

Новинки Xilinx Developer Forum 2018

Михаил Коробков (m.korobkov@inline-ctc.ru)

На прошедшем 2 октября 2018 года форуме разработчиков Xilinx Developer Forum 2018 (XDF 2018) был представлен ряд важных новинок: новые архитектура и семейства кристаллов, маршруты проектирования и методологии, новый класс изделий Xilinx и стратегия рынка, новые платы и многое другое. В статье представлен обзор наиболее интересных и значимых из них.

VERSAL – АДАПТИРУЕМАЯ ПЛАТФОРМА УСКОРЕНИЯ ВЫЧИСЛЕНИЙ

Пожалуй, наиболее интересным был анонс кристаллов с новой для Xilinx архитектурой, которые ранее были известны под кодовым названием «проект Everest», – это новое поколение, которое будет выполнено по технологии FinFET 7 нм, получило название VERSAL.

VERSAL [1] – это первая в индустрии гетерогенная платформа для ускорения вычислений, которая в прессе более известна под названием ACAP – Adaptive Compute Acceleration Platform – адаптивная платформа ускорения вычислений любых приложений, объединяющая в себе одновременно несколько различных механизмов [2] (см. рис. 1):

- Модули скалярных вычислений (Scalar Processing Engines) – это про-

цессор приложений ARM Cortex-A72, процессор реального времени ARM Cortex-R5 и специализированный контроллер управления платформой (platform management controller).

- Модули адаптируемых аппаратных средств (Adaptable Hardware Engines) – усовершенствованное по сравнению с предыдущими семействами более производительное поле FPGA (Field-Programmable Gate Array) с улучшенной способностью к частичной реконфигурации, выполняемой «на лету», скорость которой возросла до 8 раз.
- Модули векторной обработки (Intelligent Engines) – поле выполнения операций умножения с плавающей запятой с минимальными задержками (DSP Engines) и специализированное поле/модуль AI Engines с высокой пропускной способностью, а также

минимальными задержками на выполнение операций и оптимальным уровнем энергопотребления, предназначенное для решения задач в области реализации искусственного интеллекта (AI Inference) и цифровой обработки сигналов. Данное поле в архитектуре VERSAL является новым по сравнению с предыдущими семействами и будет рассмотрено отдельно.

- Интегрированные интерфейсы хост-систем – как и в предыдущих семействах ПЛИС Xilinx, в архитектуре VERSAL будут поддерживаться различные аппаратные реализации для взаимодействия с отдельно стоящими процессорами: PCIe Gen4x16, интегрированный модуль AXI-DMA, CCIX для ускорения решения задач, возлагаемых на серверные центральные процессоры (CPU – Central Processing Unit).
- Интегрированные контроллеры внешней памяти – для достижения максимальной пропускной способности и возможности работы с такими типами памяти, как DDR4-3200, LPDDR-4266 и High Bandwidth Memory (HBM). В архитектуре VERSAL предусмотрены специализированные аппаратные/интегрированные контроллеры памяти, количество которых варьируется в зависимости от кристалла.
- Интегрированные протокольные решения – для реализации высокоскоростных протоколов передачи данных предусмотрены аппаратные ядра 100 Гбит/с Multirate Ethernet, 600 Гбит/с Ethernet и Interlaken, 600 Гбит/с Cryptographic Engines (AES/IPSEC/MACSEC).
- Широкополосные трансиверы – в архитектуре VERSAL предусмотрены оптимизированные по энергопотреблению трансиверы 32/58 Гбит/с PAM4 и 112 Гбит/с PAM4.
- Интегрированные радиочастотные блоки – Multi-GSPS RF-ADC/DAC нового поколения, DDC/DUC, SD-FEC для 5G и DOCSIS.
- Программируемые блоки ввода/вывода для поддержки различных интерфейсов, среди которых MIPI D-PHY с поддержкой скорости более 3 Гбит/с на сенсор, различные низкоскоростные виды памяти типа NAND и Storage-Class, LVDS и I/O общего назначения.
- Сеть-на-кристалле (NoC – Network-on-Chip) – одна из особенностей кристал-

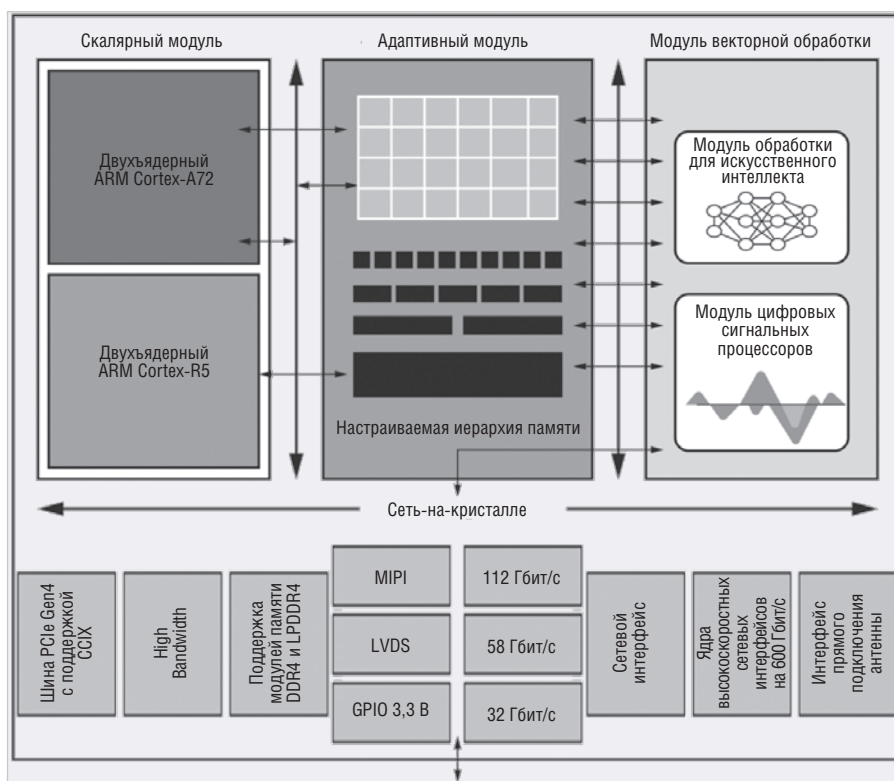


Рис. 1. Функциональная диаграмма Xilinx ACAP Versal

лов с архитектурой VERSAL, назначение которой состоит в быстрой доставке данных и обмене данными между различными частями микросхемы. Обладает пропускной способностью в несколько Тбит/с, возможностью непосредственного программирования, не требует этапа размещения и трассировки, доступна при загрузке кристалла, поскольку является интегрированной шиной. Энергопотребление такой интегрированной сети по сравнению с её реализацией на ресурсах программируемой логики ниже в 8 раз. Поддерживается возможность арбитража между различными модулями.

Наиболее важным является то, что для работы с новой архитектурой будут доступны различные наборы инструментов, такие как программное обеспечение, библиотеки, IP-ядра, промежуточное программное обеспечение (Middleware), драйверы и различные фреймворки.

Для архитектуры VERSAL анонсировано 6 серий, которые получили индивидуальные названия и разделены по целевым направлениям [2]:

- AI RF Series;
- AI Core Series;
- AI Edge Series;
- HBM Series;
- Premium Series;
- Prime Series.

ОБЛАСТИ ПРИМЕНЕНИЯ И ПРЕИМУЩЕСТВА VERSAL

Выбранные Xilinx ещё несколько лет назад целевые направления рынка остались прежними. Кристаллы архитектуры VERSAL в первую очередь ориентированы на центры обработки данных и облачные вычисления, 5G и другие беспроводные технологии, машинное обучение и высокоскоростную передачу данных по проводным каналам связи (межсерверная передача данных на скорости 400/600 Гбит/с).

Для того чтобы ответить на вопрос о выгоде, получаемой при использовании ACAP, следует рассмотреть существующие архитектуры и пояснить их предназначение [3] (см. рис. 2):

- Скалярные вычислительные элементы (CPU) эффективны для решения комплексных алгоритмов и последовательных задач принятия решений, но имеют ограниченный прирост производительности при масштабировании системы.
- Элементы векторной обработки (DSP – Digital Signal Processor, GPU – Graphics Processor Unit) наиболее эффективны в случае возможности рас-



Рис. 2. Различные вычислительные архитектуры и решаемые с их помощью задачи

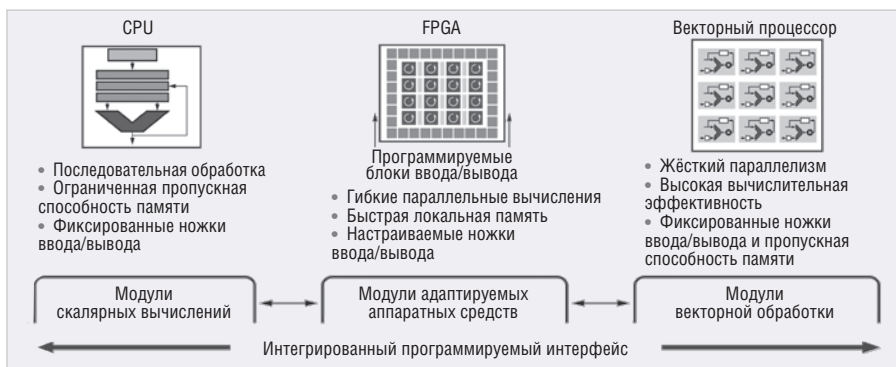


Рис. 3. Объединение в гетерогенную архитектуру

параллеливания решаемой задачи, однако теряют в производительности из-за негибкой иерархии памяти, т.е. наиболее узким местом системы является обмен данными с внешней памятью.

- Программируемая логика (FPGA) может быть сконфигурирована для решения конкретной задачи, что делает её наиболее эффективной в приложениях, где критичной является задержка вычислений (например, системы помощи водителю) или имеется нерегулярная структура данных (например, вычисление генома). При этом традиционно узким местом таких систем является время, необходимое для компиляции и имплементации кода, которое обычно составляет несколько часов, в то время как для других рассматриваемых архитектур этот процесс занимает считанные минуты.

В результате анализа всех достоинств и недостатков указанных архитектур Xilinx был сделан вывод, что наилучшим является комплексное решение, объединяющее в себе все три архитектуры (см. рис. 3), или адаптируемая платформа ускорения вычислений (ACAP).

ACAP разработана таким образом, чтобы функционировать «из короб-

ки», при этом не требуя от разработчика умений и знаний проектирования на языках RTL. По умолчанию ACAP ориентирована на проектирование с использованием высокоуровневых C-подобных языков и различных готовых фреймворков. Кристаллы с архитектурой VERSAL имеют встроенную оболочку (Shell), которая включает технологии PCIe и CCIX с интегрированными контроллерами прямого доступа к памяти DMA, NoC и интегрированными контроллерами внешней памяти, освобождая разработчиков от необходимости работы на уровне RTL. Программирование может осуществляться на уровне фреймворков, например для задач машинного обучения пользователь может непосредственно задействовать Caffe или TensorFlow, также возможно применять специализированные готовые библиотеки, например библиотеки фильтров для задач 5G. Возможно и традиционное проектирование на уровне RTL [3].

3 наиболее важных достоинства ACAP:

1. Программируемость – способность «быстрого старта» разработки приложений с помощью специализированного программного обеспечения и фреймворков.

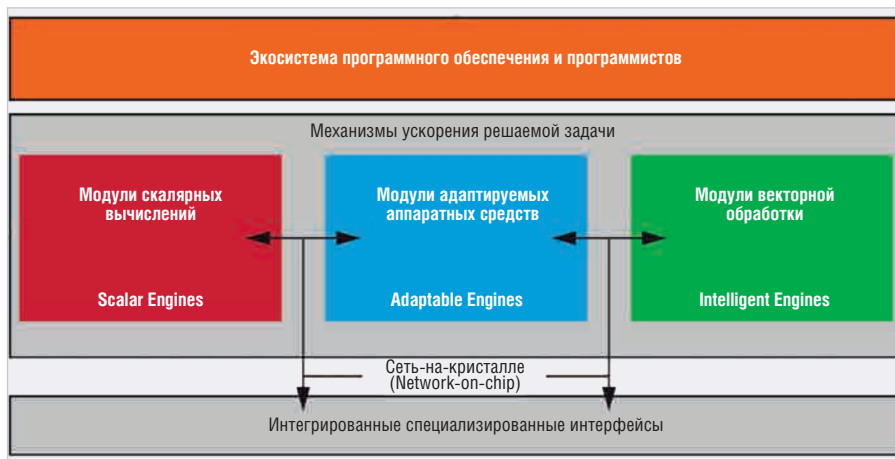


Рис. 4. VERSAL ACAP – концептуальная диаграмма

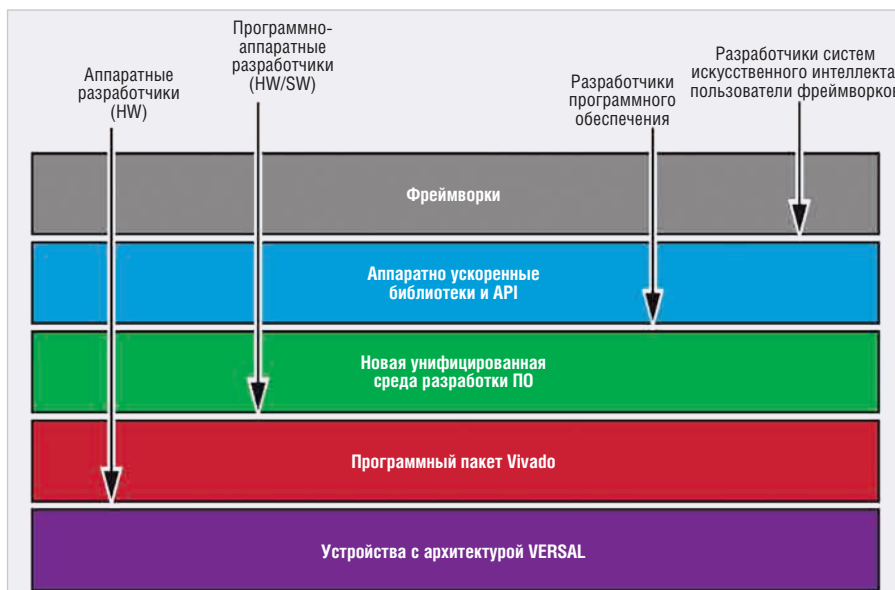


Рис. 5. Концепция программирования архитектуры VERSAL

- Ускорение широкого спектра приложений от искусственного интеллекта и сетевых смарт-карт до беспилотных автомобилей и терабитных оптических сетей.
- Динамическая адаптируемая реконфигурация – способность реконфигурировать аппаратную часть кристалла (FPGA) под решение другой задачи в течение нескольких миллисекунд.

АППАРАТНОЕ ОБЕСПЕЧЕНИЕ ПЛАТФОРМЫ ACAP

ACAP представляет собой смесь нескольких систем: скалярной, адаптируемой и интеллектуальной, связанных между собой с помощью NoC – это сеть с огромной пропускной способностью в несколько Тбит/с (см. рис. 1).

Скалярная система построена на основе двухъядерного процессора ARM Cortex-A72, имеющего в 2 раза большую производительность на ядро

по сравнению с ARM Cortex-A53, используемым в кристаллах предыдущего поколения Xilinx. Сочетание усовершенствованной архитектуры процессора и технологического процесса FinFET 7 нм позволяет в 2 раза улучшить показатель DMIP/Вт по сравнению с технологией 16 нм. Сертифицированные по стандарту ASIL-C процессоры Cortex-R5 из семейства UltraScale+™ перенесены на 7 нм техпроцесс с дополнительным уровнем безопасности и подходят для ответственных (safety) приложений.

Адаптивная система состоит из программируемой логики и ячеек памяти. Отличительной особенностью программируемой логики нового поколения является то, что её можно перепрограммировать для организации иерархии памяти, настроенной на конкретную вычислительную задачу. Такой подход позволит интеллектуальной системе достичь гораздо более высокой эффективности и большей

пропускной способности памяти на единицу вычислений по сравнению с современными CPU и GPU.

Интеллектуальная система представляет собой набор модулей с поддержкой выполнения с помощью очень длинной машинной команды (very long instruction word – VLIW), обработки множественных данных по принципу SIMD, которые соединены между собой с помощью интерконнектов, что позволяет достичь пропускной способности данных в сотни Тбит/с. Такая система обеспечивает повышение производительности в 5–10 раз в задачах цифровой обработки сигналов и машинного обучения [3].

АДАПТИВНОЕ УСКОРЕНИЕ

Гетерогенные модули VERSAL ACAP обеспечивают оптимальное ускорение программных приложений независимо от типа приложения. Интеллектуальная система ускоряет решение задачи машинного обучения и классические алгоритмы цифровой обработки сигналов. Программируемая логика нового поколения внутри адаптируемой системы ускоряет параллелизуемые алгоритмы. Многоядерный процессор предоставляет исчерпывающие встроенные вычислительные ресурсы для остальных приложений.

VERSAL спроектирован таким образом, чтобы его программирование осуществлялось без навыков разработки на уровне RTL (см. рис. 4, 5), что открывает новые возможности для специалистов из разных сфер:

- учёные, работающие с Big Data и искусственным интеллектом, могут решать свои задачи с помощью стандартных фреймворков и ускорить получение результатов;
- разработчики прикладных программ могут ускорить любое приложение, даже не имея опыта работы в низкоуровневом проектировании (RTL);
- аппаратные разработчики (HW/RTL) могут продолжать применять стандартные маршруты RTL-проектирования с использованием Vivado® Design Suite, при этом уменьшая время разработки за счёт использования архитектуры VERSAL, готовых IP-блоков и аппаратных решений.

СПЕЦИАЛИЗИРОВАННЫЕ АППАРАТНЫЕ БЛОКИ

Наличие гибких программируемых интерфейсов позволяет получить доступ к компонентам, расположенным вне кристалла, включая

стандартные интерфейсы для общения с внешним хост-процессором. В приложениях, относящихся к дата-центрам, программное приложение обычно расположено на стороне хост-процессора. Интерфейс, позволяющий подсоединить хост-процессор к программируемым ресурсам платформы VERSAL, называется оболочкой (the Shell). Встроенная оболочка включает интерфейсы CCIX (Cache Coherent Interconnect for Accelerators), PCIe Gen4 x16, контроллеры DMA, кэш-когерентную память, встроенные контроллеры памяти.

NoC выступает в качестве связующей сети IP-модулей и аппаратных компонентов архитектуры. Это открывает возможность стандартизации и масштабирования аппаратного фреймворка, позволяя достичь эффективного обмена данными между компонентами гетерогенной архитектуры VERSAL [3].

ГЕТЕРОГЕННОЕ УСКОРЕНИЕ

Несмотря на то что программируемая логика (FPGA) и векторные вычислители (DSP, GPU) имеют лучшие характеристики, чем типовые центральные процессоры, преимущество архитектуры ACAP заключается в том, что она объединяет все три типа вычислителей на одном кристалле, образуя тем самым жёстко связанный гетерогенный вычислитель. В данном случае может получиться так, что 1+1+1 будет больше 3, то есть эффект от одновременного использования различных типов вычислителей совместно будет больше, чем от каждого по отдельности.

Показатели преимущества по ключевым сегментам рынка, достигаемые при применении ACAP VERSAL [3], приведены в таблице.

В настоящее время для ознакомления с архитектурой и особенностями доступны документы для двух серий: AI Core и Prime.

СЕРИЯ AI CORE

Данная серия [4] архитектуры VERSAL обладает наивысшими вычислительными показателями и минимальными задержками выполнения операций, позволяя получить лучшие результаты в задачах искусственного интеллекта и машинного обучения. Кристаллы этой серии оптимизированы для применения в облачных вычислениях, сетевых приложениях и автономных устройствах. В серии AI Core планируется 5 устройств, в состав которых войдут:

VERSAL ACAP и ключевые сегменты рынка

Рынок	Решаемая задача	Прирост производительности в сравнении с			Примечания
		CPU	GPU	FPGA	
Центры обработки данных	Распознавание изображений – без учёта задержек	43x	2x	5x	GoogLeNet v1 (неограниченное количество примеров для обучения)
	Распознавание изображений – задержка в 2 мс	–	8x	5x	GoogLeNet v1 (<2 мс), нижняя граница для CPU 5 мс
	Анализ рисков	89x	–	>1x	По результатам исследований компании Maxeler
	Геномика (Genomics)	90x	–	>1x	Результаты анализа генома человека. По результатам исследований компании Edico Genome
	Эластичный поиск (Elastic Search)	91x	–	>1x	В 91 раз меньше задержка при анализе 1 Тбайт данных по сравнению с CPU. По результатам исследований компании BlackLynx
Беспроводная передача данных 5G	16x16 5G удалённых радиостанций (Remote Radio)	–	–	>5x	5-кратный рост пропускной способности для радиостанций 5G
	Формирование луча (Beam Forming)	–	–	>5x	5-кратный рост выполняемых вычислений
Радары специального назначения	DSP TMACs	–	–	>5x	Более 27 TMACs
	Время на итерацию алгоритма	–	–	>100x	Компиляция кода для модулей векторной обработки за считанные минуты
Автомобильная промышленность	Получение результата в системах искусственного интеллекта (<2 мс)	–	3x	15x	ResNet50 Batch=1 Лучшая масштабируемость решения в приложениях, критичных ко времени распознавания, таких как системы помощи водителю и автономное вождение
Проводная передача данных	Шифрование передаваемых данных	–	–	4x	ACAP имеет встроенные аппаратные блоки сетевых решений и криптографии, позволяя тем самым достичь терабитных значений шифрования потока данных

Примечание: TMACs – 10^{12} операций умножения/с.

- от 128 до 400 вычислительных блоков AI Engines;
 - двухъядерный процессор ARM Cortex-A72;
 - двухъядерный процессор реального времени ARM Cortex-R5;
 - 256 Кбайт встроенной памяти с функцией коррекции ошибок ECC;
 - более 1900 DSP-блоков, оптимизированных для вычислений с плавающей запятой с минимальной задержкой;
 - более 1,9 млн системных логических ячеек;
 - 130 Мбайт памяти UltraRAM;
 - до 34 Мбайт блочной памяти RAM и до 28 Мбайт распределённой памяти;
 - 32 Мбайт памяти для ускорителя, которая доступна непосредственно из любого модуля и является уникальной для устройств серии VERSAL AI.
- Также будут доступны PCIe Gen4 x8 и x16, CCIX, трансиверы 32 Гбит/с, до 4 контроллеров памяти DDR4, до 4 Multirate Ethernet-контроллеров, 650 высокосортных входов/выходов для организации интерфейсов MIPI D-PHY, NAND, LVDS, а также 78 мультиплексируемых входов/выходов, соединённых с внешними компонентами, и более чем 40 входов/выходов, поддерживающих напряжение до 3,3 В (High Density I/O). Все эти компоненты объединены сетью-на-кристалле NoC, которая имеет до 28 master/slave-

портов с пропускной способностью в несколько Тбит/с. Технические параметры и показатели производительности для данных различной точности [5] приведены в дополнительных материалах к статье на сайте журнала www.soel.ru.

ОСНОВНЫЕ ПРЕИМУЩЕСТВА СЕРИИ AI CORE

Улучшенные системные характеристики:

- новый класс вычислительных модулей для векторных операций AI Engines более чем в 100 раз превосходит в вычислениях серверные процессоры;
 - настраиваемая иерархия памяти под конкретную вычислительную задачу, которая позволит избежать большой задержки обращения к памяти – наиболее узкого места многих вычислительных систем;
 - многоядерная процессорная система на базе ARM Cortex-A72 с двукратным приростом производительности по сравнению с предыдущей версией A53;
 - шина PCIe Gen4 с пропускной способностью 25 Гбит/с и поддержкой CCIX для кэш-когерентного соединения с хост-процессором.
- Уменьшение стоимости:
- одно устройство из серии AI Core способно заменить серверную стойку с самыми производительными на сегодня CPU;

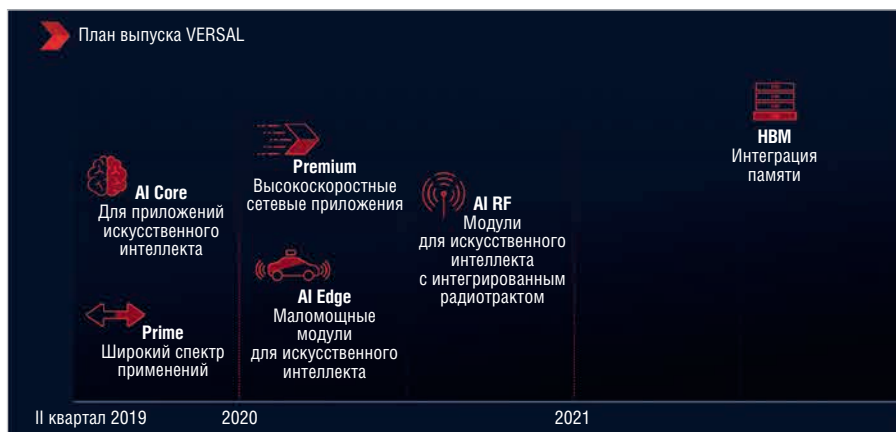


Рис. 6. План выпуска кристаллов с архитектурой VERSAL

- динамическое переключение между решаемыми задачами для объединения нескольких ускорителей в один;
- динамическое реконфигурирование частей кристалла за считанные миллисекунды для решения любых задач – от обработки и распознавания изображений до использования в научно-прикладных целях.

Повышение производительности:

- программное обеспечение оптимизировано для задач глубокого и машинного обучения;
- предварительно установленные интерфейсы (оболочка, Shell) с повышенной безопасностью, легко интегрируемые в вычислительную инфраструктуру;
- компиляция фреймворков за считанные минуты.

Также новая серия AI Core потребляет почти на 50% меньше мощности по сравнению с устройствами предыдущего поколения (при использовании новых функций).

Серия PRIME

Серия Prime [6] является основной серией среднего диапазона платформы VERSAL, применимой в нескольких сегментах рынка. Эти приложения включают разработку сетевого оборудования с пропускной способностью от 100 до 200 Гбит/с, центры обработки данных, коммуникационное тестовое оборудование. Устройства этой серии содержат трансиверы на 58 Гбит/с, оптимизированные блоки ввода/вывода и DDR, обеспечивая малую задержку и высокую производительность. Технические параметры и показатели производительности для данных различной разрядности серии Prime [7] приведены в дополнительных материалах к статье на сайте журнала www.soel.ru.

Характеристики серии Prime:

- до 2 млн системных логических ячеек;

- интегрированные Multirate MAC (Medium Access Control) с поддержкой 10/25/40/50 Гбит/с Ethernet с задержкой 1 нс для возможности применения в приложениях eCPRI и TSN;
- интегрированные высокопроизводительные многоканальные DMA-контроллеры;
- PCIe Gen4 с поддержкой CCIX (до 252 виртуальных функций);
- NoC с пропускной способностью более 1 Тбит/с;
- многоядерная процессорная система на базе ARM Cortex-A72 с двукратным приростом производительности по сравнению с предыдущей версией A53;
- многоядерный процессор реально времени ARM R5.

ПРЕИМУЩЕСТВА СЕРИИ PRIME

Повышение производительности:

- выигрыш почти в 2 раза по показателю производительность/Вт по сравнению с устройствами предыдущего поколения;
- оптимизированные по потребляемой мощности трансиверы на 32 и 58 Гбит/с, поддержка PAM4;
- 3200 Мбит/с DDR4 и 4266 Мбит/с LPDDR4 с интегрированными DMA-контроллерами.

Уменьшение стоимости и сокращение перечня компонентов:

- экономичные высокоскоростные трансиверы на 25/50/100 Гбит/с для сетевых приложений и дата-центров;
- 3 рабочих напряжения, позволяющие выполнить настройку производительности и потребляемой мощности без ущерба производительности для конечного приложения;
- корпус без крышки для улучшенного до 40% теплоотвода.

Уменьшение потребляемой мощности:

- потребляемая мощность почти на 40% ниже по сравнению с устройствами предыдущего поколения (при использовании новых функций);
- оптимизированные по потребляемой мощности трансиверы для приложений, в которых критична стоимость и потребляемая изделием мощность, таких как сетевые смарт-карты NIC, 5G-коммутаторы (eCPRI) и виртуализация сетевых функций NFV. Ускорение разработки;
- оптимизированное программное обеспечение;
- методология применения NoC;
- предварительно интегрированные интерфейсы.

Планы Xilinx по выпуску кристаллов

Xilinx планирует выпустить первые кристаллы серий AI Core и Prime во II квартале 2019 года (см. рис. 6). Кристаллы следующих семейств следует ожидать не ранее 2020 года, а именно: в I квартале 2020 года планируется выход на рынок кристаллов серий Premium и AI Edge, во II квартале того же года – кристаллов серии AI RF, затем ожидается выход серии с интегрированной высокопроизводительной памятью HBM Series, однако, согласно планам компании, это произойдет только во II квартале 2021 года [2].

Чтобы не пропустить выход новых анонсов, новостей и документов, касающихся архитектуры VERSAL, Xilinx предлагает подписку на новые обновления информации о VERSAL [8].

ЛИТЕРАТУРА

1. Versal: <https://www.xilinx.com/products/silicon-devices/acap/versal.html>
2. Versal Presentation: https://www.xilinx.com/content/dam/xilinx/imgs/press/media-kits/XDF_VERSAL_Press_Presentation_FINAL_Static.pdf
3. WP505. Versal: The First Adaptive Compute Acceleration Platform (ACAP). Xilinx. 2018.
4. Versal AI Core Series: <https://www.xilinx.com/products/silicon-devices/acap/versal-ai-core.html>
5. XMP452 – Versal AI Core Series Product Selection Guide (ver1.0.1). Xilinx inc. Oct. 2018.
6. Versal Prime Series: <https://www.xilinx.com/products/silicon-devices/acap/versal-prime.html>
7. XMP453 – Versal™ Prime Series Product Selection Guide (ver1.0.1). Xilinx inc. Oct. 2018.
8. Versal Stay Informed: <https://www.xilinx.com/products/silicon-devices/acap/versal-stay-informed.html>



Свобода проектирования



САПР электроники

В состав Delta Design, обеспечивающей сквозной цикл проектирования печатных плат, входят модули:

- Менеджер библиотек
- Схемотехнический редактор
- Схемотехническое моделирование
- HDL-симулятор
- Редактор правил
- Редактор печатных плат
- Топологический редактор плат TopoR
- Коллективная работа для предприятий