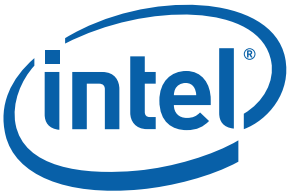


Intel® 82598 10 Gigabit Ethernet Controller Specification Update

LAN Access Division (LAD)

321040-010EN
Revision 2.93
October 2010

Legal

Revision History

Rev	Date	Comments
2.1	November 2007	Updated to support refresh
2.2	July 2007	Added Erratums 16, 25. Added Spec Clarification 1. Removed Spec Changes 1 and 2.
2.4	April 22, 2009	Combined internal and external. For old internal but numbers, search for OldRef#1.1, etc. Format udpated. *(Errata section - added or updated)* • 10. MISC: ECC On The Descriptor Completion Memory Needs To Be Disabled • 16. STAT: TX Counters Also Count Flow Control Bytes/packets • 26. PCIe: Upstream TLP Message Corruption • 35. PCIe: PCIe Elastic Buffer Noise Immunity Not Optimized *(Added - documentation to cover BSDL issues. Provides more data on JTAG testing.)* See Section 4. BSDL - JTAG Test Implications.
2.5	May 1, 2009	Combined internal and external. For old internal but numbers, search for OldRef#1.1, etc. Format udpated.
2.6	May 15, 2009	Specification Clarification added. • 2. PCIe: Completion Timeout Mechanism Compliance
2.7	June 10, 2009	Status changed to fixed. • 29. JTAG: JTDO Not Connected to Boundary Scan Shift Register During EXTEST Instruction
2.8	August 3, 2009	Specification Clarification updated. • 2. PCIe: Completion Timeout Mechanism Compliance
2.9	August 31, 2009	Specification Clarification updated. • 2. PCIe: Completion Timeout Mechanism Compliance - Workaround data updated. Errata - updated or added. • 26. PCIe: Upstream TLP Message Corruption • 32. PCIe: Missing Replay Due to Recovery During TLP Transmission • 33. PCIe: LTSSM Moves from L0 to Recovery Only When Receiving TS1/TS2 on All Lanes • 35. PCIe: PCIe Elastic Buffer Noise Immunity Not Optimized • 36. PCIe: SKP Ordered Set Resets Training Sequence Counter

2.91	February 26, 2010	Errata - added or updated. • 37. PCIe: Bus Hang if Nonexistent Register is Accessed • 38. PCIe: MSI-X Violation of PCIe Posted-Posted Rule • 39. PCIe: Completion with UR/CA Status Causes Unexpected Completion and Completion Timeout Errors to be Reported • 40. PCIe: Wrong Byte Enable Bit Used for Completion Timeout Disable Bit in Device Control 2 Register • 41. MAC: Transmitter Could Hang in 1GbE Mode if Flow Control is Enabled
2.92	7/16/2010	Specification Clarification - added. • 3. Receiver Detection Circuit Design and Established Link Width Errata - added or updated. • 42. APM Wake Up Might be Blocked if System is Shutdown Before Driver Load
2.93	10/11/2010	Specification Clarifications - added or updated. • 4. Use of Wake on LAN Together with Manageability • 5. RXDCTL.ENABLE and TXDCTL.ENABLE Will Not Change When Link Is Down Specification Change - added or updated. • 1. Update to PBA Number EEPROM Word Format Errata - updated. • 26. PCIe: Upstream TLP Message Corruption . Problem definition and Workaround #3 updated.

1. Introduction

This document applies to the Intel® 82598 10 Gigabit Ethernet Controller. It is an update to a published specification: the *Intel® 82598 10 Gigabit Ethernet Controller Datasheet*.

This document is intended for hardware system manufacturers and software developers of applications, operating systems or tools. It contains Specification Changes, Errata, and Specification Clarifications. All product documents are subject to frequent revision, and new order numbers will apply. New documents may be added. Be sure you have the latest information before finalizing your design.

1.1 Product Code and Device Identification

Product Code: JL82598EB

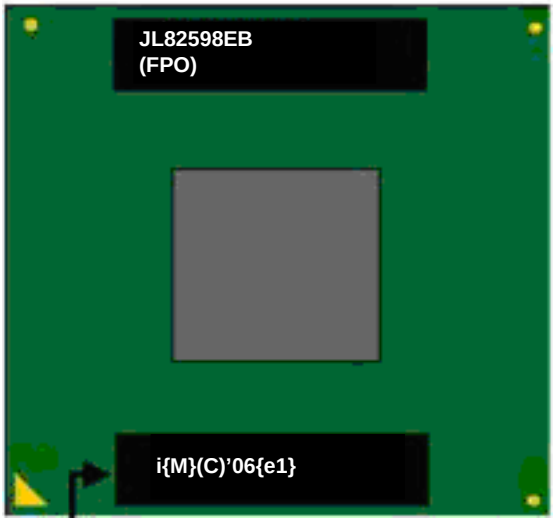
The following tables and drawings describe the various identifying markings on each device package:

Device	Stepping	MM	Top Marking	Q-Spec	Notes
82598EB	A1	890967	JL82598EB S LABE	N/A	Tape and Reel, Lead-Free
82598EB	A1	890968	JL82598EB S LABF	N/A	Tray, Lead-Free

Note: These devices can have a “GB” marking; these devices are used only on Intel network interfaces. The “GB” is functionally equivalent to the “EB” version.

Device	Vendor ID	Device ID	Revision ID
82598EB CX4 Applications	8086	10DD	0x1

1.2 Marking Diagram



Bottom area may contain two lines.

Figure 1-1. Example Showing 82598 Identifying Marks

Lead-free parts will have "JL" as the prefix for the product code (vs. "HL") and that the "Q" designator refers to the Q Specification number in the table above.

Devices can also have a "G" marking. Devices with this marking are used only on Intel network interface adapters.

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 for a description.s

Table 1-1. Terms, Codes, Abbreviation

Name	Description
Specification Changes	Modifications to the current published specifications. These changes will be incorporated in the next release of the specifications.
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.
Documentation Changes	Errors and omissions in published documents. Changes are incorporated in the next release of the documents.
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.

Table 1-1. Terms, Codes, Abbreviation

Sightings	Observed issues that are believed to be errata, but have not been completely confirmed or root caused. The intention of documenting sightings is to proactively inform users of behaviors or issues that have been observed. Sightings may evolve to errata or may be removed as non-issues after investigation completes.
A0, B1, etc.	**Stepping to which the issue applies.**
Fix	Issue to be fixed in a future stepping or DOC release.
Fixed	Issue has been fixed.
NoFix	There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.
Eval	Issue is being evaluated.

2. Sightings, Clarifications, Changes, Errata

See Section 1.3 for an explanation of terms, codes, and abbreviations used in the following tables and discussions.

Table 2-1. Summary of Sightings, Clarifications, Changes, Errata

SIGHTINGS	STATUS
None	NA
SPECIFICATION CLARIFICATIONS	STATUS
1. PCIe: End Point Request of I/O Space After Initialization	NA
2. PCIe: Completion Timeout Mechanism Compliance	NA
3. Receiver Detection Circuit Design and Established Link Width	NA
4. Use of Wake on LAN Together with Manageability	NA
5. RXDCTL.ENABLE and TXDCTL.ENABLE Will Not Change When Link Is Down	NA
SPECIFICATION CHANGES	STATUS
1. Update to PBA Number EEPROM Word Format	NA
ERRATA	STATUS
1. STAT: No “Length Error” Reported On VLAN Packets With Bad Type/Length Field	A1, NoFix
2. STAT: GPRC (Good Packet Receive Count) and GORC (Good Octets Received Count) Includes Missed Packets	A1, NoFix
3. STAT: MPRC (Multicast Packets Received Counter) Includes Received Broadcast Packets	A1, NoFix
4. INT: EITR (Extended Interrupt Throttle Register) Interval Set To Zero Causes A Write Back Per Descriptor	A1, NoFix
5. MAC/AN: Link Status Bit Is Not Self-Clearing On Read When In Link Down State	A1, NoFix

Table 2-1. Summary of Sightings, Clarifications, Changes, Errata

6. PCIe: Disabling Function0 Might Cause Some Systems to Stop	A1, NoFix
7. PCIe: Serial Number Is Not Correct	A1, NoFix
8. XAUI: CX4 Signal Detect May Violate Specification	A1, NoFix
9. XAUI: RX (Input) Return Loss Performance	A1, NoFix
10. MISC: ECC On The Descriptor Completion Memory Needs To Be Disabled	A1, NoFix
11. MAC/AN: Link Not Achieved When Two Devices, Configured To KX/KX4 Mode, Are Connected Back-to-Back, One With Auto-Negotiation Enabled And The Other Without	A1, NoFix
12. PCIe: Device Cannot Load Different Device IDs For The Two LAN Functions	A1, NoFix
13. PCIe (ANALOG): Tx Common Return Loss Violates Specification	A1, NoFix
14. DCB: Priority Flow Control Latency Specification Violation	A1, NoFix
15. MGE: NC-SI AC Timing Specification Violations	A1, NoFix
16. STAT: TX Counters Also Count Flow Control Bytes/packets	A1, Fixed
17. MAC/AN: With Lane Swap Enabled in 1G Mode Link is Not Automatically Achieved	A1, NoFix
18. PCIe: With LAN Swap Enabled, the DCA_ID.function_number Register Value Is Incorrect	A1, NoFix
19. PCIe: First PCIe Packet Is Sent With Completer ID = 0	A1, NoFix
20. MGE: Firmware Errata (NC-SI) - Additional Multicast Packets May Be Forwarded To The BMC	A1, NoFix
21. MGE: Firmware Errata (NC-SI) - Some VLAN Tagged Packets May Not Be Forwarded To The BMC While Using VLAN Mode #3	A1, NoFix
22. STAT: LED State Freezes On Entry To D3 No Wake	A1, NoFix
23. MAC/AN: Link Might Not Be Achieved in a Back-to-Back Configuration When Using Only Clause 37 Auto-Negotiation	A1, NoFix
24. JTAG: Out of Reset TAP Instruction Is Neither IDCODE Nor BYPASS	A1, NoFix
25. PCIe: Reception of Completion That Should Be Dropped May Occasionally Result In Device Hang or Data Corruption	A1, NoFix
26. PCIe: Upstream TLP Message Corruption	A1, NoFix
27. JTAG: JTDO Output is Disabled During a HIGHZ Instruction	A1, NoFix

Table 2-1. Summary of Sightings, Clarifications, Changes, Errata

28. JTAG: Boundary Scan Bypass Register is Not Loaded in Capture-DR State	A1, NoFix
29. JTAG: JTDO Not Connected to Boundary Scan Shift Register During EXTEST Instruction	A1, Fixed
30. JTAG: TAP Instruction Changes Need to be Passed Through Test Logic-Reset State	A1, NoFix
31. MAC/AN: Backplane Auto-Negotiation Does Not Work Correctly in Loose Mode	A1, NoFix
32. PCIe: Missing Replay Due to Recovery During TLP Transmission	A1, NoFix
33. PCIe: LTSSM Moves from L0 to Recovery Only When Receiving TS1/TS2 on All Lanes	A1, NoFix
34. GFLOW: TX CRC Must Be Enabled For Correct Flow Control Operation	A1, NoFix
35. PCIe: PCIe Elastic Buffer Noise Immunity Not Optimized	A1, Fixed
36. PCIe: SKP Ordered Set Resets Training Sequence Counter	A1, NoFix
37. PCIe: Bus Hang if Nonexistent Register is Accessed	A1, NoFix
38. PCIe: MSI-X Violation of PCIe Posted-Posted Rule	A1, NoFix
39. PCIe: Completion with UR/CA Status Causes Unexpected Completion and Completion Timeout Errors to be Reported	A1, NoFix
40. PCIe: Wrong Byte Enable Bit Used for Completion Timeout Disable Bit in Device Control 2 Register	A1, NoFix
41. MAC: Transmitter Could Hang in 1GbE Mode if Flow Control is Enabled	A1, NoFix
42. APM Wake Up Might be Blocked if System is Shutdown Before Driver Load	A1, NoFix
SOFTWARE CLARIFICATIONS	STATUS
1. While In TCP Segmentation Offload, Each Buffer is Limited to 64 KB	NA

2.1 Sightings

1. None active.

2.2 Specification Clarifications

1. PCIe: End Point Request of I/O Space After Initialization

Clarification: The 82598 requests I/O space if EEPROM bit 9, EEPROM PCIe General Configuration Section, PCIe Init Configuration 3 - Offset 3 is set. When this EEPROM bit is set, I/O Space is always requested.

The specification does not define a way to signal that IO BAR usage is done. When PCIe compliance tests are run, this may cause a test failure.

Implication: Failure when running PCI SIG compliance tests with EEPROM bit 9, EEPROM PCIe General Configuration Section, PCIe Init Configuration 3 - Offset 3 set.

Workaround: Disable I/O BAR requests via EEPROM bit 9, EEPROM PCIe General Configuration Section, PCIe Init Configuration 3 - Offset 3. Since various pre-boot SW tools require the I/O Space be requested, the bit is enabled by default in EEPROM images.

▼ *Return to Summary*

2. PCIe: Completion Timeout Mechanism Compliance

Clarification:PCIe Completion Timeout value must be properly set.

The 82598 Completion Timeout Value(3:0) must be properly set by the system BIOS in Intel 82598 PCIe Configuration Space Device Control 2 Register (0xC8; RW). Failure to do so can cause unpredictable system behavior.

The 82598 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 50us to 64s with a default time range of 50us-50ms. The 82598 defaults to a range of 500us – 1ms for PCIe capabilities version 1 and 2. The PCIe 2.0 Specification also strongly recommends that the default timeout value be such that the completion timeout mechanism not expire in less than 10ms.

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 82598 is installed. This will ensure that the 82598 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. In the event of a completion timeout, the device assumes the original completion is lost, and resends the original request, by default. In this condition, if the completion for the original request arrives at the 82598 device, this will result in 2 completions arriving for the same request, which may cause unpredictable system behavior.

As long as the Completion Timeout value is properly programmed by the system the completion timeout mechanism works without issue. It is expected that the system BIOS will set this value appropriately for the system.

Workaround: Alternatively a device driver could ensure the completion timeout value is set above 10ms (in order to follow the recommendation of the PCIe 2.0 specification). The driver would modify the timeout value, if and only if the default timeout value remains in configuration space. This will not impact BIOSs already changing the timeout value since the driver will not override any non-default setting of the timeout value. For extra protection against unpredictable system behavior in case the timeout setting is incorrect, it is recommended to disable the resend of the request. This can be done by clearing the Completion_Timeout_Resend bit in the EEPROM which sets the initial value of Completion_Timeout_Resend bit in the GCR Register.

New Intel drivers will implement this workaround (release 14.5 and after), modifying the completion timeout value in configuration space if the timeout value is still set to a value of 0x0 when the driver loads. The driver also disables the Completion_Timeout_Resend bit in the GCR Register.

The latest EEPROM dev_starter versions also have the Completion_Timeout_Resend bit disabled.

NOTE: The 82598 supports the ability to report a PCIe Capabilities version 1. (The PCIe Capabilities Version is loaded from the EEPROM (PCIE Init Configuration 3, bits 11:10) and reported in GCR bit 18 and in PCIe Capabilities register (0xA2) in PCIe configuration space.) PCIe v1.1 did not support a Programmable timeout in PCIe Configuration Space, therefore the timeout values are loaded from an EEPROM setting (PCIe Control, bits 6:5). This mechanism could be used to set a larger timeout value for systems in which the BIOS does not program the completion timeout value.

For details on Completion Timeout operation, see the product datasheet.

▼ *Return to Summary*

3. Receiver Detection Circuit Design and Established Link Width

Clarification: The 82598 receiver detection circuit was designed according to the PCIe Specification Rev. 1.1, which requires that an un-terminated receiver have an input impedance of at least 200 Kohm. PCIe Specification Rev. 2.0 allows the input impedance to be as low as 1 Kohm at input voltages in the range -150 - 0 mV and does not specify a minimum input impedance below -150 mV. As a result, a powered-down receiver lane with low input impedance at negative voltages could be compliant to Rev 2.0 and yet be falsely detected by the device as a terminated lane.

This is normally not an issue since any connnected lanes should be properly terminated within 5 ms after fundamental reset according to the PCIe Specification. However, there are some chipset devices that require significantly more time to prepare the termination and expect the link partner to remain in the LTSSM Detect state as long as none of the lanes are terminated. When used with such devices, the 82598 might falsely detect a receiver on one or more lanes and leave the Detect state. This can lead to establishing a link that is less than full width.

In this case, it is recommended that a Hot Reset be performed after a link has been established in order to force the 82598 to detect the receivers again when they are properly terminated. As a result, a full-width link can be established.

▼ Return to Summary

4. Use of Wake on LAN Together with Manageability

Clarification: The Wakeup Filter Control Register (WUFC) contains the NoTCO bit, which affects the behavior of the wakeup functionality when manageability is in use. Note that if manageability is not enabled, the value of NoTCO has no effect.

‘ When NoTCO contains the hardware default value of 0b, any received packet that matches the wakeup filters will wake the system. This could cause unintended wakeups in certain situations. For example, if Directed Exact Wakeup is used and the manageability shares the host's MAC address, IPMI packets that are intended for the BMC wakes the system, which might not be the intended behavior.

When NoTCO is set to 1b, any packet that passes the manageability filter, even if it also is copied to the host, is excluded from the wakeup logic. This solves the previous problem since IPMI packets do not wake the system. However, with NoTCO=1b, broadcast packets, including broadcast magic packets, do not wake the system since they pass the manageability filters and are therefore excluded.

Table 2-1. Effects of NoTCO Settings

WoL	NoTCO	Shared MAC Address	Unicast Packet	Broadcast Packet
Magic Packet	0b	-	OK	OK
Magic Packet	1b	Y	No wake	No wake
Magic Packet	1b	N	OK	No wake
Directed Exact	0b	Y	Wake even if MNG packet. No way to talk to BMC without waking host.	N/A
Directed Exact	0b	N	OK	N/A
Directed Exact	1b	-	OK	N/A

Intel Windows drivers set NoTCO by default.

▼ Return to Summary

5. RXDCTL.ENABLE and TXDCTL.ENABLE Will Not Change When Link Is Down

Clarification: The RXDCTL.ENABLE and TXDCTL.ENABLE register bits provide the internal status of the queue. (see datasheet section Receive Descriptor Control - RXDCTL and Transmit Descriptor Control - TXDCTL). In order to enable or disable a queue, the software driver should write the new ENABLE value and then poll the

bit until the change takes effect and the new value is read back. It should be noted that the change does not take effect while the link is down. Therefore, it is not possible to disable and then re-enable a queue while the link is down.

▼ Return to Summary

2.3 Specification Changes

1. Update to PBA Number EEPROM Word Format

Change: 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 EEPROM.

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.

PBA numbers have exceeded the length that can be stored as HEX values in two words. For newer NICs, the high word in the PBA Number Module is a flag (0xFAFA) indicating that the actual PBA is stored in a separate PBA block. The low word is a pointer to the starting word of the PBA block.

The following shows the format of the PBA Number Module field for new products.

PBA Number	Word 0x8	Word 0x9
G23456-003	FAFA	Pointer to PBA Block

The following provides the format of the PBA block; pointed to by word 0x9 above:

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

The new PBA block contains the complete PBA number and includes the dash and the first digit of the 3-digit suffix which were not included previously. Each digit is represented by its hexadecimal-ASCII values.

The following shows an example PBA number (in the new style):

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
	Specifies 6 words	G2	34	56	-0	03

Older NICs have PBA numbers starting with [A,B,C,D,E] and are stored directly in words 0x8-0x9. The dash in the PBA number is not stored; nor is the first digit of the 3-digit suffix (the first digit is always 0b for older products).

The following example shows a PBA number stored in the PBA Number Module field (in the old style):

PBA Number	Byte 1	Byte 2	Byte 3	Byte 4
E23456-003	E2	34	56	03

▼ *Return to Summary*

2.4 Documentation Changes

See the Revisions table in the front of the applicable Intel document.

2.5 Errata

1. STAT: No “Length Error” Reported On VLAN Packets With Bad Type/Length Field

Problem: The 82598 will 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.

Workaround: None.

Status: A1, NoFix

▼ *Return to Summary*

2. STAT: GPRC (Good Packet Receive Count) and GORC (Good Octets Received Count) Includes Missed Packets

Problem: GPRC (Good Packets Received Count) and GORC (Good Octets Received Count) includes MPC (Missed Packet Count). This is different from previous generation products.

Implication: None.

Workaround: Subtract MPC (Missed Packet Count) from GPRC for an accurate GPRC value or use QPRC. For GORC, use QBRC.

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ *Return to Summary*

3. STAT: MPRC (Multicast Packets Received Counter) Includes Received Broadcast Packets

Problem: The MPRC (Multicast Packets Received Counter) count also includes received broadcast packets.

Implication: MPRC count is incorrect.

Workaround: Subtract BPRC (Broadcast Received Packet Count) from MPRC for an accurate MPRC value.

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ *Return to Summary*

4. INT: EITR (Extended Interrupt Throttle Register) Interval Set To Zero Causes A Write Back Per Descriptor

Problem: Setting value of zero in EITR Interval register (bits [15:0]) will cause a write back per descriptor, disregarding write-back threshold value.

Implication: Will not get write back bursts as expected from setting the write-back threshold. This is the minimum inter-interrupt interval. Zero disables interrupt throttling logic.

Workaround: None.

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ *Return to Summary*

5. MAC/AN: Link Status Bit Is Not Self-Clearing On Read When In Link Down State

Problem: Link Status bit (LINKS bit 7) is asserted if there was one or more link down events since last link up. The status is to be cleared on read, but, if this bit is read when link is down, the bit will not clear.

Implication: Incorrect status is returned.

Workaround: Make sure bit is read once when link is up, then the next read is valid.

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ *Return to Summary*

6. PCIe: Disabling Function0 Might Cause Some Systems to Stop

Problem: When function0 is disabled, it becomes a dummy function that is valid for PCI. The BIOS of some systems may not handle this properly. (This is not a specification compliance issue for the 82598.)

Implication: System stops.

Workaround: Do not disable function0. Instead, cross the LAN-to-function mapping , and then disable function1.

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ *Return to Summary*

7. PCIe: Serial Number Is Not Correct

Problem: The PCIe serial number from the extended configuration space will not be correct in EEPROM-less mode and when LAN0 is disabled by the LAN-DISABLE pin. When LAN0 is disabled from EEPROM, the SN is still valid.

Implication: No impact at system level.

Workaround: None.

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ Return to Summary

8. XAUI: CX4 Signal Detect May Violate Specification

Problem: Signal meets specification if the voltage is greater than 175 mV p-p and does not meet the specification if less than 50 mV p-p. In the 82598, the signal is compared to a constant threshold (~110 mV). Variations may cause signals that are smaller than 50 mV p-p to be acceptable as long as they are greater than 42 mV p-p.

Implication: Specification non-compliance; no impact at system level.

Workaround: None.

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ Return to Summary

9. XAUI: RX (Input) Return Loss Performance

Problem: The XAUI RX fails differential return loss at frequencies >2 GHz (CX4 , KX4 pass).

Implication: Minor specification compliance issue.

Workaround: None.

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ Return to Summary

10. MISC: ECC On The Descriptor Completion Memory Needs To Be Disabled

Problem: Data errors can occur in Completion Memory when ECC is enabled and an ECC error occurs (byte enable memory does not support ECC).

Implication: Data can have errors, if enabled.

Workaround: Disable ECC on this memory area by writing 0 to reserved register bits – offset 0x110B0, bits 21,18,9,6 .

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ *Return to Summary*

11. MAC/AN: Link Not Achieved When Two Devices, Configured To KX/KX4 Mode, Are Connected Back-to-Back, One With Auto-Negotiation Enabled And The Other Without

Problem: When two devices, configured to KX/KX4 mode, are connected back-to-back and one has Auto-negotiation enabled and the other doesn't link won't be achieved.

Implication: No link in this configuration.

Workaround: None.

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ *Return to Summary*

12. PCIe: Device Cannot Load Different Device IDs For The Two LAN Functions

Problem: In the PCIe configuration space sections of the EEPROM (offset 2), there is an option to load the device id.

When this section is loaded for each LAN, the device id for lan0 is also loaded for lan1.

Implication: Support for only one device ID, loaded for both lan0 and lan1.

Workaround: None.

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ *Return to Summary*

13. PCIe (ANALOG): Tx Common Return Loss Violates Specification

Problem: The PCIe transmitter's worst-measured common mode return loss is up to -4.5 dB from 50 Mhz to 80 Mhz; the PCIe specification calls for -6dB.

Implication: Adds noise to the Tx lines; no system-level effect expected.

Workaround: None

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ Return to Summary

14. DCB: Priority Flow Control Latency Specification Violation

Problem: The specified delays from “priority pause received” until the last Tx are:

For 10 Gigabit=3072 nS (60 time slots; 3840 byte time)

For 1 Gigabit=1024 nS (two time slots; 128 byte time)

The 82598 delays are:

For 10 Gigabit=8500 nS (~5500 nS violation)

For 1 Gigabit=75, 000 nS (~74,000 nS violation)

Implication: Specification violation.

Workaround: Configure the 82598’s link partner’s Rx buffer thresholds to compensate for the violation.

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ Return to Summary

15. MGE: NC-SI AC Timing Specification Violations

Problem: Specification calls for:

- TCOmin=2.5 nS
- Thold=1 nS
- 82598 values are:
- TCOmin=1.8 nS
- Thold=1.8 nS

Implication: These values must be taken into consideration when implementing an NC-SI connection to the BMC.

Workaround: Add delay on “Data Out” and “Data In” as needed. For guidance and recommendations, please consult the Design Guidelines section of the Datasheet.

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ *Return to Summary*

16. STAT: TX Counters Also Count Flow Control Bytes/packets

Resolution: Refer to Section 4 “Programming Interface” in the Datasheet for more details.

Status: A1, Fixed

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ *Return to Summary*

17. MAC/AN: With Lane Swap Enabled in 1G Mode Link is Not Automatically Achieved

Problem: Swaping lane0 with any other lane will cause link down in 1G mode.

Implication: No link is achieved in 1G lane swap mode.

Workaround: Disable analog core lanes powerdown through EEPROM , these are the writes to analog core registers that need to be done:

0x24.5:3 <- 3’b111 -- Assert CAR_ATLAS_PWDWN_EN, PDN_TX_REG_EN and PDN_RX_REG_EN

0x0C <- 8’h00 -- De-assert PDN_TX_1G_Q0L3/2/1/0 and PDN_RX_1G_Q0L3/2/1/0

Status: A1, NoFix[1]

▼ *Return to Summary*

18. PCIe: With LAN Swap Enabled, the DCA_ID.function_number Register Value Is Incorrect

Problem: When LAN functions are crossed , register DCA_ID.funtion_select is not correct , it shows 0 for function 1 , and 1 for function 0.

Implication: Incorrect indication.

Workaround: 1. Driver reads the EEPROM , PCIe Control section (pointed to by word 0x6) - Offset 5 bit 10 to determine if functions are crossed.

2. Driver inverts DCA_ID.function_select if functions are crossed.

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ Return to Summary

19. PCIe: First PCIe Packet Is Sent With Completer ID = 0

Problem: The 82598 controller will always send first completion after PCI reset with completer ID = 0; this is instead of the bus number and device number captured from the configuration transaction.

Implication: The implication is minor; transactions are present and get correct completion.

Workaround: None.

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ Return to Summary

20. MGE: Firmware Errata (NC-SI) - Additional Multicast Packets May Be Forwarded To The BMC

Problem: If the BMC enables Multicast filtering for “IPv6 Neighbor Advertisement” and/or “IPv6 Router Advertisement”; additional Multicast packets are forwarded to the BMC. The additional packets are:

1.Packets with the ICMPv6 header’s Message Type: 135, 137.

2.IPv6 Neighbor Advertisement.

3.IPv6 Router Advertisement.

Implication: Additional packets may be forwarded to the MC.

Workaround: None.

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ Return to Summary

21. MGE: Firmware Errata (NC-SI) - Some VLAN Tagged Packets May Not Be Forwarded To The BMC While Using VLAN Mode #3

Problem: In VLAN Mode 3 (Any VLAN tagged packets & Non-VLAN tagged packets), if RCTL.VFE is set then only VLAN tagged packets that are configured in the VLAN (Host or Manageability) filter table will be forwarded to MC.

Implication: Some packets may not be forwarded to the BMC.

Workaround: Although using VLAN mode #3, the BMC should set any VLAN tag it uses with the NC-SI “Set VLAN” command.

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ *Return to Summary*

22. STAT: LED State Freezes On Entry To D3 No Wake

Problem: When transitioning to D3 no wake, LEDs will retain their last state. i.e. if link was on it will stay on in D3, even if no link.

Implication: Misleading LED output.

Workaround: Turn LEDs off before transition to D3.

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ *Return to Summary*

23. MAC/AN: Link Might Not Be Achieved in a Back-to-Back Configuration When Using Only Clause 37 Auto-Negotiation

Problem: When connected back-to-back and configured to clause 37 auto-negotiation (1G BX mode), link might not be achieved due to overlap of quiet periods in the state machine. There is a ~20% chance that auto-negotiation will succeed.

Implication: Cannot use pure clause 37 auto-negotiation in a back to back configuration.

Workaround: Enable both clause 37 and clause 73 auto-negotiation, but advertise only 1G support. This prevents the issue and clause 37 auto-negotiation completes. This workaround is not specification compliant; it is recommended only in back-to-back configuration.

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ *Return to Summary*

24. JTAG: Out of Reset TAP Instruction Is Neither IDCODE Nor BYPASS

Problem: The 82598 controller does not support the IDCODE instruction. We are supporting similar private instruction DEVSEL instead.The out of reset

instruction is not BYPASS, but DEVSEL (and this is spec violation, because DEVSEL not behaves exactly like IDCODE supposed)

Implication: The tester might mistakenly consider that the chip is in BYPASS mode while we are actually in DEVSEL mode.

Workaround: Force BYPASS mode by explicit command right after JRESET

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

See also: Section 4. BSDL - JTAG Test Implications in this document.

▼ *Return to Summary*

25. PCIe: Reception of Completion That Should Be Dropped May Occasionally Result In Device Hang or Data Corruption

Problem: This erratum can occur when the 82598 PCIe receives a completion that should be dropped, while the 82598 is starting a new request with the same TAG as the completion.

On an error-free PCIe link, this situation should never occur since the 82598 does not assert a second request with the same tag as an outstanding request.

Errors that could cause this failure:

- The TAG of a completion is corrupted due to noise on the line. This completion packet will be dropped due to LCRC error, but it could cause a failure if by chance a new request is asserted with the corrupted TAG value at the same time.
- On some platforms, it has been observed that when the upstream switch port transitions the link to L0s the line is noisy which may occasionally cause the 82598 to respond with a NAK. This NAK could cause a completion to be replayed. The 82598 will drop the duplicate packet based on the sequence number. However, the failure could occur if a new request is being asserted with the same TAG as the duplicate completion.
- An edge case of ACK timers results in a replay of a completion. This could cause the same case as above.

Implication: When the failure occurs, the actual completion data from the new request will be corrupted. The implications of this corruption of the read data depend on the type of request the 82598 was starting to send and are described below:

- TX descriptor – the 82598 may DMA the incorrect data and stops responding resulting in a device hang.
- TX data – the 82598 may transmit a packet on the network with invalid data but a valid CRC.
- RX descriptor – the 82598 may DMA a receive packet to the wrong memory address.

Workaround:

- Disabling L0s in the switch port that the 82598 is connected to prevents the duplicate completions caused by L0s.

- In the EEPROM image, keeping bit 13 (ACK/NACK Scheme) in word 0x1A of PCIe Initialization Configuration 3 set to 0b minimizes the chances of an ACK timeout.
- Set the Elastic Buffer Control bit in the EEPROM to w/a.

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ *Return to Summary*

26. PCIe: Upstream TLP Message Corruption

Problem: An internal PCIe retry buffer overflow followed by a certain sequence of messages can cause an upstream PCIe TLP corrupted message.

This failure occurs when the device is using:

[1] Legacy interrupt PCIe mode with the combined LAN receive data rate of both ports greater than the PCIe bandwidth that is effectively available.

-OR-

[2] MSI/MSI-X mode with 64-bit message addressing with the combined LAN receive data rate of both ports greater than the PCIe bandwidth that is effectively available.

Note that this issue does not occur if using a single LAN port with an x8 lane PCIe configuration since the LAN receive data rate is less than the PCIe bandwidth that is effectively available.

Implication: System hang.

Workaround#1:If using an Intel architecture system, use the MSI/MSIx interrupt scheme.

Workaround#2: If using a non-Intel architecture system, use the MSI/MSIx interrupt scheme with 32- bit message addressing.

Workaround#3: If legacy interrupt PCIe mode [1] or MSI/MSIx mode with 64-bit message addressing [2] must be used, limit PCIe posted and non-posted flow control credits advertised by the host. Note that the optimal number of credits configuration recommended is platform dependent. Typically, the total number of advertised credits should not exceed 120, with posted credits greater than non-posted, assuming credits are released only after the respective link layer ACK was sent. Additional programming details are available from your Intel representative.

Status: A1, NoFix

There are no plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ *Return to Summary*

27. JTAG: JTDO Output is Disabled During a HIGHZ Instruction

Problem: The 82598 disables JTDO outputs during a HIGHZ instruction. According to IEEE Std 1149.1-2001, "the HIGHZ instruction shall select the bypass register to be connected for serial access between TDI and TDO in the Shift-DR controller state".

Implication: If multiple devices are chained in the board, the tester won't be able to check devices behind the 82598 when it is in HIGHZ.

Workaround: Work in BYPASS mode and avoid any 82598 output contention.

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

See also: Section 4. BSDL - JTAG Test Implications in this document.

▼ *Return to Summary*

28. JTAG: Boundary Scan Bypass Register is Not Loaded in Capture-DR State

Problem: The 82598 does not load any bypass register value during a capture-DR TAP controller state. According to IEEE Std 1149.1-2001, "if bypass register is selected for inclusion in the serial path between TDI and TDO by the current instruction, the shift-register stage shall be set to a logic zero on the rising edge of TCK after entry into the Capture-DR TAP controller state".

Implication: The tester cannot recognize the BYPASS state of the 82598 while looking for a pull-down of JTDO.

Workaround: Drive zero in JTDI during the capture-DR state.

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

See also: Section 4. BSDL - JTAG Test Implications in this document.

▼ *Return to Summary*

29. JTAG: JTDO Not Connected to Boundary Scan Shift Register During EXTEST Instruction

Comment: This item has been fixed and is no longer applicable.

Status: A1, Fixed

▼ *Return to Summary*

30. JTAG: TAP Instruction Changes Need to be Passed Through Test Logic-Reset State

Problem: Changing TAP instructions in the 82598 should be passed through the test-logic-reset state which is not compliant with the IEEE Std 1149.1-2001 standard TAP controller state diagram.

Implication: Boundary scan instruction can be unpredictable.

Workaround: Pass through the test-logic-reset TAP state to change instructions.

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

See also: Section 4. BSDL - JTAG Test Implications in this document.

▼ *Return to Summary*

31. MAC/AN: Backplane Auto-Negotiation Does Not Work Correctly in Loose Mode

Problem: In loose mode, the DME alignment mechanism starts working after two PPM hops. This sometimes causes an alignment loss and a failure of the break_link state during an auto-negotiation FSM.

The wrap around in the DME aligner causes the insertion or removal of nine bits. If it occurred in an MV delimiter, the auto-negotiation process starts from the beginning.

Implication: Failure to achieve link in KX/KX4 mode.

Workaround: Disable loose mode by writing 0b to bit 24 in the AUTOC register (address 0x42A0). Disabling loose mode can also be done through the EEPROM (MAC 0/1 Section pointed by words 0x0B/0x0C, Auto Negotiation Defaults - Offset 0x04, bit 8).

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ *Return to Summary*

32. PCIe: Missing Replay Due to Recovery During TLP Transmission

Problem: If the replay timer expires during the transmission of a TLP and the LTSSM moves from L0 to Recovery during the transmission of the same TLP, the expected replay does not occur. Additionally, the replay timer is disabled, so no further replays will occur unless a NAK is received.

Implication: This situation should not occur during normal operation. If it does occur while the upstream switch is waiting for a replay, the result would be a Surprise Down error which might halt the system.

Workaround: None.

Status: A1, NoFix

▼ Return to Summary

33. PCIe: LTSSM Moves from L0 to Recovery Only When Receiving TS1/TS2 on All Lanes

Problem: According to the PCIe specification, the LTSSM should move from L0 to Recovery if a TS1 or TS2 ordered set is received on any configured Lane. The LTSSM only moves from L0 to Recovery if a TS1 or TS2 ordered set is received on all configured lanes.

Implication: This situation should not occur during normal operation since the upstream switch will transmit the TS1 or TS2 ordered sets on all lanes at the same time. If it does occur due to a broken lane, the result would be a Surprise Down error which might halt the system.

Problem does not occur under normal conditions.

Workaround: None required.

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ Return to Summary

34. GFLOW: TX CRC Must Be Enabled For Correct Flow Control Operation

Problem: The TXCRCEN bit in HLREG0 register (offset 0x04240, bit 0) enables CRC appending to Tx packets. If TXCRCEN is 0b, flow control packets will not have CRC appended and will be ignored by the partner.

Implication: Flow control is not operational.

Workaround: The HLREG0.TXCRCEN bit must be set to 1b if the 82598 is enabled to send flow control frames.

Status: A1, NoFix

There are no current plans to fix this issue in silicon. The issue may not have a significant impact or there may be a workaround. Check the release notes for your driver.

▼ Return to Summary

35. PCIe: PCIe Elastic Buffer Noise Immunity Not Optimized

Problem: The PCIe elastic buffer is used to synchronize between the clock generated by the clock recovery circuit and the internal clock. During electrical idle, in the absence of an input signal, the clock recovery circuit can be disturbed by noise and move the elastic buffer fill level away from the optimum value. EEPROM

control bits were implemented to maintain stability during electrical idle. In the default EEPROM image provided for the 82598, these bits were not set correctly.

Implication: In cases of increased noise levels during Electrical Idle, elastic buffer instability may, in rare instances, cause a link training failure when exiting from L1 state. Failure to exit L1 results in a Surprise Down error which may be a fatal error in operating systems that fully support Advanced Error Reporting. This issue is not relevant to L0s exit since L0s should be disabled in the downstream direction due to another erratum.

Workaround: Surprise Down Error reporting can be masked in the system. The system will recover after the link is re-established.

The fix is implemented in revised EEPROM version number 2.9.0; it sets PCIe Init Configuration 3 word, bits 4 and 5.

Status: A1, Fixed

▼ *Return to Summary*

36. PCIe: SKP Ordered Set Resets Training Sequence Counter

Problem: If a SKP ordered set is received during a TS1 or TS2 sequence, the TS counter is cleared. This will generally not be a problem since the upstream device should transmit at least 16 TS2 ordered sets and the 82598 only needs to detect 8 consecutive TS2 ordered sets to complete the Recovery process. A single reset of the counter will not cause a failure.

A failure can occur if the upstream device is non-compliant and transmits fewer than 16 TS2 ordered sets. In this case, could fail to compete the recovery process and then the PCIe link would go down.

Implication: There should be no failure when the upstream device functions according to the PCIe spec. If the upstream device is non-compliant, this issue could result in a Surprise Down error.

Workaround: None.

Status: A1, NoFix

▼ *Return to Summary*

37. PCIe: Bus Hang if Nonexistent Register is Accessed

Problem: In , the 0x8XXX offset in the memory BAR space is undefined. In normal operation, a PCIe access to an undefined offset in the memory BAR space should be completed after a timeout.

If an access is performed on port 0 to an address not located in the memory BAR and then an access is performed at offset 0x8XXX in the memory BAR space of port 1, this access does not complete and the PCIe bus hangs.

Implication: PCIe bus hangs.

Workaround: Do not access undefined CSR addresses.

Status: A1, NoFix

There are no current plans to fix this issue in silicon.

Updated LAD Windows* drivers are available that prevent non-existent/ undefined registers from being accessed.

▼ Return to Summary

38. PCIe: MSI-X Violation of PCIe Posted-Posted Rule

Problem: According to the PCIe Specification, “the acceptance of a Posted Request must not depend upon the transmission of any TLP from that same Upstream Port within the same traffic class.” The 82598 has a dependency between downstream posted requests to its MSI-X table and upstream MSI-X packets (MSI-X interrupt messages) that violates this rule.

Implication: Under specific stress scenarios, the upstream device might stop providing posted credits to the 82598. If the 82598 has a MSI-X message to send out and it runs out of posted credits, any upstream device access to the 82598 MSI-X table (read/write) does not complete until credits are renewed. Under this condition, the 82598 stops releasing posted credits to the upstream device, and posted data transfer stops in both directions resulting in a link deadlock. If the upstream device is able to renew its credit release flow, the 82598 is not susceptible to this erratum. If upstream device is able to renew it’s credit release flow, the 82598 is not be susceptible to this erratum.

Workaround: Implement the following:

- Use MSI instead of MSI-X interrupts. This can be accomplished via registry edits in Windows*. Intel can provide a tool that will automatically make the required registry edits in Windows. For Linux* this can be accomplished with added parameters at driver load by modifying InterruptType in /etc/modprobe.conf. InterruptType=0,0 means set both port 0 and port 1 to legacy interrupts (1,1 is MSI for both ports; 2,2 is MSI-X for both ports). Full details can be found in Linux driver README.

Contact your Intel representative for additional details or to obtain a tool for Windows* drivers.

Status: A1, NoFix

There are no current plans to fix this issue in silicon.

▼ Return to Summary

39. PCIe: Completion with UR/CA Status Causes Unexpected Completion and Completion Timeout Errors to be Reported

Problem: When the 82598 receives a PCIe completion with Unsupported Request (UR) or Completer Abort (CA) status in response to a request it generated, it reports an Unexpected Completion error. Because the completion timer is not disabled, a completion timeout error is reported when the timer expires.

Implication: This situation should not occur in systems that are operating correctly since all requests generated by the 82598 are supported.

If an UR/CA completion is received, the completion timeout error can bring down the operating system when it is reported.

Workaround: Not required for systems that are operating correctly.

Note that reporting completion timeout errors can be masked in the Uncorrectable Error Mask rgister.

Status: A1, NoFix

▼ *Return to Summary*

40. PCIe: Wrong Byte Enable Bit Used for Completion Timeout Disable Bit in Device Control 2 Register

Problem: BE[1] is used to enable the write to the Completion Timeout Disable bit in Device Control 2 register in the configuration space. It should be BE[0] since it is bit 4 in the register.

Implication: If a byte write is used, this bit is not updated since BE[1] is 0b.

The bit could be incorrectly written if a byte write to the high byte is performed. However, this is unlikely since bits 15:8 are all reserved.

Workaround: Use only word or Dword accesses to the Device Control 2 register.

Status: A1, NoFix

▼ *Return to Summary*

41. MAC: Transmitter Could Hang in 1GbE Mode if Flow Control is Enabled

Problem: If the 82598 is configured to operate at 1GbE with link flow control enabled, the transmitter might hang if a pause frame is received at a very specific timing relative to the start of a packet transmission.

Implication: Transmit hang.

Workaround: Disable link flow control when link is in 1 GbE mode.

Intel drivers disable link flow control for receive in 1 GbE.

This was implemented in Software Release 15.1.

Status: A1, NoFix

▼ *Return to Summary*

42. APM Wake Up Might be Blocked if System is Shutdown Before Driver Load

Problem: When the system is powered up and APM mode is enabled in the 82598 EEPROM, the device is able to wake correctly from a power saving state even before the software driver is loaded for the first time. According to APM specification, the 82598 is expected to be armed for further wake events even without software driver intervention.

In the 82598 implementation upon a wake event, the Magic Packet* Received bit is set in the WUS register. Also, this register needs to be cleared by the software driver before arming APM for a new wake event.

If an awake system is shutdown again before a software driver load, the Magic Packet Received bit that was not cleared might block further WoL events.

Implication: If the following events occur, in this order, this erratum might be observed:

- WoL event
- Driver doesn't load
- System transitions to S3/S5 state

For example, if after a WoL event, a BSOD occurs during system boot and the system is shutdown manually, a magic packet* might not be able to wake the system.

Workaround: None.

Status: 1, NoFix

▼ *Return to Summary*

3. Software Clarifications

Applies to Intel drivers.

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

Clarification: The 82598 supports 256 KB TCP packets; however, each buffer is limited to 64 KB since the data length field in the transmit descriptor is only 16 bits. This restriction increases driver implementation complexity if the operating system passes down a scatter/gather element greater than 64KB in length. This 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 size in all drivers to 64 KB.

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

For further details about how Intel network software drivers address this issue, refer to Technical Advisory TA-191.

▼ *Return to Summary*

4. BSDL - JTAG Test Implications

This section provides more detail on BSDL errata. The information is for manufacturers who are testing a device boundary-scan circuit and not using an ICT tester and, instead, are using a JTAG tester connected only to JTAG pins.

In each of the sections below, the errata number refers to the errata in the previous section. Click the title to return to the previous description.

24. JTAG: Out of Reset TAP Instruction Is Neither IDCODE Nor BYPASS

JTAG Tests Implication: IDCODE is commonly checked in basic JTAG test.

JTAG Tests Workaround: Disable the IDCODE check in the test configuration.

27. JTAG: JTDO Output is Disabled During a HIGHZ Instruction

JTAG Tests Workaround: Low impact. HIGHZ is not a common used command and can be avoided.

JTAG Tests Workaround: N/A.

28. JTAG: Boundary Scan Bypass Register is Not Loaded in Capture-DR State

JTAG Tests Workaround: Low impact. HIGHZ is not a common used command and can be avoided.

JTAG Tests Workaround: N/A.

30. JTAG: TAP Instruction Changes Need to be Passed Through Test Logic-Reset State

JTAG Tests Implication: JTAG tests can be combined in many different ways and instructions are changed on the fly. Test data may be re-used after changing instruction. In the 82598, the instructions SAMPLE, EXTEST and HIGHZ can be changed only by resetting the JTAG TAP state-machine which will consequently reset all the previous generated data.

JTAG Tests Workaround: Due to other 82598 JTAG limitations, the only applicable instruction with this limitation is SAMPLE. Whenever a JTAG instruction is changed from SAMPLE, a TAP RESET command should be added manually to test pattern files.

§ §