SR-IOV Configuration Guide

Intel® Ethernet CNA X710 & XL710 on SuSE Linux* Enterprise Server 12*

Technical Brief

Networking Division (ND)

August 2015

Revision History

Revision	Date	Comments
1.0	August 5, 2015	Initial release (Intel public).

Contents

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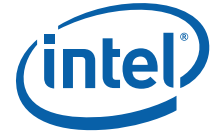
1.0   Introduction

Since its inception in 1973, Intel has been vital to the development of the Ethernet, and continues to be the Industry leader. For over 40 years, the Ethernet has been growing to accommodate increasing bandwidth needs for complex multi-media; streaming video, music and voice data, for example. Beginning with One Gigabit Ethernet, expanding to 10 Gigabit, and now introducing 40 Gigabit Ethernet, computing and storage resource needs continue to grow.

1.1   Intel and the Ethernet

Following is a brief history of Intel and the Ethernet:

- 1994: Intel ships the world's first 10/100 Mb/s Network Interface Card (NIC).

- 1997: Intel ships the first single-chip 10/100 Mb/s controller.

- 2001: Intel ships the first single-chip 10/100/1000 Mb/s controller.

- 2002: Intel ships the first XPAK Multimode Optical Transceiver, delivering 10-Gigabit Ethernet (GbE) and 10-Gigabit Fibre Channel transport for storage systems at half the cost, a third less power consumption, and a third of the size of earlier solutions.

- 2003: Intel ships the world's first 10 Gigabit Ethernet NIC.

- 2006: Intel introduces the first low-profile quad-port Ethernet NIC. By incorporating 4-Gigabit Ethernet connections in a low-profile PCI Express slot, it improved server throughput and rack density at the same time.

- 2007: Intel releases first "initiator" source code to enable Linux implementations of Fibre Channel over Ethernet (FCoE). By allowing fiber channel SAN traffic to run over Gigabit Ethernet networks, FCoE enables consolidation of storage area network (SAN) and LAN traffic, simplifying network infrastructure in data centers.

1.2   Intel® Ethernet Controller XL710

The 40 Gigabit XL710 Controller is designed for flexibility, with configurable port speeds of up to 2 x 40 GbE, or 4 x 10 GbE, ensuring a smooth transition to 40 GbE, It also provides a 222% increase in Gigabits per Watt in adapter power for approximately half the power cost when compared to using two previous generation dual-port adapters.

The XL710 offers the following features:

- 10/40 GbE Controller (Dual and Single 40 GbE, Quad and Dual 10 GbE configurations).

- PCI Express* (PCIe) 3.0, x8 including Direct I/O optimizations via TLP Processing Hints (TPH).

- Intelligent Off-load to enable high-performance with Intel® Xeon® servers.

- Network Virtualization off-loads including VXLAN and NVGRE.

- Industry-leading I/O virtualization innovations and performance with broad hypervisor and standards support.

- Intel® Ethernet Flow Director (for hardware application traffic steering).

- Excellent small packet performance for network appliances and NFV.

- Intel® Data Plane Developer Kit Optimize.

- Unified Networking providing a single wire for LAN and storage: NAS(SMB,NFS) and SAN (iSCSI, FCoE).

The following are the Intel 40 Gigabit XL710 Controller-based Dual and Quad Adapter offerings:

Note: These boards do NOT ship with optics installed. Optics must be purchased separately.

- Intel® Ethernet Converged Network Adapter X710-DA4
    - X710DA4FH, XL710DA4FHBLK (Retail, Quad Port FH)
    - X710DA4FHG1P5 (OEM Gen, Quad Port FH)
    - X710DA4G1P5 (OEM Gen, Quad Port LP)
- Intel® Ethernet Converged Network Adapter X710-DA2
    - X710DA2, XL710DA2BLK (Retail, Dual Port)
    - X710DA2G1P5 (OEM Gen, Dual Port)
- Intel® Ethernet Converged Network Adapter XL710-QDA2
    - XL710QDA2, XL710QDA2BLK (Retail, Dual Port)
    - XL710QDA2G1P5 (OEM Gen, Dual Port)
- Intel® Ethernet Converged Network Adapter XL710-QDA1
    - XL710QDA1, XL710QDA1BLK (Retail, Single Port)
    - XL710QDA1G1P5 (OEM Gen, Single Port)

Power efficiency is critical to IT specialists as energy consumption is a real concern in data center operations. The Intel Ethernet Controller provides a low-power interface to eliminate the need for additional power. It also offers the manageability IT personnel require for remote control and alerting.

This controller provides multiple interface options, a smaller footprint for reduced infrastructure and cabling costs, lower power consumption, and intelligent off-loads that do not require disabling key features and flow direction to balance high volume traffic flows.

1.3   I/O Virtualization

The Intel® Ethernet Server Adapter X710 & XL710 family of adapters delivers numerous industry-leading features that are helping data center administrators implement innovative solutions for difficult and challenging connectivity problems. I/O Virtualization is one of the fastest growing usage models within the data center.

The X710 & XL710 family of adapters provides the ability to create Virtual Functions (VFs) that are identical instantiations of the Physical Functions (PFs). VFs are capable of providing up to 10 GbE or 40 GbE connectivity to Virtual Machines (VMs) within a virtualized operating system framework. The Intel® Ethernet Server Adapter X710 supports up to 32 VFs per port, for a total of up to 128 VFs per adapter. The Intel® Ethernet Server Adapter XL710 supports up to 64 VFs per port, for a total of up to 128 VFs per adapter.

This document shows how to make use of these VFs using Xen* Virtualization, which is an integral part of SuSE Linux Enterprise Server version 12

1.3.1   Hardware Requirements

- An Intel® Ethernet Converged Network Adapter X710 or XL710.
- A server platform that supports Intel® Virtualization Technology for Directed I/O (VT-d) and the PCI-SIG* Single Root I/O Virtualization and Sharing (SR-IOV) specification.
- A server platform with an available PCI Express*: x8 5.0 Gb/s (Gen2) or x8 8.0 Gb/s (Gen3) slot.

1.3.2 Software Requirements

- SuSE Linux Enterprise Server version 12.

- Intel® Ethernet Converged Network Adapter X710 or XL710 Linux Drivers for PF and VF (http://sourceforge.net/projects/e1000/files/).
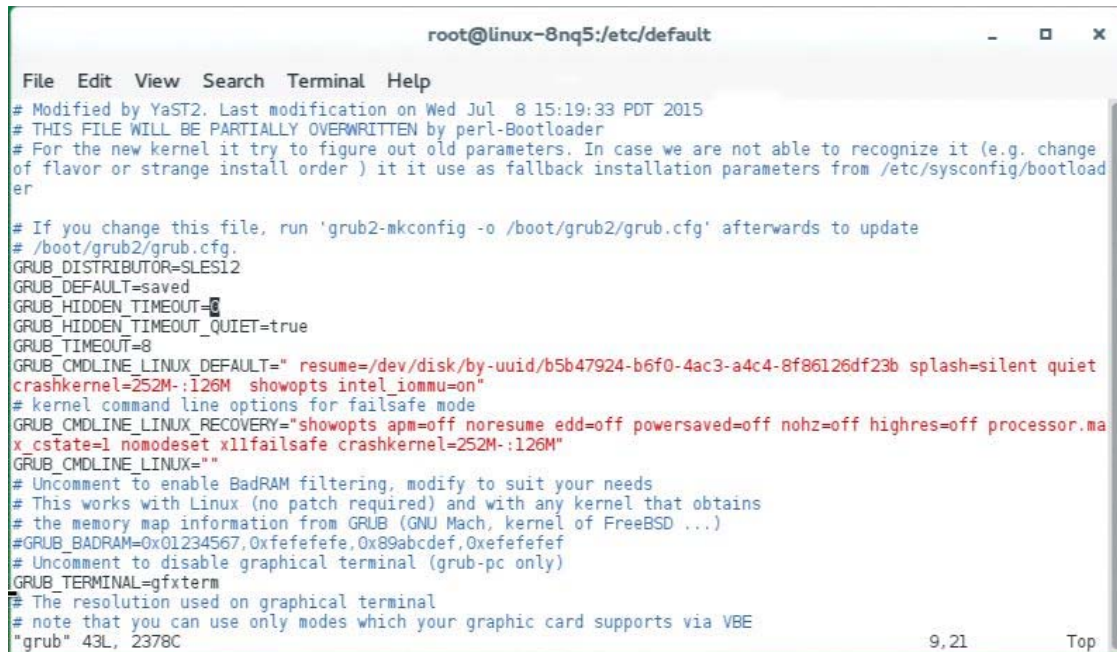
2.0 Installation and Configuration

2.1 Server Setup

This section shows various setup and configuration steps for enabling SR-IOV on Intel® Ethernet CNA X710 or XL710 server adapters.

1. Install Intel® Ethernet CNA X710 or XL710 server adapter in an available PCI-Express x8 slot. (Ensure that the x8 slot is electrically connected as x8, some slots are physically x8 but electrically support only x4. Verify this with your server manufacturer or system documentation.)

2. Power up the server.

3. Enter the server's BIOS setup and make sure the Virtualization Technology, Intel® VT-d and SR-IOV features are enabled.

4. Install SuSE Linux Enterprise Server version 12 on the server.

5. Make sure all Xen Virtualization modules, libraries, user tools, and utilities have been installed during the operation system installation.

6. The SuSE Linux Enterprise Server installation process may require a server reboot upon successful operating system install.

7. Log in to the newly-installed SuSE Linux Enterprise Server operating system using the "root" user account and password.

8. I/O Memory Management Unit (IOMMU) support is not enabled by default in SuSE Linux Enterprise Server version 12 distribution. IOMMU support is required for a VF to function properly when assigned to a VM. The following kernel boot parameter is required to enable IOMMU support for Linux kernels:

   ```
   intel_iommu=on
   ```

   This parameter can be appended to the `GRUB_CMDLINE_LINUX` entry in */etc/default/grub* configuration file, as shown in Figure 1.

Figure 1.   GRUB Configuration File

9. Update grub configuration using `grub2-mkconfig` command, as shown in Figure 2.

Figure 2.   GRUB Boot Loader Update Process Output

10. Reboot the server for the `iommu` change to take effect.

11. PF and VF drivers for the X710 and XL710 server adapters are included in the SuSE Linux Enterprise Server version 12 distribution and are named as *i40e* and *i40evf* respectively. Newer versions of these drivers are available at Intel's Open Source Linux driver site. Using latest available drivers is strongly recommended.

12. The SuSE Linux Enterprise Server version 12 installation does not create VF by default. The X710 server adapter supports up to 32 VFs per port, while the XL710 server adapter supports up to 64 VFs per port. There are two methods to create VFs depending on the Linux Kernel installed:

    a. Linux Kernel version 3.7.x and below — VFs can be created by using the *i40e* driver load time parameter called `max_vfs`.

    ```
    #modprobe i40e max_vfs=4,4
    ```

    The example in Figure 3 shows the creation of four VFs per port.

Figure 3.   i40e Driver Load Example

    b. Linux Kernel version 3.8.x and above — VF can be created by writing an appropriate value to the `sriov_numvfs` parameter via `sysfs` interface.

    ```
    #echo 4 > /sys/class/net/device name/device/sriov_numvfs
    ```

    The example in Figure 4 shows the creation of four VFs per port.

Figure 4.   VF Creation via SysFS

    The example in Figure 4 shows four VFs being created on device name `eth4`, the device name assigned to XL710 server adapter port 0 by the Linux operating system. Device name for the XL710 server adapter ports on your system may be different.

    The second command in the example above queries the `sriov_numvfs` parameter to verify the four VFs are successfully created.

13. Module options are not persistent from one boot to the next. To ensure that the desired number of VFs are created each time the server is power-cycled, append the above command to the *boot.local* file, which is located in the */etc/rc.d/* directory. The Linux OS executes the *boot.loca*l script at the end of the boot process. The example in Figure 5 shows contents of *boot.local* file.

Figure 5. *boot.local* **File Contents**

Warning: Errors and informational messages during i40e driver load are logged in the */var/log/messages* file. It is good practice to review this file to confirm that the driver loaded successfully without warnings or errors.

14. On Linux Kernel version 3.8.x and above, VF can be destroyed or disabled by writing the value 0 to the `sriov_numvfs` parameter via `sysfs` interface. Please make sure VFs are detached from the VMs before destroying the VFs.

```
#echo 0 > /sys/class/net/device name/device/sriov_numvfs
```

The example in Figure 6 shows disabling SR-IOV on a given port.

Figure 6. **Disabling/Destroying VFs Example**

15. On Linux Kernel version 3.8.x and above, the maximum number of VFs supported by the adapter can be queried by reading the `sriov_totalvfs` parameter via `sysfs` interface.

```
#cat /sys/class/net/device name/device/sriov_totalvfs
```

The example in Figure 7 shows the maximum number of VFs supported by a given port.

Figure 7.   Total VF Supported Query Output

16. Use the `lspci` command to confirm that the VF was successfully created.

```
#lspci | grep 'X710 Virtual Function'
```

The example in Figure 8 shows the result of this command.

Figure 8.   *lspci* Output

Figure 8 shows four VFs each for the physical port 0 of the XL710 server adapter. Each VF is identified by a unique bus, device, and function number. In the example, the first VF is assigned Bus #81, Device #02, and Function #0.

17. Module options are not persistent from one boot to the next. On Linux Kernel version 3.7.x and below, create the *i40e.conf* file in the */etc/modprobe.d/* folder to ensure the user-defined number of VFs are created during server boot time. Add the following content to the *i40e.conf* file to create 4 VFs per physical port.

```
options i40e max_vfs=4,4
```

Upon successful VF creation, the Linux operating system automatically loads the *i40vf* driver.

18. The VF driver automatically loads in the host operating system as soon as the VFs are created by the PF driver. The VF driver claims newly-created VFs, and these VF are not available for Virtual Machine (VM) assignment. There are two methods to overcome this scenario:

   a. Unload the VF driver from within host operating systems by executing the following command in Linux terminal with superuser (root) permission.

```
#rmmod i40evf
```

   b. Blacklist VF driver by adding `blacklist i40evf` to the */lib/modprobe.d/50-blacklist.conf* file, as shown in Figure 9.

Figure 9. *dist-blacklist* **File Contents**

This ensures that the VF driver never claims VFs in the host operating system. This method is persistent between server reboots.

19. During the creation of user-defined number of VFs, the *i40e* driver assigns MAC Address 00:00:00:00:00:00 to each VF. An application such as LibVirt or Virtual Machine Manager assigns a valid MAC Address to the VF before use. The Intel *i40e* driver has built in security feature that allows system administrators to assign a valid MAC Address to a VF from within the host operating system. Once this is done, the VM that has the VF assigned to it is not allowed to change the VF MAC Address from within the VM. Make sure each VF is assigned a unique MAC Address; duplicate MAC Addresses cause loss of communication on the network. Use the following command to set a MAC Address for each VF.

```
#ip link set eth4 vf 0 mac aa:bb:cc:dd:ee:ff
```

20. Use the following command to confirm that the VF MAC Address assignment was completed successfully.

```
#ip link show eth4
```

Figure 10 shows an example of the results of this command.

Figure 10.   VF MAC Address Query Result

21. To ensure each VF carries the same MAC Address assignment from one boot to the next, the commands from the previous step can be appended to the *boot.local* file, which is located in the */etc/rc.d/* directory. The Linux OS executes the *boot.local* script at the end of the boot process, as shown in Figure 11.


Figure 11.   *boot.local* File Contents

2.2 VM Setup

SuSE Linux Enterprise Server version 12 includes tools for creating and managing VMs. These tools offer both Command Line (CLI) and Graphical User (GUI) interfaces. Virt-Manager is a GUI tool for creating and managing VMs.

1. Use virt-manager to create a VM.

2. Install the operating system of choice on the newly-created VM. For the purpose of this document, Ubuntu 14.04 desktop Linux was installed in the VM. See the example in Figure 12:
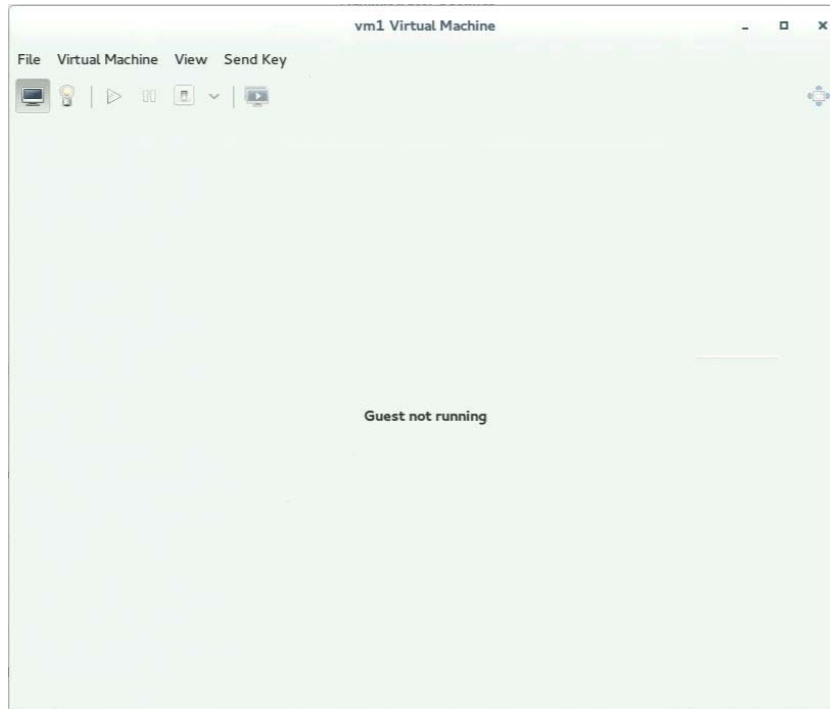
Figure 12.   Virtual Machine "vm1" Console Screen

3. Click on the [icon] icon to edit the VM properties.

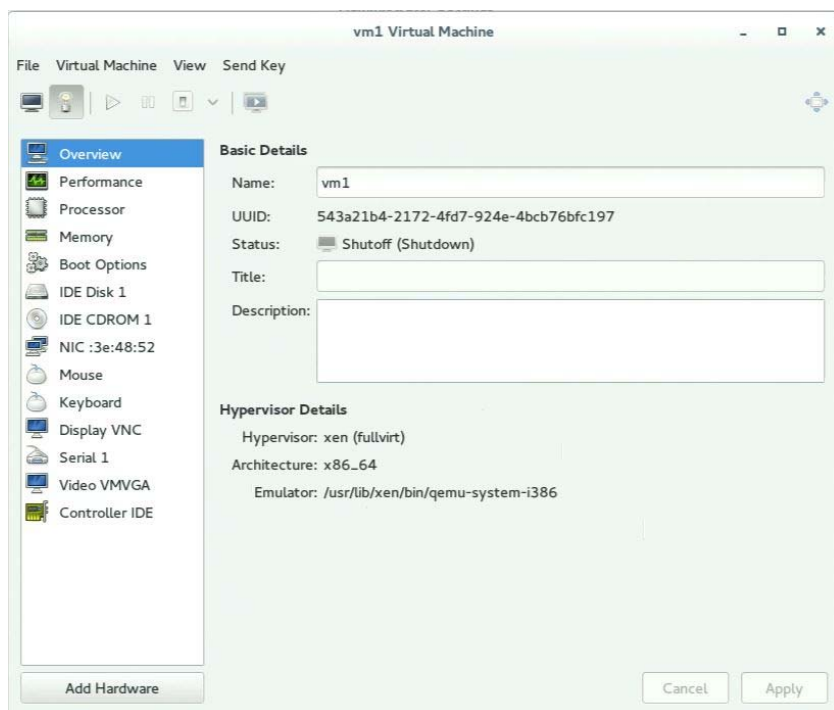4. Click on the **Add Hardware** icon to start the **Add New Virtual Hardware** wizard, as shown in Figure 13.

Figure 13.   Virtual Machine Configuration Page

5. Click **PCI Host Device** to display the **Add New Virtual Hardware** window, as shown in Figure 14.
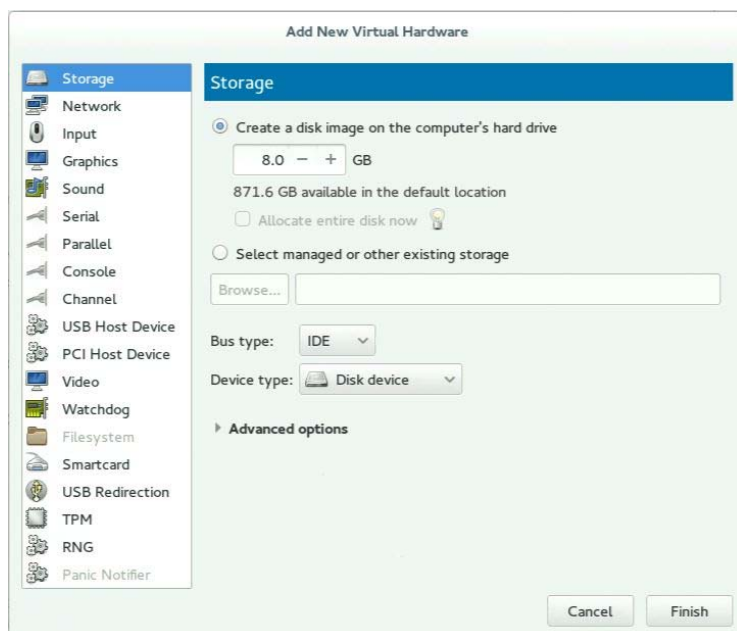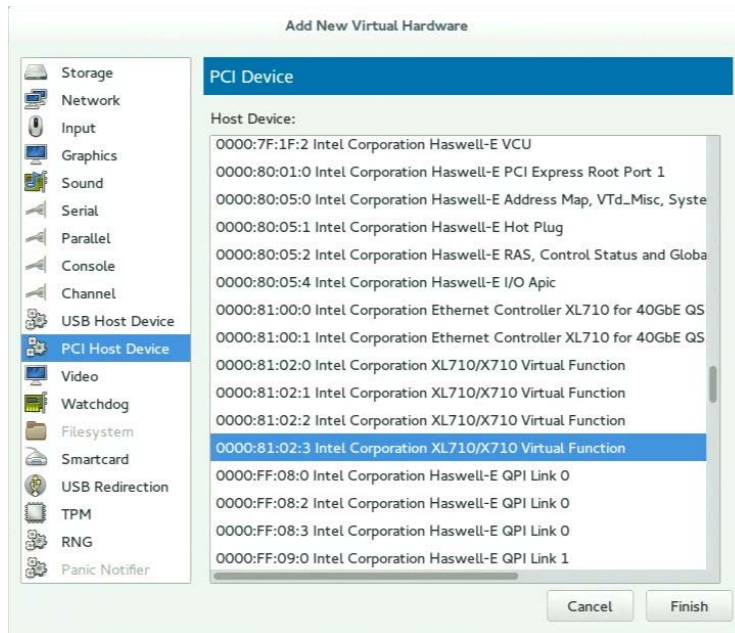

Figure 14.   Add New Virtual Hardware Page

6. Select an **XL710 X710 Virtual Function** and click **Finish**.

Figure 15.   PCI Device Selection Page

In Figure 15, the Intel Ethernet XL710 or X710 Virtual Functions are listed as "XL710 X710 Virtual Function". One or more VFs can be assigned to a VM. Upon successful assignment, the VM is ready to use.

7. Power up the Ubuntu 14.04 VM. Log into the VM using the credentials created during the VM installation process.

8. At the Linux Console, use the Linux `lspci` utility to confirm that the assigned VF is shown within the VM's PCIe hierarchy, as shown in Figure 16.
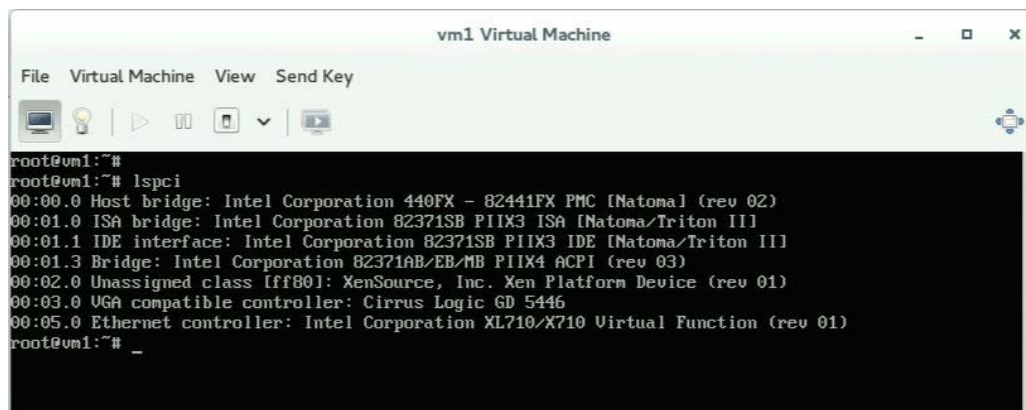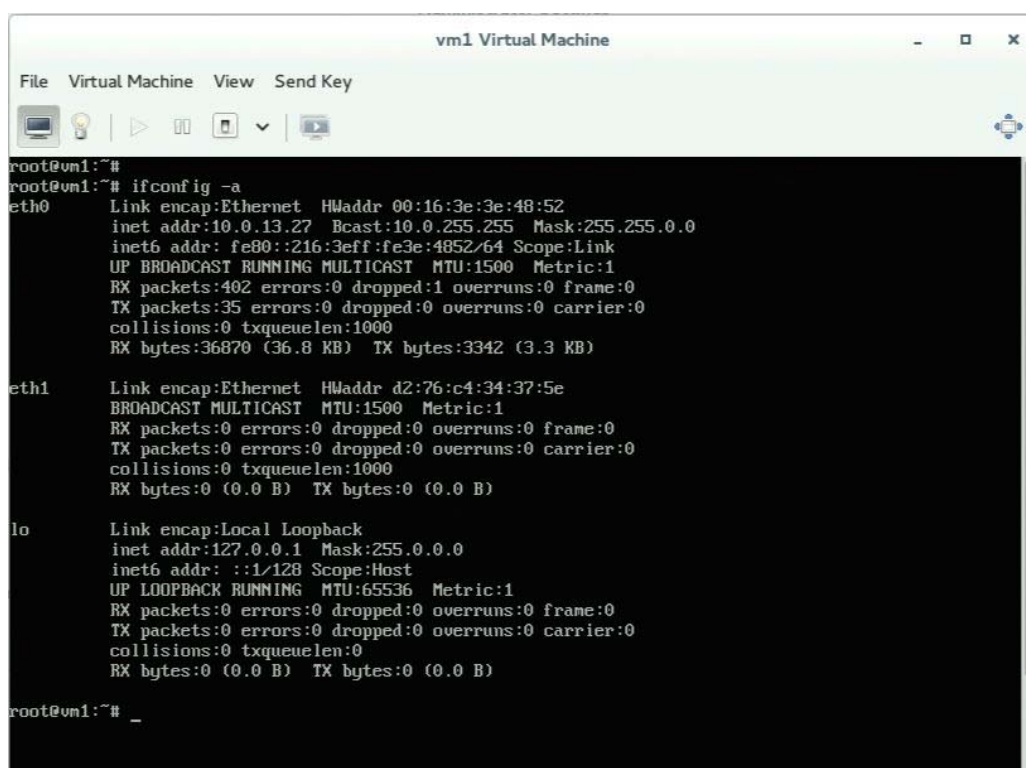

Figure 16.   "lspci" Output of the VM

9. Use the Linux `lsmod` utility to confirm that *i40evf* driver for the VF has loaded successfully, as shown in Figure 17.

Figure 17.   "lsmod" Output

10. Use the Linux `ifconfig` utility to confirm that the newly assigned VF is ready for use (Figure 18).

Figure 18.   "ifconfig" Output

11. The VF can be configured for DHCP or static IP Address assignment. The VF is ready to communicate once it has an IP Address assigned.

3.0 Summary

Intel's best-of-breed 40 GbE solutions are now available with I/O Virtualization capabilities. Customers get world-class Ethernet support along with I/O virtualization support in mainstream Linux distributions in a single adapter.

4.0 Customer Support

Intel® Customer Support Services offers a broad selection of programs, including phone support and warranty service. For more information, contact us at:

support.intel.com/support/go/network/ adapter/home.htm

Service and availability may vary by country.

5.0 Product Information

To see the full line of Intel Network Adapters for PCI Express*, visit www.intel.com/go/ethernet.

To speak to a customer service representative regarding Intel products, please call 1-800-538-3373 (U.S. and Canada) or visit support.intel.com/support/go/network/contact.htm for the telephone number in your area.
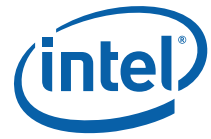
NOTE: **This page intentionally left blank.**

LEGAL

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

Intel disclaims all express and implied warranties, including without limitation, the implied warranties of merchantability, fitness for a particular purpose, and non-infringement, as well as any warranty arising from course of performance, course of dealing, or usage in trade.

This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, specifications and roadmaps.

The products and services described may contain defects or errors which may cause deviations from published specifications.

Copies of documents which have an order number and are referenced in this document may be obtained by calling 1-800-548-4725 or by visiting www.intel.com/design/literature.htm.

Intel and the Intel logo are trademarks of Intel Corporation in the U.S. and/or other countries.

* Other names and brands may be claimed as the property of others.