Networks & Communications Group

Customized COTS (C^2OTS) Series Part 2

Fabric Mezzanine Modules Push IA Blade to 160Gbps on ATCA

ADVANTECH

Enabling an Intelligent Planet

Fabric Mezzanine Modules Push IA Blade to 160Gbps on ATCA

The first white paper in this series discussed the design advantages of Fabric Mezzanine Modules (FMMs). It introduced the merits of Advantech's Customized Commercial-Off-The-Shelf (C^2OTS) concept through the use of FMM technology applied to AdvancedTCA (ATCA) blades. It also explained the importance of C^2OTS for systems integrators, security and network equipment providers as well as government and defense communications contractors.

This second paper in the series will take a closer look at the flexibility which FMMs bring to the fabric interface enabling 40GbE connectivity of Intel® Xeon® E5-2600 Processor based blades to 40G ATCA backplanes. Various system topologies will be covered which pave the way for the highest performance, consolidated control and data plane workloads under a single IA architecture.

ADVANTECH
Enabling an Intelligent Planet

Fabric Mezzanine Modules Push IA Blade to 160Gbps on ATCA

If there is one thing that never seems to change it is the requests from our customers to build even greater levels of flexibility, design reuse and performance into our products. There are many different reasons why these factors are high on the wish list of virtually every design team. It may be time to market, cost reduction or the desire to offer the latest technologies that drives a specific development project but for Network Equipment Providers connectivity and platform throughput are always high on the list. This is closely followed by the ability to create differentiated products in a variety of configurations without having to continually reinvent the wheel. It's a classic conundrum in the networking world of how to create a single blade that can support multiple different standard interfaces and configurations and at the same time allow for the creation of differentiated IP. COTS solutions and more specifically ATCA based products have seen significant growth as they have increasingly been adopted as the ideal fit to meet these challenges. There have been a number of approaches to building flexibility into an ATCA based solution. None have quite hit the right combination of flexibility, linked with the ability to maximize network capacity. None at least until this year, and Advantech's introduction of the FMM.

Fabric Mezzanine Modules

Advantech's Fabric Mezzanine Modules or FMMs have been creating a lot of buzz since the concept was first publicly introduced earlier in the year. The first paper in this series (Fabric Mezzanine Modules bring unprecedented I/O and acceleration flexibility to Intel® Xeon® ATCA Blades) discussed the fundamentals of the FMM, however, here is a brief recap.

- A Fabric Mezzanine Module is extremely compact at just 7 x 7.5 cms. This provides more than adequate space to fit 40 x 40 mm BGA ASICs and FPGAs along with associated components.
- FMC compliant connectors are used for high speed differential I/O that can exceed 10 Gbps.
- A thermal budget of up to 20W gives plenty of headroom and with the location and orientation on an ATCA blade offers improved cooling over an AMC.
- The I/O area has been designed to provide sufficient overhang for connector support on front panels or rear transition modules.
- A plethora of mass customization options are available. There are already a number of unique FMM options, however, as Advantech openly shares the specifications with customers this allows for the simple integration of custom IP.

- Two 'Types' of FMM have been defined. Type I modules do not have a faceplate and are mounted somewhere internally on a blade. Type II modules do have a faceplate and support the addition of I/O connectors.
- Due to their simplicity and compact size FMMs can easily be designed onto

ATCA Blades and RTMs (Rear Transition Modules) allowing for a wide variety of connectivity options.

- PCI Express (PCIe) x16 is used as the high speed interface to the base board. This may be used as either a full 16 lane interface or split into 2 x8 groups.

How Much Throughput Would You Like?

There are a multitude of application examples that led to the development of the FMM concept. The majority of these applications all have one thing in common and that is the need for high levels of system throughput and connectivity. To say the demand for high speed data is going through the roof is an understatement. Whether we are at home, work or on the move we expect access to all our data. The move towards cloud based infrastructure has made life easier for the consumer but only makes the headache worse for those that must build and operate the networks. Everything needs to be ramped up, not just the size of the data pipes but with a huge mix of data types and services running across the same network, traffic management is crucial. Let's also not forget the security angle. For public or private cloud data centers, creating that all important protective barrier, for their users, must be accomplished providing the highest levels of security while not impacting network throughput. It is no surprise then that when Advantech asks its customers "how much system throughput would you like?" the answer is typically "as much as you can give us!"

As external data requirements increase, one needs to be able to match this in terms of the internal switching bandwidth of a network system and at the same time manage the appropriate levels of ingress and egress. Designing any networking system platform, therefore, requires there to be a balance between the amount of external ports and the backplane switching technology, but let's not forget the processing requirement. The types of network systems we are discussing are much more than just routers. Their reason for being is to provide increased levels of network intelligence, identifying threats and/or traffic flows such that appropriate action may be taken.

Workload Consolidation and the Evolving Network Platform

These types of networking system platforms had, in the past, typically been built using a combination of network processors (NPUs) and general purpose processors. The specialist features of the NPU were relied on to handle the networking functions while the GPPs managed the overall operating system, the application and management interfaces. This division of labor made perfect sense although it did create challenges in the areas of reuse, cost and performance. When designing a system one had to utilize multiple different types of blades that were not interchangeable and could not all be used to help load balance the platform, thereby limiting the performance envelope of a given system configuration. The latest iterations of Intel® architecture (IA) processors have significantly increased performance levels and have also been designed with networking operations in mind. This now means that it is more than possible to consolidate system workloads using a single IA blade architecture. Being able to use the identical blade for both networking and application processing translates into reduced CAPEX and OPEX e.g. it removes the need for duplicate sparing. It also means that a system architecture can be utilized that allows workload to be balanced across all blades no matter what type of processing is required. IA based platforms that allow for workload consolidation are undoubtedly the way forward. To derive the highest benefit from such an approach one

must be able to configure the same IA blades for different levels of throughput and with multiple connectivity options. That brings us neatly back to the Fabric Mezzanine Module.

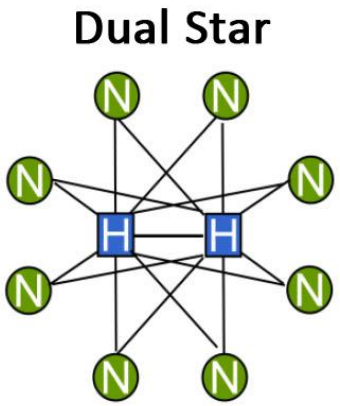
FMM vs. AMC

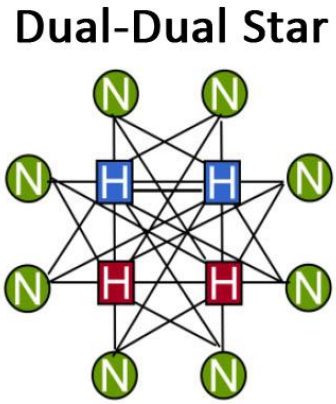
Given that this discussion is grounded in an ATCA COTS environment, one might ask why use an FMM and not the already defined AMC (Advanced Mezzanine Card)? It is a good question but one that has an equally good answer. AMCs do represent the obvious suitable open standard, however, an AMC slot eats a lot of board real estate, adds to cost, and all you can use it for on an x86 blade is some "simple" PCIe based I/O because otherwise you start to run into issues around thermal design and airflow. Our earlier discussion surrounding the need to generate as much throughput as possible raises another challenge for AMCs as most can only support a PCIe x4 interface. Increased cost, loss of board real estate and challenging thermal issues, and all for just a PCIe x4 socket, that doesn't seem like a good ROI equation. To be able to generate the desired levels of functionality, flexibility and performance the FMM is the only way to go.

Connectivity was a key criterion with the design of the FMM. The connection between the FMM and the base board or carrier is via 16 PCIe lanes which can be used as one PCIe x16 or two PCIe x8 ports. On top of that, there are up to 6 high-speed ports, each consisting of 4 SerDes lanes routed back to the carrier along with some "utility" I/O such as USB, SATA, and SGMII.

ATCA System Throughput with Dual, Dual Star Topologies

Ethernet fabric backplanes have long since been an integral part of the PICMG specification for AdvancedTCA. Leveraging the base IEEE 802.3 standard, PICMG created a series of specifications that define options for switched fabric backplanes. The first implementations were only 1Gbps although today 10Gbps fabrics are common among the leading blade and system vendors. 40Gbps backplanes were the next stepping point with early solutions utilizing 4 aggregated 10GBaseKR ports to reach 40G. Advantech has such solutions, however, using different FMM options on one of Advantech's latest ATCA blades one can elect to install a Mellanox CX-3 based FMM that supports two 40GbE ports implementing the 40GBaseKR4 standard.





The speed of the fabric is only part of the story as ATCA has an advanced topology that allows for the potential doubling of the total system throughput.

Standard ATCA systems incorporate two switch blades or hubs in a chassis and each of the node blades (typically up to 10) connects to each of the hub blades. In a topology diagram, the switch blades, or hub blades as they are called in the ATCA world, are at the center of a star, surrounded by the node blades. As there are two hub blades, there are two stars. This fabric topology is called “dual star”.

As IA architected ATCA blades have increased in performance and capacity the switch blades are now becoming the new bottleneck. For example the Advantech MIC-5332 and MIC-5333 blades that are based on dual Intel® Xeon® E5-2600 processors can easily handle 40Gbps of traffic. With 10 such blades in a system you could build a 400Gbps capable networking application platform. Life is never quite that simple. As state of the art switching silicon saturates around 600Gbps, the system’s switching capacity is limited to just 1.2Tbps. That is quite a big ‘just’ but with a 400Gbps platform that means that packets may hop through the switches a maximum of three times. As two hops are just consumed for ingress and egress traffic, three hops total present a quite big challenge.

The good news is that the ATCA specification already has a solution. The specification defines an extended topology with the ability to use four switch blades. Four switches means four stars, although it is called a dual-dual star topology not, as you might think, a quad star topology. Now with four fabrics we have 4x 600Gbps that yields 2.4Tbps switching capacity or 6 hops average for a packet in a 400Gbps system. This all sounds great, however, now each node blade needs to have four network ports – one to each hub blade.

This now brings us right back to the Advantech MIC-5332 and MIC-5333 ATCA blades both of which can support dual-dual star implementations. It is through the use of the Advantech FMM concept that these blades can support eight network ports to the backplane.

The MIC-5322 already has one Niantic down on the blade supporting 2 x 10GE ports to the backplane and using an FMM-5001B on can configure an additional Niantic to drive two additional backplane ports. The MIC-5333 has backplane ports that are connected to two FMM sites which means that a variety of FMMs can be used to implement 10GE or 40GE backplane fabric interfaces.

With 2 x FMM-50004M each of which has a Mellanox CX3 one can now drive 4 x 40Gbps across the backplane making for an aggregate 1.60Tbps solution with ten MIC-5333s. Advantech are proud to say that this configuration represents the only COTS IA blade solution that can support dual-dual star networking and offer the highest levels of network throughput available today.

Maximizing Ingress/Egress

With all the switching and network and application processing capacity let’s remember that we still must get the traffic in and out of the platform. Along with the ports built into the hubs the Fabric Mezzanine Module plays an important role here as well. If we look at a well configured Advantech ATCA system platform it would contain the following:

- 4 Switch blades e.g. T-HUB4 each with 8x10G and 2x40G ports. This is 160Gx4= 640G
- 10 MIC-5333s – Using a dual Niantic FMM each MIC-5333 can support 2x10G, times 10 blades = 200G
- Each MIC-5333 has a corresponding RTM (Rear Transition Module). Advantech’s RTM-5104 can support another 2 x10G and there is another RTM under development that can support 8x10G=80G, times 10 = 800G

All these ports combined make for a total ingress/egress capacity of up to 1.64TB. As new

FMMs are made available and/or custom modules are created this total can be increased further.

It is ultimately the combination of Advantech's FMMs and latest Intel® based ATCA blades that enables this new flexible design environment. Let's take a minute to review the features and capabilities of Advantech's two industry leading Intel® Xeon® E5-2600 Series based blades.

Advantech Dual Intel® Xeon® E5-2600 Series Processor ATCA Blades

MIC-5332

Advantech's MIC-5332 is a dual processor ATCA blade based on the Intel® Xeon® E5-2600 Processor. It enables the highest performance available in an ATCA form factor with up to 16 cores and 32 threads of processing power, fast PCI Express gen. 3 lanes running at up to 8Gbps and best in class virtualization support. Two QPI interfaces between the CPUs improve memory and I/O access throughput and latencies when one processor needs to access resources hosted by the other socket. With four DDR3 DIMMs per socket in a quad channel design running up to 1600MT/s, the MIC-5332 not only offers superior memory bandwidth over 3-channel designs, but can also support memory densities up 256GB using latest LR DIMM technology. It outperforms previous generation dual socket designs while keeping similar thermal characteristics with balanced airflow resistance. Using Intel's latest PCH with its integrated 4-port SAS controller, the need for an external storage controller is eliminated making the MIC-5332 an ideal choice for cost sensitive control plane applications. While supporting two 10GBaseKX4 interfaces in the base model, support for dual-dual star fabric implementations can be added by installing an FMM-5001, the first of Advantech's FMM designs. The FMM type II socket with PCIe x16 connectivity provides extension possibilities for additional front port I/O, offload and acceleration controllers such as Intel® QuickAssist accelerators, IPSec offload engines or customer specific logic. This unmatched flexibility combined with the highest performance Intel® Xeons available make the MIC-5332 equally well suited for application and data plane workloads.

The MIC-5332 supports hot-swappable RTMs such as the RTM-5104 for High Availability (HA)

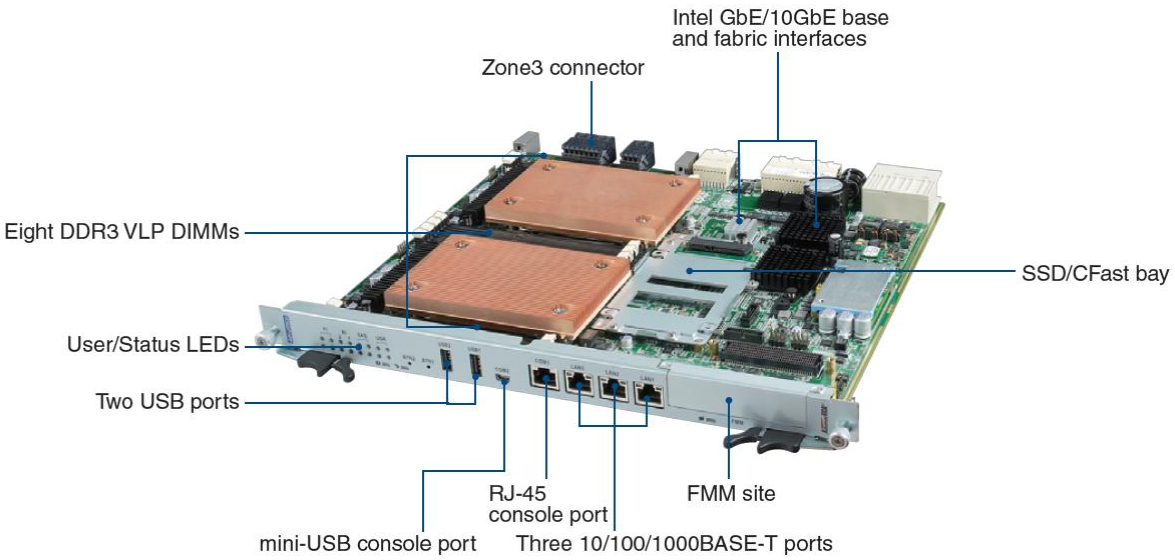


Figure 1. MIC-5332

needs, rear I/O and dual SAS storage with RAID as well as an optional FMM type II socket for further I/O and offload as shown in Figure 1.

MIC-5333

Advantech's MIC-5333 is based on the Intel® platform for communications infrastructure formerly known as "Crystal Forest." In addition to it's two high performance dual Intel® Xeon® E5-2600 Series processors, the onboard Intel® Communications Chipset 89xx Series PCH incorporates acceleration and offload features for encryption and enhanced security.

One of the unique advantages of the MIC-5333 is the increased number of available FMM sites from one to three. With two FMM Type I sites connected to the Fabric Interface and one FMM type II connected to the front panel, the flexibility in blade function personalization is extensive.

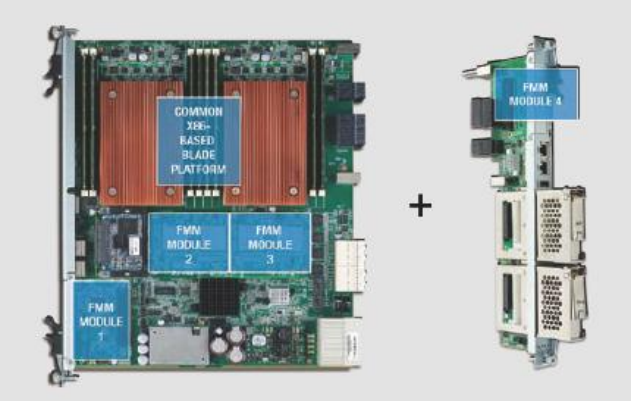


Figure 2. MIC-5333 and RTM-5104 with a total of 4 FMM sites

When coupled with the RTM-5104 Rear Transition Module shown in Figure 2., a fourth FMM site is available, making the standard COTS front ATCA blade and RTM ideal for mass customization with an abundance of I/O and acceleration options. The RTM-5104 is a single slot (6HP) ATCA rear transition module for I/O extension of Advantech ATCA CPU blades.

FMM-5001B
Intel® 82599ES with
2 x 10GBaseKR
Fabric Interface

FMM-5001F
Intel® 82599ES for
2 x 10GbE with dual SFP+
output

FMM-5001Q
Quad Intel® 82599ES with
8 x 10GBaseKR
Fabric Interface

FMM-5002
Servers Graphics Controller
with VGA connector

FMM-5004M
Mellanox CX3 with
2 x 40GBaseKR4
Fabric Interface

FMM-5006
QuickAssist Acceleration with
Intel® Communications
Chipset 89xx Series

FMM Availability continues to grow

The simple short form FMM interface specification has allowed Advantech and several OEMs to develop various FMMs in an extremely short period of time. We have discussed a number of the available options in this paper and you can see these, along with some other interesting examples in Figure 3. As we discussed, the initial modules were designed to provide maximum fabric flexibility for Advantech's growing range of Intel® Xeon® E5-2600 blades. The FMM-5001B provides dual 10GbE connectivity to ATCA backplanes using Intel® 82599 LAN controllers while dual 40GbE connectivity is enabled using the Mellanox C3 based FMM-5004M. Also available is an Intel® QuickAssist Accelerator module offering offload and acceleration for encryption and a GPU-based FMM is available with VGA graphics output.

Flexibility is an often discussed attribute of the design process but typically takes a back seat to performance and functionality. With the introduction of Advantech's FMM concept you truly can have your cake and eat it too as Advantech delivers highly functional ATCA blades, industry leading performance and the flexibility to personalize configurations to suit a multitude of applications. Early design wins have already proven how the FMM has risen to meet and exceed the expectations of customers.

If you are looking for an innovative solution for your next design, take a hard look at what an Advantech FMM could do for you. If we don't already have one that works for you we can share the short form specification and show how easy it is to create an FMM that will fit your needs like a glove. Fabric Mezzanine Modules by Advantech - Flexibility Redefined.

Figure 3. Example Fabric Mezzanine Modules

Advantech Contact Information

Hotline Europe: 00-800-248-080 | Hotline USA: 1-800-866-6008
Email: NCG@advantech.com.tw
Regional phone numbers can be found on our website at **http://www.advantech.com/contact/**

www.advantech.com/nc

ADVANTECH

Enabling an Intelligent Planet

Enabling an Intelligent Planet