

Первый TPU на углеродных трубках станет основой для суперкомпьютеров будущего

Разработка обещает ускорить развитие ИИ и сэкономить много энергии.

Ученые из Пекинского университета и других научных институтов Китая создали первый в мире тензорный процессор (Tensor Processing Unit (TPU) – это специализированный микрочип для ускорения вычислений, разработанный Google для поддержки их искусственного интеллекта и машинного обучения. Эти чипы особенно эффективны для задач, связанных с нейронными сетями, таких как глубокое обучение. Они обеспечивают более высокую производительность и эффективность по сравнению с традиционными центральными процессорами (CPU) и графическими процессорами (GPU) для специфических задач ИИ. TPU оптимизированы для быстрых вычислений с тензорами — многомерными массивами данных, которые являются ключевым компонентом алгоритмов машинного обучения." data-html="true" data-original-title="TPU" >TPU) на основе углеродных нанотрубок, который может значительно повысить энергоэффективность алгоритмов искусственного интеллекта.

Исследователи были вдохновлены быстрым развитием приложений вроде ChatGPT и Sora, а также тем, как Google разрабатывает свои TPU. Они отмечают, что традиционные кремниевые полупроводниковые технологии всё чаще не справляются с обработкой огромных объёмов данных, необходимых для работы современных систем ИИ.

Новый чип состоит из 3000 полевых транзисторов на углеродных нанотрубках, организованных в 3×3 вычислительных блока. Эти 9 блоков образуют систолическую матричную архитектуру, способную параллельно выполнять операции двухбитной целочисленной свёртки и умножения матриц.

Архитектура чипа продумана таким образом, чтобы поддерживать поток систолических входных данных. Это позволяет сократить количество операций чтения и записи в статическую оперативную память (SRAM), а значит, растет экономия энергии.

Каждый вычислительный блок получает данные от соседних блоков сверху и слева, самостоятельно вычисляет частичный результат и передает его дальше вправо и вниз. Блоки оптимизированы для выполнения 2-битных операций умножения с накоплением (MAC) и умножения матриц как для знаковых, так и для беззнаковых целых чисел.

На основе разработанного углеродного тензорного процессора исследователи построили пятислойную сверточную нейросеть, способную распознавать изображения с точностью до 88% при потреблении энергии всего 295 мкВт. Это самое низкое

Первый TPU на углеродных трубках станет основой для суперкомпьютеров будущего

энергопотребление среди всех подобных технологий.

Результаты системного моделирования показывают, что углеродный транзистор, использующий 180-нанометровый технологический узел, может достигать тактовой частоты 850 МГц, а его энергоэффективность превышает 1 TOPS/Вт (триллион операций в секунду на ватт).

Исследователи отмечают: производительность и энергоэффективность их подхода вполне можно улучшить. Для этого они планируют использовать выровненные полупроводниковые углеродные нанотрубки, уменьшить размер транзисторов, увеличить разрядность вычислительных блоков и реализовать логику КМОП.

Кроме того, учёные рассматривают возможность интеграции углеродного TPU в кремниевые микросхемы для создания трёхмерных структур. В такой конфигурации кремниевый процессор будет располагаться снизу, а углеродный TPU сверху в качестве сопроцессора.

Разработанная архитектура тесно связанных вычислительных блоков и систолический поток данных позволяют значительно сократить количество обращений к памяти, что является ключевым фактором в снижении энергопотребления. Это особенно важно в контексте растущих потребностей алгоритмов искусственного интеллекта в вычислительных мощностях.