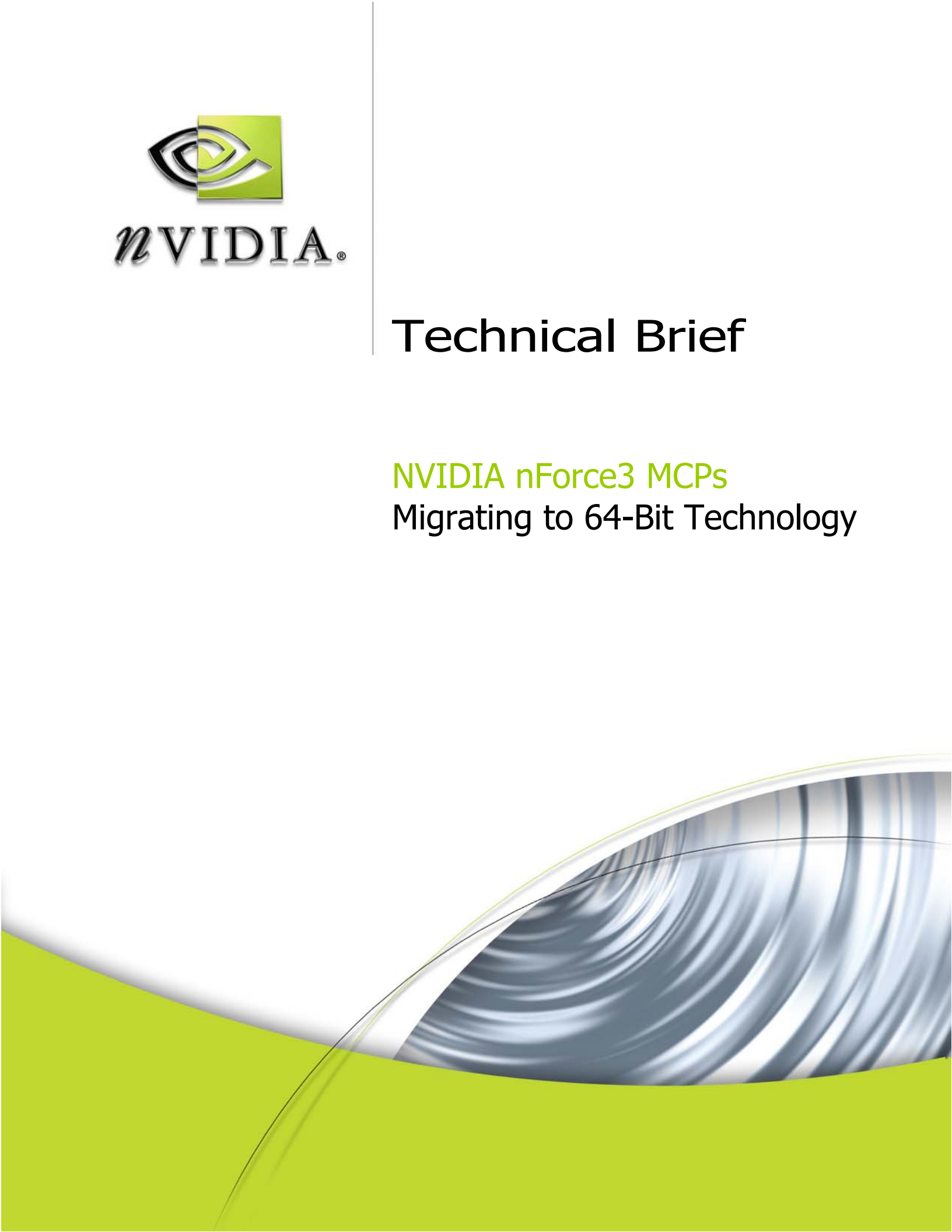

Technical Brief

NVIDIA nForce3 MCPs
Migrating to 64-Bit Technology

The Move to 64-Bit PC Architectures and Related Core Technologies

Bigger file systems, bigger models, bigger scenes—desktop systems handle bigger computing challenges every day. Designing "bigness" into the system architecture requires changes from the ground up. The proprietary world of UNIX on RISC has already made this leap forward by architecting systems with 64-bit-wide paths for memory addressing. And now, workstation and desktop PC architectures are evolving to incorporate this bigger computing paradigm.

Such a major evolution in architecture will affect every component within a system platform, including operating systems and applications. The NVIDIA nForce™3 media and communications processors (MCPs) tackle this challenge and provide full compatibility with the new 64-bit AMD Opteron™ and Athlon™64 processor architecture. In addition, they provide investment protection for existing 32-bit applications.

The third-generation NVIDIA nForce core technology maintains the NVIDIA tradition of providing users with the most robust desktop experience available. Whether users continue to use 32-bit operating environments, move to the new 64-bit environments, or chose to migrate over time, NVIDIA nForce3 MCPs will optimize existing 32-bit applications and provide users with a platform that can take advantage of 64-bit technology right from the start.

64-Bit Architectures

Today's traditional system architectures use 32-bit memory addressing. With this design, the CPU can access a data space that spans up to 4GB of virtual memory, typically split into two partitions: 2GB for the operating system and 2GB for the active application. In emerging 64-bit architectures, virtual memory addresses can be specified using the full 64 bits. This extends the addressable memory range well beyond current physical memory limitations.

Several categories of applications require and benefit from 64-bit virtual memory designs. All must be able to manipulate large data sets:

- Computer-Aided Design (CAD)
  Large engineering models can involve more than tens of thousands of discreet components. Each part requires multiple "words" of data to describe its hierarchy, attributes, and requirements. The 2GB limit for virtual memory restricts the PC's ability to process such large models. The availability of 64-bit system architectures introduces lower-cost platforms for these applications.

- Finance
  The 32-bit limitation affects this market in two dimensions. Applications that involve millions of transactions a day are limited in their ability to analyze large numbers of transactions collectively. In addition, high-precision analysis cannot be carried out for large data sets because accuracy requires multiple words to represent each floating-point number. The 64-bit architectures allow more in-depth analysis for large numbers of transactions, and offer increased accuracy. This can add up to significant savings when large amounts of securities are traded online.
- Digital Content Creation (DCC)
  Digital scenes and virtual worlds quickly consume virtual memory as more sophisticated textures, lighting, and effects are introduced. The 64-bit architectures remove restrictions on the content creators and allow scenes to integrate unlimited complexity.
- Games and Other Applications
  Games and other CPU-intensive applications written specifically for a 64-bit platform run faster and handle larger texture files more efficiently.

Migration Challenges

Previous attempts at practical 64-bit architectures have presented prohibitive challenges in the migration from 32-bit architectures. Performance, cost, and software compatibility issues have made the move from 32-bit to 64-bit impractical for many users. Not every application requires an increase in addressable virtual memory. Compared to previous 32-bit architectures, some 64-bit architectures impose a performance penalty when implementing 64-bit addressing. Most applications spend their time moving data around in memory, and performance degradation for 32-bit applications would be unacceptable to any user.

Likewise, some architectures introduce 64-bit addressing by doubling the amount of CPU hardware. The increase in cost is not acceptable to users who require more mainstream 32-bit capabilities. A third challenge—software compatibility—complicates the move to 64 bits. Operating systems must be adapted to incorporate 64-bit virtual memory capabilities, and the system and operating system combined must be able to recognize and run both 32-bit and 64-bit applications appropriately.

With its next-generation Opteron and Athlon 64 processors, AMD implements a 64-bit strategy that supports emerging 64-bit systems and applications—plus the installed base of 32-bit applications and operating systems—without adversely affecting performance, cost, or compatibility.

The AMD 64-Bit Strategy

AMD innovatively supports both 32-bit and 64-bit software. In doing so, AMD provides a painless transition for establishing an installed base of 64-bit applications while protecting investments in current 32-bit solutions.

This feat is accomplished by evolving the x86 architecture to support three modes of execution:

- Legacy mode, a 32-bit operating system with 16-bit and 32-bit applications
- 32-bit compatibility mode, a 64-bit operating system with legacy (16-bit and 32-bit) applications
- 64-bit mode, a 64-bit operating system with 64-bit compatible applications

The NVIDIA nForce3 solutions complement the newest AMD 64-bit processors, providing optimized operation in all three modes. System designers can take advantage of the AMD/NVIDIA combination today to build optimized 32-bit compatible systems for today that also get the most out of the 64-bit operating systems and applications of tomorrow.

Compatibility and 64-Bit Modes

The AMD 64-bit processors extend the x86 architecture by adding a *long mode*. When long mode is disabled, the processor operates as a standard x86 processor, compatible with all existing 16-bit and 32-bit operating systems and applications. When long mode is active, the 64-bit processor extensions are enabled. This allows the system to automatically configure according to software requirements.

Long mode encompasses two submodes of operation: *64-bit mode* and *compatibility mode*. Both submodes require a 64-bit operating system. Under the 64-bit operating system, compatibility mode supports binary compatibility with existing 16-bit and 32-bit applications. And 64-bit mode introduces the ability to execute 64-bit applications and take advantage of the new AMD 64-bit capabilities, including

- 64-bit virtual addresses
- Register extensions: Existing registers are widened to 64 bits; eight new registers are added; eight streaming Single Instruction Multiple Data (SIMD) extension registers are added
- 64-bit instruction pointer: Routing Information Protocol (RIP)
- New RIP-relative data addressing mode
- Flat address space with single code, data, and stack space

Legacy Mode

When long mode is disabled, the AMD 64-bit processors can support a 32-bit operating system and 16- or 32-bit applications, resulting in a legacy mode of operation. Essentially, the new AMD 64-bit processors allow a single-system design to support a range of operating system and application combinations. This capability lets users choose the path to 64 bits that best suits their needs and optimally protects their software investments.

NVIDIA nForce3 Technology

The new 64-bit architectures represent a major advancement for applications that manage large and complex data sets. But with 64-bit designs, the efficient movement of data throughout the system becomes even more critical if 32-bit applications are to run well and if 64-bit applications are to run without creating new bottlenecks. The need for the NVIDIA nForce3 processing capabilities becomes vital if users are to get the most out of the newest 64-bit processing platforms.

NVIDIA nForce3 media and communications processors (MCPs) offer many innovations for the emerging 64-bit world:

- The First Single-Chip Solution for 64-Bit Platforms
  The reduction of chip-to-chip interconnects results in lower overall system-level latencies. The combining of functions from two chips into one frees board space for future performance-enhancing silicon solutions and hardware-implemented capabilities.
- Advanced Technology Solutions
  The third-generation NVIDIA nForce solutions deliver built-in features for the latest network access technologies (Gigabit Ethernet), and fault-tolerant RAID 1 storage designs for parallel and serial ATA devices. In addition, they offer advanced I/O capabilities.
- Unmatched Performance
  The integrated networking solutions offer both high throughput and low CPU overhead, and the storage technology incorporates high-performance RAID 0 disk striping techniques. RAID 0+1 support allows high-performance, fault-tolerant configurations that combine striping and mirroring techniques. NVIDIA nForce3 solutions also integrate high-speed bus designs for on-chip datapaths (an optimized third-generation NVIDIA design). Plus, they integrate high-speed designs for front-edge connections (based on the HyperTransport technology for up to an unprecedented 6.4GB/sec. transfer rate to and from the CPU).
- 64-Bit Software
  The NVIDIA engineering expertise and in-depth experience working with device drivers and system-level software technologies result in the highest quality software solutions accompanying the nForce3 MCPs.

Conclusion

Combined with the AMD 64-bit strategy, the NVIDIA nForce3 MCPs provide the best path into the 64-bit world. All the NVIDIA innovations contribute to solid performance and functional benefits for today's applications, while laying a solid foundation for the emerging 64-bit operating environments and applications.

Today more than ever, the proven NVIDIA track record demonstrates the Company's unique ability to smoothly transition customers to evolving technologies.

NVIDIA wins in any competitive comparison involving

- The ability to get the **most performance** out of the AMD processor architectures (optimizing data handling for core functions)
- Solution **reliability**
- **Application and software compatibility**, both backwards and forwards

Notice

ALL NVIDIA DESIGN SPECIFICATIONS, REFERENCE BOARDS, FILES, DRAWINGS, DIAGNOSTICS, LISTS, AND OTHER DOCUMENTS (TOGETHER AND SEPARATELY, "MATERIALS") ARE BEING PROVIDED "AS IS." NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY OR OTHERWISE WITH RESPECT TO THE MATERIALS, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NONINFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE.

Information furnished is believed to be accurate and reliable. However, NVIDIA Corporation assumes no responsibility for the consequences of use of such information or for any infringement of patents or other rights of third parties that may result from its use. No license is granted by implication or otherwise under any patent or patent rights of NVIDIA Corporation. Specifications mentioned in this publication are subject to change without notice. This publication supersedes and replaces all information previously supplied. NVIDIA Corporation products are not authorized for use as critical components in life support devices or systems without express written approval of NVIDIA Corporation.

Trademarks

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

Other company and product names may be trademarks of the respective companies with which they are associated.

Copyright

©2003 NVIDIA Corporation. All rights reserved.

NVIDIA Corporation
2701 San Tomas Expressway
Santa Clara, CA 95050
www.nvidia.com