

# Российский модуль для нейронных сетей



Отечественные модули NM (NeuroMatrix) Card предназначены для персональных компьютеров и серверов. Одно из решений данной линейки, модуль NM Card Mini, используется для реализации предобученных нейронных сетей, а также широкого спектра задач цифровой обработки сигналов (ЦОС). В статье рассмотрены технические характеристики и функциональная схема модуля, представлена его микросхема K1879VM8Я.

АО НТЦ «Модуль», г. Москва

Разговоры о российской электронике обычно сопровождаются скептическими комментариями: у широкой публики принято считать, что отечественной здесь может быть разве что сборка. В настоящей статье мы познакомимся с компанией, которая решительно опровергает это предубеждение.

Научно-технический центр «Модуль» более 30 лет успешно работает на российском рынке наукоемких технологий. С 1995 года компания создает высокопроизводительные процессорные ядра и аналого-цифровые системы на кристалле. Благодаря высокой квалификации сотрудников и передовому оснащению НТЦ «Модуль» разрабатывает и производит аппаратуру управления и контроля самых совершенных авиационных и космических систем, аппаратно-программные решения в области нейронных сетей, в том числе для обработки видеопотока и изображений, навигации, связи, обнаружения и распознавания объектов, а также занимается контрактным выполнением опытно-конструкторских и научно-исследовательских работ.

В области искусственного интеллекта (ИИ) НТЦ «Модуль» разрабатывает и производит микропроцессоры и модули цифровой обработки сигналов на базе оригинальной векторно-матричной архитектуры NeuroMatrix.

В ассортименте продукции компании для систем ИИ особую роль играют модули NM (NeuroMatrix) Card, предназначенные для персональных компьютеров и серверов. Изделия обеспечивают реализацию обученных нейронных сетей, а также позволяют решать широкий спектр задач цифровой обработки сигналов (ЦОС): изображений, видеопотока, звуковой и текстовой информации.

Например, модуль NM Card Mini (рис. 1, 2) используется для реализации предобученных нейронных сетей, а также для многих задач ЦОС. Изделие может применяться в высокопроизводительных вычислительных комплексах, для автоматизации процессов производства, в системах цифровой обработки сигналов и изображений, машинного зрения, для облачной обработки данных, в робо-



Рис. 1. Модуль NM Card Mini: внешний вид

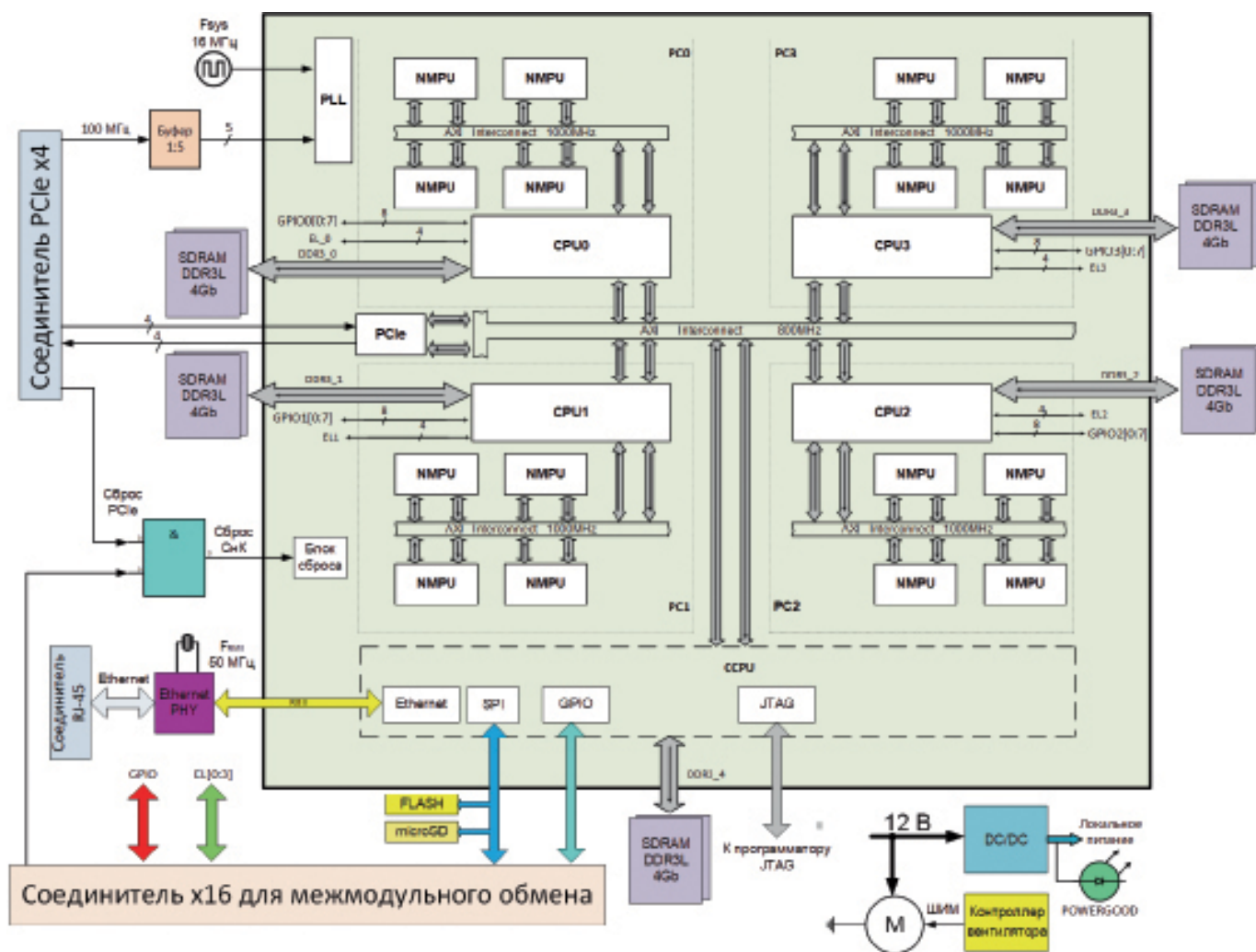


Рис. 2. Упрощенная функциональная схема модуля NM Card Mini

тотехнике, телекоммуникационных и связанных системах, а также в радиотехнических системах и комплексах.

Отметим, что NM Card Mini обладает небольшими размерами при сохранении идентичных вычислительных возможностей своего предшественника – модуля NM Card. Габаритный размер по высоте печатной платы NM Card Mini уменьшен до 121,5 мм (со 141 мм), что позволило полностью «скрыть» модуль за планкой PCIe. За счет этого появились широкие возможности применения NM Card Mini в промышленных компьютерных системах, где габаритные размеры являются важным фактором выбора оборудования.

Как и в модуле NM Card, в варианте Mini использована микросхема K1879BM8Я (NM 6408) производства НТЦ «Модуль» (рис. 3). Микросхема ориентирована на решение задач обработки больших потоков данных в реальном масштабе времени (циф-

ровая обработка сигналов, обработка изображений, навигация, связь, эмуляция нейронных сетей и т. д.).

Оперативная память модуля NM Card Mini – 5 Гб типа DDR3L (до 32 Гб/с). Такого объема памяти с запасом хватает для хранения всех данных, необходимых для реализации современных искусственных нейронных сетей.

Максимальная потребляемая мощность модуля – не более 25 Вт при типовой потребляемой мощности 10,5 Вт (измерено на испытательном тесте MLPerf inference r0.5).

Программное обеспечение NM Card Mini предоставляется для хост-систем, работающих под управлением операционных систем Linux и Windows. Разработаны драйверы для Windows 7 и Windows 10, а также для Linux (в том числе для отечественных дистрибутивов AstraLinux и «Эльбрус ОС»). Системное ПО поддерживает работу модуля под управ-

лением хост-систем всех популярных архитектур: x86 (Intel, AMD), ARM (в том числе ЦПУ «Байкал»), а также E2K (ЦПУ «Эльбрус»). На сайте разработчика представлены дистрибутивы системного и инструментального программного обеспечения.



Рис. 3. Микросхема K1879BM8Я производства НТЦ «Модуль»

Таблица 1. Производительность модуля NM Card Mini на примере некоторых современных глубоких нейронных сетей

Нейронная сеть (размер изображения в пикселях)	Режим обработки 1 кадра		Режим параллельной обработки до 4 кадров	
	Число обработанных кадров за секунду	Задержка до получения обработанного кадра, мс	Число обработанных кадров за секунду	Задержка до получения обработанного кадра, мс
alexnet (227 × 227)	13,09	76	13,6	294
inception v3 (299 × 299)	8,12	123	12,43	322
inception v3 (512 × 512)	3,93	254	5,44	735
resnet 18 (224 × 224)	27,5	36	55,32	72
squeezenet (224 × 224)	89	11	124	32
yolo v2 tiny (416 × 416)	21	47	30,4	132
yolo v3 (416 × 416)	3,7	270	4	1000

**ПРИМЕЧАНИЕ.** Пакетный режим (Batch-mode) позволяет одновременно обрабатывать несколько кадров изображения.

Кроме того, специально для реализации обученных нейронных сетей разработано фирменное прикладное программное обеспечение NMDL (NeuroMatrix Deep Learning) на ядрах NeuroMatrix.

В его состав входят следующие программные инструменты: фирменный компилятор обученных нейрон-

ных сетей, работающий с форматами DarkNet или ONNX; библиотека для применения ИНС; библиотека трансляции моделей и библиотека для подготовки изображений, а также набор утилит для применения.

Производительность модуля NM Card Mini показана на примере реализации ряда глубоких нейронных сетей

в табл. 1. Подробные данные можно посмотреть на сайте НТЦ «Модуль».

АО НТЦ «Модуль», г. Москва,  
тел.: +7 (495) 531-3080,  
e-mail: rusales@module.ru,  
сайт: www.module.ru

**HI-TECH**  
2023  
25–27 апреля  
Международная выставка инноваций и конкурс научных разработок

КВЦ «ЭКСПОФОРУМ»  
Петербургское шоссе, 64

**РЕСТЭК®** hitech-expo.ru  
+7 (965) 771 7751  
o.dorokhova@restec.ru

В рамках выставки пройдет Санкт-Петербургский промышленный конгресс

**РЕАЛИЗАЦИИ**