Technical Brief

NVIDIA nForce
Platform Processing Architecture

NVIDIA

I. Bottlenecks in PC Design

The PC as we know it today is changing faster than at any other time in its history. But, while CPU and GPU speeds, and memory and storage capacity have increased exponentially over the last 15 years, basic PC architecture has evolved more slowly. In fact, core-logic-based PC architecture has remained largely unchanged from the days when PCs were used only for the most mundane of tasks such as document processing, spreadsheets and e-mail. To address the needs of today's most demanding users—all of whom expect vastly improved system performance, high-speed networking, rich streaming media, Dolby™ Digital 5.1 audio, and high-performance graphics, NVIDIA® has created the new nForce™ Platform Processing Architecture

II. NVIDIA's nForce Platform Processing Architecture

To bridge the gap between the expectations of today's users and the limitations of current and future technologies, and to deliver unmatched system performance, it was necessary for NVIDIA to depart from the traditional "Northbridge/Southbridge" chipset architectures to create an entirely new class of PC. The NVIDIA nForce Platform Processing Architecture consists of three major innovations:

- The revolutionary system architecture has been designed around a distributed processing platform, freeing up the CPU to perform other important tasks. The system integrates NVIDIA's patent-pending system, memory, and networking technologies to improve processing efficiency and overall performance; and its balanced memory design provides the entire platform architecture with the fastest throughput possible.
- The integrated Graphics Processing Unit (GPU) and Audio Processing Unit (APU) provide unparalleled 3D graphics and audio
- Finally, by also incorporating the most complete suite of networking and communications technologies, including 10/100Base-T Ethernet and home phone-line networking (HomePNA 2.0), the NVIDIA nForce Platform Processing Architecture is the fastest, most robust and feature-rich PC platform available for the masses.

To reduce overall system latency and redefine the PC experience, NVIDIA's nForce Platform Processing Architecture is comprised of two platform processors: the nForce™ Integrated Graphics Processor (IGP) and the nForce™ Media and Communications Processor (MCP). For unmatched system performance, the nForce IGP features the TwinBank™ optimized 128-bit memory architecture providing the highest possible bandwidth; a dynamic adaptive speculative pre-processor (DASP™) for boosting CPU performance; and an integrated GeForce2™ GPU for unparalleled 3D graphics performance. The nForce MCP integrates an Audio Processing Unit (APU) with a Dolby Digital 5.1 real-time encoder; StreamThru™, enhanced data streaming technology for superior broadband and networking performance; and the industry's most complete media and communications suite, including

support for HomePNA 2.0, 10/100 Ethernet and USB. Both the nForce IGP and nForce MCP also feature built-in support for AMD®'s HyperTransport™ interconnect technology, delivering the highest continuous throughput between the two platform processors.

nForce IGP Overview

The nForce IGP redefines system and graphics performance. The core of the nForce IGP is comprised of the TwinBank Memory Architecture, NVIDIA's revolutionary 128-bit memory architecture that provides the highest memory bandwidth possible, maximizing memory efficiency so users can run multiple applications simultaneously. The addition of a dynamic adaptive speculative pre-processor (DASP) helps boost CPU and system performance; and the integration of NVIDIA's award-winning GeForce2 GPU ensure an uncompromising 3D visual experience. AMD's HyperTransport, a state-of-the-art bus interface, rounds out the IGP's core technologies, all of which are primed to make the system and graphics performance as fast as possible:

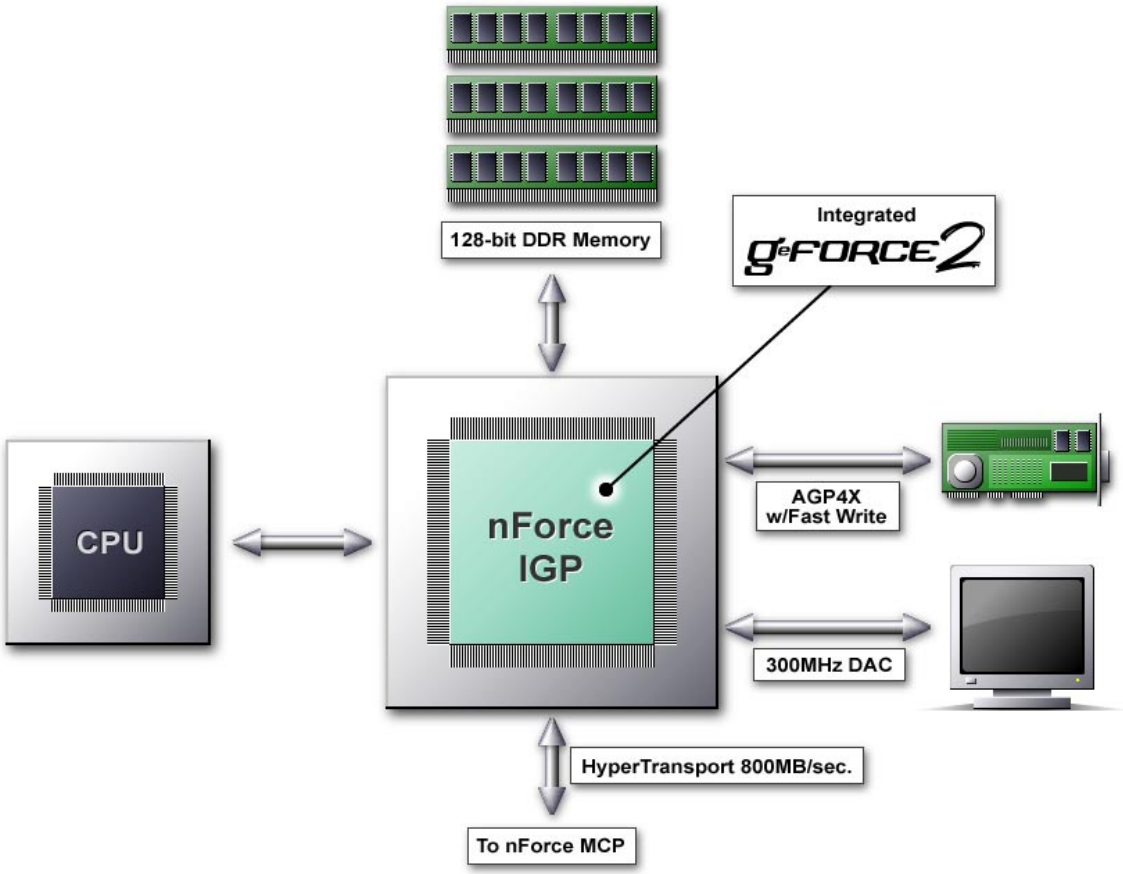

Figure 1: An illustrated overview of NVIDIA's IGP processor

nForce MCP Overview

Redefining the audio and communications experience, the nForce MCP features an integrated APU bringing unprecedented 3D positional audio and DirectX® 8.0-compatible performance to the PC platform, while providing real-time processing of up to 256-simultaneous stereo audio streams, along with a Dolby Digital real-time encoder. On the communications front is StreamThru, an innovative technology pairing that provides an optimized pipeline to enhance networking and broadband; and the most complete suite of networking and communication devices including Ethernet, home phone-line networking (HomePNA 2.0), USB, and dial-up connections. The nForce MCP also features HyperTransport for internal platform processor communications.

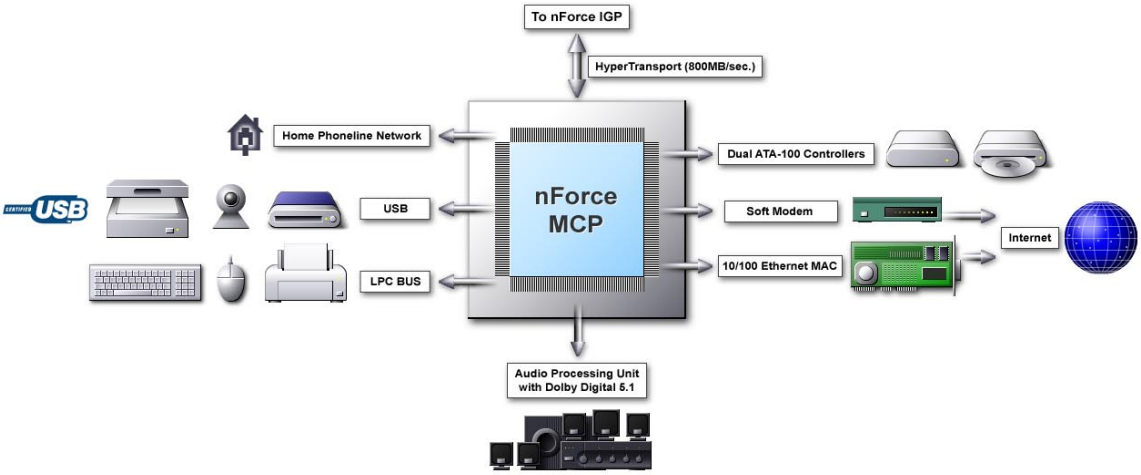

Figure 2: An illustrated overview of NVIDIA's MCP processor

To reiterate, the nForce architecture has underlying technologies that substantially increase overall system performance. They include:

- A revolutionary system architecture that's designed around a distributed processing platform, freeing up the CPU to perform other important tasks.
- TwinBank, a 128-bit, dual-channel memory architecture to eliminate memory bottlenecks
- DASP, a Dynamic Adaptive Speculative Pre-Processor to increase CPU performance
- HyperTransport, a high-speed interconnect between the IGP and MCP
- StreamThru, NVIDIA's patent-pending isochronous data transport system, providing uninterrupted data streaming for superior networking and broadband experiences.

III. Revolutionary New System Technologies

Distributed Parallel Processing

Although current CPUs have significant processing capability, they can only process instructions in a serial fashion. Yesterday's platform architecture relied on the CPU to perform all of the processing intensive tasks (see fig. 3). The example below illustrates the typical core-logic chipset that acts as a "simple" input/output (I/O) hub for the CPU, system memory, graphics accelerator *and* mastering peripherals connected to the "Southbridge".

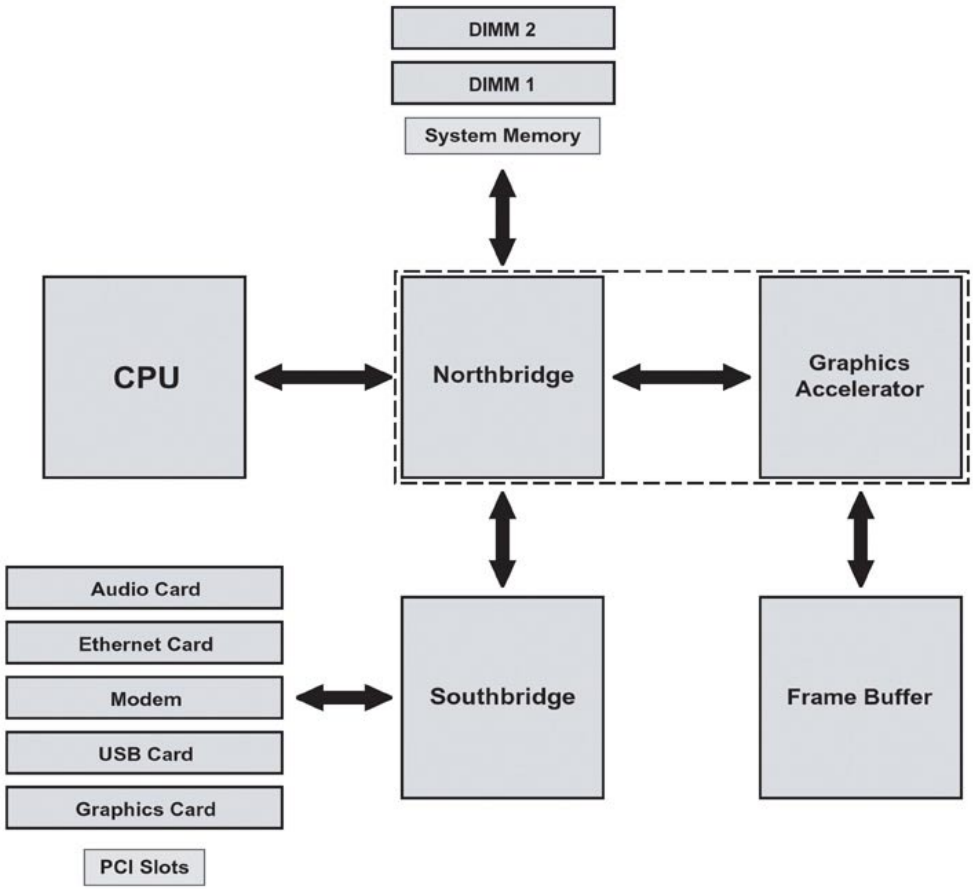

Figure 3: Yesterday's Platform Architecture

Key issues restricting the highest potential system performance include:

1. The graphics accelerators reliance on the CPU to set up calculations and compute geometry values.
2. The sound card's reliance on the CPU to process simultaneous audio streams (voices), 3D positional computation (HRTF's, elevation, direction, etc.) and various effects (chorus, reverb, obstruction, occlusion, etc.).

3. The CPU's responsibility for the other various instructions related to the applications that are running and the mastering peripherals connected to the Southbridge.

The nForce Platform Processing Architecture however, was designed to offload processing-intensive tasks from the CPU onto the platform processors (nForce IGP and nForce MCP) in order to increase the system's overall processing power. In contrast to yesterday's platform architectures, nForce distributes the processing load across the CPU, IGP, and MCP in parallel (see fig. 4), resulting in much more efficient system performance.

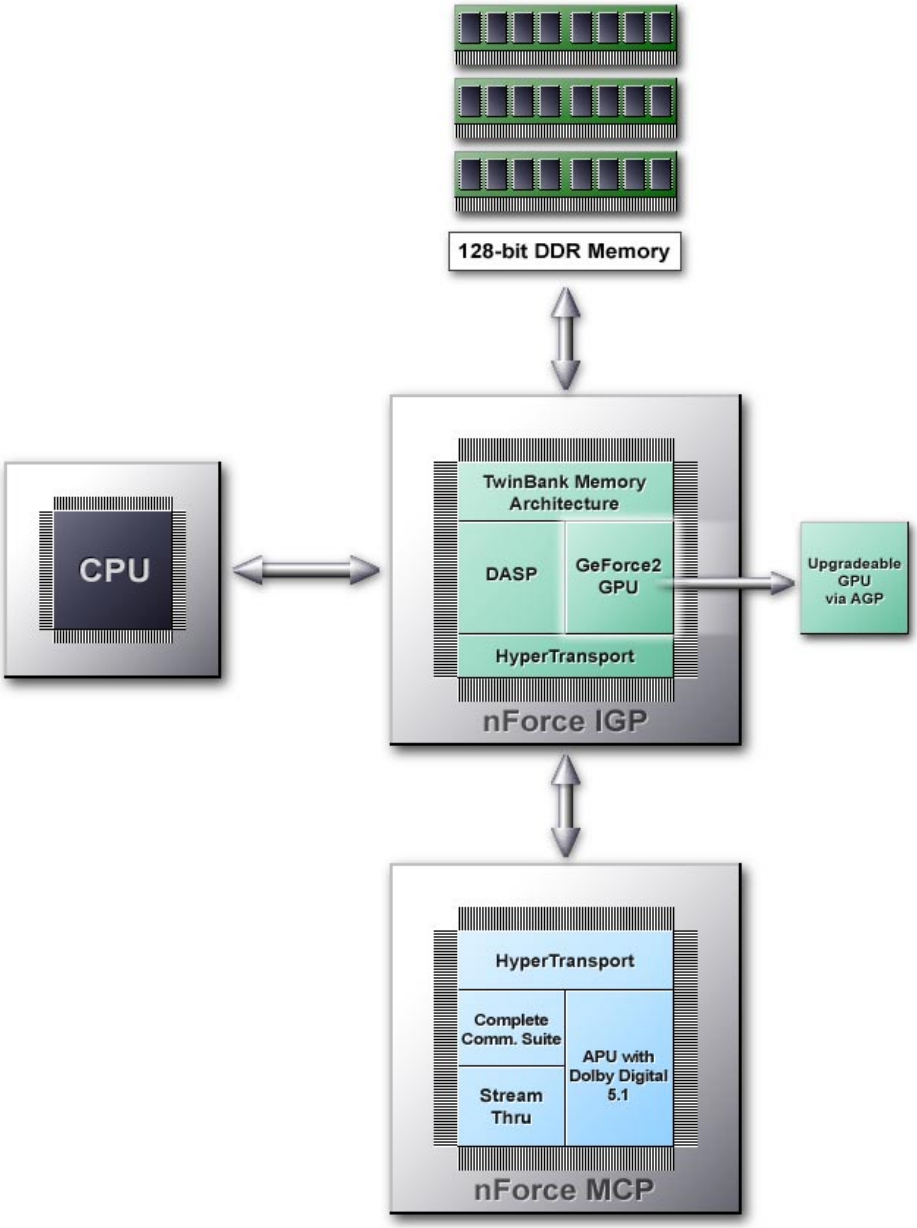

Figure 4: Today's Distributed Platform Architecture

With NVIDIA's nForce Platform Processing Architecture the highest potential system performance can be achieved because:

1. The Graphics Processing Unit (GPU) integrated into the IGP, or external via AGP4X, processes equation setup and geometry and performs rendering functions.
2. 3D positional audio, (up to and including Dolby Digital 5.1), is processed by the Audio Processing Unit (an APU is integrated into the nForce MCP).
3. The CPU is now freed up to perform other functions.
4. All processing occurs concurrently and in parallel.

TwinBank Memory Architecture

Lack of memory bandwidth also contributes to poor system and graphics performance. Mainstream "core-logic chipsets" utilize a 64-bit Single Data Rate SDRAM (SDR SDRAM) memory controller that provides only 1.06GB/s of memory bandwidth. The CPU and various masters in the "Southbridge" (audio, Ethernet, SCSI, IDE, etc.) must all compete with each other for this bandwidth resulting in slower system level performance due to increased latency and context-switching overhead.

"Core-logic chipsets" with integrated graphics accelerators further amplify this problem by allocating frame buffer memory within the system memory. This is commonly referred to as a Shared Memory Architecture (SMA). In this type of architecture, the integrated graphics accelerator must fight with the CPU and various masters for memory bandwidth through multi-step memory access arbitration.

The nForce IGP's TwinBank Memory Architecture eliminates system memory as a bottleneck by providing a 128-bit wide DDR 266MHz memory access path. This is implemented through dual-independent, 64-bit memory controllers, backed by a *single* master arbiter. The end result: 4.2GB/s peak memory bandwidth with minimum system latency. TwinBank's radical crossbar memory controller enables the CPU and GPU to access the two 64-bit memory banks concurrently, fully utilizing memory bandwidth.

For more information on the TwinBank Memory Architecture, please read NVIDIA's nForce IGP TwinBank 128-bit DDR/SDR Memory Architecture technical brief.

Dynamic Adaptive Speculative Pre-Processor (DASP)

Another revolutionary innovation introduced in NVIDIA's Platform Processing Architecture is the integration of the IGP's dynamic adaptive speculative pre-processor (DASP). The DASP increases performance by intelligently predicting possible memory accesses based on historical CPU memory access patterns, and storing them in an on-die buffer for faster retrieval by the CPU. Because the CPU can access the buffer faster than main system memory, overall system performance is greatly improved.

For more information on DASP, please read NVIDIA's nForce IGP DASP technical brief.

HyperTransport Technology

Current "core-logic chipsets" provide a maximum 266MB/sec. peak bandwidth throughput between the "Northbridge and Southbridge" at best. Unfortunately, the increased demands of PC power users have already out-paced this bandwidth capability, rendering it a platform bottleneck and severely impacting overall system performance (see fig. 5).

Device	Bandwidth Consumption
PCI Bridge	133MB/sec.
Dual ATA-100 Disk Controllers	200MB/sec.
Ethernet MAC (Full Duplex)	25MB/sec.
Dual USB Controllers	3MB/sec.
Audio Processing Unit	150MB/sec.
Audio Codec Interface	1MB/sec.
Legacy (LPC)	1MB/sec.
Total	***513MB/sec***

Figure 5: Today's Bandwidth Consumption Is More Than What Current Core-Logic Chipsets Can Provide!

With next-generation platforms the bandwidth demand will be even greater, further burdening system throughput and resulting in even more severe system bottlenecks.

AMD's HyperTransport™ technology, integrated onto both the nForce IGP and nForce MCP, improves overall system level performance by eliminating this bottleneck by providing a robust **800MB/sec.** of throughput between the two platform processors, almost 16 times more peak bandwidth than systems utilizing a 133MHz PCI bus. And, because HyperTransport™ technology is isochronous by design, time-dependent applications such as streaming video or audio can be processed seamlessly and without interruption.

StreamThru Isochronous Data Transport System

StreamThru is NVIDIA's patent-pending isochronous data transport system, providing uninterrupted data streaming for superior networking and broadband communications. By pairing the integrated 10/100Base-T Ethernet controller to an isochronous-aware internal bus, along with a single-step arbiter, StreamThru assists in making streaming video and audio smoother and jitter-free. For the first time, users can enjoy maximum efficiency from their broadband connection delivering full multimedia performance without compromise. Systems enabled with StreamThru will experience up to a 15% performance boost in these areas, depending on the applications running.

For more information on StreamThru, please read NVIDIA's nForce MCP StreamThru technical brief.

IV. Other Integrated Technologies

Graphics Processing Unit (GPU)

Leveraging its success in the graphics world, NVIDIA's nForce Platform Processing Architecture is the only one in this class to integrate the 3D graphics power of the award-winning GeForce2 GPU. With its second-generation transform and lighting capability, per-pixel shading operations, a fill-rate of up to 350M pixels/second, and an internal 8X AGP interface, the performance of the integrated GeForce2 greatly surpasses the quality and performance of all other integrated graphics solutions. More importantly, using an integrated GPU allows complex graphics calculations to be performed directly on the GPU itself. So, instead of having to communicate back and forth across the AGP bus, the CPU is freed up to perform other tasks, resulting in much more efficient—and faster—system performance. And, because the entire platform architecture is scalable, power users have the ability to override the onboard graphics chip to install even more powerful NVIDIA GPUs, such as the GeForce 3.

For more information on NVIDIA GPUs, please read the GeForce2- or GeForce3-related technical briefs.

Audio Processing Unit (APU)

The integrated Audio Processing Unit (APU) performs the world's most advanced 3D positional audio functions—and is the same NVIDIA audio technology that powers Microsoft's XBox™ game console. The APU is fully compliant with Microsoft's DirectX 8.0, and provides real-time processing of up to 256-simultaneous stereo audio streams, or 64 3D and 192 simultaneous audio streams. Furthermore, the APU can output the 3D audio streams to 2, 4 or 6 speakers. The APU is also the only to feature Dolby's new Dolby Digital Interactive Content Encoder, providing interactive 3D positional audio and Dolby Digital 5.1 channel audio for a true cinematic experience.

For more information on the APU, please read NVIDIA's nForce MCP APU technical brief.

Full Communications Suite

In addition to the audio functionality, the nForce MCP also features a full complement of communications and network connections. With an integrated 10/100Base-T Ethernet controller on-board, NVIDIA's Platform Processing Architecture is primed for broadband and network connectivity. For inexpensive home networking, there's also integrated support for HomePNA 1.0/2.0. Additional features include an integrated 56K soft modem and support for up to six USB ports and other PC peripherals, such as scanners, fingerprinting devices, digital cameras, and MP3 players.

V. Conclusion

NVIDIA's nForce Platform Processing Architecture redefines the PC platform and addresses the needs of today's most demanding users—all of whom expect vastly improved system performance, high-speed networking, rich streaming media, and Dolby Digital 5.1 audio.

Its key innovations include:

- A revolutionary, two platform processor (nForce IGP and nForce MCP) system design that distributes processing tasks across the CPU, nForce IGP and nForce MCP
- The TwinBank Memory Architecture to reduce system and memory bottlenecks technologies
- Core system technologies, including TwinBank and DASP, that help enhance overall CPU performance
- An integrated GPU for unmatched 3D graphics
- An integrated APU providing unheard 3D effects and Dolby Digital 5.1 audio
- StreamThru technology for improved networking and broadband experiences

Only with NVIDIA's nForce Platform Processing Architecture will system performance no longer be constrained by outdated core-logic chipsets. System throughput is no longer bound by outdated PCI I/O interfaces, nor is it inhibited by "simple" I/O hubs. Instead, overall system performance, including graphics, audio, networking, communications, and memory management, is enhanced by an intelligent platform design and integrated, high-performing platform processors. In short, NVIDIA's nForce Platform Processing Architecture will be responsible for transitioning the world into the truly digital, 21st century.

© 2001 NVIDIA Corporation

NVIDIA, the NVIDIA logo, nForce, TwinBank, GeForce2, and StreamThru are registered trademarks or trademarks of NVIDIA Corporation. Other company and product names may be trademarks or registered trademarks of the respective companies with which they are associated.