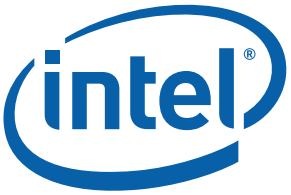

Intel® 82598 10 Gigabit Ethernet Controller Specification Update

November 2008

321040-001EN
Revision 2.3

Revisions

Date	Revision	Description
November 2008	2.3	• Updated Errata 2, 3, 7 and 25. • Added Errata 26 through 34.
July 2008	2.2	• Removed Specification Changes #1 and #2. • Removed Erratum 16. • Added Erratum 25. • Added Specification Clarification #1.
November 2007	2.1	Updated to support refresh.

Contents

Note: This page intentionally left blank.

1. Intel® 82598 10 Gigabit Controller Specification Update

This document applies to the Intel® 82598 10 Gigabit Ethernet Controller. In this document it is commonly referred to as "the device."

This document is an update to a published specification, the *82598 10 Gigabit Ethernet Controller Open Source Datasheet*.

This document is intended for hardware system manufacturers and software developers of applications, operating systems or tools. It may contain 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 & Device Identification

Product Code: JL82598EB

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

Table 1-1. Device, MM Number, Top Marking, Q-Specification (if applicable)

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.

Table 1-2. Vendor ID, Device ID, and Revision ID

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

Note: Contact your Intel representative for a complete list of 82598 device IDs.

1.2 Marking Diagram

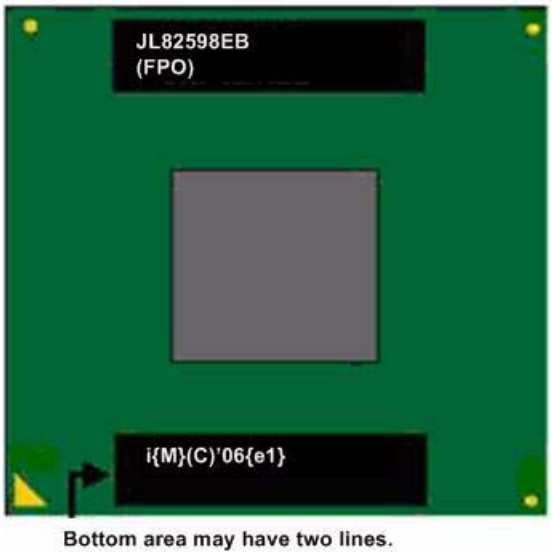

Figure 1-1. Example Showing 82598 Identifying Marks

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.

The devices can also have a “G” marking these devices are used only on Intel network interface adapters.

1.3 Summary of Changes

Table 1-3 provides definitions for the codes and abbreviations used to describe changes, errata, sightings, and clarifications listed in Table 1-4. Changes, errata, sightings and clarifications are individually described in the sections that follow.

Table 1-3. Nomenclature

Name	Description
Specification Changes	Modifications to the current published specifications. Changes will be incorporated in the next specification release.
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.
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.
Specification Clarifications	More detail about a design situation. The clarifications will be incorporated in the next release of the specifications.

Name	Description
Documentation Changes	Errors, or omissions in the current published specifications. Changes will be incorporated in the next specification release.
X	Specification Change, Erratum, or Specification Clarification that applies to this stepping. Applies to the stepping column.
Doc	Document change or update that will be implemented.
Fix	This erratum is intended to be fixed in a future stepping of the component.
Fixed	This erratum has been previously fixed.
NoFix	There are no plans to fix this erratum.
Eval	Plans to fix this erratum are under evaluation.
(No mark) or (Blank box)	This erratum is fixed in listed stepping or specification change does not apply to listed stepping.
Shaded	This Item is either new or modified from the previous version of the document.
DS	Data Sheet
DG	Design Guide
SDM	Software Developer's Manual
EDS	External Data Specification
AP	Application Note

Table 1-4. Change Summary

No.	A1	Plans or Action	Specification Changes
1	X	Doc	Removed - Refer to Section 2 "Signal Descriptions and Pinout List" in the Intel® 82598 10 Gigabit Ethernet Controller Datasheet for more details.
2	X	Doc	Removed - Refer to Section 2 "Signal Descriptions and Pinout List" in the Intel® 82598 10 Gigabit Ethernet Controller Datasheet for more details.
No.	**A1**	**Plans or Action**	**Errata**
1	X	NoFix	No "Length Error" Reported On VLAN Packets With Bad Type/Length Field.
2	X	NoFix	GPRC (Good Packet Receive Count) and GORC (Good Octets Received Count) Includes Missed Packets
3	X	NoFix	MPRC (Multicast Packets Received Counter) Includes Received Broadcast Packets
4	X	NoFix	EITR (Extended Interrupt Throttle Register) Interval Set to Zero Will Cause A Write Back Per Descriptor
5	X	NoFix	LINK_STATUS_CHECK Bit 7 Is Not Self-Clearing On Read When In Link Down State

Table 1-4. Change Summary

No.	A1	Plans or Action	Errata
6	X	NoFix	Disabling Function0 Might Cause Some Systems to Stop
7	X	NoFix	PCIe Serial Number Is Not Correct
8	X	NoFix	CX4 Signal Detect May Violate Specification
9	X	NoFix	XAUI RX (Input) Return Loss Performance
10	X	NoFix	ECC on the Descriptor Completion Memory Needs To Be Disabled
11	X	NoFix	Link Is Not Achieved When Two 82598 Devices, Configured to KX/KX4 Mode, Are Connected Back-to-Back, One WIth Auto-Negotiation Enabled And The Other Without
12	X	NoFix	82598 Cannot Load Different Device IDs For The Two LAN Functions
13	X	NoFix	PCIe Tx Common Return Loss Violates Specification
14	X	NoFix	Priority Flow Control Latency Specification Violation
15	X	NoFix	NC-SI AC Timing Specification Violations
16	X	Fixed	Removed - Refer to Section 4 “Programming Interface” in the Intel® 82598 10 Gigabit Ethernet Controller Datasheet for more details.
17	X	NoFix	With Lane Swap Enabled In 1G Mode, Link Is Not Automatically Achieved
18	X	NoFix	With LAN Swap Enabled, The DCA_ID Function_Number Register Value Is Incorrect
19	X	NoFix	First PCIe Packet Sent With Completer ID = 0
20	X	NoFix	Firmware Errata (NC-SI): Additional Multicast Packets May Be Forwarded To BMC
21	X	NoFix	Firmware Errata: (NC-SI): Some VLAN Tagged Packets May Not Be Forwarded To The BMC While Using VLAN Mode #3
22	X	NoFix	LED State Freezes On entry To D3 No Wake
23	X	NoFix	Link Might Not Be Achieved In A Back-to-back Configuration When Using Only Clause 37 Auto Negotiation
24	X	NoFix	Out-Of-Reset TAP Instruction Is Neither IDCODE nor BYPASS
25	X	NoFix	PCIe Reception of Completion That Should Be Dropped May Occasionally Result In Device Hang or Data Corruption
26	X	NoFix	Upstream TLP Message Corruption
27	X	NoFix	JTDO Output is Disabled During a HIGHZ Instruction
28	X	NoFix	Boundary Scan Bypass Register is Not Loaded in Capture-DR State
29	X	NoFix	JTDO is Not Connected to Boundary Scan Shift Register During an EXTEST Instruction
30	X	NoFix	TAP Instruction Changes Need to be Passed Through the Test-Logic- Reset State
31	X	NoFix	Backplane Auto-Negotiation Does Not Work Correctly in Loose Mode
32	X	NoFix	Missing Replay Due to Recovery During TLP Transmission

Table 1-4. Change Summary

No.	A1	Plans or Action	Errata
33	X	NoFix	LTSSM Moves from L0 to Recovery Only When Receiving TS1/TS2 on All Lanes
34	X	NoFix	TX CRC Must Be Enabled For Correct Flow Control Operation
No.		**Plans**	**Specification Clarifications**
1	X	NoFix	PCIe End Point Request of I/O Space After Initialization

1.4 Specification Changes

1. Removed - Refer to Section 2 “Signal Descriptions and Pinout List” in the Intel® 82598 10 Gigabit Ethernet Controller Datasheet for more details.

2. Removed - Refer to Section 2 “Signal Descriptions and Pinout List” in the Intel® 82598 10 Gigabit Ethernet Controller Datasheet for more details.

1.5 Errata

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

Problem: 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: No Fix: There are no plans to fix this erratum.

2. 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: No Fix: There are no plans to fix this erratum.

3. 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: No Fix: There are no plans to fix this erratum.

4. 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: No Fix: There are no plans to fix this erratum.

5. LINK_STATUS_CHECK Bit Is Not Self-Clearing On Read When In Link Down State

Problem: LINK_STATUS_CHECK 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: No Fix: There are no plans to fix this erratum.

6. 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: No Fix: There are no plans to fix this erratum.

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: No Fix: There are no plans to fix this erratum.

8. 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: No Fix: There are no plans to fix this erratum.

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: No Fix: There are no plans to fix this erratum.

10. 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 (GHECCR bits 21,18,9,6).

Status: No Fix: There are no plans to fix this erratum.

11. Link Not Achieved When Two 82598 Devices, Configured To KX/KX4 Mode, Are Connected Back-to-Back, One With Auto-Negotiation Enabled And The Other Without

Problem: When two 82598 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: No Fix: There are no plans to fix this erratum.

12. 82598 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: No Fix: There are no plans to fix this erratum.

13. PCIe 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: No Fix: There are no plans to fix this erratum.

14. 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: No Fix: There are no plans to fix this erratum.

15. NC-SI AC Timing Specification Violations

Problem: Specification calls for:

TCO_{min}=2.5 nS

T_{hold}=1 nS

82598 values are:

TCO_{min}=1.8 nS

T_{hold}=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: No Fix: There are no plans to fix this erratum.

16. Removed - Refer to Section 4 “Programming Interface” in the Intel® 82598 10 Gigabit Ethernet Controller Datasheet for more details.

17. 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 thru EEPROM , these are the writes to analog core regs 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: No Fix: There are no plans to fix this erratum.

18. 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 , PCI-E 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: No Fix: There are no plans to fix this erratum.

19. 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: No Fix: There are no plans to fix this erratum.

20. 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: No Fix: There are no plans to fix this erratum.

21. 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: No Fix: There are no plans to fix this erratum.

22. LED State Freezes On Entry To D3 No Wake

Problem: When transitioning to D3 no wake, LESs 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: No Fix: There are no plans to fix this erratum.

23. 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 will prevent the issue and clause 37 auto-negotiation will complete. This workaround is not specification compliant; it is recommended only in back-to-back configuration.

Status: No Fix: There are no plans to fix this erratum.

24. 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: No Fix: There are no plans to fix this erratum.

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 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: Ensure bit 13 “ACK/NACK_SCH”, EEPROM PCIe General Configuration Section, PCIe Init Configuration 3 - Offset 3 is set to 0b in the EEPROM image. This setting ensures ACKs are sent immediately on the PCIe bus and avoids a duplicate completion from the upstream component. This is the default setting for all 82598 EEPROM images. Setting this bit only avoids duplicate completions from being sent.

Status: No Fix: There are no plans to fix this erratum.

26. 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 under the following condition:

If using legacy interrupt PCIe mode or MSI/MSI-X mode with 64-bit message addressing while the combined LAN receive data rate of both ports is 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.

Implication: System hang.

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

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

3. If legacy interrupt PCIe mode or MSI/MSIx mode with 64-bit message addressing 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 110, 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: No Fix: There are no plans to fix this erratum.

27. 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: No Fix: There are no plans to fix this erratum.

28. 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: No Fix: There are no plans to fix this erratum.

29. JTDO is Not Connected to Boundary Scan Shift Register During an EXTEST Instruction

Problem: The 82598 does not connect the boundary scan shift register to the JTDO output during an EXTEST instruction. According to IEEE Std 1149.1-2001, "the EXTEST instruction shall select only the boundary-scan register to be connected for serial access between TDI and TDO in the Shift-DR controller state".

Implication: The tester cannot read the 82598 boundary scan shift register data during an EXTEST instruction. If multiple devices are chained in the board, the tester won't be able to load or read the boundary scan shift register for devices behind the 82598.

Workaround: Use the SAMPLE instruction to shift boundary scan data from JTDI to JTDO.

Status: No Fix: There are no plans to fix this erratum.

30. TAP Instruction Changes Need to be Passed Through the 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: No Fix: There are no plans to fix this erratum.

31. 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: No Fix: There are no plans to fix this erratum.

32. 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 required.

Status: No Fix: There are no plans to fix this erratum.

33. 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 82598 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 transmits 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.

Workaround: None required.

Status: No Fix: There are no plans to fix this erratum.

34. 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: No Fix: There are no plans to fix this erratum.

1.6 Specification Clarifications

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

Problem: 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.