Selecting Your Memory

1

2014.08.15

emi_dg_001

System architects must consider architecture, algorithms, and features of the available components.

Typically, one of the fundamental problems in high-performance applications is memory, because the challenges and limitations of system performance often reside in memory architecture. As higher speeds become necessary for external memories, signal integrity becomes more challenging; newer devices include several features to address this challenge. Altera FPGAs include dedicated I/O circuitry, various I/O standard support, and specialized intellectual property (IP).

When you select an external memory device, consider the following factors:

- Bandwidth and speed
- Cost
- Data storage capacity
- Latency
- Power consumption

Because no single memory type can excel in every area, system architects must determine the right balance for their design. The following table lists the two common types of high-speed memories and their characteristics.

Table 1-1: Differences between DRAM and SRAM

Memory Type	Description	Bandwidth and Speed	Cost	Data Storage Size and Capacity	Power consumption	Latency
DRAM	A dynamic random access memory (DRAM) cell consisting of a capacitor and a single transistor. DRAM memory must be refreshed periodically to retain the data, resulting in lower overall efficiency and more complex controllers. Generally, designers select DRAM where cost per bit and capacity are important. DRAM is commonly used for main memory.	Lower bandwidth resulting in slower speed	Lower cost	Higher data storage and capacity	Higher power consumption	Higher latency

Memory Type	Description	Bandwidth and Speed	Cost	Data Storage Size and Capacity	Power consumption	Latency
SRAM	A static random access memory (SRAM) cell that consists of six transistors. SRAM does not need to be refreshed because the transistors continue to hold the data as long as the power supply is not cut off. Generally, designers select SRAM where speed is more important than capacity. SRAM is commonly used for cache memory.	Higher bandwidth resulting in faster speed	Higher cost	Lower data storage and capacity	Lower power consumption	Lower latency

Note: The Altera IP might or might not support all of the features supported by the memory.

To compare the performance of the supported external memory interfaces in Altera FPGA devices, refer to the *External Memory Interface Spec Estimator* page on the Altera website.

Related Information

External Memory Interface Spec Estimator

DDR SDRAM Features

Double data rate (DDR) SDRAM is a 2n prefetch architecture with two data transfers per clock cycle. It uses a single-ended strobe, DQS, which is associated with a group of data pins, DQ, for read and write operations. Both DQS and DQ are bidirectional ports. Address ports are shared for read and write operations.

The desktop computing market has positioned DDR SDRAM as a mainstream commodity product, which means this memory is very low-cost. DDR SDRAM is also high-density and low-power. Relative to other high-speed memories, DDR SDRAM has higher latency-they have a multiplexed address bus, which reduces the pin count (minimizing cost) at the expense of a longer and more complex bus cycle.

DDR2 SDRAM Features

DDR2 SDRAM is a 4n prefetch architecture (internally the memory operates at half the interface frequency) with two data transfers per clock cycle. DDR2 SDRAM can use a single-ended or differential strobe, DQS or DQSn, which is associated with a group of data pins, and DQ for read and write operations. The DQS, DQSn, and DQ are bidirectional ports. Address ports are shared for read and write operations.

DDR2 SDRAM includes additional features such as increased bandwidth due to higher clock speeds, improved signal integrity on DIMMs with on-die terminations, and lower supply voltages to reduce power.

DDR3 SDRAM Features

DDR3 SDRAM is the third generation of SDRAM. DDR3 SDRAM is internally configured as an eight-bank DRAM and uses an 8n prefetch architecture to achieve high-speed operation. The 8n prefetch architecture

is combined with an interface that transfers two data words per clock cycle at the I/O pins. A single read or write operation for DDR3 SDRAM consists of a single 8n-bit wide, one-clock-cycle data transfer at the internal DRAM core and eight corresponding n-bit wide, one-half clock cycle data transfers at the I/O pins. DDR3 SDRAMs are available as components and modules, such as DIMMs, SODIMMs, RDIMMs, and LRDIMMs.

DDR3 SDRAM can conserve system power, increase system performance, achieve better maximum throughput, and improve signal integrity with fly-by topology and dynamic on-die termination.

Read and write operations to the DDR3 SDRAM are burst oriented. Operation begins with the registration of an active command, which is followed by a read or write command. The address bits registered coincident with the active command select the bank and row to be activated (`BA0` to `BA2` select the bank; `A0` to `A15` select the row). The address bits registered coincident with the read or write command select the starting column location for the burst operation, determine if the auto precharge command is to be issued (via `A10`), and select burst chop (BC) of 4 or burst length (BL) of 8 mode at runtime (via `A12`), if enabled in the mode register. Before normal operation, the DDR3 SDRAM must be powered up and initialized in a predefined manner.

Differential strobes `DQS` and `DQSn` are mandated for DDR3 SDRAM and are associated with a group of data pins, as is `DQ` for read and write operations. `DQS`, `DQSn`, and `DQ` ports are bidirectional. Address ports are shared for read and write operations.

Note: The DDR3 SDRAM high-performance controller II supports local interfaces running at full-rate, half-rate, and quarter-rate.

For more information, refer to the respective DDR, DDR2, and DDR3 SDRAM data sheets.

For more information about parameterizing the DDR2 and DDR3 SDRAM IP, refer to the *Implementing and Parameterizing Memory IP* chapter.

Related Information

Implementing and Parameterizing Memory IP

QDR, QDR II, and QDR II+ SRAM Features

Quad Data Rate (QDR) SRAM has independent read and write ports that run concurrently at double data rate. QDR SRAM is true dual-port (although the address bus is still shared), which gives this memory a high bandwidth, allowing back-to-back transactions without the contention issues that can occur when using a single bidirectional data bus. Write and read operations share address ports.

The QDR II SRAM devices are available in ×8, ×9, ×18, and ×36 data bus width configurations. The QDR II+ SRAM devices are available in ×9, ×18, and ×36 data bus width configurations. Write and read operations are burst-oriented. All the data bus width configurations of QDR II SRAM support burst lengths of two and four. QDR II+ SRAM supports only a burst length of four. Burst-of-two and burst-of-four for QDR II and burst-of-four for QDR II+ SRAM devices provide the same overall bandwidth at a given clock speed.

For QDR II SRAM devices, the read latency is 1.5 clock cycles; for QDR II+ SRAM devices, it is 2 or 2.5 clock cycles depending on the memory device. For QDR II+ and burst-of-four QDR II SRAM devices, the write commands and addresses are clocked on the rising edge of the clock, and write latency is one clock cycle. For burst-of-two QDR II SRAM devices, the write command is clocked on the rising edge of the clock, and the write address is clocked on the falling edge of the clock. Therefore, the write latency is zero because the write data is presented at the same time as the write command.

QDR II+ and QDR II SRAM interfaces use a delay-locked loop (DLL) inside the device to edge-align the data with respect to the `K` and `K#` or `C` and `C#` pins. You can optionally turn off the DLL, but the performance

of the QDR II+ and QDR II SRAM devices is degraded. All timing specifications listed in this document assume that the DLL is on. QDR II+ and QDR II SRAM devices also offer programmable impedance output buffers. You can set the buffers by terminating the ZQ pin to VSS through a resistor, RQ. The value of RQ should be five times the desired output impedance. The range for RQ should be between 175 ohm and 350 ohm with a tolerance of 10%.

QDR II/+ SRAM is best suited for applications where the required read/write ratio is near one-to-one. QDR II/+ SRAM includes additional features such as increased bandwidth due to higher clock speeds, lower voltages to reduce power, and on-die termination to improve signal integrity. QDR II+ SDRAM is the latest and fastest generation. For QDR II+ and QDR II SRAM interfaces, Altera supports both 1.5-V and 1.8-V HSTL I/O standards.

For more information, refer to the respective QDRII and QDRII+ data sheets.

For more information about parameterizing the QDRII and QDRII+ SRAM IP, refer to the *Implementing and Parameterizing Memory IP* chapter.

Related Information

Implementing and Parameterizing Memory IP

RLDRAM II and RLDRAM 3 Features

Reduced latency DRAM (RLDRAM) provides DRAM-based point-to-point memory devices designed for communications, imaging, server systems, networking, and cache applications requiring high density, high memory bandwidth, and low latency. The fast random access speeds in RLDRAM devices make them a viable alternative to SRAM devices at a lower cost.

The high performance of RLDRAM is achieved by very low random access delay (tRC), low data-bus-turnaround delay, simple command protocol, and a large number of banks. RLDRAM is optimized to meet the needs of high-bandwidth networking applications.

Contrasting with the typical four banks in most memory devices, RLDRAM II is partitioned into eight banks and RLDRAM 3 is partitioned into sixteen banks. Partitioning reduces the parasitic capacitance of the address and data lines, allowing faster accesses and reducing the probability of random access conflicts. Each bank has a fixed number of rows and columns. Only one row per bank is accessed at a time. The memory (instead of the controller) controls the opening and closing of a row, which is similar to an SRAM interface.

Most DRAM memory types need both a row and column phase on a multiplexed address bus to support full random access, while RLDRAM supports a nonmultiplexed address, saving bus cycles at the expense of more pins. RLDRAM II and RLDRAM 3 use the High-Speed Transceiver Logic (HSTL) standard with double data rate (DDR) data transfer to provide a very high throughput.

There are two types of RLDRAM II or RLDRAM 3 devices—common I/O (CIO) and separate I/O (SIO). CIO devices share a single data I/O bus, which is similar to the double data rate (DDR) SDRAM interface. SIO devices, with separate data read and write buses, have an interface similar to SRAM. Altera UniPHY Memory IP only supports CIO RLDRAM.

RLDRAM II and RLDRAM 3 use a DDR scheme, performing two data transfers per clock cycle. RLDRAM II or RLDRAM 3 CIO devices use the bidirectional data pins (DQ) for both read and write data, while RLDRAM II or RLDRAM 3 SIO devices use D pins for write data (input to the memory) and Q pins for read data (output from the memory). Both types use two pairs of unidirectional free-running clocks. The memory uses DK and DK# pins during write operations, and generates QK and QK# pins during read operations. In addition, RLDRAM II and RLDRAM 3 use the system clocks (CK and CK# pins) to sample commands and addresses, and to generate the QK and QK# read clocks. Address ports are shared for write and read operations.

RLDRAM II CIO devices are available in ×9, ×18, ×36 data bus width configurations. RLDRAM II CIO interfaces may require an extra cycle for bus turnaround time for switching read and write operations. RLDRAM 3 devices are available in ×18 and ×36 data bus width configurations.

Write and read operations are burst oriented, and all the data bus width configurations of RLDRAM II and RLDRAM 3 support burst lengths of two and four. RLDRAM 3 also supports burst length of eight at bus width ×18, and burst lengths of two and four at bus width ×36. For detailed comparisons between RLDRAM II and RLDRAM 3 for these features, refer to the *Memory Selection Overview* table.

RLDRAM II and RLDRAM 3 also inherently include the additional memory bits used for parity or error correction code (ECC).

RLDRAM II and RLDRAM 3 also offer programmable impedance output buffers and on-die termination. The programmable impedance output buffers are for impedance matching and are guaranteed to produce 25- to 60-ohm output impedance. The on-die termination is dynamically switched on during read operations and switched off during write operations. Perform an IBIS simulation to observe the effects of this dynamic termination on your system. IBIS simulation can also show the effects of different drive strengths, termination resistors, and capacitive loads on your system.

RLDRAM 3 enables a faster, more efficient transfer of data by doubling performance and reduced latency compared to RLDRAM II. RLDRAM 3 memory is suitable for operation in which high bandwidth and deterministic performance is critical, and is optimized to meet the needs of high-bandwidth networking applications. For detailed comparisons between RLDRAM II and RLDRAM 3, refer to the following table.

For more information, refer to RLDRAM II and RLDRAM 3 data sheets available from the Micron website (www.micron.com).

For more information about parameterizing the RLDRAM II and RLDRAM 3 IP, refer to the *Implementing and Parameterizing Memory IP* chapter.

Related Information

- **Implementing and Parameterizing Memory IP**
- **www.micron.com**

LPDDR2 Features

LPDDR2-S is a high-speed SDRAM device internally configured as a 4- or 8-bank memory. All LPDDR2 devices use double data rate architecture on the address and command bus to reduce the number of input pins in the system. The 10-bit address and command bus contains command, address, and bank/row buffer information. Each command uses one clock cycle, during which command information is transferred on both the positive and negative edges of the clock.

LPDDR2-S2 and LPDDR2-S4 devices use double data rate architecture on the `DQ` pins to achieve high speed operation. The double data rate architecture is essentially a 2n/4n prefetch architecture with an interface designed to transfer two data bits per `DQ` every clock cycle at the I/O pins. A single read or write access for the LPDDR2-S2/S4 consists of a single 2n-bit wide /4n-bit wide, one clock cycle data transfer at the internal SDRAM core, and two/four corresponding n-bit wide, with one-half clock cycle data transfers at the I/O pins.

Memory Selection

One of the first considerations in choosing a high-speed memory is data bandwidth. Based on the system requirements, an approximate data rate to the external memory should be determined. You must also consider other memory attributes, including how much memory is required (density), how much latency can be tolerated, what is the power budget, and whether the system is cost sensitive.

The following table lists memory features and target markets of each technology.

Table 1-2: Memory Selection Overview

Parameter	LPDDR2	DDR3 SDRAM	DDR2 SDRAM	DDR SDRAM	RLDRAM II	RLDRAM 3	QDR II/+ SRAM
Bandwidth for 32 bit interface *(1)*	25.6	59.7	25.6	12.8	25.6	35.8	44.8
Bandwidth at % Efficiency (Gbps) *(2)*	17.9	23.9	17.9	9	17.9	–	38.1
Performance / Clock frequency	167–400 MHz *(3)*	400–933 MHz	167–400 MHz *(3)*	100–200 MHz	200–533 MHz	200–800 MHz	154–350 MHz
Altera-supported data rate	Up to 1,066 Mbps	Up to 2,133 Mbps	Up to 1,066 Mbps	Up to 400 Mbps	Up to 1066 Mbps	Up to 1600 Mbps	Up to 1400 Mbps
Density	64 MB –8 GB	512 MB–8 GB,32 MB –8 GB (DIMM)	256 MB–1 GB,32 MB –4 GB (DIMM)	128 MB–1 GB, 32 MB –2 GB (DIMM)	288 MB,576 MB	576 MB – 1.1 GB	18–144 MB
I/O standard	HSUL- 12 1.2V	SSTL-15 Class I, II	SSTL-18 Class I, II	SSTL-2 Class I, II	HSTL-1.8V/ 1.5V	HSTL-1.2V and SSTL-12	HSTL-1.8V/ 1.5V
Data group width	8, 16, 32	4, 8, 16	4, 8, 16	4, 8, 16, 32	9, 18, 36	18, 36	9, 18, 36
Burst length	4, 8, 16	8	4, 8	2, 4, 8	2, 4, 8	2, 4, 8	2, 4
Number of banks	4, 8	8	8 (>1 GB), 4	4	8	16	—

Parameter	LPDDR2	DDR3 SDRAM	DDR2 SDRAM	DDR SDRAM	RLDRAM II	RLDRAM 3	QDR II/+ SRAM
Row/ column access	Row before column	Row before column	Row before column	Row before column	Row and column together or multiplexed option	Row and column together or multiplexed option	—
CAS latency (CL)	—	5, 6, 7, 8, 9, 10	3, 4, 5	2, 2.5, 3	—	—	—
Posted CAS additive latency (AL)	—	0, CL-1, CL-2	0, 1, 2, 3, 4	—	—	—	—
Read latency (RL)	3, 4, 5, 6, 7, 8	RL = CL + AL	RL = CL + AL	RL = CL	3, 4, 5, 6, 7, 8	3-16	1.5, 2, and 2.5 clock cycles
On-die termination	—	Yes	Yes	No	Yes	Yes	Yes
Data strobe	Differential bidirectional	Differential bidirectional strobe only	Differential or single-ended bidirectional strobe	Single-ended bidirectional strobe	Free-running differential read and write clocks	Free-running differential read and write clocks	Free-running read and write clocks
Refresh requirement	Yes	Yes	Yes	Yes	Yes	Yes	No
Relative cost comparison	Higher than DDR SDRAM	Presently lower than DDR2	Less than DDR SDRAM with market acceptance	Low	Higher than DDR SDRAM,less than SRAM	Higher than DDR SDRAM,less than SRAM	Highest

Parameter	LPDDR2	DDR3 SDRAM	DDR2 SDRAM	DDR SDRAM	RLDRAM II	RLDRAM 3	QDR II/+ SRAM
Target market	Mobile devices that target low operating power	Desktops, servers, storage, LCDs, displays, networking, and communication equipment	Desktops, servers, storage, LCDs, displays, networking, and communication equipment	Desktops, servers, storage, LCDs, displays, networking, and communication equipment	Main memory, cache memory, networking, packet processing, and traffic management	Main memory, cache memory, networking, packet processing, and traffic management	Cache memory, routers, ATM switches, packet memories, lookup, and classification memories

Notes to Table:

1. 32-bit data bus operating at the maximum supported frequency in a Stratix® V FPGA.
2. 70% efficiency for DDR memories, which takes into consideration the bus turnaround, refresh, infinite burst length and random access latency and assumes 85% efficiency for QDR memories.
3. The lower frequency limit depends on the higher of the DLL frequency and the minimum operating frequency of the given EMIF protocol.

Altera supports the memory interfaces, provides various IP for the physical interface and the controller, and offers many reference designs (refer to Altera's *Memory Solutions Center*).

For Altera support and the maximum performance for the various high-speed memory interfaces, refer to the *External Memory Interface Spec Estimator* page on the Altera website.

Related Information

- **Memory Solutions Center**
- **External Memory Interface Spec Estimator**

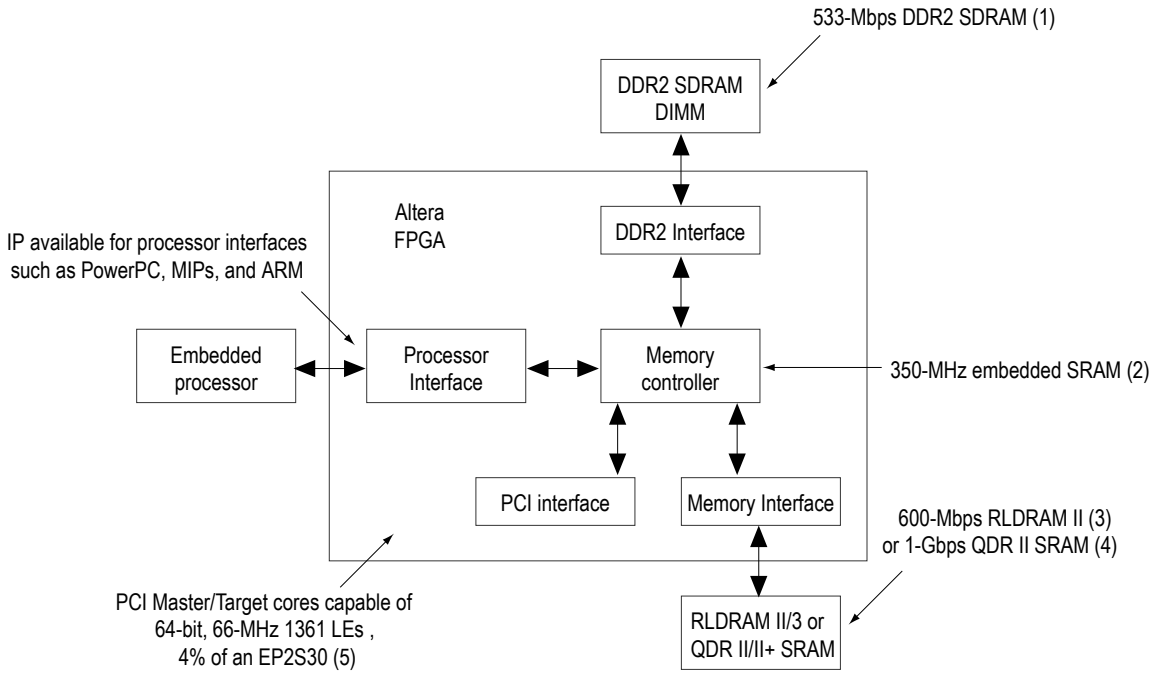
Example of High-Speed Memory in Embedded Processor

In embedded processor applications—any system that uses processors, excluding desktop processors—due to its very low cost, high density, and low power, DDR SDRAM is typically used for main memory.

Next-generation processors invest a large amount of die area to on-chip cache memory to prevent the execution pipelines from sitting idle. Unfortunately, these on-chip caches are limited in size, as a balance of performance, cost, and power must be taken into consideration. In many systems, external memories are used to add another level of cache. In high-performance systems, three levels of cache memory is common: level one (8 Kbytes is common) and level two (512 Kbytes) on chip, and level three off chip (2 Mbytes).

High-end servers, routers, and even video game systems are examples of high-performance embedded products that require memory architectures that are both high speed and low latency. Advanced memory controllers are required to manage transactions between embedded processors and their memories. Altera Arria® series and Stratix series FPGAs optimally implement advanced memory controllers by utilizing their built-in DQS (strobe) phase shift circuitry. The following figure highlights some of the features available in an Altera FPGA in an embedded application, where DDR2 SDRAM is used as the main memory and QDR II/II+ SRAM or RLDRAM II/3 is an external cache level.

Figure 1-1: Memory Controller Example Using FPGA



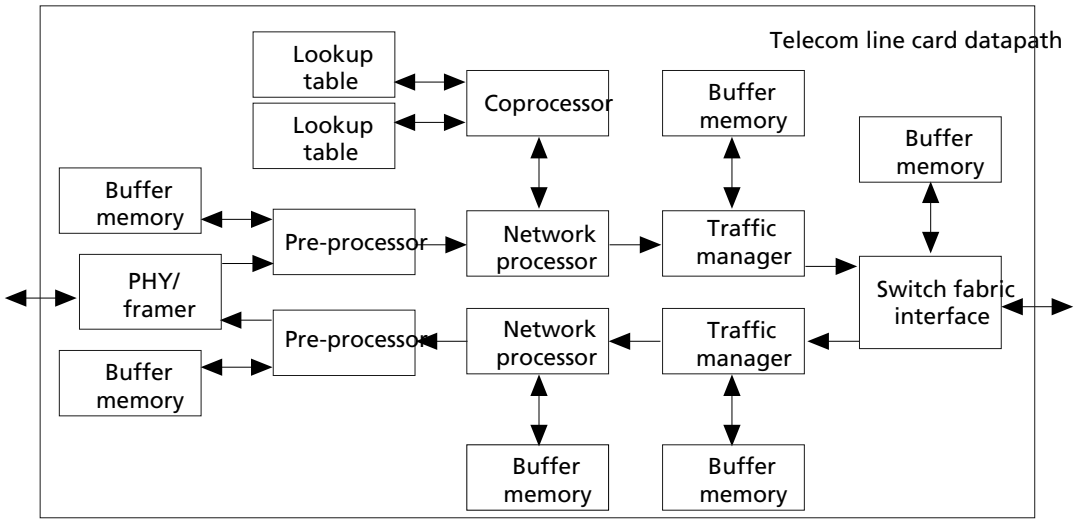
One of the target markets of RLDRAM II/3 and QDR/QDR II SRAM is external cache memory. RLDRAM II and RLDRAM 3 have a read latency close to synchronous SRAM, but with the density of SDRAM. A sixteen times increase in external cache density is achievable with one RLDRAM II/3 versus that of synchronous static RAM (SSRAM). In contrast, consider QDR and QDR II SRAM for systems that require high bandwidth and minimal latency. Architecturally, the dual-port nature of QDR and QDR II SRAM allows cache controllers to handle read data and instruction fetches completely independent of writes.

Example of High-Speed Memory in Telecom

Because telecommunication network architectures are becoming more complex, high-end network systems are running multiple 10-Gbps line cards that connect to multi-shelf switch fabrics scaling to terabits per second.

The following figure shows an example of a typical system line interface card. These line cards offer interfaces ranging from a single-port OC-192 to multi-port Gbps Ethernet, and consist of a number of devices, including a PHY/framer, network processors, traffic managers, fabric interface devices, and high-speed memories.

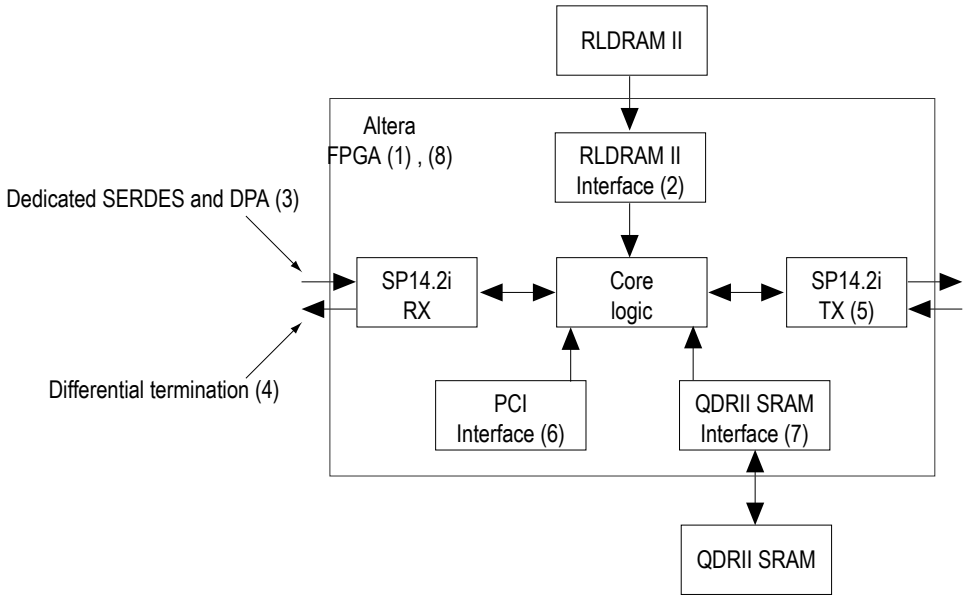
Figure 1-2: Typical Telecom System Line Interface Card



As packets traverse from the PHY/framer device to the switch fabric interface, they are buffered into memories, while the data path devices process headers (determining the destination, classifying packets, and storing statistics for billing) and control the flow of packets into the network to avoid congestion. Typically DDR/DDR2/DDR3 SDRAM and RLDRAM II/3 are used for large buffer memories off network processors, traffic managers, and fabric interfaces, while QDR and QDR II/II+ SRAMs are used for look-up tables (LUTs) off preprocessors and coprocessors.

In many designs, FPGAs connect devices together for interoperability and coprocessing, implement features that are not supported by ASIC devices, or implement a device function entirely. Altera Stratix series FPGAs implement traffic management, packet processing, switch fabric interfaces, and coprocessor functions, using features such as 1-Gbps LVDS I/O, high-speed memory interface support, multi-gigabit transceivers, and IP cores. The following figure highlights some of these features in a packet buffering application where RLDRAM II is used for packet buffer memory and QDR II SRAM is used for control memory.

Figure 1-3: FPGA Example in Packet Buffering Application



SDRAM is usually the best choice for buffering at high data rates due to the large amounts of memory required. Some system designers take a hybrid approach to the memory architecture, using SRAM to store the packet headers and DRAM to store the payload. The depth of the memories depends on the architecture and throughput of the system.

The buffer memory for the packet buffering application of an OC-192 line card (approximately 10 Gbps) must be able to sustain a minimum of one write and one read operation, which requires a memory bandwidth of 20 Gbps to operate at full line rate (more bandwidth is required if the headers are modified). The bandwidth requirement for memory is a key factor in memory selection. As an example, a simple first-order calculation using RLDRAM II as buffer memory requires a bus width of 48 bits to sustain 20 Gbps (300 MHz × 2 DDR × 0.70 efficiency × 48 bits = 20.1 Gbps), which needs two RLDRAM II parts (one ×18 and one ×36). RLDRAM II and RLDRAM 3 also inherently include the additional memory bits used for parity or error correction code (ECC). QDR and QDR II SRAM have bandwidth and low random access latency advantages that make them useful for control memory in queue management and traffic management applications. Another typical implementation for this memory is billing and packet statistics, where each packet requires counters to be read from memory, incremented, and then rewritten to memory. The high bandwidth, low latency, and optimal one-to-one read/write ratio make QDR SRAM ideal for this feature.

Document Revision History

Date	Version	Changes
August 2014	2014.08.15	• Changed some values in the *Bandwidth for 32 bits*, *Bandwidth at % Efficiency*, *Performance / Clock frequency*, and *Altera-supported data rate* rows of the *Memory Selection Overview* table.

Date	Version	Changes
December 2013	2013.12.16	Removed references to Stratix II devices.
November 2012	6.0	Added RLDRAM 3 support.
June 2012	5.0	• Added LPDDR2 support. • Added Feedback icon.
November 2011	4.0	Moved and reorganized "Selecting your Memory" section to Volume 2: Design Guidelines.
June 2011	3.0	Added "Selecting Memory IP" chapter from Volume 2.
December 2010	2.1	• Moved protocol-specific feature information to the memory interface user guides in Volume 3. • Updated maximum clock rate information for 10.1.
July 2010	2.0	• Added specifications for DDR2 and DDR3 SDRAM Controllers with UniPHY. • Streamlined the specification tables. • Added reference to web-based Specification Estimator Tool.
January 2010	1.1	Updated DDR, DDR2, and DDR3 specifications.
November 2009	1.0	First published.