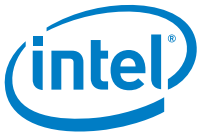


9.3.10.12 Link Control 2 Register (0xD0; RWS)

All RW fields in this register affect the device behavior only through function 0. In function 1 these fields are reserved read as zeros.

Bits	RW	Default	Description
3:0	RWS	0010b	Target Link Speed. This field is used to set the target compliance mode speed when software is using the *Enter Compliance* bit to force a link into compliance mode. Defined encodings are: 0001b = 2.5 GbE target link speed. 0010b = 5 GbE target link speed. All other encodings are reserved. If a value is written to this field that does not correspond to a speed included in the *Supported Link Speeds* field, the result is undefined. The default value of this field is the highest link speed supported by the 82599 (as reported in the *Supported Link Speeds* field of the Link Capabilities register).
4	RWS	0b	Enter Compliance. Software is permitted to force a link to enter compliance mode at the speed indicated in the *Target Link Speed* field by setting this bit to 1b in both components on a link and then initiating a hot reset on the link. The default value of this field following a fundamental reset is 0b.
5	RWS	0b	Hardware Autonomous Speed Disable. When set to 1b, this bit disables hardware from changing the link speed for reasons other than attempting to correct unreliable link operation by reducing link speed.
6	RO	0b	Selectable De-Emphasis. This bit is not applicable and reserved for endpoints.
9:7	RWS	000b	Transmit Margin. This field controls the value of the non de emphasized voltage level at the Transmitter pins. Encodings: 000b = Normal operating range. 001b = 800-1200 mV for full swing and 400-700 mV for half-swing. 010b = (n-1) — Values must be monotonic with a non-zero slope. The value of n must be greater than 3 and less than 7. At least two of these must be below the normal operating range of n: 200-400 mV for full-swing and 100-200 mV for half-swing. 111b = (n) reserved.
10	RWS	0b	Enter Modified Compliance. When this bit is set to 1b, the device transmits modified compliance pattern if the LTSSM enters Polling.Compliance state.
11	RWS	0b	Compliance SOS. When set to 1b, the LTSSM is required to send SOS periodically in between the (modified) compliance patterns.

9.3.10.13 Link Status 2 Register (0xD2; RO)

Bits	RW	Default	Description
0	RO	0b	Current De-emphasis Level. When the link is operating at 5 GT/s speed, this bit reflects the level of de-emphasis. it is undefined when the Link is operating at 2.5 GT/s speed Encodings: 0b = 6 dB. 1b = 3.5 dB.

9.4 PCIe Extended Configuration Space

PCIe configuration space is located in a flat memory-mapped address space. PCIe extends the configuration space beyond the 256 bytes available for PCI to 4096 bytes. The 82599 decodes an additional four bits (bits 27:24) to provide the additional configuration space as shown. PCIe reserves the remaining four bits (bits 31:28) for future expansion of the configuration space beyond 4096 bytes.

The configuration address for a PCIe device is computed using a PCI-compatible bus, device, and function numbers as follows:

31 28	27 20	19 15	14 12	11 2	1 0
0000b	Bus #	Device #	Fun #	Register Address (offset)	00b

PCIe extended configuration space is allocated using a linked list of optional or required PCIe extended capabilities following a format resembling PCI capability structures. The first PCIe extended capability is located at offset 0x100 in the device configuration space. The first Dword of the capability structure identifies the capability/version and points to the next capability.

The 82599 supports the following PCIe extended capabilities:

Table 9-6 Extended Capabilities list

Capability	Offset	Next Header
Advanced Error Reporting Capability	0x100	0x140/0x150/0x000[1]
Serial Number	0x140	0x150/0x000[1]
Alternative RID Interpretation (ARI)	0x150	0x160
IOV support	0x160	0x000

1. Depends on EEPROM settings enabling the serial numbers and ARI/IOV structures.

9.4.1 Advanced Error Reporting Capability (AER)

The PCIe advanced error reporting capability is an optional extended capability to support advanced error reporting. The tables that follow list the PCIe advanced error reporting extended capability structure for PCIe devices.

Byte Offset	Byte 3	Byte 2	Byte 1	Byte 0
0x100	Next Capability Ptr. (0x140)	Version (0x1)	AER Capability ID (0x0001)	
0x104	Uncorrectable Error Status			
0x108	Uncorrectable Error Mask			
0x10C	Uncorrectable Error Severity			
0x110	Correctable Error Status			
0x114	Correctable Error Mask			
0x118	Advanced Error Capabilities and Control Register			
0x11C... 0x128	Header Log			

9.4.1.1 Advanced Error Reporting Enhanced Capability Header Register (0x100; RO)

Bits	RW	Default	Description
15:0	RO	0x0001	Extended Capability ID. PCIe extended capability ID indicating advanced error reporting capability.
19:16	RO	0x1	Version Number. PCIe advanced error reporting extended capability version number.
31:20	RO	0x0140/0x0150/0x0000	Next Capability Pointer. Next PCIe extended capability pointer. See Table 9-6 for possible values of the next capability pointer.

9.4.1.2 Uncorrectable Error Status Register (0x104; RW1CS)

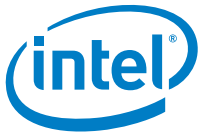
The Uncorrectable Error Status register reports error status of individual uncorrectable error sources on a PCIe device. An individual error status bit that is set to 1b indicates that a particular error occurred; software can clear an error status by writing a 1b to the respective bit. Register is cleared by LAN_PWR_GOOD.

Bits	RW	Default	Description
3:0	RO	0b	Reserved.
4	RW1CS	0b	Data Link Protocol Error Status.
11:5	RO	0b	Reserved.
12	RW1CS	0b	Poisoned TLP Status.
13	RW1CS	0b	Flow Control Protocol Error Status.
14	RW1CS	0b	Completion Timeout Status.
15	RW1CS	0b	Completer Abort Status.
16	RW1CS	0b	Unexpected Completion Status.
17	RW1CS	0b	Receiver Overflow Status.
18	RW1CS	0b	Malformed TLP Status.
19	RW1CS	0b	ECRC Error Status.
20	RW1CS	0b	Unsupported Request Error Status.
21	RO	0b	ACS Violation Status.
31:22	RO	0b	Reserved.

9.4.1.3 Uncorrectable Error Mask Register (0x108; RWS)

The Uncorrectable Error Mask register controls reporting of individual uncorrectable errors by device to the host bridge via a PCIe error message. A masked error (respective bit set in mask register) is not reported to the host bridge by an individual device. Note that there is a mask bit per bit of the Uncorrectable Error Status register.

Bits	RW	Default	Description
3:0	RO	0b	Reserved.
4	RWS	0b	Data Link Protocol Error Mask.
11:5	RO	0b	Reserved.
12	RWS	0b	Poisoned TLP Mask.
13	RWS	0b	Flow Control Protocol Error Mask.
14	RWS	0b	Completion Timeout Mask.
15	RWS	0b	Completer Abort Mask.
16	RWS	0b	Unexpected Completion Mask.



Bits	RW	Default	Description
17	RWS	0b	Receiver Overflow Mask.
18	RWS	0b	Malformed TLP Mask.
19	RWS	0b	ECRC Error Mask.
20	RWS	0b	Unsupported Request Error Mask.
21	RO	0b	ACS Violation Mask.
31:22	RO	0b	Reserved.

9.4.1.4 Uncorrectable Error Severity Register (0x10C; RWS)

The Uncorrectable Error Severity register controls whether an individual uncorrectable error is reported as a fatal error. An uncorrectable error is reported as fatal when the corresponding error bit in the severity register is set. If the bit is cleared, the corresponding error is considered non-fatal.

Bits	RW	Default	Description
3:0	RO	0b	Reserved.
4	RWS	1b	Data Link Protocol Error Severity.
11:5	RO	0b	Reserved.
12	RWS	0b	Poisoned TLP Severity.
13	RWS	1b	Flow Control Protocol Error Severity.
14	RWS	0b	Completion Timeout Severity.
15	RWS	0b	Completer Abort Severity.
16	RWS	0b	Unexpected Completion Severity.
17	RWS	1b	Receiver Overflow Severity.
18	RWS	1b	Malformed TLP Severity.
19	RWS	0b	ECRC Error Severity.
20	RWS	1b	Unsupported Request Error Severity.
21	RO	0b	ACS Violation Severity.
31:22	RO	0b	Reserved.

9.4.1.5 Correctable Error Status Register (0x110; RW1CS)

The Correctable Error Status register reports error status of individual correctable error sources on a PCIe device. When an individual error status bit is set to 1b it indicates that a particular error occurred; software can clear an error status by writing a 1b to the respective bit. Register is cleared by LAN_PWR_GOOD.

Bits	RW	Default	Description
0	RW1CS	0b	Receiver Error Status.
5:1	RO	0b	Reserved.
6	RW1CS	0b	Bad TLP Status.
7	RW1CS	0b	Bad DLLP Status.
8	RW1CS	0b	REPLAY_NUM Rollover Status.
11:9	RO	0b	Reserved.
12	RW1CS	0b	Replay Timer Timeout Status.
13	RW1CS	0b	Advisory Non-Fatal Error Status.
15:14	RO	0b	Reserved.

9.4.1.6 Correctable Error Mask Register (0x114; RWS)

The Correctable Error Mask register controls reporting of individual correctable errors by device to the host bridge via a PCIe error message. A masked error (respective bit set in mask register) is not reported to the host bridge by an individual device. There is a mask bit per bit in the Correctable Error Status register.

Bits	RW	Default	Description
0	RWS	0b	Receiver Error Mask.
5:1	RO	0b	Reserved
6	RWS	0b	Bad TLP Mask.
7	RWS	0b	Bad DLLP Mask.
8	RWS	0b	REPLAY_NUM Rollover Mask.
11:9	RO	0b	Reserved.
12	RWS	0b	Replay Timer Timeout Mask.
13	RWS	1b	Advisory Non-Fatal Error Mask.
15:14	RO	0b	Reserved.

9.4.1.7 Advanced Error Capabilities and Control Register (0x118; RO)

Bits	RW	Default	Description
4:0	ROS	0b	Vector pointing to the first recorded error in the Uncorrectable Error Status register. This is a read-only field that identifies the bit position of the first uncorrectable error reported in the Uncorrectable Error Status register.
5	RO	0b	ECRC Generation Capable. If set, this bit indicates that the function is capable of generating ECRC. This bit is loaded from EEPROM.
6	RWS	0b	ECRC Generation Enable. When set, ECRC generation is enabled.
7	RO	0b	ECRC Check Capable. If set, this bit indicates that the function is capable of checking ECRC. This bit is loaded from EEPROM.
8	RWS	0b	ECRC Check Enable. When set Set, ECRC checking is enabled.

9.4.1.8 Header Log Register (0x11C:128; RO)

The header log register captures the header for the transaction that generated an error. This register is 16 bytes.

Bits	RW	Default	Description
127:0	ROS	0b	Header of the packet in error (TLP or DLLP).

9.4.2 Serial Number

The PCIe device serial number capability is an optional extended capability that can be implemented by any PCIe device. The device serial number is a read-only 64-bit value that is unique for a given PCIe device.

Note: All multi-function devices that implement this capability must implement it for function 0; other functions that implement this capability must return the same device serial number value as that reported by function 0.

Byte Offset	Byte 3	Byte 2	Byte 1	Byte 0
0x140	Next Capability Ptr. (0x150)	Version (0x1)	Serial ID Capability ID (0x0003)	
0x144	Serial Number Register (Lower Dword)			
0x148	Serial Number Register (Upper Dword)			

9.4.2.1 Device Serial Number Enhanced Capability Header Register (0x140; RO)

Bit(s)	RW	Description
15:0	RO	PCIe Extended Capability ID. This field is a PCI-SIG defined ID number that indicates the nature and format of the extended capability. The extended capability ID for the device serial number capability is 0x0003.
19:16	RO	Capability Version. This field is a PCI-SIG defined version number that indicates the version of the capability structure present. *Note:* Must be set to 0x1 for this version of the specification.
31:20	RO	Next Capability Offset. This field contains the offset to the next PCIe capability structure or 0x000 if no other items exist in the linked list of capabilities. The value of this field is 0x150 to point to the ARI capability structure. If ARI/IOV and Serial ID are disabled in EEPROM this field is zero. See Table 9-6.

9.4.2.2 Serial Number Registers (0x144:0x148; RO)

The Serial Number register is a 64-bit field that contains the IEEE defined 64-bit Extended Unique Identifier (EUI-64*). The register at offset 0x144 holds the lower 32 bits and the register at offset 0x148 holds the higher 32 bits. The following figure details the allocation of register fields in the Serial Number register. The table that follows provides the respective bit definitions.

Bit(s)	RW	Description
63:0	RO	PCIe Device Serial Number. This field contains the IEEE defined 64-bit EUI-64*. This identifier includes a 24-bit company ID value assigned by IEEE registration authority and a 40-bit extension identifier assigned by the manufacturer.

The serial number uses the Ethernet MAC address according to the following definition:

Field	Company ID			Extension Identifier				
Order	Addr+0	Addr+1	Addr+2	Addr+3	Addr+4	Addr+5	Addr+6	Addr+7
Most Significant Byte						Least Significant Byte		
Most Significant Bit						Least Significant Bit		

The serial number can be constructed from the 48-bit Ethernet MAC address in the following form:

Field	Company ID			MAC Label		Extension identifier		
Order	Addr+0	Addr+1	Addr+2	Addr+3	Addr+4	Addr+5	Addr+6	Addr+7
Most Significant Bytes						Least Significant Byte		
Most Significant Bit						Least Significant Bit		

In this case, the MAC label is 0xFFFF.

For example, assume that the company ID is (Intel) 00-A0-C9 and the extension identifier is 23-45-67. In this case, the 64-bit serial number is:

Field	Company ID			MAC Label		Extension Identifier		
Order	Addr+0	Addr+1	Addr+2	Addr+3	Addr+4	Addr+5	Addr+6	Addr+7
	00	A0	C9	FF	FF	23	45	67
Most Significant Byte						Least Significant Byte		
Most Significant Bit						Least Significant Bit		

The Ethernet MAC address for the serial number capability is loaded from the Serial Number Ethernet MAC Address EEPROM field (not the same field that is loaded from EEPROM into the RAL and RAH registers).

Note: The official document that defines EUI-64* is: http://standards.ieee.org/regauth/oui/tutorials/EUI64.html

9.4.3 Alternate Routing ID Interpretation (ARI) Capability Structure

In order to allow more than eight functions per endpoint without requesting an internal switch, as is usually needed in virtualization scenarios, the PCI-SIG defines a new capability that allows a different interpretation of the *Bus*, *Device*, and *Function* fields. The ARI capability structure is as follows:

Byte Offset	Byte 3	Byte 2	Byte 1	Byte 0
0x150	Next Capability Ptr. (0x160)	Version (0x1)	ARI Capability ID (0x000E)	
0x154	ARI Control Register		ARI Capabilities	

9.4.3.1 PCIe ARI Header Register (0x150; RO)

Field	Bit(s)	RW	Init Val	Description
ID	15:0	RO	0x000E	PCIe Extended Capability ID. PCIe extended capability ID for the alternative RID interpretation.
Version	19:16	RO	1b	Capability Version. This field is a PCI-SIG defined version number that indicates the version of the capability structure present. Must be 0x1 for this version of the specification.
Next Capability Ptr.	31:20	RO	0x160	Next Capability Offset. This field contains the offset to the next PCIe extended capability structure. The value of the 0x160 points to the IOV structure.

9.4.3.2 PCIe ARI Capabilities and Control Register (0x154; RO)

Field	Bit(s)	RW	Init Val	Description
Reserved	0	RO	0b	Not supported in the 82599.
Reserved	1	RO	0b	Not supported in the 82599.
Reserved	7:2	RO	0b	Reserved.
NFP	15:8	RO	0x1 (func 0) 0x0 (func 1)[1]	Next Function Pointer. This field contains the pointer to the next physical function configuration space or 0x0000 if no other items exist in the linked list of functions. Function 0 is the start of the link list of functions.
Reserved	16	RO	0b	Not supported in the 82599.
Reserved	17	RO	0b	Not supported in the 82599.
Reserved	19:18	RO	00b	Reserved.

Field	Bit(s)	RW	Init Val	Description
Reserved	22:20	RO	0b	Not supported in the 82599.
Reserved	31:23	RO	0b	Reserved

1. Even if port 0 and port 1 are switched or function zero is a dummy function, this register should keep its attributes according to the function number. If LAN1 is disabled, the value of this field in function zero should be zero.

9.4.4 IOV Capability Structure

This is the new structure used to support the IOV capabilities reporting and control. The following tables shows the possible implementations of this structure in the 82599.

Byte Offset	Byte 3	Byte 2	Byte 1	Byte 0
0x150	Next Capability Ptr. (0x160)	Version (0x1)	Capability ID (0x000E)	
0x154	Control Register		Capabilities	
0x160	Next Capability Offset (0x0)	Version (0x1)	IOV Capability ID (0x0010)	
0x164	SR IOV Capabilities			
0x168	SR IOV Status		SR IOV Control	
0x16C	Total VFs (RO)		Initial VF (RO)	
0x170	Reserved	Function Dependency Link (RO)	Num VF (RW)	
0x174	VF Stride (RO)		First VF Offset (RO)	
0x178	VF Device ID		Reserved	
0x17C	Supported Page Size (0x553)			
0x180	system page Size (RW)			
0x184	VF BAR0 — Low (RW)			
0x188	VF BAR0 — High (RW)			
0x18C	VF BAR2 (RO)			
0x190	VF BAR3 — Low (RW)			
0x194	VF BAR3- High (RW)			
0x198	VF BAR5 (RO)			
0x19C	VF Migration State Array Offset (RO)			

9.4.4.1 PCIe SR-IOV Header Register (0x160; RO)

Field	Bit(s)	RW	Init Val	Description
ID	15:0	RO	0x0010	PCIe Extended Capability ID. PCIe extended capability ID for the SR-IOV capability.
Version	19:16	RO	0x1	Capability Version. This field is a PCI-SIG defined version number that indicates the version of the capability structure present. Must be 0x1 for this version of the specification.
Next pointer	31:20	RO	0x0	Next Capability Offset. This field contains the offset to the next PCIe extended capability structure or 0x000 if no other items exist in the linked list of capabilities.

9.4.4.2 PCIe SR-IOV Capabilities Register (0x164; RO)

Field	Bit(s)	RW	Init Val	Description
Reserved	0	RO	0b	Not supported in the 82599.
Reserved	20:1	RO	0x0	Reserved.
Reserved	31:21	RO	0x0	Not supported in the 82599.

9.4.4.3 PCIe SR-IOV Control/Status Register (0x168; RW)

Field	Bit(s)	RW	Init Val	Description
VFE	0	RW	0b	VF Enable/Disable. VF Enable manages the assignment of VFs to the associated PF. If *VF Enable* is set to 1b, VFs must be enabled, associated with the PF, and exists in the PCIe fabric. When enabled, VFs must respond to and can issue PCIe transactions following all other rules for PCIe functions. If set to 0b, VFs must be disabled and not visible in the PCIe fabric; VFs cannot respond to or issue PCIe transactions. In addition, if *VF Enable* is cleared after having been set, all of the VFs must no longer: • Issue PCIe transactions • Respond to configuration space or memory space accesses. The behavior must be as if an FLR was issued to each of the VFs. Specifically, VFs must not retain any context after *VF Enable* has been cleared. Any errors already logged via PF error reporting registers, remain logged. However, no new VF errors must be logged after *VF Enable* is cleared.
Reserved	1	RO	0b	Not supported in the 82599.
Reserved	2	RO	0b	Not supported in the 82599.

Field	Bit(s)	RW	Init Val	Description
VF MSE	3	RW	0b	Memory Space Enable for Virtual Functions. VF MSE controls memory space enable for all VFs associated with this PF as with the Memory Space Enable bit in a functions PCI command register. The default value for this bit is 0b. When VF Enable is 1, virtual function memory space access is permitted only when VF MSE is Set. VFs shall follow the same error reporting rules as defined in the base specification if an attempt is made to access a virtual functions memory space when VF Enable is 1 and VF MSE is zero. *Implementation Note:* Virtual functions memory space cannot be accessed when VF Enable is zero. Thus, VF MSE is "don't care" when VF Enable is zero, however, software may choose to set VF MSE after programming the VF BARn registers, prior to setting VF Enable to 1.
VF ARI	4	RW (func 0) RO (func 1)[1]	0b	VF ARI Enable. Device can locate VFs in function numbers 8 to 255 of the captured bus number.
Reserved	15:5	RO	0x0	Reserved.
Reserved	16	RO	0b	Not implemented in the 82599.
Reserved	31:17	RO	0b	Reserved.

1. Even if port 0 and port 1 are switched or function zero is a dummy function, this field should keep it's attributes according to the function number.

9.4.4.4 PCIe SR-IOV Max/Total VFs Register (0x16C; RO)

Field	Bit(s)	RW	Init Val	Description
InitialVFs	15:0	RO	64	InitialVFs. Indicates the number of VFs that are initially associated with the PF. If *VF Migration Capable* is cleared, this field must contain the same value as TotalVFs. In the 82599 this parameter is equal to the TotalVFs in this register.
TotalVFs	31:16	RO	64	TotalVFs. Defines the maximum number of VFs that can be associated with the PF. This field is loaded from the *Max VFs* field in the IOV Control Word 1 in the EEPROM.

9.4.4.5 PCIe SR-IOV Num VFs Register (0x170; RW)

Field	Bit(s)	RW	Init Val	Description
NumVFs	15:0	RW	0x0	Num VFs. Defines the number of VFs software has assigned to the PF. Software sets NumVFs to any value between one and the TotalVFs as part of the process of creating VFs. NumVFs VFs must be visible in the PCIe fabric after both NumVFs is set to a valid value and VF Enable is set to 1b.
FDL	23:16	RO	0x0 (func 0) 0x1 (func 1)[1]	Function Dependency Link. Defines dependencies between physical functions allocation. In the 82599 there are no constraints.

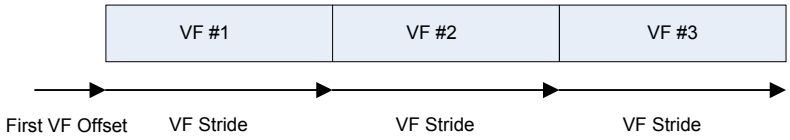
Field	Bit(s)	RW	Init Val	Description
Reserved	31:24	RO	0	Reserved.

1. Even if port 0 and port 1 are switched or function zero is a dummy function, this register should keep it's attributes according to the function number.

9.4.4.6 PCIe SR-IOV VF RID Mapping Register (0x174; RO)

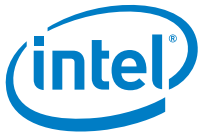
Field	Bit(s)	RW	Init Val	Description
FVO	15:0	RO	0x180	First VF offset. Defines the requestor ID (RID) offset of the first VF that is associated with the PF that contains this capability structure. The first VFs 16-bit RID is calculated by adding the contents of this field to the RID of the PF containing this field. The content of this field is valid only when *VF Enable* is set. If *VF Enable* is 0b, the contents are undefined. If the ARI *Enable* bit is set, this field changes to 0x80.
VFS	31:16	RO	0x2[1]	VF stride. Defines the requestor ID (RID) offset from one VF to the next one for all VFs associated with the PF that contains this capability structure. The next VFs 16-bit RID is calculated by adding the contents of this field to the RID of the current VF. The contents of this field is valid only when *VF Enable* is set and *NumVFs* is non-zero. If *VF Enable* is 0b or if NumVFs is zero, the contents are undefined.

1. See Section 7.10.2.6.1.



9.4.4.7 PCIe SR-IOV VF Device ID Register (0x178; RO)

Field	Bit(s)	RW	Init Val	Description
DEVID	31:16	RO	0x10ED	VF Device ID. This field contains the device ID that should be presented for every VF to the Virtual Machine (VM). The value of this field can be read from the IOV Control Word 2 in the EEPROM.
Reserved	15:0	RO	0x0	Reserved.



9.4.4.8 PCIe SR-IOV Supported Page Size Register (0x17C; RO)

Field	Bit(s)	RW	Init Val	Description
Supported page Size	31:0	RO	0x553	For PFs that supports the stride-based BAR mechanism, this field defines the supported page sizes. This PF supports a page size of 2^(n+12) if bit n is set. For example, if bit 0 is Set, the Endpoint (EP) supports 4KB page sizes. Endpoints are required to support 4 KB, 8 KB, 64 KB, 256 KB, 1 MB and 4 MB page sizes. All other page sizes are optional.

9.4.4.9 PCIe SR-IOV System Page Size Register (0x180; RW)

Field	Bit(s)	RW	Init Val	Description
Page size	31:0	RW	0x1	This field defines the page size the system uses to map the PF's and associated VFs' memory addresses. Software must set the value of the *System Page Size* to one of the page sizes set in the *Supported Page Sizes* field. As with *Supported Page Sizes*, if bit n is set in *System Page Size*, the PF and its associated VFs are required to support a page size of 2^(n+12). For example, if bit 1 is set, the system is using an 8 KB page size. The results are undefined if more than one bit is set in *System Page Size*. The results are undefined if a bit is set in *System Page Size* that is not set in *Supported Page Sizes*. When *System Page Size* is set, the PF and associated VFs are required to align all BAR resources on a *System Page Size* boundary. Each BAR size, including *VF BARn Size* (described later) must be aligned on a *System Page Size* boundary. Each BAR size, including VF *BARn Size* must be sized to consume a multiple of *System Page Size* bytes. All fields requiring page size alignment within a function must be aligned on a *System Page Size* boundary. *VF Enable* must be zero when *System Page Size* is set. The results are undefined if *System Page Size* is set when *VF Enable* is set.

9.4.4.10 PCIe SR-IOV BAR 0 — Low Register (0x184; RW)

Field	Bit(s)	RW	Init Val	Description
Mem	0	RO	0b	0b indicates memory space.
Mem Type	2:1	RO	10b	Indicates the address space size. 10b = 64-bit. This bit is loaded from the IOV Control word in the EEPROM.
Prefetch Mem	3	RO	0b*	0b = Non-prefetchable space. 1b = Prefetchable space. This bit is loaded from the IOV Control word in the EEPROM.
Memory Address Space	31:4	RW	0x0	Which bits are RW bits and which are RO to 0x0 depend on the memory mapping window size. The size is a maximum between 16 KB and the page size.

9.4.4.11 PCIe SR-IOV BAR 0 — High Register (0x188; RW)

Field	Bit(s)	RW	Init Val	Description
BAR0 — MSB	31:0	RW	0x0	MSB part of BAR0.

9.4.4.12 PCIe SR-IOV BAR 2 Register (0x18C; RO)

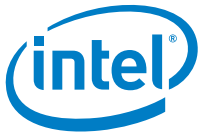
Field	Bit(s)	RW	Init Val	Description
BAR2	31:0	RO	0x0	This BAR is not used.

9.4.4.13 PCIe SR-IOV BAR 3 — Low Register (0x190; RW)

Field	Bit(s)	RW	Init Val	Description
Mem	0	RO	0b	0b indicates memory space.
Mem Type	2:1	RO	10b	Indicates the address space size. 10b = 64-bit. This bit is loaded from the IOV Control word in the EEPROM.
Prefetch Mem	3	RO	0b*	0b = Non-prefetchable space 1b = Prefetchable space This bit is loaded from the IOV Control word in the EEPROM.
Memory Address Space	31:4	RW	0x0	Which bits are RW bits and which are RO to 0x0 depend on the memory mapping window size. The size is a maximum between 16 KB and page size.

9.4.4.14 PCIe SR-IOV BAR 3 — High Register (0x194; RW)

Field	Bit(s)	RW	Init Val	Description
BAR3 — MSB	31:0	RW	0x0	MSB part of BAR3.



9.4.4.15 PCIe SR-IOV BAR 5 Register (0x198; RO)

Field	Bit(s)	RW	Init Val	Description
BAR5	31:0	RO	0x0	This BAR is not used.

9.4.4.16 PCIe SR-IOV VF Migration State Array Offset Register (0x19C; RO)

Field	Bit(s)	RW	Init Val	Description
Reserved	2:0	RO	0x0	Not implemented in the 82599.
Reserved	31:0	RO	0x0	Not implemented in the 82599.

9.5 Virtual Functions Configuration Space

The configuration space reflected to each of the VF is a sparse version of the physical function configuration space. The following table describes the behavior of each register in the VF configuration space.

Table 9-7 VF PCIe Configuration Space

Section	Offset	Name	VF behavior	Notes
PCI Mandatory Registers	0	Vendor ID	RO — 0xFFFF	
	2	Device ID	RO — 0xFFFF	
	4	Command	RW	See Section 9.5.1.1.
	6	Status	Per VF	See Section 9.5.1.2.
	8	RevisionID	RO as PF	
	9	Class Code	RO as PF	
	C	Cache Line Size	RO — 0x0	
	D	Latency Timer	RO — 0x0	
	E	Header Type	RO — 0x0	
	F	Reserved	RO — 0x0	
	10 — 27	BARs	RO — 0x0	Emulated by VMM.
	28	CardBus CIS	RO — 0x0	Not used.
	2C	Sub Vendor ID	RO as PF	
	2E	Sub System	RO as PF	
	30	Expansion ROM	RO — 0x0	Emulated by VMM.
	34	Cap Pointer	RO — 0x70	Next = MSI-X capability.
	3C	Int Line	RO — 0x0	
	3D	Int Pin	RO — 0x0	
	3E	Max Lat/Min Gnt	RO — 0x0	

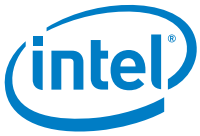


Table 9-7 VF PCIe Configuration Space (Continued)

Section	Offset	Name	VF behavior	Notes
MSI-X Capability	70	MSI-X Header	RO — 0xA011	Next = PCIe capability.
	72	MSI-x Message Control	per VF	See Section 9.5.2.1.
	74	MSI-X table Address	RO — as PF	
	78	MSI-X PBA Address	RO	
PCIe Capability	A0	PCIe Header	RO — 0x0010	Next = Last capability.
	A2	PCIe Capabilities	RO — 0x0	
	A4	PCIe Dev Cap	RO — 0x0	
	A8	PCIe Dev Ctrl	RW	As PF apart from FLR. See Table 9.5.2.2.1.
	AA	PCIe Dev Status	per VF	See Table 9.5.2.2.2.
	AC	PCIe Link Cap	RO — 0x0	
	B0	PCIe Link Ctrl	RO — 0x0	
	B2	PCIe Link Status	RO — 0x0	
	C4	PCIe Dev Cap 2	RO — 0x0	
	C8	PCIe Dev Ctrl 2	RO — 0x0	
	D0	PCIe Link Ctrl 2	RO — 0x0	
	D2	PCIe Link Status 2	RO — 0x0	
AER Capability	100	AER — Header	RO — 0x15010001	Next = ARI structure.
	104	AER — Uncorr Status	per VF	See Section 9.5.2.3.
	108	AER — Uncorr Mask	RO — 0x0	
	10C	AER — Uncorr Severity	RO — 0x0	
	110	AER — Corr Status	Per PF	
	114	AER — Corr Mask	RO — 0x0	
	118	AER — Cap/Ctrl	RO as PF	
	11C — 128	AER — Error Log	Shared two logs for all VFs	Same structure as in PF. In case of overflow, the header log is filled with ones.

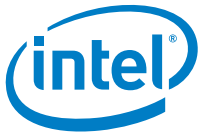
Table 9-7 VF PCIe Configuration Space (Continued)

Section	Offset	Name	VF behavior	Notes
ARI Capability	150	ARI — Header	0x0001000E	Next = Last extended Capability.
	154	ARI — Cap/Ctrl	RO — 0X0	

9.5.1 Mandatory Configuration Space

9.5.1.1 VF Command Register (0x4; RW)

Bit(s)	RW	Init Val	Description
0	RO	0b	I/O Access Enable (IOAE). RO as zero field.
1	RO	0b	Memory Access Enable (MAE). RO as zero field.
2	RW	0b	Bus Master Enable (BME). Disabling this bit prevents the associated VF from issuing any memory or I/O requests. *Note:* As MSI/MSI-X interrupt messages are in-band memory writes, disabling the bus master enable bit disables MSI/MSI-X interrupt messages as well. Requests other than memory or I/O requests are not controlled by this bit. *Note:* The state of active transactions is not specified when this bit is disabled after being enabled. The device can choose how it behaves when this condition occurs. Software cannot count on the device retaining state and resuming without loss of data when the bit is re-enabled. Transactions for a VF that has its *Bus Master Enable* set must not be blocked by transactions for VFs that have their *Bus Master Enable* cleared.
3	RO	0b	Special Cycle Enable (SCM). Hard wired to 0b
4	RO	0b	MWI Enable (MWIE). Hard wired to 0b.
5	RO	0b	Palette Snoop Enable (PSE). Hard wired to 0b.
6	RO	0b	Parity Error Response (PER). Zero for VFs.
7	RO	0b	Wait Cycle Enable (WCE). Hard wired to 0b.
8	RO	0b	SERR# Enable (SERRE). Zero for VFs.
9	RO	0b	Fast Back-to-Back Enable (FB2BE). Hard wired to 0b.



Bit(s)	RW	Init Val	Description
10	RO	0b	Interrupt Disable (INTD). Hard wired to 0b.
15:11	RO	0b	Reserved (RSV).

9.5.1.2 VF Status Register (0x6; RW)

Bits	RW	Init Val	Description
2:0	RO	0x0	Reserved (RSV).
3	RO	0b	Interrupt Status (IS). Hard wired to 0b.
4	RO	1b	New Capabilities (NC). Indicates that the 82599 VFs implement extended capabilities. The 82599 VFs implement a capabilities list, to indicate that it supports MSI-X and PCIe extensions.
5	RO	0b	66 MHz Capable (66E). Hard wired to 0b.
6	RO	0b	Reserved (RSV).
7	RO	0b	Fast Back-to-Back Capable (FB2BC). Hard wired to 0b.
8	RW1C	0b	Data Parity Reported (MPERR).
10:9	RO	00b	DEVSEL Timing (DEVSEL). Hard wired to 0b.
11	RW1C	0b	Signaled Target Abort (STA).
12	RW1C	0b	Received Target Abort (RTA).
13	RW1C	0b	Received Master Abort (RMA).
14	RW1C	0b	Signaled System Error (SSERR).
15	RW1C	0b	Detected Parity Error (DSERR).

9.5.2 PCI Capabilities

9.5.2.1 MSI-X Capability

The only registers with a different layout than the PF for MSI-X, is the control register.

Note: The message address and data registers in enhanced mode use the first MSI-X entry of each VF in the regular MSI-X table.

9.5.2.1.1 VF MSI-X Control Register (0x72; RW).

Bits	RW	Init Val	Description
10:0	RO	0x002[1]	Table Size (TS).
13:11	RO	0x0	Reserved (RSV).
14	RW	0b	Function Mask (Mask).
15	RW	0b	MSI-X Enable (En).

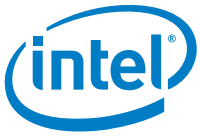
1. Default value is read from the EEPROM.

9.5.2.1.2 MSI-X PBA Register (0x78; RO)

Bits	RW	Default	Description
31:3	RO	0x400	PBA Offset. Used as an offset from the address contained by one of the function's BARs to point to the base of the MSI-X PBA. The lower three PBA BIR bits are masked off (set to zero) by software to form a 32-bit Qword-aligned offset. This value is changed by hardware to be half of the value programmed to the IOV System Page Size register.
2:0	RO	0x3	PBA BIR. Indicates which one of a function's BARs, located beginning at 0x10 in configuration space, is used to map the function's MSI-X PBA into memory space. A BIR value of three indicates that the PBA is mapped in BAR 3.

9.5.2.2 PCIe Capability Registers

The device control and device status registers have some fields which are specific per VF.



9.5.2.2.1 VF Device Control Register (0xA8; RW)

Bits	RW	Default	Description
0	RO	0b	Correctable Error Reporting Enable. Zero for VFs.
1	RO	0b	Non-Fatal Error Reporting Enable. Zero for VFs.
2	RO	0b	Fatal Error Reporting Enable. Zero for VFs.
3	RO	0b	Unsupported Request Reporting Enable. Zero for VFs.
4	RO	0b	Enable Relaxed Ordering. Zero for VFs.
7:5	RO	0b	Max Payload Size. Zero for VFs.
8	RO	0b	Not implemented in the 82599.
9	RO	0b	Not implemented in the 82599.
10	RO	0b	Auxiliary Power PM Enable. Zero for VFs.
11	RO	0b	Reserved
14:12	RO	000b	Max Read Request Size. Zero for VFs.
15	RW	0b	Initiate Function Level Reset. Specific to each VF.

9.5.2.2.2 VF Device Status Register (0xAA; RO)

Bits	RW	Default	Description
0	RO	0b	Correctable Detected. Indicates status of correctable error detection. Zero for VF.
1	RO	0b	Non-Fatal Error Detected. Indicates status of non-fatal error detection. Zero for VF.
2	RO	0b	Fatal Error Detected. Indicates status of fatal error detection. Zero for VF.
3	RO	0b	Unsupported Request Detected. Indicates that the 82599 received an unsupported request. This field is identical in all functions. The 82599 can't distinguish which function caused an error. Zero for VF.

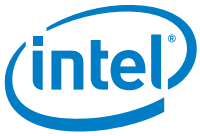
Bits	RW	Default	Description
4	RO	0b	Aux Power Detected. Zero for VFs.
5	RO	0b	Transaction Pending. Specific per VF. When set, indicates that a particular function (PF or VF) has issued non-posted requests that have not been completed. A function reports this bit cleared only when all completions for any outstanding non-posted requests have been received.
15:6	RO	0x00	Reserved.

9.5.2.3 AER Registers

The following registers in the AER capability have a different behavior in a VF function.

Uncorrectable Error Status Register (0x104; RW1C)

Bits	RW	Default	Description
3:0	RO	0x0	Reserved.
4	RO	0b	Data Link Protocol Error Status.
5	RO	0b	Surprise Down Error Status (Optional).
11:6	RO	0x0	Reserved.
12	RW1C	0b	Poisoned TLP Status.
13	RO	0b	Flow Control Protocol Error Status.
14	RW1C	0b	Completion Timeout Status.
15	RW1C	0b	Completer Abort Status.
16	RW1C	0b	Unexpected Completion Status.
17	RO	0b	Receiver Overflow Status.
18	RO	0b	Malformed TLP Status.
19	RO	0b	ECRC Error Status.
20	RW1C	0b	Unsupported Request Error Status. When caused by a function that claims a TLP.
21	RO	0b	ACS Violation Status.
31:21	RO	0x0	Reserved.



10.0 Manageability

Network management is an important requirement in today's networked computer environment. Software-based management applications provide the ability to administer systems while the operating system is functioning in a normal power state (not in a pre-boot state or powered-down state). The Intel® System Management Bus (SMBus) Interface and the Network Controller Sideband Interface (NC-SI) fill the management void that exists when the operating system is not running or fully functional. This is accomplished by providing mechanisms by which manageability network traffic can be routed to and from a Baseboard Management Controller (BMC), or simply MC.

This section describes the supported management interfaces and hardware configurations for platform system management. It describes the interfaces to an external BMC, the partitioning of platform manageability among system components, and the functionality provided by the 82599 in each platform configuration.

10.1 Platform Configurations

This section describes the hardware configurations for platform management. It describes the partitioning of platform manageability among system components and the functionality provided by the 82599 in each of the platform configurations.

The 82599 supports pass-through manageability to an on-board BMC. The link between the 82599 and the BMC is either SMBus or the DMTF NC-SI.

10.1.1 On-Board BMC Configurations

Figure 10-1 (left option) depicts an SMBus-only connection between the 82599 and the BMC. The SMBus is used for all communication between the 82599 and the BMC (pass-through traffic, configuration, and status). The protocol details for this configuration follow the SMBus commands described in Section 10.2.2. Figure 10-1 (right option) depicts an NC-SI-only connection between the 82599 and the BMC. The NC-SI is used for all communication between the 82599 and the BMC (pass-through traffic, configuration, and status). The protocol details for this configuration follow the DMTF NC-SI protocol.

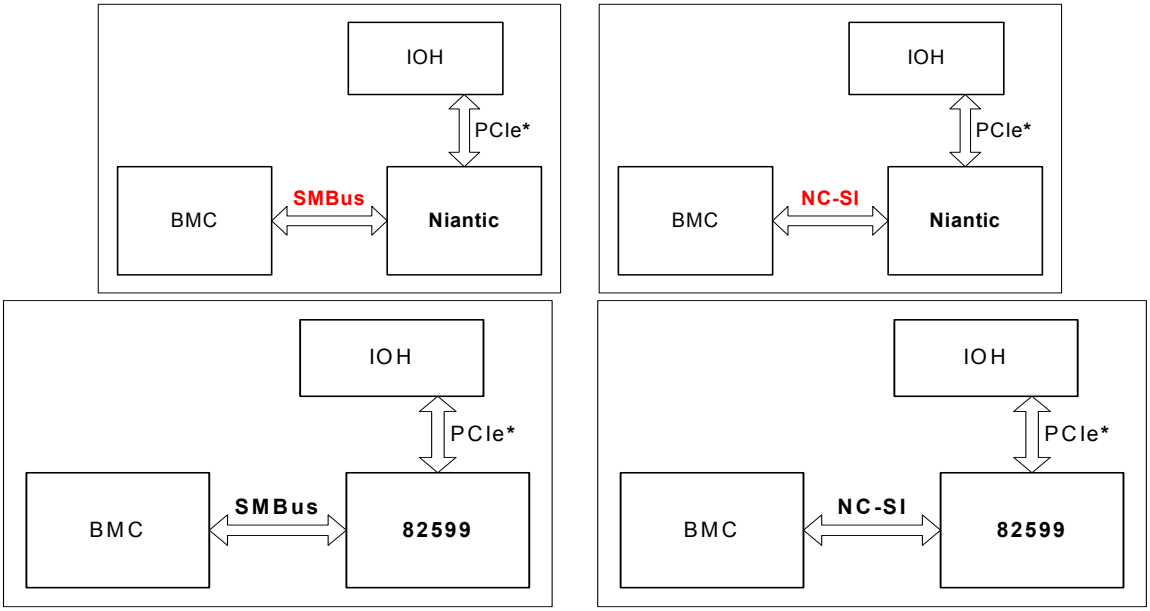


Figure 10-1 the 82599 to BMC Connectivity Through SMBus Link or NC-SI Link

Refer to the sections that follow for a description of the traffic types that use the NC-SI and/or SMBus interfaces.

10.1.2 82599 NIC

BMC connection to a NIC is not expected.

10.2 Pass Through (PT) Functionality

The 82599 supports traffic pass through to an external BMC. The pass-through traffic is carried through an NC-SI interface or SMBus (legacy devices) based on the Redirection Sideband Interface setting in the EEPROM (loaded on power up). The usable bandwidth for either direction is up to 100 Mb/s in NC-SI mode and up to 400 Kb/s in SMBus mode. Supplemental descriptions on SMBus and NC-SI interfaces can be found in Section 3.2 and in Section 3.3. The following list describes usage models for the pass through traffic:

- BMC management traffic
- Keyboard or mouse traffic for KVM (low data rate)
- Video traffic for KVM (low average rate of 150 Kb/s to 200 Kb/s) — transmit only
- USB 2.0 redirect (up to 50 Mb/s)
- IDE redirect for remote CD/floppy (rate — priority 1 — CDx7 = 1.05 Mb/s. Priority 2 — CDx24 = 64 Mb/s