

How fast is Gigabit Over Copper?

Prepared by Anthony Betz and Paul Gray

April 2, 2002

Given the relatively low cost, backwards-compatibility, and widely-availability solutions for gigabit over copper network interfaces, the migration to commodity gigabit networks has begun. Copper-based gigabit solutions are now providing an alternative to the often more expensive fiber-based network solutions that are typically integrated in high performance environments such as today's tightly-coupled cluster systems. But how do these cards compare with their fiber based counterparts? Are the Linux-based drivers ready for prime-time? The intent of this paper is to provide an extensive comparison of the various Gigabit over copper network interface cards available. Since performance is based on numerous factors such as bus architecture and the network protocol being used, these are the two main subjects of our investigation.

Initially the goal was to compare the performance of DMA access to the network interface via MVIA. Upon reviewing our preliminary data it was observed that DMA access was actually slower than the traditional communication mechanisms. Given this turn of events our focus changed to benchmarking the various available gigabit over copper cards. Throughput and reliability emerged as the two most important factors to investigate since they play key roles in clustering. Given that high throughput isn't always indicative of high performance reliability and predictability were also concerns. While it may be nice to send static sized packets back and forth at high speeds there is always the question of if I increase the packet size will I continue to get the same throughput or will the network choke. Both questions are valid concerns and at times require trade-offs between the two.

In order to collect sufficient information to gauge the performance of the cards we needed a program that benchmarked specifically what we were interested in. Ames Labs has a nice utility that measures the performance of various protocols called Netpipe. Netpipe is rather simple program that combines tcp and netperf to give an applications view of end-to-end behavior. Netpipe allows testing of various protocols such as TCP, MPI and PVM. Basically all that goes on with Netpipe is packets of increasing size are sent from a transmitter to a receiver during which the time for the transmission is recorded. In our case we chose to look at TCP since it is the most widely used protocol by choice or not. Our bandwidth benchmarks look at sustained throughput using TCP. While other communication protocols are available, indeed preferred, for high-performance computing, TCP-based benchmarks provide an immediate insight into the expected performance of the cards.

The method used in our testing was an out of the box approach. Simply put, how will a given card perform when just taken from its packaging and installed without any tweaking. Regrettably many users are under the naïve assumption that aside from installing the card nothing else needs to be done. With this in mind we set out to carry out our tests from this standpoint. Hardcore computer users know that there is very little in a computer that can not be tweak and gigabit cards are no assumption. While tweaking of individual cards may provide enhanced performance it is far too time consuming and nearly a project in itself. That is why the only modifications to card parameters that were examined were bus speed, communication sub layer and protocol.

Changing bus speed is probably the first intuitive modification that most users think about and is just about the most basic. In many cases increasing the bus speed will give an increase in performance albeit it small in some circumstances. In our case the change would be from a 33 MHz to 66 MHz bus. This modification is any easy change that didn't require any changes on the card itself, only the placement on the motherboard. Given that a number of the tested cards support the new PCI-X standard this would be a worthwhile test to see how well the performance scales at a higher clock speed. In computer hardware doubling the clock speed rarely doubles the performance of the device being changed. Often one aspect of performance increase while another decreases. In the case of throughput it generally increases but the reliability and predictability of the card decreases.

Modifying the communication sub layer isn't as drastic as it may sound. This modification is similar to changing the bus speed, but instead of the bus clock being doubled the number of data lines or data bus is doubled. The simplified idea behind this is that on a given bus of 32-bit width only 32 bits can be transferred at one time. If the width is doubled to 64 bits a total of 64 bits can be transferred in a single clock cycle. In our case we didn't actually add any pins or make changes to the motherboard. Instead we placed a 32-bit card into a 64-bit slot or a 64-bit slot into a 32-bit slot. In cases where the 32-bit architecture card was used in a 64-bit slot a noticeable increase in throughput was observed. Instances where a 64-bit architecture card was tested in a 32-bit bus throughput was always reduced considerably.

The final modification made involved the actual communication protocol. Changing the protocol is not as intuitive as the previous two methods since the protocol is often viewed as static. The TCP protocol allows users to either increase or decrease the amount of data within the packet. This accomplished by changing the MTU (maximum transmission unit) value (fig 1). Under IPv6 there is a minimum MTU value of 1280 bytes. Given the lower bound and that 1500 bytes is the default under Linux, the range tested values were 1500, 3000, 4000, 6000 and 9000 bytes. All cards with the exception of those based off of the ns8320 and ns 83821 chipsets were tested at 9000. Those based off the ns8382x chipset had an upper limit of 6000 due to hardware limitations. Expectations on throughput were based from the assumption that increasing the packet size will also increase throughput. In most cases this was true up to a point. Most cards reached their peak throughput at 6000 bytes. While changing the MTU performance with respect to throughput may increase but it is at the expense of reliability and predictability (fig 2, fig 3). Another problem associated with increasing the MTU size is that it can lead to severe fragmentation problems. Placing a limit on packet sizes isn't really the problem. The problem is when a large piece of data needs to be sent and it exceeds this limit. Imagine that you need to send a message that is 1600 bytes in size. By having the MTU set to 1500 bytes limits all packets to a size of 1500 bytes. In order to send your message you need to use two separate packets now as oppose to one. When the receiver receives one of these packets it sees that there is a flag set in the packet header indicating that this is just part of the message. The receiver no needs to wait for the other packet to arrive. Packets don't always arrive in order so it can't be reassembled on the fly. Now the receiver receives the last packet and sees that it has received the last packet in the series and now needs to reassemble all of the packets to form the original 1600 bytes message. While this is a streamlined and shortened example it shows the overhead

involved when large packets are sent. Also fragmentation rears its ugly head when switches and routers are brought into the mix. The majority of switches on the market don't support jumbo frames and thus become a bottleneck. In order to transmit large packets these switches do one of two things. Either they break up the packet into their own maximum size based off their MTU or they refuse to relay the packet and thus force the communicating parties to lower their MTUs.

With PCI-X coming into the marketplace in more and more motherboards as well as the multitude of systems with more traditional 32-bit PCI subsystems, numerous cards are available for today's 64bit and 32bit computer systems. The 64bit cards tested were as follows: Syskonnect SK9821, Syskonnect SK9D21, Asante Giganix, Ark Soho-GA2000T, 3Com 3c996BT and Intel's E1000 XT. The 32bit cards were Ark Soho-GA2500T, D-Link DGE500T.

The test environment consisted of two testbeds. The first testbed consisted of two server-class Athlon systems with a 266MHz FSB. Specifications are as follows: Tyan S2466N Motherboard, AMD 1500MP, 2x64-bit 66/33MHz jumper-able PCI slots, 4x32-bit PCI slots, 512MB DDR Ram, 2.4.17 Kernel, RedHat 7.2. The second testbed consisted of typical desktop/workstation Pentium-based systems. Specifications are: Pentium III 500 Mhz, 128MB Ram, 5x32-bit PCI slots, 3x16-bit ISA slots.

D-Link DGE-500T was the first of the gigabit cards tested. This card is based on National Semiconductor's dp83820 chipset and is designed for a 32-bit bus. The chipset in this card turned out performance nearly identical to the two Ark cards and the Giganix cards tested in our test suite, since all utilize the dp83820 chipset from National Semiconductor. Notable hardware features are 33/66 Mhz local bus master, 8KB transmit and 32KB receive data buffers. The Linux driver used was the ns83820 as included in the 2.4.17 kernel. Latency on both platforms was .0002 seconds.

Peak throughput while operated in a 32bit bus was 192.21 Mbps. This was achieved in the Dell systems. The Athlon systems only obtained a peak of 172.21 Mbps when these cards were inserted into the 32-bit bus. Both systems show a slight drop in throughput but eventually level out (fig 4). Peak throughput while operated in a 64bit bus running at 33Mhz with an MTU of 6000 was 607.19 Mbps. When the bus was jumpered to auto select 66/33Mhz, the performance increase was negligible. Peak throughput was 606.88 Mbps. Comparing the plots of the 66Mhz and 33Mhz run reveals that they are essentially identical (fig 5).

Price: \$45

The cost per Mbps is as follows:

32-bit 33Mhz: $\$45 / ((192.21 + 172.21) / 2) = \$.25$

64-bit 33Mhz: $\$45 / 315.96 = \$.14$

64-bit 66Mhz: $\$45 / 316.40 = \$.14$

The Ark Soho-GA2500T is also a 32-bit PCI card design. Like the D-Link DGE-500T and the Asante Giganix cards, this card is based on the National Semiconductor dp83820 chipset. Hardware is identical to the DGE-500T. With that in mind the performance was estimated to be close to the D-Link DGE500T. The driver used was the generic ns83820 included in the 2.4.17 kernel. The latency for both test systems was .0002 seconds.

Peak throughput achieved while in a 32-bit 33Mhz bus was in the Dell system:

192.62 Mbps. While the Athlon system in the same bus setup only reached 172.19 Mbps. As before, there is a performance drop at the 1Kb and 5-10Kb packet sizes. Peak throughput while operated in a 64-bit bus running at 33Mhz was 610.83 Mbps and 609.98 Mbps when running at 66Mhz respectively(fig 6). As with the Soho-GA2000T, there is no noticeable difference between a 33Mhz and a 66Mhz bus(fig 7).

Price: \$44

The cost per Mbps is as follows:

32-bit 33Mhz: $\$44 / ((192.62+172.19) / 2) = \$.24$

64-bit 33Mhz: $\$44 / 610.83 = \$.07$

64-bit 66Mhz: $\$44 / 609.98 = \$.07$

Our transition into cards designed for a 64-bit PCI bus began with the Ark Soho-GA2000T. Like its 32-bit counterpart, this card was designed around the ns83820 chipset, which will allow us to examine the performance benefits, if any, in moving from a 32-bit slot to a 64-bit slot.

Designed to run in a 64-bit 66Mhz slot, this card is backwards compatible to 32-bit and 33Mhz slots. This card is based off of National Semiconductor's dp83820 chipset so performance was expected to be similar to the DGE500T and the Soho-GA2500T. The driver used was the generic ns83820 included in the 2.4.17 kernel. Latency was .0002 seconds on both test platforms.

Peak throughput for a 32-bit 33Mhz slot was 189.93 Mbps in the Dell system. The Talons were only able to reach 172.26 Mbps. Peak throughput for 64-bit 33Mhz was 665.06 Mbps with an MTU of 6000. Peak throughput while running at 66Mhz was 640.60 Mbps(fig 8). With the exception of the 6000MTU tests, there is no noticeable difference between bus speeds of 33 and 66Mhz(fig 9).

Price: \$69

The cost per Mbps is as follows:

32-bit 33Mhz: $\$69 / ((172.26+189.93)/2) = \$.38$

64-bit 33Mhz: $\$69 / 665.06 = \$.10$

64-bit 66Mhz: $\$69 / 640.60 = \$.11$

The second 64bit card tested was Asante's Giganix. This card is designed for a 64-bit bus but, is backwards compatible to 32bit and 33Mhz configurations. Giganix is based off of the dp83821 chipset. The driver supplied by Asante was unable to compile due a bug in the code. In order to get the card to work the generic ns83820 driver was used again. Performance was expected to be similar to the GA2000T. Latency was .0002 seconds on both systems.

Peak throughput for a 32-bit 33Mhz configuration was 238.75 Mbps in the Dell systems, with a peak of 172.19 in the Athlons. When comparing to the GA2000T, the Athlon results stay about the same whereas the Dell systems increase by 50Mbps. Peak throughput for 64-bit 33Mhz was 641.02 Mbps with an MTU of 6000. When running at 66Mhz, the peak is 651.51 Mbps with the MTU at 6000(fig 10).

An interesting spike in throughput on the 64-bit 66Mhz tests was when the MTU was set to 3000. Aside from the 40Mbps difference between the two bus speeds, the plots look very similar. The main difference is the spike at 8KB packets(fig 11).

Price: \$138

The cost per Mbps is as follows:

32bit 33Mhz: $\$138 / ((238.75+172.19) / 2) = \0.67

64bit 33Mhz: $\$138 / 641.02 = \0.22

64bit 66Mhz: $\$138 / 651.51 = \0.21

The first of the Syskonnect cards tested was the SK9821. This card is designed for a 64-bit bus. The SK9821's are backwards compatible to 32-bit and 33Mhz configurations. The driver used was sk98lin from the kernel source. Notable hardware : 512 byte VPD memory and 1MB SRAM. Latency was .000048 on the Dells and .000025 seconds on the Athlons. Of all the 64bit cards tested, the SK9821 is the first to have a noticeable difference in performance between the two bus speeds.

Of all cards tested, the Syskonnect SK9821 gave the most consistent throughput over all packet sizes, and was far-and-away the overall performance leader. In the server-class testing environment, peak throughput in our 64-bit 33Mhz setup was 782.27Mbps with the MTU set to 9000. The peak for 66Mhz tops off at roughly 940Mbps with jumbo frame MTU sizes of 6000 and 9000(fig 12,13). Peak throughput on 32-bit 33Mhz was 365.27 Mbps on the Dells. After the peak, is reached there is a noticeable drop in throughput as it levels off to the 330Mbps range.

Price: \$570

The cost per Mbps is as follows:

32-bit 33Mhz: $\$570 / ((365.27+163.97) / 2) = \2.15

64-bit 33Mhz: $\$570 / 782.27 = \0.73

64-bit 66Mhz: $\$570 / 938.97 = \0.61

The second card tested from Syskonnect was the SK9D21. The SK9D21 is aimed at the desktop/workstation market. While support for this card under Windows environments appears to be solid, there were too many technical issues. The testing environment's mix of kernel, motherboard, Athlon chipset, and Syskonnect drivers made for too many components to successfully debug the problems with this card thoroughly. This card is designed for a 64-bit bus the card is backwards compatible with 32-bit and 33Mhz configurations. While an exhaustive analysis of the cards was unavailable, it should be noted that the latency was successfully determined at .000123 seconds.

Our difficulties with this card were limited to the 64-bit bus. Our tests were successful in analyzing the performance in both the Athlon-based systems and the Pentium-based systems in 32-bit busses.

When driver issues for this card are resolved, performance evaluations in this section will be amended.

Peak throughput in the Dell system was 377.53 Mbps. As with the SK9821, there is a drop off after the peak is reached.

Price: \$228

The cost per Mbps is as follows:

32-bit 33Mhz: $\$228 / 377.53 = \0.60

The next card in the test suite was the 3Com's 3c996BT. This card is designed as a 64-bit 133Mhz card, but is backwards compatible to 32-bit, 33 and 66Mhz

configurations. The driver used was the bcm5700, version 2.0.28, as supplied by 3Com. Latency was .000103 in the Dells and .000078 in the Athlons.

The peak throughput achieved in this card while in a 32-bit 33Mhz slot was 436.23 Mbps in the Dell systems. In the Athlon system, the same bus configuration only reached 184.02 Mbps. Peak throughput while running in a 64-bit 33Mhz slot was 884.09 Mbps this was with an MTU of 4000. While running at 66Mhz, the peak was only 546.16 Mbps with an MTU of 6000. These plots are all relatively smooth when compared to the other plots for this card(fig 14). Performance in a 66Mhz slot is actually lower for all MTU sizes as compared to a 33Mhz slot(fig 15).

Price: \$138

The cost per Mbps is as follows:

32-bit 33Mhz: $\$138 / ((436.23+184.02) / 2) = \$.44$

64-bit 33Mhz: $\$138 / (884.09) = \$.16$

64-bit 66Mhz: $\$138 / (546.16) = \$.25$

The final 64bit card tested was Intel's E1000 XT. As with the 3c996BT this card is designed for future PCI-X bus speeds running at 133Mhz. It is compatible with a variety of configurations running at 33 and 66Mhz as well as 32-bit. The card uses Intel's e1000 module, version 4.1.7. Latency in the Athlon systems was .000091 seconds. Due to time constraints, we have yet to test this card in the Dell testbed.

Peak throughput achieved was 743.14 Mbps while running in a 64-bit 66Mhz slot with the MTU set to 9000(fig 16). Performance in a 32-bit configuration turned out the lowest throughput for all cards tested coupled with the most erratic throughput. During the throughput tests, the card would drop 100% of packets for extended lengths of time. Initial testing in the 64-bit setup showed performance similar to the Giganix card with regards to a 64-bit bus. Once the MTU was set to 9000 performance became very erratic, stagnated several times, then stabilized once the packet size reached an upper threshold peak. Note that the drop in performance was not associated with the (expected) phenomena of packet reassembly when the TCP packet size exceeds the MTU.

As testing continued to the 66Mhz phase things only got worse. Once the MTU exceeded 3000, performance was no longer predictable. During the 4000 MTU tests, the throughput plummeted to around .4 Mbps for several TCP packet sizes(fig 16). At an MTU of 6000 and at 9000 the same problem occurred as before in the 64-bit 33Mhz test.

Price: \$169

The cost per Mbps is as follows:

32-bit 33Mhz: $\$169 / 142.02 = \1.18

64-bit 33Mhz: $\$169 / 624.41 = \$.27$

64-bit 66Mhz: $\$169 / 743.14 = \$.22$

Of the eight cards tested, the clear performance champion was the SK9821 with regard to throughput and consistency. The 3Com 3c996BT has a modest price tag and respectable performance for the entry-level server configuration. If price per megabit is the main concern, the Ark Soho-GA-2500T has the lowest cost per Mbps, making it a viable solution for entry-level systems requiring higher throughput than fast ethernet.

The D-Link DGE500T and the Soho-GA2500T show nearly identical peaks, which is to be expected since the drivers and the chipsets were the same. The 3Com 3C996BT has results when compared to the 64-bit 33MHz results were surprising

inasmuch as these cards showed better performance at 33MHz bus than at the higher 66MHz bus. Of all of the cards tested, the Intel E1000 TX proved to be comparable to the comparable to the Asante Giganix card in peak performance, but the erratic overall performance proved too much to overcome. Some general comparisons that can be derived from the above results include the notion of "cost per peak megabit." Depending upon the environment that the network device is to be installed, the cost per peak megabit varies greatly. For example, if one would wish to upgrade their P-III-based desktop system with a 32-bit, 33MHz PCI, the GA25000T is the clear cost-effective solution, but would not be able to provide throughput at the level of the 3Com 3C996BT. In an HPC environment, where sustained throughput is critical and the switch is capable of Jumbo frames, the SK9821 would be the best performer. In light of gigabit switching hardware that lacks Jumbo Frame support, a comparison of the 1500MTU results shows the SK9821 is still a viable choice, as is the 3Com 3C996BT which provides a more cost-effective solution..

Throughout the testing numerous anomalies cropped up. Probably the most interesting and consistent was the higher throughput in the Dell systems as opposed to the Athlon systems. Throughout all cards tested the Dell's out performed the Athlons(fig 18). The spread in performance ranged from only a few Mbps as in the DGE-500T to the drastic as with the 3c996BT. Further testing is definitely needed to address this phenomena. A few ideas have been tossed around as to the cause for this. Two main ideas are maturity and or drivers. Given that the MPX chipset is relatively new there may exist bottlenecks in the connection between the Northbridge and Southbridge chipsets. The other may simply be drivers with regards to Linux for this given chipset. In order to determine the actual cause for this performance gap several avenues exists. The first would be to drop back to the older Tyan MP chipset based motherboard. Another would be to test an altogether different motherboard manufactures implementation of the MPX chipset. This method would help to narrow down the problem to the specific motherboard and or manufacture. Also testing a uniprocessor Athlon board could help determine if it is related to SMP configurations. Testing also should be done on the other side of processor field. Uniprocessor and SMP Pentium based systems should be tested to find out if the performance edge is only evident in on that specific motherboard used by Dell. Another avenue is to test if it is operating system related. A comparison with respect to at least the 32-bit 33Mhz configuration should be made. If the performance gap is gone it leads to the likelihood of driver shortcomings under Linux. In all fairness a comparison to Windows or other operating systems should be made.

The second phenomena observed dealt with the 3Com 3c996BT cards. Operating within a 64-bit 33Mhz configuration performance is excellent with regards to throughput and reliability. Once the bus speed is bumped up to 66Mhz the throughput drops considerably(fig 15). Up to this increase in bus speed the 3c996BT had the highest throughput of all cards tested. Given that the card is designed to take advantage of PCI-X systems the drop was unexpected. The only reason we have been able to come up with is that the card has been tweaked to have its best performance under the most widely available configuration. In this case 64-bit 33Mhz has been the most common 64-bit configuration for server class systems. Given more time it is likely that the card could be re-tweak to perform in the newer 64-bit 66Mhz configuration.

524291 708.4019
 786429 712.2845
 786432 379.0397

786435 272.1815
 1048573 725.1468

Table 2

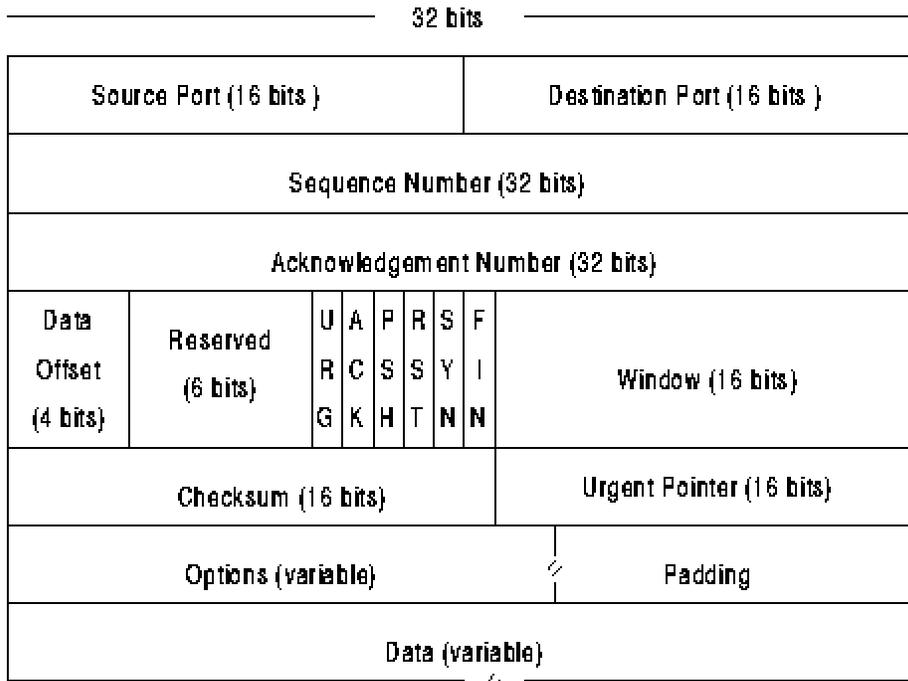


Figure 1

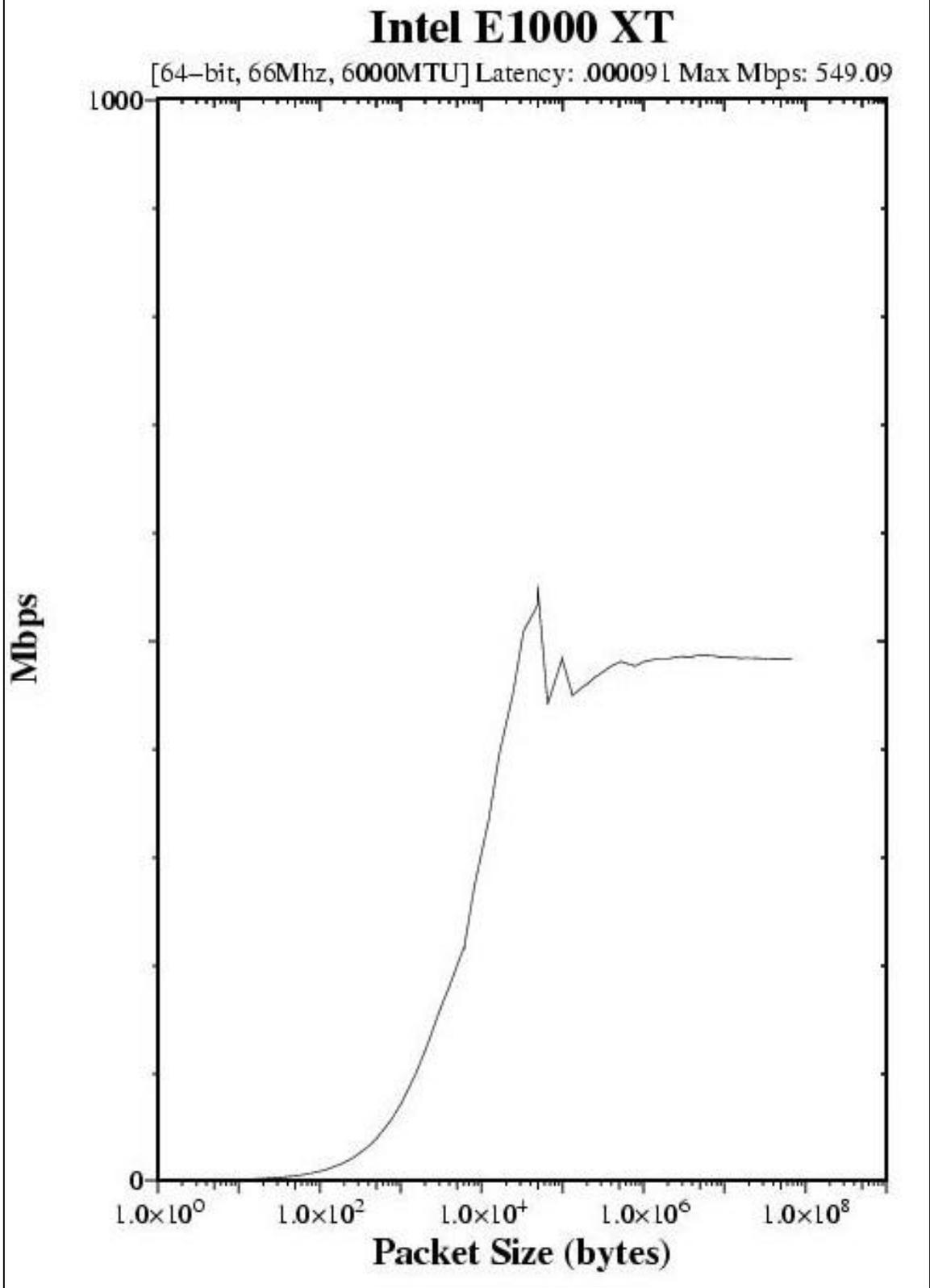


Figure 2

Intel E1000 XT

[64-bit, 66Mhz, 9000MTU] Latency: .000091 Max Mbps: 743.14

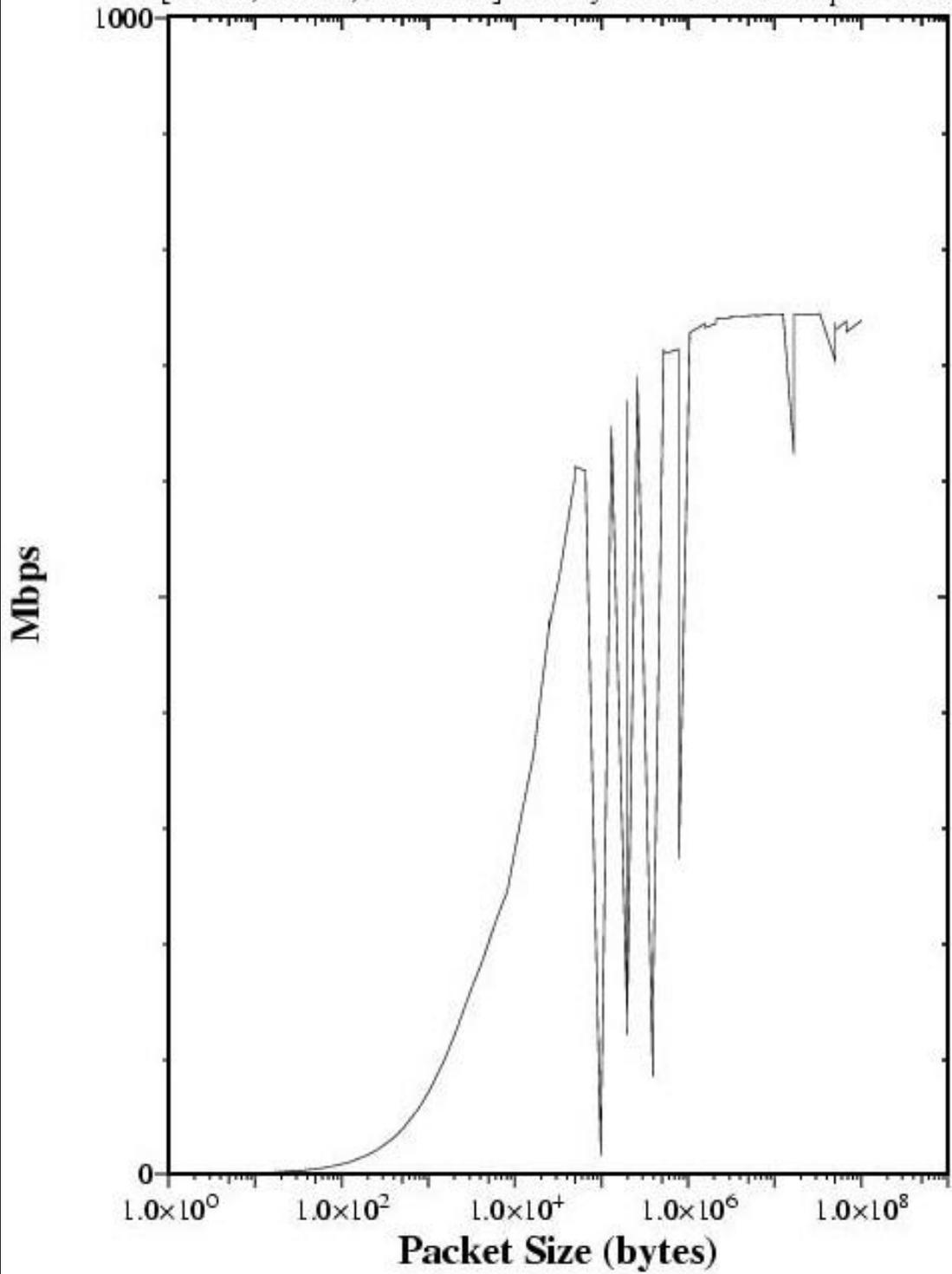


Figure 3

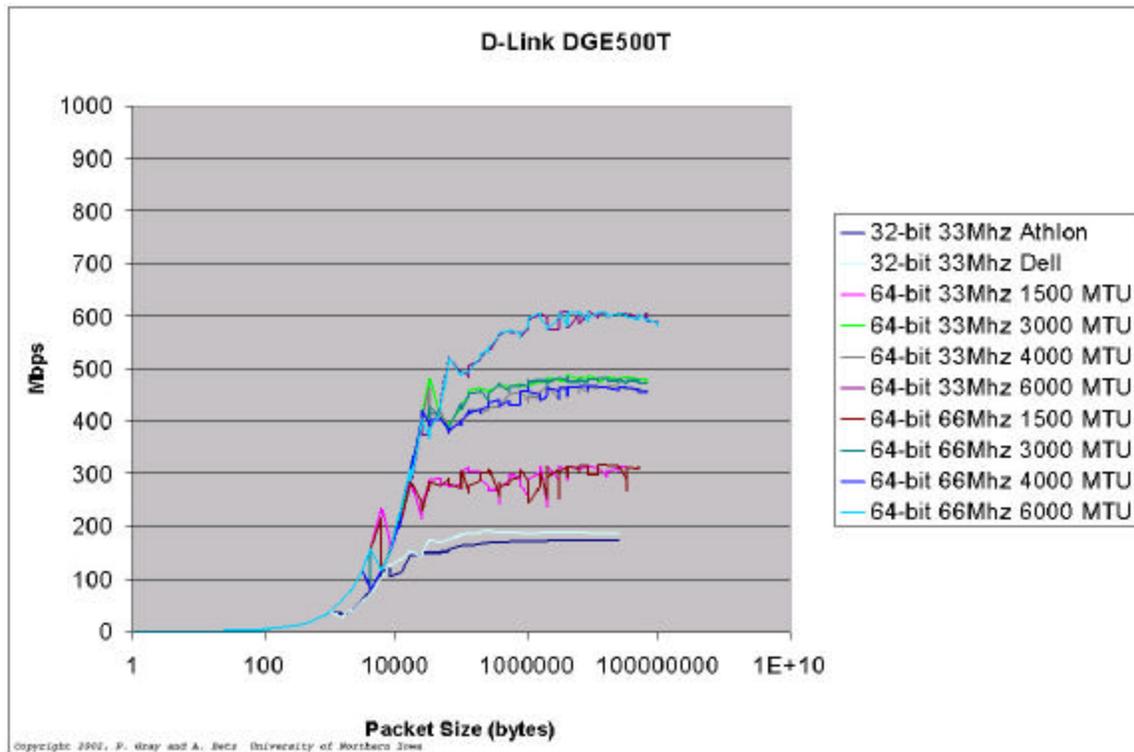


Figure 4

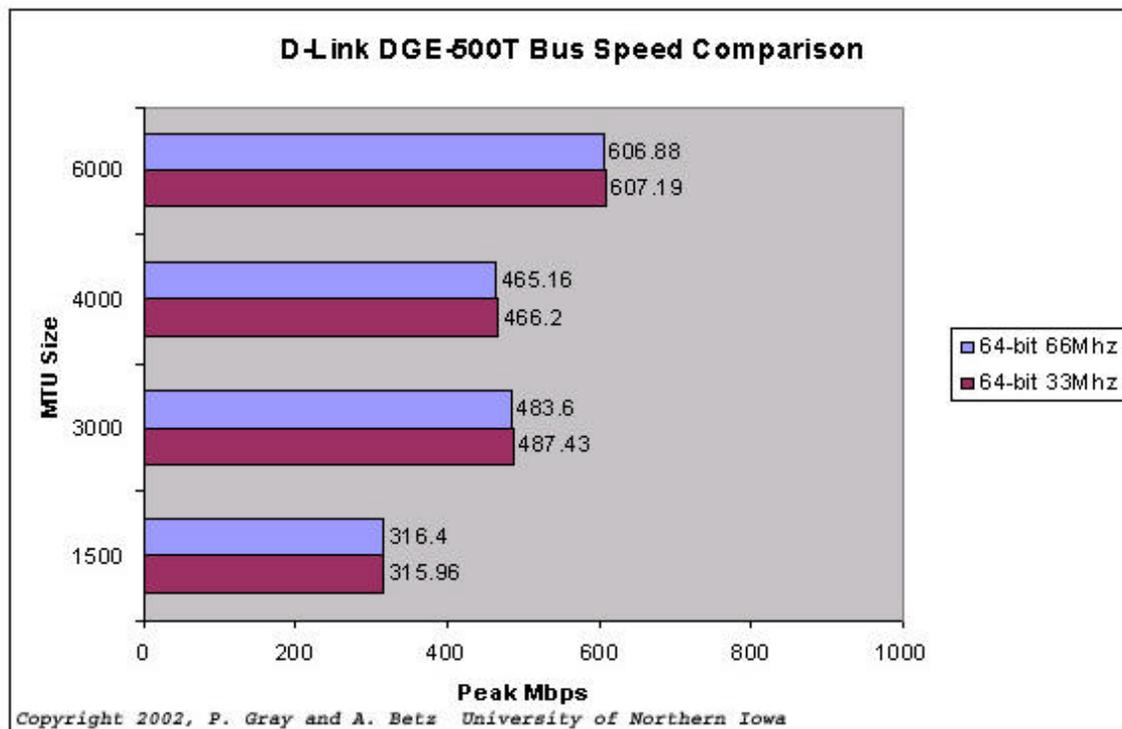


Figure 5

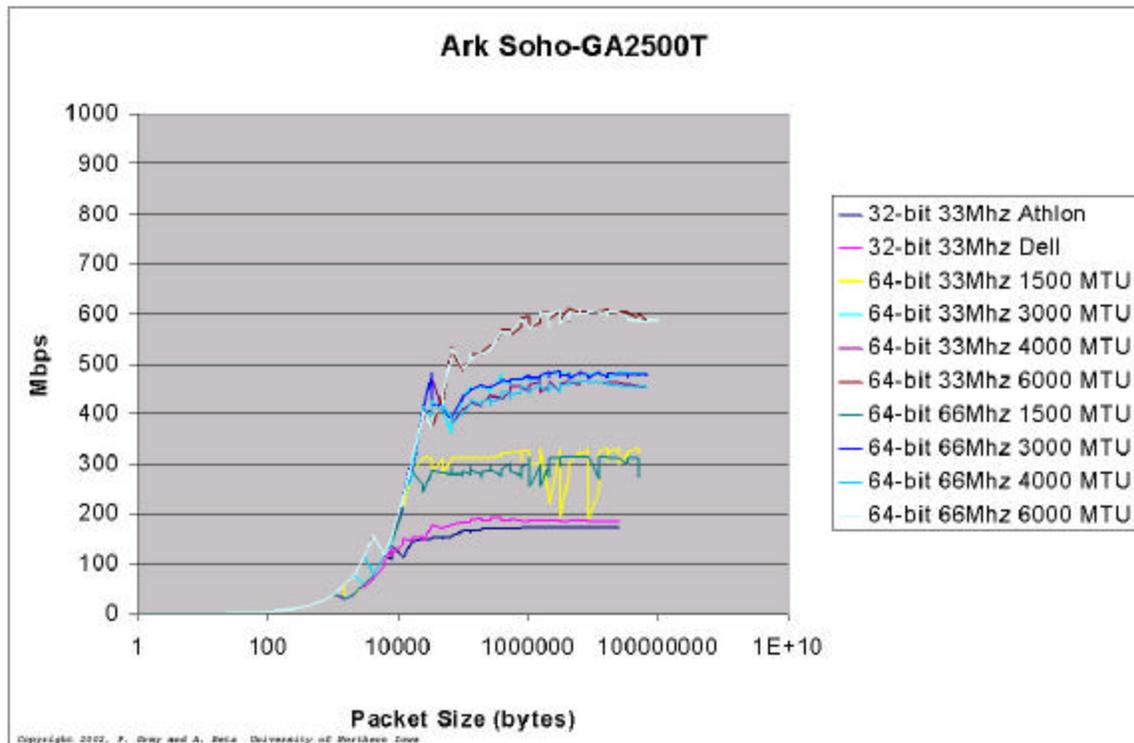


Figure 6

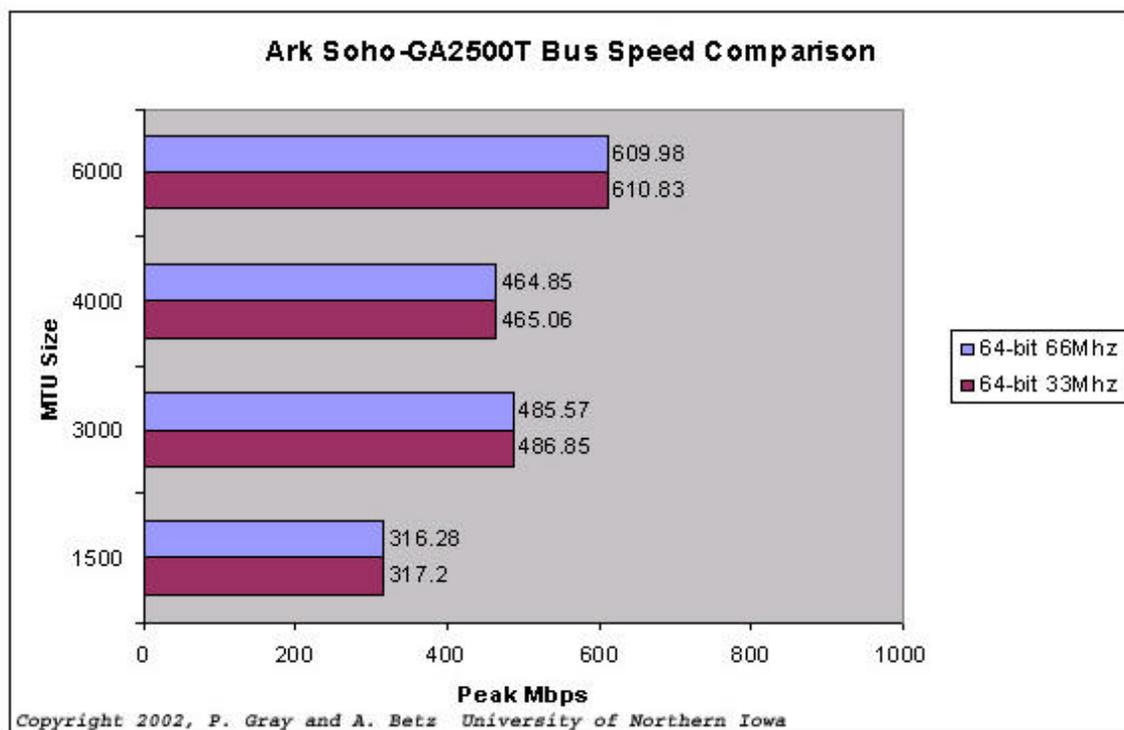


Figure 7

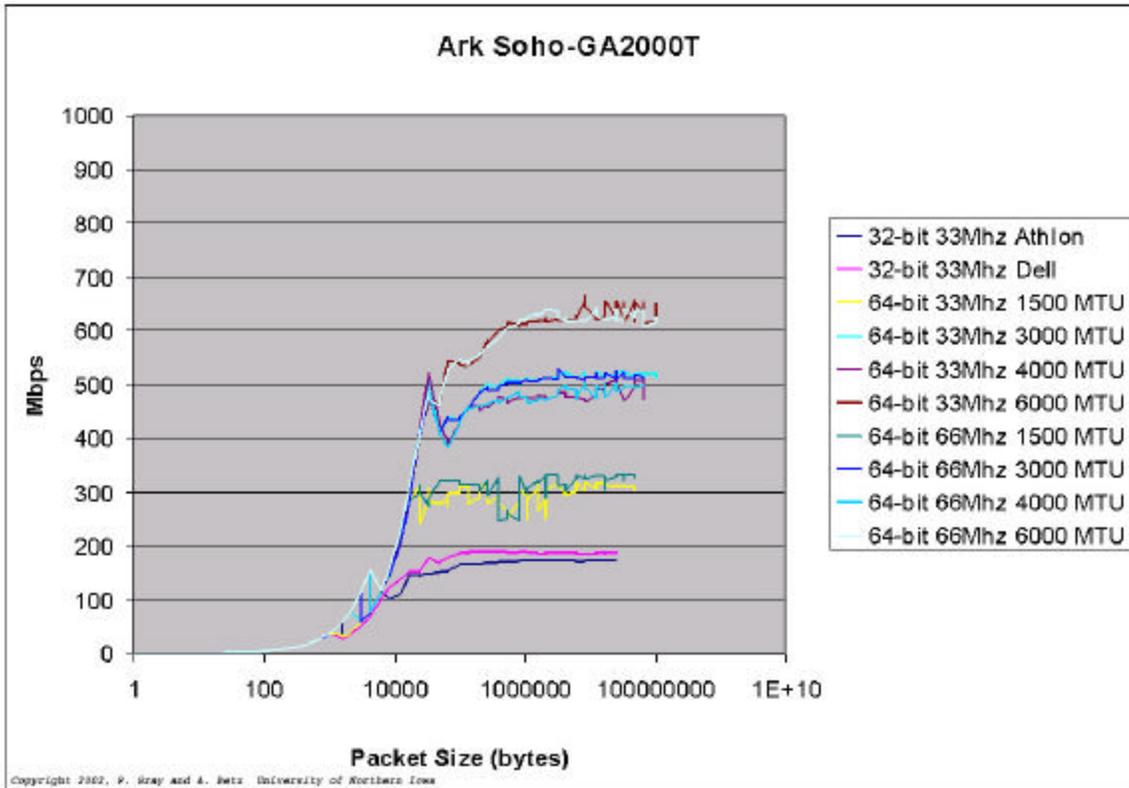


Figure 8

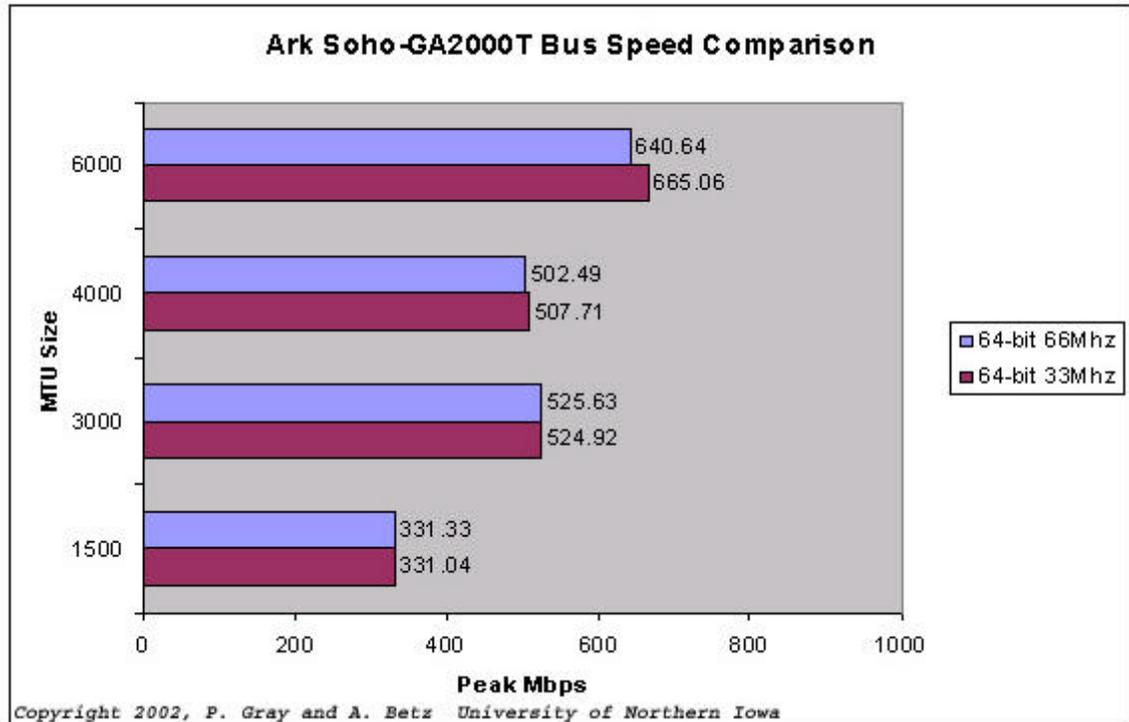


Figure 9

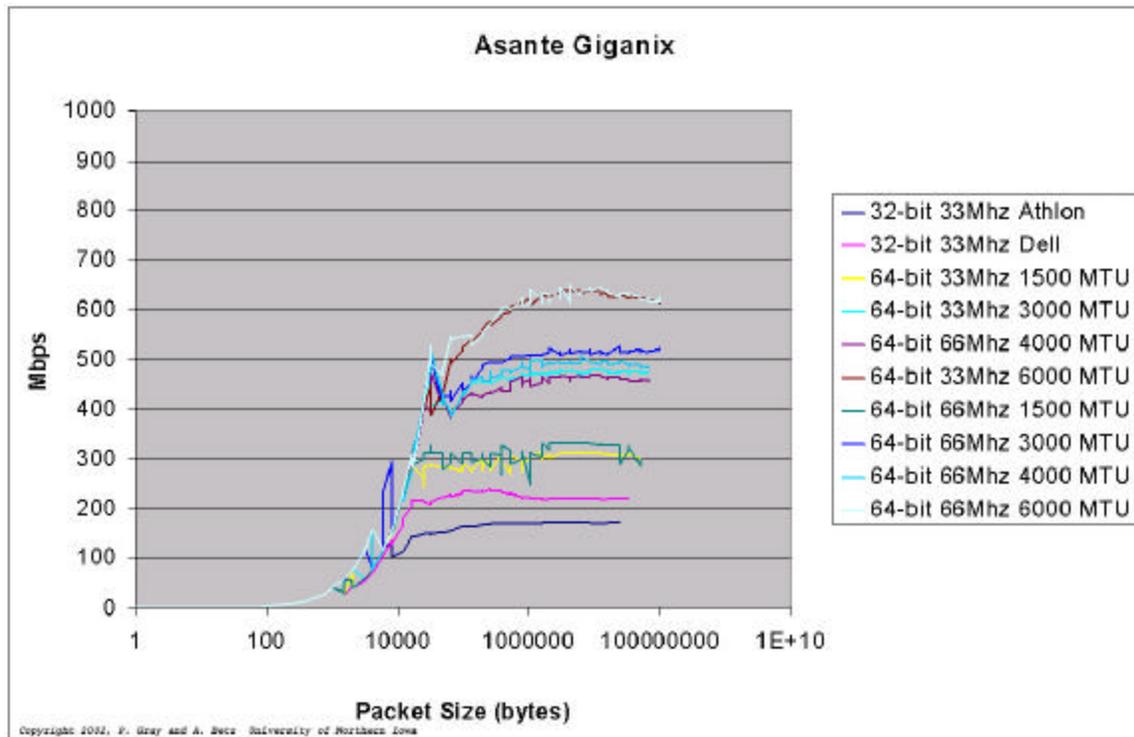


Figure 10

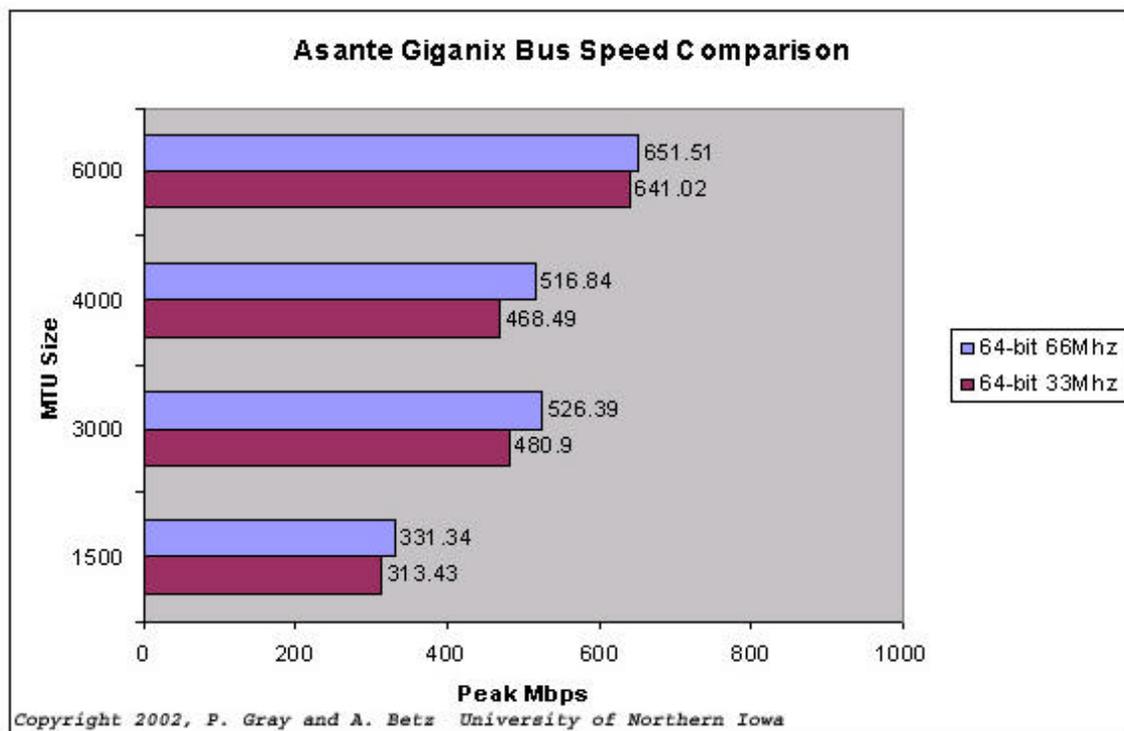


Figure 11

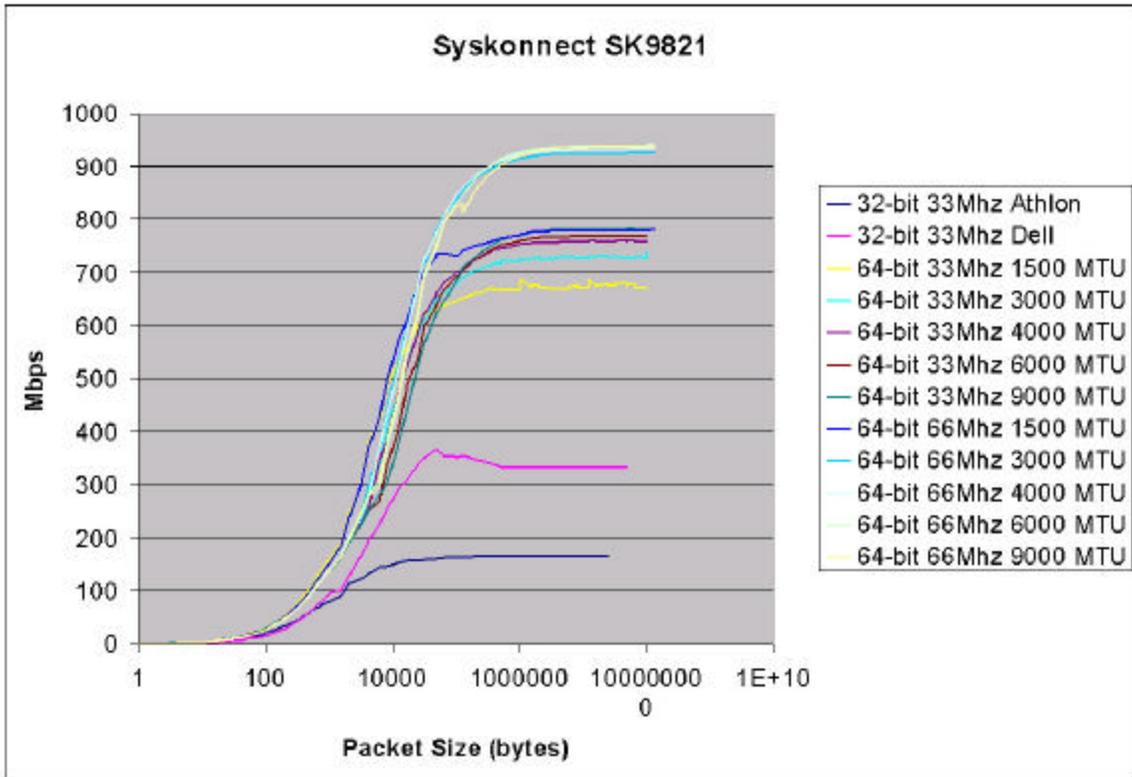


Figure 12

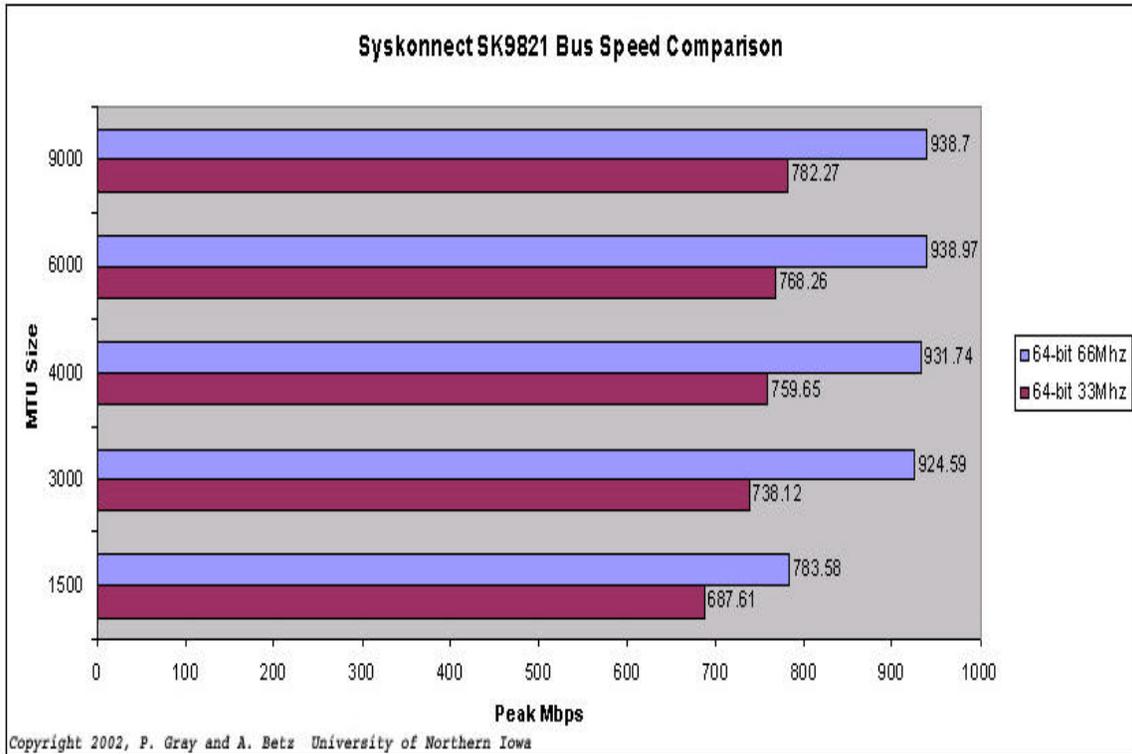


Figure 13

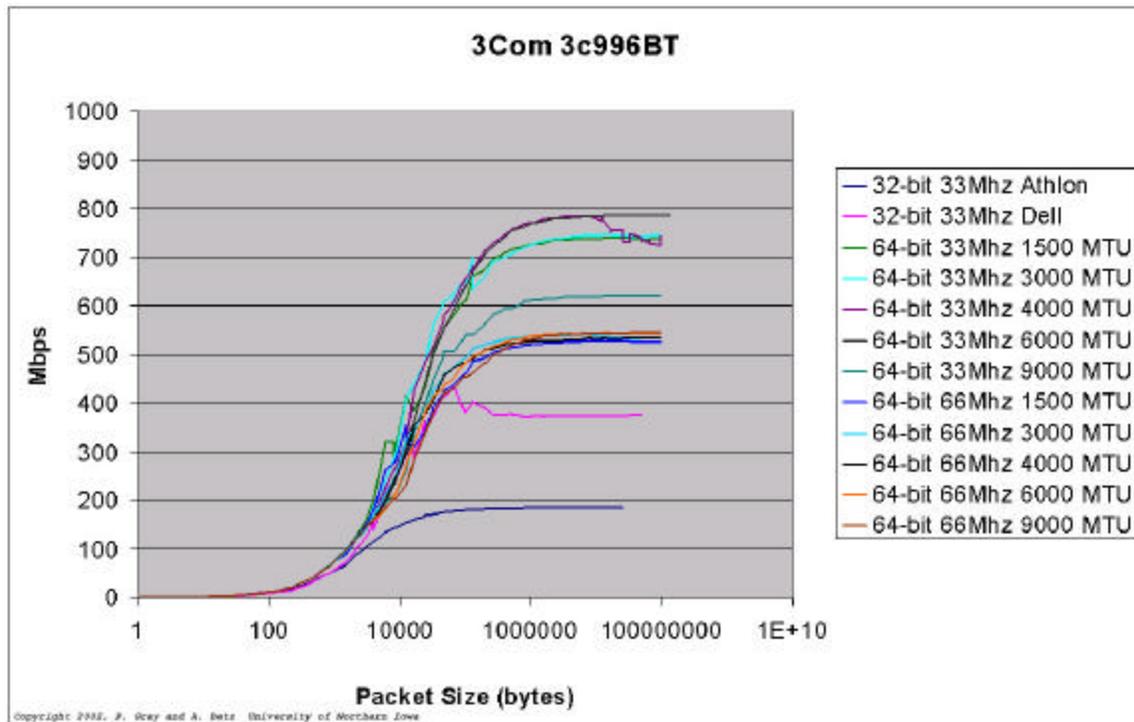


Figure 14

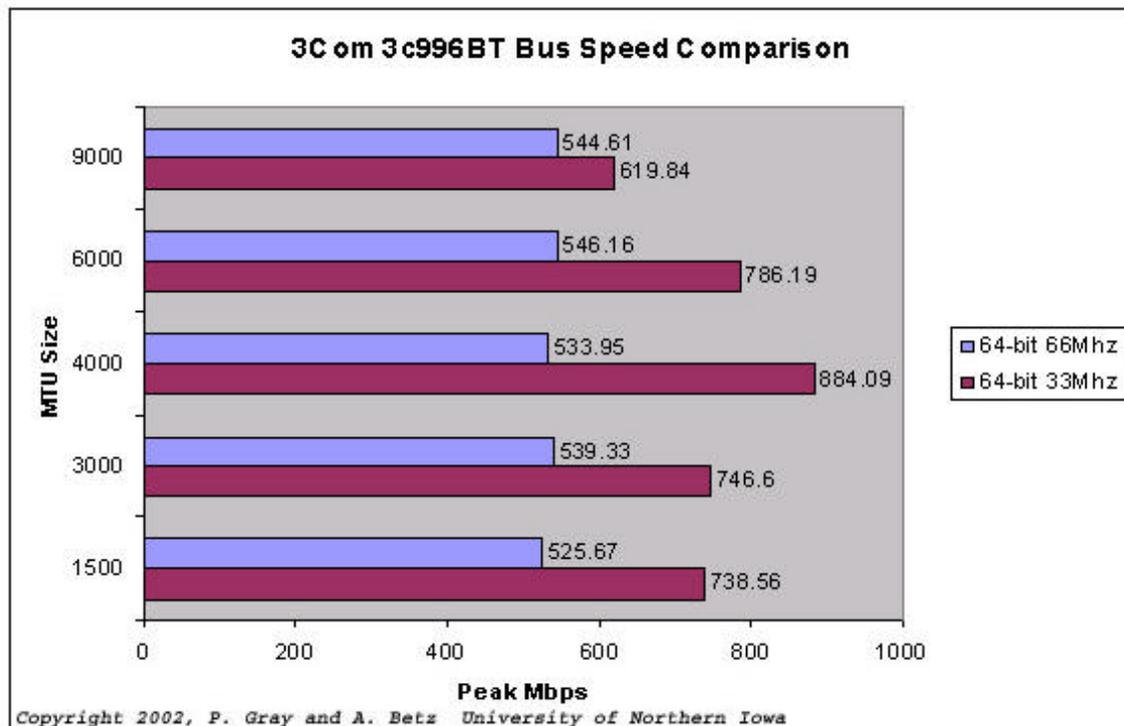


Figure 15

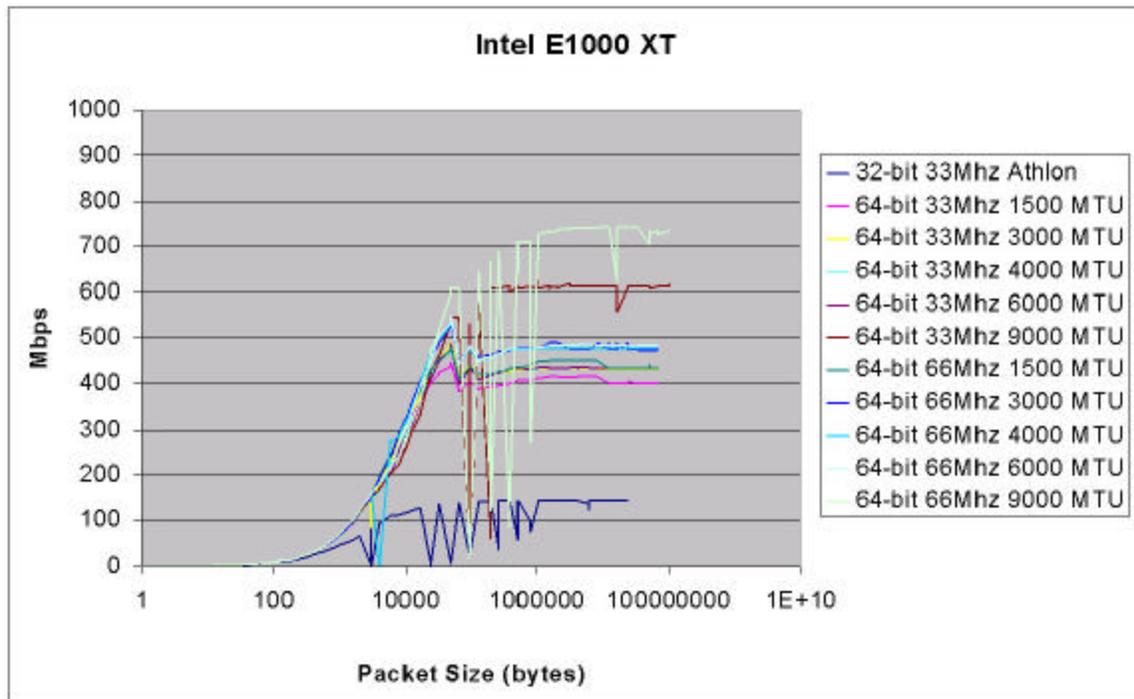


Figure 16

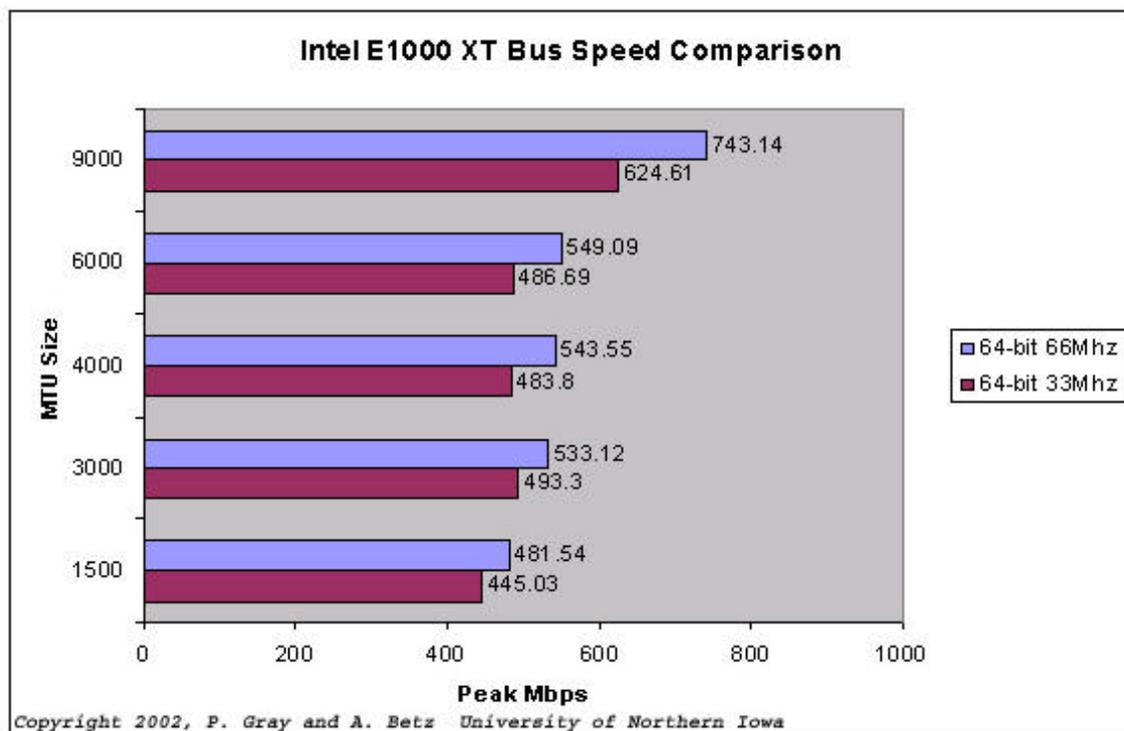
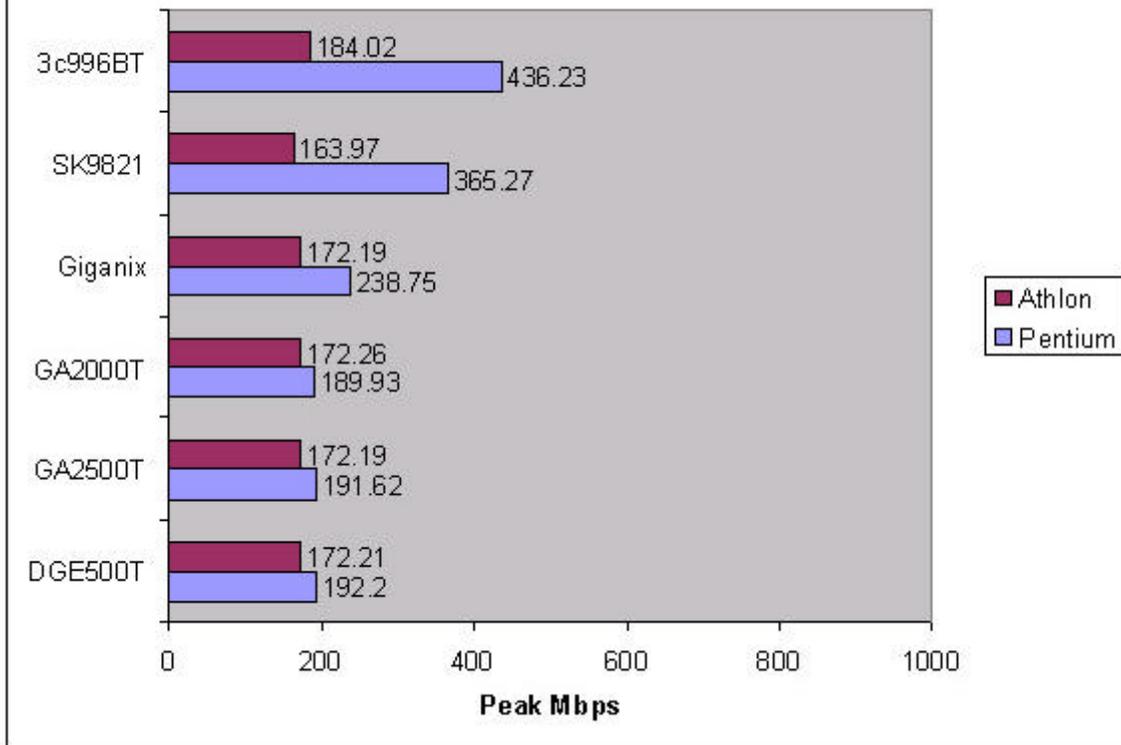
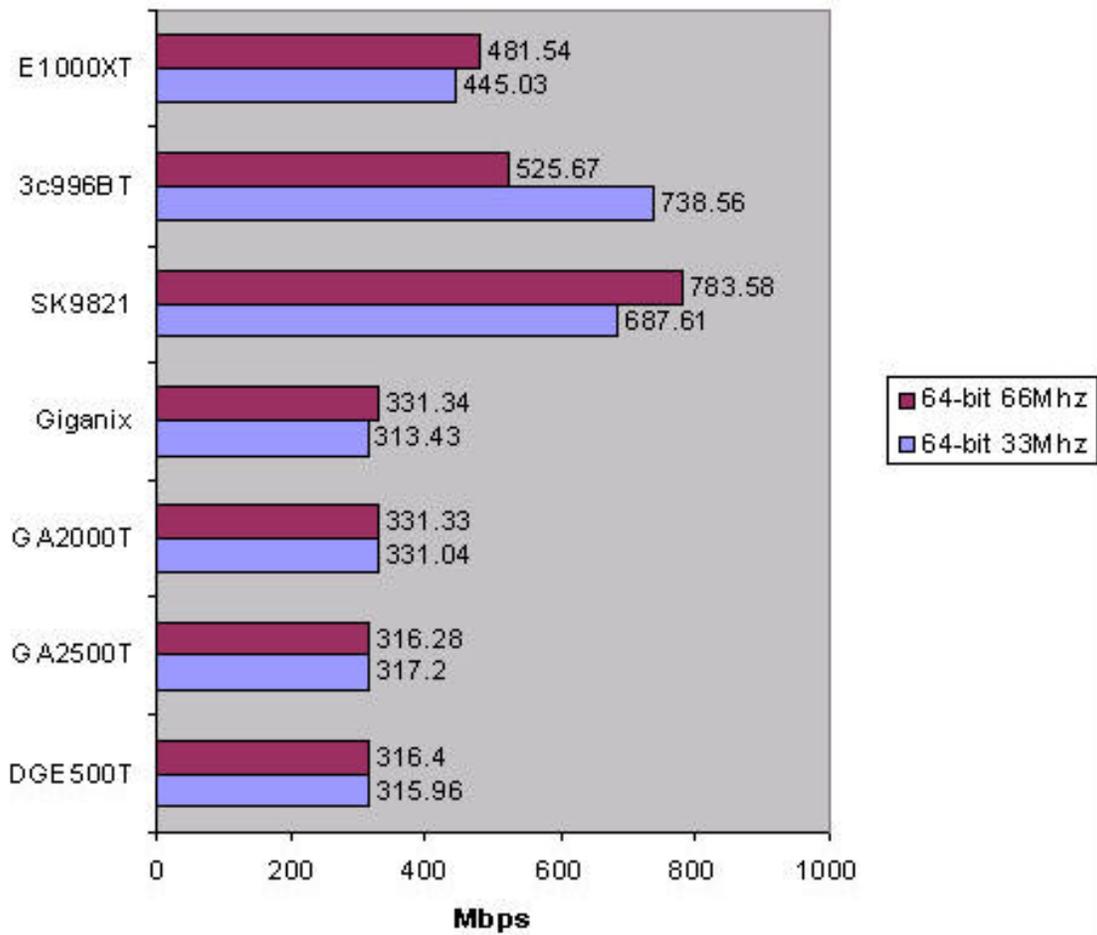


Figure 17

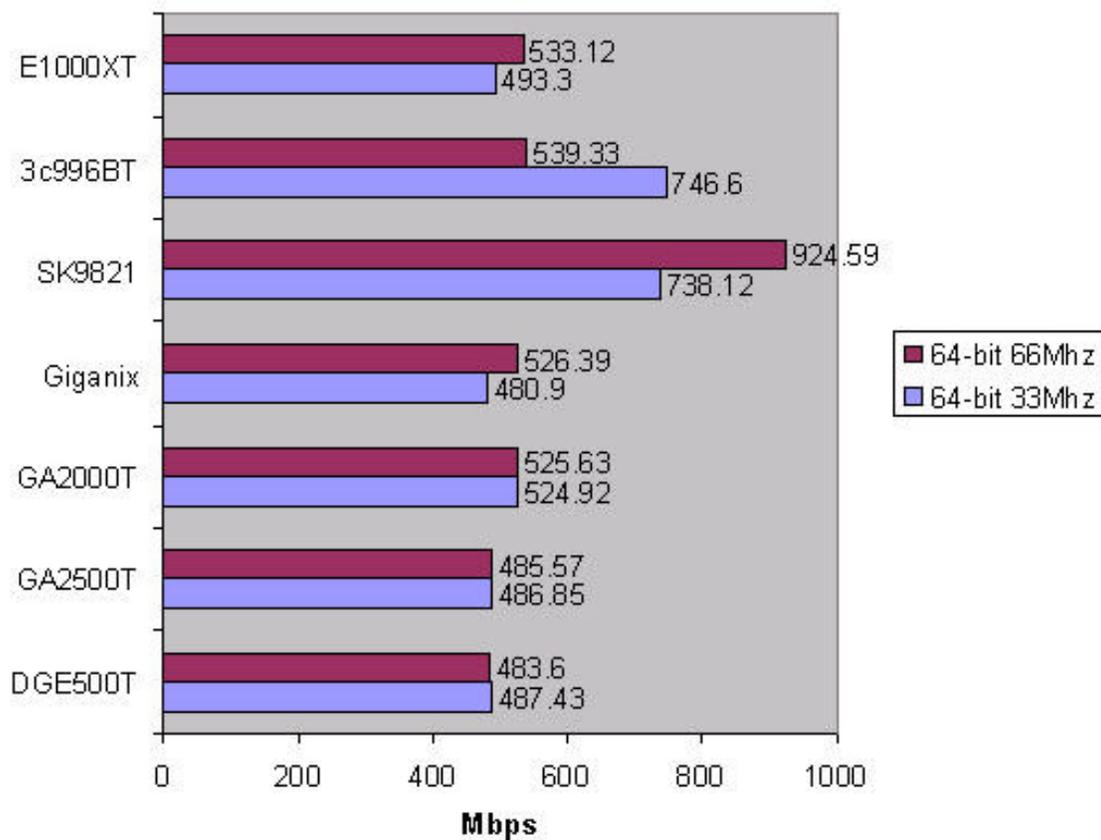
32-bit 33Mhz Bus Comparison



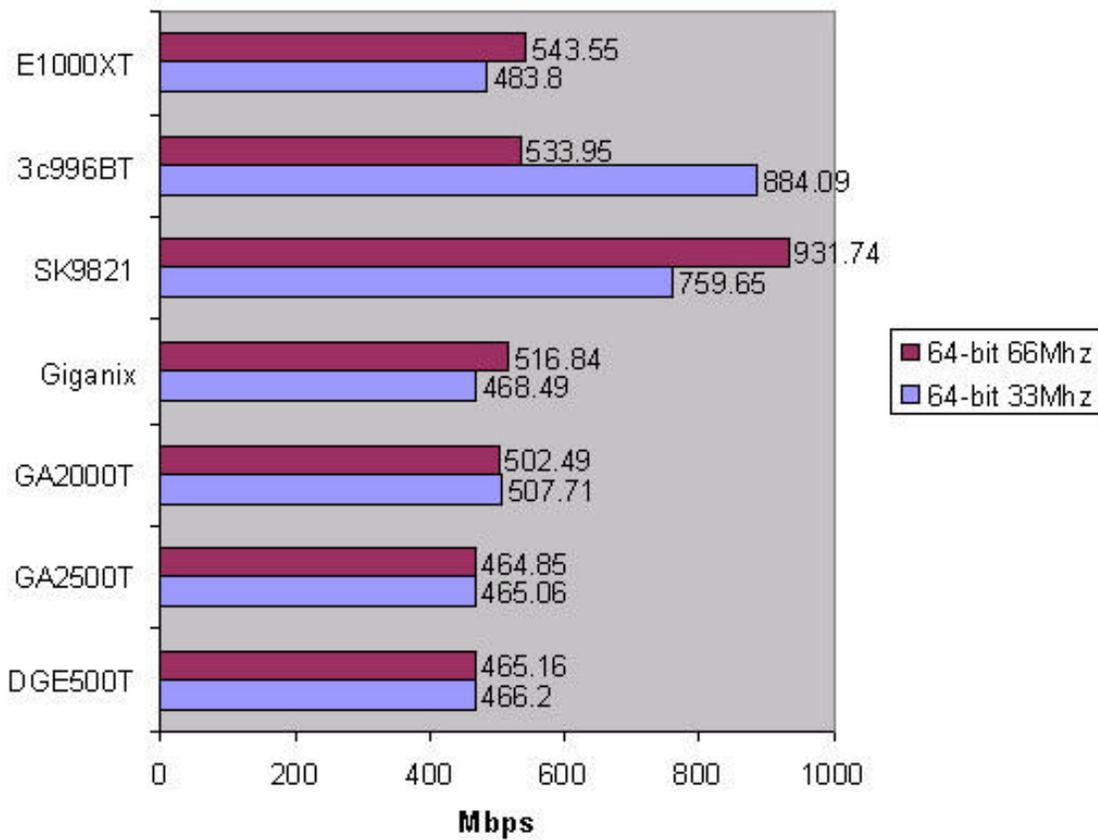
1500 MTU Card Comparison



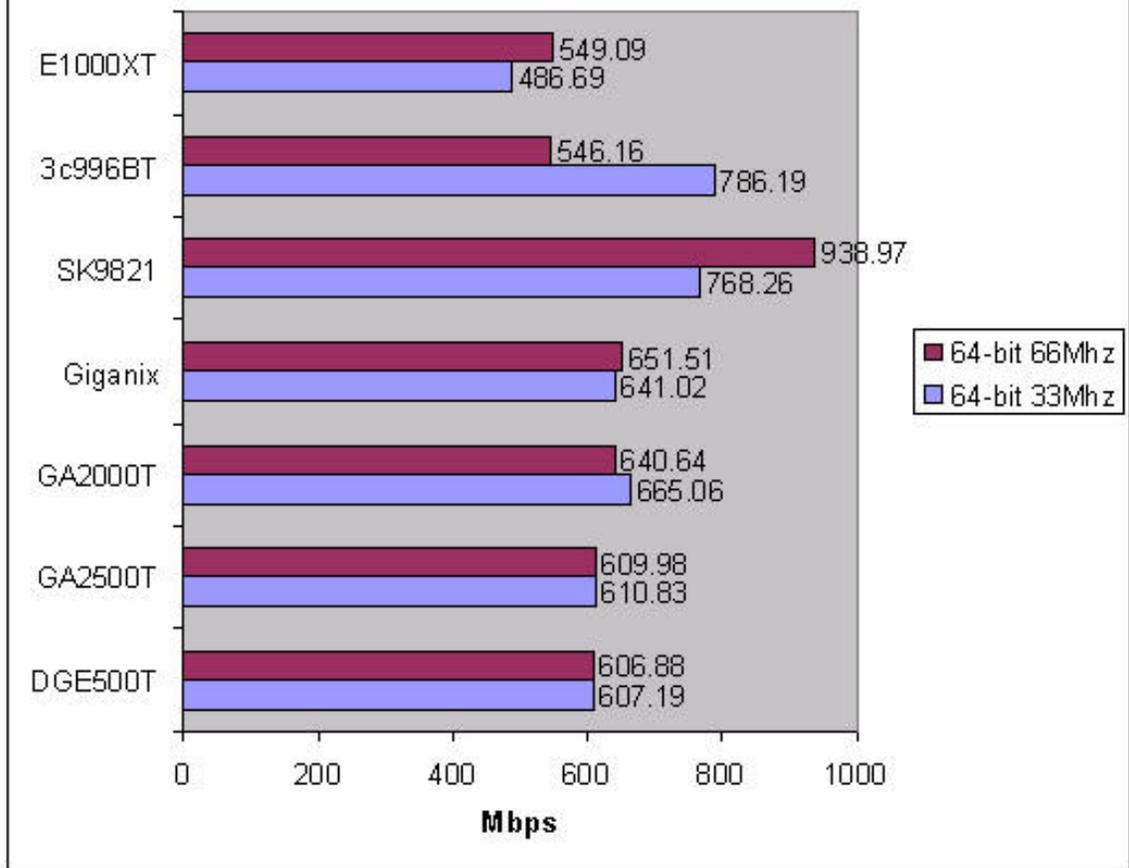
3000 MTU Card Comparison



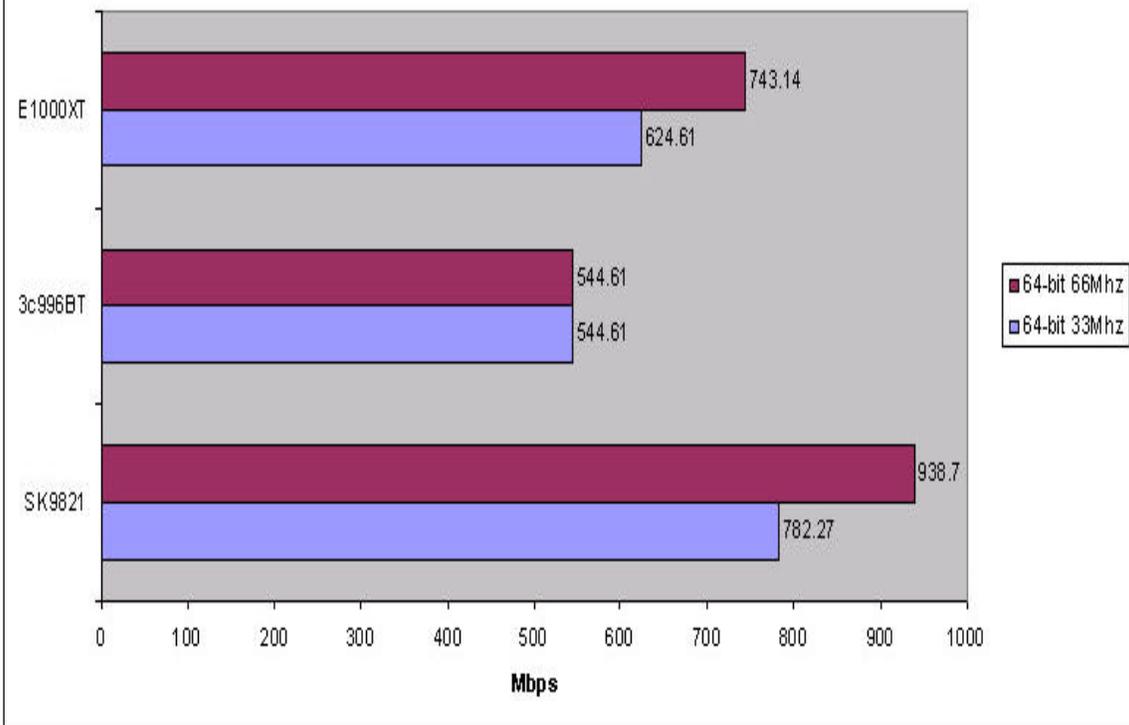
4000 MTU Card Comparison

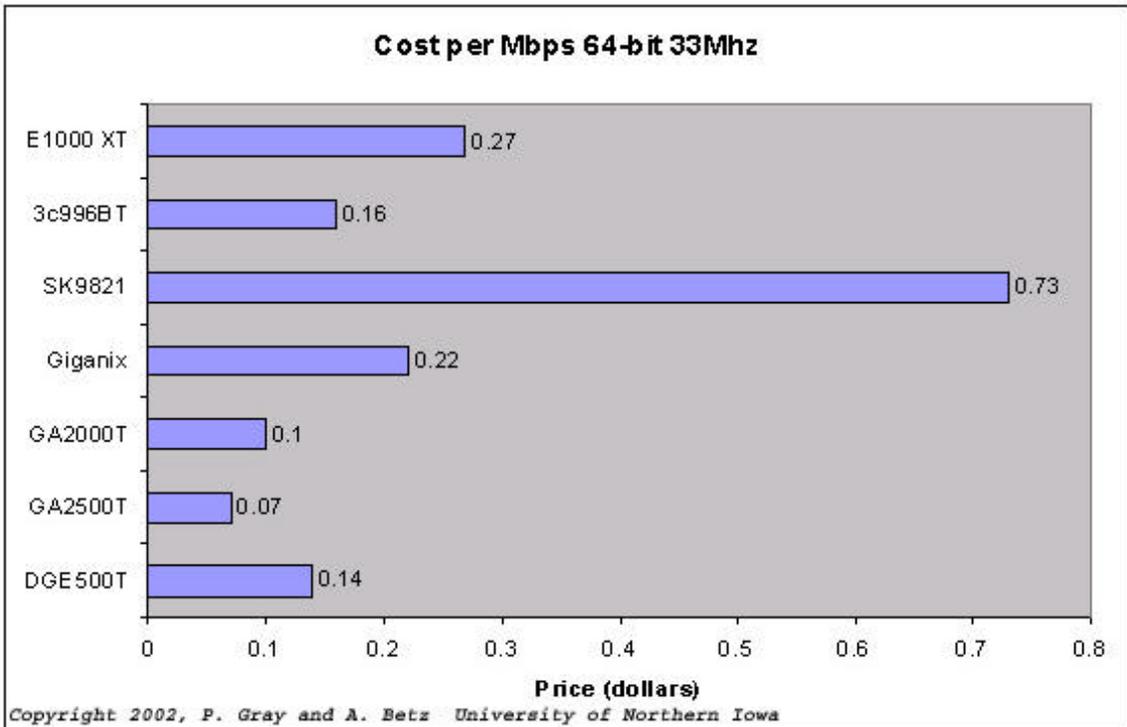
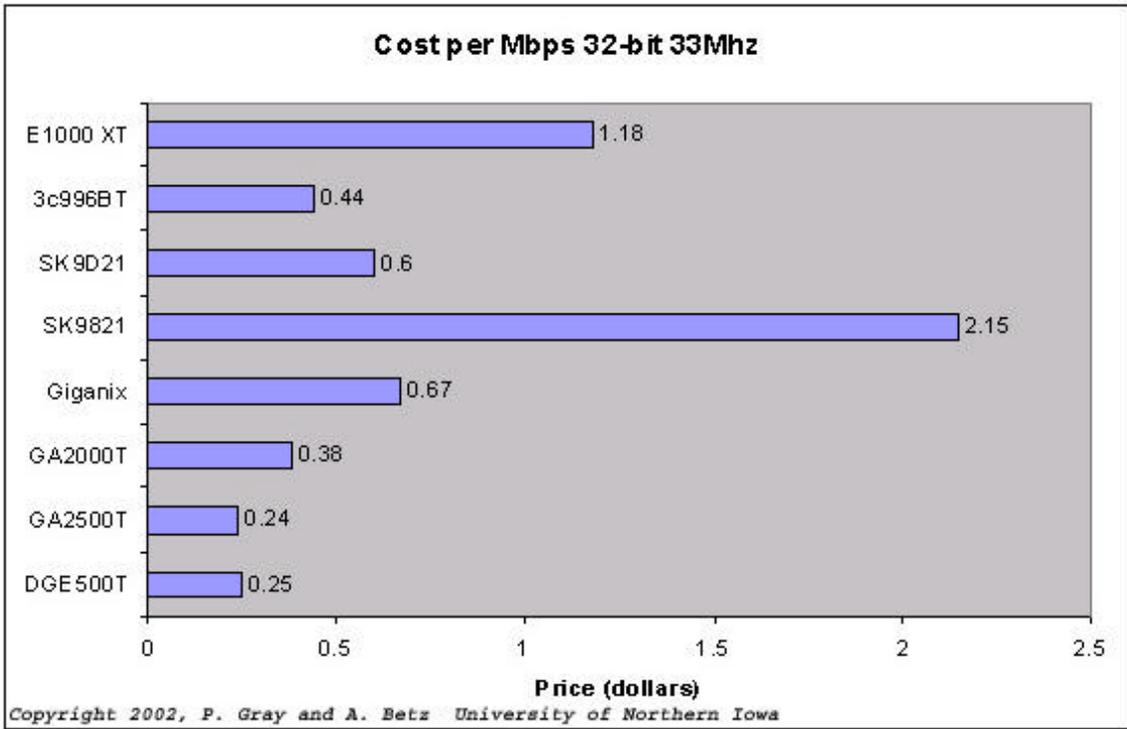


6000 MTU Card Comparison

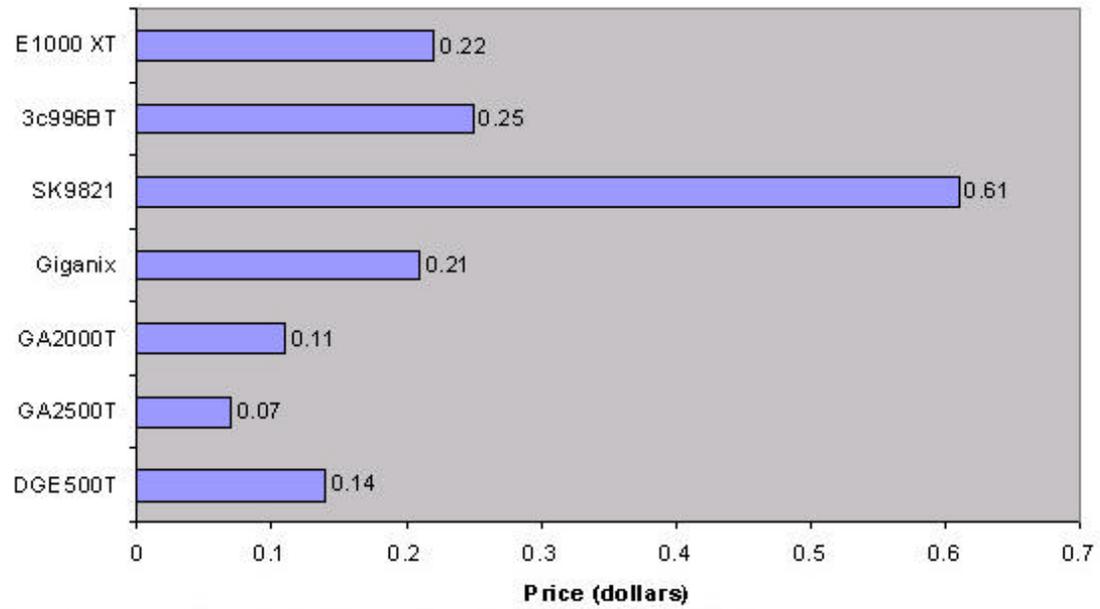


9000 MTU Card Comparison





Cost per Mbps 64-bit 66Mhz



Copyright 2002, P. Gray and A. Betz University of Northern Iowa