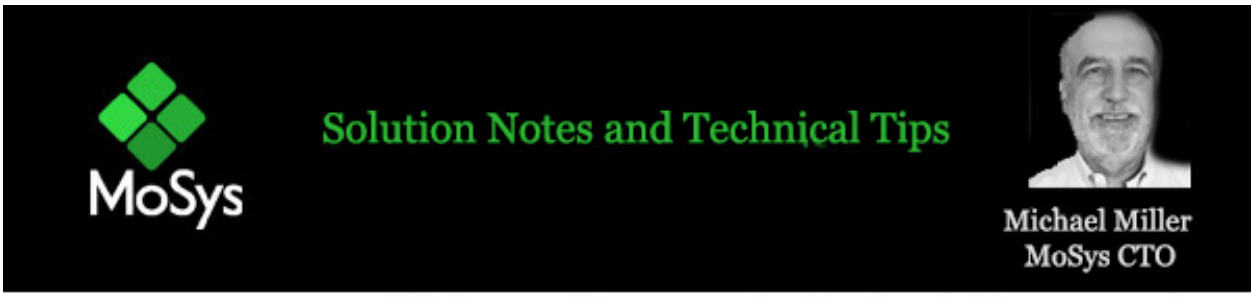


SOLUTION NOTE #1002: Achieving 6.5 Billion Transactions/Sec in a Single Memory Device

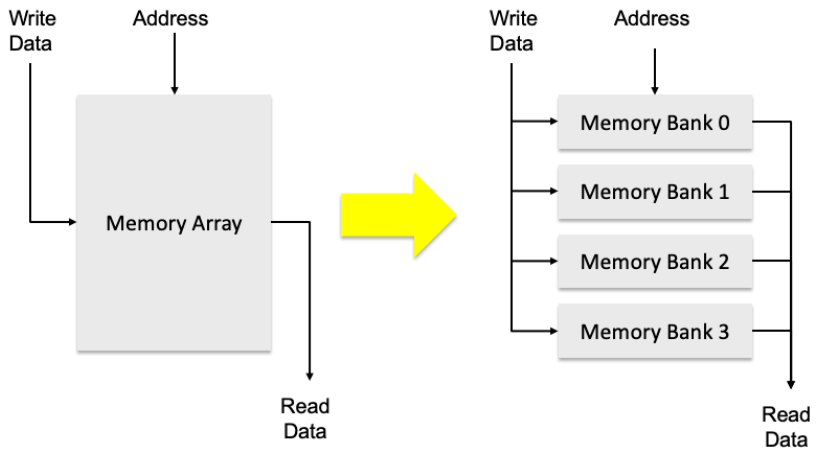
TOPIC SUMMARY:

Emerging applications are demanding more random transactions than ever before. Many now need to access a large address space at 2B random transactions per second or more. Let's take, for example:

- You are designing a system to handle ethernet traffic on 8 ports @ 100G, then you will need (8) x (150Mpps) x (the number of routing or switching features you need to search), meaning, you will require a memory system that provides for greater than 2Bts.
- In stateful firewall applications, the ability to read and write a memory location in less than 6ns is key.
- Machine Learning algorithms, which perform classification operations using a random forest of trees approach, will be limited by traditional memories' access rates.
- Video editing on the fly for streaming 4K HD frames with one write and 2 reads can reach 3Bts. The worst case for 8K HD can be four times that found in 4K HD.

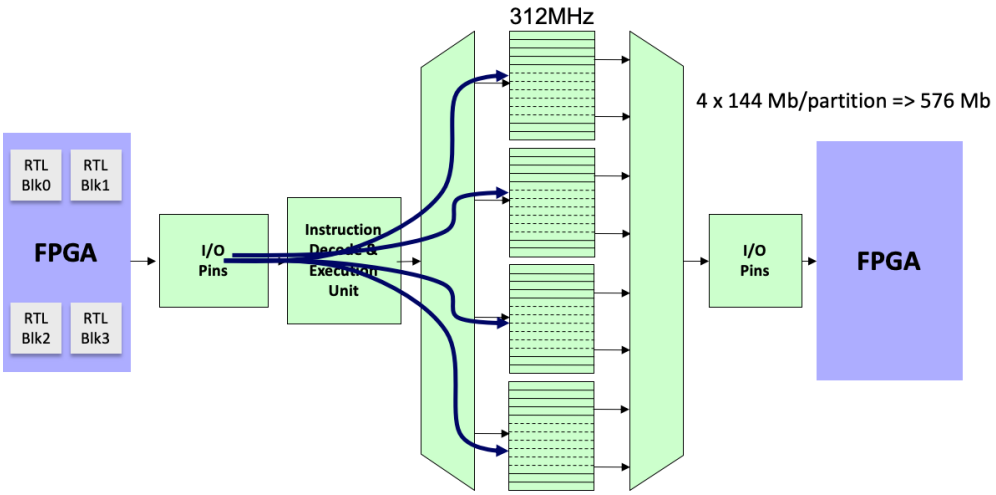
As process nodes have progressed to ever smaller geometries, FPGAs and ASICs have incorporated an increasing amount of on-chip memory. While having on-chip memory fits many applications well, often times the designer cannot fit all of the application data set in the available memory or it may not be the correct configuration or random-access profile. For these reasons and others, off-chip memory is still an important option. To keep up with demand, there are many more emerging applications that will exceed the random access rates of memories such as QDR IV, SigmaQuad IVe, RLDRAM, DDR and HBM.

Today's memory technologies have evolved from a single bank to multi-bank architectures. Doing so allows the address and data bus I/O pins to cycle at a multiple of the internal memory native cycle rate (tRC). This has evolved because memory cycle times have not kept up with I/O technology due to the intrinsic characteristics of the on-chip transistor and wire resistance/ capacitance. Today, SRAMs can cycle around 600MHz to 1.5GHz, DRAM cycles at 20MHz and embedded DRAM at 375MHz.



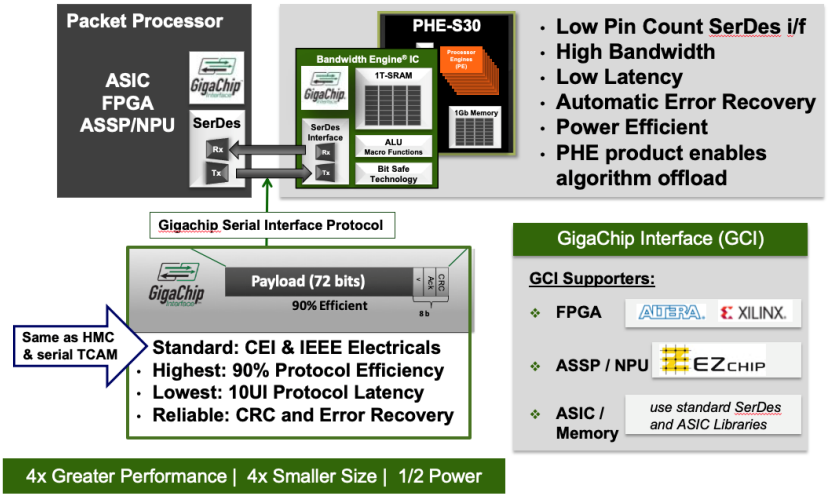
In multi-banked architectures, the I/O clock rate is the bank cycle rate x the number of banks. Double Data Rate schemes can cut the data rate on the bus by a factor of two. The architecture of the application inside the FPGA then interacts with the different blocks in the memory in a Time Division Multiplexing (TDM) fashion. This means that each block of logic (Blk0, Blk1, Blk2 in the drawing below) gets a round robin time slice of the I/O pins to transmit a memory transaction. In many synchronous architectures, if there is no transaction required in a given time slot, then a "no operation" command is transmitted. In the same fashion, all read data is sent back to the FPGA in the same TDM ordering. In this way, all transactions can be kept in order. To make the best use of the TDM architecture, layout of data must then be considered.

Architectural features of the memory such as bank size and number of banks become important. Too few banks can translate into coarse allocation options which, in turn, can lead to poor utilization. This can also lead to scheduling concerns wherein tables can be accessed by the logic blocks in the FPGA. The number of internal ports can also lead to scheduling conflicts. This often forces designers to lay down many more memory devices in order to achieve the performance needed. This in turn, complicates design layouts, board space, power and other factors all of which contribute to additional costs and can delay product introductions.

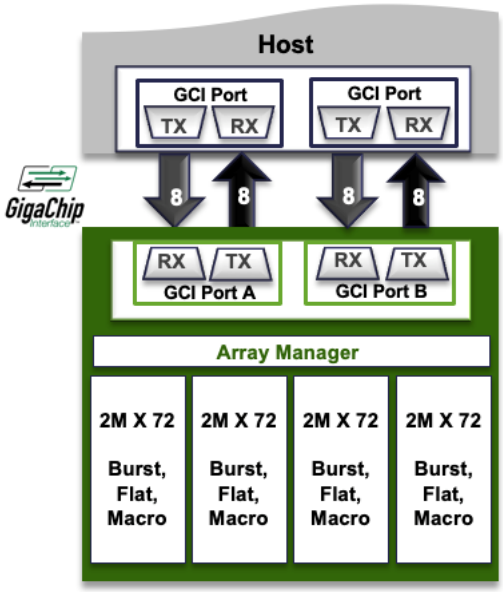


MoSys© Bandwidth Engine Family

At 1.333GHz, the buses are a challenge to layout even when utilizing new techniques such as matched trace length, training at reset, inversion and error correction. To take buses to even higher frequencies, SerDes signaling can provide a proven and reliable path. To that end, MoSys has developed a reliable transport protocol called Giga Chip Interface (GCI) which operates on top of the typical CEI 11 and CEI 25 signaling standards used by all ASICs and FPGAs today.

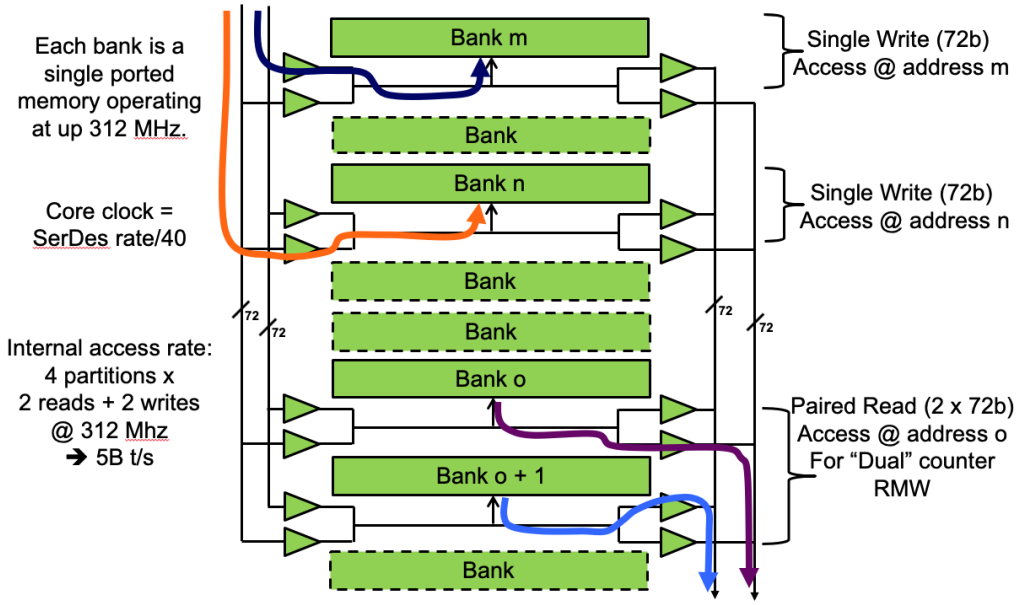


Using the GCI interface, the MoSys Bandwidth Engine 2 (BE2) offers 576Mb of high speed 1T-SRAM available over 16 SerDes lanes with a random-access rate of up to 3.33 Bts (2.5Bts read + .83Bts write) or 2.5Bts balanced reads and writes. The memory is organized as 4 partitions with 64 banks in each partition.



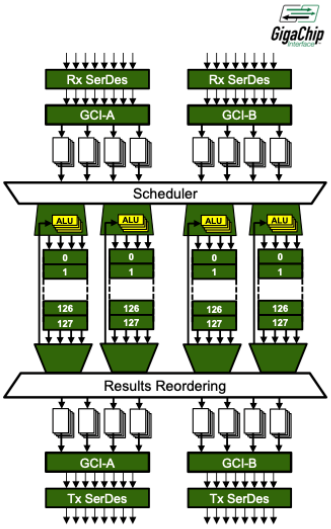
Because the MoSys BE2 uses 1T-SRAM with a tRC as low as 2.66ns, it is highly suitable for high access rate applications such as packet header processing for stateful applications. For example, at 100Gbps, packets can arrive at 150Mpps or every ~6ns. With a tRC <3ns, both a read and a write can be accomplished within the 6ns boundary. With the MoSys BE2, there can be 16 (8 read and 8 write) memory transactions inflight at a core rate of 1.25GHz (4 x 312.5 MHz). To realize the full internal random-access rates and avoid limitations of the I/O, the designer can make use of on-chip ALUs to perform read modify write (RMW) operations such as add/subtract, dual counter and Boolean operations. More details are provided in other technical notes on the RMW in-memory functions. (See the BE-2 User Guide which is available on the MoSys Customer Portal site)

In the BE2 architecture, there are two read ports and two write ports per partition. This allows for dual accesses as well as the ability for two RTL blocks to access each partition simultaneously.



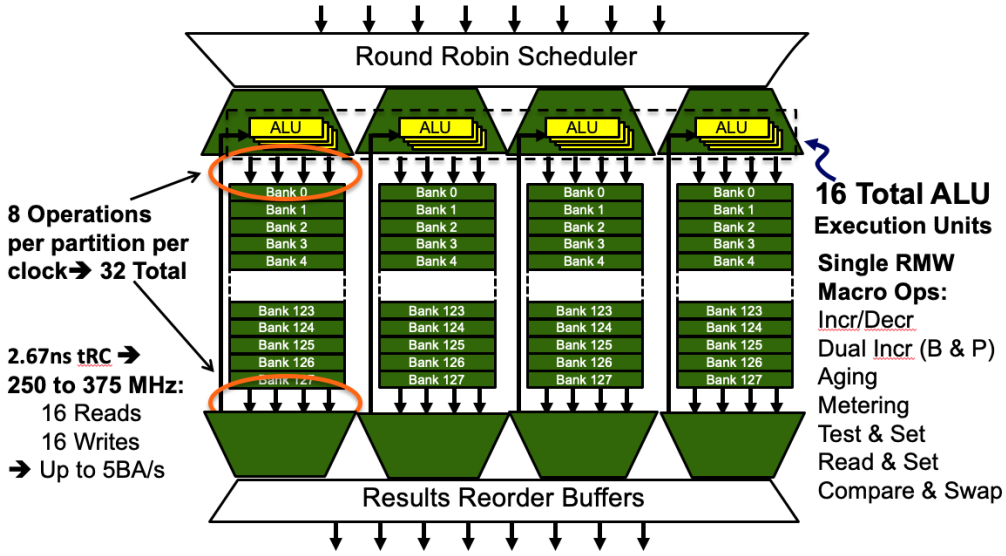
Bandwidth Engine 3

The MoSys Bandwidth Engine 3 (BE3) device not only doubles capacity to 1 Gb and increases the SerDes lane rates to 25Gbps, but it also allows for up to 32 transactions (16 read + 16 write) to be handled simultaneously on each clock. BE3 is useful for applications where higher bandwidth is required per pin, higher capacity on a single interface, higher transaction rate or the user needs the more powerful in-memory operations (i.e., 400G ENet, 8K HD Video processing, Machine Learning and high-speed decision trees such as TCAMs).



The BE3 architecture incorporates an on-chip scheduler and re-order buffer to allow users to take advantage of features such as internal over clocking and true dual port operation between the two GCI ports. To keep up with the I/O, the internal banks have 4

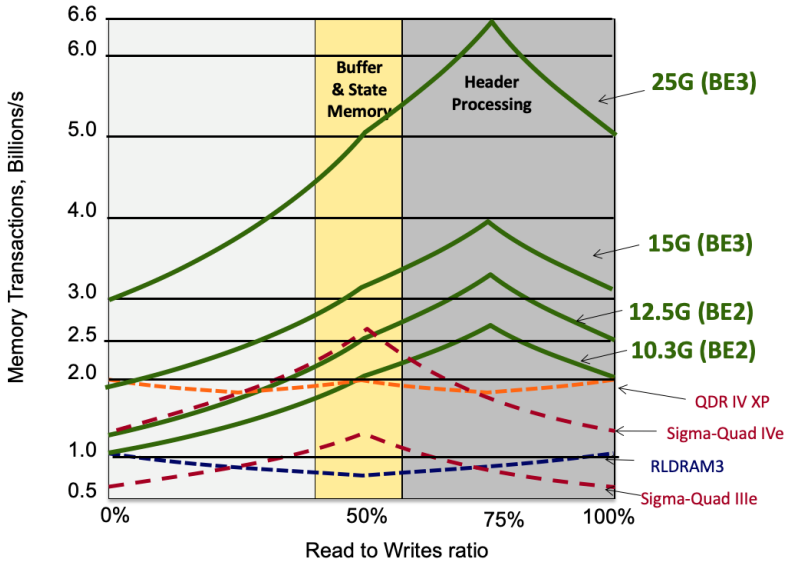
read and 4 write data buses. This allows for a doubling of the internal RMW ALU operations for in-memory operations. The BE3 architecture increases both the number of ports per partition to 4 reads and 4 writes and the number of banks to 128 which in turn, reduces the probability of collision by four.



In the BE-3 device, it is possible to operate the internal core at 1.5GHz, allowing 12Bts (6B read and 6B write) which supports a peak external performance of 5Bts per second total at 25Gbps SerDes rates. From these possible operations, there is the possibility of many more internal accesses than the external I/O allows for. The internal overprovisioning can therefore be used by on-chip in-memory processing operations including RMW or the 32 internal RISC cores with its purpose-defined ISA. (The RISC processors will be covered in a future solutions note.)

Comparing Traditional Network Memories

For traditional comparisons of networking memories, the application must be analyzed with respect to the ratio of reads-to-writes. For applications that demand an equal number of reads-to-writes, the devices with separate buses for read data and write data shine (various QDR and SigmaQuad) whereas memories with shared I/O pins (RLDRAM, DDR) are lost while the bus is being turned around. The MoSys BE product family not only shows good balanced read/write performance, but they also have the unique capability to provide even greater performance by optimizing for Read dominated applications (3:1 ratio of reads-to-writes). For packet processing, this can be the difference between look up tables in header processing vs packet buffering or state tables. In video editing this can be useful, where the controller writes once but reads several times.



The number of banks is dramatically different. The BE3 has 512 banks and the BE2 has 256. This is in comparison to RLDRAM and SigmaQuad which have 8. Therefore, the allocation efficiency and collision probability are much lower in the BE products.

Comparing with HBM

On today's FPGAs, High Bandwidth Memory (HBM) is available either in one or two 4 high stacks. HBM is an outstanding high-density memory featuring great burst bandwidth (202 GBs) characteristics. The limitation of HBM is similar to that of DRAM in that the tRC is ~50ns. The random-access rate is 2.7 Bts theoretically, and more like 1.8 Bts in practice when measured with single word transfers using a random-access pattern. This requires the organization of many parallel banks and often duplicated data when allowed. On the other hand, the BE3 device has 4x time the number of banks with a tRC as low as 2.67ns and a total external access rate of 6.5Bts and internal of 12Bts. Thus, MoSys BE2/BE3 is complementary to HBM by providing a high random-access rate while HBM features high bandwidth.

HBM > BE3/PHE for streaming data
PHE > 5x HBM for Random Access Rate
PHE > 3x the Random Access Bandwidth

	DDR4 DIMM	Single 4x High HBM2	BE3	PHE
tRC (Row Cycle time in a bank)	50ns	45ns	3ns	3ns
# Banks	16	128	512	512
Maximum Random Access Read Only Rate	320 M/s	2.7 B/s	5B/s	5B/s
Measured Random Access Rd + Wr Rate	200 M/s	1.8 B/s	6.5B/s	10B/s
Measured Random Access Rd + Wr Bandwidth	12.8GB/s	57.6GB/s	58GB/s	180GB/s
Measured Maximum Rd + Wr BL8 Bandwidth	20GB/s	202 GB/s	90GB/s	180GB/s

Note:
- HBM - 900MHz clock, 32B Random Accesses, 256B for Maximum Read Bandwidth Accesses
- BE3– 1.25GHz clock, 9B for Maximum external measured Random Access Rd + Wr Bandwidth
- PHE– 1.25GHz clock, 18B for Maximum internal measured Random Access Rd + Wr Bandwidth
- HBM maximum access bandwidth as reported in HC29.22.530, maximum random access rate extrapolated

Conclusion

MoSys Bandwidth Engine products provide random access rates that extend well beyond currently available traditional networking memories. When the application requires a moderately large memory (128GB) with access rates well in excess of 2Bts, the MoSys BE products offer the industry's highest performance.

If you are looking for more technical information or need to discuss your technical challenges with an expert, we are happy to help! Email us and we will arrange to have one of our technical specialists speak with you. You can also sign up for our newsletter. Already convinced? You can request a quote from sales. Finally, please follow us on social media so we can keep in touch.