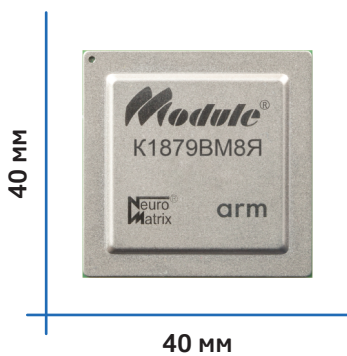
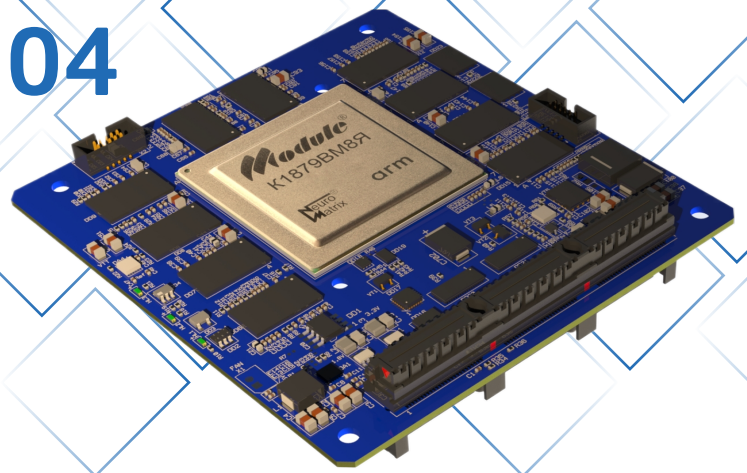




# NM Mezzo PCIe/104



на базе  
СБИС K1879BM8Я



Модуль расширения, выполненный в промышленном форм-факторе PCIe/104 type 2, поддерживает как верхнее, так и нижнее расположение относительно модуля с центральным процессором.

Модуль решает задачи реализации нейронных сетей, цифровой обработки сигналов и изображений. 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> 16 тензорных ядер NMC4 – 1 ГГц.</p> <p><b>RISC:</b> 5 RISC ядер Arm Cortex A5 – до 800 МГц.</p> <p><b>Производительность:</b> FP32 – 512 GFLOP/s. FP64 – 128 GFLOP/s.</p> <p><b>Техпроцесс:</b> 28 нм КМОП.</p> <p><b>Мощность (типичная/максимальная):</b> 12 Вт / 25 Вт.</p> <p><b>Напряжение питания:</b> 12 В (5 В - опция).</p> <p><b>Температурный диапазон:</b> -40С .... +60°С.</p>	<p>5 Гб памяти DDR3L (до 32 Гб/с). 2 x 4Гбит SPI NAND FLASH (Опция) PCIe 2.0x4 (Endpoint). Системный разъем: PCIe/104 type 2. Ethernet 100 Мб/с с поддержкой протокола EDCL (опция). JTAG.</p>	<p><b>Драйверы:</b> Windows (7,10), Linux (включая "AstraLinux" и "Эльбрус ОС").</p> <p><b>NMC SDK:</b> IDE (open VS-code), компиляторы, отладчики ARM и NMC. Bare Metal SDK. БЗИО, БУПВ (MPI, GAS), OpenCL. Набор специализированных библиотек математических функций (BLAS, ЦОС).</p> <p><b>Программный модуль NMDL+:</b> Фирменный компилятор ИНС DarkNet и ONNX. Библиотека для применения ИНС. Библиотека трансляции моделей. Библиотека для подготовки изображений. Набор утилит для применения. Поддержка ЯП: C/C++, Python.</p>

## Решает задачи



# Производительность в реальных тестах

Ниже приведено сравнение производительности реализации (Inference) некоторых глубоких нейронных сетей на аппаратных платформах построенных на базе СБИС K1879BM8Я (NM Card/Card Mini, NM Mezzo/Mezzo Mini, NM Mezzo PCIe/104) и Nvidia Jetson Nano.

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

**Latency** – задержка до получения первого обработанного кадра.

**Multi unit mode** – режим параллельной обработки одного кадра.

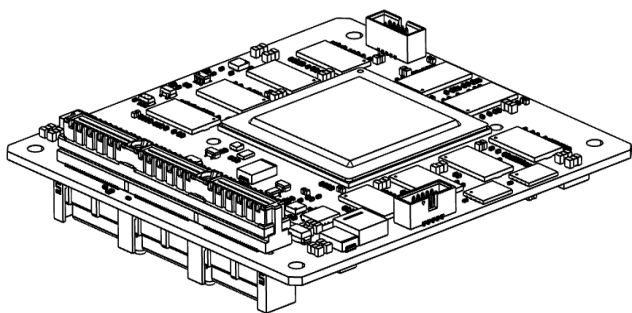
**Batch-mode** – режим одновременной обработки нескольких кадров.

## Сравнение производительности реализации глубоких нейронных сетей.

Модель нейросети (размер изображения)	Аппаратная платформа (режим работы с кадрами)	FPS	Latency, мс	Удельная производительность, FPS/Вт
resnet_50_imagenet (224x224x3)	K1879BM8Я (multi unit mode)	13	76	1.08
	K1879BM8Я (batch mode)	21.19	188	1.77
	Nvidia Jetson Nano	37	27	3.70
inception_v3_imagenet (299x299x3)	K1879BM8Я (multi unit mode)	14.8	67	1.23
	K1879BM8Я (batch mode)	22.42	178	1.87
	Nvidia Jetson Nano	6.45	155	0.65
yolo_v3_coco (416x416x3)	K1879BM8Я (multi unit mode)	3.94	253	0.33
	K1879BM8Я (batch mode)	4.61	865	0.38
	Nvidia Jetson Nano	4.93	139	0.49
yolo_v3_tiny_coco (416x416x3)	K1879BM8Я (multi unit mode)	29.42	33	2.45
	K1879BM8Я (batch mode)	35	114	2.92
	Nvidia Jetson Nano	15	67	1.50
yolo_v5s_coco (640x640x3)	K1879BM8Я (multi unit mode)	5.06	197	0.42
	K1879BM8Я (batch mode)	5.45	733	0.45
	Nvidia Jetson Nano	10	100	1.00
yolo_v7 (640x640x3)	K1879BM8Я (multi unit mode)	1.33	750	0.11
	K1879BM8Я (batch mode)	1.36	2941	0.11
	Nvidia Jetson Nano	3	333	0.30
yolo_v7_tiny (640x640x3)	K1879BM8Я (multi unit mode)	11.28	88	0.94
	K1879BM8Я (batch mode)	12.53	319	1.04
	Nvidia Jetson Nano	16	63	1.60

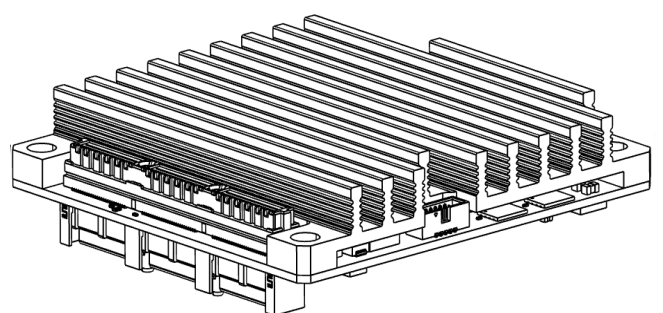
## Габариты модуля

Исполнение без радиатора

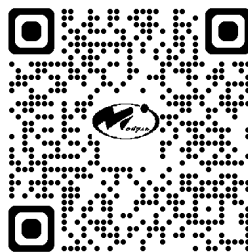


96x90,2x22 мм  
Масса 80 гр.

Исполнение с радиатором



96x90,2x37 мм  
Масса 210 гр.



Дополнительная информация  
доступна по ссылке

