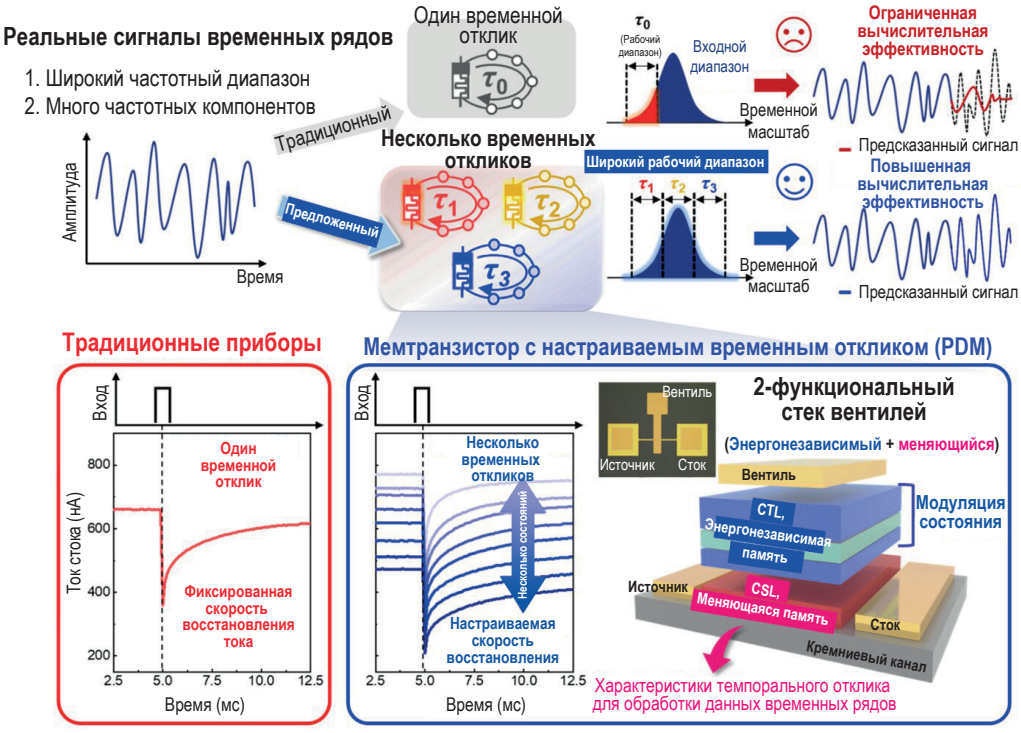
Чип Chameleon – 40-кратное уменьшение ошибок предсказания

По материалам Корейского института передовых наук и технологий

Искусственный интеллект – это не только новые возможности и кардинальное повышение эффективности в различных сферах человеческой деятельности, но и новые подходы к вычислительным ресурсам, на которые опирается AI и без которых он не способен существовать. Довольно быстро стало понятно, что традиционные полупроводниковые приборы, успешно применявшиеся многие десятилетия, уже не отвечают потребностям искусственного интеллекта. Это стимулирует ученых и инженеров ведущих стран мира разрабатывать новые концепции и создавать полупроводниковые приборы, существенно отличающиеся от того, что было сделано ранее.

Полупроводниковые AI-приборы становятся все более программируемыми. Исследователи из Корейского института передовых наук и технологий – KAIST (Korean Advanced Institute of Science and Technology) – разработали прибор, характеристики отклика которого можно программировать для обработки данных, меняющихся с разными скоростями. Эта технология обеспечивает до 40-кратного уменьшения количества ошибок предсказания для меняющихся во времени данных. Ожидается, что она позволит улучшить эффективность работы средств искусственного интеллекта, применяемых в беспилотных транспортных средствах, роботах и носимых устройствах, в режиме реального времени.

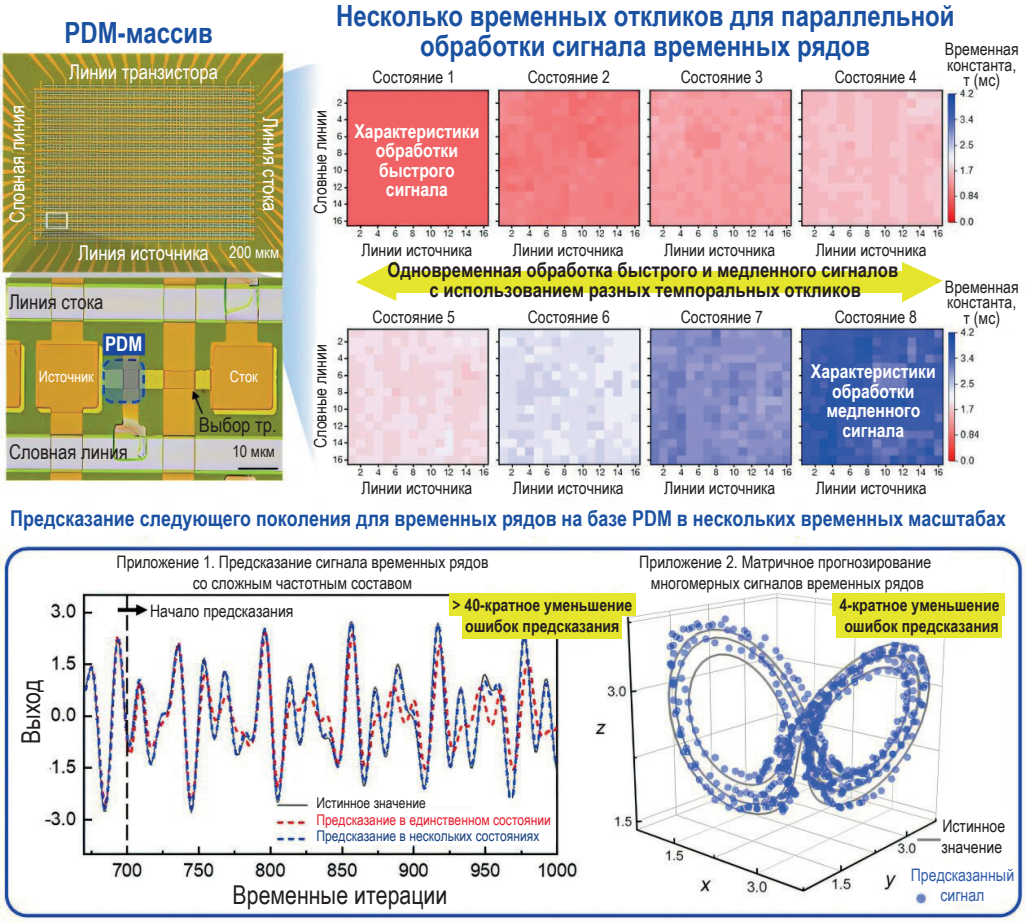
Группа исследователей, возглавляемая именным профессором Чхве Синхеном из Школы электротехники и Высшей школы полупроводниковых технологий, разработала программируемый динамический мемтранзистор (PDM) – полупроводниковый прибор, характеристики временного отклика которого можно устанавливать в различные состояния и поддерживать созданную конфигурацию. Есть также возможность создавать массивы из таких приборов. Научная работа об этом была опубликована в журнале Nature Communications.



Концепция и принцип работы программируемого динамического мемтранзистора PDM (programmable dynamic memtransistor) для обработки сигналов временных рядов в разных временных масштабах. Данные временных рядов в реальном мире содержат и быстро, и медленно меняющуюся информацию, требуя для осуществления эффективного анализа различных характеристик реагирования во времени. Традиционные полупроводниковые приборы обладают фиксированным единым временем отклика, что ограничивает диапазон временных масштабов, которые они способны обработать. В отличие от них, приборы PDM содержат 2-функциональную вентильную структуру, которая позволяет программировать их характеристики времени отклика на нескольких уровнях, от быстрого до медленного, и сохранять эти характеристики энергонезависимым способом. Применяя эти различные темпоральные отклики совместно, приборы PDM могут эффективно извлекать информацию из широкого спектра временных масштабов и повышать точность предсказания сигналов последовательных рядов (диаграмма KAIST)

Мемтранзистор представляет собой полупроводниковый прибор следующего поколения, в котором функция хранения информации, присущая памяти, сочетается с вычислительной функцией транзистора. В PDM способность обработки данных при сохранении ранее содержавшейся информации позволяет корректировать характеристики отклика этого прибора и сохранять их для поступающих данных.

Современные компьютеры и смартфоны нуждаются в сложной программной обработке для анализа данных, меняющихся с те-



Структура интегрированного массива PDM с разными характеристиками темпорального отклика и его эффективность в обработке сигналов временного ряда разного масштаба. За счет присвоения разных временных констант – от быстрого до медленного отклика – отдельным приборам массив PDM способен обрабатывать несколько сигналов временного ряда параллельно, одновременно извлекая информацию в разных временных масштабах. Опираясь на этот подход, массив обеспечивает сокращение ошибок предсказания более чем в 40 раз для составных по частоте сигналов временных рядов и более чем в 4 раза для многомерных хаотических сигналов по сравнению с методами, основанными на едином фиксированном темпоральном отклике (диаграмма KAIST)

чением времени. Это выливается в большие вычислительные нагрузки и высокое энергопотребление. Для решения проблемы исследователи изучали технологии, которые позволяют полупроводниковым устройствам обрабатывать данные напрямую. Однако традиционные приборы имеют фиксированные скорости отклика, которые нельзя изменить после изготовления прибора.

Исследователи преодолели это ограничение, разработав двухслойную структуру внутри транзистора, объединяющую слой хранения заряда, в котором аккумулируются и обрабатываются данные, со слоем захвата электронов, который управляет скоростью отклика, опираясь на энергонезависимый метод.

В приборе PDM, созданном группой исследователей, входящие данные обрабатываются в слое хранения заряда, тогда как слой захвата электронов управляет (на нескольких уровнях) скоростью восстановления, с которой полупроводник возвращается в свое первоначальное состояние. В ходе экспериментов исследователям удалось изменять время восстановления тока примерно в 5 раз, а характеристическую частоту – более чем в 10 раз.

В частности, в опытах, преусматривавших предсказание данных, когда быстрые и медленные изменения были причудливо перемешаны, прибор PDM показал уменьшение числа ошибок предсказания в 40 раз по сравнению с традиционными полупроводниковыми приборами с фиксированным временем отклика.

PDM обеспечивает точную обработку информации даже при изменении скорости написания от руки или движения объектов, используя характеристики отклика, настроенные в соответствии с различными временными масштабами входных сигналов. Как только характеристики отклика заданы, прибор запоминает их без необходимости непрерывной подачи питания от внешнего источника, равно как не нуждается в сложной предварительной обработке входных данных. Поскольку прибор полностью совместим с материалами, используемыми и широко применяемыми в процессах изготовления массовых полупроводниковых приборов, он тоже полностью подходит для массового производства и коммерциализации.

Коллектив исследователей изготовил PDM-массив и использовал его для предсказания сложных данных, подтверждая достижение точности, сравнимой с той, что обеспечивают традиционные системы на базе программного обеспечения, но с существенно меньшим энергопотреблением.

«Наше исследование демонстрирует AI-полупроводник, характеристики отклика которого можно программировать для эффективной обработки данных, скорость которых меняется по-разному, – отметил Чхве. *– Мы ожидаем, что эта технология станет ключевой, улучшая производительность AI-устройств, таких как автономные транспортные средства, роботы и носимые приборы, параллельно сокращая их энергопотребление».*

В завершение надо отметить, что хотя корейские исследователи вполне конкретно определили первоначальные сферы применения созданной ими технологии, высока вероятность, как это часто бывает, что технология выйдет за планируемые пределы и найдет более широкое применение, в том числе и в сфере создания медиаконтента.