

General Description

Versal devices are the industry's first adaptive compute acceleration platforms (ACAP), combining adaptable processing and acceleration engines with programmable logic and configurable connectivity to enable customized, heterogeneous hardware solutions for a wide variety of applications in Data Center, Automotive, 5G Wireless, Wired, and Defense. ACAP devices feature transformational features like an integrated silicon host interconnect shell and Intelligent Engines (AI and DSP), Adaptable Engines, and Scalar Engines, providing superior performance/watt over conventional FPGAs and GPUs.

AI Core Series: The high-compute series with medium density programmable logic and connectivity capability coupled with AI and DSP acceleration engines.

Prime Series: The mid-range series with medium density programmable logic, signal processing, and connectivity capability.

Series Comparisons

Table 1: Device Resources

ACAP Resources and Capabilities	AI Core Series	Prime Series
NoC	✓	✓
Aggregate INT8 TOP/s	49–147	3–27
System Logic Cells (K)	540–1,968	352–2,154
Hierarchical Memory (Mb)	68–191	40–324
DSP Engines	928–1,968	472–3,984
AI Engines	128–400	–
Processing System	✓	✓
Serial Transceivers	8–44	12–66
Max. Serial Bandwidth (full duplex) (Tb/s)	2.9	4.2
I/O	424–770	316–856
Memory Controllers	2–4	1–6

Summary of Features

Architecture

Versal ACAPs are built around an integrated shell composed of a network on chip (NoC), which enables seamless memory-mapped access to the full height and width of the device. ACAPs comprise: a multicore scalar processing system (PS); interconnect for CCIX/PCIe (CPM); SIMD VLIW AI Engine accelerators for artificial intelligence and complex signal processing; and Adaptable Engines in the programmable logic (PL). Together, these form the shell—a platform for fast time-to-market (TTM) compute acceleration for cloud, edge, and networking applications. The platform management controller (PMC), adjacent to the PS, is responsible for booting and configuring the device from data stored in an off-chip flash memory device. Versal devices typically have I/O and memory controllers on the north and south edges of the device and serial transceivers on the east and west edges. The NoC spans full height and width of the device.

Compute and Acceleration

The Versal AI Core series has an array of signal processing cores that are highly optimized for functions in machine learning, convolutional neural networks, wireless radio, backhaul, cable, and radar applications. The array consists of a number of AI Engines, each comprising a 32-bit scalar RISC processor, fixed and floating point vector units, data memory, and interconnect. AI Engines can be used as a single tile, as the complete array, and any granularity in between. The creation of custom acceleration and compute engines in the AI Engine array is done at high-level through C and C++.

Every Versal ACAP has Scalar Engines that comprise a dual-core Arm Cortex-A72 (APU) and a dual-core Arm Cortex-R5 (RPU) in the PS. The PS includes a number of peripherals for communication standards, including gigabit Ethernet and USB 2.0, and controllers for SPI, I2C, UART, and CAN-FD. The PS accesses the DDR memory controllers on the top and bottom of the device through the NoC. In addition to interfacing to external memory, the APU includes: Level-2 (L2) cache; the RPU includes tightly coupled memory (TCM); and both APU and RPU have access to the on-chip memory (OCM).

The programmable logic (PL) is made up of configurable logic blocks, containing 6-input look-up tables (LUTs) and flip-flops; different-sized memory blocks; 36Kb block RAM and 288Kb UltraRAM; digital signal processing (DSP) blocks; and a wealth of interconnect, switches, and muxes to connect blocks together. All resources are arranged in columns. The PL is divided into regions that are a fixed height. Each region has its own clocking capabilities and NoC access points.

Platform Management

The PMC resides adjacent to, but is independent from, the PS. It is responsible for the boot and configuration of the device from the primary boot source. The PMC is also responsible for configuring the PL, which can be configured before or after the PS. It also controls encryption, authentication, system monitoring, and device debug capabilities of the ACAP.

Connectivity

The north and south edges of Versal ACAPs typically contain a number of XPIO banks and associated memory controllers to read from and write to DDR4 and LPDDR4 memory. XPIO can be used independently from the dedicated memory controllers for many functions, including any with soft memory controllers created in the PL. The east and west edges of the device typically contain serial transceivers capable of communicating up to 58Gb/s.

Feature Summary

Table 2: Versal AI Core Series

	VC1352	VC1502	VC1702	VC1802	VC1902
AI Engines	128	217	310	300	400
AI Engine Data Memory Blocks	1,024	1,736	2,480	2,400	3,200
AI Engine Data Memory (Mb)	32	54.25	77.5	75	100
DSP Engines	928	1,312	1,272	1,600	1,968
System Logic Cells	539,840	797,440	1,020,880	1,585,938	1,968,400
CLB Flip-Flops	493,568	729,088	933,376	1,450,000	1,799,680
LUTs	246,784	364,544	466,688	725,000	899,840
Distributed RAM (Mb)	7.5	11.1	14.2	22.1	27.5
Block RAM Blocks	499	547	826	800	967
Block RAM (Mb)	17.5	19.2	29.0	28.1	34.0
UltraRAM Blocks	151	215	402	325	463
UltraRAM (Mb)	42.5	60.5	113.1	91.4	130.2
Accelerator RAM Blocks	1	0	1	0	0
Accelerator RAM (Mb)	32	0	32	0	0
APU	Dual-core Arm Cortex-A72; 48KB/32KB L1 Cache w/ parity and ECC; 1MB L2 Cache w/ ECC				
RPU	Dual-core Arm Cortex-R5; 32KB/32KB L1 Cache, and TCM w/ECC				
Memory	256KB On-Chip Memory w/ECC				
Connectivity	Ethernet (x2); UART (x2); CAN-FD (x2); USB 2.0 (x1); SPI (x2); I2C (x2)				
NoC Master / Slave Ports	10	14	18	28	28
DDR Bus Width	128	128	128	256	256
DDR Memory Controllers	2	2	2	4	4
CCIX & PCIe (CPM)	–	1 x Gen4x16, CCIX	–	1 x Gen4x16, CCIX	1 x Gen4x16, CCIX
PCI Express	1 x Gen4x8	4 x Gen4x8	1 x Gen4x8	4 x Gen4x8	4 x Gen4x8
Multirate Ethernet MAC	1	4	3	4	4
XPIO	378	378	378	648	648
HDIO	44	44	44	44	44
GTY Transceivers (32.75Gb/s)	8	44	24	44	44

Table 3: Versal AI Core Series: Device-Package Combinations and Maximim I/O

	VC1352	VC1502	VC1702	VC1802	VC1902
	XPIO, HDIO, MIO, GTY	XPIO, HDIO, MIO, GTY	XPIO, HDIO, MIO, GTY	XPIO, HDIO, MIO, GTY	XPIO, HDIO, MIO, GTY
VBVA1024	378, 22, 78, 8	378, 22, 78, 8			
VSVE1369	378, 44, 78, 8		378, 44, 78, 24		
VSVA1596		378, 44, 78, 32	378, 44, 78, 16		
VIVA1596				378, 44, 78, 32	378, 44, 78, 32
VSVD1760				648, 44, 78, 24	648, 44, 78, 24
VSVA2197		378, 44, 78, 44		648, 44, 78, 44	648, 44, 78, 44

Table 4: Versal Prime Series

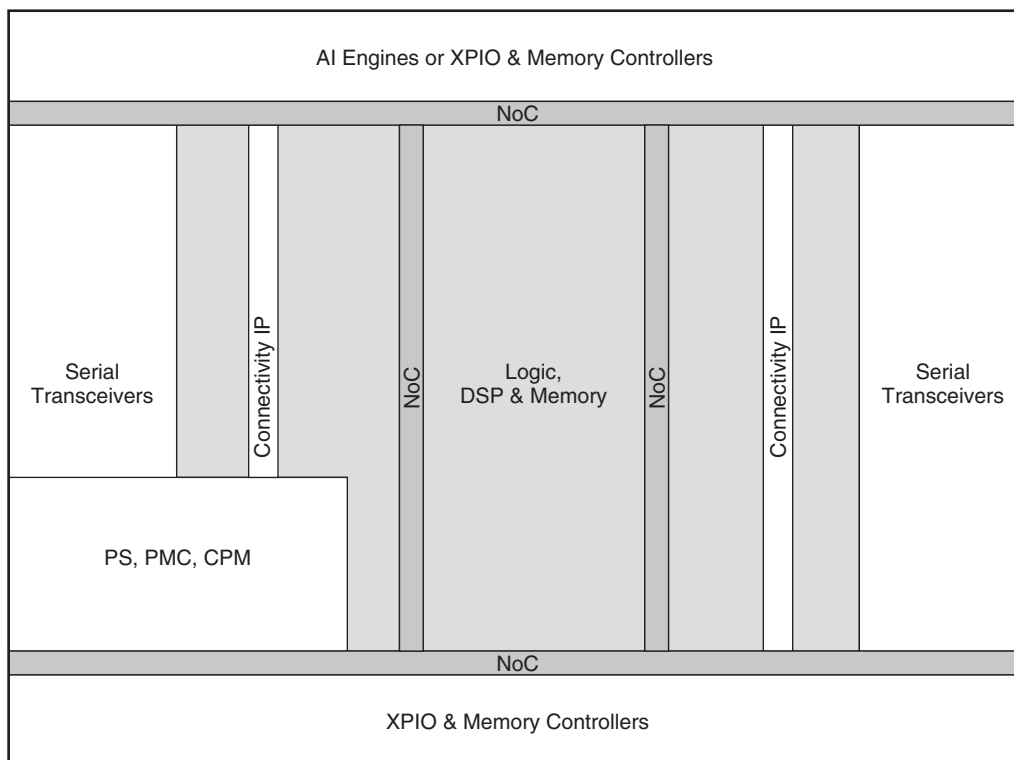
	VM1102	VM1302	VM1402	VM1502	VM1702	VM1802	VM2502	VM2602	VM2702	VM2902
System Logic Cells	352,240	571,760	1,001,840	799,440	1,662,640	1,968,400	2,029,720	1,263,360	1,804,688	2,153,760
CLB Flip-Flops	322,048	522,752	915,968	729,088	1,520,128	1,799,680	1,855,744	1,155,072	1,650,000	1,969,152
LUTs	161,024	261,376	457,984	364,544	760,064	899,840	927,872	577,536	825,000	984,576
Distributed RAM (Mb)	4.9	8.0	14.0	10.7	23.2	27.5	28.3	17.6	25.2	30.0
Block RAM Blocks	216	452	1,124	547	2,856	967	1,365	1,552	2,100	2,572
Block RAM (Mb)	7.6	15.9	39.5	19.2	100.4	34.0	48.0	54.6	73.8	90.4
UltraRAM Blocks	96	168	168	215	228	463	701	424	600	724
UltraRAM (Mb)	27.0	47.3	47.3	60.5	64.1	130.2	197.2	119.3	168.8	203.6
DSP Engines	472	736	1,504	1,312	2,096	1,968	3,984	1,880	2,500	3,080
APU	Dual-core Arm Cortex-A72; 48KB/32KB L1 Cache w/ parity and ECC; 1MB L2 Cache w/ ECC									
RPU	Dual-core Arm Cortex-R5; 32KB/32KB L1 Cache, and TCM w/ECC									
Memory	256KB On-Chip Memory w/ECC									
Connectivity	Ethernet (x2); UART (x2); CAN-FD (x2); USB 2.0 (x1); SPI (x2); I2C (x2)									
NoC Master / Slave Ports	5	16	16	14	22	28	28	16	26	26
DDR Bus Width	64	128	256	128	128	256	288	384	384	384
DDR Memory Controllers	1	2	4	2	2	4	5	6	6	6
CCIX & PCIe (CPM)	–	–	–	–	1 x Gen4x16, CCIX	1 x Gen4x16, CCIX	1 x Gen4x16, CCIX	1 x Gen4x16, CCIX	1 x Gen4x16, CCIX	1 x Gen4x16, CCIX
PCI Express	1 x Gen4x8	2 x Gen4x8	2 x Gen4x8	2 x Gen4x8	4 x Gen4x8	4 x Gen4x8	1 x Gen4x8	1 x Gen4x8	2 x Gen4x8	2 x Gen4x8
Multirate Ethernet MAC	1	2	2	4	2	4	1	2	2	2
XPIO	216	432	648	378	378	648	702	756	702	702
HDIO	22	44	44	44	22	44	44	22	44	44
GTY Transceivers (32.75Gb/s)	12	24	24	44	32	44	16	20	32	40
GTM Transceivers (58Gb/s)	0	0	0	0	0	0	28	32	44	52

Table 5: Versal Prime Series: Device-Package Combinations and Maximum I/O

	VM1102	VM1302	VM1402	VM1502	VM1702	VM1802	VM2502	VM2602	VM2702	VM2902
	XPIO, HDIO, MIO GTY, GTM	XPIO, HDIO, MIO GTY, GTM	XPIO, HDIO, MIO GTY, GTM	XPIO, HDIO, MIO GTY, GTM	XPIO, HDIO, MIO GTY, GTM	XPIO, HDIO, MIO GTY, GTM	XPIO, HDIO, MIO GTY, GTM	XPIO, HDIO, MIO GTY, GTM	XPIO, HDIO, MIO GTY, GTM	XPIO, HDIO, MIO GTY, GTM
SFVBB625	216, 22, 78 4, 0									
VFVB1024	216, 22, 78 12, 0	216, 44, 78 16, 0	324, 44, 78 16, 0		324, 22, 78 16, 0					
VFVB1369		216, 44, 78 24, 0	324, 44, 78 24, 0	324, 44, 78 24, 0	378, 22, 78 32, 0					
VFVA1760		432, 44, 78 24, 0	648, 44, 78 24, 0					756, 22, 78 20, 0		
VFVC1760				378, 44, 78 44, 0	378, 22, 78 24, 0	378, 44, 78 44, 0		378, 22, 78 20, 32	378, 44, 78 24, 32	378, 44, 78 24, 32
VSVD1760						648, 44, 78 24, 0				
VSVA2197						648, 44, 78 44, 0	648, 44, 78 16, 16			
VSVA2785							702, 44, 78 16, 28	702, 22, 78 20, 32	702, 44, 78 32, 44	702, 44, 78 40, 52

Device Layout (Architecture and Interconnect)

Versal devices are built from a library of building blocks dedicated to processing, compute, acceleration, and connectivity. Figure 1 shows the layout of an ACAP with the NoC connecting to the host processor via CPM and the various heterogeneous processing elements: programmable logic (PL), vector-based accelerators (AI Engines), and scalar processing accelerators.



DS950_01_102618

Figure 1: ACAP Device Layout

Serial transceivers are located on the east and west edges of the device with XPIO and memory controllers on south and north of the device. In the Versal AI Core Series, there is an acceleration array on the north edge of the device in place of the XPIO and memory controllers. Connectivity IP is located in columns close to the serial transceivers. Resources are connected together through a matrix of programmable interconnect routes for local and regional signal connectivity as well as the NoC for high bandwidth and long distance communication around the device.

NoC

The NoC is an AXI-4 based network of interconnect within the ACAP architecture that easily enables high-bandwidth connections to be routed around the device. The NoC extends in both horizontal and vertical directions to the edges of the device. It exists to connect together areas of the device that demand and use large quantities of data alleviating any resource burden on the local and regional device interconnect. The NoC is a full blocking crossbar between memory controllers, programmable logic, processing system, AI Engines, and platform management controller. Examples of NoC connections include:

- Sharing device access to DRAM (DDR memory)
- PL to PL connections
- Memory mapped access to the AI Engine array
- Connecting between PS and PL
- Connecting between PS and DDR memory

In devices built using stacked silicon interconnect (SSI) technology, the vertical NoC columns connect between adjacent super logic regions (SLRs), which allows device configuration data to travel between master and slave SLRs.

Platform Management Controller

The PMC is responsible for managing the ACAP with the following main categories of responsibility: securely booting and configuring the platform; and life-cycle management, which includes device integrity and debug, and system monitoring.

Boot and Configuration

The PMC is responsible for booting the ACAP from the primary boot source in a multi-stage boot process that supports both a non-secure and a secure boot. For a secure boot, the AES-GCM, SHA3-384 decryption/authentication, and ECDSA/ RSA blocks decrypt and authenticate the image. Upon reset, the mode pins are read to determine the primary boot memory device, such as quad-SPI, octal-SPI, SD, or eMMC. The PMC then proceeds to execute the code out of on-chip BootROM and copies the first stage boot loader (FSBL) from the boot device to the on-chip memory while undergoing authentication and decryption. The configuration of the PL is also undertaken by the FSBL. The bitstream is loaded from its storage medium, and after authentication and decryption, is sent to the PL configuration interface.

It is also possible to reconfigure portions of the PL using partial reconfiguration. A new bitstream for a portion of the PL can be loaded from the PS, through the primary or secondary boot interfaces, e.g., PCIe or Ethernet. Upon reconfiguration, a portion of the PL provides the new functionality determined by the new configuration, enabling users to quickly adapt the functionality of their ACAP to changing system requirements.

System Monitoring

The PMC contains system monitoring capability for monitoring voltage and temperature in the PS and PL to enhance the overall safety, security, and reliability of the system. The core of the system monitor is a 10-bit 200kSPS ADC, which can be accessed via JTAG, PMBus, or I2C interfaces, via the PS directly, and via the PL through the NoC.

Device Integrity and Debug

JTAG is the primary interface for Versal ACAP debug features. The JTAG architecture has two IEEE Std 1149.1 compliant TAP controllers that are connected in series: the Arm DAP controller and the PMC tap controller. The Arm DAP controller is the main controller for debug functions supporting: PS CoreSight debug architecture, debug of the PL, nonvolatile flash programming, and eFUSE/BBRAM programming. The PMC TAP controller supports: device IDCODE, programming of the PL, and boundary scan.

The PMC also contains a high-speed debug port (HSDP) that can be used as a faster debug method than the primary JTAG interface. The HSDP interface is a high-throughput interface consisting of separate ingress and egress simplex Aurora channels that leverage the transceivers to the north of the PS. The HSDP allows daisy-chaining of channels from different devices. The HSDP can also be accessed by the serial transceivers in the PL via an Aurora bridge also in the PL.

External Flash Memory Interfaces

The SD/eMMC controller supports 1- and 4-bit data interfaces at low, default, high-speed, and ultra-high-speed (UHS) clock rates. This controller also supports 1-, 4-, or 8-bit-wide eMMC interfaces that are compliant to the eMMC 4.51 specification. eMMC is one of the primary boot and configuration modes and supports boot from managed NAND devices. The controller has a built-in DMA for enhanced performance.

The quad SPI controller is one of the primary boot and configuration devices. It supports 4-byte and 3-byte addressing modes. In both addressing modes, single, dual-stacked, and dual-parallel configurations are supported. Single mode supports a quad serial NOR flash memory, while in double stacked and double parallel modes, it supports two quad serial NOR flash memories.

The octal SPI controller is one of the primary boot and configuration devices. It has an 8-pin interface and provides 200Mb/s of bandwidth in double data rate mode and up to 166Mb/s in single data rate mode. It has two chip-selects to support deeper memory and a built-in DMA for enhanced performance.

Slave Boot Modes

In addition to JTAG, SelectMAP is also a slave boot mode. SelectMAP is a high bandwidth, stream oriented, parallel interface that can be configured as 8-, 16- or 32-bit wide. It runs up to 200MHz.

Compute and Acceleration Engines

AI Engine

Versal AI Core devices contain an AI Engine array on the north edge of the device. The AI Engine array is a two dimensional array of AI Engine tiles that each contain: an AI Engine, a high-performance VLIW vector (SIMD) processor; integrated memory; and interconnects for streaming, configuration, and debug. Alongside the AI Engine tiles is the AI Engine array interface that provides the necessary interface logic to connect the AI Engine array to the other resources in the PL, PS, and the NoC.

AI Engine

The AI Engine contains a scalar unit, a vector unit, load units, and a memory interface. The scalar unit contains: a 32-bit scalar RISC processor with register files for general-purpose, pointer, configuration, and backup registers; and a 32x32-bit scalar multiplier and supports non-linear functions including sine/cosine, squareroot, and inverse-squareroot. Three address generator units (AGUs) are available: two dedicated as load units, and one dedicated as a store unit. The vector unit contains: 512-bit vector fixed-point / integer unit and a single-precision floating point vector unit—both supporting concurrent operation on multiple vector lanes. Within each AI Engine is a dedicated, single-port, 16KB program memory 128-bit wide and 1k word deep. The program memory supports instruction compression and has ECC protection and reporting.

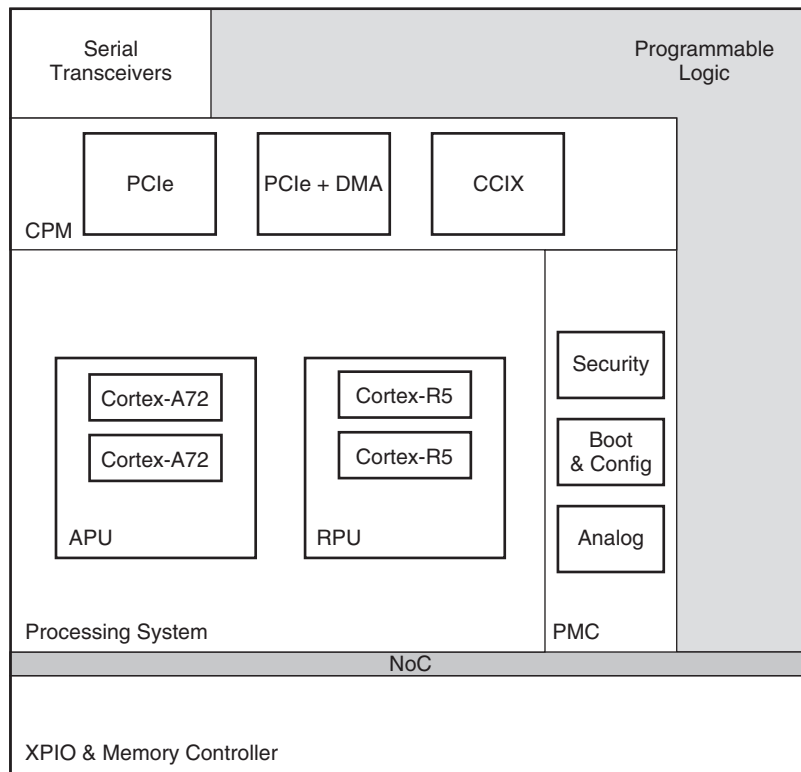
AI Engine Data Memory

Separate from the AI Engine, each AI Engine tile contains a 32KB data memory divided into eight single-port banks, each 256-bit wide and 128b deep. This structure allows up to eight parallel memory access transactions every clock cycle, with five cycle access latency. Stall signals identify memory access conflicts during which time any outstanding memory operations are buffered. Each data memory module supports memory error detection (parity) and reporting.

The data memory also contains DMA logic that supports incoming stream to local memory, outgoing stream from local memory, and buffered streams in local memory. Support for two-dimensional stride access enables any AI Engine to access data memories in adjacent AI Engine tiles in the north, south, east, and west directions, allowing a single AI Engine to access up to 128KB of data memory.

Processing System

All Versal devices contain a processing system (PS) consisting of Scalar Engines (APU and RPU) and peripherals. The PS is part of a group of architectural elements that include the platform management controller (PMC), interconnect for CCIX and PCIe (CPM) block, NoC, and integrated memory controllers that are tightly coupled, but are also capable of operating independently from each other. The simplified layout is shown in [Figure 2](#).



DS950_02_061418

Figure 2: PS and Surrounding Blocks

The PMC is responsible for booting the PS from one of its primary boot sources. The PS also has direct access to the PCIe®, PCIe-DMA, and CCIX features inside the CPM, which talks to the serial transceivers directly to the north. Programmable logic can be configured at any stage of the process and can be performed before or after the PS is booted.

Application Processing Unit (APU)

The APU has a feature-rich dual-core Arm Cortex-A72 processor. Cortex-A72 cores are 64-bit-wide application processors based on Arm-v8A architecture. The Armv8 architecture supports hardware virtualization. Each of the Cortex-A72 cores has: 48KB of instruction L1 cache and 32KB of data L1 cache, with parity and ECC protection respectively; a NEON SIMD engine; and a single and double precision floating point unit. In addition to these blocks, the APU consists of a snoop control unit and a 1MB L2 cache with ECC protection to enhance system-level performance. The snoop control unit keeps the L1 caches coherent thus eliminating the need of spending software bandwidth for coherency. The APU also has a built-in interrupt controller supporting virtual interrupts. The APU communicates to the rest of the PS through the 128-bit AXI coherent extension (ACE) port via Cache Coherent Interconnect (CCI) block, using the system memory management unit (SMMU). The APU is also connected to the PL through the 128-bit accelerator coherency port (ACP), providing a low latency coherent port for accelerators in the PL. To support real-time debug and trace, each core also has an Embedded Trace Macrocell (ETM) that communicates with the Arm CoreSight™ Debug System.

Real-Time Processing Unit (RPU)

The RPU in the PS contains a dual-core Arm Cortex-R5 processor. Cortex-R5 cores are 32-bit real-time processor cores based on Arm-v7R architecture. Each of the Cortex-R5 cores has 32KB of level-1 (L1) instruction and data cache with ECC protection. In addition to the L1 caches, each of the Cortex-R5 cores also has a 128KB tightly coupled memory (TCM) interface for real-time single cycle access. The RPU also has a dedicated interrupt controller and floating point unit. The RPU can operate in either split or lock-step mode. In split mode, both processors run independently of each other. In lock-step mode, they run in parallel with each other, with integrated comparator logic, and the TCMs are used as 256KB unified memory. The RPU communicates with the rest of the PS via the 128-bit AXI-4 ports connected to the low power domain switch. It also communicates directly with the PL through 128-bit low latency AXI-4 ports. To support real-time debug and trace, each core also has an embedded trace macrocell (ETM) that communicates with the Arm CoreSight Debug System.

Connectivity Peripherals

In the PS, many peripherals are used to connect to external devices over industry-standard protocols, including CAN-FD, SPI, USB, Ethernet, I2C, and UART. Many of the peripherals support clock gating and power gating modes to reduce dynamic and static power consumption. These peripherals either use multiplexed I/O (MIO) to connect to the external components or if required they can also be pinned using the extended multiplexed I/O (EMIO) through the PL.

With the adjacent CPM and PMC providing access to the high-speed and configuration interfaces respectively, the number of peripherals required directly in the I/O unit is relatively small, containing:

- 2 gigabit Ethernet controllers
- 2 SPIs controllers
- 2 I2Cs controllers
- 2 CAN/CAN-FD controllers
- 2 UARTs
- GPIO
- 1 USB 2.0 (device, host, and OTG) controller

The following functions are included in the I/O unit so they can share MIO:

- 4 Triple-timer counters
- 1 Watchdog Timer

All peripherals within the I/O unit have Trustzone support through system control registers. The I/O unit has master and slave AXI interface ports to the LPD interconnect. One APB bus at top level is used to control AXI bridges within the I/O unit.

USB 2.0

The USB controller can be configured as host, device, or On-The-Go (OTG). The core is compliant to USB 2.0 specification and supports high, full, and low-speed modes in all configurations. In host mode, the USB controller is compliant with the Intel XHCI specification. In device mode, it supports up to 12 end points. The Universal Low Peripheral Interface (ULPI) is used to connect the controller to an external PHY operating up to 480Mb/s.

Ethernet MAC

The pair of tri-speed Ethernet MACs support 10Mb/s, 100Mb/s, and 1Gb/s operations. They also support jumbo frames and timestamping through the interfaces based on IEEE Std 1588 v2. Time Sensitive Network (TSN), which either uses IEEE Std 1588 or 802.1AS-REV, is also supported. The Ethernet MACs can be connected through the MIO (RGMII), or through EMIO (GMII). The GMII interface can be converted to a different interface within the PL.

Configurable Logic Block (CLB)

Every configurable logic block (CLB) contains 32 look-up tables (LUTs) and 64 flip-flops. The LUTs can be configured as either one 6-input LUT with one output, or as two 5-input LUTs with separate outputs but common inputs. Each LUT can optionally be registered in a flip-flop. In addition to the LUTs and flip-flops, the CLB contains arithmetic carry logic and multiplexers to create wider logic functions. Within each CLB, 16 LUTs can be configured as 64-bit RAM, as 32-bit shift registers (SRL32), or as two SRL16s. For every group of 64 flip-flops, there are four clocks signals, four set/reset signals, and 16 clock enables. Within every CLB are dedicated interconnect paths for connecting LUTs together without having to exit and re-enter a CLB and cascade muxes. This enables a flexible carry logic structure that allows a carry chain to start at any bit in the chain.

Internal Memory

Each ACAP contains several programmable, internal storage capabilities. In addition to the distributed RAM capability in the CLB, there are dedicated blocks for building various size storage elements.

On-Chip Memory (OCM)

In addition to the 32KB of L1 data cache, the RPU contains 256KB OCM with ECC. The OCM is accessed through two 128-bit AXI interfaces with one AXI interface dedicated to the two Cortex-R5 processors and the other AXI interface available to the APU and other masters. Memory accesses from the RPU are treated with higher priority than memory accesses through the general 128-bit AXI interface.

Some Versal ACAPs include Accelerator RAM, an additional 4MB of on-chip memory with ECC located outside of the PS. This memory provides direct access from the RPU via a 128-bit AXI interface and can also be accessed from the PL through two 256-bit AXI interfaces. The memory is divided into three banks supporting concurrent read or write accesses from the PL and RPU to different banks.

Block RAM

True dual-port block RAMs, each having 36Kb of storage capacity, can be configured as either one 36Kb RAM, or two completely independent 18Kb RAMs. Each port can be configured as $4K \times 9$, $2K \times 18$, $1K \times 36$, or 512×72 in simple dual-port mode. The two ports can have different aspect ratios. Also, the read port width can be different from the write port width for each port.

Synchronous operation: Each memory access, read, and write is controlled by the clock. All inputs, data, address, clock enable, and write enable are registered. The data output is always latched, retaining data until the next operation. An optional output data pipeline register allows higher clock rates at the cost of an extra cycle of latency. During a write operation, the data output can be made to reflect the previously stored data, the newly written data, or remain unchanged. There is independent reset control of output latches and registers.

Asynchronous operation: The data outputs can also be set/reset asynchronously. Sleep input (places array in low power state) can be optionally asynchronous.

True dual-port operation: The block RAM has two completely independent ports that share nothing but the stored data.

Simple dual-port operation: One port is dedicated as a write port and the other as a read port. The data width can thus be extended to 72 bits for the 36Kb full block RAM or 36 bits for the "split" 18Kb block RAM.

Cascade mode supports all configurations available in 36Kb RAM or 18Kb RAM. Cascading refers to combining multiple block RAMs to build larger ones, without using additional logic resources.

Each 64-bit-wide block RAM can generate, store, and use eight additional bits to perform single-bit error correction and double-bit error detection (ECC) during the read process. The ECC logic can also be used when writing to, or reading from, external 64- or 72-bit-wide memories. Block RAM contents can be initialized or cleared by the configuration bitstream.

UltraRAM

Dual-port UltraRAMs, each having 288K bits of storage capacity, can be configured as one 288Kb RAM. Each port can be configured as 32K x 9, 16K x 18, 8K x 36, or 4K x 72. The two ports can have different aspect ratios.

Synchronous operation only: Each memory access, read, and write is controlled by the clock. All inputs, data, address, clock enable, and write enable are registered. The data output is always latched, retaining data until the next operation. An optional output data pipeline register allows higher clock rates at the cost of an extra cycle of latency.

Asynchronous control: The data outputs can also be set/reset asynchronously. Sleep input (places the memory array in low-power state) can be optionally asynchronous.

Pseudo dual-port operation: There are two ports on the memory. Each is capable of reading or writing in a single cycle. The ports are sequenced in a fixed order, allowing up to two transactions per cycle. (Both ports write, both ports read, or one port reads while the other writes.) This necessitates that the two ports share a common clock. During a write operation, the data output remains unchanged on a given port. There is independent reset control of output latches and registers.

ECC logic in the UltraRAM supports real time error checking and correction. Both ports have dedicated ECC for either read or write. The ECC logic is organized for 64-bit-wide data, which can generate, store, and use eight additional bits to perform single-bit error correction and double-bit error detection (ECC) during the read process.

It is possible to cascade the address and data of adjacent blocks to build deeper memories. Optional pipelining is also available to maintain the clock rate through tall cascades of UltraRAM.

Digital Signal Processing (DSP)

DSP applications use many binary multipliers and accumulators, best implemented in dedicated DSP Engines. Versal devices have many dedicated, low-power DSP Engines, combining high speed with small size while retaining system design flexibility.

Each DSP Engine fundamentally consists of a dedicated 27×24 bit two's complement multiplier and a 58-bit accumulator. The multiplier can be dynamically bypassed, and two 58-bit inputs can feed a single-instruction-multiple-data (SIMD) arithmetic unit (dual 24-bit add/subtract/accumulate or quad 12-bit add/subtract/accumulate), or a logic unit that can generate any one of ten different logic functions of the two operands.

The DSP Engine includes an additional pre-adder, typically used in symmetrical filters. This pre-adder improves performance in densely packed designs and reduces the DSP Engine count by up to 50%. The 116-bit-wide XOR function, programmable to 12, 22, 24, 34, 58, or 116-bit widths, enables performance improvements when implementing forward error correction and cyclic redundancy checking algorithms.

The DSP Engine also includes a 58-bit-wide pattern detector that can be used for convergent or symmetric rounding. The pattern detector is also capable of implementing 116-bit-wide logic functions when used in conjunction with the logic unit.

The DSP Engine provides extensive pipelining and extension capabilities that enhance the speed and efficiency of many applications beyond digital signal processing, such as wide dynamic bus shifters, memory address generators, wide bus multiplexers, and memory-mapped I/O register files. The accumulator can also be used as a synchronous up/down counter.

The DSP Engine layout enables new modes of operation in addition to the conventional fixed-point operation.

Three element vector / INT8 dot product: The DSP Engine can be used in vector fixed-point ALU mode in which the 27 x 24 multiplier is replaced by a three-dimensional vector dot-product unit. The dot-product unit supports element-wise product negation with NEGATE pins.

Complex 18b x 18b: Using two back to back DSP Engines, the Versal architecture enables creation of an 18 x 18 + 58 twos complement complex multiply accumulator in which each of the two complex inputs can be optionally conjugated.

Single precision floating point: The DSP Engine contains a floating-point multiplier and a floating-point adder with separate outputs in binary32 format. Each floating-point multiplier input can be in either binary32 (single-precision or FP32) or binary16 (half-precision or FP16) format.

Soft-Decision Forward Error Correction (SD-FEC)

The soft-decision forward error correction (SD-FEC) is an integrated IP block that provides high throughput LDPC, polar, turbo code implementations. The LDPC decode and encode functionality is capable of covering a range of customer specified quasi-cyclic (QC) codes. The turbo decode functionality principally covers codes used by LTE. The SD-FEC offers significant power and area savings versus implementations done in the programmable logic.

The SD-FEC block is integrated within the ACAP as part of the integrated IP column and interfaces to other blocks using interconnect and global clock resources. The memory requirements of the block are handled within the block itself and do not use block RAM or UltraRAM resources.

Connectivity

Transceivers

Two types of transceivers are in Versal devices: GTY transceivers support data rates up to 32.75Gb/s; GTM transceivers support data rates up to 58Gb/s. Minimum data rate for all transceivers is 1.2Gb/s but lower data rates can be achieved by utilizing oversampling in the programmable logic.

GTY Transceivers

The serial transmitter and receiver are independent circuits that use an advanced phase-locked loop (PLL) architecture to multiply the reference frequency input by certain programmable numbers between 4 and 25 to become the bit-serial data clock. Each transceiver has a large number of user-definable features and parameters. All of these can be defined during device configuration, and many can also be modified during operation.

Transmitter (GTY)

The transmitter is fundamentally a parallel-to-serial converter with a conversion ratio of 16, 20, 32, 40, 64, 80, 128, or 160. This allows the designer to trade off datapath width against timing margin in high-performance designs. These transmitter outputs drive the PC board with a single-channel differential output signal. TXOUTCLK is the appropriately divided serial data clock and can be used directly to register the parallel data coming from the internal logic. The incoming parallel data is fed through an optional FIFO and has additional hardware support for the 8B/10B, 64B/66B, or 64B/67B encoding schemes to provide a sufficient number of transitions. The bit-serial output signal drives two package pins with differential signals. This output signal pair has programmable signal swing as well as programmable pre- and post-emphasis to compensate for PC board losses and other interconnect characteristics. For shorter channels, the swing can be reduced to reduce power consumption.

Receiver (GTY)

The receiver is fundamentally a serial-to-parallel converter, changing the incoming bit-serial differential signal into a parallel stream of words, each 16, 20, 32, 40, 64, 80, 128, or 160. This allows the designer to trade off internal datapath width against logic timing margin. The receiver takes the incoming differential data stream, feeds it through programmable DC automatic gain control, linear and decision feedback equalizers (to compensate for PC board, cable, optical and other interconnect characteristics), and uses the reference clock input to initiate clock recognition. There is no need for a separate clock line. The data pattern uses non-return-to-zero (NRZ) encoding and optionally ensures sufficient data transitions by using the selected encoding scheme. Parallel data is then transferred into the device logic using the RXUSRCLK clock. For short channels, the transceivers offer a special low-power mode (LPM) to reduce power consumption by approximately 30%.

The receiver DC automatic gain control and linear and decision feedback equalizers can optionally "auto-adapt" to automatically learn and compensate for different interconnect characteristics. This enables even more margin for 10G+ and 25G+ backplanes.

Out-of-Band Signaling

The transceivers provide out-of-band (OOB) signaling, often used to send low-speed signals from the transmitter to the receiver while high-speed serial data transmission is not active. This is typically done when the link is in a powered-down state or has not yet been initialized. This benefits PCIe and SATA/SAS and QPI applications.

GTM Transceivers

The serial transmitter and receiver are independent circuits that use an advanced phase-locked loop (PLL) architecture to multiply the reference frequency input by certain programmable numbers between 16 and 160 to become the bit-serial data clock. Each transceiver has a large number of user-definable features and parameters. All of these can be defined during device configuration, and many can also be modified during operation. The GTM transceiver supports a PAM4 data range of 19–30Gb/s and 38–58Gb/s, NRZ data range of 9.5–15Gb/s and 19–30Gb/s.

Transmitter (GTM)

The transmitter is fundamentally a parallel-to-serial converter. These transmitter outputs drive pulse amplitude modulated signals with either four levels (PAM4) or two levels (NRZ) to the PC board with a single-channel differential output signal. TXOUTCLK is the appropriately divided serial data clock and can be used directly to register the parallel data coming from the internal logic. The incoming parallel data can optionally leverage a Reed-Solomon, RS(544,514) Forward Error Correction encoder and/or 64b66b data encoder. The bit-serial output signal drives two package pins with PAM4 differential signals. This output signal pair has programmable signal swing as well as programmable pre- and post-emphasis to compensate for PC board losses and other interconnect characteristics. For shorter channels, the swing can be reduced to reduce power consumption.

Receiver (GTM)

The receiver is fundamentally a serial-to-parallel converter, changing the incoming PAM4 differential signal into a parallel stream of words. The receiver takes the incoming differential data stream, feeds it through automatic gain compensation (AGC) and a continuous time linear equalizer (CTLE), after which it is sampled with a high-speed analog to digital converter. Further equalization is completed digitally via a decision feedback equalizer (DFE) and feed forward equalizer (FFE) implemented in DSP logic before the recovered bits are parallelized and provided to the PCS. This equalization provides the flexibility to receive data over channels ranging from very short chip-to-chip to high loss backplane applications across all supported rates. Clock recovery circuitry generates a clock derived from the high-speed PLL to clock in serial data and provides an appropriately divided and phase-aligned clock, RXOUTCLK, to internal logic.

Parallel data can optionally be transferred into an RS-FEC and/or 64b/66b decoder before being presented to the programmable logic interface.

PCIe

There are two methods of communicating over PCI Express in Versal ACAPs: the integrated block for PCI Express that resides in the connectivity IP column illustrated in [Figure 1](#) and the cache coherent PCI Express block that resides in the CPM shown in [Figure 2](#).

Integrated Block for PCI Express

The user-accessible integrated block for PCIe resides in the connectivity IP column within the programmable logic. This block communicates with the adjacent serial transceivers and supports Gen1, Gen2, Gen3 data rates at link widths of x1, x2, x4, x8, or x16, and Gen4 data rates at link widths of x1, x2, x4, or x8. Each block can be configured as an Endpoint or Root Port. The Root Port can be used to build the basis for a compatible Root Complex, to allow custom chip-to-chip communication via the PCI Express protocol, and to attach ASSP Endpoint devices, such as Ethernet controllers or Fibre Channel HBAs, to the ACAP. For high-performance applications, advanced buffering techniques of the block offer a flexible maximum payload size of up to 1,024 bytes. The integrated block interfaces to the integrated high-speed transceivers for serial connectivity and to block RAMs for data buffering. Combined, these elements implement the Physical Layer, Data Link Layer, and Transaction Layer of the PCI Express protocol.

PCI Express and CCIX

The primary PCIe interface for the processing system resides in the CPM. There are two integrated blocks for PCIe in the CPM supporting up to Gen4 x16. Both of the integrated blocks for PCIe can be configured as an Endpoint and one can also be configured as a Root Port and contains DMA logic. The CPM also incorporates CCIX functionality to allow a programmable logic accelerator to act as a CCIX compliant accelerator. In this configuration, the integrate block for PCIe is configured as an Endpoint and connects to the serial transceivers to the north and the PS to the south of the CPM. The PS provides paths to connect to the NoC and, therefore, to the DDR memory controllers. The integrated block for PCIe can also be configured as a CCIX-only 20 or 25 GT/s x8 mode.

Ethernet

The Versal architecture contains integrated blocks for Ethernet functionality capable of operating at different data rates.

Multirate Ethernet MAC (MRMAC)

The multirate Ethernet MAC (MRMAC) provides high-performance, low latency Ethernet ports supporting a wide range of customization and statistics gathering. Supported configurations are: 1 x 100GE; 2 x 50GE; 1 x 40GE; 4 x 25GE; and 4 x 10GE.

The MRMAC supports the following FECs defined and required by IEEE standards: Clause 91 RS(528,514) KR4 FEC, for 25/50/100GE NRZ support; Clause 91 RS(544,514) KP4 FEC for 50/100GE PAM4 support; and Clause 74 FEC, for 10/25/40/50GE low-latency support. The MRMAC has a rich set of bypass modes to enable access to FEC-only mode (for custom protocols) and FEC+PCS (for protocol testers).

The MRMAC also supports a new high-precision timestamping feature to enable sub-nanosecond accuracy on IEEE Std 1588 timestamps. This provides hardware support for new IEEE Std 1588-based time-sensitive networks (TSN) as well as the next generation Ethernet-based wireless fronthaul protocol (eCPRI).

I/O

Two types of programmable I/Os exist in the programmable logic with additional I/Os available in the PS. See [Table 6](#).

Table 6: Programmable I/O

I/O Type	XPIO	HDIO	MIO
Voltage	1.0V–1.5V	1.8V–3.3V	1.8V–3.3V
Purpose	Highest performance, DDR memory interface	Lower performance, wider voltage range	Support PS peripherals

XPIO

XPIO are always located on the bottom and also sometimes on the top of the device, depending on the device and family. XPIO are optimized for high-performance communication including, but not limited to, interfacing to DDR4 memory through the integrated memory controller blocks. XPIO are arranged in banks of 54 I/O and organized as nine 6-bit nibbles. XPIO support standards with maximum supply voltage of 1.5V. Every XPIO bank includes a physical layer interface (PHY) that can operate in 4:1 mode for use with the integrated memory controllers or 8:1 mode for use with custom circuitry.

HDIO

High-density I/O (HDIO) banks are designed to be a cost-effective method for supporting lower speed, higher voltage range I/O standards. Arranged in banks of 22, the number of HDIO varies depending on device and package size. HDIOs offer single-ended I/O including 3.3V and 2.5V LVTTTL and LVCMOS, SSTL from 1.8V to 1.2V, and HSTL from 1.8V to 1.5V. SSTL and HSTL can use the uncalibrated on-die termination. HDIO also offer differential receiver for low-speed clock inputs and pseudo-differential transmitter. There is Internal V_{REF} support. The system designer can specify the slew rate and the output strength. The input is always active but is usually ignored while the output is active. Each pin can optionally have a weak pull-up, weak pull-down resistor, or weak keeper.

MIO

Multiple banks of general-purpose I/O are implemented within the PS and PMC, each with a dedicated power supply. The main category of I/O are the three banks of multiplexed I/O (MIO), which can be accessed by the PS, the PMC, and the PL. Fixed-function I/O are also available for control and configuration functions.

Clocking

Multiple clock generation blocks are used to synthesize clock frequencies. Clock buffers and routing connect the signals to their destinations.

PS Clocking

All clocks in the PS belong to one of three groups: the main PLL clocks; the internal ring oscillator clock and interface clocks.

Main PLL Clocks

The majority of the logic in the PS is clocked from the three PLLs in the PS and one PLL in the PMC through user configurable clock divider circuits. These divider circuits generate clocks to all CPUs, main interconnects, the PMC, and all peripherals. The clocks and their associated PLLs are spread across three power domains: the PMC domain, containing the PMC; the low-power domain, containing the RPU and all peripheral clocks; and the full-power domain, containing all other clocks and their PLLs.

Internal Ring Oscillator

The PMC operates as the security manager for the device and requires a 400MHz clock provided by an internal ring oscillator.

Interface Clocks

This category includes clocks that are directly supplied from outside the PS and includes clocks for the external interfaces, including Ethernet, USB, SWDT and CAN-FD.

PL Clocking

Clock signals travel around the ACAP architecture on a network of bidirectional, horizontal, and vertical routing tracks that support many independent clock networks. The vertical tracks reside adjacent to the NoC columns. The programmable logic is divided into clock regions that each have a horizontal clock spine through the middle that can carry 24 clock signals. Clock signals travel along these horizontal clock spines and are then driven into the individual clocked elements within the PL such as flip-flops, DSP Engines, block RAM, and UltraRAM. Clock buffers and clock management components reside adjacent to the XPIO rows on south and (sometimes) north edges of the device.

Clock Management

To generate multiple clock frequencies and phases from an input clock source, Versal devices contain mixed-mode clock managers (MMCMs) and phase-locked loops (PLLs). MMCMs reside adjacent to the horizontal NoC row adjacent to the XPIO and PLLs reside in the XPIO banks. The MMCM and PLL share many characteristics. Both can serve as a frequency synthesizer for a wide range of frequencies and as a jitter filter for incoming clocks. At the center of both components is a voltage-controlled oscillator (VCO), which speeds up and slows down depending on the input voltage it receives from the phase frequency detector (PFD). There are three sets of programmable frequency dividers: D, M, and O. The predivider D, programmable by configuration and afterwards via the dynamic reconfiguration port (DRP), reduces the input frequency and feeds one input of the traditional PLL phase/frequency comparator. The feedback divider M (programmable by configuration and afterwards via DRP) acts as a multiplier because it divides the VCO output frequency before feeding the other input of the phase comparator. D and M must be chosen appropriately to keep the VCO within its specified frequency range. The VCO has eight equally spaced output phases (0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°). Each can be selected to drive one of the output dividers (six for the PLL, O0 to O5, and seven for the MMCM, O0 to O6), each programmable by configuration to divide by any integer from 1 to 128.

MMCM additional programmable features: The MMCM has a fractional counter in either the feedback path (acting as a multiplier) or in one output path. Fractional counters allow non-integer increments of 1/8 and can thus increase frequency synthesis capabilities by a factor of 8. The MMCM can also provide fixed or dynamic phase shift in small increments that depend on the VCO frequency.

Memory Controllers

Dedicated controllers that can support either DDR4 or LPDDR4 are located on the south and (sometimes) north edges of the ACAP. Each memory controller has four bidirectional 128-bit system ports and contains a scheduler with transaction reordering capability to improve memory access efficiency. The memory controller operates at half the DRAM clock rate. For example, DRAM data rate per bit is 3200Mb/s, DRAM clock rate is 1600MHz, and the memory controller clock rate is 800MHz. The memory controllers talk to the dedicated memory PHY (XPHY) in the XPIO banks which, in turn, interface with the I/O pins.

Ordering Information

This document is a pre-release document, provided ahead of silicon ordering availability. Therefore, Xilinx Versal ACAP ordering information is not yet available. Please contact your Xilinx sales representative for more information on Versal ACAP Early Access Programs.

Revision History

The following table shows the revision history for this document:

Date	Version	Description of Revisions
05/16/2019	1.1	Updated Table 1 , Table 3 , Table 4 , Table 5 , NoC , and Connectivity Peripherals .
10/02/2018	1.0	Initial Xilinx release.

Disclaimer

The information disclosed to you hereunder (the "Materials") is provided solely for the selection and use of Xilinx products. To the maximum extent permitted by applicable law: (1) Materials are made available "AS IS" and with all faults, Xilinx hereby DISCLAIMS ALL WARRANTIES AND CONDITIONS, EXPRESS, IMPLIED, OR STATUTORY, INCLUDING BUT NOT LIMITED TO WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT, OR FITNESS FOR ANY PARTICULAR PURPOSE; and (2) Xilinx shall not be liable (whether in contract or tort, including negligence, or under any other theory of liability) for any loss or damage of any kind or nature related to, arising under, or in connection with, the Materials (including your use of the Materials), including for any direct, indirect, special, incidental, or consequential loss or damage (including loss of data, profits, goodwill, or any type of loss or damage suffered as a result of any action brought by a third party) even if such damage or loss was reasonably foreseeable or Xilinx had been advised of the possibility of the same. Xilinx assumes no obligation to correct any errors contained in the Materials or to notify you of updates to the Materials or to product specifications. You may not reproduce, modify, distribute, or publicly display the Materials without prior written consent. Certain products are subject to the terms and conditions of Xilinx's limited warranty, please refer to Xilinx's Terms of Sale which can be viewed at <http://www.xilinx.com/legal.htm#tos>; IP cores may be subject to warranty and support terms contained in a license issued to you by Xilinx. Xilinx products are not designed or intended to be fail-safe or for use in any application requiring fail-safe performance; you assume sole risk and liability for use of Xilinx products in such critical applications, please refer to Xilinx's Terms of Sale which can be viewed at <http://www.xilinx.com/legal.htm#tos>.

This document contains preliminary information and is subject to change without notice. Information provided herein relates to products and/or services not yet available for sale, and provided solely for information purposes and are not intended, or to be construed, as an offer for sale or an attempted commercialization of the products and/or services referred to herein.

Automotive Applications Disclaimer

AUTOMOTIVE PRODUCTS (IDENTIFIED AS "XA" IN THE PART NUMBER) ARE NOT WARRANTED FOR USE IN THE DEPLOYMENT OF AIRBAGS OR FOR USE IN APPLICATIONS THAT AFFECT CONTROL OF A VEHICLE ("SAFETY APPLICATION") UNLESS THERE IS A SAFETY CONCEPT OR REDUNDANCY FEATURE CONSISTENT WITH THE ISO 26262 AUTOMOTIVE SAFETY STANDARD ("SAFETY DESIGN"). CUSTOMER SHALL, PRIOR TO USING OR DISTRIBUTING ANY SYSTEMS THAT INCORPORATE PRODUCTS, THOROUGHLY TEST SUCH SYSTEMS FOR SAFETY PURPOSES. USE OF PRODUCTS IN A SAFETY APPLICATION WITHOUT A SAFETY DESIGN IS FULLY AT THE RISK OF CUSTOMER, SUBJECT ONLY TO APPLICABLE LAWS AND REGULATIONS GOVERNING LIMITATIONS ON PRODUCT LIABILITY.