Hitech
высокие технологии

LinQ

РОССИЙСКИЕ РЕШЕНИЯ ДЛЯ АППАРАТНОГО УСКОРЕНИЯ РАБОТЫ НЕЙРОННЫХ СЕТЕЙ

2026

О компании

Компания «ХайТэк» – российский разработчик решений на базе отечественного процессора собственной архитектуры и полного набора необходимого программного обеспечения под брендом LinQ.

Созданные компанией аппаратные ускорители для искусственного интеллекта опираются на запатентованную архитектуру и проприетарные алгоритмы инференса нейросетей.

В РЕЕСТРЕ РОССИЙСКОЙ ПРОДУКЦИИ

Процессор LinQ H1M
(К1945ВМ3Я)

BGA1296
Корпус

28 нм
Техпроцесс

16 МБ
Встроенная память

≤ 30 Вт
Энергопотребление

750 МГц
Частота блока TPU

Gen3, EndPoint, 8 линий
Контроллер PCIe

2 канала 72bit до 32Gb
Контроллер DDR4+ECC

RISC-V
Управляющий контроллер

УСКОРИТЕЛИ

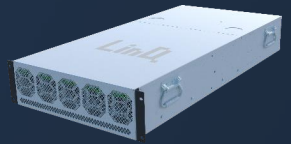
EDGE

PCIe

Акселератор

ПАК

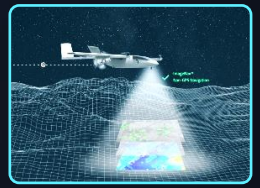
Линейка модулей LinQ для ускорения нейронных сетей

Характеристики	HP	HPD	HPQ	HPS	HX
Внешний вид					
Производительность	24 (int8) ТОПС	48 (int8) ТОПС	96 (int8) ТОПС	960 (int8) ТОПС	24 (int8) ТОПС
ResNet-50 (batch 1)	460 FPS	920 FPS	1840 FPS	-	460 FPS
ResNet-50 (batch 8)	1046 FPS	2092 FPS	4184 FPS	-	1046 FPS
Частота TPU	до 812 МГц	до 812 МГц	до 812 МГц	до 812 МГц	до 812 МГц
Интерфейсы	PCIe 2.0 x8	PCIe 3.0 x16	PCIe 3.0 x16	2x PCIe x16 Gen4	мулти
Форм-фактор	PCIe x16, 1 слот	PCIe x16, 1 слот	PCIe x16, 1 слот	3U Rackmount	SBC3,5”
Энергопотребление	<30 Вт	<55 Вт	<105 Вт	<1150 Вт	<45 Вт

Робототехника

Медицина

Транспорт


Космос и ДЗЗ

Биометрия

Безопасность

Производительность*

Сети (выборочно)		Batch	LinQ HP		LinQ HPQ	
			Throughput, FPS	Latency, ms	Throughput, FPS	Latency, ms
1	anti_spoof_mn3	1	1561,1	985,72	6244,4	985,72
2	inception_v3	1	287,1	4,77	1148,4	4,77
3	inception_v3	8	480,8	-	1923,2	-
4	lenet5	1	12519,7	128,90	50078,8	128,90
5	lenet5	8	84961,6	-	339846,4	-
6	mobilenet_v2	1	1201,7	1,44	4806,8	1,44
7	mobilenet_v2	8	2456,8	-	9827,2	-
8	resnet152	1	228,4	5,34	913,6	5,34
9	resnet50	1	432,9	2,99	1731,6	2,99
10	resnet50	8	1112,8	-	4451,2	-
11	ssd_mobilenet_v2	1	556,7	3,54	2226,8	3,54
12	ssd_mobilenet_v2	8	428	-	1712	-
13	yolo5s	1	106,1	23,71	424,4	23,71
14	yolov8n_pose	1	104,8	21,51	419,2	21,51

**Измерение выполнено по методике MLCommons MLPerf (Intel Core i7-10700K CPU @ 3.80GHz, Ubuntu 24.04, TPU Runtime 16.18.6, TPU Framework 2026.8.43*

Поддерживаемые сети

- ✓ densenet121
- ✓ efficientnet_b0
- ✓ efficientnet_b1
- ✓ efficientnet_b2
- ✓ efficientnet_b3
- ✓ inception_v1
- ✓ inception_v2
- ✓ inception_v3
- ✓ inception_v4
- ✓ lenet5
- ✓ mobilenet_v1
- ✓ mobilenet_v2
- ✓ nasnet_mobile
- ✓ pnasnet_large
- ✓ pnasnet_mobile
- ✓ resnet152
- ✓ resnet34
- ✓ resnet50_mlperf
- ✓ resnet50_v2
- ✓ ssd_mobilenet_v1
- ✓ ssd_mobilenet_v2
- ✓ scrfd 2.5g
- ✓ tiny_yolo2
- ✓ tiny_yolo3
- ✓ torch_densenet169
- ✓ torch_mobilenet_v2
- ✓ torch_resnet50
- ✓ vgg19
- ✓ xception
- ✓ YOLO2
- ✓ YOLO3
- ✓ YOLO4
- ✓ YOLO5s
- ✓ YOLO8
- ✓ YOLO11
- ✓ … и другие

Кросс-платформенность и совместимость с ОС

Поддерживаемые архитектуры

x86

intel

AMD

LOONGSON

MIPS

arm

эльбрус

AMPERE

Поддерживаемый стек ОС

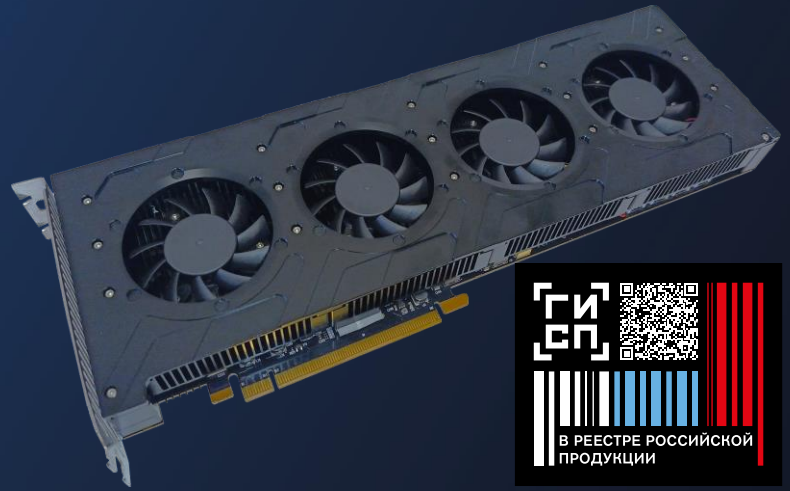
ubuntu

alt linux

ALT LINUX TEAM

ASTRA LINUX®

Ускоритель LinQ HPQ для туманных вычислений и центров обработки данных



Модуль LinQ HPQ входит в линейку модулей ускорения LinQ стандарта PCI Express (PCIe) и построен на базе четырех тензорных процессоров LinQ H1.

Включен в Реестр российской промышленной продукции Минпромторга РФ.

КЛЮЧЕВЫЕ ХАРАКТЕРИСТИКИ

- ✓ Сверхнизкая задержка при малом батче (текущий: 2,3 мс, возможность оптимизации до 1,5 мс для ResNet-50)
- ✓ Высокая эффективность на малом батче на CNN /Transformer моделях
- ✓ Нет сторонних IP в вычислителях (ARM, etc.)
- ✓ Оптимизаторы компилятора с элементами ИИ
- ✓ Программный стек собственной разработки
- ✓ Интеграция PyTorch через ONNX
- ✓ Пиковая производительность 96 TOPS (int8)
- ✓ Поддержка инференса нескольких моделей

Ускоритель LinQ HPS для высокоплотных вычислений

Акселератор LinQ HPS построен на базе модулей ускорения LinQ HPQ, имеет пиковую производительность 960 ТОПС и предназначен для выполнения инференса тяжелых нейронных сетей в реальном времени с низкой задержкой, обеспечивая потоковую передачу входных данных и результатов инференса.

КЛЮЧЕВЫЕ ХАРАКТЕРИСТИКИ

- ✓ Сверхнизкая задержка при малом батче (текущий: 2,3 мс, возможность оптимизации до 1,5 мс для ResNet-50)
- ✓ Высокая эффективность на малом батче на CNN /Transformer моделях
- ✓ Нет сторонних IP в вычислителях (ARM, etc.)
- ✓ Оптимизаторы компилятора с элементами ИИ
- ✓ Программный стек собственной разработки
- ✓ Интеграция с PyTorch через ONNX
- ✓ Пиковая производительность 960 TOPS (int8)
- ✓ Поддержка инференса 40 и более нейронных сетей
- ✓ Форм-фактор: 3U Rackmount
- ✓ PCIe интерфейс: x32 Gen4

ПАК LinQ HPC для инференса сложных нейросетей

Высокопроизводительный ПАК LinQ HPC для инференса сложных нейронных сетей (CNN, Transformer и т. д.) в реальном времени. Создан на основе двух ускорителей LinQ HPS и управляющего сервера.

КЛЮЧЕВЫЕ ХАРАКТЕРИСТИКИ

Производительность и масштабируемость

- ✓ до **1920 TOPS** (int8) на двух юнитах (node)
- ✓ одновременный запуск **более 100 моделей** нейронных сетей
- ✓ линейная масштабируемость при установке нескольких ускорителей
- ✓ форм-фактор **12U Rackmount**, поддержка стандартной **19" стойки**
- ✓ дополнительные опции **ASIC-плат** для аппаратного видеокодирования

Низкая задержка

- ✓ **~2,3 мс** для ResNet-50 (batch=1), при оптимизациях – до **1,5 мс**
- ✓ спецоптимизаторы в компиляторе (DNN Quant, Graph Optimizer)

Программный стек собственной разработки

- ✓ поддержка **PyTorch** через **ONNX**
- ✓ собственные инструменты анализа производительности
- ✓ облачный сервис (Cloud TPU) для инференса нейронных сетей

Надёжность и защита

- ✓ ECC во всех блоках памяти
- ✓ Watchdog Timer, Power Monitor, мониторинг температуры
- ✓ гибкое управление тактовой частотой для оптимизации TDP

Энергопотребление

- ✓ 3000 Вт (зависит от частоты и режима охлаждения)

Ускорители Linq HPx: области применения

Помощь в принятии решений

Детектирование и классификация объектов

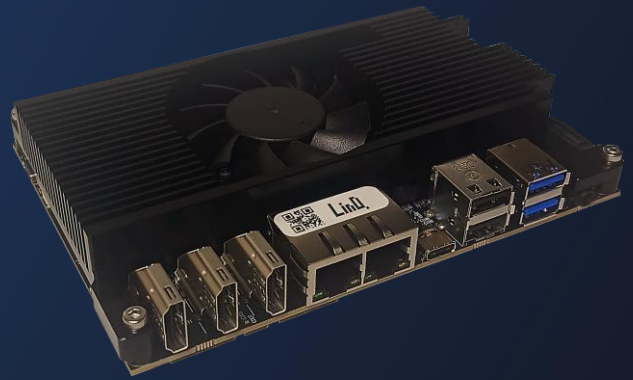
Предиктивное техобслуживание

Распределенные вычисления

Аналитика и безопасность

Научные исследования

Ускоритель LinQ HX для инференса нейронных сетей в режиме реального времени



Модуль LinQ HX входит в линейку модулей ускорения LinQ и построен на базе тензорного процессора LinQ H1M и встроенного процессора общего назначения RK3588 с широким набором интерфейсов.

Ускоритель предназначен для запуска ИИ-приложений и применения в беспилотных системах, выполняющих задачи в области мониторинга, доставки, картографии, поисково-спасательных операций и др.

КЛЮЧЕВЫЕ ХАРАКТЕРИСТИКИ

Производительность

- ✓ до **24 TOPS** (int8) и 6 TOPS на одном модуле
- ✓ одновременный запуск **более 5 моделей** нейронных сетей

Гибкая интеграция

- ✓ **широкий набор интерфейсов** обеспечивает управление и отладку модуля, подключение внешних устройств и различных компонентов, прием/передачу данных и пр.

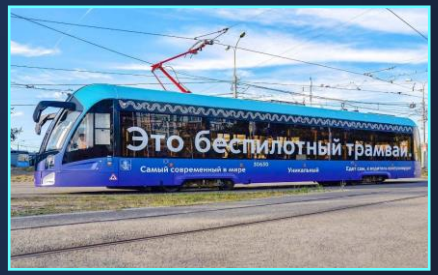
Низкая задержка

- ✓ **~2,3 мс** для ResNet-50 (batch=1), при оптимизациях – **до 1,5 мс**
- ✓ специализированные оптимизаторы в компиляторе (DNN Quant, Graph Optimizer)

Программный стек собственной разработки

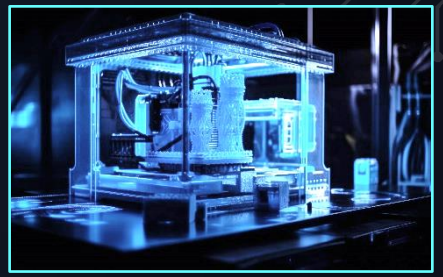
- ✓ поддержка **PyTorch** через **ONNX**

Ускорители LinQ HX: области применения



Беспилотный рельсовый транспорт

Автономный транспорт



Промышленные роботы и робототехника

Промышленный интернет вещей (IIoT)

Беспилотные летательные аппараты

Системы помощи водителю (ADAS)

Спасибо за внимание!

ООО «ХайТэк»

Москва, Волгоградский пр-т, д. 42,
ОЭЗ Технополис «Москва»

info@lin-q.ru