counters.



Workaround:

None.

Status: B0=Yes; NoFix

3. Flow Director: L4 Packet Type Gives Wrong Indication

Problem:

The MSB of the L4 Packet Type (L4TYPE) field in the Flow Director Filters Command Register (FDIRMC[6]) might give a wrong value during read access.

The flow director filters operate with the correct parameters.

Implication:

No impact on functionality. Software should ignore the read result of this bit.

Workaround:

None. Make sure that in a read to verify successful write, this bit is ignored.

Status: B0=Yes; NoFix

4. Replaced by Erratum #37

5. No Length Error on VLAN Packets with Bad Type/Length Field

Problem:

The X540 does not assert length error for VLAN packets that have a bad Type/Length field in the MAC header.

Implication:

There is no impact on system level performance. The packets are posted to the host as with any other packets.

Workaround:

None.

Status: B0=Yes; NoFix

6. GPRC and GORCL/H Also Count Missed Packets

Problem:

GPRC (Good Packets Received Count) and GORCL/H (Good Octets Received Count) count missed packets and missed packets bytes.

Implication:

None.

Workaround:

Statistics are available indirectly for these registers. This workaround is included in Intel drivers.

- For GPRC — Subtract MPC (Missed Packet Count) from GPRC. Alternatively, use QPRC.
- For GORCL/H — Use QBRCL/H (Quad Bytes Received).

Status: B0=Yes; NoFix

7. FCoE: In Order to Read DMA-Rx FCoE Context, CSRs Need to Add a Dummy Write

Problem:

There is a need to add a dummy write before the read of an FCoE context CSRs (FCDMARW) to avoid context corruption.

Implication:

No impact.

Workaround:

Write FCDMARW twice while having the required FCoE read index valid and 0b in the RE and WE bits.

Note: No workaround in current Intel drivers.

Status: B0=Yes; NoFix

8. In 100 Mb/s Link Mode, CSR Access to DMA-Rx Might Reach an Internal Timeout

Problem:

In 100 Mb/s link mode, internal clocks are slower, and access of an internal register can lead to timeout.

Implication:

An unknown value is returned on the PCI Express* (PCIe*) interface.

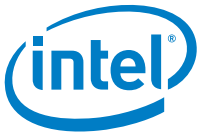
Workaround:

Software — In 100 Mb/s link mode, programmers need to disable aggregation in DMA-Rx (set RDRXCTL.*AGGDIS*=1b) and to extend the PCIe timeout extension to 32 μs (set PCIEMISC.*TO_EXTENSION* to 011b).

When aggregation is disabled, expect an impact on performance for packets below 128 bytes in length.

Note: Programmers should not increase the timeout extension beyond 32 μs to avoid PCIe system issues.

Status: B0=Yes; NoFix



9. MACSec: When PN=0b, a Packet is Not Dropped

Problem:

According to the MACSec specification, frames with PN=0 (packet number) in the sectag should be counted as bad tags/packets. The X540 considers these packets as late packets and they are incorrectly identified as a late packets instead of a bad tag/ packets. So they are dropped, but for the wrong reason (late packet instead of bad tagged).

Implication:

MACSec Rx statistic counters might report inaccurate values.

Workaround:

None.

Status: B0=Yes; NoFix

10. MACSec Statistics: LSECRXUC, LSECRXNUSA and LSECRXUNSA Statistics Counters Not implemented According to Specification

Problem:

InPktsUnchecked (LSECRXUC) statistic is not provided (LSECRXUC does not count correctly).

InPktsNotUsingSA (LSECRXNUSA) and InPktsUnusedSA (LSECRXUNSA) should be defined per SA. In this implementation, these are captured by a single counter.

Implication:

Statistics defined in the MACSec standard cannot be provided.

Workaround:

None.

Note: No workaround in current Intel drivers. Once MACSec is included in Intel drivers, this workaround will be applied.

Status: B0=Yes; NoFix

11. Cause of an Interrupt Might Never be Cleared

Problem:

If the cause of an interrupt is set by the Extended Interrupt Cause Set (EICS) register writing just before the interrupt line is set, then it might not be cleared. This means that there might be a deadlock that prevents the interrupt line from rising.

This erratum only occurs when all three modes referenced are used at the same time: non-PBA mode, Auto Clear (of the cause), No Auto Mask.

PBA is Pending Bit Array mode. During this mode the device is able to capture additional interrupts during the interval between initial interrupt and driver access to the device.

Implication:

The X540 stops issuing interrupts.

Workaround:

When operating using the above configurations, software should manually clear the cause by writing a 1b to the specific bit in the relevant EICR/EICR1/EICR2/VTEICR0-63 register (after the interrupt occurs and the EICS was written). This workaround is included in Intel drivers.

Status: B0=Yes; NoFix

12. The X540 Does Not Meet the Timing Requirements for PAUSE Operation in 1 GbE Speed

Problem:

In 1 GbE speed, the X540 responds to a received pause frame after a longer time than defined in the IEEE 802.3 specification.

Implication:

Specification conformance. The response gap is small.

Workaround:

None.

Status: B0=Yes; NoFix

13. The X540 Does Not Meet the Timing Requirements for PAUSE Operation in 100 Mb/s

Problem:

In 100 Mb/s speed, the X540 responds to a received pause frame after a longer time than defined in the IEEE 802.3 specification.

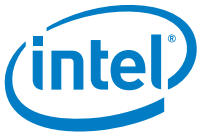
Implication:

Specification conformance. No system impact with low traffic.

Workaround:

None.

Status: B0=Yes; NoFix



14. NC-SI Additional Multicast Packets Might be Forwarded to the MC

Problem:

If the MC enables multicast filtering for IPv6 neighbor advertisement and/or IPv6 router advertisement, additional multicast packets are forwarded to the MC. The additional packets forwarded are:

- Packets with the ICMPv6 header's message type: 135, 137.
- IPv6 neighbor advertisement.
- IPv6 router advertisement.

Implication:

Additional packets might be forwarded to the MC.

Workaround:

The MC should filter the different multicast packets.

Status: B0=Yes; NoFix

15. SMBus: Unread Packets Received on One Port Might Cause Loss of Ability to Receive on Other Port

Problem:

The X540's two ports share an internal memory. When packets are received by one of the ports and not read by the MC, they are stored in the shared memory. When this memory fills up, no more packets can be received from either ports.

Implication:

Loss of packets. The MC should be aware of the previous behavior.

Workaround:

1. Make use of a SMBus alert timeout mechanism.
2. Momentarily disable receives by the other port.

Status: B0=Yes; NoFix

16. NC-SI: Packet Loss When the MC Sends Packets to Both Ports and One Port Has Link Down

Problem:

NCSI Rx (MC-to-LAN) FIFO is shared between both ports. When one of the LAN port’s Tx buffer is congested because of link failure or flow control, the NCSI Rx FIFO gets congested and as a result the packets for the second port also get dropped and are not sent to the LAN.

Implication:

Loss of packets. The MC should be aware of the problem.

Workaround:

The MC should monitor the link status and stop sending packets to a specific port if the link is down.

Status: B0=Yes; NoFix

17. FCoE: Exhausted Receive Context is Not Invalidated if Last Buffer Size is Equal to User Buffer Size

Problem:

If the last buffer of an FCoE context does not have sufficient room for the FC payload, the context is considered exhausted and must be invalidated by hardware.

The FCoE context is not invalidated as required under the following scenarios:

- FCoE last buffer size (FCDMARW.*LASTSIZE*) equals the exact user buffer size (FCBUFF.*BUFFSIZE*).
- FCoE DDP last payload byte in a mid packet written to the last byte of the last allocated buffer (the packet fills in the exact buffer value).
- Extra FCoE packet(s) are received in the problematic context.

Implication:

- Invalid host memory access.
- Hardware does not invalidate FCoE context when exhausted and does not assert error status to software.

Workaround:

FCoE context last buffer must be smaller than the context buffer size.

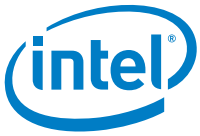
If it is necessary to configure a last buffer to equal buffer size, the following flow should be used:

- Allocate the extra user-buffer in the context list. Set it in the context buffer list and then increment FCBUFF.*BUFFCNT* to reflect a possible usage of an additional buffer.
- Set FCDMARW.*LASTSIZE* = 0x1.
- If flow ends and the extra buffer is used, the flow is invalid and exhausted.

If FCDMARW.*LASTSIZE* = FCBUFF.*BUFFSIZE*, the number of used DDP buffers is limited to 255. The FCBUFF.*BUFFCNT* value should be programmed for less than 256.

The workaround is included in ixgbe v3.2.10 and in Intel's Windows* drivers, starting with Release 16.4 version 2.9.66.0.

Status: B0=Yes; NoFix



18. Gen1 Tx Compliance Pattern Test Mode (Wrong Disparity)

Problem:

Tx compliance is a test mode in which the X540 transmits a continuous sequence of symbols for electrical characterization purposes. In X540 A0, Gen1, this pattern is not compliant with the specification. The transmitted symbol sequence is COM-, D21.5, COM-, D10.2 (instead of COM-, D21.5, COM+, D10.2). Note that in Gen2, the test mode works correctly.

Implication:

PCIe specification compliance issue.

Workaround:

Use JTAG flow to configure the PCIe core to transmit the right pattern.

Status: B0=Yes; NoFix

19. LEDs Cannot be Configured to Blink in LED_ON Mode

Problem:

When the *LEDx_Mode* field of a specific LED is set to 1110b in the LEDCTL register (0x00200), the respective LED is in LED_ON mode. This LED should be always asserted when the mode is set to LED_ON. The LED should also blink based on the *LEDx_BLINK* setting; however, due to a device limitation, the LED does not blink regardless of the *LEDx_BLINK* value.

Implication:

LEDs cannot be configured to blink in LED_ON mode.

Workaround:

The software driver should switch between LED_ON and LED_OFF mode to make the LED blink.

Status: B0=Yes; NoFix

20. NVM Missing or Blank

Problem:

Due to analog defaults (overridden with NVM load), the X540 might not reach PCIe link with certain link partners if an NVM is missing or blank.

Implication:

The X540 might fail to show up on the PCIe bus when an NVM is missing or blank.

Workaround:

Pre-program the NVM prior to powering on the X540.

Status: B0=Yes; NoFix

21. Removed

22. LED Does Not Blink in Invert Mode

Problem:

LEDx_IVRT bit in LEDCTL register (offset 0x00200) is ignored if the respective *LEDx_BLINK* bit is set. This issue is relevant only if *LEDx_MODE* is programmed to one of the modes where *LEDx_BLINK* is used (MAC_ACTIVITY, FILTER_ACTIVITY, LINK_UP, LINK_1G, LINK_10G and LINK_100).

Implication:

LED stays lit during idle time.

Workaround:

If *LEDx_IVRT* must be set together with a blink effect, use LINK_ACTIVITY mode instead of the modes using *LEDx_BLINK* (MAC_ACTIVITY, FILTER_ACTIVITY, LINK_UP, LINK_1G, LINK_10G and LINK_100).

Status: B0=Yes; NoFix

23. In Certain Configurations, LPLU (at S5) Can Link at 1 GbE

Problem:

In certain configurations, LPLU (at S5) can link at 1 GbE instead of 100 Mb/s if the following conditions take place:

1. Setting LPLU disable 10 GbE and LPLU disable 1 GbE in the NVM.
2. Linking at 1 GbE at S0.
3. Going down to S5.
4. Link is 1 GbE instead of 100 Mb/s.

Implication:

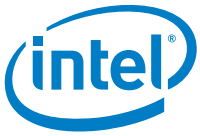
LPLU links at 1 GbE instead of 100 Mb/s.

Workaround:

1. Set LPLU to 10 GbE disable in the NVM (links at 100 Mb/s if the link partner is capable).
2. Set both LPLU disable 10 GbE and LPLU disable 1 GbE in the NVM. Make sure your S0 link is 10 GbE and when going down to S5 it ends with a 100 Mb/s link.

Status: B0=Yes; Fixed in NVM 4.3

Note: NVM starting with v4.3 and later will incorporate this fix.



24. External POR Assertion

Problem:

When the BYPASS_POR signal is asserted high (Power On Reset bypass), the X540 might not established link due to PHY reset limitations.

Implication:

When set to 1b, BYPASS_POR disables the internal POR circuit and uses the LAN_PWR_GOOD pin as a POR indication. Note that this might not work due to PHY reset limitations.

Workaround:

Setting BYPASS_POR to 0b (instead of 1b), maintains the functionality of using the LAN_PWR_GOOD pin as a POR indication.

Status: B0=Yes; NoFix

25. The Allow Link Down (ALD) Feature Does Not Work While Using Function Swap

Problem:

The ALD feature does not work when function swap is applied.

When one function in D3, a second in D0, and when ALD is applied to the D3 function.

Implication:

The ALD feature doesn't work while using port swap.

Workaround:

None.

Status: B0=Yes; NoFix

26. PCIe Gen2 Tx Common Return Loss

Problem:

Spec com: < -6 dB @ 50 MHz-2.5 GHz; Some lanes showed failures @ 2.5 GHz - the violation is 0.5-2.45 db above specification limit.

Implication:

No impact on customers.

Workaround:

None.

Status: B0=Yes; NoFix

27. Double Image Policy Flow is Not Applicable to PHY Image Module

Problem:

NVM double image policy flow is used to protect the update of big Flash modules, but it is not applicable to the PHY image module.

Implication:

The PHY image module update is not protected by the double image policy.

Workaround:

None.

Status: B0=Yes; NoFix

28. Flow Director Filters Configuration Issue

Problem:

Before an X540 receive path enable, the default value of both RXCTRL.*RXEN* and SECRXCTL.*RX_DIS* is zero. If the flow director filters are configured in this state, the receive data buffer might not be configured correctly.

Implication:

Receive hang.

Workaround:

If RXCTRL.*RXEN* is clear, set SECRXCTL.*RX_DIS* and wait for a SECRXSTAT.*SECRX_RDY* indication before configuring the flow director filters.

This workaround is implemented in the Intel ixgbe driver 3.8.21.

Status: B0=Yes; NoFix

29. PCIe Compliance Pattern is Not Transmitted When Connected to a x4/x2/x1 Slot

Problem:

If the PCIe compliance pattern is activated by setting the Enter Compliance bit in the Link Control 2 register, the X540 is able to transmit the compliance pattern only if it is connected to a x8 slot. If it is connected to a x4, x2 or x1 slot, the unconnected lanes falsely cause a premature exit from the compliance state and the pattern is not transmitted.

If a passive test load is applied on all lanes, the X540 goes to a compliance state and transmits the pattern accordingly, regardless of the internal lane width configuration.

Implication:

A PCIe compliance pattern cannot be transmitted if the X540 is connected to a x4 or narrower PCIe slot.

Workaround:

The X540 is still able to transmit the compliance pattern when connected to a x4/x2/x1 slot if entering the Polling.Compliance state due to detecting eight consecutive TS1 Ordered Sets in Polling.Active with the Compliance Receive bit (bit 4 of Symbol 5) asserted.

Status: B0=Yes; NoFix

30. PF's MSI TLP Might Contain the Wrong Requester ID When a VF Uses MSI-X

Problem:

When using IOV, if a PF uses MSI interrupts and one or more VFs use MSI-X interrupts, some of the MSI TLPs for the PF might contain the wrong Requester ID.

Implication:

There could be missing interrupts on the PF since the incorrect Requester ID could result in the virtualization mechanism mis-routing or dropping TLPs.

Workaround:

If any VFs use MSI-X, all PFs should also use MSI-X.

Status: B0=Yes; NoFix

31. PCIe Rx Termination During Power-Up

Problem:

According to the PCIe Specification, the receiver is required to present a high-impedance termination any time adequate power is not provided to the receiver until the device is out of reset. The X540 does not present high-impedance termination for a period of a few hundred microseconds right after power on, and then restores the high-impedance termination right after that for at least 100 ms until it is out of reset.

Typically, this issue is unseen if the root complex exits reset at the same time or after the X540. In systems where the root complex exits reset before the X540 is powered up, this issue might cause a false detection of the X540 Rx by the root complex, and its LTSSM moves to the polling state. However, since the X540 is under reset for at least 100 ms, the root complex LTSSM should time out and it will not affect the PCIe link connection.

Implication:

Momentary detection of the X540 receiver when it is still under reset. However, for a fully compliant root complex there is no implication.

Workaround:

Synchronizing the reset of the X540 and an upstream device avoids any false Rx detection.

Status: B0=Yes; NoFix

32. EICR Bit 23 Can be Read as Set

Problem:

EICR bit 23 is reserved but can sporadically be read as 1b. If erroneously enabled by the software device driver, it can generate an interrupt.

Implication:

Unexpected software device driver behavior.

Workaround:

The software device driver should ignore this bit.

Status: B0=Yes; NoFix

33. NC-SI: Get NC-SI Pass-Through Statistics Response Might Contain Incorrect Packet Counts

Problem:

The X540 maintains packet counters that are used in the Get NC-SI Pass-through Statistics Response. These counters are halted during PCIe reset.

Implication:

If a PCIe reset has occurred since the previous Get NC-SI Pass-through Statistics Response, the packet count values could be lower than the actual packet counts.

Workaround:

The packet counts in the Get NC-SI Pass-through Statistics Response can be used for debug purposes, but they should not be used for maintaining reliable statistics.

Status: B0=Yes; NoFix

34. IPv4 Checksum Error Might be Reported for Multicast Frames Over 12 KB

Problem:

IPE (IPv4 Checksum Error) might be rarely set in the Rx descriptor of multicast frames over 12 KB even though their checksum is valid.

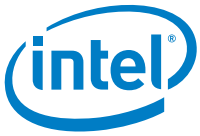
Implication:

An IPE (IPv4 Checksum Error) error can incorrectly be reported by the X540.

Workaround:

To avoid the erratum condition, limit the size of jumbo frames to less than or equal to 12 KB.

If using jumbo frames over 12 KB, software should re-calculate the IPV4 Header Checksum if RDESC.*IPE* is set.



The Intel Windows* and Linux* drivers limit the size of jumbo frames to less than or equal to 9 KB and are not exposed to this erratum.

Status: B0=Yes; NoFix

35. Flow Director: Collision Indication Can be Cleared by Adding a New Filter

Problem:

A Flow Director collision indication of the last Signature filter can be unintentionally cleared by adding a subsequent Signature filter.

Implication:

Flow Director collision indication is missing.

Workaround:

None.

Status: B0=Yes; NoFix

36. RXMEMWRAP Register Content is Inaccurate

Problem:

RXMEMWRAP register (0x03190) content is inaccurate:

- Rx Buffer Wrap Around Counter values could be inaccurate.
- Rx Buffer Empty bits are not reliable in the presence of FCoE or TCP-no-payload packets.

Implication:

Incorrect status read.

Workaround:

Use the RXUSED register for an indication as to whether or not the Rx buffer is empty.

Status: B0=Yes; NoFix

37. Flow Director Statistics Inaccuracy

Problem:

- FDIRMATCH (0x0EE58) should count the number of packets that matched any flow director filter.
- FDIRMISS (0x0EE5C) should count the number of packets that missed matching any flow director filter.
- FDIRFSTAT.*FADD* (0x0EE54, bits 7:0) should count the number of failed added filters due to no space in the filter table.

These counters might be incremented by two instead of one.

Implication:

The counters cannot be used for exact statistics. Counters should be used as an approximate indication on miss/match/failed addition of filters.

Workaround:

None.

Status: B0=Yes; NoFix

38. Clearing RXEN During VM-to-VM Loopback Traffic Might Cause an Rx Hang

Problem:

If the RXCTRL.*RXEN* bit is cleared during the reception of VM-to-VM loopback data traffic, the Rx path might hang.

Implication:

Rx hang.

Workaround:

The PFDTXGSWC.*LBE* bit should be cleared before clearing RXCTRL.*RXEN*, and can be set again after setting RXCTRL.*RXEN*.

Status: B0=Yes; NoFix

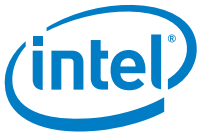
39. NC-SI Hardware Arbitration Issues

Problem:

1. During normal operation, the X540 might get FLUSH commands with a smaller ID then the device ID. The X540 should pass on the received FLUSH; but it sends it's own ID for ~2 µs and then passes on the lower ID FLUSHes.
2. The time from received-idle (while in a wait idle state) until the X540 sends IDLE on ARB_OUT is 1.7 µs; the maximum time allowed (by the specification) is T9 = 640 ns.
3. The X540 sends XON opcode after the end of the Master Assignment process, even if the XOFF time (~300 ms) has expired. If the X540 exits the congested mode during the Master Assignment process it sends XON opcode at the end of the Master Assignment process even if the XOFF has expired. The X540 does not consider the Master Assignment duration in the XOFF expiration time.
4. When the X540 enters congestion mode it sends XOFF opcode and also makes a request for TOKEN in order to send a XOFF frame. When the X540 enters congestion mode it should send a XOFF frame if it has the TOKEN or XOFF opcode even if it has not received the TOKEN. The X540 should not send both of them (opcode and frame) in any case.

Implication:

1. No implication in actual operation. Eventually, the lower IDs pass and arbitration succeeds.
2. The issue is not expected to cause problems because the timeout period is longer. Minor NC-SI compliance violation related to hardware arbitration.



3. No implication.
4. Slight delay in traffic coming from the MC but no platform implication.

Workaround:

None.

Status: B0=Yes; NoFix

40. PCIe SR-IOV Reserved Bits are Writable

Problem:

According to the PCIe Specification, RsvdP register fields must be read only and must return 0 (all 0's for multi-bit fields) when read.

In this device the following reserved bits are writable:

- SR-IOV Capability Structure offset 0x08 - SR-IOV Control/Status Register (0x168), bits 15:5.
- SR-IOV Capability Structure offset 0x13 (0x173), bits 7:0

Implication:

No functional implication. Note that software should not write to any reserved bits.

Workaround:

None.

Status: B0=Yes; NoFix

41. PCIe Advanced Error Reporting: First Error Pointer

Problem:

The First Error Pointer in the Advanced Error Capabilities and Control Register (PCIe register 0x118 bits 4:0) is a field that identifies the bit position of the first error reported in the Uncorrectable Error Status register. In the X540 implementation, the following bits of the Uncorrectable Status Register are not covered by this field:

- Bit 4 -Data Link Protocol Error Status.
- Bit 13 - Flow Control Protocol Error Status.
- Bit 14 - Completion Timeout Status.

Implication:

PCIe specification compliance issue.

Workaround:

None.

Status: B0=Yes; NoFix

42. IPv4 Checksum Error Might be Reported for a Fragmented Packet

Problem:

In rare cases, IPE (IPv4 Checksum Error) might be set in the Rx descriptor of a fragmented packet even though the checksum is valid.

The issue can happen only in a packet with the IPv4 header followed by payload data with no TCP/UDP header, and the first payload bytes looks like the SNAP packet header – AA AA 03 00 00 00.

Implication:

An IPE (IPv4 Checksum Error) error can incorrectly be reported by the device.

Workaround:

If an IPv4 checksum error is reported by the device, the software driver should validate the checksum if the first payload bytes looks like the SNAP packet header – AA AA 03 00 00 00.

Status: B0=Yes; NoFix

43. LLC Packet without SNAP Header

Problem:

If FCoE filtering is enabled, and an LLC Header is recognized in the packet, the X540 always skips the 3 bytes of the presumed SNAP Header and looks for the FCoE Ether-Type (0x8906). If a SNAP Header is not present, and the data at this offset is 0x8906, the packet is falsely recognized as FCoE.

Implication:

In general, a valid LLC packet incorporates a SNAP Header and there is no impact. The problematic packet might be seen in vendor-specific traffic. In this case, the packet is dropped and FCCRC counter is incremented.

Workaround:

None.

Status: B0=Yes; NoFix

44. Internal PHY Slow Initialization

Problem:

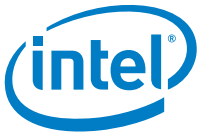
Under very rare conditions, the internal PHY power-up might be slower than usual.

In this case, if APM is enabled in the NVM, the device will try to establish link in a very early stage. Since the PHY is not ready at this stage, the PHY might hang and the link will remain down.

An AC Power cycle should recover the link.

Implication:

If APM is enabled, the Link will stay down. If APM is disabled in the NVM there is no issue.



Workaround:

Enable APM by FW:

- APM bits are cleared in NVM (at offset 0x38) and ACPI capabilities bits set (offset 0x2C).
- The FW reads ACPI capabilities and enables also APM (WoL) during Dr.

This flow will delay APM enable and will allow PHY full initialization.

This workaround suits only users that will always have APM & ACPI enabled in the NVM.

Status: B0=Yes; NoFix

If the device is not able to link after power on, please contact your Intel Representative to get an NVM with the workaround implemented.

3. Software Clarifications

Table 3-1 Summary of Software Clarifications

Software Clarification	Status
1. While in TCP Segmentation Offload, Each Buffer is Limited to 64 KB	N/A
2. RSC Performance Trade-Off	N/A
3. Serial Interfaces Programmed by Bit Banging	N/A
4. Identity Network Adapter Port by Blinking LED	N/A
5. PF/VF Drivers Should Configure Registers That are Not Reset by VFLR	N/A

1. While in TCP Segmentation Offload, Each Buffer is Limited to 64 KB

The X540 supports 256 KB TCP packets. However, each buffer is limited to 64 KB since the data length field in the descriptor is only 16 bits. This restriction can complicate things for the driver if the operating system passes down a scatter/gather element greater than 64 KB in length. This issue can be avoided by limiting the offload size to 64 KB.

Investigation has concluded that the increase in data transfer size does not provide any noticeable improvements in LAN performance. As a result, Intel network software drivers limit the data transfer in all drivers to 64 KB.

Please note that Linux operating systems only support 64 KB data transfers.

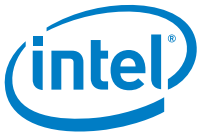
2. RSC Performance Trade-Off

The RSC feature is used to merge receive frames into the same descriptor structure with a shared header, improving receiving packet performance.

It should be noted that if small Rx data buffers are used (2 KB), RSC may involve a high rate of partial cache line PCIe transactions, which have a performance penalty from a memory access perspective.

In overloaded systems (more than 2 x 10 Gb/s LAN ports traffic load) the use of RSC may adversely affect Rx data throughput. Therefore, there is a performance trade off regarding the usage of the RSC feature.

To improve throughput in overloaded systems, the user can use large receive data buffers (larger than 2 KB or may opt to turn of RSC.



3. Serial Interfaces Programmed by Bit Banging

When bit banging on a serial interface (such as SPI, I^2C, or MDIO), it is often necessary to perform consecutive register writes with a minimum delay between them. However, simply inserting a software delay between the writes can be unreliable due to hardware delays on the CPU and PCIe interfaces. The delay at the final hardware interface might be less than intended if the first write is delayed by hardware more than the section write. To prevent such problems, a register read should be inserted between the first register write and the software delay. For example: write, read, software delay, write.

4. Identity Network Adapter Port by Blinking LED

Intel device drivers and supported tools include a feature that provides network adapter port identification by blinking LED2. This feature assumes that LED2 is connected as the Link/Activity LED as recommended in the reference schematics.

5. PF/VF Drivers Should Configure Registers That are Not Reset by VFLR

The following registers are not reset by VFLR and need to be configured by PF or VF in case of a change to a new configuration (such as VF OS transition):

- VFRDH/T
- VFTDH/T
- VFPSRTYPE
- VFSRRCTL
- VFRXDCTL
- VFMBMEM
- VFTXDCTL
- VFTDWBAL/H
- VFDCA_RXCTRL
- VFDCA_TXCTRL

LEGAL