



# Гетерогенные платформы для передовых систем искусственного интеллекта

Наделение устройств IoT элементами искусственного интеллекта считается одним из перспективнейших направлений развития автоматизации. Тайваньская компания ADLINK, являющаяся ведущим членом альянса Intel® Internet of Things Solutions, представляет масштабируемые современные решения на базе гетерогенных вычислительных платформ. Предлагаемый ADLINK подход ускоряет и удешевляет развёртывание интеллектуальных устройств, значительно облегчая работу системных проектировщиков, инженеров и программистов.

Компания ADLINK создала гибкие гетерогенные вычислительные платформы и помогает потребителям оптимизировать системную архитектуру для лучшего соответствия решаемым задачам.

## ОПТИМИЗИРОВАННОЕ РЕШЕНИЕ ДЛЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Многие отрасли, и тому есть огромное число примеров, сегодня рассматривают искусственный интеллект (Artificial Intelligence – AI) в качестве пути трансформации своего бизнеса через более высокий уровень автоматизации и машинного обучения. В частности, производители экспериментируют с AI в области машинного зрения для распознавания и классификации дефектов, а также AI с поддержкой оптического распознавания символов для извлечения данных из устаревших систем. Тем не менее, AI всё ещё переживает период младенчества, а сложность и разнообразие оборудования и программных решений могут стать камнем преткновения на пути реализации проектов.

Для достижения оптимального решения системные архитекторы должны сначала определиться, стоит ли

большую часть алгоритмов AI сосредоточить вблизи конечных устройств (то есть «на переднем крае») или в облаке. Это концептуальное решение впоследствии повлияет на их выбор аппаратных платформ в отношении производительности, габаритов, веса и требований к энергопотреблению (SWaP – Size, Weight and Power). Чтобы максимально увеличить производительность решений AI, всё чаще рассматривают гетерогенную вычислительную платформу. Это означает, что она объединяет два или более различных типов вычислительных ядер, таких как:

- универсальный процессор (Central Processor Unit – CPU);
- программируемая вентильная матрица (Field Programmable Gate Array – FPGA);
- графический процессор (Graphics Processing Unit – GPU);
- специализированная интегральная схема (Application-Specific Integrated Circuit – ASIC).

В данной статье рассматриваются возможности реализации передовых решений AI с применением основных перечисленных типов ядер. Кроме того, в ней описаны методы, используемые ADLINK для поддержки клиентов в проектах оптимизации решений AI.

## ЗАЧЕМ УСТРОЙСТВАМ ИНТЕЛЛЕКТ

Интернет вещей (Internet of Things – IoT) эволюционировал от простых устройств, передающих данные в облако, до интеллектуальных устройств, автономно действующих по сложным алгоритмам. Локальная обработка алгоритмов AI на интеллектуальном устройстве даёт много преимуществ, в том числе:

- более быстрый отклик, минимизирующий задержку благодаря снижению объёмов трафика между устройством и облаком;
- высокую безопасность и снижение риска искажения данных благодаря отправке меньшего количества данных по сетям;
- лучшую мобильность за счёт уменьшения зависимости от неполадок в беспроводных сетях (мёртвые зоны, перебои в обслуживании), когда функции AI выполняются локально на мобильном устройстве;
- снижение стоимости связи: передавая меньше данных, вы тратите меньше на сетевые сервисы.

## ПРОБЛЕМЫ РАЗРАБОТКИ

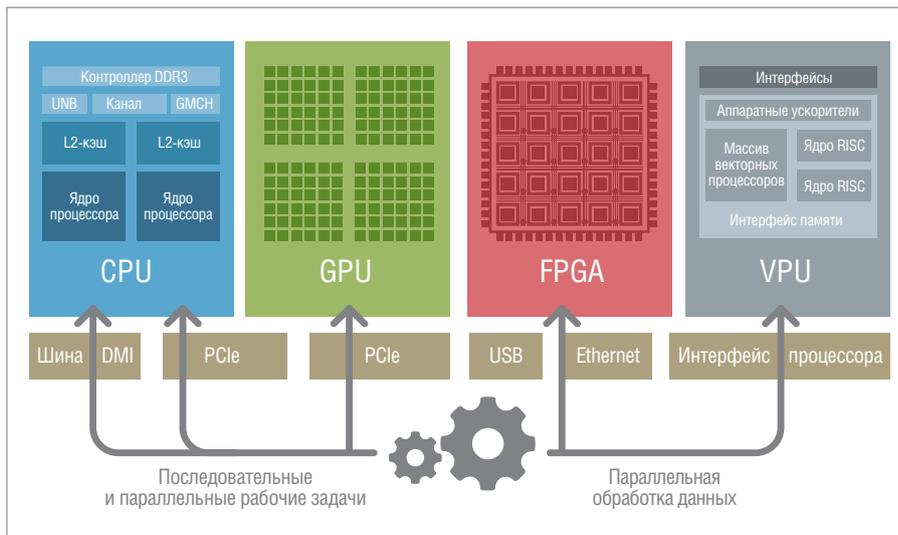
Области применения AI невероятно разнообразны. С целью улучшения качества принятия решений системные

архитекторы применяют AI для обработки широкого спектра входных данных, таких как видео, текст, голос, изображения и данные от датчиков. При этом они должны выбрать наилучший из методов принятия решений, реализуемых различными алгоритмами глубокого обучения (например, TensorFlow, Torch и Caffe) и нейросетевыми алгоритмами (например, рекуррентными и свёрточными) с различным числом слоёв. Конкретные комбинации нейронных сетей и программных пакетов, работающих на специализированных вычислительных ядрах, подходят для решения определённых задач, таких как обработка изображений, распознавание символов и классификация объектов.

Многие рабочие алгоритмы AI требуют большого объёма памяти, параллельных вычислений и вычислений с низкой точностью [1]. Задача системных архитекторов состоит в том, чтобы определить оптимизированную платформу для AI, которая бы сочетала экономическую эффективность с достаточными вычислительными ресурсами, удовлетворяющими требованиям скорости и точности. При разработке платформ для решений «на грани» системные архитекторы должны учитывать дополнительные требования, такие как надёжность оборудования и строгие SWaP-ограничения.

### ADLINK ПРЕДЛАГАЕТ ПЛАТФОРМУ

При проектировании платформы AI системным архитекторам следует рассмотреть возможность использования



#### Условные обозначения:

UNB – унифицированный северный мост; GMCH – центр управления графикой и памятью; CPU – центральный процессор; GPU – графический процессор; FPGA – программируемая вентильная матрица; VPU – процессор обработки зрительных образов.

Рис. 1. Гетерогенная платформа ADLINK для систем искусственного интеллекта

гетерогенной вычислительной архитектуры, содержащей несколько типов ядер, включая CPU, GPU, FPGA и ASIC. Цель такого подхода состоит в распределении рабочих нагрузок AI между наиболее подходящими для этого ядрами, что приводит к более быстрым вычислениям и меньшему энергопотреблению при реализации конкретных функций по сравнению с однородной платформой.

Разработка такой платформы на практике является более сложной задачей, но ADLINK упрощает процесс проектирования, предлагая гетерогенные платформы, представляющие

собой своеобразный конструктор из базовых платформ (рис. 1). Системные архитекторы могут настраивать платформы ADLINK в соответствии с вычислительными потребностями задач AI, что сокращает их усилия по разработке и обеспечивает масштабируемое решение.

### СРАВНЕНИЕ ОСНОВНЫХ ТИПОВ ЯДЕР

Далее приводится краткий обзор некоторых сильных сторон и ограничений для разных типов ядер, используемых при обработке задач AI. Дополнительная информация представлена в табл. 1.

Сравнение типов ядер, применяемых для решения задач AI

Таблица 1

Тип ядра	Специализированные ASIC	Типовое энергопотребление	Описание	Сильные стороны	Ограничения
CPU		Высокое	Гибкие универсальные процессоры	<ul style="list-style-type: none"> <li>Сложные инструкции и задания</li> <li>Управление системой</li> </ul>	<ul style="list-style-type: none"> <li>Возможные узкие места доступа к памяти</li> <li>Немного ядер (4–16)</li> </ul>
GPU		Высокое	Параллельные ядра для высококачественного рендеринга графики	<ul style="list-style-type: none"> <li>Высокопроизводительная обработка задач AI</li> <li>Высокопараллельное ядро со 100 или 1000 ядрами</li> </ul>	<ul style="list-style-type: none"> <li>Высокое энергопотребление</li> <li>Большие габариты</li> </ul>
FPGA		Среднее	Конфигурируемые логические элементы	<ul style="list-style-type: none"> <li>Гибкость</li> <li>Перепрограммируемость в полевых условиях</li> </ul>	<ul style="list-style-type: none"> <li>Высокое энергопотребление</li> <li>Сложность программирования</li> </ul>
ASIC		Низкое	Индивидуальная логика на основе специализированных библиотек	<ul style="list-style-type: none"> <li>Быстрая работа и низкое энергопотребление</li> <li>Малые габариты</li> </ul>	<ul style="list-style-type: none"> <li>Ограниченная функциональность</li> <li>Дорогой индивидуальный дизайн</li> </ul>
	VPU	Сверхнизкое	Процессор обработки изображений и искусственного зрения	<ul style="list-style-type: none"> <li>Низкая потребляемая мощность и небольшие габариты</li> <li>Процессор предназначен для ускорения обработки изображений</li> </ul>	<ul style="list-style-type: none"> <li>Ограниченный набор данных и их комбинаций</li> <li>Ограниченная поддержка сети</li> </ul>
	TPU	От среднего до низкого	Специализированная интегральная схема ASIC, разработанная Google	<ul style="list-style-type: none"> <li>Специализированная поддержка инструмента</li> <li>Процессор оптимизирован для TensorFlow</li> </ul>	<ul style="list-style-type: none"> <li>Уникальный дизайн</li> <li>Очень ограниченная поддержка фреймворка</li> </ul>

### ЦПУ общего назначения

Как правило, каждая платформа AI имеет ЦПУ (CPU) для управления платформой, исполнения вспомогательных приложений и, возможно, пользовательского интерфейса. Кроме того, центральные процессоры хорошо работают со смешанным потоком данных (например, аудио, текстом, изображениями и т.д.), процессами извлечения, преобразования и загрузки (Extract, Transform and Load – ETL).

### Графический процессор

Графический процессор (GPU) – это специализированное ядро, используемое для обработки графики, и его архитектура хорошо подходит для рабочих нагрузок AI. Состоящий из сотен или тысяч микроядер, используемых для параллельного выполнения сложных математических и статистических вычислений, графический процессор справляется как с задачами глубокого обучения на основе нейронных сетей (Deep Neural Networks – DNN), так и с выводом; однако графические процессоры могут занимать много места и отличаются высоким энергопотреблением.

### Программируемая вентильная матрица

Программируемая вентильная матрица (FPGA) имеет настраиваемые логические элементы, которые могут быть запрограммированы/перепрограммированы для пользовательского приложения, обеспечивая высокий уровень гибкости.

### Специализированная интегральная схема

ASIC – это специализированные логические чипы, разработанные с ис-

пользованием особых функциональных библиотек логических цепей. Такие ядра могут быстро выполнять сложные повторяющиеся вычисления, но они дороги, поскольку финансовые вложения и затраты времени на их инженерию весьма значительны, а тиражируемость изделий невысока.

- **Процессор обработки зрительных образов (Vision Processing Unit – VPU)** – это маломощные специализированные ASIC. Они компактны и применяются для решения задач компьютерного зрения и обработки изображений. Хорошо подходят для предобученных моделей и в меньшей степени для машинного обучения.
- **Тензорный процессор (Tensor Processing Unit – TPU)** разработан Google для нейронных вычислений (например, вывод). Эти специализированные ASIC оптимизированы для алгоритма машинного обучения Google, который называется TensorFlow.

### ПРИМЕРЫ ПРИМЕНЕНИЯ AI

ADLINK стремится помочь системным архитекторам заставить AI эффективно работать на гетерогенных компьютерах (рис. 2). Приведём некоторые примеры задач компьютерного зрения.

#### Автоматический оптический контроль

Автоматический оптический контроль (Automated Optical Inspection – AOI) используется для выявления дефектов продукта в ходе производственного процесса, помогая обслуживающему персоналу быстро решать проблемы с выходом продукта и его качеством. Машины AOI, основанные на высокопроизводительных периферийных вычислительных платформах ADLINK,

обеспечивают обнаружение и идентификацию дефектов практически в реальном времени и выполняют приложения AI, используемые для накопления знаний в предметной области с целью лучшей классификации дефектов.

#### Оптическое распознавание символов

Ещё одним приложением компьютерного зрения является оптическое распознавание символов (Optical Character Recognition – OCR), используемое для считывания данных с графических дисплеев устаревших машин, не подключённых к системе. Встроенные графические процессоры, использующие алгоритмы искусственного интеллекта на оборудовании гетерогенных вычислений ADLINK, значительно увеличивают скорость и точность рабочих процессов OCR.

#### Автономные мобильные роботы

В новом поколении автономных мобильных роботов (Autonomous Mobile Robots – AMR) используются вычисления AI с ускорением VPU для визуального управления и предотвращения столкновений. Эти возможности позволяют им приспосабливаться к изменениям в плане здания или в процессах за счёт простого обновления программного обеспечения, позволяющего роботам правильно ориентироваться и выполнять новые задачи. Мобильные роботы завтрашнего дня будут управляться программным обеспечением роя, назначающим задачи роботам на основе их доступности и местоположения, тем самым повышая их эффективность, производительность и способность работать совместно друг с другом и с людьми.



Рис. 2. Проектирование AI на базе гетерогенных вычислительных платформ

Дисплей  
машиниста

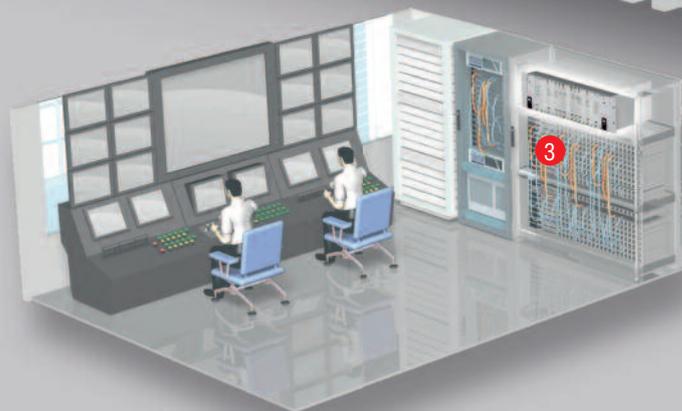
Информационный дисплей  
для пассажиров

Информационно-развлекательная  
система

Детектор дыма

IP-камера

Внутренняя  
связь



### Для построения систем

- 1 Управление поездом
- 2 Хранения данных
- 3 Диспетчерских центров

## CompactPCI®/PlusIO/Serial



### cPCI-A3515

Процессорная плата 3U CompactPCI Serial  
с процессором Intel Core i7 4/5-го поколения и ECC



### cPCI-3510 (BL)

Процессорная плата 3U CompactPCI PlusIO  
с процессором Intel Core i7 4/5-го поколения и ECC



### cPCI-3620

Процессорная плата 3U CompactPCI  
с процессором Intel Atom E3800 SoC и ECC



### cPS-H325/WDC

3U CompactPCI 8HP модуль питания PICMG 2.11  
с диапазоном рабочих температур -40...+85°C



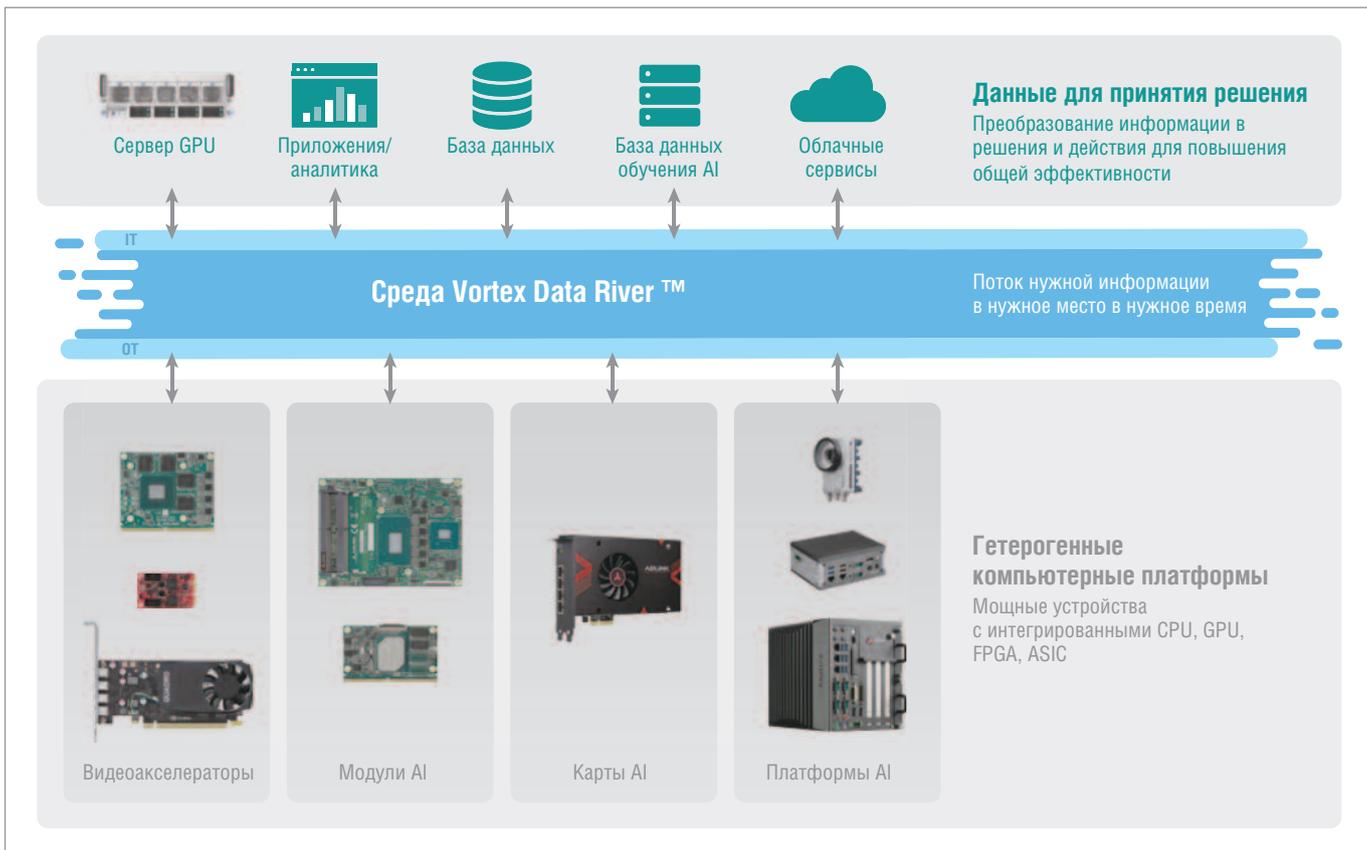


Рис. 3. Архитектура гетерогенных вычислительных платформ ADLINK

### ПЕРИФЕРИЙНЫЕ УСТРОЙСТВА С AI

Технология ADLINK позволяет IoT создавать инновационные встраиваемые решения для периферийных вычислений. Теперь благодаря оптимизированным для AI гетерогенным вычислительным платформам ADLINK выводит встраиваемые вычисления на новый уровень. Гетерогенные вычислительные платформы ADLINK состоят из продуктов уровня плат, систем и серверов, ускоренных благодаря GPU и VPU, что позволяет системным архитекторам создавать и оптимизировать системную архитектуру как для приложений AI, так и для обучения, как показано на рис. 3. В дополнение к энергоэффективности и надёжности оборудование ADLINK обеспечивает высокую производительность, необходимую для быстрой обработки данных в задачах глубокого обучения, принятия решений по паттернам и автономного машинного обучения. Умные вычислительные платформы ADLINK повышают интеллектуальные возможности и обеспечивают потоковую передачу данных между периферийными устройствами и системами в режиме реального времени, что в конечном итоге приводит к более эффективному принятию решений.

### СЕРВИСЫ ПО ОПТИМИЗАЦИИ СИСТЕМЫ ADLINK

Чтобы помочь пользователям выбрать оптимальную, с точки зрения экономической эффективности, платформу для своих приложений, в дополнение к большому разнообразию продуктов для гетерогенных вычислений ADLINK предлагает консультационные услуги.

ADLINK может дать рекомендации по аппаратному обеспечению в части оптимизации по критериям производительности/потребляемая мощность и производительность/стоимость для приложений AI в интеллектуальном производстве, умном городе и обороне. Для выявления узких мест платформ AI ADLINK осуществляет анализ производительности систем, а также сотрудничает с исследовательскими организациями и академическими учреждениями. С помощью анализа можно определить, не использует ли система неоправданно много памяти и увеличит ли производительность за счёт добавления ресурсов (например, при увеличении объёма памяти).

Воспользуйтесь преимуществами встраиваемых вычислительных решений ADLINK и глубокого обучения, чтобы оптимизировать производитель-

ность периферийных устройств с поддержкой AI.

### ЗАКЛЮЧЕНИЕ

Встраиваемые интеллектуальные решения промышленного класса для жёстких условий эксплуатации — давняя специализация тайваньской компании ADLINK, имеющей представительства в США, Сингапуре, Пекине, Шанхае, Шэньчжэне, Японии, Корее и Германии. Продукты ADLINK пользуются спросом более чем в 40 странах на пяти континентах, что, несомненно, свидетельствует об их высоком качестве и надёжности. Хорошо зарекомендовала себя продукция ADLINK и в России. Мы рекомендуем разработчикам ознакомиться с параметрами и возможностями гетерогенных вычислительных платформ ADLINK и рассмотреть перспективы применения их в будущих проектах. ●

### ЛИТЕРАТУРА

1. Sundeep Bajikar. Why AI Workloads Require New Computing Architectures — Part 1 [Электронный ресурс] // Режим доступа : <http://blog.appliedmaterials.com/ai-workloads-computing-architectures-part-1>.

Авторизованный перевод  
Юрия Широкова  
E-mail: [textoed@gmail.com](mailto:textoed@gmail.com)

# SEZAM

## ТАМ, ГДЕ ИБП БЕССИЛЬНЫ



## Сетевой защитный модуль SEZAM

### Параметры

- вход 220, 380 В
- мощность 3, 5, 10, 15 кВт
- рассеиваемая энергия импульсов перенапряжения до 20 кДж

### Защита от

- повышенного напряжения
- импульсов от 4,5 до 10 кВ и разрядов молнии
- последствий обрыва нулевого провода
- преднамеренных электромагнитных воздействий

**PROSOFT**®

ОФИЦИАЛЬНЫЙ ДИСТРИБЬЮТОР

(495) 234-0636  
INFO@PROSOFT.RU

[WWW.PROSOFT.RU](http://WWW.PROSOFT.RU)