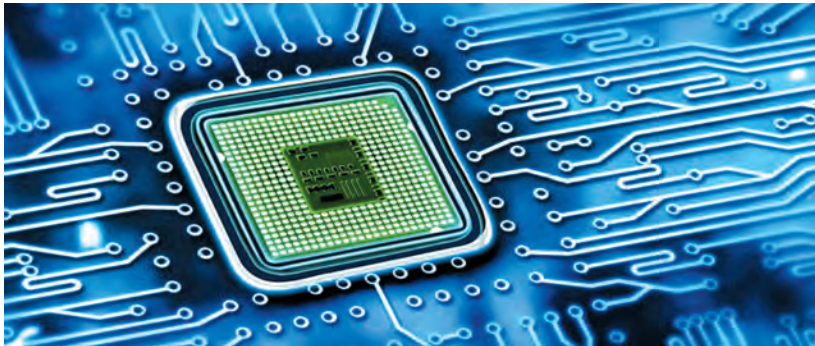


# Дела процессорные: претенденты бросают вызов



Прошлый год принес немало интересных изменений в сегмент серверных процессоров. Существенно укрепляет свои позиции AMD, чипы на базе микроархитектуры ARM находят все более широкое применение, хотя гегемонии Intel пока ничто не угрожает.

Сегмент серверных процессоров развивается очень динамично. Несмотря на все сложности, связанные с разработкой и производством чипов, здесь есть целый ряд компаний, мечтающих оспорить безоговорочное лидерство Intel. Существенно потеснить гегемона в ближайшие годы вряд ли удастся, но отвоевать заметные доли рынка вполне под силу, особенно если речь идет об AMD, которая в 2019 году триумфально вернулась в серверный сегмент. Чипы на базе микроархитектуры ARM не только становятся производительнее, но и обрастают партнерской инфраструктурой, которая позволяет им проникать в новые сегменты глобального серверного рынка. При этом, несмотря на общее лидерство, у Intel на данный момент отмечается определенное технологическое отставание от конкурентов, что, очевидно, заставляет руководство процессорного гиганта немного нервничать. Пока что это не слишком большая проблема, но рынок высоких технологий развивается быстро и ситуация может измениться в любой момент.

## AMD — новый дзен

В 2019 году AMD отпраздновала 50-летний юбилей. Но это не единственная хорошая новость. За последние несколько лет компания совершила решительный прорыв на рынке серверных процессоров. Успехи производителя напоминают о временах Opteron и позволяют говорить о полноценном возвращении AMD в сегмент корпоративных решений. Все благодаря новой микроархитектуре Zen. Первые чипы на ее основе, выполненные по 14-нм техпроцессу, появились в 2017 году. Но уже в 2018-м появилась Zen 2, и в августе 2019-го было представлено широкое семейство коммерческих процессоров на ее основе под общим названием EPYC Rome.

Новая микроархитектура подразумевает уже 7-нм технологии, в угоду чему AMD даже сменила производителя — вместо GlobalFoundries, которая отказалась переходить на новый техпроцесс, чипы для компании теперь производит тайванская TSMC. При этом основной

конкурент — Intel — в своих новых серверных процессорах, которые также были представлены в 2019 году, остался на 14-нм.

В зависимости от модели EPYC Rome (рис. 1) может содержать от 8 до 64 двухпоточных вычислительных ядер и до 256 Мбайт кэш-памяти третьего уровня (по 4 Мбайт на ядро). Тактовая частота составляет 2–3,1 ГГц при обычной работе и 3,2–3,4 ГГц в режиме ускорения. При этом среднее тепловыделение самых мощных моделей не превышает 225 Вт (пиковое — не более 240 Вт).



Рис. 1. Новый 64-ядерный процессор AMD EPYC Rome 7742

Есть версии для 1- и 2-процессорных серверов. Каждый чип, вне зависимости от числа ядер, поддерживает 128 линий PCI Express 4.0, что стало важным достижением для новой микроархитектуры. Любой EPYC Rome имеет встроенный 8-канальный контроллер оперативной памяти с поддержкой DDR4 ECC (вплоть до 3200 МГц). Но возможно, главным конкурентным преимуществом новых чипов является цена, которая существенно ниже, чем у «одноклассников» от Intel — в ряде случаев отличие в два раза и более.

Например, в рамках официальной презентации процессоров на базе Zen 2 были представлены результаты

сравнения производительности старших моделей — Intel Xeon Platinum 8280M и AMD EPYC 7742 с помощью теста SPECrate 2017\_int\_peak. По итогу получены результаты в 180 баллов для первого процессора и 385 для второго (здесь чем больше — тем лучше). При этом их цена отличается также почти вдвое не в пользу Intel — \$13 тыс. для Xeon и чуть менее \$7 тыс. для EPYC. Конечно, к результатам тестов, обнародованных на презентации самой же AMD, можно относиться с определенной долей скепсиса. Но и независимые проверки дают похожие показатели. Например, профильный ресурс **ServeTheHome** протестировал сервер, оснащенный двумя флагманскими чипами EPYC 7742 с 64 ядрами в каждом. Используя ОС Linux (Ubuntu 19.04) и бенчмарк Geekbench 4.4.2 Pro, удалось получить устойчивый результат в диапазоне 183–188 тыс. баллов, притом что 4-процессорный сервер Dell PowerEdge R840 с четырьмя 28-ядерными процессорами Intel Xeon Platinum 8180M первого поколения показал результат 155 тыс. баллов. Правда, у Intel есть еще более мощные и производительные процессоры, чем 8280M, а в случае AMD на EPYC 7742 (и специальной модификации 7H12, которая не сильно от него отличается) ее серверный модельный ряд заканчивается.

Похоже, AMD поймала кураж на рынке. Так, если в конце 2017 года ее доля в серверном сегменте составляла 0,8%, то в 2018-м выросла сразу в четыре раза — до 3,2%, а по итогам 2019-го должна составить не менее 5–6%. Если так пойдет и дальше, то вполне возможно, что к концу 2020-го компания завоюет 10% рынка. Во всяком случае, такие прогнозы озвучивает отраслевой ресурс Digitimes. Это выглядит вполне реалистично ввиду того, что о поддержке EPYC в своих решениях заявили все крупнейшие мировые производители серверов — HPE, Dell, Supermicro, Lenovo, ASUS, Tyan и многие другие. Спешно подтягиваются и разработчики программного обеспечения.

Серверы на базе Rome заказывают для своих дата-центров Google, Twitter, Facebook, присматривается к ним и потоковая видеослужба Netflix. К тому же на базе процессоров AMD строится все больше суперкомпьютеров. Так, они будут использованы для перспективной системы Cray Frontier вычислительной мощностью 1,5 эксафлопс, создание которой должно завершиться в 2021 году (к моменту создания это будет, возможно, самый мощный суперкомпьютер в мире). Но что более важно для популяризации чипа — EPYC все больше проникает в публичные облачные платформы. Например, в 2019 году Amazon Web Services (AWS) запустила новые виртуальные машины (инстансы) T3a на базе этих чипов в составе публичного облака EC2. В конце 2019 года инстансы на базе EPYC Rome стали доступны также и в Microsoft Azure.

Но AMD не останавливается на достигнутом. В середине нынешнего года готовится презентация очередной версии микроархитектуры Zen 3 и нового поколения процессоров на ее основе — EPYC Milan. Основным преимуществом этих 7-нм чипов будет использование

четырёх потоков на каждом ядре вместо нынешних двух. В 2021–2022 годах ожидается появление Zen 4, а процессоры EPYC Genoa, которые планируется создать на ее основе, будут поддерживать интерфейс PCIe 5.0 и смогут работать с оперативной памятью типа DDR5. Возможно также, что их начнут производить по 5-нм техпроцессу, но это пока неподтвержденная информация.

## Intel — конец Itanium и новые горизонты

Несмотря на очевидные успехи AMD, позициям Intel в серверном сегменте пока ничто не угрожает — компания уверенно контролирует около 95% этого рынка, и даже в случае потенциального ухудшения ситуации у нее есть немало технологических козырей, которые помогут выровнять положение. Тем не менее Intel испытывает и целый ряд проблем, связанных с освоением новых технологий, которые становятся все более сложными. Например, периодически возникают скандалы, связанные с внезапным обнаружением уязвимостей в чипах производителя — в 2017 году были выявлены угрозы Meltdown и Spectre (которые приобрели широкую известность в 2018-м), а в прошлом году этот список пополнился целой серией «дырок», выявленных специалистами по кибербезопасности. Новые уязвимости получили свои названия — RIDL, Fallout и ZombieLoad, и по данным Intel, уже нейтрализованы. Но сам факт их появления заставил настояться многих заказчиков.

Еще одна существенная трудность, с которой столкнулась компания, вызвана переходом на новый техпроцесс. После освоения 14-нм технологии планировался масштабный переход на 10-нм, но с этим возникли проблемы. По данным зарубежных информационных ресурсов, продукты Intel, созданные по новому техпроцессу, имеют вдвое более высокий уровень отбраковки по сравнению с 14-нм. В связи с этим, как ожидается, 10-нм изделия будет производить только одна фабрика вместо четырех, как планировалось изначально. Кое-кто из экспертов вообще считает, что 10-нм технология Intel обречена на провал, поскольку не способна обеспечить существенный выигрыш в производительности по сравнению с процессорами предыдущего поколения. Возможно, это лишь домыслы, поскольку Intel готовит в 2020 году презентацию семейства серверных чипов под кодовым названием Ice Lake-SP, которые будут выпускаться именно по 10-нм техпроцессу.

Но прежде чем перейти к новым решениям, которых в 2019 году было представлено изрядное количество, скажем несколько слов о том, что уже отжило. Так, в прошлом году, наконец, была поставлена финальная точка в истории с агонизирующей серией процессоров Itanium. Чипы, появившиеся на рынке в 2001 году, не оправдали ожиданий по продажам с самого начала, но все же их использование удалось продавать в отдельных узких сферах. В 2017 году появилась последняя серия семейства — Itanium 9700 под кодовым именем Kittison.

И вот в начале прошлого года стало официально известно о том, что прием заказов на эти процессоры прекратится 30 января 2020-го, а все поставки будут полностью свернуты 29 июля 2021 года. На этом 20-летняя история чипа, можно считать, окончена, а все задачи, характерные для Itanium, будут выполнять новые Xeon.

Также в 2019 году начал окончательно сворачиваться еще один интересный проект — в мае Intel распространила уведомление о старте программы по прекращению поставок последних моделей ускорителей вычислений Xeon Phi (выпускаемых с 2017 года по 14-нм техпроцессу). Эти решения в свое время активно развивались, находили применение в суперкомпьютерах и других системах, которым требовалась повышенная производительность. Но, похоже, здесь сказался результат конкуренции с Nvidia, чьи решения семейства Tesla фактически подмяли под себя весь массовый рынок ускорителей. Вместе с тем Intel не отказывается от данного сегмента вообще, разрабатывая новые решения. В частности, в ноябре 2019 года на отраслевой конференции SC19 компания представила принципиально новую архитектуру ускорителей X<sup>e</sup> (Exascale for everyone), коммерческие продукты на базе которой должны появиться в ближайшие годы. Как сообщается, к 2021 году за счет использования X<sup>e</sup> HPC-системам удастся в 500 раз повысить производительность вычислительных узлов на базе технологий Intel. Из более близких перспектив — в 2019 году были выпущены ускорители Nervana Spring Crest и Spring Hill, предназначенные для задач, связанных с искусственным интеллектом.

Отказалась Intel и от развития собственной технологии интерконнекта Omni-Path, которая позиционировалась как решение для соединения узлов суперкомпьютеров и гиперконвергентных систем. Первое поколение, выпущенное несколько лет назад, обеспечивало скорость передачи данных на уровне 100 Гбит/с, планировался выпуск моделей на 200 Гбит/с. Но в 2019 году проект закрыли, очевидно, посчитав, что более перспективным вариантом является использование стандартных технологий Ethernet и Infiniband. В пользу последнего допущения говорит тот факт, что в марте 2019 года Intel представила свой первый сетевой адаптер для серверов 100 GbE (до этого у компании были лишь многоканальные решения на 25/50 GbE, которые давали 100-гигабитный канал только путем агрегации).

Отметим, что первые адаптеры 100 GbE других производителей начали появляться на рынке еще в 2014–2015 годах, так что Intel не является здесь пионером, но важно, что только сейчас такие продукты начинают получать более-менее массовое признание. Все больше производителей предлагают такие адаптеры на рынке, что в свою очередь, ведет к росту конкуренции и общему снижению цены.

Самым заметным событием 2019 года от Intel стала масштабная презентация целого спектра новых продуктов, которая прошла в начале апреля. Обновилось не только практически все семейство серверных процессоров

Xeon Scalable, но также были представлены новые образцы энергонезависимой оперативной памяти Optane DC Persistent Memory и ПЛИС FPGA Agilex второго поколения.

Главным новшеством для чипов Xeon Scalable второго поколения стал переход на микроархитектуру Cascade Lake, которая, как и предыдущая Skylake, подразумевает 14-нм техпроцесс. Радикальных отличий между первым и вторым поколениями процессоров не слишком много — слегка увеличились тактовые частоты (при сохранении среднего тепловыделения на прежнем уровне), появилась поддержка оперативной памяти DDR4–2933. В остальном базовые характеристики и набор возможностей остались прежними — до 28 двухпоточных ядер, по 1 Мбайт кэш-памяти второго уровня на ядро и до 38 Мбайт общей памяти L3 на весь процессор. Наиболее скоростные версии в режиме Turbo Boost способны работать с тактовой частотой 4,4 ГГц, есть также чипы, которые штатно обеспечивают 3,8 ГГц. В продуктовом ряду, охватывающем сегодня более 60 моделей, имеются комбинации количества ядер, тактовых частот и объема кэш-памяти на любой вкус.

Позже была представлена флагманская серия процессоров Xeon Platinum 9200 (Cooper Lake), в котором количество ядер может достигать 56, объем памяти L3 — 77 Мбайт, тактовая частота — до 3,8 ГГц при среднем тепловыделении 400 Вт. Фактически это два отдельных 16–28 ядерных чипа, упакованных в одном корпусе (рис. 2). Размер процессора составляет 76×72,5 мм (что примерно соответствует размеру игровой карты), это делает его одним из самых крупных в своем сегменте. Впечатляет также и цена — хотя открытого прайс-листа на эти модели нет, поскольку они будут поставляться только в составе комплексных систем, но учитывая, что стоимость 28-ядерной флагманской модели Xeon Platinum 8280L составляет \$13 тыс., логично предположить, что два таких кристалла в одном корпусе будут стоить как минимум вдвое дороже. По сведениям профильных зарубежных СМИ, стоимость Xeon Platinum 9200 составляет не менее \$26–30 тыс. (а возможно и больше) в зависимости от варианта исполнения.



Рис. 2. Процессор Intel Xeon Platinum 9200 второго поколения — его стоимость может доходить до нескольких десятков тысяч \$.

Что касается дальнейших планов компании, то уже к 2021 году для серверных процессоров планируется обеспечить поддержку восьмиканальной оперативной памяти типа DDR5 и шины PCI Express 5.0, а к 2022-му намечен переход на 7-нм техпроцесс.

## ARM утверждается в серверном клубе

Доля серверов на базе процессоров ARM в мировом объеме продаж все еще незначительна, тем не менее различные варианты чипов ARM уже прочно заняли место на рынке решений для суперкомпьютеров — например, они будут использованы в новых системах Cray Shasta, Atos BullSequana, Fujitsu Post-K и многих других. В ноябре 2018 года HPC-кластер на базе ARM впервые попал в список TOP 500. Но более интересно то, что год от года эта микроархитектура все более решительно заявляет о себе в корпоративном сегменте и получает поддержку крупных игроков рынка. Так, еще в 2018 году компания Huawei представила серверный ARM-процессор HiSilicon Hi1620. А в начале 2019-го появился новый более мощный чип — Kunpeng 920. Модель произведена по 7-нм техпроцессу, содержит 64 ядра, работает на частоте 2,6 ГГц, имеет встроенный восьмиканальный контроллер памяти DDR4, поддерживает шину PCI Express 4.0 и сетевой контроллер 100 GbE. Процессоры HiSilicon используются, в частности, для флагманской серии серверов Huawei TaiShan и СХД OceanStor Dorado V6.

При этом сама британская компания ARM, которая является основным разработчиком и держателем патентов на одноименную архитектуру, стремительно наращивает преимущество. В начале прошлого года была объявлена новая стратегия Neoverse, рассчитанная на несколько лет вперед. Она подразумевает три основные вехи, каждая из которых ознаменует выход доступных для лицензирования 64-битных ядер: Ares (7 нм), Zeus (7 и 5 нм), Poseidon (только 5 нм). Благодаря освоению новых техпроцессов и вводимым усовершенствованиям предполагается, что производительность чипов будет возрастать на 30% ежегодно. Семейство процессоров Ares (они же Neoverse N1) уже представлено — модели на его основе могут вмещать до 128 вычислительных ядер при суммарном тепловыделении всего 200 Вт.

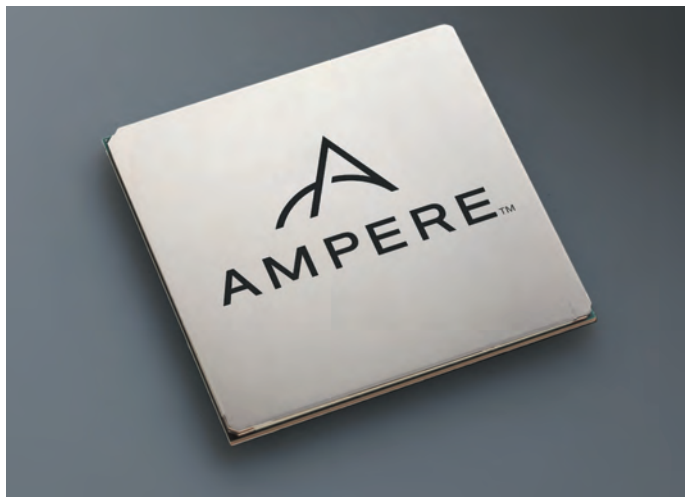


Рис. 3. Новый 80-ядерный ARM-процессор Ampere QuickSilver

Также компания показала чипы для телекоммуникационного оборудования сетей 5G — Neoverse E1. Первые мощные 80-ядерные чипы на базе Neoverse N1 разрабатывает Ampere Computing. Новая серия получила название QuickSilver (рис. 3) — ее производством займется тайванский чипмейкер TSMC. Разработки Zeus и Poseidon будут представлены, соответственно, в 2020 и 2021 годах.

Кстати, с Ampere связана еще одна интересная история последнего времени — в апреле 2019 года стало известно, что Oracle планирует вложить в эту компанию \$40 млн. Рынок воспринял этот шаг в качестве косвенного подтверждения того, что инвестор собирается вскоре отказаться от собственной процессорной архитектуры SPARC (доставшейся «в наследство» в результате покупки Sun Microsystems) и сосредоточится на ARM-решениях для своих будущих серверов. Если это произойдет, то последним крупным разработчиком, поддерживающим SPARC для серверов, останется Fujitsu. Но и эта компания активно развивает свой ARM-процессор A64FX, который, очевидно, является более перспективным.

Напомним также, что виртуальные машины под названием A1, на базе 16-ядерных ARM-процессоров Graviton, уже всюю предлагает публичный облачный сервис AWS EC2, а в конце года уже были доступны инстансы на базе нового Graviton 2 с 64 ядрами (разработанные на базе Neoverse N1). По словам представителей Amazon, в общем случае эти ARM-чипы на 40% превосходят решения Intel по показателю «производительность в расчете на потраченный доллар». Вслед за Amazon, которая является сегодня безоговорочным лидером облачного рынка, виртуальные машины на базе ARM с конце прошлого года стали доступны на платформе Microsoft Azure, где для этой цели используются процессоры Marvell ThunderX2.

Важным событием для ARM стало также и то, что с лета 2019 года о поддержке данной микроархитектуры объявила компания Nvidia, чьи графические ускорители стали фактическим стандартом во многих вычислительных системах, требующих повышенной производительности, в частности решений на базе ИИ. Такой ход существенно расширил области применения ARM и окончательно оформил эти процессоры в качестве полноценного участника мирового «серверного клуба».

То, что произошло на рынке серверных процессоров в 2019 году, в какой-то степени можно считать прелюдией к настоящей борьбе, которая может развернуться в последующие годы. Позиции Intel сильны, но первые крупные успехи позволяют конкурентам чувствовать себя гораздо увереннее, тем более им есть что предложить заказчику — не только сегодня, но и в перспективе ближайших нескольких лет. Все это создает действенные стимулы для дальнейшего развития сегмента и здоровой конкурентной борьбы, от результатов которой, будем надеяться, индустрия только выиграет.

Игорь КИРИЛЛОВ, Сиб