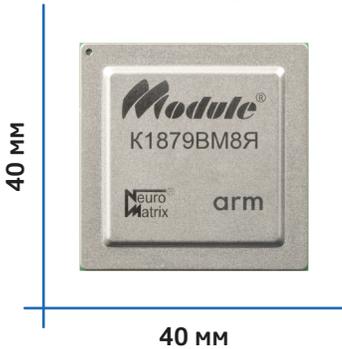


Module

# NM Quad



на базе  
4 СБИС K1879BM8Я



Высокопроизводительный вычислительный модуль на основе четырех многоядерных DSP-процессоров на базе оригинальной архитектуры NeuroMatrix Core 4 с управляющим RISC-процессором ARM Cortex A5 в форм-факторе PCIe add-in card x4 (2 слота).

Решает задачи цифровой обработки больших массивов данных.

Работает с 32- и 64-разрядными данными с плавающей точкой одинарной и двойной точности.

## Технические характеристики

Процессор и топология:	Интерфейсы и память:	ПО и NN:
<p><b>DSP:</b> 64 ядра NMC4 – 1 ГГц</p> <p><b>RISC:</b> 20 ядер Arm Cortex A5 – до 800 МГц L2 Cache – 512 Кб</p> <p><b>Производительность:</b> FP32 – 2 TFLOP/s FP64 – 0.5 TFLOP/s</p> <p><b>Техпроцесс:</b> 28 нм КМОП</p> <p><b>Мощность (типичная/максимальная):</b> 50 Вт / 100 Вт</p> <p><b>Температурный диапазон:</b> 0°C .... +40°C</p>	<p>20 Гб памяти DDR3L (до 32 Гб/с) PCIe 2.0 (4 lanes)</p> <p>Форм-фактор PCIe x16, 2 слота расширения</p> <p><b>Суммарная пропускная способность интерфейсов межпроцессорного обмена одного процессора:</b> 160 Гбит/с</p> <p><b>Внутренняя память одного процессора:</b> 76 Мбит</p>	<p><b>NMC SDK:</b> IDE (open VS-code) Компиляторы ARM и NMC Отладчик ARM Bare Metal SDK БЗИО, БУПВ (MPI, GAS), OpenCL Набор специализированных библиотек математических функций (BLAS, ЦОС)</p> <p><b>Драйверы:</b> Windows (7,10), Linux (включая "AstraLinux" и "Эльбрус ОС")</p> <p><b>Neuro Matrix Deep Learning:</b> Фирменный компилятор ИНС DarkNet и ONNX. Библиотека для применения ИНС. Библиотека трансляции моделей. Библиотека для подготовки изображений. Набор утилит для применения.</p>

## Решает задачи



## Benchmark NN inference

Ниже приведены значения производительности модуля при реализации (Inference) некоторых глубоких нейронных сетей.

**FPS** – число обработанных кадров изображения в секунду.

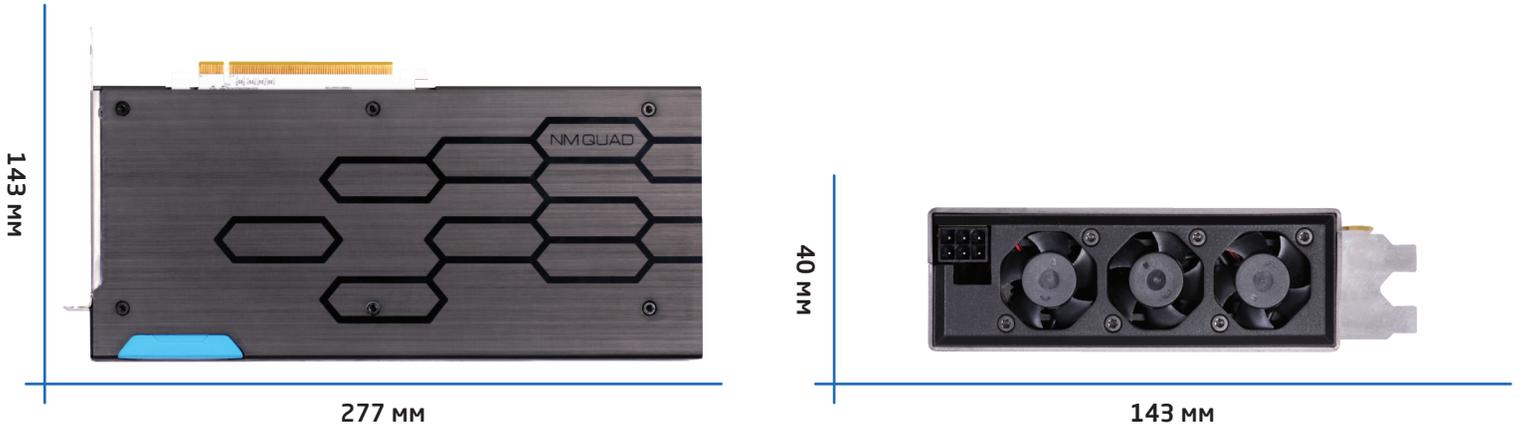
**Latency** – задержка до получения обратного кадра.

**Batch-mode** – режим одновременной обработки нескольких кадров.

Нейронная сеть (размер изображения в пикселях)	FPS NM Quad (multi unit mode)	Latency NM Quad (multi unit mode)	FPS NM Quad (batch mode)	Latency NM Quad (batch mode)
alexnet (227x227)	50,4	79	52	308
inception v3 (299x299)	51,2	78	81,2	197
inception v3 (512x512)	15,72	254	21,76	735
resnet 18 (224x224)	100	40	188	85
resnet 50 (224x224)	48,8	82	82,4	194
squeezenet (224x224)	297,6	13	400	40
u-net (512x512)*	8	500	8	2000
yolo v2 tiny (416x416)	84	47	121,6	132
yolo v3 (416x416)	14,8	270	18	889
yolo v3 tiny (416x416)	109,2	36	141,2	113
yolo v5s (640x640)	18,8	212	21,2	754

\* В модели u-net произведена замена слоев transposed convolution на upsampling

## Габариты модуля



Масса модуля < 1,5 кг.

