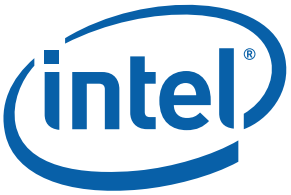

Intel® 82583V Gigabit Ethernet Controller Specification Update

July 2015
Revision 3.2

Revision History

Date	Revision	Description
July 2015	3.2	Added Erratum#15.
June 2014	3.1	Added Erratum#14.
October 2013	3.0	Added Errata #12 and #13.
June 2012	2.9	Revised Erratum #11.
June 2012	2.8	Added Erratum #11.
January 2012	2.7	Revised Erratum #8. Added Software Clarification #2.
July 2011	2.6	Added Erratum #10. Added Specification Change #3.
February 2011	2.5	Added Erratum #7, #8, and #9. Added Software Clarification #1. Added Specification Change #1 and #2.
March 2010	2.4	Added Erratum #6.
November 2009	2.3	Changed title of Specification Clarification #1.
August 2009	2.2	Added Erratum #5. Updated Section 1.3.
June 2009	2.1	Added Specification Clarification #1.
June 2009	2.0	Initial Public Release.
April 2009	1.0	Initial Release (Intel Confidential).

Note: This page intentionally left blank.

1.1 Introduction and Scope

This document applies to the 82583V GbE Ethernet Controller.

This document is an update to a published specification, the *Intel® 82583V GbE Controller Datasheet*. It is intended for use by system manufacturers and software developers. 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.

References to PCIe* in this document refer to PCIe Rev. 1.1 (2.5 GHz) x1.

1.2 Product Code and Device Identification

Product Code: WG82583V.

The following tables and drawings describe the identifying markings for the 82583V:

Table 1. Marking

Device	Stepping	Top Marking	Description
82583V	A1	WG82583V	Production (Lead Free)

Table 2. Device ID

82583V Device ID Code	Vendor ID	Device ID	Revision ID
82583V	0x8086	0x150C	N/A

Table 3. MM Numbers

Product	MM Number	Tray/Tape and Reel
WG82583V	903008	Tape and Reel; SLGVC
WG82583V	903072	Tray; SLGVD

1.3 Marking Diagram

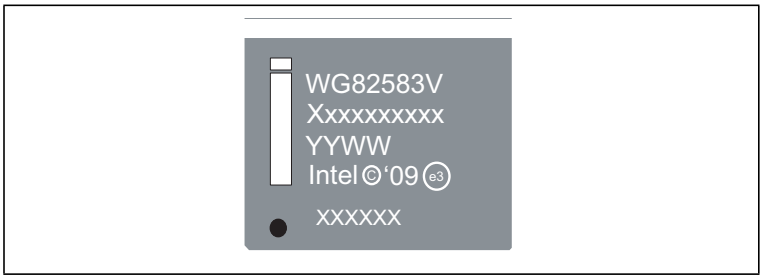

Figure 1. Top Marking Example With Identifying Marks

Notes:

Line 1: Marketing Name (WG82583V)
Line 2: TSMC Fab Lot Number "Xxxxxxxxx" or "Xxxxxxxx.x"
Line 3: Assembly Date Code "YYWW"
Line 4: "INTEL", © Copyright including two number date code, circled "e3" lead-free mark
Line 5: Country of Origin

1.4 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 4 for a description.

Table 4. Terms, Codes, Abbreviations

Name	Description
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.
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	Typos, errors, or omissions from the current published specifications. These changes will be incorporated in the next release of the specifications.
Name	**Description**

Yes or No	If the errata applies to a stepping, "Yes" is indicated for the stepping (for example: "A0=Yes" indicates errata applies to stepping A0). If the errata does not apply to stepping, "No" is indicated (for example: "A0=No" indicates the errata does not apply to stepping A0).
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.
Red Change Bar / or Bold	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

1.5 Hardware Clarifications, Changes, Updates and Errata

See Section 1.4 for an explanation of terms, codes, and abbreviations.

Table 5. Summary of Hardware Clarifications, Changes and Errata; Errata Include Steppings

Specification Changes	Status
1. PBA Number Module — Word Address 0x8-0x9	N/A
2. Updates to PXE/iSCSI EEPROM Words	N/A
3. TARC Register Setting Change	N/A
Specification Clarifications	**Status**
1. PCIe Completion Timeout Mechanism Compliance	N/A
Documentation Updates	**Status**
None.	N/A
Errata	**Status**
1. 10BASE-T IEEE-Specified Harmonic Content Level Issue	A1=Yes; No Fix
2. 100BASE-TX Marginal Rise/Fall Time Performance	A1=Yes; No Fix
3. Revision ID is Zero for the 82583V A1 Stepping	A1=Yes; No Fix
4. Missing Interrupt Following ICR Read	A1=Yes; No Fix
5. PCIe: Missing Replay Due to Recovery During TLP Transmission	A1=Yes; No Fix
6. PCIe: Completion with UR/CA Status Causes Unexpected Completion and Completion Timeout Errors to be Reported	A1=Yes; No Fix
7. Tx Data Corruption When Using TCP Segmentation Offload	A1=Yes; No Fix
8. Disabling Receive Logic During Packet Reception Might Cause Receive Failures	A1=Yes; No Fix
9. PCIe Hang or System Crash when L0s is Enabled in the Upstream PCIe Switch Port	A1=Yes; No Fix
10. Packets Received With an L2 + L3 Header Length Greater than 256 Bytes Can Incorrectly Report a Checksum Error	A1=Yes; No Fix
11. Device Transmit Operation Might Halt in TCP Segmentation Offload (TSO) Mode When Multiple Requests (MULR) Are Enabled	A1=Yes; No Fix
12. Dropped Rx Packets	A1=Yes; No Fix
13. PCIe: Common Mode Voltage Shift During L1 Exit	A1=Yes; No Fix
14. TimeSync: Value Read from SYSTIM Registers Might be Incorrect	A1=Yes; No Fix
15. Incorrect 64-bit Statistics Counter Value	A1=Yes; No Fix

1.5.1 Specification Changes

1. PBA Number Module — Word Address 0x8-0x9

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

Note that 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 0x08	Word 0x09
G23456-003	FAFA	Pointer to PBA Block

The PBA block is pointed to by word 0x09.

Word Offset	Description	Reserved
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 0x08 and 0x09. 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

Word 0x30 is 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. TARC Register Setting Change

In the Transmit Arbitration Count register of Tx Queue 0 (TARC[0]:x3840), bit 26 should be set to 1b for proper 82583 operation.

1.5.2 Specification Clarifications

1. PCIe Completion Timeout Mechanism Compliance

If the latency for PCIe completions in a system is above 21 ms and PCIe completion timeout mechanism is enabled, there may be unpredictable system behavior.

The 82583V complies with the PCIe 1.1 specification for completion timeout mechanism. The PCIe 1.1 specification provides a timeout range between 50 µs to 50 ms with a strong recommendation that it be at least 10 ms. The 82583V uses a range of 21-42 ms.

The completion timeout value in a system MUST be above the expected maximum latency for completions in the system in which the 82583V is installed. This will ensure that the 82583V receives the completions for the requests it sends out, avoiding a completion timeout scenario. If the latency for completions is above 21 ms this can result in the device timing out prior to a completion returning. In the event of a completion timeout, per direction in the PCIe specification the device assumes the original completion is lost, and resends the original request. In this condition, if the completion for the original request arrives at the 82583V devices, this will result in two completions arriving for the same request, which may cause unpredictable system behavior.

Therefore, if the PCIe completion latency for a system cannot be guaranteed to be lower than 21 ms, the PCIe completion timeout mechanism should be disabled by setting GCR.Disable_timeout_mechanism.

For more details on Completion Timeout operation in the 82583V refer to the *Intel® 82583V GbE Controller Datasheet*.

1.5.3 Documentation Changes

None.

1.5.4 Errata

1. 10BASE-T IEEE-Specified Harmonic Content Level Issue

Problem: On some board designs, the 82583V might not meet the IEEE specification (1411.10.03) that states that the harmonic content is to be at least 27 dB below the 10 MHz fundamental frequency.

Implication: IEEE conformance is marginal. There is no impact on system level performance; however, care should be taken to verify the impact of radiated EMI (Electromagnetic Interference) on system-level EMI tests.

Workaround: There is no silicon/firmware/software workaround; however, using short low-resistance traces (less than four inches and without a LAN switch) can help reduce harmonic content.

Status: A1=Yes; No Fix

2. 100BASE-TX Marginal Rise/Fall Time Performance

Problem: The 82583V rise/fall time has been marginal compared to the IEEE specification (5 ns).

Implication: IEEE conformance is marginal. Depending on system topology (LAN switch/no LAN switch), MDI trace lengths, and configuration (docked/undocked), the 100BASE-TX rise/fall time might not meet the IEEE specification. Note that there is no impact on system level performance.

Workaround: There is no silicon/firmware/software workaround. However, using short low-resistance traces (less than four inches and without a LAN switch) can help improve rise time. In some designs it's better to use two 82583V's instead of a LAN switch.

Status: A1=Yes; No Fix

3. Revision ID is Zero for the 82583V A1 Stepping

Problem: The Revision ID field in the PCIe configuration space is zero instead of one.

Implication: Error counters may not be accurate.

Workaround: None.

Status: A1=Yes; No Fix

4. Missing Interrupt Following ICR Read

Problem: If the Interrupt Cause Register (ICR) is read when at least one bit is set in the interrupt mask register and INT_ASSERTED is set to 0b, a new interrupt event occurring on the same clock cycle as the ICR read is ignored.

Implication: Missed interrupts leading to delays in responding to interrupt events. Specifically, this can cause a delay in processing a received packet.

Typically, the ICR is only read in response to an interrupt so this problem does not occur. However, when using legacy interrupts and sharing interrupts between devices, software might poll all the devices to find the source of the interrupt, including those devices that did not assert an interrupt. There might also be other situations in non-Intel drivers where ICR is polled even when no interrupt has been asserted.

Workaround: If reading ICR when there is no active interrupt cannot be avoided, clear the mask register (by writing 0xFFFFFFFF to IMC) before reading ICR. Note that in this case the ICR is cleared when read even if INT_ASSERTED is set 0b.

Status: A1=Yes; No Fix

5. 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 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=Yes; No Fix

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

Problem: When the 82583V 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 82583V 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 register.

Status: A1=Yes; No Fix

7. Tx Data Corruption When Using TCP Segmentation Offload

Problem: When using TSO, a situation can occur where a PCIe MRd request is repeated with the same address, resulting in data corruption. At the end of the TCP packet, the Tx DMA hangs because the length doesn't match. This can only occur when the following are true:

- The first buffer of the packet is larger than [3 * (max_read_request - 4)].
- There is a 4 KB boundary within 64 bytes following the end of the header bytes in the buffer

Implication: Possible data corruption since a TCP packet is transmitted containing the wrong data but with the correct checksum.

Data transmission halts as the Tx DMA module enters a hang state.

Workaround: The failure can be avoided by ensuring at least one of the following:

- The buffer containing the headers should not be larger than [3 * (max_read_request - 4)]. To meet this requirement even for the minimum value of 128 bytes for max_read_request, the buffer should not be larger than 372 bytes.
- The alignment of the buffer containing the headers should be such that there is no 4 KB boundary within 64 bytes following the end of the header bytes. Assuming standard Ethernet/IP/TCP headers of 54 bytes, this means that the buffer should not start 54-118 bytes before a 4 KB boundary. For example, 128-byte alignment for this buffer could be used to fulfill this condition.

This problem has not been reported when using an Intel Linux* or Windows* drivers. Current analysis shows it is very unlikely for a situation to exist that would cause the 82583 to be at risk for the errata when using the Intel Linux or Windows drivers.

Status: A1=Yes; No Fix

8. Disabling Receive Logic During Packet Reception Might Cause Receive Failures

Problem: Clearing the RCTL.EN bit while a packet is being received might cause a failure of the receive packet buffer pointer mechanism.

Implication: Unpredictable behavior and/or hang of receive packet buffer control logic.

Workaround: Use one of the following options:

- Do not clear RCTL.EN. Operations are halted by using CTRL.RST and then re-initializing the device.

-OR-

- In order to disable receive operation, all receive filters should first be disabled such that no packets are stored in the packet buffer. All of the following must be true before clearing RCTL.EN.
- RCTL.UPE, RCTL.MPE, RCTL.VFE, RCTL.PMCF, and RCTL.BAM are 0b.
- RCTL.DPF is 1b.
- The Address Valid bit is 0b in all 16 RAH registers.
- The Multicast Table Array is clear.

Once RCTL.EN has been cleared, the register values can be restored to the normal values.

-OR-

In order to disable receive operation, the PHY should be held in reset such that no packets are stored in the packet buffer. Use the following sequence:

- Set CTRL.PHY_RST.
- Clear RCTL.EN.
- Clear CTRL.PHY_RST.

Note: This option should not be used if there is a critical manageability session in progress. For example, if the MANC.KEEP_PHY_LINK_UP bit is 1b.

Intel drivers implemented the workaround starting from Release 16.4 (Linux e1000e v1.4.4).

Status: A1=Yes; No Fix

9. PCIe Hang or System Crash when L0s is Enabled in the Upstream PCIe Switch Port

Problem: When the 82583V's PCIe receive lane goes in and out of the L0s state, there might be failures in the PCIe logic that result in a PCIe hang or a PCIe uncorrectable error. These issues have been reported when the 82583V is connected to the Intel® 5 Series Express Chipset ports, but they do not occur on all PCIe switch ports.

Problematic connections to the 82583V can be identified by clearing and polling the Correctable Error Status register in the PCIe config space on the 82583V while passing LAN traffic. Repeated assertions of the Bad DLLP and/or Replay Timer Timeout bits more than once per minute indicate a problematic situation in which the following workaround should be applied.

Implication: Either a PCIe hang that requires a PCIe reset assertion to recover, or a system crash due to an uncorrectable error reported by the 82583V.

Workaround: Disable L0s in the Intel® 5 Series Express Chipset port connected to the 82583V. Note there is no requirement to disable ASPM-L1.

Status: A1=Yes; No Fix

10. Packets Received With an L2 + L3 Header Length Greater than 256 Bytes Can Incorrectly Report a Checksum Error

Problem: L2/L3 packets with long/multiple next header extensions incorrectly report a receive checksum error when the length from a Destination Address (DA) to the beginning of the TCP/UDP header is greater than 256 bytes.

Implication: A receive checksum error can incorrectly be reported by the 82574, even if there is no checksum error.

Workaround: When the driver receives a packet with a checksum error reported by hardware, software should check the L2/L3 header length. If the L2/L3 header length is 256 bytes or greater, software should verify the checksum.

The Intel Windows and Linux drivers address this issue by passing packets with bad checksums to the network stack for further examination.

Status: A1=Yes; No Fix

11. Device Transmit Operation Might Halt in TCP Segmentation Offload (TSO) Mode When Multiple Requests (MULR) Are Enabled

Problem: The device transmit flow stops and the 82583V hangs when operating in TSO with MULR enabled.

Implication: When operating in TCP segmentation offload mode and with multiple request enabled, the following workaround must be in place, or the transmit flow might stop unexpectedly.

Workaround: The software device driver must ensure that the first descriptor points to the (L2+L3+L4) header and at least two bytes of the data (payload). This has been implemented in the Intel Windows drivers in Release 16.4. Linux (e100e) has had this workaround implemented since the 82583V has been supported in the software device driver.

Status: A1=Yes; No Fix

12. Dropped Rx Packets

Problem: When receiving a mixture of Ethernet packets, some of which should pass the L2 address filtering and some of which should not, a few of the packets that should pass are actually dropped. The rate of dropped packets might vary, but it is on the order of 1 in 30,000 with random traffic.

Implication: A small number of missing Rx packets. On TCP connections, this might not be noticeable. On UDP connections, the application layer might be affected.

Workaround: Do one of the following:

- Set NVM word 0x0F, bit 1. This is the preferred workaround since it does not significantly increase the power consumption when the device is idle. This workaround is implemented in the 82583V Dev_Starter NVM v2.1.4. Contact your Intel representative to obtain updated NVM images.

Status: Clear GCR.L1_act_without_L0s_rx and disable ASPM L0s. These settings have the side effect of preventing the device from entering the ASPM L1 state, thereby eliminating the power savings when the device is idle. A1=Yes; No Fix

13. PCIe: Common Mode Voltage Shift During L1 Exit

Problem: When using the PCIe ASPM L1 state, there might be a shift in the common mode voltage on the 82583V PCIe Tx lines when exiting the L1 state. This shift only occurs if the 82583V initiates the L1 exit and does not occur if the upstream device initiates the L1 exit.

Implication: In rare cases, the voltage shift on the 82583V PCIe Tx lines can cause the upstream device's receiver to incorrectly detect some bit values at a later time as it responds to the voltage shift. This could cause a failure to exit the L1 state and a Surprise Down error on the PCIe link.

Workaround: Either disable ASPM L1 or set bit 10 of word 0x1E in the NVM.

The NVM change is implemented in NVM release 2.1.4.

Status: A1=Yes; No Fix

14. TimeSync: Value Read from SYSTIM Registers Might be Incorrect

Problem: Due to a synchronization error, the value read from SYSTIML/SYSTIMH might be incorrect. The probability of failure depends on several factors, but it is always less than 50%.

Implication: It is problematic to use the SYSTIM clock as a system clock because the read values are unreliable.

Workaround: Do one of the following:

To get a reliable value, read the SYSTIM registers twice and check that both of the following are true:

- The difference between the 2 values is a multiple of TIMINCA.inc_value..
- The difference between the 2 values is positive and the magnitude is reasonable for consecutive reads on the system. For example, 100 µs might be a good threshold.

If either of these conditions is false, discard the value and try again.

The Intel e1000e driver included this workaround starting with v3.0.4.

Status: A1=Yes; No Fix

15. Incorrect 64-bit Statistics Counter Value

Problem: As documented in the datasheet, the 64-bit statistics counters are cleared when reading the upper 32-bit register. As a result, any increments to the counter that occurred between reading the lower 32-bit register and reading the upper 32-bit register are lost.

This applies to the following statistics counters:

Good Octets Received

Good Octets Transmitted

Total Octets Received

Total Octets Transmitted

Implication: The counter values could be slightly lower than the actual number of octets received or transmitted.

Workaround: To minimize the probability of this issue occurring, read the counters as infrequently as possible. (At 1 Gb/s, the octet counters cannot saturate in less than 4675 years.) Also, ensure that the upper 32-bit register is read as soon as possible after reading the lower 32-bit register.

To prevent any loss of information, ignore the upper 32-bit register and treat the lower 32-bit register as a 32-bit counter that is not cleared by read and wraps to zero when it reaches its maximum value. Read this counter at least once every 30 seconds, and maintain the high portion of the counter in software by incrementing it each time the hardware counter value has wrapped since the previous read.

Status: A1=Yes; No Fix

2. Software Clarifications

Table 2-1. Summary or Software Clarifications

Software Clarifications	Status
1. While in TCP Segmentation Offload, Each Buffer is Limited to 64 KB	N/A
2. Serial Interfaces Programmed by Bit Banging	N/A

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

Problem Description:

The 82583V 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 64 KB 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.

2. Serial Interfaces Programmed by Bit Banging

Problem Description:

When bit banging on a serial interface (such as SPI, I2C, 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.

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