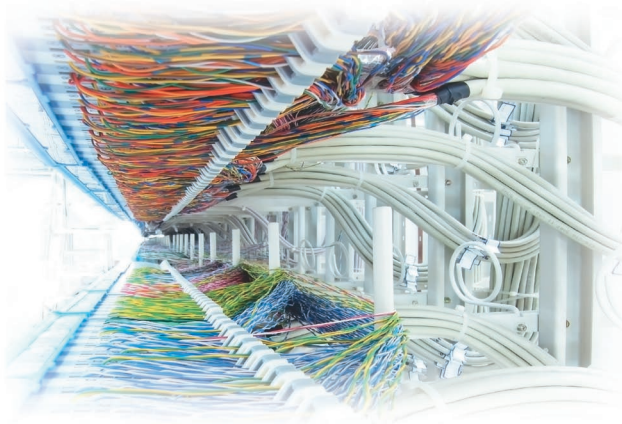
Networks on Processors Improve On-Chip Communications



David Geer

As microprocessors have become more complex, on-chip communications, traditionally handled by buses, have become a performance bottleneck. Chip makers are looking into network-on-chip systems to address this issue.

The microprocessor has been a critical engine for computer technology's advancement. As processors have become more complex, though, a critical performance-related issue has become on-chip communications.

Basically, a microprocessor connects a computational engine to a memory system, typically via a bus. However, as chip technology has advanced, it has used multiple topologies and added dedicated functional units such as application-specific integrated circuits (ASICs) to the basic computational engine, creating systems-on-chip (SoCs).

These developments have complicated on-chip communications and made it difficult for buses to maintain performance at a reasonable cost. They have also hurt design productivity.

Researchers and chip designers have thus begun considering replacing buses with network-on-chip approaches.

An NoC connects a network of computational engines, distributed storage, and programmable I/O via a network of routers or switches that communicate via addressed data packets.

Unlike buses, NoC proponents say, the networking approach enables reliable, robust on-chip communications on complex processors, even in the face of problems.

Several manufacturers are already using NoC technology with their chips. There are also ongoing research projects.

Nonetheless, the technology must overcome several challenges before it is ready for widespread use in multiple settings.

DRIVING FORCES

Single-chip embedded systems such as IBM's Cell and those used in high-definition TVs are becoming increasingly complex.

Such SoCs include numerous cores that perform distinct functions—such as digital-signal and graphics processing—and that operate at different clock frequencies.

This complicates on-chip communications. For example, as the number of cores on a processor increases, the number of potential communications paths between them rises exponentially.

And with more components, a single bus must address communications over a larger area, which leads to latency, according to David Gwilt, interconnects-products manager for chip maker ARM.

Also, as processors' feature sizes shrink and they contain more transistors and cores, interconnect technology more profoundly affects chip performance and power usage.

Therefore, improving on-chip communications technology has become increasingly important to researchers and processor manufacturers.

0018-9162/09/$25.00 © 2009 IEEE Published by the IEEE Computer Society

Scalability

Bus technology doesn't scale well enough to provide high performance and effective on-chip communications at a reasonable cost on complex processors, said Carlo Pistritto, STMicroelectronics' on-chip communication system R&D director.

Typically, chip makers have tried to segment buses for use with the different elements on SoCs. This process turns buses from being long wires that are globally clocked and stretch the entire length of the chip to being separate groups of locally clocked, bundled wires connected to bridges at each end.

However, ARM's Gwilt said, segmenting buses is difficult for complex SoCs and multicore chips. Buses must be manually designed with different segments that suit specific chip architectures, with their different core placements and configurations. This can be an expensive and time-consuming process, he explained.

Technical issues

As the number of on-chip processing and storage arrays increases, the wires on the processor get closer to one another. This situation causes parasitic capacitance, which degrades bus performance.

Parasitic capacitance is the unwanted storage of an electrical charge between closely packed parallel wires on long buses. This unintended storage of charge can lead to bus faults, explained Arizona State University associate professor Karam Chatha.

With arbitration complexity, a bus has trouble directing all the traffic going to and from the many on-chip cores, which slows performance.

Design productivity gap

The development of increasingly smaller transistors enables chip makers to put more of them on processors. However, figuring out how to do this with efficient communications between the transistors is increasingly difficult and time-consuming for designers. Buses are not easily scalable and thus complicate this problem.

NOC IN A NUTSHELL

Philips Electronics' researchers conducted the first NoC work in late 2001, according to Kees Goossens, NXP Semiconductors' principal research scientist for SoC architectures and infrastructure.

Chip designers have begun considering replacing buses with network-on-chip approaches.

How it works

Some high-performance computing and data-networking processors could benefit from circuit-switching NoC techniques, according to Goossens. The technique's use of dedicated circuits for communications increases bandwidth efficiency.

However, the approach doesn't provide routing flexibility and thus doesn't avoid network congestion well. Therefore, today's NoCs use packet switching.

Packet-switched systems. NoC systems use packet switching over multiplexed links. Network interface units manage the packetization of native intercore data. The system then uses routers or, much less frequently, switches to connect an SOC's various processing cores and other elements via copper wiring, as Figure 1 shows.

As in telecommunications systems, NoC technology uses routing algorithms and tables to decide the optimal way to send packets, according to Goossens.

With NoCs, the algorithms should entail simple logic to avoid hurting host-chip performance and consuming excessive power, according to Gwilt.

To meet the requirements of on-chip communications and to function within host processors' limitations, the routers must be small, energy-efficient, and fast.

Networking techniques. On-chip communications must be very fast, latency-free, and flexible. Thus, the networking techniques must be simple. The trade-off is that simple networks offer fewer capabilities.

Not all traditional network technologies work with NoCs. For example, TCP/IP entails too much latency.

The Open Core Protocol International Partnership thus developed the OCP standard for on-chip communications, with which some NoC vendors work.

Some companies use proprietary approaches. For example, chip designer Arteris works with its proprietary NoC Transaction and Transport Protocol, which functions like a universal data-translation protocol for intercore communication, said company president K. Charles Janac.

Architectures and topologies

NoC technology gives chip designers flexibility in choosing the network topology that best fits their processors' needs, according to University of Bologna professor Luca Benini, founder of and scientific advisor for iNoCs, a start-up provider of on-chip interconnection technology.

For example, a mesh NoC topology can function with a high degree of parallelism and thus is well-suited for multiprocessor SoCs, whose cores must run in parallel.

Prototype NoCs by the Electronics and Information Technology Laboratory of the French Atomic Energy Commission's Faust, the Swedish Royal Institute of Technology's Nostrum, and the Technion-Israel Institute of Technology's QNoC work with a mesh topology.

The Korea Advanced Institute of Science and Technology's BONE (Basic On-chip Network) prototype uses a clustered mesh topology.

The Pierre & Marie Curie University's SPIN uses a fat-tree topology.

GALS

According to Benini, most NoCs work with SoCs that have a globally asynchronous, locally synchronous architecture.

With GALS, an SoC is divided into internally synchronous subsystems that have different clock cycles from one another and that thus must communicate using asynchronous techniques.

NoCs enable this communication by synchronizing transmissions between subsystems in the wires.

Advantages

By handling intercore data transfers, an NoC enables processor cores to function without having to be involved in the communications process.

NoCs work with multiple routers or switches connected by shorter wires than buses use. The use of shorter wires makes NoCs more energy-efficient.

And the shorter connections reduce the complexity of designing the wires to yield predictable speed, power, noise, and reliability. NoCs' regular, well-controlled structure of short, bundled wires is much more predictable than buses' single long wires that run all over the chip.

NoCs increase engineering and design productivity and reduce design risk by providing a malleable framework that adjusts to any configuration and topology of chip elements, with the help of available tools. This lets designers shape the NoC topology around the needs of the SoC. Buses don't provide such a framework or tools.

Manufacturing costs. An NoC can reduce SoC manufacturing costs via simulation tools that make sure communications between the chip cores work before fabricating the chips. Buses don't come with such tools.

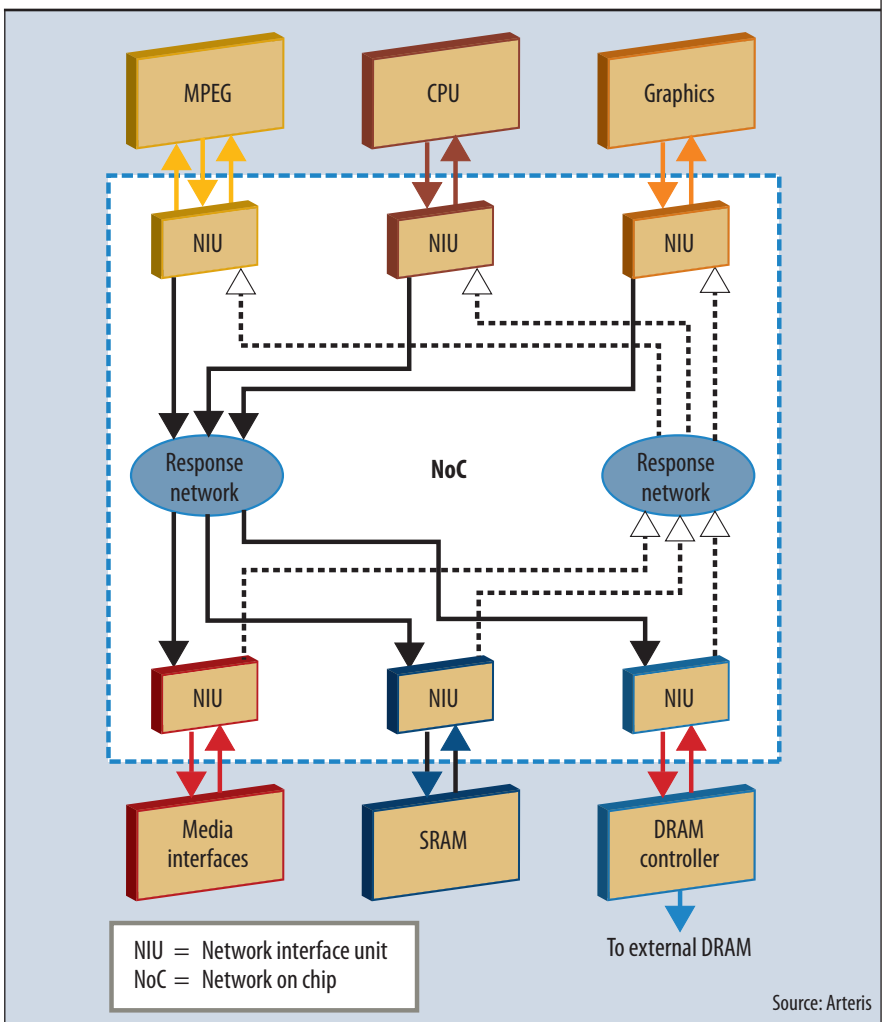


Figure 1. NoC systems apply networking, rather than bus, technology, to communications among the cores on a processor. Network interface units manage the packetization of native intercore data. Routers then connect the various processing cores and other elements.

Performance. NoCs increase SoC performance by removing the communications bottlenecks that buses cause.

And with packet switching, signals from multiple sets of communications share the wires running between routers in an NoC, explained Carnegie Mellon University professor Radu Marculescu. This provides high throughput.

Arteris says its NoC Solution system makes chips perform three times as fast as they would using conventional bus systems.

Scalability. NoC technology is scalable because designers can simply add more networking elements if manufacturers add cores to their chips.

On large SoCs, buses don't scale easily because they are not as adaptable to different on-chip topologies and they create increasing communications bottlenecks.

Time to market. NoCs reduce SoC design time. Vendors provide chips with an open framework and design tools that automate a lot of the process of making an NoC that fits their processor design.

Variability. NoCs conduct error detection, containment, and correc-

tion efficiently. They thus effectively take most errors out of the process of sending signals across the wire, reducing unpredictability.

Commercial implementations

Most chip makers use other companies' NoC approaches, although some utilize their own.

STMicroelectronics designs and produces its own system, VSTNoC, which the company uses in its chips, including those for high-definition TVs.

iNoCs' NoC platform synthesizes on-chip communications networks with topologies tailored to the needs of target SoC platforms, explained the company's Benini.

About 300 million SoCs that use Sonics Inc.'s SMART Interconnects NoC technology appear in laptops, PCs, high-definition TVs, smart phones, gaming consoles, and other products from companies such as Canon, Dell, Motorola, Nokia, Samsung, Sony, and Toshiba, according to Drew Wingard, Sonics' chief technology officer and founder.

NXP is working on its Aethereal Network on Chip for consumer electronics and embedded systems such as set-top boxes, TVs, mobile phones, and automobile applications.

Companies also provide tools that manufacturers can use to incorporate and customize NoCs within their chip designs. For example, Arteris provides assemblers, compilers, component libraries, and traffic analyzers and simulators.

"Texas Instruments is using them in its OMAP4 family of [mobile] processors," said Janac. Pixelworks is also using them in the chips it designs for the advanced-display industry, he added.

Research projects

Several researchers are working on NoC-related projects.

For example, Chatha's Arizona State team is researching NoCs, as well as tools for customizing the systems for specific applications.

Carnegie Mellon's Marculescu is working on SlickNets, a project researching NoCs for multiprocessor SoCs.

THE KNOCK ON NOC

Chip makers don't always recognize the advantages of NoCs, said George Michelogiannakis, deputy director of the Institute of Computer Science at the Foundation for Research and Technology-Hellas.

When they recognize the need, he added, companies often don't have the necessary expertise, so they continue working with bus technology.

In addition, he said, today's NoCs aren't always reliable enough to continue operating if a part of the system fails.

And vendors must be careful to design NoCs efficiently or they can consume a lot of power.

Existing network models are a good start for NoCs, according to Richard Savoie, electrical/software project manager with Boston Engineering, an engineering services firm. However, he said, the process and information necessary to enable intelligent routing creates data-packet overhead that can be burdensome for on-chip communications.

Also, he said, routing algorithms still must be designed to offer more parallelism, to take advantage of NoCs' performance potential.

Added NXP's Goossens, researchers must better determine how to design and use the existing components to maximize NoCs' performance.

And the lack of a standard interface could pose challenges for different types of NoCs when they need to communicate with one another.

Moreover, said Goossens, as SoC vendors begin to address issues like security and virtualization, NoC researchers must do the same.

During the next five years, Goossens predicted, NoC adoption will expand to applications such as symmetric multiprocessors, ASICs, field-programmable gate arrays, and Internet routers.

The technology will become very popular during that time because it will work with chips in widely used devices such as laptops, PCs, TVs, networking equipment, game consoles, and smart phones, according to Sonics' Wingard.

NoC technology will also thrive because its improved communications and scalability are good for the increasingly popular multicore chips, said Carnegie Mellon's Marculescu.

"In five years, we will be using this NoC technology massively," said STMicroelectronics' Pistritto.

For this to occur, though, Goossens said, the technology will have to be standardized and researchers must address the technology's problem areas.

David Geer *is a freelance technology writer based in Ashtabula, Ohio. Contact him at david@geercom.com.*

Editor: Lee Garber, *Computer*, l.garber@computer.org