The **BLAZAR Family of Accelerator Engines** are a family of high capacity, high-speed memories that support high bandwidth, fast random memory access rates and includes the BE-2 and BE-3 products with In-Memory Functions (IMF). The PHE which is a next generation device is a 1Gb Memory (equivalent to BE-3) with the addition of 32 RISC processors which allow full custom algorithms to be processed on the same silicon as the associated high speed memory. By combining the power of programmable cores with internal memory bandwidth that allows for up to 1.5Tb of memory bandwidth all within the one 27 x27 mm package allows a system architect to place significant compute power wherever it is needed for maximum performance benefit.

The MoSys PHE is supported with a full Integrated Development Environment (IDE) that is Linux based and supports a full assembler and debugger.

**BASE FEATURES**

The Programmable HyperSpeed Engine (PHE)
- **1Gb of tRC of 2.7 ns memory**
  - 4 partitions of 4M x 72
  - A partition has 128 Banks of Memory
- **In-Memory User Programmable functions**
  - 32 RISC core processor with instruction memory and data registers
  - Multi-threaded
- **In-Memory Fixed Intelligent functions**
  - BURST functions (same as BE3-RMW)
  - RMW functions (same as BE3-RMW)

**APPLICATIONS FOCUS**

- Algorithm based applications that benefit from localized high-performance RISC computing
- High bandwidth data acquisition and manipulation
- Security (Stateful Firewall, Anomaly Detect, DOS)
- High bandwidth data access
- Networking (TCAM, LPM, Statistics)
- Genomics (pattern match)
- Compute acceleration (Database, KVS, …)
- ML (Decision Trees, Random Forest, …)
- Video (Row/Column manipulation)
- FPGA Acceleration for Xilinx and Intel

**KEY FEATURES - Memory**

- **1 Gb SRAM (16M x 72b)**
- High Bandwidth, low pin count serial interface
  - Highly optimized, efficient and reliable serial transport command and data protocol
  - Large high access rate SRAM class memory (sub 3ns tRC)
    - Up to 6.5 Billion transactions/sec
- Internal RMW Macro functions for statistics, metering, filtering, and atomic operations
- 32 Programmable cores for in-memory search and compute
- Eases board layout and signal integrity, -minimal trace length matching required, operates over connectors
  - Reduction of I/O pins from 5x to 45x depending on equivalent memory density and type

**KEY FEATURES – In-Memory Functions**

A list of Acceleration options are available in every PHE device

**BURST** In-Memory Function
- For sequential Read or Write functions for DATA MOVEMENT
- Burst length: 1, 2, 4 or 8 words

**RMW** In-Memory Function
- RMW are Read/Modify/Write functions
- Includes may functions for COMPUTE and DECISION
- Examples: ADD, SUB, Compare, INC plus 15 other functions
- Increase execution speed and bandwidth

**USER PROGRAMMABLE** In-Memory functions
- Embedded functions or algorithms
**PROGRAMMABLE HYPERSPEED ENGINE (PHE) ARCHITECTURE**

Acceleration Engines give Software and Hardware System Architects Acceleration Options not previously available

---

### High speed serial I/O
- GCI serial I/O versions of 12.5, 15.6 and 25 Gbps for high bandwidth (up to 640Gbps)
- Device can operate with a minimum of 4 lanes.
- Has two, full duplex 8 lane ports that operate independently. Device can be used as a dual port memory
- Reduces number of signal pins over traditional memories, increases signal integrity allowing longer board traces to ease board signal routing
- Operates across connectors

---

### Memory/Function Controller
- Directs read/write function execution to selected bank of memory
- Manages the sequence of all In-Memory Functions
- Directs User Defined Function execution
- Multiple scheduling domains minimize blocking short latency operation by long latency In-Memory function operations
- Controls parallel function execution

---

### ALU
- Embedded RMW Functions utilize ALUs for in memory computational functions
  - 32 ALUs
  - Simultaneous operations

---

### Main Memory
- 1Gb
  - 4 partitions/128 banks
  - 16 READ & 16 WRITE ports
  - From 0.67 to 3.2 ns tRC
  - Allows parallel partition & Bank execution
  - Up to 6B rd/s + 6B wr/s simultaneously (internal)

---

### Result Reordering
- Parallel Processing Reorder buffers ensure that results are returned in same order as commands are received

---

### RISC Core CLUSTER ARRAY
- Comprised of 8 Clusters (total of 32 PE RISC processors & memory with 12Bt/s)
- A Cluster has
  - 4 PE RISC processors
  - 4k x 72b memory @ 1.5B transactions/sec
- PE RISC Processor, 1.5 GHz
  - 1k x 72b Instruction memory
  - 128 x72b Internal Registers
  - 8 threads (total of 256 threads managed by PHE device)
- Location of User Defined In Memory Compute algorithms or Bandwidth Functions (IMF)
**PHE CLUSTER / CLUSTER ARRAY ARCHITECTURE**

**CLUSTER ARCHITECTURE**

A Cluster has 4 RISC Processor Engines (PE) and Data Memory

- Each RISC Processor Engine
  - 1.5 GHz
  - 128 Internal Registers (IR)
  - 1k x 72b Instruction Memory (IM)
  - Optimized ISA instruction set

The 4 PEs are connected to Data Memory by a cross point switch

1. Random Access Data Memory
   - 4k x 72b
2. Connected to the 4 Cluster PEs with a cross point switch
   - 0.67ns tRC to the 4 PEs in the Cluster
   - Allow simultaneous access by all four PEs

**CLUSTER ARRAY ARCHITECTURE**

Eight (8) Clusters are combined to form the Cluster Array

Base architecture of the PHE is the same as the BE3-RMW (Reference BE3-RMW Product Brief)

1. One of eight Clusters
2. Memory Controller with Scheduler is extended to manage the algorithms that are executed in the Cluster Array
   - Domain priority supports 4 levels
   - 256 thread management
3. Cross point switch connects all cluster memories together.
   - Any PE can access all of the combined Cluster Array memory
   - Allows simultaneous access by all PE to all memory
4. Cross point switch allows transfer to/from any main memory to/from and Cluster Memory
ACCELERATING FPGA APPLICATIONS

With more and more demand on performance and the increasing need for additional features, an FPGA’s resources can easily be consumed.

The Blazar PHE gives the hardware/software architects many options to speed application performance or create unique system architectures that execute faster.

Options to increase the performance of an FPGA design with BLAZAR Engines:
- Offload as many functions as possible
- Simplify software by moving complex functions or algorithms that consume execution time and RTL
- Take advantage of the High-Speed Serial Interface to the high random-access rate 1Gb memory (tRC 2.7 ns device dependent)

Power of the PE Instruction Set

The instruction set is very powerful and versatile. It allows the programming of simple operations to very powerful functions and algorithms.

The MoSys applications team is ready to help you define the best utilization of the FPGA and PHE resource to Accelerate your application.

“Software Defined... Hardware Accelerated”

Candidate FPGA tasks to move to the PHE

The 32 RISC cores provides flexibility for the System Architect to sort tasks between hardware RTL and software tasks

- Simplify
  - High usage RTL
- Common system tasks
- Complex algorithms/functions
  - Application specific
  - Requiring higher throughput
  - Can utilize multiple copies operating in parallel
  - Requiring complex RTL designing
- General tasks
  - Need more memory than FPGA on-chip can support
  - Memory access bottlenecked tasks
  - Time consuming memory sequential tasks
  - Time & sequence critical tasks
- Speed increase
  - Parallel processing (32 processors)
    - Optimized ISA & 256 threads
    - Higher access rates, lower latency, higher bandwidth
    - Utilize Engine scheduler to optimize execution
    - Install multiple copies of same algorithm/functions and scheduler will find available processors
SDK sends in GCI traffic with Python calls.

HDK sends in GCI traffic with Verilog pins on a Verilog module.

Development Steps:
1. Single Thread Debug on Behavioral Model
2. Multi Thread Debug on Behavioral Model
3. Performance Tuning w/ Behavioral Model
Code Development Environment

Development Steps

1. Single Thread Debug on Behavioral Model
2. Multi Thread Debug on Behavioral Model
3. Performance Tuning w/ Behavioral Model
4. Performance Benchmark w/ NPS Model
5. Multi Thread Validation w/ PSE + NPS

Integrated Development Environment (IDEm)

Development Steps

1. Single Thread Debug on Behavioral Model
2. Multi Thread Debug on Behavioral Model
3. Performance Tuning w/ Behavioral Model
4. Multi Thread Validation w/ PSE + NPS
5. Performance Benchmark w/ PSE + NPS
# Accelerator Engine Family Overview

Software Defined - Hardware Accelerated

<table>
<thead>
<tr>
<th>In-Memory</th>
<th>Part Number</th>
<th>Description</th>
<th>Package</th>
<th>Interface</th>
<th>Memory</th>
<th>Access Rate</th>
<th>In-Memory Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>BURST</td>
<td>MSR822</td>
<td>Bandwidth Engine 2 Burst Serial 0.5Gb High Access Memory</td>
<td>FCBGA 19x19</td>
<td>16</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>MSR830</td>
<td>Bandwidth Engine 3 Burst Serial 1Gb High Access Memory</td>
<td>FCBGA 27x27</td>
<td>16</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>RMW</td>
<td>MSR820</td>
<td>Bandwidth Engine 2 RMW Serial 0.5Gb High Access Memory with ALU for RMW functions</td>
<td>FCBGA 19x19</td>
<td>16</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>MSR830</td>
<td>Bandwidth Engine 3 RMW Serial 1Gb High Access Memory with ALU for RMW functions</td>
<td>FCBGA 27x27</td>
<td>16</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Program</td>
<td>MSP530</td>
<td>Programmable HyperSpeed Engine Serial Interface, 1Gb Memory, 32 RISC Processor cores for custom algorithms, compute, functions</td>
<td>FCBGA 27x27</td>
<td>16</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>RTL-AE</td>
<td>RTL</td>
<td>RTL Memory Controller for Bandwidth Engine and Programmable HyperSpeed Engine. Manages memory and the serial interface signals. Presents a QDR like parallel RTL interface to the user.</td>
<td>FPGA RTL Code</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

BW MAX. = At the highest serial interface speed

MoSys Provides Reference Designs to Speed Customer Design Process

Memory Controller & App Functions
GCI PCS / Framer

CONTACT MOSYS TO LEARN ABOUT THESE ADVANCED FEATURES

[www.mosys.com](https://www.mosys.com)
[https://mosys.com/blazar-family-of-accelerator-engines/]

MoSys is a registered trademark of MoSys, Inc. in the US and/or other countries. Blazar, Bandwidth Engine, HyperSpeed Engine, IC Spotlight, LineSpeed and the MoSys logo are trademarks of MoSys, Inc. All other marks mentioned herein are the property of their respective owners.