
The Need for a High-Bandwidth Memory Architecture in Programmable Logic Devices

Introduction

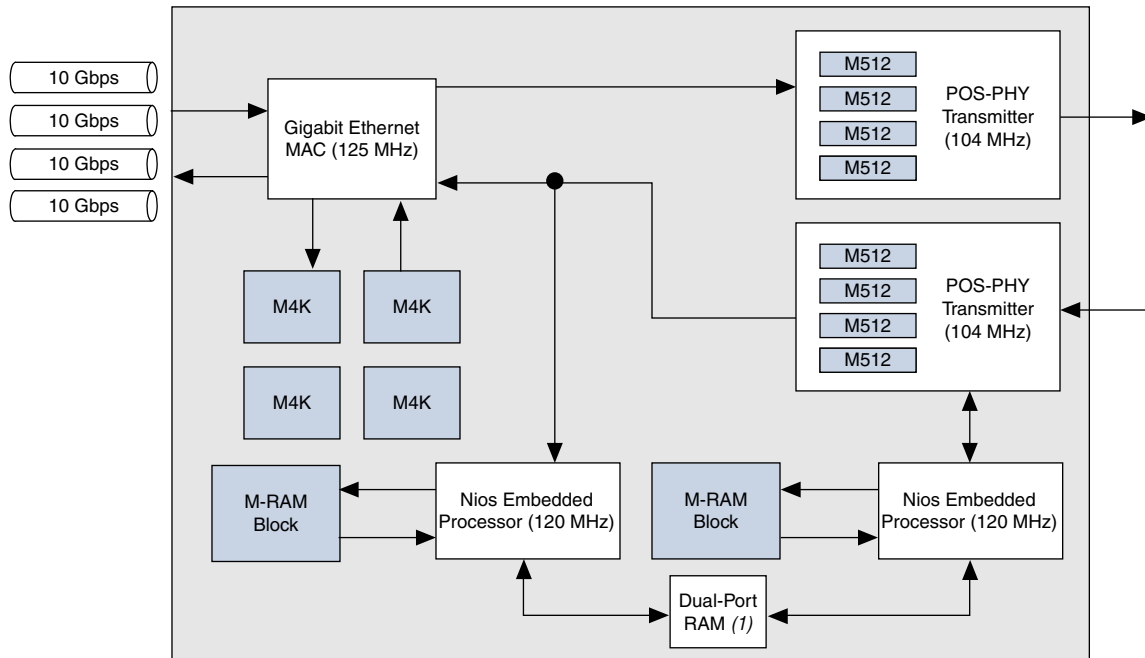
One of the challenges M-RAM faced by engineers designing communications equipment is that memory devices were adopted from PC computer platforms. These devices, although inexpensive and easy to source, are suboptimal for communication systems because they were designed with a different end system in mind. SDRAM memory is a good example of a “borrowed” PC technology. Fundamentally, SDRAM memory was designed for microprocessor storage, not to accept TCP/IP packets travelling at 10 gigabits per second (Gbps) through very high-performance switching equipment.

Although discrete memory suppliers introduced new architectures with increased bandwidth such as double data rate (DDR), zero-bus turnaround (ZBT), and quad data rate (QDR), these architectures have not kept pace with the performance demands of transmission media standards such as OC-48 and OC-192. Since these discrete devices formerly were the only source of system memory, they defined the bottleneck in the system.

The industry adapted by developing sophisticated memory architectures embedded within programmable logic devices (PLDs). Over the last several years, the memory architectures in PLDs have become as complex as many ASIC system designs, including application-optimized features such as dual-port, mixed-width, and parity bits.

The latest generation of high-bandwidth, multi-functional memory architectures available in programmable logic supports higher system speeds than the discrete memory devices currently available. Furthermore, significant performance, time-to-market, and cost advantages are associated with embedded memory in a multi-million-gate programmable logic. [Figure 1](#) shows how embedded memory architectures on PLDs can increase system performance by supporting various functions. [Figure 1](#) also illustrates that although system bandwidth may be 40 Gbps, a significantly higher memory bandwidth may be required to support clock domain translation because each clock domain translation requires two memory transactions.

Figure 1. The Role of PLD Embedded Memory Architectures in Communications Equipment

**Note to Figure 1:**

(1) M4k, M512, and M-RAM™ blocks support dual-port RAM mode.

In the new Stratix™ family of high-density PLDs, Altera offers the TriMatrix™ memory architecture, a technology that meets the memory bandwidth requirements of emerging transmission and processing technologies unavailable with discrete memory devices. The TriMatrix memory architecture in Stratix devices offers up to 10 Mbits of RAM and up to 12 terabits per second peak device memory bandwidth, which makes this family an ideal choice for memory-intensive applications. The TriMatrix memory architecture consists of an array of three embedded RAM block types that can be configured to support a wide range of applications:

- M512 blocks (32 × 18): you can use the smaller M512 blocks of RAM for first-in first-out (FIFO) applications
- M4K blocks (128 × 36): you can use the medium size M4K for channelized functions (a channelized function consolidates smaller channels into a larger channel)
- M-RAM blocks (4,096 × 144): you can use the large M-RAM blocks with 512K bits of RAM for network processor cache

The different memory block sizes are the foundation of the Stratix high-bandwidth memory architecture. The organization of these resources enables you to customize these abundant memory resources distributed throughout the programmable logic at a lower overall system cost.

Communications systems contain devices which use multiple standards at multiple clock frequencies

Communications systems require memory to perform a variety of functions in the system, including buffering data between clock domains. As various telecommunications and/or networking standards intersect in hardware, memory resources are required to translate or “shake hands” between their unique clock domains, and slow down or speed up the rate of data transfer.

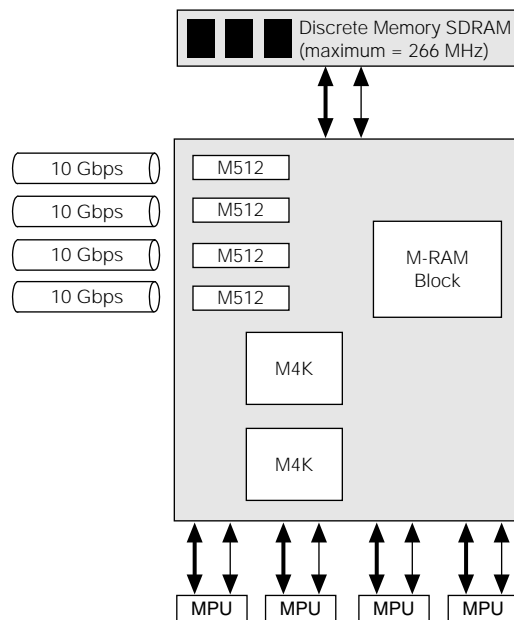
For example, the OC-48 and OC-192 telecommunications standards perform the same function, but at different frequencies. However, many systems must support both standards, requiring translation between them. Memory on a PLD effectively supports this translation (see Figure 1).

Data packet sizes are growing, and the speed at which we can transport them between devices is accelerating

As applications such as video-on-demand become mainstream, the packaging of transmitted data is changing to accommodate increasing packet sizes. At the same time, wide area networks (WANs) are beginning to adopt multiple-wavelength fibers with each wavelength running at 10 Gbps, and have started planning for 40-Gbps networks. Increasing data sizes and rates create greater demands for larger storage buffers on end systems and devices.

With the TriMatrix memory architecture, you can bring this data into a Stratix device, store (in large packet buffers), synchronize (with small wide FIFO buffers), process (with multiple Nios™ embedded processor cores using M-RAM blocks for data storage), and then send the data back out (see Figure 2).

Figure 2. Major Elements in Communications Design that Define the Role of Memory



Processor performance is growing at a slower rate than transmission media I/O speeds

Since 1999, the roadmaps for microprocessor and network processor devices have been one step behind the requirements of current networking research and development. This lag in development results in system implementations in which the number of system processors is a function of the system bandwidth, leading to more overhead, less efficiency, additional board space, additional devices, higher costs, increased board complexity, and higher power consumption. Network designers are forced to accept PC technologies that never quite do exactly what they want.

Network processor suppliers' current roadmaps show the ability to process 10 Gbps, approximately 25% of the current bandwidth of optical transport technologies under development. Therefore, to achieve bandwidth objectives, the system must incorporate multiple processors to support the anticipated network packet loads. The additional complication caused by more processors is typically addressed through complex memory-management techniques. Multiple ports, byte masking, and large buffers are all integral parts of these techniques. With every additional

processor, additional data storage requirements are introduced, including standard CPU cache and main CPU memory.

The TriMatrix memory architecture incorporates memories of multiple sizes that can be configured to serve a multitude of system-level functions, reducing the total number of memory solutions required in a given system. The result is a reduction in system complexity and associated unit costs, and a “cleaner” design.

Off-chip memory speeds for storing those packets are also growing at a slower rate than PLD speeds

Further exacerbating the multi-processor memory complexity, the dedicated-memory devices such as SDRAM, quad data rate QDR SRAM, and DDR SDRAM have not accelerated at a pace comparable to the bandwidth capacity of the transmission media. Once again, a PLD is required to accept packets of data at very high speeds and store it quickly in wide memories, reducing the transmission frequency of the data packets to a level at which discrete memory devices can successfully receive and store them.

For example, a 40-Gbps input data rate has to be translated to a 256-bit bus running at 156 MHz. This 256-bit bus needs to communicate with a back-end interface, a microprocessor, and an external memory on the input and output side (see [Figure 2](#)). Therefore, the memory architecture must support four (interfaces) \times two (input/output) \times 256 bits (bus width) for basic interfacing.

Very high memory bandwidth is required at the intersection of microprocessor, memory, and transmission technologies, which is the PLD

One important metric for evaluating a memory architecture is its peak memory bandwidth. The peak memory bandwidth defines the upper limit of the device’s ability to move data through memory.

To calculate the peak memory bandwidth of a Stratix device, each TriMatrix memory block is assumed to be at maximum clock frequency in simple dual-port mode (i.e., the devices can simultaneously read and write data). The calculation used to determine the peak memory bandwidth in the Stratix family is as follows:

Peak Bandwidth of Each Block Type

- *M512 Block Peak Bandwidth:* (maximum achievable frequency \times maximum configurable bit width)/number of clock cycles to write a word \times number of write ports
- *M4K Block Peak Bandwidth:* (maximum achievable frequency \times maximum configurable bit width)/number of clock cycles to write a word \times number of write ports
- *M-RAM Block Peak Bandwidth:* (maximum achievable frequency \times maximum configurable bit width)/number of clock cycles to write a word \times number of write ports

Peak Bandwidth of Each Stratix Device

The calculation used to determine the peak memory bandwidth in each Stratix device is as follows:

(Peak bandwidth of M512 \times number of M512 blocks on Device) + (peak bandwidth of M4K \times number of M512 blocks on device) + (peak bandwidth of M-RAM block \times number of M512 blocks on device)

The peak bandwidth of each Stratix device equation calculates the memory architecture's bandwidth within the device. The calculation also assumes that each block type contains the standard number of ports (i.e., no special techniques for increasing the number of ports are utilized). See [Tables 1 and 2](#).

Table 1. Peak Memory Bandwidth by Stratix Block Type

File	Peak Memory Bandwidth
M512 Blocks	6 Gbps
M4K Blocks	11 Gbps
M-RAM Blocks	43 Gbps

Table 2. Peak Memory Bandwidth by Stratix Device

Device	M512 Resources	M4K Resources	M-RAM Resources	Peak Memory Bandwidth
EP1S10	95	60	1	1 terabit per second
EP1S20	194	82	2	2 terabits per second
EP1S25	224	138	2	3 terabits per second
EP1S30	295	171	4	4 terabits per second
EP1S40	384	183	4	4 terabits per second
EP1S60	574	292	6	7 terabits per second
EP1S80	767	364	8	9 terabits per second
EP1S120	1,118	520	12	12 terabits per second

The PLD must manage packet speeds and buffer them while the microprocessor and off-chip memories prepare themselves for more data. For more information on Stratix I/O capabilities, see [Application Note 202 \(Using High-Speed Differential I/O Interfaces in Stratix Devices\)](#).

Conclusion

Emerging technologies in PLDs such as the TriMatrix memory architecture, address the need of next-generation communications systems. By incorporating high-density memory architectures into devices with the I/O capabilities to receive and transmit data at high rates, Stratix devices are well suited for communications system designs.



101 Innovation Drive
San Jose, CA 95134
(408) 544-7000
<http://www.altera.com>

Copyright © 2002 Altera Corporation. All rights reserved. Altera, The Programmable Solutions Company, the stylized Altera logo, specific device designations, and all other words and logos that are identified as trademarks and/or service marks are, unless noted otherwise, the trademarks and service marks of Altera Corporation in the U.S. and other countries. All other product or service names are the property of their respective holders. Altera products are protected under numerous U.S. and foreign patents and pending applications, maskwork rights, and copyrights. Altera warrants performance of its semiconductor products to current specifications in accordance with Altera's standard warranty, but reserves the right to make changes to any products and services at any time without notice. Altera assumes no responsibility or liability arising out of the application or use of any information, product, or service described herein except as expressly agreed to in writing by Altera Corporation. Altera customers are advised to obtain the latest version of device specifications before relying on any published information and before placing orders for products or services.