

Математические модели социально-экономических процессов

Проблемы повышения энергоэффективности аппаратных средств в области информационных технологий

Л. Е. ВАРШАВСКИЙ

Аннотация. В статье проводится исследование проблем повышения энергоэффективности аппаратных средств для информационных технологий (ИТ). Рассматриваются основные факторы и направления инновационного развития микро- и нанoeлектроники, определяющей перспективы элементной базы аппаратных средств. Исследуются проблемы энергоэффективности, характерные для центров обработки данных (ЦОД), а также для суперкомпьютеров. Рассматриваются подходы к решению проблемы повышения энергоэффективности аппаратных средств, связанные с созданием квантовых компьютеров, а также с развитием нанofотоники.

Ключевые слова: *информационные технологии, центры обработки данных, суперкомпьютеры, энергоэффективность, квантовые компьютеры, кремниевая фотоника.*

Введение

В настоящее время информационные технологии превратились в важную отрасль, определяющую будущее развитие экономики. По оценке компании Gartner в 2012 г. объем мирового рынка информационных технологий составил около 2 трлн долл. (без коммуникационных технологий). В структуре расходов на информационные технологии около 44 % составляют расходы на информационные услуги, а 40 % — расходы на аппаратные средства (devices) и на системы центров обработки данных (ЦОД) [1].

В России сектор информационных технологий также вошел в стадию зрелости. Объем рынка ИТ-продуктов и услуг в нашей стране составил в 2012 г. 36 млрд долл. По некоторым оценкам в ближайшие

5 лет среднегодовой темп прироста отечественного рынка ИТ-продуктов и услуг составит 5–7 %.

Большая часть объема продаж (свыше 50 % по данным Минэкономразвития) наблюдалась в сегменте аппаратных средств. Ключевая роль этого сегмента сохранится и в будущем, хотя в утвержденном Правительством Российской Федерации прогнозе долгосрочного социально-экономического развития страны на период до 2030 г. предполагается некоторое уменьшение этой доли до 41–44 % в 2020 г. и до 31–35 % в 2030 г. [2].

Этот сегмент характеризуется повышенной капиталоемкостью и ресурсоемкостью. Так, по данным компании DCD Intelligence в 2012 г. капитальные вложения только в ЦОД в мире составили 105 млрд долл. Электрическая мощность всех ЦОД в мире

уже составляет 38 ГВт, что эквивалентно 38 крупным энергоблокам АЭС. По прогнозу DCD Intel-ligence в текущем 2013 г. электрическая мощность ЦОД должна повыситься еще на 17 %.

Существенным фактором, определяющим величину инвестиций в ЦОД, является и величина занимаемой ими площади (white space), которая в конце 2012 г. составляла 26 млн м². В текущем году она может возрасти до 31 млн м² [3], что примерно эквивалентно величине жилого фонда такого российского города, как Новосибирск (1,5 миллиона жителей).

Таким образом, информационные технологии превратились в важную отрасль, которая не только стимулирует инновационное развитие, но и влияет на энергоемкость и капиталоемкость экономики. В связи с этим повышается актуальность анализа возможных направлений повышения эффективности этой отрасли. В настоящей статье исследуются проблемы повышения энергоэффективности аппаратных средств в области информационных технологий. Проводится анализ основных направлений инновационного развития микро- и наноэлектроники, определяющих перспективы аппаратных средств для информационных технологий. Исследуется ряд ключевых экономических проблем, сопутствующих развитию ЦОД, а также суперкомпьютеров. Рассматриваются революционный и эволюционный пути решения проблемы повышения энергоэффективности аппаратных средств, связанные с реализацией квантовых компьютеров, а также с интеграцией кремниевой электроники и фотоники.

1. Основные факторы и проблемы инновационного развития производственной инфраструктуры информационных технологий (на примере микро- и наноэлектроники)

90-е годы прошедшего столетия ознаменовались ускоренным технологическим развитием микроэлектроники и средств вычислительной техники. Менее чем за десятилетний период существенно улучшились все основные параметры, характеризующие качественные и количественные характеристики микроэлектронных схем, долговременной памяти и создаваемых на их основе вычислительных устройств.

Как известно, производительность вычислительных устройств определяется размерами полупроводниковых схем. Чем меньше размеры и выше степень интеграции схем, тем выше их быстродействие. Уменьшение размеров до определенных пределов

ведет также к снижению рассеиваемой энергии, что, в свою очередь, способствует повышению качества и надежности работы схем. Этим объясняется стремление ученых и инженеров снижать топологические размеры схем и повышать степень их интеграции до пределов, определяемых фундаментальными физическими ограничениями.

Еще в 70-х гг. прошлого века видным специалистом в области микроэлектроники, одним из основателей компании Intel Г. Муром была высказана широко известная ныне гипотеза об увеличении в 2 раза степени интеграции интегральных схем через каждые 2 года, впоследствии ставшая драйвером последующих изменений в микроэлектронике [4].

За последние 20 лет прогресс в области миниатюризации и повышения производительности микроэлектронных устройств и компонентов ЭВМ идет значительно быстрее, чем предполагалось даже в конце 1980-х – начале 1990-х гг. С 1995 по 2011 гг. топологические размеры полупроводниковых схем (их изменение не подчиняется закону Мура) уменьшились почти в 17 раз, причем только за последние 8 лет текущего десятилетия — в 3 раза (табл. 1). В конце 2011 г. компанией начато крупномасштабное производство микропроцессоров Ivy Bridge на основе трехмерных транзисторов 3-D Tri-Gate с топологическим размером 22 нм, а в текущем 2013 г. она планирует перейти к выпуску схем с топологическим размером 14 нм. В 2015 г. Intel предполагает перейти на топологический размер 10 нм, а впоследствии — на 7 нм и 5 нм [5].

Следует отметить, что у лидера отрасли, компании Intel, в последнее время переход на новые топологические размеры схем происходит в строгом соответствии с разработанными дорожными картами (roadmaps) [6]. В производстве микропроцессоров Intel придерживается стратегии развития tick-tock. На первом этапе перехода на производство схем с меньшим топологическим размером (этап tick) архитектура микропроцессоров несущественно меняется или остается прежней. Переход к новой платформе процессоров происходит по мере освоения производства чипов с новым, уменьшенным топологическим размером схем (этап tock) [7].

Так, после перехода на производство схем размером 32 нм (с 45 нм), на первом этапе (tick) стали производиться процессоры Westmere, являющиеся усовершенствованным вариантом процессоров Nehalem, которые производились на схемах с размером 45 нм. На этапе tock стали выпускаться процессоры Sandy Bridge с новой архитектурой. После освоения схем с размером 22 нм было начато производство процессоров Ivy Bridge, представляющих собой модификацию Sandy Bridge. Наконец, на этапе tock

Таблица 1

Эволюция технологических процессов фирмы Intel (Источник: [8], с. 249.)

Внедрение, год	1995	1997	1999	2001	2003	2005	2007	2009	2011
Топологический размер, нм	350	250	180	130	90	65	45	32	22

(для схем размером 22 нм) был осуществлен переход к производству процессоров Haswell с новой архитектурой.

Однако в конце первого десятилетия XXI века традиционная полупроводниковая технология стала подходить к границам развития, которые обусловлены фундаментальными законами физики. В результате этого, по мнению экспертов закон Мура, может перестать действовать уже в ближайшие годы [9]. Это связано прежде всего с повышением энергоемкости электронных приборов при уменьшении размеров и ростом капиталоемкости их производства. Здесь необходимо в первую очередь выделить следующие проблемы.

Во-первых, очень серьезной становится проблема отвода тепла, выделяемого из-за непрерывно повышающейся температуры по мере повышения плотности монтажа элементов (уже сейчас в чипах выделяется свыше 100 Вт тепла на 1 см^2 , что уже сопоставимо с величиной плотности энергии, выделяемой на поверхности Солнца) [8].

Во-вторых, физические размеры отдельных элементов приблизились к размерам атомов. Так, компанией Intel планировалось довести толщину оксидных полупроводниковых электродов в своих изделиях всего лишь до 3 атомов (уже в 2005 г. она составляла 1,2 нм). Однако основная трудность в использовании тонких оксидных электродов и диэлектриков связана с возможностью возникновения квантового туннелирования, в результате которого ток управляющего электрода может достигнуть или превзойти ток канала, что не позволит управлять работой транзистора. Эта проблема до сих пор не полностью решена, несмотря на существенный прогресс в разработке и использовании изолирующих слоев затворов транзисторов на основе диэлектриков с высоким значением коэффициента диэлектрической постоянной k (high- k dielectric) [10].

Рассмотрим эту проблему на примере процессоров для компьютеров. На высоких частотах в чипах малых топологических размеров возникают неконтролируемые токи, обусловленные, в частности, тем, что проводники, соединяющие активные элементы, обладают заметными взаимной емкостью и индуктивностью, величина которых, в свою очередь, зависит и от диэлектрической проницаемости используемых в микросхемах материалов. Кроме того, ис-

пользуемые в чипах транзисторы являются источниками и приемниками электромагнитного излучения, что отражается на работе p-n-переходов. Уменьшение топологических размеров схем приводит к возникновению утечек и за счет проявления квантовых эффектов, в частности, туннельного эффекта [8]. Для снижения утечек токов в последнее время используются: медные проводники вместо алюминиевых, материалы с низкими показателями диэлектрической проницаемости — low k , технологии напряженного кремния — Strained Si и кремний-на изоляторе — SOI. Недавно начата разработка кремний-фотонных приборов на одном чипе, в которых проводящие металлические соединения заменяются на оптические (см. п. 4).

Таким образом, несмотря на стремление разработчиков к сдерживанию роста напряжения питания ядер процессоров, тепловая мощность, выделяемая даже в штатных режимах работы процессоров, имеет тенденцию к увеличению. При этом из-за постоянства площади выпускаемых чипов возрастает плотность выделяемой тепловой энергии, что приводит к необходимости использования мощных средств охлаждения. Для решения проблемы повышения энергоэффективности микросистемных схем используется широкий спектр методов, начиная с совершенствования архитектуры создаваемых приборов и заканчивая использованием программных средств [8]. Так, компанией Intel в 2011 г. созданы трехмерные транзисторы с тремя затворами 3-D Tri-Gate transistors (о начале разработки было объявлено в 2002 г.). В этих транзисторах, относящихся к конструкции FinFET, кремниевое тело, в котором находятся исток и сток, выполнено в форме вертикально расположенного «плавника» (fin). «Плавник» обернут тройным затвором, каждая сторона которого управляет током в «своей» области. За счет такой архитектуры обеспечиваются: увеличение доступной для прохождения токов площади транзистора; уменьшение токов утечки; высокое быстродействие процессов переключения схем; уменьшение используемой мощности и электроэнергии; более высокая плотность размещения элементов в схеме. В частности, в трехмерных транзисторах 3-D Tri-Gate с топологическим размером 22 нм при работе с низким электрическим напряжением (примерно 0,7 В) производительность на 37 % выше, а расход энергии на 50 % меньше, чем у транзисторов с топо-

логическим размером 32 нм, выполненных по традиционной планарной технологии. При этом увеличение производственных затрат составляет, по оценкам, всего 2–3 % [11, 12].

Кроме двух указанных проблем можно выделить также проблему, связанную с повышением требований к чистоте при уменьшении топологического размера схем и к точности изготовления соединений, а также проблему, связанную с экономическим фактором — ростом капиталоемкости производства. Рассмотрим последнюю проблему подробнее.

Стоимость производственной линии по производству чипов удваивается каждые три года (в этом состоит так называемый второй закон Мура). Так, переход на производство 450 нм кремниевых подложек (пластин), предусмотренный в дорожной карте полупроводниковой индустрии, потребует, по оценкам специалистов, сооружения фабрик стоимостью 10 млрд долл., что в 2,5 раза выше стоимости передовых фабрик, выпускающих пластины размером 300 нм. В результате такого перехода, который считается выгодным для потребителей пластин (будущих производителей чипов с топологическим размером схем 10 нм), но невыгодным для производителей (так, переход на производство пластин диаметром 300 нм оказался менее эффективным, чем ожидалось, а для некоторых фабрик — убыточным), повышается риск того, что рынок производителей пластин станет чисто олигополистическим с небольшим числом крупных участников. О величине производственных инвестиций в обновление производства свидетельствует и тот факт, что только для перехода на выпуск чипов по технологии 32 нм компания Intel планировала в течение двух лет осуществить капитальные вложения в размере 7 млрд долл. [13]. Капитальные вложения этой компании в сооружение в штате Аризона фабрики по производству чипов с топологическим размером схем 14 нм, которую предполагается ввести в текущем 2013 г., оцениваются в 5 млрд долл. [14].

Указанные проблемы подчеркивают необходимость перехода элементной базы информационных технологий на новые технологические принципы.

2. Центры обработки данных и суперкомпьютеры — драйверы повышения эффективности информационных технологий

2.1. Особенности развития ЦОД за рубежом

Компании, связанные с информационными технологиями и, прежде всего, с Интернетом, осуществляют значительные капиталовложения в сооруже-

ние мощных центров обработки данных (англ. data centers) и их электроснабжение. Так, компания Microsoft затратила на сооружение ЦОД, расположенного неподалеку от Чикаго, 500 млн долл. При этом потребовалось построить 3 электрические подстанции мощностью 198 МВт. Еще один из крупнейших ЦОД, SuperNap в Лас-Вегасе, занимает общую площадь 407 тыс. кв. футов (приблизительно 40 тыс. м²); его электрическая мощность составляет 250 МВт, т. е. для его питания требуется выделять отдельный энергоблок мощной электростанции.

Компания Google, вкладывающая ежеквартально сотни миллионов долларов на создание ЦОД и их инфраструктуры и ориентирующаяся на «зеленую энергетику», начала регулярные долгосрочные закупки больших объемов энергомощностей ветряных электростанций (примером этого служит соглашение Google с компанией Next Era о приобретении у последней более 100 МВт электрической мощности в течение 20 лет) [15].

В результате инвестирования значительного объема средств возросли вычислительные мощности ЦОД, характеризуемые числом серверов. В связи с тем, что в настоящее время отсутствует официальная статистика о показателях деятельности ЦОД, в том числе об их мощности, при проведении исследований приходится использовать оценки отдельных организаций и специалистов. Так, по оценкам профессора Дж. Куми (Jonathan Koomey) с 2000 по 2010 гг. общее количество серверов в ЦОД мира увеличилось с 14 млн до почти 32 млн. Приблизительно треть серверного парка была сосредоточена в США (в 2000 г. 5,5 млн, а в 2010 г. около 12 млн) [16].

По данным Uptime Institute с 2000 по 2006 гг. средняя производительность серверов возросла в 27 раз, в то время как потребление ими энергии — в 3,9 раза. В результате этого средние затраты на электроэнергию за трехлетний период эксплуатации серверов в 2006 г. составляли 50 % от их первоначальной стоимости (в 2000 г. — только 13 %). В настоящее время этот показатель, по-видимому, сравнялся с первоначальной стоимостью серверов, а доля затрат на электроэнергию в общей стоимости владения (total cost ownership — ТСО, которая включает в себя текущие затраты и амортизацию капитальных вложений) приближается к 30 % [17].

В структуре потребляемой серверами мощности наибольшая доля приходится на процессоры (они потребляют от 45 до 200 Вт) и оперативную память (модуль памяти DIMM потребляет от 5 до 21 Вт). О росте потребляемой в мире различными классами серверов электроэнергии свидетельствуют данные табл. 2.

Таблица 2

Оценки объемов потребления электроэнергии серверами ЦОД (млрд кВт.ч.) (Источник: [16].)

Класс серверов/ Год	2000	2005	2010
Массовый (volume)	19,7	50,5	67,2–81,5
Средний (mid-range)	6,7	6,7	5,1–7,3
Высокопроизводительный (high-end)	2,8	4,4	7,8–13,0

Несмотря на то, что рост производительности серверов и суперкомпьютеров существенно опережает рост их энергоемкости, увеличение мощностей ЦОД сопровождается возрастанием объемов энергопотребления. Так, по данным Дж. Куми, в 2010 г. ЦОД мира потребляли 203–271 млрд кВт.ч. электроэнергии, что составляет 1,1–1,5 % от мирового потребления электроэнергии (в 2000 г. соответственно около 71 млрд кВт.ч, или 0,5 %). В США соответствующие показатели составляли в 2010 г. 67–86 млрд кВт.ч (1,5–2,2 %), а в 2000 г. — 28 млрд кВт.ч (0,8 %) [16].

Следует отметить, что значительные объемы электроэнергии ЦОД потребляются объектами их инфраструктуры, главным образом, системой охлаждения (на их долю приходится до 50 % используемой в центрах электроэнергии).

Так, по некоторым оценкам, в США большая часть роста энергопотребления в ЦОД в 2000–2005 гг. (62 %) была связана с увеличением числа серверов и лишь 38 % — с увеличением энергопотребления в расчете на 1 сервер. Это, в частности, обусловлено тем, что вычислительные мощности большей части ЦОД загружены лишь на 10–15 %. В то же время эффективная работа преобразовательных устройств (трансформаторов, адаптеров и др.) достигается при загрузке 50–75 %. При меньшей нагрузке КПД преобразователей падает до 20–40 %, что приводит к значительным потерям энергии [17, 18].

Следует учитывать и то, что на основной потребитель энергии в ЦОД (кроме собственно оборудования информационных технологий (ИТ-оборудования): серверов и суперкомпьютеров) — систему охлаждения приходится также значительная часть помещений ЦОД, что вызывает существенное увеличение капитальных затрат. В целом, по расчетам Uptime Institute, доля объектов инфраструктуры в капитальных вложениях в ЦОД составляет примерно 50 % [19]. В последние годы объем инвестиций в источники бесперебойного питания, электрические генераторы, системы охлаждения, системы обеспечения безопасности и др. уже значительно превышает инвестиции в ИТ-оборудование (серверы, системы хранения информации, коммутаторы

и др.). В 2012 г. из 105 млрд долл., затраченных в мире на ЦОД, на них приходилось соответственно 49 и 35 млрд долл. [20].

В связи с высокой долей объектов инфраструктуры в потреблении электроэнергии ЦОД, при анализе энергоэффективности центров широко используется показатель PUE (power usage effectiveness — отношение полной подведенной мощности к реально затраченной на работу вычислительных систем). Примерные значения этого показателя для некоторых распространенных систем охлаждения приведены в табл. 3.

Выбор системы охлаждения на конкретном ЦОД зависит от природно-климатических условий, наличия и стоимости ресурсов воды и электроэнергии, экологических и ряда других факторов. В последнее время наиболее перспективным способом охлаждения ЦОД и суперкомпьютеров принято считать жидкостное, особенно водяное охлаждение. Это обусловлено тем, что ввиду существенно более высокой теплоемкости воды (более чем в 3200 раз выше теплоемкости воздуха) уменьшаются размеры оборудования и вспомогательных систем, что в свою очередь ведет к снижению площади, а следовательно и капитальных затрат на строительство ЦОД [21].

В качестве примера успешного продвижения современной технологии водяного охлаждения следует отметить использование в центре CINECA системы Aurora Hot Water Cooling technology для суперкомпьютера Euroga. За счет применения высокопроизводительных графических сопроцессоров (ускорителей) NVIDIA Tesla K20, а также перехода на эту систему охлаждения, в центре удалось уменьшить коэффициент PUE до рекордно низкого уровня 1,05 и в то же время добиться рекордного для начала текущего 2013 г. уровня удельной производительности суперкомпьютеров в 3150 Мегафлопс/Вт. По оценкам разработчиков, за счет реализации данных мероприятий размер потенциальной экономии показателя TCO (стоимости владения) в ЦОД может достигнуть 30–50 %, главными составляющими которой являются экономия энергии (до 50 % относительно прежнего уровня) и пятикратное уменьшение пло-

Таблица 3

Показатели энергоэффективности PUE некоторых систем охлаждения ЦОД и суперкомпьютеров.
(По данным источника [21])

Вид охлаждения	Энергоэффективность (PUE)
Классическая система прецизионного кондиционирования с фреоновыми кондиционерами	2,0
Чиллеры с применением свободного охлаждения (Free-cooling)	1,7
Системы непосредственного охлаждения наружным воздухом (Direct Free Cooling (DFC); Natural Free Cooling (NFC))	1,3
Жидкостное охлаждение	До 1,06

щади из-за повышения плотности размещения оборудования [22].

Проведенный в 2012 г. Uptime Institute опрос 1100 организаций, связанных с эксплуатацией ЦОД, показал, что среднее значение показателя PUE в США составляет 1,8–1,9. Это несколько ниже, чем в 2007 г., когда значение этого показателя составляло 2,5 [23]. В то же время в ЦОД таких компаний, как Google, Yahoo, Facebook и Microsoft, значение PUE доходит до 1,07–1,20 [24].

По оценкам Uptime Institute, потенциал энергосбережения в ЦОД в США составляет 50 % от текущего расхода энергии. В числе наиболее энергоэффективных мероприятий, которые могут способствовать достижению этой цели, называются: виртуализация и консолидация серверов; создание энергосберегающих серверов; выключение неработающих серверов и др. [17].

Так, в компании Intel за счет виртуализации ЦОД, связанных с проведением работ по серийному дизайну приборов и схем, удалось повысить загрузку центров обработки данных с 55 % в 2006 г. до 80 % в 2008 г. Переход на новые серверы с процессором Intel Xeon Processor X5570 позволил компании сэкономить 90 % пространства и 86 % энергии по сравнению с более старыми моделями серверов с процессором Intel Xeon Processor. В итоге компании удалось за 3 года уменьшить капитальные затраты на ЦОД на 65 % [25].

В перспективе значительная экономия электроэнергии и затрат в ЦОД связывается с переходом широкого круга пользователей компьютеров к так называемым облачным вычислениям (cloud computing). По оценкам некоторых экспертов, к 2020 г. большая часть пользователей откажется от настольных компьютеров и будет пользоваться вычислительными ресурсами общего пользования с помощью обычных бытовых приборов [26].

Так, по прогнозу аналитической компании Forrester Research следует ожидать существенный рост

объема мирового рынка облачных вычислений — с 40,7 млрд долл. в 2011 г. до более 241 млрд долл. в 2020 г.

По расчетам специалистов Google переход на облачные вычисления может обеспечить общую экономию электроэнергии в размере 68–87 %. Расход электроэнергии на работу серверов и их систему охлаждения может снизиться на 70–90 %. Так, переход на предоставляемый этой компанией облачный сервис Google Apps одного из правительственных агентств США, U. S. General Services Administration (GSA), позволил последнему уменьшить число и мощность серверов соответственно с 324 до 61 и с 163 кВт до 22 кВт, что обеспечивает снижение годовых затрат на электроэнергию с 307,4 тыс. долл. до 22,4 тыс. долл. [27].

В связи с высокой эффективностью облачных сервисов американским правительством принято решение (Federal Data Center Consolidation Initiative — FDCCI) к концу 2015 г. закрыть 800 правительственных ЦОД из 2000, или 40 % от общего числа. Предполагаемый при этом переход на облачные сервисы позволит по оценкам правительству экономить на ИТ-обслуживании более 3 млрд долл. в год. Закрытие ЦОД позволит также высвободить земельные ресурсы и снизить энергозатраты [28]. С начала 2010 г. по январь 2013 г. уже закрыто 420 ЦОД [29].

Следует однако отметить, что широкомасштабный переход потребителей компьютерных услуг на облачные сервисы будет сдерживаться нерешенностью проблем компьютерной безопасности.

2.2. Особенности развития ЦОД в России

Рассмотренные тенденции имеют непосредственное отношение и к нашей стране, хотя у нас ЦОД находятся еще на начальном этапе развития. Так, объем рынка ЦОД в России в 2008 г. оценивался в 2,3 млрд рублей, причем доля Московского региона составляла 78 %, Санкт-Петербурга — 12 %, а регионов — не более 10 % [30].

В 2010 В России насчитывалось около 130 коммерческих ЦОД, а в 2012 г. по оценке агентства iKS-Consulting — 170. Объем рынка в 2012 г. составил 6,2 млрд руб. (рост по отношению к 2011 г. — 22,8 %). Суммарная величина полезной площади помещений достигла 62,6 тыс. кв. м, а число стоек — 18,7 тыс. [31, 32]. Предполагается, что в 2013 году суммарная площадь ЦОД увеличится еще на 22 тыс. кв. м.

Следует отметить неравномерность территориального распределения в стране. Так, на Московский регион приходится 77 % доходов российского рынка коммерческих ЦОД и 64,7 % коммерческих площадей. При этом объем рынка ЦОД Москвы в 2012 г. составил 4,8 млрд руб. (в 2011 г. — 3,9 млрд руб.). Вместе с тем суммарная площадь всех коммерческих центров в Московском регионе сравнима с площадью одного ЦОД SuperNap в Лас-Вегасе.

Характерно, что 9 из 10 крупнейших операторов российских ЦОД (на них приходится 59,5 % всего рынка по площади, 62,1 % по количеству стоек и 61,6 % по доходам) — это московские компании [33]. Вместе с тем, некоторые аналитики отмечают постепенный перенос петербургскими компаниями своих ИТ-ресурсов из ЦОД Москвы в Санкт-Петербург [34].

Следует отметить, что в общем объеме рынка ЦОД обычно не учитывается корпоративный сектор. Вместе с тем крупные банки, нефтегазовые и энергетические компании, а также государственные организации, обладающие мощными собственными ИТ-службами, развивают свою сеть ЦОД. Большая часть корпоративных ЦОД также сосредоточена в Москве (80 % по оценке CNews Analytics) [35].

Характерным примером является Сбербанк, сдавший в эксплуатацию в конце 2011 г. ЦОД «Южный порт», считающийся крупнейшим в Европе. Общая площадь здания центра составляет 16 500 кв. м, причем площадь ИТ-залов — 5000 кв. м. Используемая электрическая мощность составляет приблизительно 25 МВт. В созданном ЦОД интегрируются информационные ресурсы всех территориальных отделений Сбербанка, что позволяет повысить производительность вычислительных процессов и улучшить информационную поддержку подразделений. Инвестиции в объект, на создание которого ушло 15 месяцев, оцениваются в 1,2 млрд долл. Величина годовой экономии, связанной с реализацией этого проекта составит, по оценкам банка, 500 млн долл., а срок окупаемости — 2,5 года [36].

Во многом с активизацией деятельности крупных компаний в области ИТ связан рост объема капитальных вложений в ЦОД в России. По мнению специалистов, в перспективе на базе корпоративных ЦОД крупнейшие компании России с су-

щественной долей государственного участия начнут предоставлять услуги не только собственным подразделениям, но и другим организациям.

Определенное развитие получают в нашей стране и облачные вычисления. Так, если объем рынка облачных вычислений в России в 2009 г. составил 4,8 млн долл., то в 2010 г. уже 35 млн долл., а в 2011 г. — свыше 59 млн долл. (по данным консалтинговой компании IDC). Несмотря на то, что российский рынок пока находится на начальной стадии развития, в стране наблюдается растущий спрос к облачной модели предоставления ИТ-услуг. По недавнему прогнозу компании Orange Business Services, рынок облачных услуг для бизнеса в России увеличится с 4,5 млрд руб. в 2012 г. до 19 млрд руб. к 2016 г. Наибольшим спросом будут пользоваться решения IaaS (инфраструктура как сервис) и SaaS (программное обеспечение как сервис) [37].

По мнению специалистов, для полномасштабной реализации облачных вычислений необходимо решение ряда практических вопросов, связанных с обеспечением безопасности, повышением пропускной способности каналов связи, совершенствованием оборудования и протоколов передачи данных, качественным изменением методов обработки данных, повышением эффективности ЦОД.

Важно отметить, что по показателю энергоэффективности PUE наши ЦОД значительно отстают от зарубежных. Так, по некоторым оценкам, в нашей стране значение этого показателя составляет в среднем 2,2–2,5. Вместе с тем следует отметить, что в ряде новых проектов отечественных ЦОД значение этого показателя снижено до 1,3–1,4 [38].

Проблема повышения энергоэффективности ЦОД является весьма актуальной для нашей страны и ввиду дефицита энерго мощностей, а также высокой стоимости электроэнергии. В то же время строительство новых коммерческих и корпоративных ЦОД сдерживается ростом стоимости подключения 1 кВт электроэнергии (например, в Москве за последние годы она выросла с 1000 до 3000–4000 долларов). В связи с этим в отечественных ЦОД часто используются собственные дизель-генераторы, из-за чего могут возникать экологические проблемы, особенно в случае расположения центров в непосредственной близости от жилых районов (это, впрочем, характерно и для ЦОД зарубежных стран)

Существуют также проблемы, связанные с сертификацией ЦОД. Так, в настоящее время многие отечественные компании присваивают без должного основания вводимым ЦОД уровень Tier 3 (один из наиболее высоких среди 4 уровней надежности). Вместе с тем к 2012 г. сертификаты Uptime Institute получили всего три коммерческих ЦОД в России [39].

Следует отметить и не совсем благоприятные экономические показатели отечественных ЦОД. Так, по данным исследования компании «Мира» срок окупаемости коммерческих ЦОД у нас составляет около шести лет, а инвестиции в ЦОД уровня Tier 3 — 15–25 тыс. долл. на кв. м. [40].

2.3. Проблемы повышения энергоэффективности суперкомпьютеров

В последние годы со стороны широкого круга организаций устойчиво повышается интерес к использованию высокопроизводительных вычислений (high-performance computing, HPC). Важным сегментом рынка высокопроизводительных вычислений является рынок суперкомпьютеров, доходы от продажи которых в 2012 г. составили 5,6 млрд долл. [41].

Наиболее производительные суперкомпьютеры сконцентрированы в исследовательских центрах государственных структур и крупных компаний. Так, 15 из 23 суперкомпьютеров производительностью свыше 1 Петафлопс (Пфлопс)¹, входивших в ноябре 2012 г. в список 500 наиболее производительных суперкомпьютеров мира Top 500 (он публикуется 2 раза в год [42]), эксплуатировалось в исследовательских центрах.

Однако суперкомпьютеры все более широко используются не только в исследовательском и академическом секторах, но и в бизнесе. Так, если в 1997 г. в Top 500, входил 161 суперкомпьютер из коммерческого сектора (industry), то в 2007 г. — уже 248. Примерно такое же число суперкомпьютеров из этого сектора было представлено и в ноябрьском списке Top 500 за 2012 г. (247 суперкомпьютеров). Драйверами активного применения суперкомпьютеров в бизнесе являются: 1) рост затрат на создание и испытание образцов (прототипов) изделий, 2) резкое снижение затрат на виртуальное моделирование, а также на вычисления на суперкомпьютерах [43].

Повышение производительности суперкомпьютеров в последние годы сопровождается ростом их электрической мощности. Так, суммарная электрическая мощность, потребляемая всеми суперкомпьютерами из списка Top 500, устойчиво возрастала со 161 МВт в 2007 г. до 443 МВт в 2012 г. (по данным ноябрьских выпусков Green 500) [44].

В частности, наиболее производительный в мире в 2011 г. суперкомпьютер SPARC64 компании Fujitsu с быстродействием 10 Пфлопс, потребляет около 10 МВт эл. мощности. Согласно рейтингу, прове-

денному в июне 2012 г., необходимая электрическая мощность для наиболее производительного тогда и значительно менее энергоемкого суперкомпьютера Sequoia — система IBM BlueGene/Q, установленного в Ливерморской национальной лаборатории им. Лоуренса, США, имеющего быстродействие 16,32 Пфлопс, также составляет немалую величину — 7,9 МВт. Наконец, в соответствии с последним рейтингом Green 500 за ноябрь 2012 г., самый производительный суперкомпьютер Titan с быстродействием 17,59 Пфлопс имеет электрическую мощность 8,2 МВт [44, 45].

Следует также отметить, что из-за высокой плотности размещения вычислительных мощностей в ограниченном пространстве проблема отвода тепла в суперкомпьютерах стоит более остро, чем в ЦОД. В настоящее время, по мнению специалистов, наилучшие характеристики, как и для ЦОД, имеет технология водяного охлаждения, которая позволяет снизить показатель PUE до 1,05–1,06 [22, 46].

В связи с высокой энергоемкостью суперкомпьютеров, начиная с 2007 г. регулярно публикуются списки Green 500, в которых проводится ранжировка наиболее производительных суперкомпьютеров по величине соотношения между их производительностью R_{max} и используемой электрической мощностью. В соответствии с ноябрьским за 2012 г. списком Green 500, максимальная удельная производительность суперкомпьютеров составляла 2499 Мегафлопс/Вт. В то же время, как показывает анализ приведенных в этом списке данных, у 56 % суперкомпьютеров из 500 самых мощных показатель энергоэффективности составляет менее 10 % от максимального уровня. Следует отметить, что по уровню энергоэффективности и отечественные суперкомпьютеры долгие годы почти на порядок уступали лучшим зарубежным образцам, хотя в 2012 г. ситуация улучшилась (у нас достигнут уровень 1949 Мегафлопс/Вт) [44].

Бурный рост производительности суперкомпьютеров стимулировал возрастание спроса на высокопроизводительные вычисления HPC со стороны исследовательского сектора, государственных структур и бизнеса. В связи с этим перед специалистами ряда высокоразвитых стран, в частности в США, Японии и Китае, поставлена проблема создания суперкомпьютеров эксафлопной производительности (т. е. производительностью более 10^{18} флопс в секунду). При этом необходимо иметь в виду, что в настоящее время не существует принципиальных технических ограничений для создания суперкомпьютеров такого класса на основе существующего оборудования [47]. Однако при достигнутом к настоящему времени уровне энергоэффективности необходимая

¹ В списке Top 500 приводится ранжировка суперкомпьютеров по производительности R_{max} в соответствии с тестом Linpack, связанным с решением линейных уравнений большой размерности.

электрическая мощность экзафлопного суперкомпьютера составила бы 400–500 МВт. С учетом систем охлаждения, при значении PUE, равном 2, для работы такого суперкомпьютера потребовалось бы 800–1000 МВт электрической мощности, что равно мощности крупного энергоблока тепловой или атомной электростанции. Этой мощности достаточно для обеспечения электроэнергией крупного города с населением несколько сот тысяч человек.

Поэтому, например, в США перед разработчиками ставится задача создания экзафлопного суперкомпьютера электрической мощности порядка 20 МВт. Так, с целью ограничения роста масштабов энергопотребления, а также повышения его эффективности, агентством DARPA Министерства обороны США была в 2010 г. разработана программа «Вездесущие высокоэффективные вычисления» (Ubiquitous High Performance Computing — УНПС). В рамках этой программы ставилась задача обеспечить размещение в одном кабинете компьютерные мощности в 1 Пфлопс с энергоэффективностью 50 Гигафлопс/Вт (это в 20 раз больше текущего уровня). Первый образец системы УНПС должен быть создан в 2018 г. [48].

Рассмотрим, насколько реальна такая задача. Статистический анализ динамических рядов показывает, что для наиболее производительных суперкомпьютеров из списка Top 500 показатель $R_{\max t}$ в млрд. Мфлопс связан следующей функциональной зависимостью с числом ядер процессоров PR_t в суперкомпьютере и топологическим размером чипов hp_t в нм:

$$\ln R_{\max t} = a_1 + b_1 \times \ln(PR_t/hp_t^2). \quad (1)$$

В то же время необходимая для работы суперкомпьютера электрическая мощность E_t (в кВт) оказывается связанной с отношением PR_t/hp_t :

$$\ln E_t = a_2 + b_2 \times \ln(PR_t/hp_t). \quad (2)$$

В таблице 4 приведены результаты оценивания параметров соотношений (1)–(2) методом наименьших квадратов, проведенного на основе данных из Top 500 и Green 500 за 2005–2012 гг. В выборку были включены наиболее производительные суперкомпьютеры с процессорами на основе схем разного топологического размера (по 3 суперкомпьютера с процессорами на основе схем размером от 130 нм до 32 нм; эти размеры приведены в табл. 1). В скобках представлены значения среднеквадратических ошибок оценок параметров.

Расчеты с использованием полученных зависимостей показывают, что при сохранении существующей тенденции увеличения производительности за счет использования большего числа ядер процессоров и уменьшения топологического размера схем

до 10 нм, 7 нм и 5 нм, мощность суперкомпьютера экзафлопной производительности может составить не 400–500 МВт, а соответственно 40 МВт, 33 МВт и 29 МВт. Следует, однако, отметить, что в рамках статистического анализа не удастся в полной мере отразить такие способы снижения энергоемкости будущего суперкомпьютера, которые связаны, в частности, с начавшимся недавно широким использованием сопроцессоров-ускорителей, а также с повышением производительности сетей (межсоединений).

Так, в настоящее время одним из эффективных направлений увеличения производительности и снижения энергоемкости (повышения энергоэффективности) суперкомпьютеров является использование сопроцессоров (ускорителей) на основе высокопроизводительных графических процессоров — GPU (главным образом, GPU NVIDIA), а также сопроцессоров Intel Xeon Phi и ряда других производителей. Наглядным примером целесообразности такого направления служит суперкомпьютер Titan, являющийся наиболее производительным в списке Top 500, опубликованном в ноябре 2012 г.² В этом суперкомпьютере, созданном путем апгрейда суперкомпьютера Jaguar, половину процессоров составляют GPU NVIDIA K20х. Подсчитано, что если бы апгрейд суперкомпьютера Jaguar осуществлялся на основе обычных микропроцессоров (CPU), то электрическая мощность суперкомпьютера Titan составила бы 8,2 МВт, как сейчас, а 30 МВт [49]!

Если учесть, что в суперкомпьютере Titan удалось снизить электроемкость одного процессорного ядра до 14,6 Вт, то, используя метод аналогий, нетрудно приблизительно оценить необходимую для экзафлопного суперкомпьютера электрическую мощность. Для процессоров на основе схем с топологическим размером 10 нм, 7 нм и 5 нм она составит соответственно 46 МВт, 22 МВт и 11 МВт. Таким образом, полученные оценки свидетельствуют о реальности создания суперкомпьютера производительностью 1 экзафлоп и электрической мощностью 20 МВт при переходе на производство схем с малыми топологическими размерами. Сроки такого перехода, сопряженного с высокими затратами, зависят, в частности, от прогресса в создании экономически эффективных установок экстремальной ультрафиолетовой фотолитографии (EUV).

Таким образом, в целом перспективы развития суперкомпьютеров связаны с успешным решением проблемы снижения стоимости. Роль этого показателя особенно высока ввиду дороговизны высокопроизводительных суперкомпьютеров, так и отно-

² См. также данные о суперкомпьютере Euroga в подразделе «Особенности развития ЦОД за рубежом».

Таблица 4

Оценки параметров a_i, b_i зависимостей (1)–(2), $i=1,2$

Зависимость	a_i	b_i	Коэффициент детерминации R^2
(1)	10,940	0,852	0,934
	(0,280)	(0,065)	
(2)	4,476	0,443	0,766
	(0,580)	(0,071)	

сительно малой продолжительности их жизненного цикла по сравнению с другими высокотехнологичными объектами сопоставимой стоимости.

Наглядным примером может служить судьба суперкомпьютера Roadrunner в Лос-Аламосской национальной лаборатории, который первым превзошел петафлопный рубеж производительности в 2008 г. Эксплуатация этого суперкомпьютера стоимостью 121 млн долл. была прекращена в конце марта текущего года, т. е. менее чем через 5 лет после ввода в строй. Вместо него в лаборатории будет использоваться более производительный суперкомпьютер Cielo стоимостью 53 млн долл. [50].

Стоимость введенных в эксплуатацию несколькими годами позже, а также разрабатываемых высокопроизводительных машин на порядок выше, чем у Roadrunner. Так, по некоторым данным стоимость первого в списке Top 500 в ноябре 2011 г. японского суперкомпьютера K Computer превысила миллиард долларов [47]. Желаемое ограничение по стоимости для суперкомпьютера нового поколения экзафлопной производительности, который намечено создать к 2020 г. в Японии, также составляет порядка 1 млрд долл. (11 млрд иен) [51].

В связи с отмеченными выше тенденциями актуальным является исследование новых энергоэффективных технологий для работы ЦОД и отдельных электронных устройств и систем.

Одно из кардинальных решений этой проблемы связано с анализом возможностей создания неэнергоёмких квантовых компьютеров и проведения квантовых вычислений. Другое принципиально возможное направление по-видимому состоит в переходе к углеродной наноэлектронике. Однако работы в этой области пока не вышли из стадии поисковых исследований.

В среднесрочной перспективе наиболее надёжным путем решения проблемы повышения энергоэффективности в полупроводниковой электронике представляется эволюционное развитие существующих технологий, в частности, по пути интеграции кремниевой электроники и фотоники [52].

3. Квантовые компьютеры и повышение эффективности ИТ-оборудования

В отличие от обычного классического компьютера, оперирующего с битами, квантовый компьютер оперирует с состояниями квантовых систем. Простейшей квантовой системой, которая может выполнять функцию, аналогичную битам в классических компьютерах, является система с двумя возможными состояниями — кубит (qubit) — квантовый бит информации. Любое нормированное состояние однокубитной квантовой системы ψ является суперпозицией двух базисных состояний $|0\rangle$ и $|1\rangle$:

$$|\psi\rangle = a|0\rangle + b|1\rangle, \quad \text{где } |a|^2 + |b|^2 = 1.$$

Измерение состояния кубита приводит к одному из базисных состояний $|0\rangle$ и $|1\rangle$ с вероятностями соответственно $|a|^2$ и $|b|^2$. Суперпозиция состояний квантовой системы, состоящей из n кубитов, также является состоянием квантовой системы. В связи с возможностью суперпозиции состояний квантовый компьютер с n кубитами позволяет проводить вычисления в гильбертовом пространстве размерности 2^n , что резко повышает его вычислительные мощности (при $n = 1000$ размерность пространства составляет $2^{1000} \approx 10^{300}$) [53].

Другим важным свойством является запутанность (entanglement) состояний квантовых систем. Оно характеризует коррелированность состояний кубитов и позволяет говорить о перспективных направлениях развития (квантовые вычисления, надёжная связь и телепортация, датчики электромагнитного излучения высокого разрешения и др.) [54, 55].

Логические операции над квантовыми состояниями в квантовых компьютерах выполняются путем унитарных преобразований, не нарушающих квантовые суперпозиции в процессе вычислений. Унитарные преобразования (квантовые гейты) по сути представляют собой логические блоки (вентили), составляющие процесс вычислений. Для реализации любой унитарной операции на n кубитах достаточно иметь универсальный конечный набор однокубитовых операций (например, операции НЕ, преобра-

зования Адамара и др.) и двухкубитовой операции CNOT (условного НЕ) [50, 56].

Схематично работу квантового компьютера можно представить в виде последовательности трех операций: приведение системы в начальное состояние; проведение вычислений путем унитарных преобразований начальных условий; измерение конечного состояния [57].

Особенность квантового компьютера состоит в том, что правильное решение в нем может быть определено лишь с некоторой вероятностью. Специалистами отмечается, что разработанные к настоящему времени алгоритмы для квантовых вычислений приводят, как правило, к правильным результатам с достаточно большой вероятностью за небольшое число итераций [53].

Повышенный интерес к разработке квантового компьютера, наблюдаемый в последние годы, обусловлен тем, что с его помощью может быть решен ряд задач, для которых не существует эффективных алгоритмов решения (например, имеющая важное значение для криптографии задача о факторизации чисел, проблема коммивояжера и связанные с ней задачи исследования операций и др.). Интенсификации исследований в области квантовых компьютеров способствовало и построение в 1990-х гг. эффективных алгоритмов квантовых вычислений. Важным стимулом к повышению интереса к квантовым компьютерам и квантовым вычислениям явилась разработка Гровером (Grover) в 1995 г. алгоритма поиска в неструктурированной базе. Квантовый алгоритм Гровера позволяет существенно снизить число итераций, необходимых для поиска N случайно заданных объектов (с $N/2$ до $N^{1/2}$).

На основе этого можно надеяться, что важной сферой применения квантовых компьютеров станут вычисления для ЦОД, например, для поисково-информационных систем. Реализация квантовых алгоритмов на квантовых компьютерах позволила бы существенно снизить парк серверов, а, следовательно, и капитальные и текущие затраты (включая затраты на энергию) в центрах обработки данных [58].

Другой перспективной областью может быть применение квантовых компьютеров для моделирования квантовых систем. Следует учитывать, что во многих практически важных случаях исследование квантовых систем не удастся провести даже на современных суперкомпьютерах [55].

Создание квантовых компьютеров сдерживается, по мнению специалистов, тем, что измерение кубита сопряжено с решением сложных технологических проблем, требующих преодоления трудностей измерения состояния отдельной атомной частицы:

атома, иона, электрона, спина электрона или атомного ядра, фотона.

Имеется также целый ряд других проблем. Так, по терминологии академика К. Валиева, квантовый компьютер представляет собой цифровой компьютер с аналоговым управлением. В нем управление эволюцией кубитов осуществляется с помощью обычного компьютера и генераторов импульсов. При этом параметры сигналов, управляющих кубитами, должны контролироваться с погрешностью 10^{-5} – 10^{-4} , которая на 2–3 порядка меньше погрешности, характерной для сигналов, используемых в традиционной аналоговой технике.

Необходимо также преодоление эффектов декогерентизации, которые возникают при взаимодействии квантового компьютера с окружающей средой, при неточности задания параметров управляющих сигналов или вследствие неконтролируемого взаимодействия между собой кубитов и т. д. (декогерентизация препятствует проявлению суперпозиции и запутывания состояний, то есть использованию именно тех свойств квантового компьютера, которые обуславливают его преимущество по сравнению с традиционным). Для решения этой проблемы в настоящее время разрабатываются различные варианты метода *квантовой исправления ошибок* (*quantum error correction*) [53].

В последнее время появились сообщения об определенном прогрессе в создании кубитов в металлах, находящихся в сверхпроводящем состоянии (в таких кубитах в качестве металла обычно используется ниобий или алюминий), в частности, на основе технологии RSFQ (rapid single flux quantum). Перспективы этой технологии связаны с тем, что переключение логических схем в сверхпроводящем состоянии осуществляется с существенно меньшим выделением тепла, чем в кремниевых чипах. Кроме того, несмотря на необходимость использования криогенной техники, технологии охлаждения материалов до температуры жидкого гелия уже хорошо освоены [58]. В университетских лабораториях созданы прообразы простейших квантовых процессоров на основе сверхпроводящих кубитов, для которых получены запутанные состояния, а также более высокое, чем ранее, время поддержания когерентности (см., например, [59]– [63]).

Значительный интерес проявляется к адиабатическим квантовым компьютерам (AQC) на основе технологии RSFQ, разработку которых осуществляет канадский стартап D-Wave Systems. Особенностью адиабатических квантовых компьютеров является то, что в них не используются квантовые гейты. Адиабатические квантовые вычисления основаны на использовании адиабатической теоремы кванто-

вой механики, вытекающей из уравнения Шредингера. Решение оптимизационных задач осуществляется в следующей последовательности. Вначале формируется гамильтониан системы (с взаимодействующими кубитами), минимальное энергетическое состояние которой (ground state) может быть решением проблемы (гамильтониан конечной системы). На следующем шаге синтезируется система с простым гамильтонианом, для которого может быть достаточно просто определено минимальное энергетическое состояние (например, система с невзаимодействующими кубитами). В это начальное состояние переводится система. Далее, простой гамильтониан медленно (адиабатически в квантово-механическом смысле) преобразуется в гамильтониан конечной системы. При этом, в соответствии с адиабатической теоремой, начальное минимальное энергетическое состояние простой системы переводится в минимальное энергетическое состояние для конечной системы. Полученное состояние и является решением задачи [58].

В мае 2011 г. состоялась первая продажа квантового компьютера D-Wave One, созданного D-Wave Systems, аэрокосмической компании Lockheed Martin за 10 млн долл. Компьютер из 128 кубитов, находящихся в сверхпроводящем состоянии, предназначался для решения специализированных задач, связанных с созданием сложных радиолокационных и аэрокосмических систем. В частности, предполагалось использовать D-Wave One для проведения исследований по автоматическому обнаружению ошибок в программных продуктах для управления сложными системами (в рамках проекта создания истребителя F-35). После двухлетней эксплуатации специалисты компании Lockheed Martin стали проявлять больший оптимизм в отношении возможностей коммерческого использования приобретенного компьютера [64, 65].

В мае текущего года появилось сообщение о том, что NASA, Google и консорциум университетов (Universities Space Research Association — USRA) организовали Лабораторию квантового искусственного интеллекта (Quantum Artificial Intelligence Lab), которая будет проводить исследование возможностей адиабатического квантового компьютера D-Wave Two. Квантовый компьютер D-Wave Two создан путем апгрейда первого коммерческого квантового компьютера D-Wave One со 128 до 512 кубитов.

Этот компьютер будет установлен в подразделении Advanced Supercomputing Facility исследовательского центра Эймса, который входит в NASA. Предполагается, что квантовый компьютер будет использоваться широким кругом исследователей из правительственных организаций промышленности и уни-

верситетов (на последнюю группу будет бесплатно отводиться до 20 % вычислительного времени). Принятию решения об учреждении Лаборатории предшествовало тестирование D-Wave Two независимыми экспертами и всеми сторонами-учредителями. Оно показало, что с помощью адиабатического квантового компьютера ряд проблем может быть выполнен на 3 порядка быстрее, чем с помощью обычного.

Учредители Лаборатории квантового искусственного интеллекта намерены использовать D-Wave Two в исследованиях по машинному обучению, которые, по их оценкам, эффективнее проводить не на обычных, а на квантовых компьютерах. В Google надеются, что квантовый компьютер поможет усовершенствовать систему поиска в Интернете и технологию распознавания речи. В то же время в NASA предполагают использовать D-Wave для моделирования широкого круга явлений (от атмосферных явлений до столкновений галактик) [66].

Интересно отметить, что инвесторами стартапа D-Wave являются Джефф Безос (Jeff Bezos, основатель Amazon.com), инвестиционный банк Goldman Sachs и инвестиционная фирма In-Q-Tel, имеющая тесные связи со спецслужбами США. Подобный интерес к квантовым компьютерам со стороны бизнес-сообщества во многом обусловлен опасением относительно того, что в ближайшие 10 лет, по мере приближения к пределам кремниевой электроники, завершится эра экспоненциального роста производительности классических компьютеров [67, 68].

Однако в целом работы по созданию универсальных квантовых компьютеров находятся в стадии фундаментальных исследований и еще далеки от перехода в стадию внедрения в производство. Несмотря на проявляемый в последние месяцы со стороны ряда исследователей оптимизм, разработка необходимых для решения широкого круга задач квантовых компьютеров потребует достаточно длительное время, причем для получения положительного результата необходима консолидация усилий в нанoeлектронике, других областях физики, а также в математике [53, 55, 69].

4. Эволюционные направления инновационного развития элементной базы информационных технологий

Ввиду нерешенности многих научно-технических проблем, связанных с созданием квантовых компьютеров и углеродной электроники, перспективы дальнейшего развития элементной базы ИТ-оборудования связываются с эволюционным путем совершенствования технологий на основе кремния [70].

Так, в 2007 г. в компании IBM была создана технология 3-мерной интеграции чипов (3-D Chip Stacking Technique). Трехмерная стековая многослойная структура (например, состоящая из чипов процессора и памяти, наложенных один на другой), выполненная по технологии TSV (through silicon vias — сквозные отверстия в кремнии), позволяет провести 100 000 вертикальных отверстий на см². Это дает возможность существенно (на 2–3 порядка) уменьшить протяженность металлических соединений по сравнению с традиционной планарной технологией и значительно (на порядки) увеличить объемы и скорость передачи информации. 3-мерные структуры обеспечивают снижение потребления энергии в устройствах беспроводной связи (на 40 %) и в многоядерных процессорах (на 20 %) [71]. В настоящее время данная технология является одним из наиболее привлекательных направлений повышения быстродействия вычислительных устройств, которое, однако, имеет свои границы повышения эффективности [72, 73].

Перспективы дальнейшего развития технологии производства 3-мерных структур связываются с созданием электронно-фотонных интегральных схем на основе кремния. Некоторые специалисты считают, что нанофотоника может предложить прорывные технологии, которые не только повысят показатели работы систем связи и компьютеров, но и смогут коренным образом изменить архитектуру компьютеров [74].

В настоящее время компании IBM и Intel ведут работы по созданию кремниевых фотонных систем, совместимых с широко используемыми КМОП-структурами (см. п. 1), что позволит существенно повысить быстродействие, а также уменьшить стоимость и расход энергии в электронных схемах по сравнению с уже создаваемыми 3-мерными структурами.

Так, в одном из проектов, разрабатывавшихся компанией IBM, предлагалось разместить в нижней части трехмерной структуры процессоры и память, а в верхней части, — так называемый «фотонный» слой, состоящий из тысяч оптоэлектронных приборов (модуляторов, детекторов, переключателей) и аналоговых электрических цепей. Этот слой должен выполнять не только функции высокоскоростного соединения между компонентами на чипе, но и маршрутизатора [75]. К настоящему времени уже получены первые результаты. В начале 2010 г. IBM объявила о создании лавинного фотодетектора на основе кремния и германия (avalanche photodetector), который способен принимать оптические сигналы со скоростью 40 Гб/с. Этот детектор работает при напряжении 1,5 вольт, что в 15–20 раз ниже, чем в существующих приборах. Кроме того, в нем

обеспечивается на 50–70 % более высокий уровень подавления шумов, чем в традиционных детекторах [76].

В конце 2010 г. было объявлено о создании технологии интеграции электрических и оптических приборов на одном кремниевом чипе (CMOS Integrated Silicon Nanophotonics), позволяющей на порядок увеличить плотность интеграции по сравнению с существующими методами. Существенным является то, что новые чипы могут производиться на передовых фабриках, выпускающих КМОП-структуры [77].

В декабре 2012 г. компания IBM сообщила об успешной интеграции оптических компонентов (модуляторов, германиевых фотодетекторов, волноводов и др.) в кремниевые КМОП-схемы с топологическим размером 90 нм. Выбор интегральных схем с таким топологическим размером обусловлен, по мнению разработчиков, тем, что их характеристики соответствуют перспективным потребностям технологии, а также экономическими соображениями (низкой стоимостью производства подобных КМОП-схем). Разработанная технология позволяет обеспечить передачу оптических сигналов по каждому из четырех используемых каналов со скоростью 25 Гб/с [78].

Важные результаты в области кремниевой нанофотоники получены и в компании Intel. Там разработаны основные составные блоки систем передачи оптической информации на основе кремния (в частности, гибридный кремниевый лазер, кремниевые детектор и модулятор оптических сигналов, превосходящие по характеристикам аналогичные приборы, изготовленные на основе других элементов) [79].

Летом 2010 г. Intel объявила о создании прототипа оптического канала передачи данных на базе полупроводниковых компонентов с использованием гибридного кремниевого лазера со скоростью передачи данных по оптическому волокну, достигающей 50 Гб/с, что являлось мировым рекордом. Линия связи включает кремниевый передатчик с четырьмя лазерами и чип-приемник. Каждый из четырех лазерных лучей поступает в оптический модулятор для последующего кодирования. Приемник разделяет лучи и направляет их в фотодетекторы, которые преобразуют данные в электрические сигналы. Весьма существенно, что передающий и приемный чипы могут производиться на основе отработанных в полупроводниковой промышленности процессов [80].

Важным этапом многолетних исследований Intel в области нанофотоники явилась демонстрация в апреле текущего 2013 г. на конференции Intel в Пекине первого кремниевого фотонного чипа, представляющего собой интегрированный модуль, кото-

рый состоит из гибридного лазера, модуляторов, детекторов, волноводов и электрических цепей. Разработанный чип обеспечивает скорость передачи данных 100 ГБ/с, но в будущем компания намерена достичь рубежа в 1 терабит в секунду (1ТБ/с). Кроме этого чипа, Intel совместно с компанией Corning разработал новый оптический (фотонный) кабель и соединение, пропускной способностью 1,6 ТБ/с.

Так как производство кремниевого фотонного чипа основывается на использовании хорошо развитой кремниевой технологии, его стоимость должна быть существенно ниже, чем чипа на основе обычно используемого арсенида галлия (GaAs). По некоторым данным в перспективе Intel предполагает продавать кремниевый фотонный чип по цене 5 долл. [81].

Достижения Intel и IBM позволяют надеяться на замену уже в ближайшем будущем медных кабелей и соединений тонкими и легкими оптическими волокнами, по которым очень большие объемы данных могут быть переданы с существенно большей скоростью, без существенных искажений и без выделения тепла. Актуальность отмеченных разработок этих компаний обусловлена, в частности, переходом к производству многоядерных процессоров с десятками ядер, необходимостью обеспечения связи между отдельными серверами и серверными стойками в ЦОД и суперкомпьютерах, а также целесообразностью обеспечения раздельного проведения модернизации отдельных блоков серверов (например, процессора и памяти) [82].

В связи с успехами, достигнутыми в последнее время в 3-D интеграции и в кремниевой нанофотонике, у крупных производителей микроэлектронной продукции укрепилась уверенность в том, что темпы уплотнения чипов удастся сохранить и после 2020 г.

5. Заключение

Информационные технологии превратились в важную отрасль, которая не только стимулирует инновационное развитие, но и оказывает значительное

влияние на энергоемкость и капиталоемкость экономики, а также на экологию территорий. В перспективе следует ожидать усиления такого влияния.

Дальнейшее развитие традиционной кремниевой полупроводниковой технологии требует разработки комплекса принципиально новых подходов, направленных на повышение энергоэффективности электронных приборов и систем.

Для повышения энергоэффективности ЦОД необходимо проведение широкого круга комплексных исследований в области создания энергосберегающего ИТ-оборудования и инфраструктуры, в частности экономичных систем охлаждения. В настоящее время наметился консенсус относительно того, что наилучшие характеристики для ЦОД, как и для суперкомпьютеров, имеет технология водяного охлаждения.

Существенная экономия электроэнергии и затрат в ЦОД может быть обеспечена путем их виртуализации и перехода к облачным вычислениям. Вместе с тем широкомасштабный переход потребителей компьютерных услуг на облачные сервисы будет сдерживаться нерешенностью проблем компьютерной безопасности.

К настоящему времени достигнут определенный прогресс в области создания адиабатических квантовых компьютеров, ориентированных на решение задач эффективного поиска и искусственного интеллекта. Однако в целом работы по созданию универсальных квантовых компьютеров находятся в стадии фундаментальных исследований и еще далеки от перехода в стадию внедрения в производство.

Один из наиболее реальных путей развития элементной базы и аппаратных средств для информационных технологий на среднесрочную перспективу состоит в интеграции кремниевой электроники и фотоники.

В связи со значительными энергоемкостью и капиталоемкостью развивающейся быстрыми темпами индустрии ЦОД, необходима организация систематической государственной статистической отчетности о технико-экономических показателях деятельности этих центров.

Литература

1. [Электронный ресурс], <http://www.gartner.com/newsroom/id/2394415>
2. [Электронный ресурс], http://www.tadviser.ru/index.php/%D0%A1%D1%82%D0%B0%D1%82%D1%8C%D1%8F%3A%D0%98%D0%A2-%D1%80%D1%8B%D0%BD%D0%BE%D0%BA_%D0%A0%D0%BE%D1%81%D1%81%D0%B8%D0%B8
3. [Электронный ресурс], <http://www.cbronline.com/news/data-centre-investment-reach-105bn-in-2012-report-081012>
4. [Электронный ресурс], Gordon Moore. Progress in Digital Integrated Electronics. http://download.intel.com/museum/Moores_Law/Articles-Press_Releases/Gordon_Moore_1975_Speech.pdf
5. [Электронный ресурс], <http://www.cnews.ru/news/top/index.shtml?2012/05/15/489441>
6. [Электронный ресурс], <http://www.eetimes.com/news/latest/showArticle.jhtml?articleID=51000381&printable=true>
7. *Lammers D.* Intel Ramping 32 nm Manufacturing in Oregon // Semiconductor International, 9.14.2009.
8. *Рудометов Е.* Материнские платы и чипсеты. 4-е изд. Анатомия ПК. СПб.: Питер, 2007. 368 с.
9. [Электронный ресурс], <http://www.cnews.ru/news/top/index.shtml?2008/07/08/307658>
10. [Электронный ресурс], <http://www.intel.com/technology/silicon/high-k.htm?iid=SEARCH>
11. *Шука А. А.* Нанoeлектроника. М.: Физматкнига, 2007.
12. [Электронный ресурс], http://newsroom.intel.com/community/intel_newsroom/blog/2011/05/04/intel-reinvents-transistors-using-new-3-d-structure
13. [Электронный ресурс], <http://nanotechwire.com/news.asp?nid=7404>
14. [Электронный ресурс], <http://www.nanowerk.com/news/newsid=20205.php>
15. [Электронный ресурс], <http://www.datacenterknowledge.com/archives/2011/04/21/google-buys-wind-power-to-green-oklahoma-grid/>
16. *Koomey J.* Growth in Data Center Electricity Use 2005 to 2010, August 2011, [Электронный ресурс], <http://www.analyticspress.com/datacenters.html>
17. *Brill K. G.* The Invisible Crisis in the Data Center: The Economic Meltdown of Moore's Law. Uptime Institute.
18. *Minas L.* The Problem of Power Consumption in Servers. 2009. [Электронный ресурс], <http://www.intel.com>
19. A Simple Model for Determining True Total Cost of Ownership for Data Centers. Version 2.1, March 31, 2008. Uptime Institute.
20. [Электронный ресурс], <http://www.cbronline.com/news/data-centre-investment-reach-105bn-in-2012-report-081012>
21. *Дружинин Е.* Энергоэффективное охлаждение суперкомпьютеров. Краткий обзор технологий охлаждения ЦОД // Суперкомпьютеры. 2012. № 11. С. 30–33.
22. [Электронный ресурс], <http://efytimes.com/e1/99599/Eurotech-Supercomputer-At-CINECA-Breaks-World-Record-For-Energy-Efficiency>
23. [Электронный ресурс], <http://blog.uptimeinstitute.com/category/data-center-energy-efficiency/page/2/>
24. [Электронный ресурс], <http://www.datacenterknowledge.com/archives/2011/05/10/uptime-institute-the-average-pue-is-1-8/>
25. Intel Data Centers Solutions: Strategies to Improve Efficiency, 2009. [Электронный ресурс], <http://www.intel.com>
26. *Miller C.* Pew: Cloud Services Will Dominate By 2020. June 15th, 2010. [Электронный ресурс], <http://www.datacenterknowledge.com>
27. [Электронный ресурс], http://static.googleusercontent.com/external_content/untrusted_dlcp/www.google.com/en/us/green/pdf/google-apps.pdf
28. [Электронный ресурс], <http://www.cnews.ru/news/top/index.shtml?2011/10/10/459220>
29. [Электронный ресурс], <http://www.datacenterdynamics.com/ru/focus/archive/2012/07/most-us-federal-agencies-track-data-center-consolidation>
30. [Электронный ресурс], <http://www.osp.ru/lan/2008/09/5535759/>
31. [Электронный ресурс], <http://www.cnews.ru/reviews/free/datacenter/articles/articles3.shtml>
32. [Электронный ресурс], <http://www.cnews.ru/reviews/free/datacenter/articles/articles1.shtml>
33. [Электронный ресурс], <http://telecomblogger.ru/14418>
34. [Электронный ресурс], <http://spbit.su/news/n135588/>
35. [Электронный ресурс], <http://www.cnews.ru/reviews/free/datacenter/articles/articles5.shtml>
36. [Электронный ресурс], <http://www.cnews.ru/news/top/index.shtml?2011/11/12/464294>
37. [Электронный ресурс], <http://cloud.cnews.ru/news/line/index.shtml?2013/02/20/519735>
38. [Электронный ресурс], <http://www.cnews.ru/reviews/free/datacenter/int/hp/>
39. [Электронный ресурс], <http://www.cnews.ru/reviews/free/datacenter/articles/articles6.shtml>
40. [Электронный ресурс], <http://www.irn.ru/articles/31390.html>

41. [Электронный ресурс], [http://www.tadviser.ru/index.php/%d0%a1%d1%82%d0%b0%d1%82%d1%8c%d1%8f:%d0%a1%d1%83%d0%bf%d0%b5%d1%80%d0%ba%d0%be% d0%bc%d0%bf%d1%8c%d1%8e%d1%82%d0% b5%d1%80%d1%8b_\(%d0%bc%d0%b8%d1% 80%d0%be%d0%b2%d0%be%d0%b9_%d1% 80%d1%8b%d0%bd%d0%be%d0%ba](http://www.tadviser.ru/index.php/%d0%a1%d1%82%d0%b0%d1%82%d1%8c%d1%8f:%d0%a1%d1%83%d0%bf%d0%b5%d1%80%d0%ba%d0%be% d0%bc%d0%bf%d1%8c%d1%8e%d1%82%d0% b5%d1%80%d1%8b_(%d0%bc%d0%b8%d1% 80%d0%be%d0%b2%d0%be%d0%b9_%d1% 80%d1%8b%d0%bd%d0%be%d0%ba)
42. [Электронный ресурс], <http://www.top500.org>
43. *Turek D.* The Age of the Supercomputer. [Электронный ресурс], <http://www.businessweek.com/stories/2007-05-07/the-age-of- the-supercomputerbusinessweek-business-news- stock-market-and-financial-advice>
44. [Электронный ресурс], <http://www.green500.org/lists.php>
45. [Электронный ресурс], <http://www.cnews.ru/news/top/index.shtml?2012/06/27/494386>
46. *Левшин И.* Из жизни алгоритмов. HPC Superstar // Суперкомпьютеры. 2013. № 13. С. 10–12.
47. *Левшин И.* Из жизни звезд // Суперкомпьютеры. 2013. № 13. С. 6–9.
48. *Feldman M.* DARPA Sets Ubiquitous HPC Program in Motion. [Электронный ресурс], http://www.hpcwire.com/hpcwire/2010-08-10/darpa_sets_ubiquitous_hpc_program_in_motion.html
49. [Электронный ресурс], <http://www.economist.com/news/science-and-technology/21566611-world-has-new-fastest-computer-thanks-video-games-deeper-thought>
50. [Электронный ресурс], <http://www.bbc.co.uk/news/technology-21993132>
51. [Электронный ресурс], <http://efytimes.com/e1/fullnewsp.asp?edid=105870>
52. Экономические проблемы развития революционных технологий. Нанотехнологии / Рук. авт. колл. В. Л. Макаров, А. Е. Варшавский. М.: Наука, 2012.
53. *Валиев К. А.* Квантовые компьютеры и квантовые вычисления // Успехи физических наук. 2005. Т. 175. № 1. С. 1–39.
54. [Электронный ресурс], <http://www.carolla.com/quantum/QuantumComputers.htm>
55. Quantum computers. [Электронный ресурс], http://ewh.ieee.org/r10/bombay/news4/Quantum_Computers.htm
56. *Нильсен М., Чанг И.* Квантовые вычисления и квантовая информация. М.: Мир, 2006.
57. *Кулин С. Я.* Квантовая информация // Успехи физических наук. 1999. Т. 169. № 5. С. 508–526.
58. *Wolf E. L.* Quantum Nanoelectronics: An Introduction to Electronic Nanotechnology and Quantum Computing, Wiley, 2009.
59. Scientists create first electronic quantum processor. [Электронный ресурс], <http://www.physorg.com/news165418586.html>
60. Three Solid-State Qubits Entangled: Big Step Toward Quantum Error Correction. [Электронный ресурс], <http://www.sciencedaily.com/releases/2010/09/100929132533.htm>
61. Quantum Computing Research Edges Toward Practicality. [Электронный ресурс], <http://www.sciencedaily.com/releases/2010/10/101005104446.htm>
62. [Электронный ресурс], <http://www.nanowerk.com/news2/newsid=26603.php>
63. [Электронный ресурс], <http://www.nanowerk.com/news/newsid=24827.php>
64. *Rachel Courtland.* A First for Quantum Computing? [Электронный ресурс], <http://spectrum.ieee.org/tech-talk/computing/hardware/a-first-for-quantum-computing>
65. *Hardy Q.* A Strange Computer Promises Great Speed. [Электронный ресурс], http://www.nytimes.com/2013/03/22/technology/testing-a-new-class-of-speedy-computer.html?_r=0&pagewanted=print
66. *Choi Ch.* Google and NASA Launch Quantum Computing AI Lab. <http://www.technologyreview.com/news/514846/google-and-nasa-launch-quantum-computing-ai-lab/>, May 16, 2013.
67. [Электронный ресурс], http://www.nytimes.com/2013/03/22/technology/testing-a-new-class-of-speedy-computer.html?_r=0&pagewanted=print
68. *Simonite T.* The CIA and Jeff Bezos Bet on Quantum Computing. [Электронный ресурс], <http://www.technologyreview.com/news/429429/the-cia-and-jeff-bezos-bet-on-quantum-computing/>
69. Pitt professor lays groundwork for quantum computers. [Электронный ресурс], <http://www.post-gazette.com/pg/10284/1094223-115.stm>
70. The Coming Chip Revolution // Business Week. April 18. 2005.
71. IBM moves Moore's Law into the third-dimension. [Электронный ресурс], http://domino.watson.ibm.com/comm/pr.nsf/pages/news.20070412_3dchip.html
72. *Emma P. G., Kursun E.* Is 3D Chip Technology the Next Growth Engine for Performance Improvement? // IBM J. Res.&Dev., 2008. Vol. 52. № 6. P. 541–551.
73. *Васильев А.* Современные технологии 3D-интеграции // Степень интеграции. Информационный бюллетень. Октябрь, 2009. № 2. С. 6–9.
74. *Saif Islam M., Shih-Yuan Wang.* Nanophotonics for Communications // Journal of Nanophotonics. 2008. Vol. 2. P. 1–5.
75. [Электронный ресурс], http://domino.research.ibm.com/comm/research_projects.nsf/pages/photonics.projects.html
76. IBM Scientists Create Ultra-Fast Device Which Uses Light for Communication between Computer Chips. [Электронный ресурс], <http://www-03.ibm.com/press/us/en/pressrelease/29595.wss>

77. [Электронный ресурс], http://www-03.ibm.com/press/us/en/pressrelease/33115.wss#_release#release
78. [Электронный ресурс], http://researcher.ibm.com/researcher/view_project.php?id=2757
79. [Электронный ресурс], <http://techresearch.intel.com/articles/Tera-Scale/1419.htm>
80. Фотоника: Intel добился передачи данных на скорости 50 ГБ/с. [Электронный ресурс], <http://www.cnews.ru/news/top/index.shtml?2010/07/28/402924>
81. [Электронный ресурс], <http://www.theinquirer.net/inquirer/news/2260771/intel-shows-off-first-fully-functional-photonics-chip>; http://www.theregister.co.uk/2013/04/11/intel_silicon_photonics_breakthrough/
82. [Электронный ресурс], <http://slashdot.org/topic/datacenter/intel-silicon-photonics-moving-to-market/>

Варшавский Леонид Евгеньевич. Зав. лаб. ИСА РАН. Гл. н. с. ЦЭМИ РАН. Д. э. н. Окончил в 1970 г. МИФИ. Количество печатных работ — более 90. Область научных интересов: математическое моделирование рыночных процессов и прогнозирование показателей рынков высокотехнологичной и капиталоемкой продукции, математическое моделирование инвестиционных и инновационных стратегий промышленных фирм, математическое моделирование макроэкономических процессов, методы управления динамическими системами, экономика науки. E-mail: hodvar@mail.ru