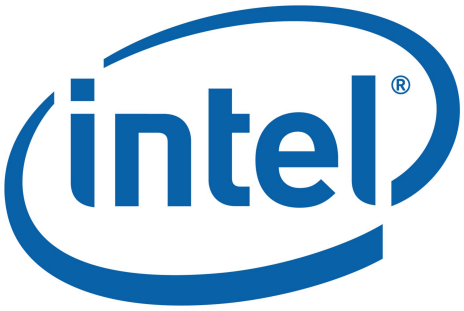


Intel® 82599 10 GbE Controller Datasheet

Networking Division (ND)

PRODUCT FEATURES

General

- Dual port 10 GbE device or Single Port device (82599EN)
- Serial Flash Interface
- 4-wire SPI EEPROM Interface
- Configurable LED operation for software or OEM customization of LED displays
- Protected EEPROM space for private configuration
- Device disable capability
- Package Size - 25 mm x 25 mm
- Networking
- Complies with the 10 Gb/s and 1 Gb/s Ethernet/802.3ap (KX/KX4/KR) specification
- Complies with the 10 Gb/s Ethernet/802.3ae (XAUI) specification
- Complies with the 1000BASE-BX specification
- Complies with the IEEE 802.3x 100BASE-TX specification
- Support for jumbo frames of up to 15.5 KB
- Auto negotiation Clause 73 for supported mode
- CX4 per 802.3ak
- Flow control support: send/receive pause frames and receive FIFO thresholds
- Statistics for management and RMON
- 802.1q VLAN support
- TCP segmentation offload: up to 256 KB
- IPv6 support for IP/TCP and IP/UDP receive checksum offload
- Fragmented UDP checksum offload for packet reassembly
- Message Signaled Interrupts (MSI)
- Message Signaled Interrupts (MSI-X)
- Interrupt throttling control to limit maximum interrupt rate and improve CPU usage
- Receive packet split header
- Multiple receive queues (Flow Director) 16 x 8 and 32 x 4
- 128 transmit queues
- Receive header replication
- Dynamic interrupt moderation
- DCA support
- TCP timer interrupts
- Relaxed ordering
- Support for 64 virtual machines per port (64 VMs x 2 queues)
- Support for Data Center Bridging (DCB)(802.1Qaz, 802.1Qbb, 802.1p)

Host Interface

- PCIe Base Specification 2.0 (2.5GT/s) or (5GT/s)
- Bus width — x1, x2, x4, x8
- 64-bit address support for systems using more than 4 GB of physical memory

MAC Functions

- Descriptor ring management hardware for transmit and receive
- ACPI register set and power down functionality supporting D0 and D3 states
- A mechanism for delaying/reducing transmit interrupts
- Software-controlled global reset bit (resets everything except the configuration registers)
- Eight Software-Definable Pins (SDP) per port
- Four of the SDP pins can be configured as general-purpose interrupts
- Wake up
- Ipv6 wake-up filters
- Configurable flexible filter (through EEPROM)
- LAN function disable capability
- Programmable memory transmit buffers (160 KB/port)
- Default configuration by EEPROM for all LEDs for pre-driver functionality
- Support for SR-IOV

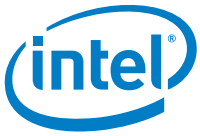
Manageability

- Eight VLAN L2 filters
- 16 flex L3 port filters
- Four Flexible TCO filters
- Four L3 address filters (IPv4)
- Advanced pass through-compatible management packet transmit/receive support
- SMBus interface to an external manageability controller
- NC-SI interface to an external manageability controller
- Four L3 address filters (IPv6)
- Four L2 address filters

March 2016
Revision 3.3
331520-004

Revision History

Rev	Date	Comments
0.5	May 2008	Initial release (Intel Confidential). This release contains advanced information.
0.6	October 2008	Updated to reflect developments, corrections.
0.75	February 2009	Major update (all sections) — Reflects latest device developments and corrections.
0.76	March 2009	Updated the following sections: Programming Interface, Manageability, NVM, Initialization, Power Management, and Interconnects.
1.0	March 2009	Major update (all sections) — Reflects latest device developments and corrections.
1.5	May 2009	Major update (all sections) — Reflects latest device developments and corrections.
1.9	June 2009	Minor update (all sections) — Reflects latest device developments and corrections.
2.0	July 2009	Initial release (Intel Public).
2.01	July 2009	Added x8 lane note to Section 1.2.1.
2.1	October 2009	• Changed jumbo frame size from KB to bytes (all occurrences). • Changed "XTAL_25_MODE" to "RSVDAC6_VCC". • Updated section 2.1.4 (changed type from T/s to O). • Added F20 and H7 to the table in section 2.1.12. • Changed "OSC_FREQ_SEL" to "RSVDAC6_VCC". • Corrected PCIe versions to "PCIe V2.0 (2.5GT/s or 5GT/s)". • Updated the table in section 3.2.7.2.1 (added text to the vendor ID column). • Updated the jumbo frame calculations in sections 3.7.7.3.3, 3.7.7.3.4, and 3.7.7.3.5. • Added section 4.6.13 "Alternate MAC Address Support". • Updated section 5.2.2 "Auxiliary Power Usage". • Added text to section 6.3.6 "Alternate Ethernet MAC Address - Word Address 0x37". • Updated Table 6.1 (added /1 to row 4). • Updated section 6.4.5.8. • Added L34TIMIR register name to the Queue Enable bit in section 8.2.3.7.19. • Corrected the D10GMP and LMS bit descriptions in section 8.2.3.22.19. • Corrected the LP AN page D low bit description in section 8.2.3.22.23. • Updated the PRDC bit description is section 8.2.3.23.75. • Changed the bit length (31 to 8 to 31 to 0) to the table heading in section 8.2.3.25.12. • Updated the Restart_AN bit description in section 8.2.3.23.22. • Corrected the bit 8 description in section 9.3.7.1.4. • Updated section 10.2.2.2.4 (bits RAGEN and TFOENODX; read/write value). • Added text "Jumbo packets above 2 KB . . . to Filtering exceptions in section 10.3.1. • Correct the Buffer Length (byte 1) description in section 10.5.3.8.2. • Changed the title of table 11.6, 11.7, and 11.8. • Changed Watts to mW in the Power row of table 11.6. • Updated the power values in table 11.7 and 11.8. • Updated the mechanical package drawing in section 11.5.4.



Rev	Date	Comments
2.1 (cont.)		• Added power summary table (table 11.6). • Updated section 1.2.1, 3.1.4.5.3, 5.2.5.3.2 (note), and 6.4.5.2.2 (bit descriptions). • Updated bit descriptions for MRQE, RRM, TDRM, and PRDC. • Updated tables in sections 10.3.1, 10.5.1.13.1, and 10.5.2.1.5. • Added Single Port Power table (table 11.8) • Added SFI optics references. • Changed the bit name in section 5.3.1 from APM Wake Up (APM) to APM Enable (APME).
2.2	January 2010	• Updated BX4 spec reference (changed 1000BASE-BX4 to 10GBASE-BX4). • Added jumbo frame KB value to note after Table 1.2. • Added new section 1.6.2 “Byte Count”. • Added BX4 and CX4 references. • Updated the note in section 2.1.8. • Updated pin name (SDP0_6) in section 2.1.10. • Updated section 3.1.4.5.3 (Relaxed Ordering); last paragraph. • Added BX4 info to section 3.7. • Added new BX4 section (3.7.1.5). • Updated section 3.7.4.4 (link speed). • Updated section 3.7.7.3.3 and 3.7.7.3.5 (jumbo frame values). • Added note after table 3.27 (IPG pacing feature). • Added VFLR note after table 4.6. • Added BX4 reference to section 4.6.4.2. • Added IPG pacing feature note at the end of section 4.6.11.4. • Added jumbo frame value to section 4.6.11.4 and table 4.9 (KB value). • Changed the bit name in section 5.3.1 from APM Wake Up (APM) to APM Enable (APME). • Updated the note in section 5.2.5.3.2 (DMA completions). • Changed GIO Master Disable to PCIe Master Disable (throughout entire EAS). • Changed GIO Master Enable Status to PCIe Master Enable Status (throughout entire EAS). • Updated bullet list in section 5.3.1 and added WKEN bit note at the end of section 5.3.1. • Swapped fields “Possible Len/LLC/SNAP Header” and “Possible VLAN Tag” in sections 5.3.3.1.4. through 5.3.3.1.7 and sections 5.3.3.2.1 and 5.3.3.2.2. • Updated section 6.3.5.4 (changed GIO to PCIe; bit 3 description). • Changed the default setting for CDQMH in section 6.3.6.5 to 0x1404. • Updated section 6.3.5.22 (MSIX and CDO bit definitions). • Removed old 6.3.6.7 section title (Spare 0/1 - Offset 0x05). • Added 5-tuple note to section 7.1.2.5. • Removed sub-bullet under 4-bit RSS Type field in section 7.1.2.8. • Removed TcpIPv6Ex, IPv6Ex and UdpIPV6E info from section 7.1.2.8.1. Updated TCP segment bullet and IPv4 packet sub-bullet in section 7.1.2.8.1. • Updated table 7.10 (Destination Address/Port and Source Address/Port; first row). • Changed RXCTL to DCA_RXCTL[n] under table 7.15 (Packet Buffer Address (64) paragraph). • Changed descriptors per queue value from 64 to 40 in section 7.2.3.3. • Updated figure 7.39 (changed BCN to transmit rate scheduler). • Updated SecTag bullet description in section 7.8.4. • Updated the APME bit description in section 8.2.3.2.9. • Updated the MRQE bit description in section 8.2.3.7.12. • Added a note to the Queue Enable bit description in section 8.2.3.7.19. • Removed the note from section 8.2.3.8.5.

Rev	Date	Comments
2.2 (cont.)		• Changed GIO to PCIe in section 8.2.2.1.1 (bit 2 description). • Updated the RRM bit description in section 8.2.2.11.1. • Updated SECTX_OFF_DIS and ECC_TXERR bit descriptions in section 8.2.2.13.2. • Updated SECRX_OFF_DIS and ECC_RXERR bit descriptions in section 8.2.2.13.7. • Added a note to the KX_support bit description in section 8.2.2.23.22. • Updated the PRDC bit description in section 8.2.2.24.75. • Updated bit 4 description (WKEN) in section 8.2.2.25.1. • Added a VF Mailbox note to section 8.3.5.1.5. • Changed RW to RO in section 9.3.10.13 title. • Updated the Filters table in section 10.3.1. • Added note to section 10.5.1.13.1 (TCO Mode reference). • Updated the TCO Mode table in section 10.5.2.1.4. • Updated section 11.3.1.1 (rise time relationships). • Added Single Port Power table (Table 11.8) • Changed all SFI Optics references to unconditional text (now exposed to external customers). • Added single port power numbers (table 11.8). • Added BX4 to section 11.4.4. • Changed crystal load capacitance to 27 pF.
2.3	April 2010	• Updated section 3.7.7.1.4 (changed TXOFF to TC_XON). • Changed VMBMEM to VFMBMEM. • Updated section 5.3.2 (last paragraph). • Added a note after the table in section 6.4.2.3. • Updated section 8.2.3.5.13 - changed VT31 to VT32. • Changed all occurrences of SPD to SDP in section 8.2.3.1.4. • Updated the TC_XON field description. • Updated Table 9.6 - Address Space (low register for 64-bit memory BARs) description. • Added recommended and minimum EEPROM sizes to section 12.6.2.
2.4	September 2010	The following was updated and or changed for this release: • Section 4.6.11.3.1 (changed MRQC.VT_Ena to MTQC.VT_Ena). • Section 4.6.11.3.3 (changed “via setting RTTDQSEL first for the lowest indexed queue of a pool” to “via setting RTTDQSEL first for the pool index”). • Section 4.6.11.6.1 (updated first step under “Refill Credits”). • Section 4.6.12 (updated Security Offload description). • Section 6.3.2.3 (APM Enable Port 1/0 bit descriptions). • Section 6.3.3 (PBA Number Module — Word Address 0x15-0x16). • Section 6.3.8 (Checksum Word Calculation (Word 0x3F)). • Section 6.4.5.5 (PCIe Control 1 —Offset 0x04). • Section 6.4.5.8 (PCIe Control 3 — Offset 0x07). • Section 7.1.2.3 (L2 Ethertype Filters, step 9). • Section 7.1.2.5 (L3/L4 5-tuple Filters, removed “If the packet is mirrored or replicated . . .”. • Section 7.1.2.7 (Flow Director Filters, removed “In case of mirroring or replication. . .”. • Section 7.2.5.3 (added a note to Tx SCTP CRC Offload). • Section 8.2.3.11.4 (TXDQ_IDX bit description). • Section 8.2.3.11.5 (register RTTDT1C description). • Section 8.2.3.10.3 (VT bit description). • Section 8.2.3.8.2 (VET bit description). • Section 8.2.3.11.9 (DCB Transmit Descriptor Plane T2 Config bit descriptions).

Rev	Date	Comments
2.4 (cont.)		• Section 8.2.3.13 (updated Security Offload description). • Section 8.2.3.13.5 (updated MINSECIFG and SECTXDCB bit descriptions). • Section 8.2.3.21.22 (updated Rx Queue Index bit description).
2.5	November 2010	The following was updated and or changed for this release: • Section 2.1.8 (changed pull-up to pull-down in the note following the table). • Section 6.4.2 (updated bit 15 bit description). • Section 7.1.2.2 (updated RSS queues reference). • Section 7.1.11 (updated IPv6 filter description). • Section 7.7.2.2 (added a note about using advanced transmit descriptors in DCB mode). • Section 8.2.3.6.1 (added notation about the EICR register). • Section 8.2.3.8.4 (updated the RQPL bit description). • Section 8.2.3.25.3 (updated the WUS register description). • Section 9.3.10.7 (updated bit description for bits 9:4). • Section 11.4.5.1 (changed load capacitance value to 20 pF). • Added new Table 12-1 (Microstrip Trace Dimensions for SFI Using Different Dielectric Materials). • Section 12.12.1.1 (updated the part numbers for recommended crystals). • Updated Figures 12-20 and 12-21 (changed 10 KΩ to 100 Ω). • Section 13.11.4 (updated the maximum static normal load value).
2.6	December 2010	• Updated section 3.4.7 EEPROM Recovery (changed Data Byte value from 0xD8 to 0xB6). • Added reference clock specifications note to section 11.4.3. • Updated table 11.25 (changed duty cycle values and added p-noise for non-high serial speed parameter. • Added new figure 11.16 (refclk phase noise as a function of frequency).
2.7	April 2011	• Updated Table 1.5 (Flow Director Filters). • Revised section 2.1.13 (LAN1_DIS_N and LAN1_DIS_N name and function description). • Revised section 3.1.4.6.1 (changed “two credits” to “four credits” under “Rules for FC updates”). • Revised table 4.4 (LAN Disable Strapping Pins row; removed “X” from PCIe PERST# and In-band PCIe Reset • columns). • Added SECTXMINIFG.SECTXDCB field reference to sections 4.6.11.3.1 4.6.11.3.2. • Revised section 6.4.5.11 (PCIe Dummy Device ID — Offset 0x0A; changed default value to 0x10A6). • Revised section 7.1.2.7 (Flow Director Filters). • Revised table 7.5 (Flow Director Filters). • Revised section 7.1.2 (added cross reference to last bullet). • Revised section 7.1.2.1 (Queuing in a Non-virtualized Environment). • Revised section 7.1.2.2 (Queuing in a Virtualized Environment). • Revised table 7.19 (Receive Errors (RDESC.ERRORS) Layout). • Revised section 7.1.7.1 (Fetch On Demand; removed Figure 12 reference). • Revised section 8.2.3.21.1 (Flow Director Filters Control Register; bits 1:0 description). • Added a FTFT register note to section 8.2.3.25.12. • Revised the tables at the end of section 8.2.3.24.9 (Flexible Host Filter Table Registers — FHFT).

Rev	Date	Comments
2.7 (cont.)		• Revised section 8.2.3.22.8 (MAC Core Control 0 Register; changed MDCSPD default value to 1b). • Revised section 8.2.3.8.6 (Receive Descriptor Control — RXDCTL[n]; corrected bit assignments).
2.71	September 22, 2011	• Section 3.1.3.1.2. Case 1 table updated. Tag ID 30 information added. • Section 6.3.5.8, PCIe Control 3 — Offset 0x07. PREFBAR, Bit 14 exposed. • Section 6.2.10, Software Reserved Word 16 — Alternate SAN MAC Block Pointer — Word Address 0x27 and Section 6.2.11, Software Reserved Word 17 — Active SAN MAC Block Pointer — Word Address 0x28 updated. • Table 6-7, Usable Flash_Size. Exposed in Datasheet. • Section 7.1.2.7.4; note at the end of section updated. See phrase "RXPBSIZE[0...3] to 0x18000 (96 K) and RXPBSIZE[4...7] to zero." • Section 8.2.3.7.4, Packet Split Receive Type Register — PSRTYPE[n] (0x0EA00 + 4*n, n=0...63 / 0x05480 + 4*n, n=0…15; RW), Additional bullet added to note. See: "PSR_type4 should be set to enable RSC, regardless header split mode." • Table 9-4, Bit 3 description field updated. New text: "This bit should be set only on systems that do not generate prefetchable cycles." • Table 10-1, "Clear Ethernet MAC Address" command removed from list of supported commands. This command no longer exists in specification. Also, "Set Ethernet MAC Address" command corrected; now called "Set MAC Address".
2.72	October 18, 2011	• Section 6.2.7, Alternate Ethernet MAC Address — Word Address 0x37. In table, word "port" changed to "function". • Section 6.2.11.1, Active SAN MAC Address Block. Added this section. Was missing from Datasheet. • Section 6.4.4.7, NC-SI Configuration Offset 0x06. Updated. Description added to bits 4:0.
2.73	December 7, 2011	• Section 2.1.16. Pin name assignments corrected for SDP0[7:0] and SDP0[7:0]. • Section 5.2.5.3.2. Sentence changed in section. See the new wording: "The driver then reads the change made to the PCIe Master Disable bit and then polls the PCIe Master Enable Status bit." • Section 5.2.4.2. Sentence changed. See the revised sentence: "When a XAUI interface is in low-power state, the 82599 asserts the respective SDP pin to enable an external PHY device to power down as well." • Section 5.2.5.4.1. Sentence updated. See the revised sentence: "Note that the state of the **SDP pins** is undefined once power is removed from the device." • Section 6.2.9, Software Reserved Word 15 — Ext. Thermal Sensor Configuration Block Pointer — Word Address 0x26. Section added: Pointer to External Thermal Sensor Configuration block. • Figure 7-37 updated to correct rendering problem in figure. • Section 7.2.1.2. Under discussion of PAYLEN field in FCoE; revised sentence. New text is: "In FCoE TSO offload, the PAYLEN field defines the FC payload size." • Table 11-25, SerDes Crystal Specifications. In table, Shunt Capacitance recommendation changed from 6[pf] maximum to 7[pf] maximum. • Section 8.2.3.1.4. Table footnote (#2) added. • Fix needed in flow control Statistic Counters:
2.74	February 03, 2012	• Section 1.2, Product Overview. Note explaining single and dual port information context added. • Section 7.1.5, Legacy Receive Descriptor Format; see VP (VLAN Packet subsection. This text has been updated. New text is: "When set, the VP field indicates that the incoming packet's type is a VLAN (802.1q, matching the VLNCTRL.VET). If the RXDCTL.VME bit is set as well, then active VP field also means that the VLAN has been stripped from the packet to the receive descriptor. See further description of 802.1q VLANs in Section 7.4."

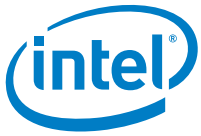
Rev	Date	Comments
2.74 (cont.)		• Section 8.2.3.23.13, Priority XON Transmitted Count — PXONTXC[n] (0x03F00 + 4*n, n=0...7; RC) - New description for XONTXC field - "Number of XON packets transmitted per TC. Sticks to 0xFFFF". • Section 8.2.3.23.14, Priority XON Received Count — PXONRXCNT[n] (0x04140 + 4*n, n=0...7; RC) - New description for XONRXC field - "Number of XON packets received per UP. Sticks to 0xFFFF". • Section 8.2.3.23.15, Priority XOFF Transmitted Count — PXOFFTXCNT[n] (0x03F20 + 4*n, n=0...7; RC) - New description for XOFFTXC field - "Number of XOFF packets transmitted per TC. Sticks to 0xFFFF". • Section 8.2.3.23.16, Priority XOFF Received Count — PXOFFRXCNT[n] (0x04160 + 4*n, n=0...7; RC) - New description for XOFFRXC field - "Number of XOFF packets received per UP. Sticks to 0xFFFF". • Section 8.2.3.27.7, PF VF Receive Enable — PFVFRE[n] (0x051E0 + 4*n, n=0...1; RW) and Section 8.2.3.27.8, PF VF Transmit Enable — PFVFTE[n] (0x08110 + 4*n, n=0...1; RW). Text changed in both descriptions. New text is: "Respective bits per VF are reset on VFLR, BME bit clear or on VF software reset." • Tables modified: Table 11-6, Power Summary for Dual Port Devices (82599ES, 82599EB) and Table 11-7, Power Summary for Single Port Device (82599EN). These tables now provide clear power summaries for single port and dual port devices.
2.75	April 24, 2012	• Section 2.1.8, NC-SI. Note corrected; now specifies correct pull-ups/downs used when NC-SI is disconnected. • Section 4.2.3, Reset Effects. Note #11. PSRTYPE removed from the list. • Table 4-3, Power-Up Timing Guarantees. The t_{opll} and t_{pcipll} descriptions were updated. The t_{pgres} MAX value was added. • Section 5.2.5.3.2, Master Disable. There is new material in the section. The new text begins: "In the above situation, the data path must be flushed before the software resets the 82599. The recommended method to flush the transmit data path is...."; the discussion continues with a methodology presentation. • Table 5-3, Start-up and Power State Transition Timing Parameters. Footnote with link to Table 4.4 added. t_{pres} MIN value entered at 100 ms. • Section 8.2.3.5.1, Extended Interrupt Cause Register- EICR (0x00800; RW1C). Bit 31 exposed (Other Cause Interrupt bit). • Section 7.10.2.2.7, Serial ID. Section updated. New text is "The serial ID capability is not supported in VFs." • Section 8.2.3.1.3, Extended Device Control Register — CTRL_EXT (0x00018; RW). Note text has been added to internal version of the bit 16 discussion (which has also been exposed for external use). The new text is: "The bit must be set during Rx flow initialization for proper device operation." • Section 8.2.3.4.12, PCIe Control Extended Register — GCR_EXT (0x11050; RW). Exposed Buffers Clear Function (bit 30) in the Datasheet. Was RESERVED. • Section 8.2.3.22.8, MAC Core Control 0 Register — HLREG0 (0x04240; RW). Exposed RXCRCSTRP bit in the Datasheet. Was RESERVED. • Section 8.2.3.22.19, Auto Negotiation Control Register — AUTOC (0x042A0; RW). Description of bits 11 &10 updated (DI0GMP, RATD). • Section 12.2.5, Trace Geometries. The inadequate data provided in this section has been replaced by a reference to a document that contains complete information.
2.76	September 6, 2012	• Section 3.7.7.3.1, Priority Flow Control. The first sentence in the section was updated for clarity. • Figure 7-6 and Figure 7-7. These have been updated for clarity. • Section 7.1.2.3, L2 Ethertype Filters: — Text has been updated in items 8 & 9. — Check the mirroring rules bullet in the same section; both second-level bullets after this bullet have been updated. • Section 7.1.2.7.11, Query Filter Flow. The table in this section was updated. Note the N/A entries.

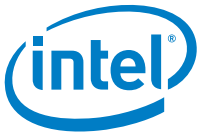
Rev	Date	Comments
2.76 (cont.)		• Section 7.1.2.7, Flow Director Filters. A note has been added. See: "Note: IPv6 extended headers are parsed by the 82599, enabling TCP layer header recognition. Still the IPv6 extended header fields are not taken into account for the queue classification by Flow Director filter. This rule do not apply for security headers and fragmentation header. Packets with fragmentation header miss this filter. Packets with security extended headers are parsed only up to these headers and therefore can match only filters that do not require fields from the L4 protocol. • Section 7.1.2.8.1, RSS Hash Function. A note has been added. See "Note: IPv6 extended headers are parsed by the 82599, enabling TCP layer header recognition. Still the IPv6 extended header fields are not taken into account for the queue classification by RSS filter. This rule do not apply for security headers and fragmentation header. Packets with fragmentation header miss this filter. Packets with security extended headers are parsed only up to these headers and therefore can match only filters that do not require fields from the L4 protocol. • Section 7.1.5, Legacy Receive Descriptor Format. A paragraph has been updated for clarity. Search for "The VP field indicates whether the incoming packet's type is a VLAN (802.1q). It is set if the packet type matches VLNCTRL.VET. Furthermore, if the RXDCTL.VME bit is set then active VP bit also indicates that VLAN has been stripped in the 802.1q packet..." • Section 7.2.1.2, Transmit Path in the 82599. A sentence in one of the subsections was rephrased. Search for "Each on-die descriptor queue contains up to 40 descriptors..." • Section 7.2.3.2.4, Advanced Transmit Data Descriptor. A sentence has been updated. Search for "optional VLAN tagging, the FCoE trailer containing the FC CRC and EOF (for FCoE packets), Ethernet CRC or Ethernet padding." • Section 7.1.10, Header Splitting. A note (indicating a restriction) has been added to this section. See "Note: Header Splitting mode might cause unpredictable behavior and should not be used with the 82599. For more information, see the product specification update errata on this subject." • Section 7.13.3.3.6, DDP Context. A sentence in a subsection was updated. Search for "Hardware uses the SEQ_CNT for checking in order reception." • Section 8.2.3.21.20, Flow Director Filters VLAN and FLEX Bytes — FDIRVLAN (0x0EE24; RW). Find the VLAN Field; the description for this field has been updated. • Section 8.2.3.22.23, Auto Negotiation Link Partner Link Control Word 1 Register — ANLP1 (0x042B0; RO). The description for the ANAS field has been updated. • Section 8.2.3.21.22, Flow Director Filters Command Register — FDIRCMD (0x0EE2C; RW). The description has been updated for the FDIRCMD,Drop bit. • Section 9.3.10.7, Link Capabilities Register (0xAC; RO). A value for bits 14:12 has been updated. The new value is "111b = More than 64 ms." • Section 12.11.1, LAN Disable. A paragraph was deleted because it referred to an obsolete function. The current second paragraph is also new. • Section 12.2.1, MAUI Channels Lane Connections. A sentence in the second paragraph was deleted because it did not apply. • Section 15.0, Glossary and Acronyms. The list was updated.Obsolete entries were removed.
2.8	June 21, 2013	• Section 1.2, Product Overview - Modified version information. • Section 3.2.5.1, Transmit Errors in Sequence Handling - Fixed typos in note at end of section. • Section 3.7.4.2, MAC Link Setup and Auto Negotiation - Added content from Specification Update as note to auto-negotiation discussion. • Section 4.6.9, FCoE Initialization Flow - Exposed TSOFF, TEOFF, RSOFF and REOFF registers to external documentation. • Section 6.2.4, Software Reserved Word — PXE VLAN Configuration Pointer — Word Address 0x20 - Add PXE VLAN NVM words to NVM Maps. • Added Section 6.2.6.1, PXE Setup Options PCI Function 0 — Word Address 0x30. • Section 6.3.5.6, PCIe Control 2 — Offset 0x05 - Changed default value of Bit 2 (Dummy Function Enable) from 1b to 0b. • Section 7.1.2.7.2, Flow Director Filters Status Reporting - Replaced paragraph regarding packets that do not match a flow director filter.

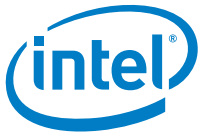
Rev	Date	Comments
2.8 (cont.)		• Section 7.1.2.7.7, Update Filter Flow - Add content from Specification update regarding internal memory space requirements. • Section 7.1.6.2, Advanced Receive Descriptors — Write-Back Format - Corrected typos in Table 7-16 and related paragraph. • Section 7.2.3.2.3, Advanced Transmit Context Descriptor - Added titles to tables 7-35, 7-36 and 7-37, and added respective cross references in related body text. • Section 7.2.3.2.3, Advanced Transmit Context Descriptor - Changed text in FCoEF description from "EOFF to "TEOFF." • Section 7.13.2.7.4, Dynamic End Of Frame Fields - Replaced Table 7-93, EOF Codes in TSO. • Section 7.7.2.4.1, Definition and Description of Parameters - Modified text in first paragraph. • Section 7.3.4.3.2, MSI-X Vectors Used by Virtual Functions (VFs) - Corrected typos in Figure 7-25 through Figure 7-27 . • Section 8.2.3.1.6, LED Control — LEDCTL (0x00200; RW) - Modified text for LINK/ ACTIVITY description. • Section 8.2.3.5.1, Extended Interrupt Cause Register- EICR (0x00800; RW1C), Removed Step 3 from Flow Director description. • Section 8.2.3.20, FCoE Registers - Exposed TSOFF, TEOFF, RSOFF and REOFF registers to external documentation. • Section 8.2.3.21.10, Flow Director Filters Free — FDIRFREE (0x0EE38; RW) - Changed bits 30:16 to Reserved. • Section 8.2.3.21.11, Flow Director Filters Length — FDIRLEN (0x0EE4C; RC) - Changed bits 30:16 to Reserved. • Section 8.2.3.21.13, Flow Director Filters Failed Usage Statistics — FDIRFSTAT (0x0EE54; RW/RC) - Changed description for bits 7:0. • Section 8.2.3.22.19, Auto Negotiation Control Register — AUTOC (0x042A0; RW) - Added note preceding register description. • Section 8.2.3.22.22, Auto Negotiation Control 2 Register — AUTOC2 (0x042A8; RW) - Exposed bit 28. Bits 27:19 and bit 29 Reserved. • Section 11.6.2, EEPROM - Based on minimum and recommended EEPROM sizes presented in Section 11.6.2.1 and Section 11.6.2.2, removed 8, 16, 32 and 64 Kb devices from Table 11-29 • Section 12.3.1, Supported EEPROM Devices - Based on minimum and recommended EEPROM sizes presented in Section 11.6.2.1 and Section 11.6.2.2, removed 8, 16, 32 and 64 Kb devices and accompanying note from Table 12-1 • Section 11.3.1.1, Power On/Off Sequence - Added rows to Table 11-5 for Tlpgw, Tlpg-per and Tlpg.
2.9	January 8, 2014	• Section 1.2, Product Overview — Updated product version information. • Section 3.4.2, EEPROM Device — Updated EEPROM compatibility information, and added reference to table of support EEPROM devices. • Section 4.6.11.4, Transmit Rate Scheduler — Fixed typo. • Section 6.3.5.13, IOV Control Word 1 — Offset 0x0C — Updated default value for Max VFs filed, and added note to field description. • Section 7.2.3.1, Introduction — Modified legacy descriptors information in Transmit Descriptors "Introduction" section. • Section 7.2.3.2.2, Legacy Transmit Descriptor Format — Corrected typo from "Rx" to "Tx" in Report Status (RS) description. • Section 7.2.3.2.4, Advanced Transmit Data Descriptor — Updated list in "Check Context bit" description.
3.0	November 5, 2014	• Section 4.6.3.2, Global Reset and General Configuration — Updated text related to FCRTH[n].RTH fields. • Section 4.6.9, FCoE Initialization Flow — Text updates. • Section 7.1.2.3, L2 Ethertype Filters — Text updates. • Section 8.2.3.23.3, Error Byte Packet Count — ERRBC (0x04008; RC) — Changed long register name from "Error Byte Count".

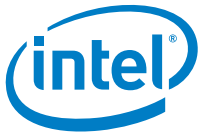
Rev	Date	Comments
3.1	February 1, 2015	• Section 4.6.6, Interrupt Initialization — Provided additional information on "Operating with Legacy or MSI Interrupts". • Section 6.4, Firmware Module — Made changes associated with the addition of Appendix B. • Section 6.4.4.7, NC-SI Configuration Offset 0x06 — Exposed Enable Channel Swap field (Bit 13). • Section 8.2.3.1.2, Device Status Register — STATUS (0x00008; RO) — Added text describing the *LinkUp* bit as Read/Write. • Section 11.5.1, Mechanical — Corrected typo (FCGBA -> FCBGA) • Section 12.2.1, MAUI Channels Lane Connections — Added note regarding unused pins and design with 82599EN single port SKU. • Added Appendix A, "Packets and Frames". • Added Appendix B, "LESM - Link Establishment State Machine for the 82599".
3.2	October 19, 2015	• Section 6.2.3, iSCSI Boot Configuration — Word Address 0x17 - Updated section.
3.3	March 11, 2016	• Section 4.2.1.5.3, Virtual Function FLR (VFLR) — Added note related to VFMBMEM. • Removed VFMBMEM from Note #11 related to Table 4-6. • Section 14.1, Link Loopback Operations — Removed a textual reference to a non-existent register.

Contents

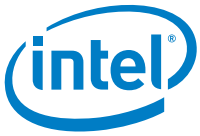








NOTE: ***This page intentionally left blank.***



1.0 Introduction

1.1 Scope

This document describes the external architecture (including device operation, pin descriptions, register definitions, etc.) for the 82599, a dual 10 Gigabit Ethernet (GbE) Network Interface Controller.

This document is intended as a reference for logical design group, architecture validation, firmware development, software device driver developers, board designers, test engineers, or anyone else who may need specific technical or programming information about the 82599.

1.2 Product Overview

The 82599 is a derivative of previous generations of Intel 1 GbE and 10 GbE Network Interface Card (NIC) designs. Many features of its predecessors remain intact; however, some have been removed or modified as well as new features introduced.

Note: This device is available in dual port and single port variants. The single port device is limited to port 0 and to the SFI interface. If you are using the single port device, information in this document pertaining to port 1 and other interfaces does not apply.

Three versions are available:

- 82599EB — PCI Express* (PCIe*) 2.0, dual-port 10 Gigabit Ethernet controller for XAUI, KX, KX4, BX, BX4 and CX4 interfaces.
- 82599ES — Dual-port serial 10 GbE backplane interface for blade implementations (includes the 82599EB SKU functionality plus KR and SFI interfaces).
- 82599EN — Single-port serial 10 GbE SFI interface for blade implementations.

Note: For addition information, see the introductory material in the *Intel® 82599 10 Gigabit Ethernet Controller Specification Update.*

1.2.1 82599 Silicon/Software Features

The base software device driver supports the following interfaces:

- XAUI (BX4)
- SFI
- KX/KX4
- KR

Linux software features include:

- LLI — Low Latency Interrupts
- DCA — Direct Cache Access
- RSC — Receive Side Coalescing
- All sleep states (S0 through S5); however, for sleep states S3 through S5, there are power and airflow conditions that need to be met. Refer to Section 5.0 and Section 11.0 for more details.
- Header Split — This feature consists of splitting a packet header to a different memory space and help the host to fetch headers only for processing.
- Flow Director (SW ATR only) — A large number of flow affinity filters that direct receive packets by their flows to queues for classification, load balancing, and matching between flows and CPU cores.

Windows software features include:

- LLI
- DCA
- Wake on LAN (WoL) support:
  - WoL from S3 and S4 are not currently supported for the 82599. Also, there are no plans to support it in the future.
  - WoL from S5 is supported for connections capable of KR to KX transitions only. Implementation of this feature has special requirements, contact your Intel representative for more details.
- Header Split — This feature consists of splitting a packet header to a different memory space and help the host to fetch headers only for processing.

Note: Some PCIe x8 slots are actually configured as x4 slots. These slots have insufficient bandwidth for full 10 GbE line rate with dual port 10 GbE devices. If a solution suffers bandwidth issues when both 10 GbE ports are active, it is recommended to verify that the PCIe slot is indeed a true PCIe x8.

1.2.2 System Configurations

The 82599 is targeted for system configurations such as rack mounted or pedestal servers, where it can be used as an add-on NIC or LAN on Motherboard (LOM). Another system configuration is for blade servers, where the 82599 can be used in a LOM or mezzanine card.

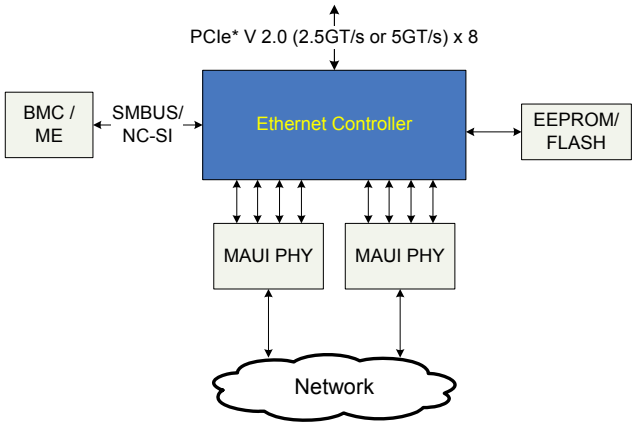


Figure 1-1 Typical Rack / Pedestal System Configuration

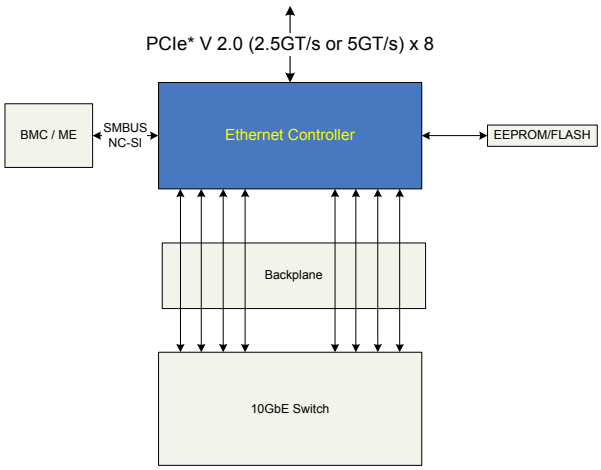


Figure 1-2 Typical Blade System Configuration

1.2.3 External Interfaces

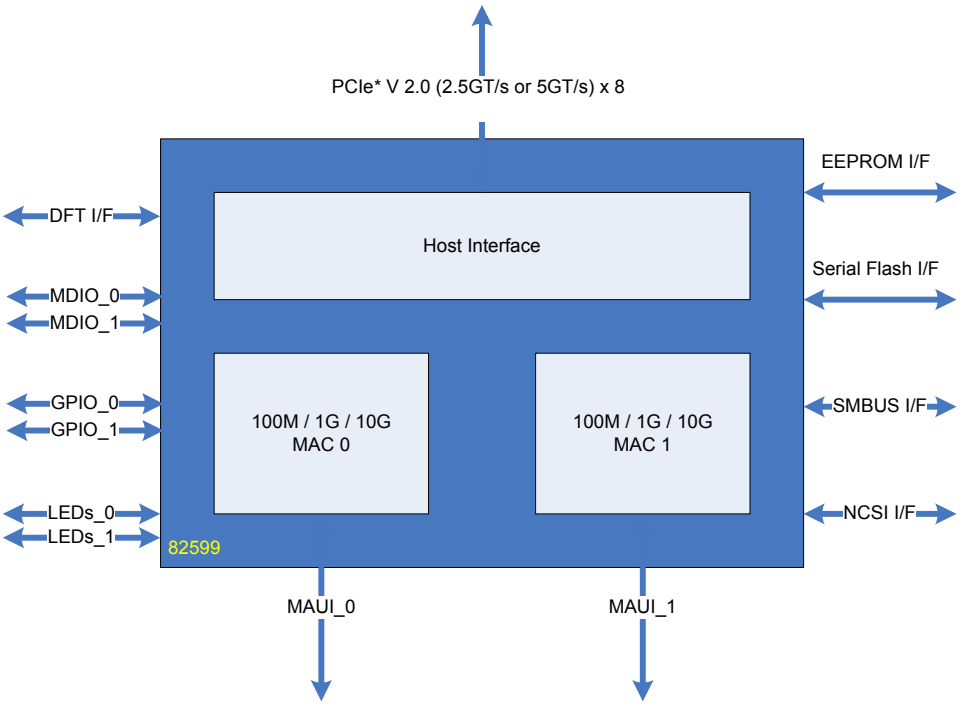


Figure 1-3 82599 External Interfaces Diagram

1.2.4 PCI-Express* (PCIe*) Interface

The 82599 supports PCIe V2.0 (2.5GT/s or 5GT/s). See Section 2.1.2 for full pin description and Section 11.4.3 for interface timing characteristics.

1.2.5 Network Interfaces

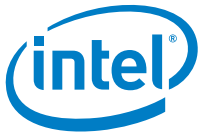
The 82599 interfaces the network through the Multi-Speed Attachment Unit Interface also referred to as the MAUI interface.

Two independent MAUI interfaces are used to connect two the 82599 ports to external devices. Each MAUI interface can be configured to interface using the following high speed links:

a. XAUI for connection to another XAUI compliant PHY device or optical module.

b. SGMII for connection to another SGMII compliant PHY using 1000BASE-BX or 1000BASE-KX electrical signaling.

c. 1000BASE-KX for connection over a backplane to another 1000BASE-KX compliant device.



d. 10GBASE-KX4 for connection over a backplane to another 10GBASE-KX4 device.

e. 1000BASE-BX for connection over a backplane to another 1000BASE-BX compliant device.

f. 10GBASE-CX4 for connection over a CX4 compliant cable to another 10GBASE-CX4 compliant device.

g. SFI for connection to another SFI compliant PHY or optical module.

h. 10GBASE-KR for connection over a backplane to another 10GBASE-KR compliant device.

i. 10GBASE-BX4 for connection over a backplane to another 10GBASE-BX4 device.

The 82599 also supports:

- IEEE 802.3ae (10 Gb/s) implementations. It performs all of the functions required for transmission and reception handling called out in the standards for a XAUI Media interface.
- IEEE 802.3ak, IEEE 802.3ap Backplane Ethernet (KX, KX4, or KR), and PICMG3.1 (BX only) implementations including an Auto-Negotiation layer and PCS layer synchronization.
- SFP+ MSA (SFI) implementations.

These interfaces can be configured to operate in 100 Mb/s mode (SGMII), 1 Gb/s mode (SGMII, BX and KX) and 10 Gb/s mode (XAUI, CX4, KX4, KR and SFI). In 100 Mb/s mode, 1 Gb/s mode and in KR and SFI 10 Gb/s modes, only one of the four MAUI lanes (lane 0) is used and the remaining lanes (lanes 1 to 3) are powered down. For more information on how to configure the 82599 for 100 Mb/s,
1 Gb/s or 10 Gb/s operating modes, refer to Section 3.7.

Refer to Section 2.1.3 for full-pin descriptions and to the respective specifications (IEEE802.3, optical module MSAs...). For the timing characteristics of those interfaces see the relevant external specifications as listed in Section 11.4.4 for interface timing characteristics.

1.2.6 EEPROM Interface

The 82599 uses an EEPROM device for storing product configuration information. Several words of the EEPROM are accessed automatically by the 82599 after reset in order to provide pre-boot configuration data that must be available to it before it is accessed by host software. The remainder of the stored information is accessed by various software modules used to report product configuration, serial number, etc.

The 82599 uses a SPI (4-wire) serial EEPROM devices. Refer to Section 2.1.4 for the I/O pin descriptions; Section 11.4.2.4 for timing characteristics of this interface and Section 11.6.2 for a list of supported EEPROM devices.

1.2.7 Serial Flash Interface

The 82599 provides an external SPI serial interface to a Flash (or boot ROM) device. The 82599 supports serial Flash devices with up to 64 Mb (8 MB) of memory. The size of the Flash used by the 82599 can be configured by the EEPROM. See Section 2.1.5 for full pin description and Section 11.4.2.3 for interface timing characteristics.

Note: Though the 82599 supports devices with up to 8 MB of memory, bigger devices can also be used. Accesses to memory beyond the Flash device size results in access wrapping as only the lower address bits are used by the Flash control unit.

1.2.8 SMBus Interface

SMBus is an optional interface for pass-through and/or configuration traffic between an external MC and the 82599.

The 82599's SMBus interface supports standard SMBus, up to a frequency of 400 KHz. Refer to Section 2.1.6 for full-pin descriptions and Section 11.4.2.2 for timing characteristics of this interface.

1.2.9 NC-SI Interface

NC-SI is an optional interface for pass-through traffic to and from a MC. The 82599 meets the NC-SI version 1.0.0a specification.

Refer to Section 3.3 for an additional description of the NC-SI interface, Section 2.1.8 for the pin descriptions, Section 10.5.1 for NC-SI programming and Section 11.4.1.4 for the timing characteristics.

1.2.10 MDIO Interfaces

The 82599 implements two serial management interfaces known as the Management Data Input/Output (MDIO) Interface that controls and manages PHY devices (master side). This interface provides the Media Access Controller (MAC) and software with the ability to monitor and control the state of the PHY. The 82599 supports the MDIO frame formats specified in both IEEE802.3 clause 22 and IEEE802.3 clause 45 using the electrical specification defined in IEEE802.3 clause 22 (LVTTL signaling). The MDIO interface can be controlled by software via a MDI single command and address register — MSCA (see Section 8.2.3.22.11 for more details).

Each MDIO interface should be connected to the relevant PHY as shown in the following example (each MDIO interface is driven by the appropriate MAC function).

Refer to Section 3.7.6 for complete description of the MDIO interface, Section 2.1.9 for the pin descriptions, the MSCA register in Section 8.2.3.22.11, and Section 11.4.2.7 for the timing characteristics.