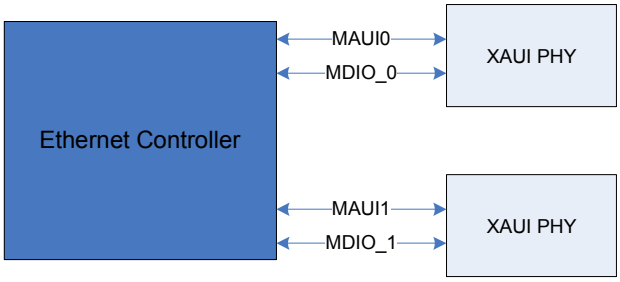


Figure 1-4 MDIO Connection Example

1.2.11 I²C Interfaces

The 82599 implements two serial management interfaces known as I²C Management Interfaces for the control and management of external optical modules (XFP and SFP+). This interface provides the MAC and software with the ability to monitor and control the state of the optical module. The use, direction, and values of the I²C pins are controlled and accessed using fields in the I2C Control (I2CCTL) register.

Each I²C interface should be connected to the relevant PHY as shown in the following example (each I²C interface is driven by the appropriate MAC function).

Refer to Section 2.1.7 for the pin descriptions, I2CCTL register information in Section 8.2.3.1.5 for I²C programming, and Section 11.4.2.2 for timing characteristics.

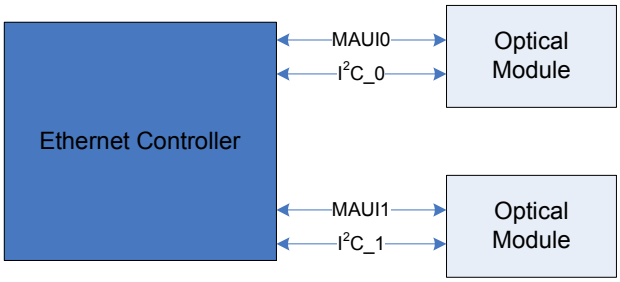


Figure 1-5 I²C Connection Example

1.2.12 Software-Definable Pins (SDP) Interface (General-Purpose I/O)

The 82599 has eight SDP pins per port that can be used for miscellaneous hardware or software-controllable purposes. These pins can each be individually configured to act as either input or output pins. Via the SDP pins, the 82599 can support IEEE1588 auxiliary device connections, control of the low speed optical module interface, and other functionality. For more details on the SDPs see Section 3.6 and the ESDP register information in Section 8.2.3.1.4.

1.2.13 LED Interface

The 82599 implements four output drivers intended for driving external LED circuits per port. Each of the four LED outputs can be individually configured to select the particular event, state, or activity, which is indicated on that output. In addition, each LED can be individually configured for output polarity as well as for blinking versus non-blinking (steady-state) indications.

The configuration for LED outputs is specified via the LEDCTL register (see Section 8.2.3.1.6). In addition, the hardware-default configuration for all LED outputs can be specified via an EEPROM field (see Section 6.3.7.3), thereby supporting LED displays configured to a particular OEM preference.

See Section 2.1.11 for a full pin description.

1.3 Features Summary

Table 1-1 to Table 1-7 list the 82599's features in comparison to previous dual-port 1 Gb/s and 10 Gb/s Ethernet controllers.

Table 1-1 General Features

Feature	82599	82598	Reserved
Serial Flash Interface	Y	Y	
4-wire SPI EEPROM Interface	Y	Y	
Configurable LED Operation for Software or OEM Customization of LED Displays	Y	Y	
Protected EEPROM Space for Private Configuration	Y	Y	
Device Disable Capability	Y	Y	
Package Size	25 x 25 mm	31 x 31 mm	
Watchdog Timer	Y	N	
Time Sync (IEEE 1588)	Y	N	

Table 1-2 Network Features

Feature	82599	82598	Reserved
Compliant with the 10 Gb/s and 1 Gb/s Ethernet/802.3ap (KX/KX4) Specification	Y	Y	
Compliant with the 10 Gb/s 802.3ap (KR) specification	Y	N	
Support of 10GBASE-KR FEC	Y	N	

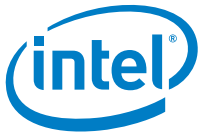


Table 1-2 Network Features (Continued)

Feature	82599	82598	Reserved
Compliant with the 10 Gb/s Ethernet/802.3ae (XAUI) Specification	Y	Y	
Compliant with SFI interface	Y	N	
Support for EDC	N	N	
Compliant with the 1000BASE-BX Specification	Y	Y	
Half-duplex at 10/100 Mb/s Operation and Full-Duplex Operation at all Supported Speeds	Y (100 Mb FDX)	NA	
10/100/1000 Copper PHY Integrated On-chip	N	N	
Support Jumbo Frames of up to 15.5 KB (15872 bytes)	Y [1]	Y	
Auto-Negotiation Clause 73 for Supported Modes	Y	Y	
Flow Control Support: Send/Receive Pause Frames and Receive FIFO Thresholds	Y	Y	
Statistics for Management and RMON	Y	Y	
802.1q VLAN Support	Y	Y	
SerDes Interface for External PHY Connection or System Interconnect	Y	Y	
SGMII Interface	Y (100 M/1G only)	N	
SerDes Support of non Auto-Negotiation Partner	Y	Y	
Double VLAN	Y	N	

1. The 82599 supports full-size 15.5 KB (15872-byte) jumbo packets while in a basic mode of operation. When DCB mode is enabled, or security engines enabled or virtualization is enabled, the 82599 supports 9.5 KB (9728-byte) jumbo packets.

Table 1-3 Host Interface Features

Feature	82599	82598	Reserved
PCIe* Host Interface	PCIe V2.0 (2.5GT/s or 5GT/s)	PCIe v2.0 (2.5GT/s)	
Number of Lanes	x1, x2, x4, x8	x1, x2, x4, x8	
64-bit Address Support for Systems Using More Than 4 GB of Physical Memory	Y	Y	
Outstanding Requests for Tx Data Buffers	16	16	
Outstanding Requests for Tx Descriptors	8	8	
Outstanding Requests for Rx Descriptors	8	4	
Credits for P-H/P-D/NP-H/NP-D (shared for the 2 ports)	16/16/4/4	8/16/4/4	

Table 1-3 Host Interface Features (Continued)

Feature	82599	82598	Reserved
Max Payload Size Supported	512 Bytes	256 Bytes	
Max Request Size Supported	2 KB	256 Bytes	
Link Layer Retry Buffer Size (shared for the 2 ports)	3.4 KB	2 KB	
Vital Product Data (VPD)	Y	N	
End to End CRC (ECRC)	Y	N	

Table 1-4 LAN Functions Features

Feature	82599	82598	Reserved
Programmable Host Memory Receive Buffers	Y	Y	
Descriptor Ring Management Hardware for Transmit and Receive	Y	Y	
ACPI Register Set and Power Down Functionality Supporting D0 & D3 States	Y	Y	
Integrated LinkSec security engines: AES-GCM 128-bit; Encryption + Authentication; One SC x 2 SA per port. Replay Protection with Zero Window	Y	N	
Integrated IPsec security engines: AES-GCM 128bit; AH or ESP encapsulation; IPv4 and IPv6 (no option or extended headers)	1024 SA / port	N	
Software-Controlled Global Reset Bit (Resets Everything Except the Configuration Registers)	Y	Y	
Software-Definable Pins (SDP); (per port)	8	8	
Four SDP Pins can be Configured as General Purpose Interrupts	Y	Y	
Wake-on-LAN (WoL)	Y	Y	
IPv6 Wake-up Filters	Y	Y	
Configurable (through EEPROM) Wake-up Flexible Filters	Y	Y	
Default Configuration by EEPROM for all LEDs for Pre-Driver Functionality	Y	Y	
LAN Function Disable Capability	Y	Y	
Programmable Memory Transmit Buffers	160 KB / port	320 KB / port	
Programmable Memory Receive Buffers	512 KB / port	512 KB / port	

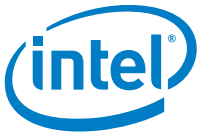


Table 1-5 LAN Performance Features[1]

Feature	82599	82598	Reserved
TCP/UDP Segmentation Offload	256 KB in all modes	256 KB in legacy mode, 32 KB in DCB	
TSO Interleaving for Reduced Latency	Y	N	
TCP Receive Side Coalescing (RSC)	32 flows / port	N	
Data Center Bridging (DCB), IEEE Compliance to: Priority Groups (up to 8) and Bandwidth Allocation (ETS) IEEE802.1Qaz Priority-based Flow Control (PFC) IEEE802.1Qbb	 Y Y	 Y Y	
Transmit Rate Scheduler	Y	N	
IPv6 Support for IP/TCP and IP/UDP Receive Checksum Offload	Y	Y	
Fragmented UDP Checksum Offload for Packet Reassembly	Y	Y	
FCoE Tx / Rx CRC Offload	Y	N	
FCoE Transmit Segmentation	256 KB	N	
FCoE Coalescing and Direct Data Placement	512 outstanding Read — Write requests / port	N	
Message Signaled Interrupts (MSI)	Y	Y	
Message Signaled Interrupts (MSI-X)	Y	Y	
Interrupt Throttling Control to Limit Maximum Interrupt Rate and Improve CPU Use	Y	Y	
Rx Packet Split Header	Y	Y	
Multiple Rx Queues (RSS)	Y (multiple modes)	8x8 16x4	
Flow Director Filters: up to 32 K - 2 Flows by Hash Filters or up to 8 K- 2 Perfect Match Filters	Y	N	
Number of Rx Queues (per port)	128	64	
Number of Tx Queues (per port)	128	32	
Low Latency Interrupts DCA Support TCP Timer Interrupts Relax Ordering	Yes to all	Yes to all	
Rate Control of Low Latency Interrupts	Y	N	

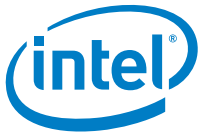
1. The 82599 10 GbE operation is focused on performance improvement; note that the 82599's 1GbE mode is optimized for power saving.

Table 1-6 Virtualization Features

Feature	82599	82598	Reserved
Support for Virtual Machine Device Queues (VMDq)	64	16	
L2 Ethernet MAC Address Filters (unicast and multicast)	128	16	
L2 VLAN filters	64	-	
PCI-SIG SR IOV	Y	N	
Multicast and Broadcast Packet Replication	Y	N	
Packet Mirroring	Y	N	
Packet Loopback	Y	N	
Traffic Shaping	Y	N	

Table 1-7 Manageability Features

Feature	82599	82598	Reserved
Advanced Pass Through-Compatible Management Packet Transmit/Receive Support	Y	Y	
SMBus Interface to an External MC	Y	Y	
NC-SI Interface to an External MC	Y	Y	
New Management Protocol Standards Support (NC-SI)	Y	Y	
L2 Address Filters	4	4	
VLAN L2 Filters	8	8	
Flex L3 Port Filters	16	16	
Flexible TCO Filters	4	4	
L3 Address Filters (IPv4)	4	4	
L3 Address Filters (IPv6)	4	4	



1.4 Overview of New Capabilities Beyond the 82598

1.4.1 Security

The 82599 supports the IEEE P802.1AE LinkSec specification. It incorporates an inline packet crypto unit to support both privacy and integrity checks on a packet by packet basis. The transmit data path includes both encryption and signing engines. On the receive data path, the 82599 includes both decryption and integrity checkers. The crypto engines use the AES GCM algorithm, which is designed to support the 802.1AE protocol. Note that both host traffic and Manageability Controller (MC) management traffic might be subject to authentication and/or encryption.

The 82599 supports IPsec offload for a given number of flows. It is the operating system's responsibility to submit (to hardware) the most loaded flows in order to take maximum benefits of the IPsec offload in terms of CPU utilization savings. Main features are:

- Offload IPsec for up to 1024 Security Associations (SA) for each of Tx and Rx
- AH and ESP protocols for authentication and encryption
- AES-128-GMAC and AES-128-GCM crypto engines
- Transport mode encapsulation
- IPv4 and IPv6 versions (no options or extension headers)

1.4.2 Transmit Rate Limiting

The 82599 supports Transmit Rate Scheduler (TRS) in addition to the Data Center Bridging (DCB) functionality provided in the 82598. TRS is enabled for each transmit queue. The following modes of TRS are used:

- Frame Overhead — IPG is extended by a fixed value for all transmit queues.
- Payload Rate — IPG, stretched relative to frame size, provides pre-determined data (bytes) rates for each transmit queue.

1.4.3 Fibre Channel over Ethernet (FCoE)

Fibre Channel (FC) is the predominant protocol used in Storage Area Networks (SAN). Fibre Channel over Ethernet (FCoE) enables a connection between an Ethernet storage initiator and legacy FC storage targets or a complete Ethernet connection between a storage initiator and a device.

Existing FC Host Bus Adapters (HBAs) used to connect between FC initiator and FC targets provide full offload of the FC protocol to the initiator that enables maximizing storage performance. The 82599 offloads the main data path of I/O Read and Write commands to the storage target.

1.4.4 Performance

The 82599 improves on previous 10 GbE products in the following performance vectors:

- Throughput — The 82599 aims to provide wire speed dual-port 10 Gb/s throughput. This is accomplished using the PCIe physical layer (PCIe V2.0 (5GT/s), by tuning the internal pipeline to 10 Gb/s operation, and by enhancing the PCIe concurrency capabilities.
- Latency — The 82599 reduces end-to-end latency for high priority traffic in presence of other traffic. Specifically, the 82599 reduces the delay caused by preceding TCP Segmentation Offload (TSO) packets. Unlike previous products, a TSO packet might be interleaved with other packets going to the wire. Interleaving is done at the Ethernet packet boundary, therefore reducing the maximum delay due to a TSO from a TSO-worth of data to an MTU-worth of data.
- CPU utilization — The 82599 supports reduction in CPU utilization, mainly by supporting Receive Side Coalescing (RSC)
- Flow affinity filters

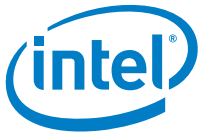
1.4.4.1 Receive Side Coalescing (RSC)

RSC coalesces incoming TCP/IP packets into larger receive segments. It is the inverse operation to TSO on the transmit side. It has the same motivation, reducing CPU utilization by executing the TCP/IP stack only once for a set of received Ethernet packets. The 82599 can handle up to 32 flows per port at any given time. See Section 7.11 for more details on RSC.

1.4.4.2 PCIe V2.0 (5GT/s)

Several changes are defined in the size of PCIe transactions to improve the performance in virtualization environments. Larger request sizes decrease the number of independent transactions on PCIe and therefore decreases trashing of the IOTLB cache. Changes include:

- Increase in the number of outstanding requests (data, descriptors) to a total of 32 requests
- Increase in the number of credits for posted transaction (such as for tail updates) to 16
- Increase in the maximum payload size supported from 256 bytes to 512 bytes
- Increase in the supported maximum read request size from 256 bytes to 2 KB. Note that the amount of outstanding request data does not change. That is, if the 82599 supports N outstanding requests of 256 bytes, then it would support N/2 requests of 512 bytes, etc.
- Retry buffer size — The link layer retry buffer size increases to 3.4 KB to meet the higher speed of PCIe V2.0 (5GT/s).



1.4.5 Rx/Tx Queues and Rx Filtering

The 82599 Tx and Rx queues have increased in size to 128 Tx queues and 128 Rx queues. Additional filtering capabilities are provided based on:

- L2 Ethertype
- 5-tuples
- SYN identification
- Flow Director — 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.

See Section 7.0 for a complete description.

1.4.6 Interrupts

Several changes in the interrupt scheme are available in the 82599:

- Control over the rate of Low Latency Interrupts (LLI)
- Extensions to the filters that invoke LLIs
- Additional MSI-X vectors for the five-tuple filters and for IOV virtualization

See Section 7.3 for more details.

1.4.7 Virtualization

See Section 7.10 for more details.

1.4.7.1 PCI- IOV

The 82599 supports the PCI-SIG single-root I/O Virtualization initiative (SR-IOV), including the following functionality:

- Replication of PCI configuration space
- Allocation of BAR space per virtual function
- Allocation of requester ID per virtual function
- Virtualization of interrupts

The 82599 provides the infrastructure for direct assignment architectures through a mailbox mechanism. Virtual Functions (VFs) might communicate with the Physical Function (PF) through the mailbox and the PF can allocate shared resources through the mailbox channel.

1.4.7.2 Packet Filtering and Replication

The 82599 adds extensive coverage for packet filtering for virtualization by supporting the following filtering modes:

- Filtering by unicast Ethernet MAC address
- Filtering by VLAN tag
- Filtering of multicast Ethernet MAC address
- Filtering of broadcast packets

For each of the above categories, the 82599 can replicate packets to multiple Virtual Machines (VMs). Various mirroring modes are supported, including mirroring a VM, a Virtual LAN (VLAN), or all traffic into a specific VM.

1.4.7.3 Packet Switching

The 82599 forwards transmit packets from a transmit queue to an Rx software queue to support VM-VM communication. Transmit packets are filtered to an Rx queue based on the same criteria as packets received from the wire.

1.4.7.4 Traffic Shaping

Transmit bandwidth is allocated among the virtual interfaces to avoid unfair use of bandwidth by a single VM. Allocation is done separately per DCB traffic class so that bandwidth assignment to each traffic class is partitioned among the different VMs.

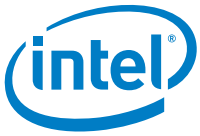
1.4.8 VPD

The 82599 supports VPD capability defined in the PCI Specification, version 3.0. See Section 3.4.9 for more details.

1.4.9 Double VLAN

The 82599 supports a mode where all received and sent packets have at least one VLAN tag in addition to the regular tagging that can optionally be added. This mode is used for systems where the switches add an additional tag containing switching information.

When a port is configured to double VLAN, the 82599 assumes that all packets received or sent to this port have at least one VLAN. The only exception to this rule is flow control packets, which don't have a VLAN tag. See Section 7.4.5 for more details.



1.4.10 Time Sync — IEEE 1588 — Precision Time Protocol (PTP)

The IEEE 1588 International Standard lets networked Ethernet equipment synchronize internal clocks according to a network master clock. The protocol is implemented mostly in software, with the 82599 providing accurate time measurements of special Tx and Rx packets close to the Ethernet link. These packets measure the latency between the master clock and an end-point clock in both link directions. The endpoint can then acquire an accurate estimate of the master time by compensating for link latency. See Section 7.9 for more details.

The 82599 provides the following support for the IEEE 1588 protocol:

- Detecting specific PTP Rx packets and capturing the time of arrival of such packets in dedicated CSRs
- Detecting specific PTP Tx packets and capturing the time of transmission of such packets in dedicated CSRs
- A software-visible reference clock for the above time captures

1.5 Conventions

1.5.1 Terminology and Acronyms

See Section 15.0 for a list of terminology and acronyms used throughout this document.

1.5.2 Byte Count

When referencing jumbo packet size, 1 KB equals 1024 bytes.

For example:

- 9.5 KB equals 9.5 x 1024 = 9728 bytes
- 15.5 KB equals 15.5 x 1024 = 15872 bytes

1.5.3 Byte Ordering

This section defines the organization of registers and memory transfers, as it relates to information carried over the network:

- Any register defined in big endian notation can be transferred as is to/from Tx and Rx buffers in the host memory. Big endian notation is also referred to as being in network order or ordering.

- Any register defined in little endian notation must be swapped before it is transferred to/from Tx and Rx buffers in the host memory. Registers in little endian order are referred to being in host order or ordering.

Tx and Rx buffers are defined as being in network ordering; they are transferred as is over the network.

Note: Registers not transferred on the wire are defined in little endian notation. Registers transferred on the wire are defined in big endian notation, unless specified differently.

1.6 Register/Bit Notations

This document refers to device register names with all capital letters. To refer to a specific bit in a register the convention REGISTER.BIT is used. For example CTRL.GIO Master Disable refers to the GIO Master Disable bit in the Device Control (CTRL) register.

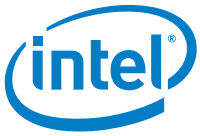
This document also refers to bit names as initial capital letters in an italic font. For example, *GIO Master Disable*.

1.7 References

The 82599 implements features from the following specifications:

IEEE Specifications

- IEEE standard 802.3-2005 (Ethernet). Incorporates various IEEE Standards previously published separately. Institute of Electrical and Electronic Engineers (IEEE).
- 10GBASE-X — An IEEE 802.3 physical coding sublayer for 10 Gb/s operation over XAUI and four lane PMDs as per IEEE 802.3 Clause 48.
- 1000BASE-CX — 1000BASE-CX over specially shielded 150 Ω balanced copper jumper cable assemblies as specified in IEEE 802.3 Clause 39.
- 10GBASE-LX4 — IEEE 802.3 Physical Layer specification for 10 Gb/s using 10GBASE-X encoding over four WWDM lanes over multimode fiber as specified in IEEE 802.3 Clause 54.
- 10GBASE-CX4 — IEEE 802.3 Physical Layer specification for 10 Gb/s using 10GBASE-X encoding over four lanes of 100 Ω shielded balanced copper cabling as specified in IEEE 802.3 Clause 54.
- 1000BASE-KX — IEEE 802.3ap Physical Layer specification for 1 Gb/s using 1000BASE-X encoding over an electrical backplane as specified in IEEE 802.3 Clause 70.
- 10GBASE-KX4 — IEEE 802.3ap Physical Layer specification for 10 Gb/s using 10GBASE-X encoding over an electrical backplane as specified in IEEE 802.3 Clause 71.



- 10GBASE-KR — IEEE 802.3ap Physical Layer specification for 10 Gb/s using 10GBASE-R encoding over an electrical backplane as specified in IEEE 802.3 Clause 72.
- 1000BASE-BX — 1000BASE-BX is the PICMG 3.1 electrical specification for transmission of 1 Gb/s Ethernet or 1 Gb/s Fibre Channel encoded data over the backplane.
- 10GBASE-BX4 — 10GBASE-BX4 is the PICMG 3.1 electrical specification for transmission of 10 Gb/s Ethernet or 10 Gb/s Fibre Channel encoded data over the backplane.
- IEEE standard 802.3ap, draft D3.2.
- IEEE standard 1149.1, 2001 Edition (JTAG). Institute of Electrical and Electronics Engineers (IEEE).
- IEEE standard 802.1Q for VLAN.
- IEEE 1588 International Standard, Precision clock synchronization protocol for networked measurement and control systems, 2004-09.
- IEEE P802.1AE/D5.1, Media Access Control (MAC) Security, January 19, 2006.

PCI-SIG Specifications

- PCI Express 2.0 Base specification, 12/20/2006.
- PCI Express™ 2.0 Card Electromechanical Specification, Revision 0.9, January 19, 2007.
- PCI Bus Power Management Interface Specification, Rev. 1.2, March 2004.
- PICMG3.1 Ethernet/Fibre Channel Over PICMG 3.0 Draft Specification January 14, 2003 Version D1.0.
- Single Root I/O Virtualization and Sharing, Revision 0.7, 1/11/2007.

IETF Specifications

- IPv4 specification (RFC 791)
- IPv6 specification (RFC 2460)
- TCP specification (RFC 793)
- UDP specification (RFC 768)
- ARP specification (RFC 826)
- RFC4106 — The Use of Galois/Counter Mode (GCM) in IPsec Encapsulating Security Payload (ESP).
- RFC4302 — IP Authentication Header (AH)
- RFC4303 — IP Encapsulating Security Payload (ESP)
- RFC4543 — The Use of Galois Message Authentication Code (GMAC) in IPsec ESP and AH.
- IETF Internet Draft, Marker PDU Aligned Framing for TCP Specification.
- IETF Internet Draft, Direct Data Placement over Reliable Transports.

Other

- Serial-GMII Specification, Cisco Systems document ENG-46158, Revision 1.7.
- Advanced Configuration and Power Interface Specification, Rev 2.0b, October 2002
- Network Controller Sideband Interface (NC-SI) Specification, Version cPubs-0.1, 2/18/2007.
- System Management Bus (SMBus) Specification, SBS Implementers Forum, Ver. 2.0, August 2000.
- EUI-64 specification, http://standards.ieee.org/regauth/oui/tutorials/EUI64.html.
- Backward Congestion Notification Functional Specification, 11/28/2006.
- Definition for new PAUSE function, Rev. 1.2, 12/26/2006.
- GCM spec — McGrew, D. and J. Viega, "The Galois/Counter Mode of Operation (GCM)", Submission to NIST. http://csrc.nist.gov/CryptoToolkit/modes/proposedmodes/gcm/gcm-spec.pdf, January 2004.
- FRAMING AND SIGNALING-2 (FC-FS-2) Rev 1.00
- Fibre Channel over Ethernet Draft Presented at the T11 on May 2007
- Per Priority Flow Control (by Cisco* Systems) — Definition for new PAUSE function, Rev 1.2, EDCS-472530

In addition, the following document provides application information:

- 82563EB/82564EB Gigabit Ethernet Physical Layer Device Design Guide, Intel Corporation.

1.8 Architecture and Basic Operation

1.8.1 Transmit (Tx) Data Flow

Tx data flow provides a high-level description of all data/control transformations steps needed for sending Ethernet packets over the wire.

Table 1-8 Tx Data Flow

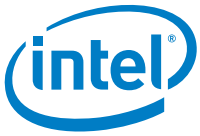
Step	Description
1	The host creates a descriptor ring and configures one of the 82599's transmit queues with the address location, length, head, and tail pointers of the ring (one of 128 available Tx queues).
2	The host is requested by the TCP/IP stack to transmit a packet, it gets the packet data within one or more data buffers.
3	The host initializes the descriptor(s) that point to the data buffer(s) and have additional control parameters that describes the needed hardware functionality. The host places that descriptor in the correct location at the appropriate Tx ring.
4	The host updates the appropriate Queue Tail Pointer (TDT)
5	The 82599's DMA senses a change of a specific TDT and as a result sends a PCIe request to fetch the descriptor(s) from host memory.
6	The descriptor(s) content is received in a PCIe read completion and is written to the appropriate location in the descriptor queue.
7	The DMA fetches the next descriptor and processes its content. As a result, the DMA sends PCIe requests to fetch the packet data from system memory.
8	The packet data is being received from PCIe completions and passes through the transmit DMA that performs all programmed data manipulations (various CPU offloading tasks as checksum offload, TSO offload, etc.) on the packet data on the fly.
9	While the packet is passing through the DMA, it is stored into the transmit FIFO. After the entire packet is stored in the transmit FIFO, it is then forwarded to transmit switch module.
10	The transmit switch arbitrates between host and management packets and eventually forwards the packet to the MAC.
11	The MAC appends the L2 CRC to the packet and sends the packet over the wire using a pre-configured interface.
12	When all the PCIe completions for a given packet are complete, the DMA updates the appropriate descriptor(s).
13	The descriptors are written back to host memory using PCIe posted writes. The head pointer is updated in host memory as well.
14	An interrupt is generated to notify the host driver that the specific packet has been read to the 82599 and the driver can then release the buffer(s).

1.8.2 Receive (Rx) Data Flow

Rx data flow provides a high-level description of all data/control transformation steps needed for receiving Ethernet packets.

Table 1-9 Rx Data Flow

Step	Description
1	The host creates a descriptor ring and configures one of the 82599's receive queues with the address location, length, head, and tail pointers of the ring (one of 128 available Rx queues)
2	The host initializes descriptor(s) that point to empty data buffer(s). The host places these descriptor(s) in the correct location at the appropriate Rx ring.
3	The host updates the appropriate Queue Tail Pointer (RDT).
6	A packet enters the Rx MAC.
7	The MAC forwards the packet to the Rx filter.
8	If the packet matches the pre-programmed criteria of the Rx filtering, it is forwarded to an Rx FIFO.
9	The receive DMA fetches the next descriptor from the appropriate host memory ring to be used for the next received packet.
10	After the entire packet is placed into an Rx FIFO, the receive DMA posts the packet data to the location indicated by the descriptor through the PCIe interface. If the packet size is greater than the buffer size, more descriptors are fetched and their buffers are used for the received packet.
11	When the packet is placed into host memory, the receive DMA updates all the descriptor(s) that were used by the packet data.
12	The receive DMA writes back the descriptor content along with status bits that indicate the packet information including what offloads were done on that packet.
13	The 82599 initiates an interrupt to the host to indicate that a new received packet is ready in host memory.
14	The host reads the packet data and sends it to the TCP/IP stack for further processing. The host releases the associated buffer(s) and descriptor(s) once they are no longer in use.



2.0 Pin Interface

2.1 Pin Assignment

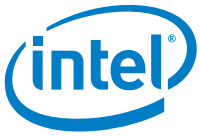
2.1.1 Signal Type Definition

Signal	Definition	DC Specification
In	Input is a standard input-only signal.	Section 11.4.1.2
Out (O)	Totem Pole Output (TPO) is a standard active driver.	Section 11.4.1.2
T/s	Tri-state is a bi-directional, tri-state input/output pin.	Section 11.4.1.2
O/d	Open drain enables multiple devices to share as a wire-OR.	Section 11.4.1.3
A-in	Analog input signals.	Section 11.4.3 and Section 11.4.4
A-out	Analog output signals.	Section 11.4.3 and Section 11.4.4
B	Input BIAS.	-
CML-in	CML input signal.	Section 11.4.5
NCSI-in	NC-SI input signal.	Section 11.4.1.4
NCSI-out	NC-SI output signal.	Section 11.4.1.4
Pu	Internal pull-up.	-
Pd	Internal pull-down.	-

2.1.2 PCIe Symbols and Pin Names

See AC/DC specifications in Section 11.4.3.

Reserved	Pin Name	Ball #	Type	Name and Function
	PE_CLK_p PE_CLK_n	AB23 AB24	A-in	PCIe Differential Reference Clock In. A 100 MHz differential clock input. This clock is used as the reference clock for the PCIe Tx/Rx circuitry and by the PCIe core PLL to generate clocks for the PCIe core logic.
	PET_0_p PET_0_n	Y23 Y24	A-out	PCIe Serial Data Output. A serial differential output pair running at 5 Gb/s or 2.5 Gb/s. This output carries both data and an embedded 5 GHz or 2.5 GHz clock that is recovered along with data at the receiving end.
	PET_1_p PET_1_n	V23 V24	A-out	PCIe Serial Data Output. A serial differential output pair running at 5 Gb/s or 2.5 Gb/s. This output carries both data and an embedded 5 GHz or 2.5 GHz clock that is recovered along with data at the receiving end.
	PET_2_p PET_2_n	T23 T24	A-out	PCIe Serial Data Output. A serial differential output pair running at 5 Gb/s or 2.5 Gb/s. This output carries both data and an embedded 5 GHz or 2.5 GHz clock that is recovered along with data at the receiving end.
	PET_3_p PET_3_n	P23 P24	A-out	PCIe Serial Data Output. A serial differential output pair running at 5 Gb/s or 2.5 Gb/s. This output carries both data and an embedded 5 GHz or 2.5 GHz clock that is recovered along with data at the receiving end.
	PET_4_p PET_4_n	J23 J24	A-out	PCIe Serial Data Output. A serial differential output pair running at 5 Gb/s or 2.5 Gb/s. This output carries both data and an embedded 5 GHz or 2.5 GHz clock that is recovered along with data at the receiving end.
	PET_5_p PET_5_n	G23 G24	A-out	PCIe Serial Data Output. A serial differential output pair running at 5 Gb/s or 2.5 Gb/s. This output carries both data and an embedded 5 GHz or 2.5 GHz clock that is recovered along with data at the receiving end.
	PET_6_p PET_6_n	E23 E24	A-out	PCIe Serial Data Output. A serial differential output pair running at 5 Gb/s or 2.5 Gb/s. This output carries both data and an embedded 5 GHz or 2.5 GHz clock that is recovered along with data at the receiving end.
	PET_7_p PET_7_n	C23 C24	A-out	PCIe Serial Data Output. A serial differential output pair running at 5 Gb/s or 2.5 Gb/s. This output carries both data and an embedded 5 GHz or 2.5 GHz clock that is recovered along with data at the receiving end.
	PER_0_p PER_0_n	AC20 AC21	A-in	PCIe Serial Data Input. A serial differential input pair running at 5 Gb/s or 2.5 Gb/s. An embedded clock present in this input is recovered along with the data.
	PER_1_p PER_1_n	AA20 AA21	A-in	PCIe Serial Data Input. A serial differential input pair running at 5 Gb/s or 2.5 Gb/s. An embedded clock present in this input is recovered along with the data.
	PER_2_p PER_2_n	U20 U21	A-in	PCIe Serial Data Input. A serial differential input pair running at 5 Gb/s or 2.5 Gb/s. An embedded clock present in this input is recovered along with the data.



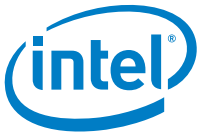
Reserved	Pin Name	Ball #	Type	Name and Function
	PER_3_p PER_3_n	R20 R21	A-in	PCIe Serial Data Input. A serial differential input pair running at 5 Gb/s or 2.5 Gb/s. An embedded clock present in this input is recovered along with the data.
	PER_4_p PER_4_n	K20 K21	A-in	PCIe Serial Data Input. A serial differential input pair running at 5 Gb/s or 2.5 Gb/s. An embedded clock present in this input is recovered along with the data.
	PER_5_p PER_5_n	H20 H21	A-in	PCIe Serial Data Input. A serial differential input pair running at 5 Gb/s or 2.5 Gb/s. An embedded clock present in this input is recovered along with the data.
	PER_6_p PER_6_n	D20 D21	A-in	PCIe Serial Data Input. A serial differential input pair running at 5 Gb/s or 2.5 Gb/s. An embedded clock present in this input is recovered along with the data.
	PER_7_p PER_7_n	B20 B21	A-in	PCIe Serial Data Input. A serial differential input pair running at 5 Gb/s or 2.5 Gb/s. An embedded clock present in this input is recovered along with the data.
	PE_WAKE_N	AA18	O/d	Wake. Pulled to 0b to indicate that a Power Management Event (PME) is pending and the PCIe link should be restored. Defined in the PCIe specifications.
	PE_RST_N	AD18	In	Power and Clock Good Indication. Indicates that power and PCIe reference clock are within specified values. Defined in the PCIe specifications; also called: PCIe Reset and PERST.
	PE_RBIAS PE_RSENSE	M24 N24	B	PCIe BIAS. A 24.9 Ω ±0.5%, 50 ppm resistor should be connected from PE_RBIAS to the chip's 1.2V Analog PCIe supply rail (VCC1P2_PE). Connection should be as close as possible to the chip. Resistor is used for internal impedance compensation and BIAS current generation circuitry. PE_RSENSE is used as sensing node and should be shorted on board to PE_RBIAS as close as possible to the external resistor's pad.

2.1.3 MAUI

See AC/DC specifications in Section 11.4.4 and Section 11.4.5.

Reserved	Pin Name	Ball #	Type	Name and Function
	XA_RBIAS_p XA_RBIAS_n	L2 L1	B	MAUI BIAS. A 1 KΩ ±0.5%, 50 ppm resistor should be connected between XA_RBIAS_p and XA_RBIAS_n and located close to the chip. Resistor generates internal BIAS currents used for impedance compensation. XA_RBIAS_n is internally connected to ground.
	REFCLKIN_p REFCLKIN_n	P2 P1	CML-in	External Reference Clock Input/Crystal Oscillator Input. If an external clock is applied, it must be 25 MHz ±0.01%.
	RX0_L3_p RX0_L3_n	B4 A4	A-in	XAUI Serial Data Input for Port 0. A serial differential input pair running at up to 3.125 Gb/s. An embedded clock present in this input is recovered along with the data.

Reserved	Pin Name	Ball #	Type	Name and Function
	RX0_L2_p RX0_L2_n	D4 D5	A-in	XAUI Serial Data Input for Port 0. A serial differential input pair running at up to 3.125 Gb/s. An embedded clock present in this input is recovered along with the data.
	RX0_L1_p RX0_L1_n	F4 F5	A-in	XAUI Serial Data Input for Port 0. A serial differential input pair running at up to 3.125 Gb/s. An embedded clock present in this input is recovered along with the data.
	RX0_L0_p RX0_L0_n	H4 H5	A-in	XAUI Serial Data Input for Port 0. A serial differential input pair running at up to 3.125 Gb/s. An embedded clock present in this input is recovered along with the data. This lane is also used in BX, BX4, CX4, KX, KR, and SFI modes.
	TX0_L3_p TX0_L3_n	C1 C2	A-out	XAUI Serial Data Output for Port 0. A serial differential output pair running at up to 3.125 Gb/s. This output carries both data and an embedded clock that is recovered along with data at the receiving end.
	TX0_L2_p TX0_L2_n	E1 E2	A-out	XAUI Serial Data Output for Port 0. A serial differential output pair running at up to 3.125 Gb/s. This output carries both data and an embedded clock that is recovered along with data at the receiving end.
	TX0_L1_p TX0_L1_n	G1 G2	A-out	XAUI Serial Data Output for Port 0. A serial differential output pair running at up to 3.125 Gb/s. This output carries both data and an embedded clock that is recovered along with data at the receiving end.
	TX0_L0_p TX0_L0_n	J1 J2	A-out	XAUI Serial Data Output for Port 0. A serial differential output pair running at up to 3.125 Gb/s. This output carries both data and an embedded clock that is recovered along with data at the receiving end. This lane is also used in BX, BX4, CX4, KX, KR, and SFI modes.
	RX1_L3_p RX1_L3_n	U4 U5	A-in	XAUI Serial Data Input for Port 1. A serial differential input pair running at up to 3.125 Gb/s. An embedded clock present in this input is recovered along with the data.
	RX1_L2_p RX1_L2_n	W4 W5	A-in	XAUI Serial Data Input for Port 1. A serial differential input pair running at up to 3.125 Gb/s. An embedded clock present in this input is recovered along with the data.
	RX1_L1_p RX1_L1_n	AA4 AA5	A-in	XAUI Serial Data Input for Port 1. A serial differential input pair running at up to 3.125 Gb/s. An embedded clock present in this input is recovered along with the data.
	RX1_L0_p RX1_L0_n	AC4 AD4	A-in	XAUI Serial Data Input for Port 1. A serial differential input pair running at up to 3.125 Gb/s. An embedded clock present in this input is recovered along with the data. This lane is also used in BX, BX4, CX4, KX, KR, and SFI modes.
	TX1_L3_p TX1_L3_n	T1 T2	A-out	XAUI Serial Data Output for Port 1. A serial differential output pair running at up to 3.125 Gb/s. This output carries both data and an embedded clock that is recovered along with data at the receiving end.
	TX1_L2_p TX1_L2_n	V1 V2	A-out	XAUI Serial Data Output for Port 1. A serial differential output pair running at up to 3.125 Gb/s. This output carries both data and an embedded clock that is recovered along with data at the receiving end.



Reserved	Pin Name	Ball #	Type	Name and Function
	TX1_L1_p TX1_L1_n	Y1 Y2	A-out	XAUI Serial Data Output for Port 1. A serial differential output pair running at up to 3.125 Gb/s. This output carries both data and an embedded clock that is recovered along with data at the receiving end.
	TX1_L0_p TX1_L0_n	AB1 AB2	A-out	XAUI Serial Data Output for Port 1. A serial differential output pair running at up to 3.125 Gb/s. This output carries both data and an embedded clock that is recovered along with data at the receiving end. This lane is also used in BX, BX4, CX4, KX, KR, and SFI modes.

2.1.4 EEPROM

See AC specifications in Section 11.4.2.4.

Reserved	Pin Name	Ball #	Type	Name and Function
	EE_DI	B18	O	Data output to EEPROM.
	EE_DO	A18	In Pu	Data input from EEPROM.
	EE_SK	B19	O	EEPROM serial clock operates at maximum of 2 MHz.
	EE_CS_N	C19	O	EEPROM chip select output.

2.1.5 Serial Flash

See AC specifications in Section 11.4.2.3.

Reserved	Pin Name	Ball #	Type	Name and Function
	FLSH_SI	B6	T/s	Serial data output to the Flash.
	FLSH_SO	A7	In Pu	Serial data input from the Flash.
	FLSH_SCK	A8	T/s	Flash serial clock operates at 12.5 MHz.
	FLSH_CE_N	B7	T/s	Flash chip select output.

2.1.6 SMBus

See the AC specifications in Section 11.4.2.2.

Reserved	Pin Name	Ball #	Type	Name and Function
	SMBCLK	AC19	o/d	SMBus Clock. One clock pulse is generated for each data bit transferred.
	SMBD	AB19	o/d	SMBus Data. Stable during the high period of the clock (unless it is a start or stop condition).
	SMBALRT_N	AA19	o/d	SMBus Alert. Acts as an interrupt pin of a slave device on the SMBus.

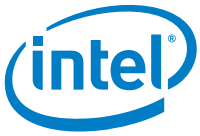
Note: If the SMBus is disconnected, an external pull-up should be used for the SMBCLK, SMBD pins.

2.1.7 I^2C

See the I^2C specification and Section 11.4.2.2 for AC specifications.

Reserved	Pin Name	Ball #	Type	Name and Function
	SCL0	AB12	o/d	I^2C Clock. One clock pulse is generated for each data bit transferred.
	SDA0	AA12	o/d	I^2C Data. Stable during the high period of the clock (unless it is a start or stop condition).
	SCL1	AD17	o/d	I^2C Clock. One clock pulse is generated for each data bit transferred.
	SDA1	AC18	o/d	I^2C Data. Stable during the high period of the clock (unless it is a start or stop condition).

Note: If the I^2C is disconnected, an external pull-up should be used for the clock and data pins.



2.1.8 NC-SI

See AC specifications in Section 11.4.2.5.

Reserved	Pin Name	Ball #	Type	Name and Function
	NCSI_CLK_IN	AC11	NCSI-In	NC-SI Reference Clock Input. Synchronous clock reference for receive, transmit, and control interface. It is a 50 MHz clock ± 50 ppm.
	NCSI_CRS_DV	AB11	NCSI-Out	Carrier Sense/Receive Data Valid (CRS/DV).
	NCSI_RXD_0 NCSI_RXD_1	AA11 AC10	NCSI-Out	Receive Data. Data signals to the BMC.
	NCSI_TX_EN	AB10	NCSI-In	Transmit Enable.
	NCSI_TXD_0 NCSI_TXD_1	AA10 AD11	NCSI-In	Transmit Data. Data signals from the BMC.

Note: If NC-SI is disconnected: an external pull-down should be used for the NCSI_CLK_IN and NCSI_TX_EN pins; a pull-up (10 kΩ) should be used for NCSI_TXD[1:0].

2.1.9 MDIO

See AC specifications in Section 11.4.2.7.

Reserved	Pin Name	Ball #	Type	Name and Function
	MDIO0	AD12	T/s	Management Data. Bi-directional signal for serial data transfers between the 82599 and the PHY management registers for port 0. *Note:* Requires an external pull-up device.
	MDC0	AC12	O	Management Clock. Clock output for accessing the PHY management registers for port 0. MDC clock frequency is Proportional to link speed. At 10 Gb/s Link speed MDC frequency can be set to 2.4 MHz (default) or 24 MHz.
	MDIO1	AC17	T/s	Management Data. Bi-directional signal for serial data transfers between the 82599 and the PHY management registers for port 1. *Note:* Requires an external pull-up device.
	MDC1	AB18	O	Management Clock. Clock output for accessing the PHY management registers for port 1. MDC clock frequency is Proportional to link speed. At 10 Gb/s Link speed MDC frequency can be set to 2.4 MHz (default) or 24 MHz.

2.1.10 Software Defined Pins (SDPs)

See AC specifications in Section 11.4.2.1.

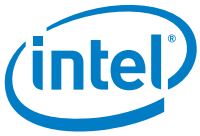
See Section 3.6 for more details on configurable SDPs.

Reserved	Pin Name	Ball #	Type	Name and Function
	SDP0_0 SDP0_1 SDP0_2 SDP0_3 SDP0_4 SDP0_5 SDP0_6 SDP0_7	AD8 AC8 AB8 AA8 AD7 AC7 AB7 AA7	T/s Pu	General Purpose SDPs. 3.3V I/Os for function 0. Can be used to support IEEE1588 Auxiliary devices, Low speed optical module interface SDP0_4 is dedicated input pin for Security enablement. Security offload on both ports is enabled if the Security Enablement flags in the SKU Fuses register are set to 1b and SDP0_4 input pin is driven high. See Section 3.6 for possible usages of the pins.
	SDP1_0 SDP1_1 SDP1_2 SDP1_3 SDP1_4 SDP1_5 SDP1_6 SDP1_7	AC16 AB16 AB17 AA17 AA16 AC15 AB15 AA15	T/s Pu	General purpose SDPs. 3.3V I/Os for function 1. Can be used to support IEEE1588 auxiliary devices, low speed optical module interface See Section 3.6 for possible usages of the pins.

2.1.11 LEDs

See AC specifications in Section 11.4.2.1.

Reserved	Pin Name	Ball #	Type	Name and Function
	LED0_0	AD14	O	Port 0 LED0. Programmable LED that indicates Link-Up (default).
	LED0_1	AC14	O	Port 0 LED1. Programmable LED that indicates 10 Gb/s Link (default).
	LED0_2	AB14	O	Port 0 LED2. Programmable LED that indicates a Link/Activity indication (default).
	LED0_3	AA14	O	Port 0 LED3. Programmable LED that indicates a 1 Gb/s Link (default).
	LED1_0	AD13	O	Port 1 LED0. Programmable LED that indicates Link-Up (default).
	LED1_1	AC13	O	Port 1 LED1. Programmable LED that indicates 10 Gb/s Link (default).
	LED1_2	AB13	O	Port 1 LED2. Programmable LED that indicates a Link/Activity indication (default).
	LED1_3	AA13	O	Port 1 LED3. Programmable LED that indicates a 1 Gb/s Link (default).



2.1.12 RSVD and No Connect Pins

Connecting RSVD pins based on naming convention:

- NC – pin is not connected in the package
- RSVD_NC – reserved pin. Should be left unconnected.
- RSVD_VSS – reserved pin. Should be connected to GND.

Reserved	Pin Name	Ball #	Name and Function
	RSVDA11_NC RSVDA12_NC RSVDA17_NC RSVDA20_NC RSVDA21_NC RSVDB10_NC RSVDB11_NC RSVDB12_NC RSVDB17_NC	A11 A12 A17 A20 A21 B10 B11 B12 B17	RSVD* pins.
	RSVDB8_NC RSVDB9_NC RSVDC10_NC RSVDC11_NC RSVDC12_NC RSVDC13_NC RSVDC14_NC RSVDC15_NC RSVDC16_NC	B8 B9 C10 C11 C12 C13 C14 C15 C16	RSVD* pins.
	RSVDC17_NC RSVDC18_NC RSVDC7_NC RSVDC8_NC RSVDC9_NC RSVDD10_NC RSVDD11_NC RSVDD12_NC RSVDD13_NC	C17 C18 C7 C8 C9 D10 D11 D12 D13	RSVD* pins.
	RSVDD14_NC RSVDD15_NC RSVDD16_NC RSVDD17_NC RSVDD18_NC RSVDD7_NC RSVDD8_NC RSVDD9_NC RSVDE11_NC	D14 D15 D16 D17 D18 D7 D8 D9 E11	RSVD* pins.

Reserved	Pin Name	Ball #	Name and Function
	RSVDE13_NC RSVDE15_NC RSVDE9_NC RSVDJ6_NC RSVDJ7_NC RSVDL23_NC RSVDL24_NC	E13 E15 E9 J6 J7 L23 L24	RSVD* pins.
	RSVDM1_NC RSVDM2_NC RSVDM20_NC RSVDM21_NC RSVDN1_NC RSVDN2_NC RSVDN20_NC RSVDN21_NC RSVDN4_NC	M1 M2 M20 M21 N1 N2 N20 N21 N4	RSVD* pins.
	RSVDN5_NC RSVDT6_NC RSVDT7_NC RSVDW20_NC RSVDW21_NC	N5 T6 T7 W20 W21	RSVD* pins.
	RSVDY11_NC RSVDY13_NC RSVDY15_NC RSVDY17_NC RSVDY18_NC	Y11 Y13 Y15 Y17 Y18	RSVD* pins.
	NCY16 NCY14 NCY12 NCY10 NCY8 NCU7 NCE18 NCE16 NCE14 NCE12 NCE10 NCE8 NCP4 NCL4 NCF20 NCH7	Y16 Y14 Y12 Y10 Y8 U7 E18 E16 E14 E12 E10 E8 P4 L4 F20 H7	NC pins.
	RSVDY9_VSS RSVDV16_VSS RSVDW16_VSS RSVDF21_VSS RSVDE17_VSS	Y9 V16 W16 F21 E17	RSVD* pins.