

HyperTransport™ Technology: Optimized for Digital Video & Audio

AMD Technology Evangelism Group

ADVANCED MICRO DEVICES, INC.

One AMD Place
Sunnyvale, CA 94088

Table of Contents

Introduction 3

 Inadequate Bus Architectures 4

 HyperTransport™ Technology Solution 4

System Bandwidth Challenges Presented by Digital Video 5

 Bandwidth, Bandwidth, Bandwidth 5

 Compression Issues 5

 Canopus DVStormSE Capture Card 6

Figure 1: Canopus DVStormSE Capture Card 7

System Bandwidth Challenges Presented by Digital Audio 8

 Reduced Latency is Key 8

 NVIDIA®'s SoundStorm Audio Solution with Integrated Dolby Digital Encoder 8

Figure 2: NVIDIA's SoundStorm Technology Supports Six Speakers, Plus Dolby Digital Output via SPDIF 9

How HyperTransport™ Technology Helps 10

 Very High Bandwidth 10

 Concurrency 10

 Full Support for PCI 11

 Low Latency 11

 Isochronous Support 11

 New HyperTransport Features 13

AMD Opteron™ and AMD Athlon™ Processors based on Hammer Technology 14

 Multiprocessor Support 14

Figure 3: Quad-Processor Platform Based on the AMD Opteron™ Processor. 15

 Integrated DDR DRAM Memory Controller 15

Figure 4: Functional Unit Diagram of an AMD Processor Based on Hammer Technology. 16

Summary 17

Introduction

Digital video and audio applications are putting strains on PC systems like never before. System performance is pivotal for compressing digital video and audio into smaller, more manageable sizes and for rendering effects. For consumer-level systems, the processor and accompanying system architecture must be able to decompress data pumped into the system, move it quickly between the processor and memory so the stream can be rendered with real-time special effects, and then recompress the stream into a suitable format for output. In professional systems, the movement of large *uncompressed* video within the system is the issue. If the system contains a bottleneck in any one of these areas, playback will be compromised and the compression of the data will cost extra precious time. Even the majority of the special-purpose hardware used for non-linear editing (NLE), where digitized video can be manipulated under software control and pieces of video can be cut, pasted, and copied anywhere in the timeline of a project in real-time, continue to rely heavily on processors and system architectures.

The primary issue is not getting the digital video and/or audio data into and out of the computer system; the issue is moving data around inside the system without compromising the integrity of the data stream. This is becoming more of a concern because processing power has continued to double every 18 months while the performance of the I/O bus architecture—the path along which the computer transfers data between system components—has lagged, doubling in performance approximately only every three years. Clearly, these two areas of technology are not being innovated at a comparable rate. To provide an example, matching fast components with a slow I/O bus is like driving a sports car on a crowded highway; if the system I/O bus cannot adequately move video frames from the memory to the processor to be modified, and then back to memory again so it can be displayed in a cohesive manner, the data will become congested and will suffer traffic jams much like the sports car during rush hour. Creators typically solve this problem by fully rendering the video stream prior to viewing it, wasting large amounts of time and possibly money in the process. This is tantamount to having the crowded lanes cleared of traffic prior to the arrival of the sports car to ensure it will not collide with other vehicles. This comparison may seem a bit drastic, but it makes sense when the subject matter relies on real-time special effects or is comprised of multiple tracks tied into a tightly choreographed timeline.

Typically, the inherent scalability of the system means that the more processing power provided by a system, the better the overall performance. But as systems continue to be based on outdated I/O technology; this will not be the case.

Inadequate Bus Architectures

Currently, systems use a limited bandwidth link comprised of a proprietary interface, between the Northbridge and the Southbridge. The Northbridge controls communication between the processor/ memory/video, and the Southbridge controls communication from hard drives and other external devices used for importing video and audio information. A limited I/O interface of this sort essentially creates a bottleneck as the chipset interconnect is forced to use a low bandwidth bus that must be used to share all external connections.

While these proprietary Northbridge and Southbridge connection technologies are capable of transferring data at throughputs of 266 megabytes to approximately one gigabyte per second and may contain smart electronics for better utilization of the bus, they still cannot sufficiently handle high-bandwidth data such as uncompressed digital video, high-definition video, or more importantly, the compression and decompression of digital video at the same time. Add to that concurrent audio encoding and broadband traffic and the buses all but grind to a halt.

HyperTransport™ Technology Solution

The core-logic architectures of today, particularly the interface between the Northbridge and the Southbridge, do not support isochronous data transfers in which data packets are time-dependent and must be streamed in a way that ensures fluid playback. Lack of isochronous support is the leading cause for the inconsistent playback and jittery images of digital video and the stuttering of digital audio. Another issue is the lack of concurrency in today's interfaces. The inability to transfer data simultaneously in both directions is a major limiting factor in non-linear video editing, whether using a software-based solution that taxes the processor and memory, or a hardware-based solution that relies on a dedicated capture card. When using a hardware-based solution, the system must be capable of handling very large files while maintaining a continuous, uninterrupted flow of data both to and from the drive. Software-based solutions make this equation even tougher. HyperTransport™ I/O technology supports both isochronous and

concurrent connections and has the capability to handle very large amounts of data at very high speeds.

System Bandwidth Challenges Presented by Digital Video

Bandwidth, Bandwidth, Bandwidth

The motion picture, television, and broadcast industries are just beginning to work with digital video. The growth of DVD, high-definition, and theatrical-quality video is placing an ever larger burden on the computers that are used for adding and mixing special effects and for compressing data for broadcast. With powerful NLE software tools available today, creators and producers can manipulate digitized video like never before—as long as the hardware is up to par.

I/O buses need to be capable of moving large amounts of digital video. For full-screen, full-motion broadcast quality NTSC digital video, the system must be capable of successfully moving 29.97 frames per second at 720 by 480 pixels in RGB (red, green, blue) 24-bit color, which equates to 31 megabytes of data per second. A system built to handle high-definition HDTV (1080i digital video) must be capable of successfully moving 24 frames per second at 1920 by 1080 pixels for a grand total of 149 megabytes of data per second. Theatrical quality resolutions are reached via a process called “Up-RESing” where lower resolution material is upped to higher resolution material through a process of interpolation where new pixels are created by comparing and averaging existing ones. Interpolative Up-RESing requires images be moved to 2048 by 1151 pixels in size (at 24 frames per second) to allow for images to be fit into larger resolution areas of the 35mm negative. As expected, this process requires much more processing power and time, with data rates equating to a whopping 170 megabytes of data per second.

Of course, these bandwidths are for single streams of video. The need for multiple video streams in high-production work, especially when multiple video streams are used for creating transitions and when special effects are applied, places an even greater bandwidth burden on the system.

Compression Issues

Compression is the translation of source material, comprised of video, audio, or a combination of both, using a variety of computer algorithms to reduce the amount of data

required to accurately represent the content. In simpler terms, compression reduces the sheer size of uncompressed video to a size that is manageable for storage and distribution.

While dedicated capture cards support hardware-based special effects and digital video (DV) and MPEG compression codecs, they do not have all of the special effects and compression schemes required by the developer. These chores are left to software based plug-ins and applications typically bundled with the card while placing the majority of the processing load on the host processor. A typical scenario would be a DV capture card using a software codec to encode a video stream into an MPEG format, and sending the finished product to either the hard drive or an external device. An example of this is the **DVStormSE** capture card from Canopus.

Canopus DVStormSE Capture Card

The Canopus **DVStormSE** capture card provides real-time DV editing and boasts capabilities including the use of up to five independent video tracks, more than 20 title and graphic tracks, and 24 filters. It also employs Canopus's Scalable Technology Architecture to provide more features and increased performance as processor power increases. By relying on the processor and system architecture, the card can offer more real-time capabilities over time, ultimately increasing the number of real-time video tracks beyond five, and increasing the number of title and graphics tracks indefinitely and delivering more real-time creativity.

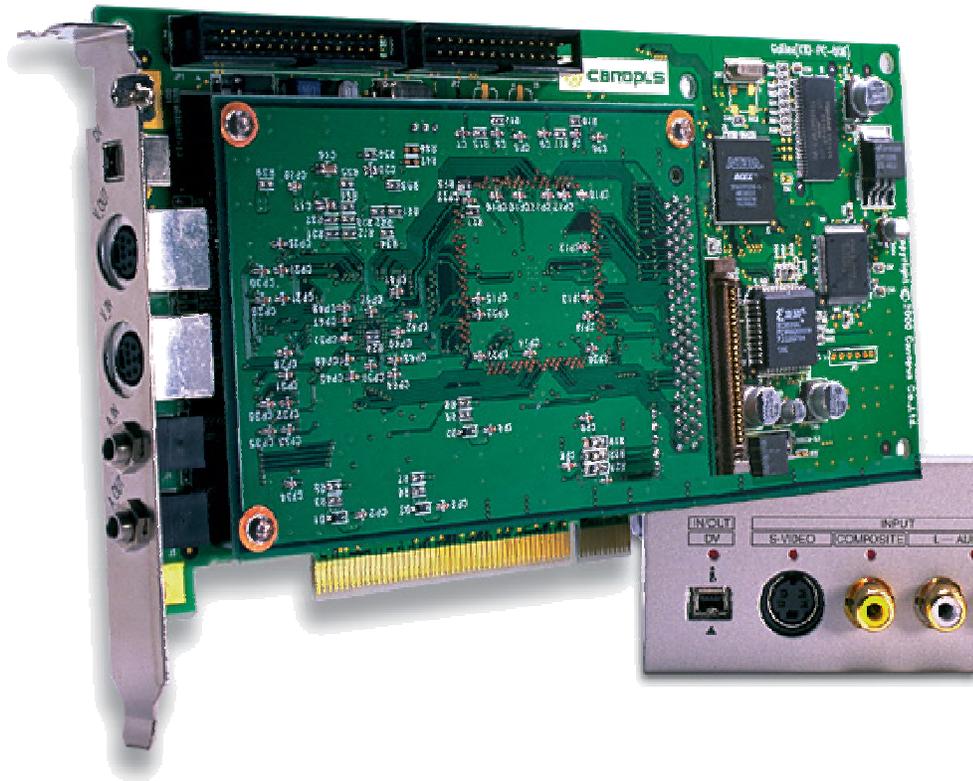


Figure 1: Canopus DVStormSE Capture Card.

I/O bandwidth continues to be an issue, whether digital video relies on a hardware or software codec to handle compression and decompression chores. The data still needs to be moved within the system, and the bottleneck issue will continue to be a problem. This is especially true as digital video sources become larger, and current bus technology continues to be unable to support high-speed transfers and concurrent transactions where data can be sent in both directions at the same time. Digital video is not the only area where this is important.

System Bandwidth Challenges Presented by Digital Audio

Reduced Latency is Key

The high volumes, increased computing power, and reliability of personal computers has contributed to its outgrowth from the home studio into the professional studio as the platform used for the recording, editing, and mixing of high fidelity audio content creation. In this environment, the real-time recording of multiple independent tracks, the mixing of special effects, and the real-time editing and playback of these tracks place tremendous strains on systems that must be capable of handling these streams without skipping or glitches. While systems utilizing older multi-drop, shared buses such as PCI can sustain today an aggregate of 20 megabytes per second of raw audio data comprised of 48 tracks (24 in and 24 out simultaneously), each sampled at 24 bits at 96kHz, the reality is these shared buses cannot always reliably handle these audio streams when other devices are sharing the bus.

Older multi-drop, shared I/O buses like PCI are half-duplex links where data can only be sent in a single direction at a time while other devices must wait their turn. Multi-drop buses, which can be connected to individual devices but divide the total available bandwidth between them, use an arbitration method that works on an interrupt basis with shared devices to guarantee bandwidth is distributed among all devices. To compound the issue further, many different arbitration schemes exist among different architectures that can lead to inconsistencies between devices and dissimilar platforms. Also, the inefficiency, bus turnaround, and overhead imposed on data sent across older shared buses like PCI greatly decreases the amount of overall bandwidth available to the bus and create latencies, further encumbering fluid communication. As can be expected, this is not the optimal scenario for moving time critical audio streams without glitches. HyperTransport I/O technology, with its low latency and guaranteed high bandwidth attributes, is designed to solve these very issues.

NVIDIA®'s SoundStorm Audio Solution with Integrated Dolby Digital Encoder

Digital audio is experiencing the same fundamental issues as digital video, albeit to a lesser extent. Audio innovation is making many new inroads as NVIDIA's nForce™ and nForce2 family of platform processors are a testament. Part of the overall nForce

digital media platform solution, including the nForce Integrated Graphics Processor (IGP) and the nForce System Platform Processor (SPP). The Media and Communications Processor (MCP/MCP-T) incorporates the NVIDIA nForce Audio Processing Unit (APU) which features the industry's first Dolby Digital Interactive Content Encoder. This breakthrough technology can dynamically encode any multi-channel 2D or 3D audio source into Dolby Digital 5.1 in real-time and output it digitally. The nForce family of platform processors utilizes a high-speed HyperTransport I/O bus interface for internal communications, delivering unprecedented levels of PC performance.

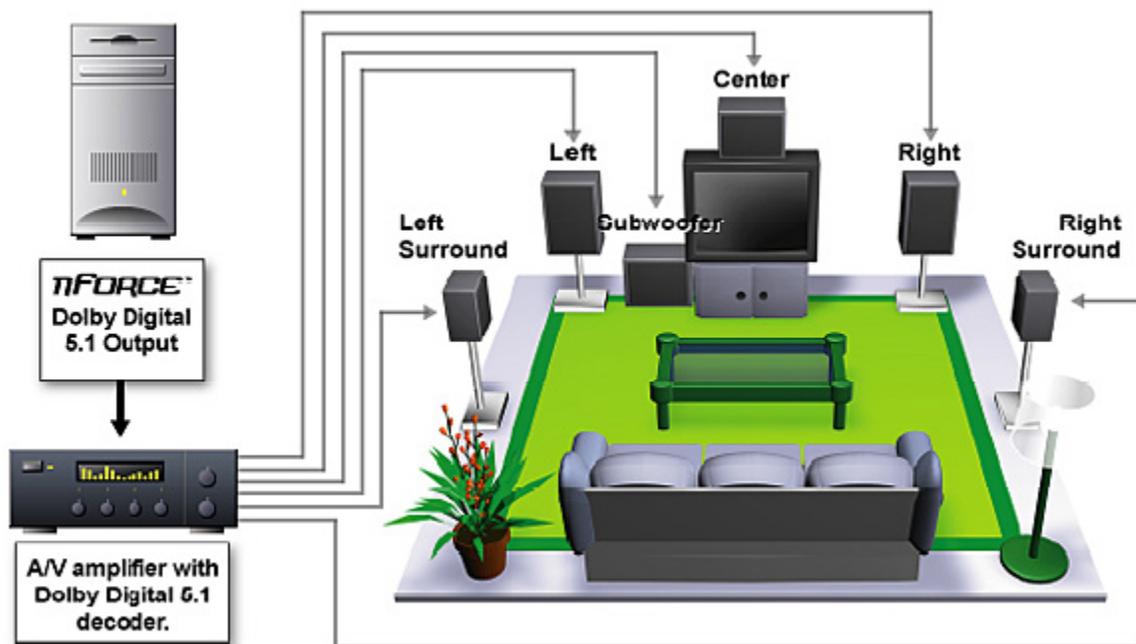


Figure 2: NVIDIA's SoundStorm Technology Supports Six Speakers, Plus Dolby Digital Output via SPDIF

NVIDIA, finding that today's PC applications are increasingly complex with advanced 3D graphics, high-speed networking, streaming video, and cinematic 3D audio, found that no other current technology allowed for the implementation of full Dolby Digital 5.1 2D or 3D audio processing and broadband networking in the MCP. The nForce Audio Processing Unit (APU) integrates the Dolby Digital Interactive Content Encoder into a programmable DSP with a fix-to-float format engine. This engine is used to take the output of the Global Processor and encode it into a Dolby Digital (AC-3) stream, allowing users to experience true theater-quality, multi-channel surround sound,

rendered in real-time, on their Dolby Digital-equipped PCs, Mini-Disc players, and home theater systems.

How HyperTransport™ Technology Helps

Very High Bandwidth

HyperTransport links are capable of extremely fast signaling, and are designed to operate at clock speeds ranging from 200MHz up to 800MHz. HyperTransport links utilize double data rate (DDR) technology, and transfers two bits of data per clock cycle, for an effective transfer rate of up to 1,600 megabytes per second in each direction. Since transfers can occur in both directions simultaneously, an aggregate transfer rate of 6.4 gigabytes per second in a 16-bit HyperTransport I/O link and an aggregate transfer rate of 12.8 gigabytes per second in a 32-bit HyperTransport I/O link can be achieved.

Basically, the HyperTransport I/O link uses two point-to-point unidirectional links instead of the single-ended signaling employed by parallel buses like PCI. These two wires employ a differential signaling technique that reads data as the difference between the two signals sent, allowing it to operate at very high clock rates without suffering from electrical issues susceptible to parallel buses such as bouncing signals, interference, and cross-talk from adjacent signals; all of which can potentially disrupt audio and video streams. The HyperTransport link is also a “packetized” bus, which means addresses, data, and commands are sent along the same wires allowing designers to implement much narrower links.

Concurrency

Concurrency is the ability of the I/O bus to transmit data in both directions at the same time. In the networking world, this is referred to as full-duplex operation where signals can be transmitted and received simultaneously. HyperTransport I/O links support full-duplex operation and can benefit systems by offering concurrent support to areas where it has been absent in the past. Systems based on current I/O technologies like PCI suffer arbitration issues and are not capable of delivering the bandwidth and concurrency needed for professional level digital video and audio applications. When time-critical data is of the utmost importance, support for concurrency is vital.

Full Support for PCI

HyperTransport technology provides high speeds while maintaining full software and operating system compatibility with the PCI bus used in most systems today.

HyperTransport technology has been designed for concurrent connections, running PCI cycles and other types of I/O cycles at the same time. HyperTransport technology is designed to interface with today's I/O standards including AGP, PCI, PCI-X, IEEE-1394, USB 2.0, PL-3, SPI-4.2, and Gigabit Ethernet as well as next generation buses including AGP-8X, InfiniBand architecture, PCI-X 2.0, PCI Express, SPI-5, and 10 Gigabit Ethernet among others.

Low Latency

HyperTransport technology provides low latency access into main memory and high bandwidth to I/O devices. Because of the high speeds of the HyperTransport I/O link and the low latency of the channel, multiple chips can be daisy-chained together without a significant performance impact. Data streams that are latency and bandwidth critical receive isochronous data support to ensure ultra fast access and bandwidth to main memory.

Isochronous Support

HyperTransport technology includes support for time-dependent isochronous data such as streaming digital video and real-time voice. Isochronous data is characterized as a stream of data with packets scheduled at regular, periodic intervals. Ensuring that synchronous data reaches its destination is critical, especially when digital video is being encoded and compressed in real-time, and when the video is of movie quality. Current system architectures, particularly the interface between the traditional Northbridge and Southbridge chips, do not support the concept of an isochronous data stream.

To maximize throughput of digital video and audio, the destination must receive its data with minimal delay, requiring the need for guaranteed isochronous latency and dedicated bandwidth. Without these guarantees, latency can occur, ultimately resulting in more required bus bandwidth. The result is sub-par frame rates, audio sync problems, and inadequate playback of video and jittery images. To provide an example, imagine 100 people piling onto a bus or an airplane without any regard for order; some people might be blocked for a long time. This is the inefficient way in which PCI moves data today. With HyperTransport technology, a ticket is provided to each individual so they can enter

in an orderly manner while tickets for the next trip are being lined up before the plane has even left the gate. PCI, by taking a first-come first-served approach, lowers the efficiency of a bus that doesn't have much bandwidth to start with.

To support high-priority isochronous communication, HyperTransport technology includes support for an operating mode in which the number of virtual channels is doubled, and associated flow control buffer types are doubled. Transactions also have an isochronous bit associated with them that must be maintained by tunnels even when isochronous mode is disabled.

In Isochronous mode, there are two classes of service defined. The high-priority service class is intended to support isochronous traffic, and the low-priority service class is intended for all other traffic. High-priority traffic is serviced before low-priority traffic, and is prioritized in a queue so that low-priority traffic is not gridlocked and the isochronous priority stream is not taking advantage of 100 percent of the bus. The overall available bandwidth of HyperTransport technology should be able to handle both isochronous and low-priority traffic. This eliminates the need for a fairness algorithm to regulate the insertion of isochronous traffic. Isochronous flow control is enabled on a per-link basis to allow isochronous requests and responses to "tunnel" through non-isochronous devices on a chain.

Of course, overall bandwidth within the system is still limited by the maximum and typical data rate the hard disk can provide. A HyperTransport I/O link capable of providing a data throughput of 12.8 gigabytes per second would have problems actually communicating with slower devices at this speed because the other parts within the system could not fully utilize the bandwidth of the HyperTransport technology bus. But a bus with this type of capability also opens doors of opportunity such as the ability to add multiple independent links to ensure each connection can communicate at its highest possible speed. For example, multiple PCI-X slots can be implemented so that the bus is not shared. Also, extra bandwidth can be used when dealing with large data types concurrently.

NVIDIA's StreamThru technology is an excellent example of the capabilities that HyperTransport technology provides. StreamThru is NVIDIA's isochronous data transport system, providing uninterrupted data streaming for networking and broadband

communications. By interfacing the integrated 10/100Base-T Ethernet controller to an isochronous-aware internal bus and single-step arbiter, StreamThru assists in making streaming video and audio smoother and jitter-free.

The HyperTransport link between NVIDIA's nForce and nForce2 platform processors; the Integrated Graphics Processor (IGP), the System Platform Processor (SPP), and the Media and Communications Processor (MCP/MCP-T), support multiple virtual channels of isochronous data streams. The HyperTransport I/O link controller on the IGP and SPP side will dispatch both isochronous and non-isochronous requests to an intelligent arbiter, which supports memory latency and bandwidth delivery for both the read and write data paths, resulting in fast media streaming, packet transfers, and data downloads.

New HyperTransport Features

Work on the HyperTransport I/O link specification continues to add new features that will help in areas of digital video and audio. New HyperTransport technology networking extensions have been included, adding peer-to-peer communications where packets can be sent directly between peer devices without having to be reflected via the host device, and 64-bit addressing that supports large memory models in excess of one terabyte. The addition of these extensions also includes a message-passing protocol that adds the ability to stream a sequence of packets to a given address, and the addition of 16 streaming point-to-point flow controlled virtual channels that support millions of end-to-end flow controlled individual streams. It also creates the ability to bridge Systems Packet Interface (SPI) 4.2 traffic that is typically used in communications data plane chips, and an enhanced error recovery protocol that automatically detects and recovers from data errors that are likely to occur when HyperTransport technology links become even faster in the future.

AMD Opteron™ and AMD Athlon™ Processors based on Hammer Technology

Multiprocessor Support

HyperTransport I/O technology is more than a system bus; it is a processor bus as well. AMD multiprocessor designs based on the AMD Opteron™ processor and HyperTransport I/O technology offer a great deal of performance for digital video and audio applications. Unlike other typical configurations, AMD multiprocessor-based systems with HyperTransport technology do not share a single bus to system memory with the other devices within the system. Depending on how they are arranged, an AMD multiprocessor-based system with the AMD Opteron processor using two, four, or even eight processors, can seamlessly communicate amongst themselves using three HyperTransport links that are built directly into the processor die. And since HyperTransport technology offers bandwidth to spare, the latencies from processor to memory are significantly low while performance yields will continue to increase as advances are made in memory technology.

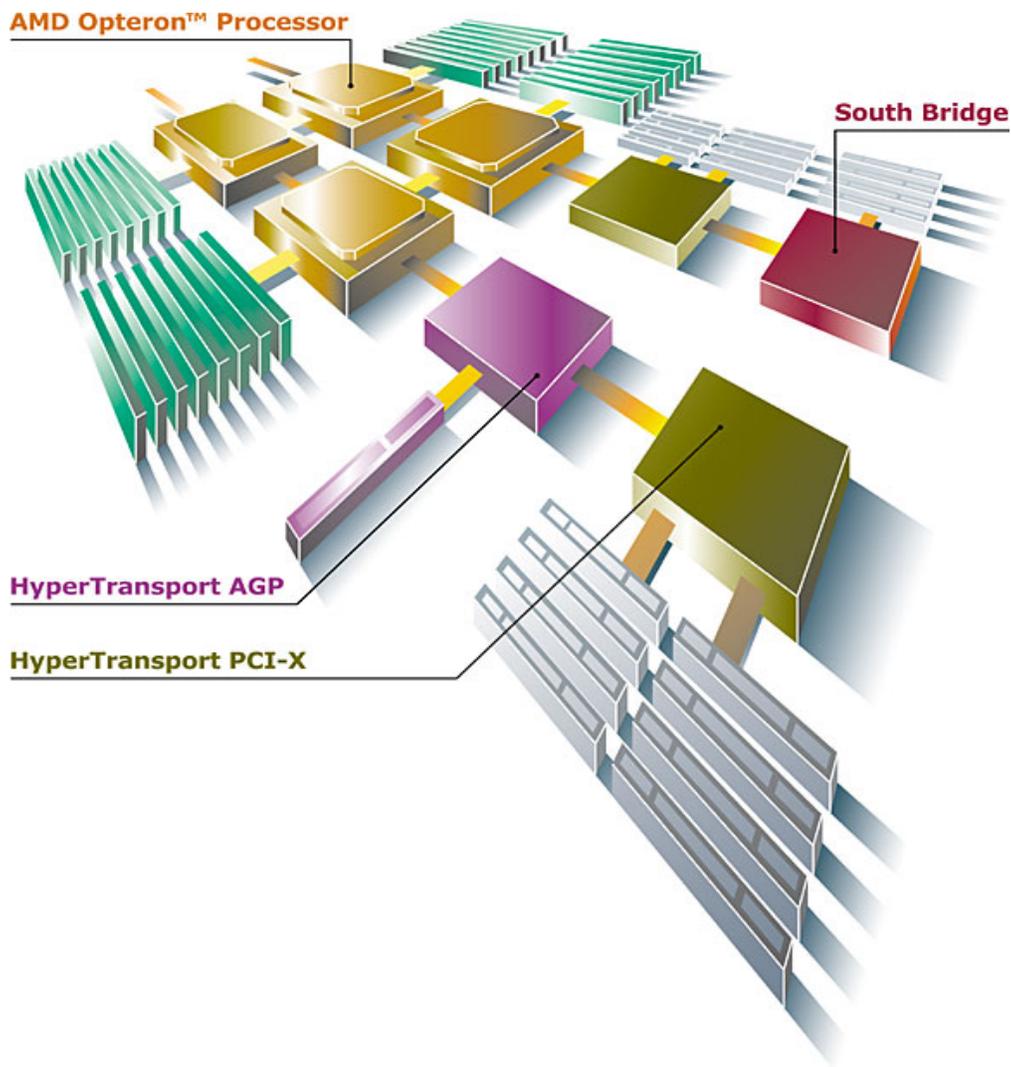


Figure 3: Quad-Processor Platform Based on the AMD Opteron™ Processor.

Integrated DDR DRAM Memory Controller

Memory bandwidth between the system memory and the processor core has been one of the greatest limiting factors in performance with regards to the real-time editing of high-resolution video and audio. The AMD Opteron processor and AMD Athlon™ processor based on Hammer technology directly address this bottleneck by integrating a memory controller into the processor, completely changing and revolutionizing the method for the way x86-based processors access main memory. By attaching memory directly to the processor, moving the memory controller from the Northbridge to reside

directly on the processor, and eliminating the front side bus altogether, the processor can benefit from increased memory bandwidth with an overall reduced latency.

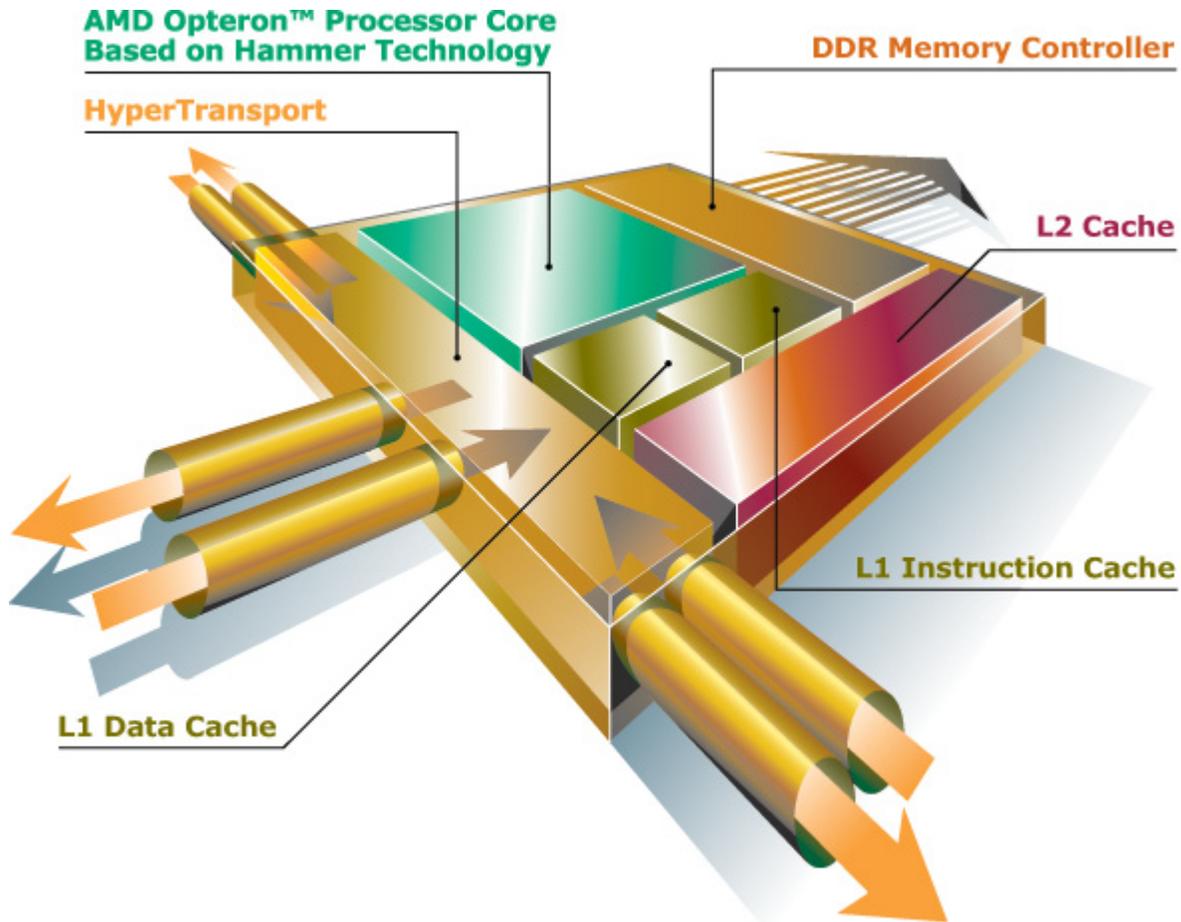


Figure 4: Functional Unit Diagram of an AMD Processor Based on Hammer Technology.

AMD processors based on Hammer technology may incorporate a dual-channel DDR DRAM controller with a 128-bit interface capable of supporting up to eight DDR DIMMs (four per channel). When used in conjunction with PC2700 memory, rated at speeds of 333MHz, the available memory bandwidth available to the processor becomes equivalently 5.3 gigabytes per second. Since the memory controller is now operating at the same gigahertz speeds as the processor, as processor frequency scales, the latency is further reduced.

But it doesn't stop there. The integrated memory controller delivers even more scalability in multiprocessor designs. Taking the example above with PC2700 memory but this time within a four-processor multiprocessing system based on the AMD Opteron processor and support for up to 32 DIMMs, the overall memory bandwidth is designed to deliver an astonishing 21.3 gigabytes per second of available memory bandwidth. This level of bandwidth would exceed the requirements for the largest of digital video and audio streams.

Summary

HyperTransport technology, thanks to its high-bandwidth, low latency connections, isochronous transport system, and support for concurrent communications, offers the bandwidth and performance necessary for the next generation of digital video and audio. HyperTransport technology provides backwards compatibility for PCI software, drivers, and operating systems, while helping to eliminate bottlenecks and providing the bandwidth necessary for future high-speed chips and interconnects standards.

HyperTransport technology is licensed royalty-free to all members of the HyperTransport Technology Consortium. More information about HyperTransport technology can be found by visiting the consortium's web site at www.hypertransport.org.

AMD Overview

AMD is a global supplier of integrated circuits for the personal and networked computer and communications markets with manufacturing facilities in the United States, Europe, and Asia. AMD produces microprocessors, flash memory devices, and support circuitry for communications and networking applications. Founded in 1969 and based in Sunnyvale, California, AMD had revenues of \$3.9 billion in 2001. (NYSE: AMD).

AMD, the AMD Arrow logo, AMD Opteron, AMD Athlon and combinations thereof, are trademarks of Advanced Micro Devices, Inc. HyperTransport is a licensed trademark of the HyperTransport Technology Consortium. NVIDIA is a registered trademark and nForce is

a trademark of NVIDIA Corporation. Other names used in this publication are for identification purposes only and may be trademarks of their respective companies.