Characterizing 10 Gbps Network Interface Energy Consumption

Ripduman Sohan, Andrew Rice, Andrew W. Moore
Computer Laboratory, University Of Cambridge
Email: firstname.lastname@cl.cam.ac.uk

Kieran Mansley
Solarflare Communications
Email: kmansley@solarflare.com

***Abstract*—This paper quantifies the energy consumption in six 10 Gbps and four 1 Gbps interconnects at a fine-grained level, introducing two metrics for calculating the energy efficiency of a network interface from the perspective of network throughput and host CPU usage. It further compares the energy efficiency of multiport 1 Gbps to 10 Gbps interconnects.**

I. Introduction

Communication is a fundamental function of the modern server; the energy efficiency of any server is intrinsically linked to how *quickly* and *efficiently* data can be moved between it and other devices. Considering that data traffic is continually increasing both over the Internet and private networks [1], it follows that a power efficient network subsystem can result in significant runtime energy cost savings [2].

An important first step in optimizing energy consumption is quantifying its use. In this work, we set out to examine the energy efficiency of 10 Gbps (10G) server interconnects. In particular, we make the following contributions: (*i*) we measure and characterize the idle and active power consumption for a number of production 10 Gbps Network Interface Cards (NICs) of varying makes, models, architectures and utilizing different physical media; (*ii*) we compare their energy efficiency from a throughput and host CPU utilisation perspective and (*iii*) we compare the power efficiency of 10G NICs to single, dual and quad port 1 Gbps (1G) configurations.

In the remainder of this paper Section II outlines the NIC test set and the measurement infrastructure and methodology. Section III details our measurement results and analysis in the areas of idle (Section III-A) and active (Section III-B) energy efficiency while Section III-C compares the energy efficiency of 1G and 10G configurations. Section IV concludes.

II. NICs & Measurement Platform

This section briefly describes our NIC test set, and measurement platform and methodology. A more detailed description is available in the technical report accompanying this work [3].

A. NICs

Table I lists all the NICs in the test set. We measured six production 10G NICs from four manufacturers and an additional four 1G NICs for the 1G-10G comparison discussed in Section III-C. A detailed description of the chipset and physical layer properties of all devices are available in our extended technical report [3].

The 10G devices span the most common physical media types: CX4 (IEEE standard 802.3ak), short range Fibre (IEEE standard 802.3ae) and Base-T (IEEE standard 802.3an). This is of interest because there is a clear tradeoff between the cost of the NIC and the physical media: CX4 is a simple, low power copper wire standard designed to connect over short distances (up to 15 metres). The simplicity of the standard means the physical layer is cheap to implement, however, interconnections are complex and expensive to manufacture. Base-T is able to utilize existing twisted pair cabling, however signal processing overheads at 10G result in complicated physical layer designs. Finally, Fibre is relatively cheap but mandates the use of expensive transceivers.

B. Hardware and Software

All measurements were taken on a pair of SuperMicro machines consisting of an 6025W-NTR+B board based on the Intel 5400 chipset, equipped with two Xeon 5482 dual die 3.20 GHz quad core CPUs for a total of 8 logical processors. Every core has 32KB of level one data cache and every die has 6MB of shared level two cache. The system was equipped with 4GB of RAM on a quad-pumped 1600 MHz memory bus. NICs interface with the host through a (version 2) PCI-Express bus and are connected via an 8 lane connector theoretically capable of sustaining a 8GB/s transfer rate.

For the duration of the measurements the OS used was Windows Server 2008/Enterprise running in 32bit mode. Every NIC was measured using the latest drivers available (as provided on the product support website) at the time of measurement. Ethernet frame size is 1500 bytes. We used the IXIA Chariot [4] tool to generate realistic traffic streams when taking measurements that required the NIC to be active.

C. Measurement Apparatus & Methodology

We measure energy consumption by intercepting the 3.3 and 12v power lines of the PCI-Express connector to measure the current in the circuit. Figure 1 illustrates the 12v measurement apparatus in detail. PCI-Express connectors supply a (single sourced) 12v voltage on pins 2,3 of the Side A rail and 1,2 on the Side B rail of the connector. We intercept these pins and common them, feeding the resulting line through a 0.01Ω series resistor, R, before re-splitting the line to feed identical pins on a riser card into which the NIC is fitted.

Using Ohm's law the current in the circuit (I) is determined by measuring the potential difference across R using voltmeter V_1. As the current in the circuit is constant, it follows that the power being consumed in the circuit may be calculated as the product of I and the potential difference across the entire circuit as measured by voltmeter V_2.

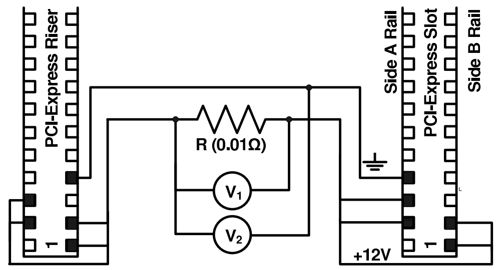

Fig. 1. 12v Power Measurement Apparatus

NIC	Link Rate (Gbps)	Physical Medium	Part Number
Solarflare(Fibre)	10	Fibre	SFE4002
Solarflare(Base-T)	10	Base-T	SFE4001
Solarflare(CX4)	10	CX4	SFE4003
Broadcom(Fibre)	10	Fibre	PE10G2T-SR
Intel(CX4)	10	CX4	PE10G2I-CX4
Intel(Base-T)	10	Base-T	PE10G1-T
Intel 1G	1	Base-T	EXPI9400PT
Broadcom Multiport(2x1G)	1	Base-T	NC380T
Intel Multiport(2x1G)	1	Base-T	EXPI9402PT
Intel Multiport(4x1G)	1	Base-T	PEG4I-RoHS

TABLE I
NIC TEST SET

NIC	Offload	Media	Idle Power (W)		
			3.3v	12v	Total
Intel(Base-T)	No	Base-T	6.0	15.2	21.2
Solarflare(Base-T)	No	Base-T	1.0	17.0	18.0
Broadcom(Fibre)	Yes	Fibre	5.9	7.2	13.1
Solarflare(Fibre)	No	Fibre	2.6	3.1	5.7
Intel(CX4)	No	CX4	5.6	0.0	5.6
Solarflare(CX4)	No	CX4	1.6	3.0	4.6

TABLE II
10 GBPS NICS - IDLE POWER CONSUMPTION

A similar setup forms the other half of the measurement apparatus by binding pins 9,10 of the Side A rail and 8,10 of the Side B rail thereby enabling the calculation of power drawn on the 3.3v circuit.

Some of our analysis required measurement of whole server power consumption. For this we used two standard off-the-shelf digital power meters with a resolution of 0.1W. The results reported in this paper are the average of at least three independent measurements. All related measurements were verified to be within 3% of one other. All results are reported to one decimal place.

III. CHARACTERIZING ENERGY CONSUMPTION

In this section we analyze idle energy consumption (Section III-B), and the runtime energy cost and power efficiency of 1G and 10G NIC deployments (Section III-C).

A. Idle Energy Consumption

Idle energy is defined as the energy consumed by the card when powered, with all links connected (and OS driver loaded) but not transferring any data. In practice it is the least amount of energy required to keep the card functional. Table II lists the idle power profiles of the 10G NICs in our test set. Our measurements lead us to make the following observations:

1) NICs may contribute significantly to server energy consumption: While NIC power consumption may seem insignificant, it is high enough that we consider it worth factoring in when designing server farms. Typical modern servers have a baseline power draw of between 150–250W depending on hardware configuration. The measured NICs, on the other hand, show a power consumption of between 5–20W. The presence of a 10G NIC increases baseline power consumption between 2.0–13.3%, a figure large enough to warrant careful consideration of which 10G interconnect should be used in the servers.

2) Physical media influences power consumption: As Table II shows, there is an order of magnitude difference in the idle power consumed by all the NICs in the test set. Various reasons may account for this difference, most significantly the internal design of the NIC and the CMOS processing technology may influence power draw.

To determine the power consumption attributable to adaptation for the physical layer we focus on the Solarflare NICs. All the Solarflare NICs measured in this work are based on an identical internal design and manufactured using the same CMOS processing technology. The only major differences in the design of the measured NICs are due to adaptation for the physical layer.

Results highlight that the CX4 model has the lowest power consumption due to the simple and straightforward wire-like design of the CX4 interconnect. This is followed by the Fibre model which consumes an additional watt due to the transceiver (as explained in Section III-A4). Finally, the Base-T model consumes the most energy due to the signal processing component of the card which is responsible for generating the pulse-amplitude-modulated waveform in the physical media.

While our physical media analysis are based on the Solarflare NICs, the results in Table II verify our claims. In general, for all cards CX4 devices consume the least energy followed by Fibre and Base-T variations respectively.

3) Offload is more power expensive: A common design optimization involves offloading network processing onto the NIC for the purposes of increased performance or reduced host CPU usage. It is commonly expected that the increased functionality and complexity of offload NICs will result in devices that have a significantly larger power footprint than more traditional designs.

While our NIC test set only includes a single offload device (Broadcom(Fibre)), our measurements confirm expectations. This device has an order of magnitude larger power draw than any other NIC adapted for CX4 or Fibre. The increased power consumption is due primarily to relatively high power usage in the 12v circuit. This is attributable to the CPU and RAM on the NIC which draw power even when the NIC is idle.

4) Link connection status has little effect on power consumption: We set out to measure the power consumption of multiport NICs (popular due to the economy and space advantages they offer) in regard to link connection status. Specifically, we quantified device power consumption in relation to the number of active links. We tested by physically removing the transceiver in the case of Fibre and disconnecting the link in the case of Base-T. We measured the power consumption for all (1G and 10G) NICs in our test set and observed that link connection state has very little impact on device idle energy consumption.

Table III illustrates the results for the Broadcom(Fibre) and Intel Multiport(4x1G) devices. As the table shows, for both the 10G Fibre and 1G Base-T devices, link connection only marginally increases power consumption (approximately 1W). For the sake of brevity we omit reporting the results of the other multiport NICs in the test set. However, we verify that we observed similar results in all cases.

Our measured results indicate that between 40–85% (Intel Multiport(2x1G) and Intel Multiport(4x1G) respectively) of

NIC	Link Speed	Media	Number Of Active Links	Idle (W) Power
Broadcom(Fibre)	10 Gbps	Fibre	0	11.1
			1	12.1
			2	13.1
Intel Multiport(4x1G)	1 Gbps	Base-T	0	7.9
			1	9.0
			2	10.1
			3	11.1
			4	12.3

TABLE III
MULTIPORT NICS - IDLE POWER CONSUMPTION

NIC	Active Power (W)	Throughput (Gbps)	CPU Usage(%)
Intel(Base-T)	21.4	11.0	369.6
Solarflare(Base-T)	18.2	15.8	508.3
Broadcom(Fibre)	14.0	18.7	264.7
Solarflare(Fibre)	5.9	15.9	508.3
Intel(CX4)	5.6	10.3	302.3
Solarflare(CX4)	4.9	16.5	484.4

TABLE IV
MEASURED 10 GBPS NICS - ACTIVE POWER, THROUGHPUT AND CPU USAGE

the overall power consumed by multiport NICs is attributable to the system electronics and remains constant regardless of the number of connected links.

B. Active Energy Efficiency

This section presents results of the active (in use) energy consumption of the 10G NICs in the test set and analyses their energy efficiency with respect to device throughput and host CPU usage.

1) Active Energy Consumption: Active power consumption is obtained by taking measurements while transferring data over 5 bidirectional TCP streams. Table IV lists the measured results for the 10G NICs in the test set along with the host CPU required to sustain maximum achievable throughput (the total amount of host CPU available in the system is 800%, defined as 8 logical processors each of which can be fully dedicated to the experiment).

There is very little difference in the power usage of an active NIC compared to an idle one. For all measured NICs the difference in power usage is less than 1W with the largest difference being only 0.9W (Broadcom(Fibre)). This means very little additional energy is required to transmit data.

Finally, the results also show that though throughput performance varies widely across the tested NICs (ranging from 11–18.7 Gbps), there is no correlation between power usage and performance – some low performing NICs have a high power draw while other higher performing NICs have a low power draw.

2) NIC Performance Per Watt: For any set of NICs able to sustain a required level of performance, the most power efficient can be defined as the one that is able to provide the most performance for the least amount of energy consumed. Using this requirement, we define the performance per watt of a NIC as the throughput in Gbps per watt of energy consumed.

We analyzed all the 10G NICs in our test for the purposes of determining NIC performance per watt. Figure 2 provides the results. As the figure indicates, the best performance is provided by the Solarflare(CX4) due to its high throughput and low power footprint. This is followed by the Solarflare(Fibre) which has near identical performance to the CX4 variation of the NIC but consumes 1W more of power in the physical layer due to the Fibre transceiver (Section III-A4).

While the Broadcom(Fibre) has the best throughput performance of all measured NICs, it fares poorly from a performance per watt perspective due to the high energy consumption of the offload engine on the NIC. Unsurprisingly, the Base-T NICs have the lowest performance in the measured set due to their high power overhead at the physical layer.

3) Server Performance Per Watt: Conventionally, all data transferred through the NIC is subject to processing in the host OS network subsystem. As has been measured previously, this processing requires substantial amounts of host CPU, especially for high speed links [5].

High host CPU usage has inspired the development of offload NIC designs which move some or all network processing onto the network card for the purpose of reducing host CPU utilization. However, Section III-A3 has also highlighted that offload designs have higher energy consumption.

There is clearly a tradeoff between the throughput performance of the NIC, the amount of power it consumes and the amount of host CPU used to service the network interconnect. An ideal NIC will provide high throughput, use little power and consume a minimum amount of host CPU.

Given a set of NICs that can be serviced within a maximum threshold of host CPU dedicated to network processing, an administrator will likely select the one able to provide the best performance for the least power *and* host CPU consumption. However, correlating NIC power consumption, throughput and host CPU consumption is non-trivial; all three parameters are independent variables as listed in Table IV.

We introduce *server* performance per watt as a simple metric that enables reasoning about NIC host CPU consumption. Server performance per watt is defined as the throughput obtained per watt of *server* energy consumed. It is based on our observations that: (*i*) an idle powered server has a constant power draw and (*ii*) server power consumption increases in proportion to CPU load[1]. In effect, this metric incorporates the utilisation of host CPU for servicing the network. If the NIC requires a large amount of host CPU server power consumption increases and server performance per watt reduces.

We analyzed all the 10G NICs in our test set to determine server performance per watt. Figure 3 presents the results. The Broadcom(Fibre) NIC is the most efficient from a whole-server perspective due to its low CPU load in relation to the extra power consumed by its offload engine. This is followed by the Solarflare NICs which have a better server performance per watt result than the Intel NICs in spite of consuming more CPU. This is due to their higher throughput characteristics. Finally the Intel NICs have the lowest performance overall due to their low throughput. Server performance per watt results are also interesting as they show that, from a system perspective, overall server energy efficiency is still dominated by host CPU utilisation.

C. Multiport 1G vs 10G

In this section we compare a number of single and multiport (dual and quad) 1G configurations with the 10G NICs in our test set in order to determine those that provide the best performance-to-power ratios. We focus on NICs adapted for

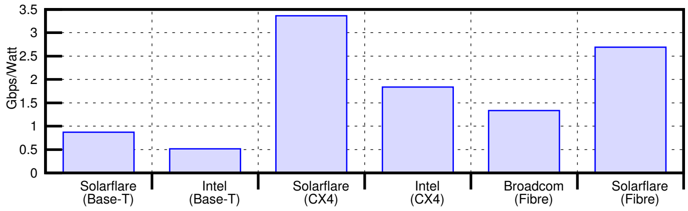

Fig. 2. 10 Gbps NICs – NIC Performance Per Watt

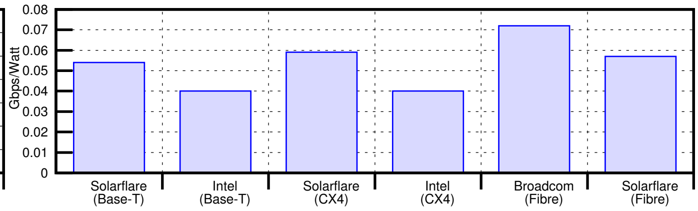

Fig. 3. 10 Gbps NICS – Server Performance Per Watt

NIC	Media	Throughput (Gbps) Theoretical	Actual	Active Power (W)
Intel 1G	Base-T	2	1.7	1.9
Broadcom Multiport(2x1G)	Base-T	4	3.3	7.0
Intel Multiport(2x1G)	Base-T	4	3.3	3.6
Intel Multiport(4x1G)	Base-T	8	5.7	12.5

TABLE V
1G NICs - Performance And Power Characteristics

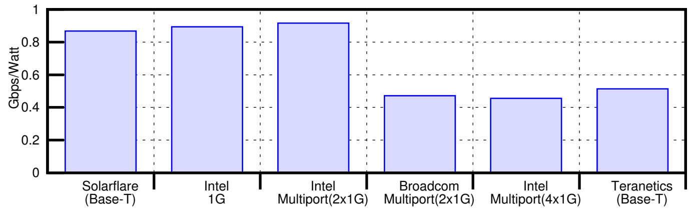

Fig. 4. 1 vs 10 Gbps - Performance-Per-Watt

the Base-T physical layer as this is the most prevalent wiring infrastructure in modern datacenters. Table V presents the characteristics of our measured 1G NICs. The results lead to the following observations:

1) Throughput efficiency decreases as the number of ports increase: Our measurements show that throughput does not scale in relation to the number of ports. While it is unlikely that any NIC will achieve its theoretical throughput (due to host and protocol overheads), we found that the single port NIC is able to achieve 85% of theoretical bandwidth, dual ports are able to achieve 82.5% of theoretical bandwidth but quad port devices are only able to achieve 70% of theoretical bandwidth. In comparison our 10G NICs are able to achieve up to 93.5% (Broadcom(Fibre)) of theoretical bandwidth.

2) Power consumption increases in correlation to the number of ports: As Table V illustrates, the power footprint of the multiport NICs increases in relation to the number of ports on the device. Focusing on the Intel single and dual port NICs (chosen as devices from the same manufacturer are likely to contain common design elements and be implemented using similar technology), we notice that the the average active power[2] consumed per port remains approximately the same (1.8–1.9W) for the single and dual port variations. Furthermore, power consumption actually increases to 3.125W for the quad port NIC. However, this increase is likely to be due to the fact that the quad port NIC is manufactured by Silicom and thus uses a different physical layer implementation to the single and dual port NICs

While confirmation would require detailed instrumentation and measurement, power consumption measurements suggest that there is little electronic integration on the device. NIC datasheets list a single controller but physical power draw seems to suggest a duplication of functionality (and associated electronics) in a single packaging. In some cases (e.g. Broadcom Multiport(2x1G)) the multiport NIC is actually composed of multiple 1G NICs coupled on the same printed circuit board. From a technical perspective, the only advantage of using 1G multiport devices in comparison to single port NICs is the PCI-Express slot savings efficiency.

3) 1G NICs Possess Similar NIC Performance Per Watt Characteristics as 10G NICs: Next, we evaluated the efficiency of the multiport 1G NICs by calculating their efficiency in terms of NIC performance per watt. Figure 4 provides the results of this analysis. As illustrated, the relatively low power consumption and high throughput achieved by the Intel Multiport(2x1G) NIC ensures it has the highest performance per watt of the measured set. This is followed closely by the Intel 1G and then the Solarflare(Base-T). The Broadcom Multiport(2x1G) and Intel Multiport(2x1G) both have much lower performance per watt due to their low throughput and high power draw.

While the NIC performance per watt metric provides a simple, efficient and abstract mechanism for comparing the power efficiency of different NICs it is important to note that practical factors such as PCI-Express slot availability and high host CPU requirements may limit NIC configurations.

IV. Conclusions

This paper measured and analysed the power consumption of six 10 Gbps and four multiport 1 Gbps NICs spanning a range of design, manufacturer and physical media types. Our results found that, generally, 10 Gbps NICs consume between 4.5–20W of power depending on design and physical transmission media while 1 Gbps NICs consume between 2–13W. Furthermore, there is very little difference in the power consumption of an idle or loaded NIC. Higher link speeds have high host CPU requirements (between 250–500% CPU). Finally, the work determined that the current generation of 10 Gbps NICs are able to match mature 1 Gbps NICs in performance per watt energy efficiency.

References

[1] A. Odlyzko, "Internet traffic growth: sources and implications," in *Proceedings of SPIE (Optical Transmission Systems and Equipment for WDM Networking II)*, vol. 5247, pp. 1–15, Oct. 2003.

[2] K. Christensen, B. Nordman, and R. Brown, "Power management in networked devices," *IEEE Computer*, vol. 37, pp. 91–93, Aug 2004.

[3] R. Sohan, A. Rice, A. W. Moore, and K. Mansley, "Characterizing 10 Gbps Network Interface Energy Consumption," Tech. Rep. TR-784, University of Cambridge, Computer Laboratory, July 2010. http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-784.html.

[4] "IxChariot - Test Device and System Performance under Real-World Load Conditions." http://www.ixchariot.com/. Retrieved 1st April 2010.

[5] A. Gallatin, J. Chase, and K. Yocum, "Trapeze/IP: TCP/IP at near-gigabit speeds," in *Proceedings of the 1999 USENIX Annual Technical Conference, FREENIX Track*, pp. 109–120, USENIX Association, Jun 1999.

Notes

[1] We assume little or no disk activity as the workload is CPU bound

[2] Similar to the 10G NICs, active power is only marginally larger than idle power for the 1G NICs