

# Архитектура и области применения ПЛИС Xilinx Versal

И. Тарасов, д. т. н.<sup>1</sup>

УДК 004.2 | ВАК 05.13.05

Versal – новая программируемая платформа компании Xilinx, выпускаемая с соблюдением 7-нм технологических норм на производственных мощностях фирмы TSMC. Семейство Versal отнесено производителем к новому подклассу ACAP (adaptive compute acceleration platform – адаптивная платформа ускорения вычислений). Особенностью устройств Versal является добавление на кристалл матрицы процессоров с архитектурой VLIW, ориентированных на реализацию алгоритмов машинного обучения и обработку сигналов в сетях программно-определяемого радио. Об архитектуре, характеристиках и областях применения ПЛИС Xilinx Versal рассказывается в статье.

Сегодня разработчики высокопроизводительных вычислительных устройств могут выбирать архитектуру из нескольких основных классов:

1. CPU (центральное процессорное устройство) представляет собой, как правило, процессор с универсальной системой команд. Для них характерны наиболее широкие возможности по реализации алгоритмов, однако универсальный характер архитектуры обуславливает и функциональную избыточность [1].
2. GPU (графическое процессорное устройство) представляет собой показательный пример подхода, когда возможности одного вычислительного узла существенно сокращаются, чтобы в полной мере соответствовать требованиям определенного класса задач. Уменьшение площади одного узла позволяет разместить на кристалле больше таких узлов, что хорошо видно на примере современных GPU.
3. FPGA (Field-Programmable Gate Array) – это наиболее распространенная сегодня архитектура ПЛИС, состоящая из матрицы реконфигурируемых логических ячеек (с добавлениями аппаратных компонентов статической памяти и умножителей). По сравнению с GPU размер ячейки еще меньше, что позволяет конфигурировать цифровые устройства вплоть до отдельных триггеров. На практике это означает возможность реализации на базе FPGA произвольной цифровой схемы.

На протяжении эволюции ПЛИС в микросхемы с архитектурой FPGA добавлялись аппаратные устройства, что приводило к введению новых терминов. Производители ПЛИС (в частности, Xilinx) обращали внимание разработчиков на то, что новые микросхемы были чем-то большим,

чем «ПЛИС с аппаратными ядрами». Например, при помещении на кристалле аппаратной подсистемы процессора ARM проектирование для такой ПЛИС достаточно существенно изменяется. Микросхема получает новые возможности, переставая быть просто процессором с очень большой матрицей конфигурируемых ячеек. В этой матрице можно реализовать автономно работающий ускоритель, для поддержания работы которого не требуется постоянный обмен данными с ARM. Соответственно, процесс проектирования для такой микросхемы начинается с разработки системной архитектуры с целью определить, какие задачи будут реализованы аппаратно, а какие – программно.

Вполне успешным и до сих пор актуальным продуктом такого типа являются семейства ПЛИС Xilinx Zynq-7000 и Xilinx Zynq UltraScale+. Производитель позиционирует эти ПЛИС как отдельный подкласс – ППЧНК (полностью программируемые системы на кристалле). Важным свойством этих ПЛИС является готовность к работе подсистемы ARM вне зависимости от проекта, реализуемого для конфигурируемой части. Поскольку контроллеры памяти DDR3, Ethernet, USB и др. не содержат конфигурируемых ресурсов, их характеристики являются предсказуемыми.

Серия Versal стала следующим важным этапом эволюции архитектуры ПЛИС, добавив подсистему, уникальную не только для программируемой логики, но и в большой степени для цифровых микросхем как таковых [2]. Два семейства (Versal AI и Versal Edge) содержат матрицу процессорных ядер с архитектурой VLIW. Речь идет не о подобию подсистемы ARM, а о количествах до 400 процессоров, выполняющих до шести команд за такт. При такой организации коммуникационные возможности процессоров ограничены, а локально доступная память измеряется десятками килобайт, поэтому их преимущественным

<sup>1</sup> РТУ МИРЭА, профессор, ilya\_e\_tarasov@mail.ru.

назначением являются алгоритмы машинного обучения и обработки сигналов в системах программно-определяемого радио. Разумеется, это не строгое ограничение, а те направления, на которые VLIW-процессоры были ориентированы в процессе проектирования.

Следует помнить и о том, что в матрице FPGA находятся сотни или тысячи аппаратных компонентов DSP, обеспечивающих суммарную производительность, на порядок превышающую возможности CPU или сигнальных процессоров. Таким образом, в новой платформе, обозначенной уже как «адаптивная платформа ускорения вычислений», присутствуют три качественно различающихся подсистемы:

- матрица ресурсов FPGA, включающая в себя программируемые логические ячейки, блоки статической памяти и компоненты для аппаратного выполнения операции «умножение с накоплением»;
- многоядерный процессор ARM;
- матрица VLIW-процессоров.

Эволюция системной архитектуры ПЛИС Xilinx показана на рис. 1.

### СИСТЕМНАЯ АРХИТЕКТУРА VERSAL

Подсистемы ПЛИС Versal схематично показаны на рис. 2. Кроме трех основных подсистем, описанных выше, на кристалле также реализованы интерфейсные компоненты, контроллеры динамической памяти, накристалльная шина (Programmable Network On Chip, NoC) и ряд других устройств. Также в семействе Versal HBM в корпусе интегрирована и динамическая память большого объема с интерфейсом High Bandwidth Memory (HBM), который и дал название семейству.

Как было отмечено, наиболее важным изменением является добавление подсистемы VLIW-процессоров. Эти процессоры, однако, присутствуют только в двух выпускаемых (и одном планируемом) семействах, для которых в название добавлено обозначение AI. Этим подчеркивается, что основное назначение таких процессоров – ускорять некоторые подклассы алгоритмов, например алгоритмы машинного обучения

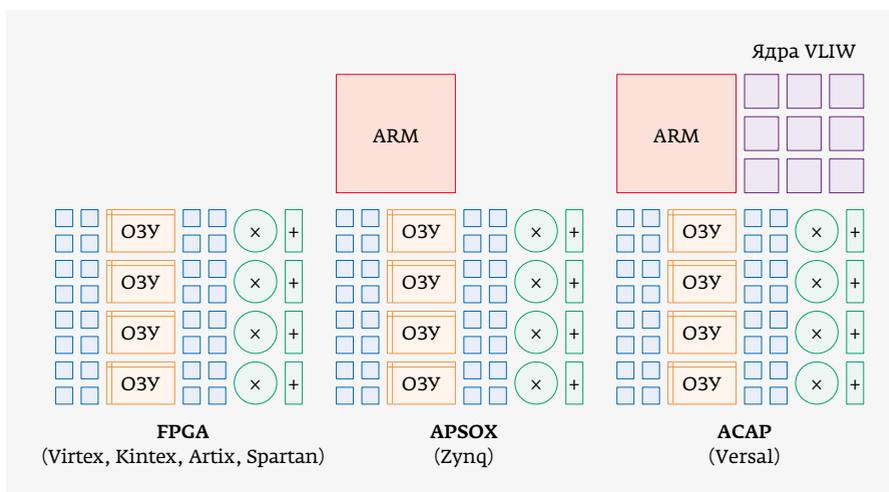


Рис. 1. Эволюция системной архитектуры ПЛИС Xilinx

(«искусственного интеллекта», AI). В то же время процессоры не являются жестко запрограммированными и подразумевают не только программирование пользователем, но и гибкие варианты взаимодействия с процессорами ARM и матрицей ресурсов FPGA, используя для этого программируемую сеть на кристалле. Эта сеть основана на многоканальной шине AXI4, широко используемой в ПЛИС Xilinx для аппаратных и программно реализуемых процессорных ядер, поэтому для нее существует достаточное количество методических материалов и подключаемых IP-ядер.

Логические ячейки Versal во многом основаны на архитектуре ПЛИС серии 7. Выпущенные в 2012 году 28-нм ПЛИС семейств Virtex-7, Kintex-7 и Artix-7 оказались довольно удачными, создав хорошую основу для дальнейшего развития, причем в последовавших сериях UltraScale

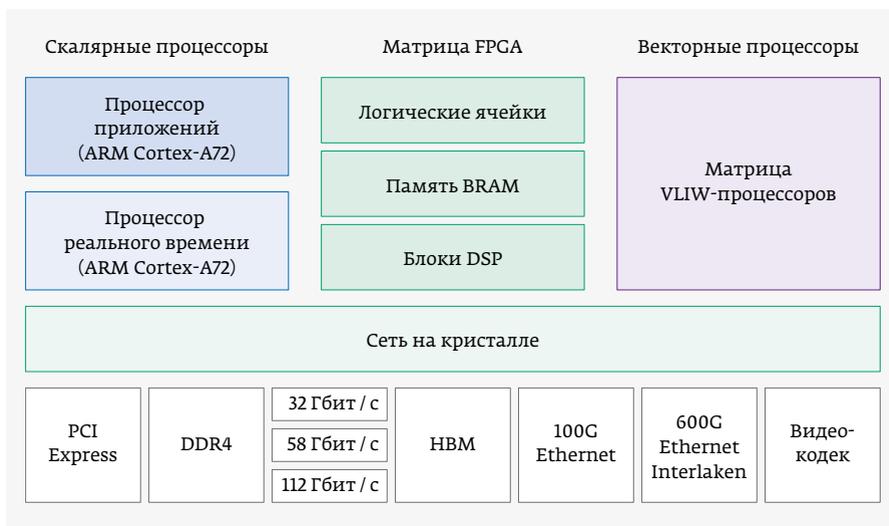


Рис. 2. Подсистемы ПЛИС Xilinx Versal

(20 нм) и UltraScale+ (16-нм FinFET) общая организация логических ячеек, модулей статической памяти и компонентов DSP была в целом сохранена, получая только эволюционные улучшения.

Базовая логическая структура, конфигурируемый логический блок (КЛБ), состоит из 32-х 6-входовых таблиц истинности (Look-Up Table, LUT) и 64 триггеров. Для LUT возможен режим, в котором одна таблица разбивается на две 5-входовых, с частичным обобщением входов. Также часть LUT может быть сконфигурирована как 64-битная память или 32-битный сдвиговый регистр. Все эти возможности унаследованы от серии 7 и хорошо зарекомендовали себя в практических проектах, поскольку их применение не требует специального вмешательства разработчика, обеспечиваясь во многих случаях синтезатором на основе анализа исходного текста на языках описания аппаратуры.

Компоненты памяти являются стандартными для матрицы FPGA Xilinx. Хорошо известным является режим распределенной памяти (distributed memory), характерный именно для ПЛИС Xilinx. Он предполагает использование ресурсов LUT для реализации блоков с мелкой гранулярностью. Доступ на чтение при этом является асинхронным вследствие особенностей работы LUT.

Блочная память (BRAM, Block RAM) имеет архитектуру, унаследованную еще от серии 7. Это синхронная статическая двупортовая память, причем интерфейс является «истинно двупортовым» – никакие сигналы портов A и B не обобщены. Блоки такой памяти удобны в том числе для реализации систем ресинхронизации данных, обеспечивая независимый доступ к одному массиву ячеек со стороны независимых интерфейсов, тактируемых собственными сигналами, относящимися к разным тактовым регионам. Базовой организацией является 36 Кбит  $\times$  1, однако блок может быть сконфигурирован как 4 Кбит  $\times$  9, 2 Кбит  $\times$  18 или 512  $\times$  72 бит. Для двух интерфейсов можно использовать разные режимы адресации.

Блоки памяти UltraRAM – более крупные по сравнению с BRAM блоки, имеющие объем 288 Кбит. Эти блоки также имеют два интерфейса, однако тактируются общим тактовым сигналом. Сочетание компонентов BRAM и UltraRAM позволяет реализовать подсистему с очень высокой пропускной способностью (поскольку речь идет о сотнях и тысячах независимо работающих блоков).

Компоненты цифровой обработки сигналов, которые в ПЛИС Xilinx долгое время обозначались DSP48 (из-за 48-разрядного аккумулятора), в Versal получили название DSP58. Очевидно, причиной этого стало увеличение разрядности аккумулятора. Исходная архитектура, введенная еще в семействе Virtex-4, предусматривала умножение независимых 18-разрядных операндов и 48-разрядный аккумулятор. По мере эволюции ПЛИС увеличивалась разрядность перемножаемых операндов, и в итоге

для Versal она составляет 27 и 24 бит. Поскольку результат их умножения может иметь 51 бит, потребовалось и увеличение разрядности аккумулятора.

Современные нейросети используют в том числе и сокращенные разрядности синаптических весов. Например, постоянное представление данных в нейросети в формате single prevision floating point (fp32) может хорошо соответствовать возможностям CPU и GPU, но в ряде задач избыточно. При этом блоки DSP58 в Versal могут реализовать одновременно три операции умножения над операндами формата INT8. В сочетании с ядрами VLIW ресурсы Versal образуют высокопроизводительный и взаимно дополняющий набор подсистем, пригодный как для реализации нейросетей и цифровой обработки сигналов, так и для подобных задач, требующих высокой производительности и гетерогенной вычислительной платформы.

## ХАРАКТЕРИСТИКИ И ОБЛАСТИ ПРИМЕНЕНИЯ СЕМЕЙСТВ VERSAL

На данный момент анонсированы пять из шести планируемых семейств платформы Versal. Семейство Versal RF AI будет содержать также высокоскоростные АЦП и ЦАП и предназначено для построения беспроводных коммуникационных устройств. В табл. 1 приведены сравнительные характеристики семейств Versal.

Важным отличием ПЛИС для реализации нейросетей является возможность сокращения разрядности коэффициентов и отказа от неиспользуемых связей с соответствующим сокращением аппаратных ресурсов, требуемых для реализации такой нейросети. Вычислительные платформы с фиксированной разрядностью данных обладают заведомо менее гибкими возможностями для таких преобразований, хотя их сильными сторонами и являются более высокая тактовая частота и меньший размер компонентов на кристалле. В то же время материалы компании Xilinx [3] показывают, что сочетание сокращения разрядности коэффициентов и прореживания (pruning) нейросети положительно влияют на характеристики ПЛИС в сравнении с GPU и TPU (Tensor Processing Unit).

В обзоре [4] рассматривается производительность Versal в сравнении с CPU и GPU (а также «чистыми» FPGA) в ряде вычислительных задач.

В классе «центры обработки данных» рассмотрены:

- распознавание изображений;
- анализ рисков;
- расшифровка генома человека;
- эластичный поиск.

В рассмотренных задачах упоминается ускорение относительно CPU в 43–90 раз. опережение GPU не такое существенное, однако GPU оказались пригодными не для всех рассмотренных задач.

Для сетей 5G наблюдается 5-кратное ускорение относительно FPGA, причем CPU и GPU в сравнении не участвуют.

Таблица 1. Сравнительные характеристики семейств ПЛИС Versal

	Versal AI Edge	Versal AI Core	Versal Prime	Versal Premium	Versal HBM
Процессорная подсистема	2-ядерный ARM Cortex-A72; 48 Кбайт/32 Кбайт L1 кэш; 1 Мбайт L2 кэш 2-ядерный ARM Cortex-R5F; 32 Кбайт/32 Кбайт L1 кэш				
Число процессорных ядер AI	8-304	128-400	-	-	-
Число логических ячеек, тыс.	44-1 139	540-1 968	329-2 233	1 575-7 352	3 507-5 148
Число блоков DSP	90-1 312	928-1 968	464-3 984	1 904-1 4352	7 392-10 848
Память, Мбит	40-177	90-191	54-282	198-994	509-751
Число приемопередатчиков	0-44	8-44	8-48	72-168	88-128
Память HBM, Гбайт	-	-	-	-	8-32
Суммарная производительность над данными INT8, трлн операций/с	7-228	57-228	8-57	36-206	107-157

Для автомобильной электроники рассмотрены приложения «помощник водителя». Например, распознавание дорожной ситуации сетью ResNet50 оказалось быстрее по сравнению с GPU в три раза. При этом наличие в составе Versal процессорных ядер ARM позволяет решать на той же микросхеме и другие задачи – например, мониторинга, поддержки бортовой мультимедийной и навигационной подсистем и т. д.

### МАРШРУТ ПРОЕКТИРОВАНИЯ УСТРОЙСТВ НА БАЗЕ VERSAL

Очевидно, что появление новой подсистемы в составе ПЛИС изменяет маршрут проектирования для таких микросхем. Например, совершенно нерационально разрабатывать программу для каждого из VLIW-ядер. Кроме того, к разработке программного обеспечения добавляется процесс системного проектирования, в рамках

## Xilinx VERSAL – платформа адаптивного ускорения вычислений с аппаратными ядрами ИИ для «конвейерных» вычислений – AI Engines

**В составе Xilinx Versal**

- ◆ программируемая логика
  - увеличение логической емкости CLB в 4 раза\*
  - повышение производительности за счет локальной маршрутизации
  - усовершенствованный Dual LUT режим, конфигурация половины LUT в CLB в качестве LUTRAM
- ◆ увеличенная производительность\* модуля скалярных вычислений:
  - процессора приложений в 2 раза с ARM Cortex-A72
  - процессора реального времени в 1,4 раза с ARM Cortex-R5
- ◆ DSP блоки процессоров с разрядностью 58 бит и поддержкой вычислений с плавающей точкой
- ◆ аппаратные ядра искусственного интеллекта AI Engines для ускорения нейросетей и цифровой обработки сигнала с возможностью проведения «конвейерных» вычислений



\* в сравнении с Zynq UltraScale+





**МАКРО  
ГРУПП**  
www.macrogroup.ru

**Макро Групп – официальный дистрибьютор**  
Санкт-Петербург, ул. Свеаборгская, дом 12  
тел.: (812) 370 60 70 e-mail: fpga@macrogroup.ru

которого необходимо распределить задачи между подсистемами Versal. Процессоры ARM, подсистема ядер VLIW и матрица ресурсов FPGA являются взаимодополняющими, поэтому на ранних этапах проектирования следует правильно распределять решаемые системой задачи.

Большую помощь здесь оказывает многоуровневый маршрут проектирования, который в настоящее время включает в себя и адаптированные фреймворки на языках программирования высокого уровня. На рис. 3 показан пример взаимодействия инструментов разработки при проектировании нейросетей. Видно, что на верхних уровнях абстрагирования можно использовать такие популярные фреймворки, как Caffe, PyTorch и TensorFlow, данные из которых могут быть переданы в стандартный маршрут проектирования на основе САПР Xilinx Vitis. В основе реализации лежит «процессорный модуль глубокого обучения» (DPU), который использует комбинацию ресурсов Versal для реализации нейронов и связей между ними.

Применение высокоуровневых средств описания проекта, с одной стороны, способствует появлению некоторой функциональной избыточности (разработчики могут рассчитывать на более эффективную реализацию отдельных узлов), однако, с другой – существенно снижает порог вхождения в данную технологию. Возможность загрузить САПР Vitis бесплатно означает, что даже без приобретения дорогостоящих программных продуктов и тем более аппаратных средств можно оценить перспективы использования Versal в конкретном проекте.



Рис. 4. Внешний вид отладочной платы Versal AI Core Series VCK190

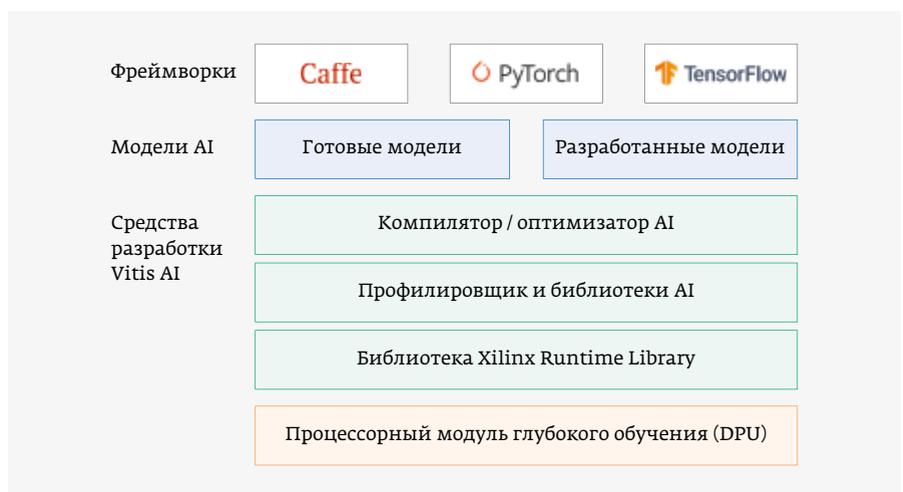


Рис. 3. Взаимодействие инструментов разработки при проектировании нейросетей для ПЛИС Versal

В настоящее время Xilinx предлагает несколько вариантов отладочных плат. Например, на рис. 4 показана отладочная плата VCK190 с установленной ПЛИС Versal AI Core Series [5].

\* \* \*

В заключение можно сказать, что новое семейство ПЛИС, представленное Xilinx, является важной вехой на пути эволюции этой аппаратной платформы. Изменения, ожидаемые от Versal, представляются сопоставимыми с теми изменениями в практике проектирования, которые произошли с распространением ПЛИС серии 7. Можно отметить возрастающую значимость высокоуровневых инструментов описания проектов, в том числе HLS (High-Level Synthesis – высокоуровневый синтез), готовых фреймворков и библиотек, ориентированных на обработку изображений и нейросети.

### ЛИТЕРАТУРА

1. **Patterson D.** A New Golden Age for Computer Architecture: History, Challenges, and Opportunities. – UC Berkeley and Google, December 5, 2018. <https://www.acm.org/hennessy-patterson-turing-lecture>
2. System-Level Benefits of the Versal Platform: [https://www.xilinx.com/content/dam/xilinx/support/documentation/white\\_papers/wp539-versal-system-level-benefits.pdf](https://www.xilinx.com/content/dam/xilinx/support/documentation/white_papers/wp539-versal-system-level-benefits.pdf)
3. <https://www.xilinx.com/products/design-tools/vitis/vitis-ai.html>
4. Versal: The First Adaptive Compute Acceleration Platform (ACAP): [https://www.xilinx.com/support/documentation/white\\_papers/wp505-versal-acap.pdf](https://www.xilinx.com/support/documentation/white_papers/wp505-versal-acap.pdf)
5. Versal AI Core Series VCK190 Evaluation Kit: <https://www.xilinx.com/products/boards-and-kits/vck190.html>

ПРИ ПОДДЕРЖКЕ И УЧАСТИИ:

Министерства промышленности, связи,  
цифрового и научно-технического развития  
Омской области,

Министерства региональной безопасности  
Омской области,

Администрации города Омска,

ОРО ООО «Союз машиностроителей России»,

Представительства ГК «Ростех» в Омской области,

Союза «Омская Торгово-промышленная палата»,

Омская ассоциация промышленников и предпринимателей  
(Региональное отделение Российского союза  
промышленников и предпринимателей),

Ассоциации «Омский региональный центр компетенций ТЭК»,

АНО «Омский НОЦ»

2 - 4 марта 2022г. **ОМСК**

# **XXIII СИБИРСКИЙ ПРОМЫШЛЕННО- ИННОВАЦИОННЫЙ ФОРУМ**

# **ПРОМТЕХЭКСПО**

**В ЭКСПОЗИЦИИ ФОРУМА:**

**ОМСКГАЗНЕФТЕХИМ**

**МАШИНОСТРОЕНИЕ**

**МЕТАЛЛООБРАБОТКА**

**СВАРКА**

**ЭНЕРГОСИБ, СИБМАШТЭК**

**ИНЭКСПО**

**АВТОМАТИЗАЦИЯ, ЭЛЕКТРОНИКА,**

**ИЗМЕРЕНИЯ**

**IT-ТЕХНОЛОГИИ**

**СВЯЗЬ**

**Одновременно проводится выставка  
«ИНДУСТРИЯ БЕЗОПАСНОСТИ И СВЯЗИ»**

**ОМСК-ЭКСПО**  
ВЫСТАВОЧНАЯ КОМПАНИЯ

 **ИнтерСиб**

**тел./факс: +7 (3812) 23-23-30**

**e-mail: [expo@intersib.ru](mailto:expo@intersib.ru)**

**[www.intersib.ru](http://www.intersib.ru)**