

УДК 002.53:002.55

В.В. Вычужанин

**РАЗВИТИЕ ИНФРАСТРУКТУРНЫХ РЕШЕНИЙ  
ДЛЯ ТЕХНОЛОГИИ ОБЛАЧНЫХ ВЫЧИСЛЕНИЙ**

*Проанализированы имеющиеся на рынке высокопроизводительные сетевые решения. Рассмотрена реализация стандарта 100 Гбит Ethernet в сетевых устройствах на программируемых логических интегральных схемах и оптических модулях для технологии облачных вычислений.*

**Ключевые слова:** *облачные вычисления, 100 Гбит Ethernet, оптические модули, программируемые логические интегральные схемы.*

*Проаналізовано наявні на ринку високопродуктивні мережеві рішення. Розглянуто реалізація стандарту 100 Гбіт Ethernet в мережевих пристроях на програмованих логічних інтегральних схемах і оптичних модулях для технології хмарних обчислень.*

**Ключові слова:** *хмарні обчислення, 100 Гбіт Ethernet, оптичні модулі, програмовані логічні інтегральні схеми.*

*Analyzed on the market high-performance networking solutions. We consider the implementation of the standard 100 G Ethernet in the network devices in the programmable logic integrated circuits and optical modules for cloud computing.*

**Keywords:** *cloud computing, 100 Gb Ethernet, optical modules, programmable logic integrated circuits.*

**Введение.** Облачные вычисления (Cloud Computing), представляют собой комплекс методологий и технологий с организацией распределенной обработки данных, рассматривающих программно-аппаратное обеспечение и информационную инфраструктуру в качестве интернет-сервиса. В масштабах предприятия облачные вычисления позволяют отказаться от собственной программно-аппаратной инфраструктуры. Все необходимые данные пользователей хранятся на удаленной площадке провайдера услуг с доступом к ним при помощи интернет-браузера с персонального компьютера (ПК) или мобильного устройства, подключенного к сети Ethernet [1-4].

К технологиям облачных вычислений предъявляются требования:

- соответствие стандартам IEEE Std 802.3ba GbE и ИТ-услуг ISO 20000 [5];
- учет индивидуальных особенностей ИТ – услуг клиентов;
- наличие дата-центра, например Tier 3 – уровня;
- бесперебойность системы хранения данных при пиковых нагрузках;
- доступность сервисов 99,9 %;

– отсутствие влияния потребляемых мощностей друг на друга на технологическом уровне.

Сегодня все больше клиентов прибегают к облачным вычислениям, чтобы избежать затрат на приобретение новой инфраструктуры. Компании предлагают своим заказчикам не только облачные приложения или инфраструктуру, но и полноценные сервисы, ничем не отличающиеся от внутренних программных и аппаратных решений. Организуется система предоставления облачных услуг, обеспечивающая производительность и бесперебойность сервисов при демократичных ценах.

**Актуальность.** Растущий спрос на облачные услуги обусловлен их максимальным соответствием современным бизнес-задачам – заказчик платит только за фактически используемые мощности в конкретный момент.

Облачные вычисления, несмотря на преодолеваемые недостатки связанные с вопросом приватности данных, обладают рядом преимуществ. Это пониженные требования к вычислительной мощности ПК, приводящие к меньшим затратам на программно-аппаратное обеспечение, на обслуживание и потребляемую электроэнергию, а также к экономии дискового пространства. Повышение скорости обработки данных, так как не требуется модернизации клиентской части. Облачные вычисления также характеризуются большой надежностью, поскольку отказ одного из серверов, предоставляющего услуги, не ведет к краху системы.

Существуют следующие модели облаков как сервисов: виртуальный сервер (Infrastructure as a Service – IaaS); платформа (Platform as a Service – PaaS); программное обеспечение (Software as a Service – SaaS); виртуальное рабочее место (Desktop as a Service – DaaS); коммуникации (Communications as a Service – CaaS). Модели облачных вычислений SaaS, PaaS и IaaS (рис.1) предоставляют программные компоненты, остальные модели – аппаратное обеспечение, серверы, сети и пр.



Рис. 1. Модели облачных вычислений [5]

На рынке для профессиональных разработчиков программного обеспечения представлены, основанные на концепции облачных вычислений, среды Amazon AWS, Microsoft Azure, Google App Engine, Stax,

10gen, Google Wave, Force.com и др. Среды являются универсальными или ориентированными на какой-либо тип приложений. Универсальная среда Google App Engine поддерживает языки программирования Python и Java, а среды Engineyard.com и Heroku.com. Engineyard, основанные на Amazon EC2 и Amazon S3, поддерживают язык программирования Ruby. Microsoft Azure предоставляет ориентированную инфраструктуру, инструменты разработки и поддерживает PHP, MySQL, Ruby, Python, Java, Eclipse и Zend [6]. Варианты реализации вычислений в облаке представлены на рис. 2, а в табл. 1 – сравнительные характеристики вычислений.

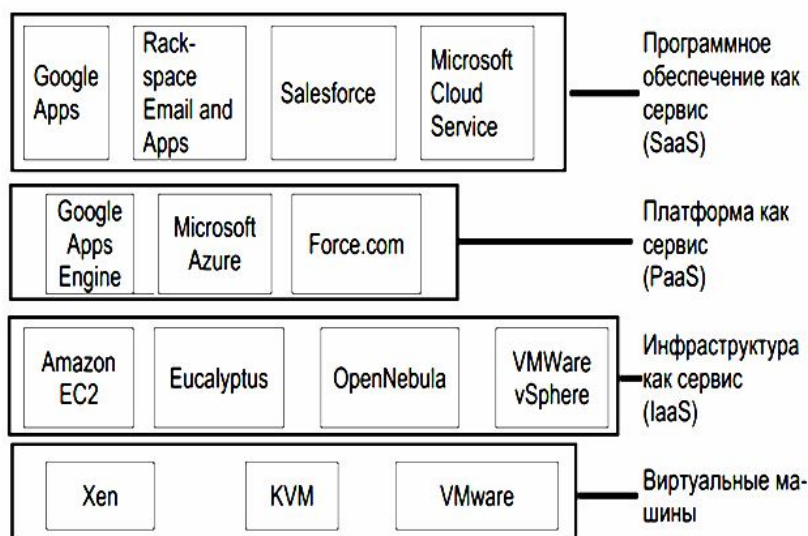


Рис. 2. Варианты реализации вычислений в облаке

Переход клиентов к потреблению облачных услуг требует ответа на следующие вопросы. Во-первых, будет ли стоимость перехода на облачные вычисления достаточно низкой, чтобы ожидаемый результат проявился уже в среднесрочной перспективе? Второе – являются ли облачные вычисления платформой для роста и развития предприятия? Третье – обеспечит ли переход на облачные вычисления уровень безопасности, достаточный для того, чтобы считать отличающие его риски приемлемыми на новом уровне развития предприятия? В этой связи реализация быстрого доступа к информации, хранящейся в облаках, разработка сетевых инфраструктур, позволяющих передавать широкополосные данные на большие расстояния со скоростью не менее 100 Гбит/с, обладающих высокой плотностью портов, низким энергопотреблением, а также сочетающих требуемую масштабируемость с невысокой стоимостью сетевых инфраструктур является актуальной научно-технической задачей.

Таблиця 1

*Характеристики облачных вычислений*

Характеристика	Amazon Web Services	Google App Engine	Microsoft Azure	Force.com
1	2	3	4	5
Вычислительная архитектура	Elastic Compute Cloud (EC2) поддерживает загрузку образов виртуальных машин и предоставляет клиентские API для создания экземпляров и управления ними	Распределенная архитектура Google	Размещается в центрах обработки данных Microsoft, предоставляя операционную систему и сервисы разработки	"Коммунальная" (Multitenant) архитектура
Управление виртуализацией	Работа на уровне операционной системы в гипервизоре Xen	Приложения работают на экземплярах виртуальной машины Java или среды выполнения Python	Гипервизор (на основе Hyper-V)	Поддерживается инфраструктурой Force.com. Встроенные регуляторы резервируют и распределяют ресурсы между пользователями
Сервис	IaaS	PaaS	PaaS	SaaS через среду разработки Force.com
Балансировка нагрузки	Гибкая балансировка нагрузки	Автоматическое масштабирование и балансировка нагрузки	Встроенная балансировка нагрузки	Встроенная балансировка нагрузки между участниками
Интероперабельность	Опубликованные API для взаимодействия с экземплярами; размещаемые приложения могут быть написаны на стандартных языках программирования, работающих в виртуальных машинах	Размещаемые приложения, написанные на Java и Python, доступны через предоставленные ими интерфейсы; доступ к интернет-ресурсам из кода и сервиса Fetch в App Engine	Размещаемые приложения работают в Windows и доступны через предоставляемые ими интерфейсы, приложения могут также использовать простой API для доступа к структуре Azure	Web Services API в Force.com предоставляет извне доступ к данным, хранящимся в Force.com; приложения Web-логики могут быть представлены как стандартный Web-сервис; для сервисов REST предлагается HTTP API

*Продолжение табл. 1*

1	2	3	4	5
Хранение	Simple Storage Service (S3), Amazon Elastic Block Storage (EBS), Amazon SimpleDB, Amazon Relation Database Service (RDS)	Хранилище данных App Engine (нереляционная, встроенная Bigtable); объекты с поддержкой транзакции	Постоянные данные, хранящиеся в нереляционных блоках, таблицах и очередях; ресурсы хранения SQL предоставленные SQL Azure	Постоянные данные хранятся в объектах, экземпляры объектов аналогичны таблице реляционной базы данных
Безопасность	SAS 70 Type II Certification, межсетевой экран, сертификат X.509. API для поддержки защиты SSL, список контроля доступа	SAS 70 Type II Certification, защищенный доступ к интранет через Secure Data Connector компании Google	SAS 70 Type II Certification, приложения работают в 64-разрядной MS-Windows Server 2008	SAS 70 Type II Certification, контроль доступа к данным на основе идентификационной информации пользователей и их ролей в организации
Платформа программирования	Amazon Machine Image (AMI), MapReduce	Java и Python, планируемые задачи и очереди, доступ к таким сервисам, как получение URL, почта, Memcache, работа с изображениями	.net и неуправляемый код, если он работает в Windows	Модель разработки на базе метаданных, платформа Visualforce для создания пользовательского интерфейса, язык программирования Apex

**ЦЕЛЮЮ РАБОТЫ** является развитие инфраструктурных решений для технологии облачных вычислений, обеспечивающих реализацию стандартов IEEE Std 802.3ba GbE и ИТ-услуг ISO 20000.

**ИЗЛОЖЕНИЕ ОСНОВНОГО МАТЕРИАЛА.** Расходы на переход в облако оправданы только тогда, когда инициатор перехода оказывается в состоянии разместить в облаке сервис, востребованный не только им одним. Условия, при которых переход в облако становится двигателем развития:

- модульность и интероперабельность сервисов, позволяющие объединять их понятным для любого внешнего потребителя образом;
- масштабируемость и законченность, дающие возможность обрабатывать большой поток входных запросов и при отсутствии необходимости производить настройку внутренних процессов сервиса;
- истинная облачность – сервис должен быть в целом перенесен в облако.

Стандарт IEEE Std 802.3ba GbE регламентує при функціонуванні мережних пристроїв використовувати паралельно- послідовальний спосіб передачі даних на швидкості 100 Гбіт/с., а також:

- підтримувати швидкість передачі даних 100 Гбіт/с на логічному MAC (Media Access Control)-рівні управління доступом до середовища передачі даних;
- підтримувати тільки повнодуплексні режими MAC – рівня;
- зберігати формат кадру Ethernet 802.3 MAC – рівня;
- забезпечувати для інтерфейсу між MAC – і фізичним рівнями підтримку величини коефіцієнта бітових помилок не гірше  $10^{-12}$ ;
- зберігати мінімальний і максимальний розміри кадрів стандарту IEEE Std 802.3ba GbE;
- забезпечувати сумісність з оптичними мережами OTN.

Фізичний рівень 100 Гбіт Ethernet утворений трьома основними PCS (Physical Coding Sublayer), PMA (Physical Medium Attachment), PMD (Physical Medium Dependent Sublayer) і двома опціональними підрівнями [7, 8] (рис.3). С MAC – рівня дані потрапляють на підрівень узгодження RS (Reconciliation Sublayer), де послідовальний потік даних перетворюється в паралельний 64-бітний (64В) потік і через інтерфейс CGMI (100 Gb Media Independent Interface) потрапляє на підрівень кодування PCS. На PCS функціонує схема кодування 64В/66В, в якій 66-бітне слово переадресується в паралельні індивідуальні потоки.

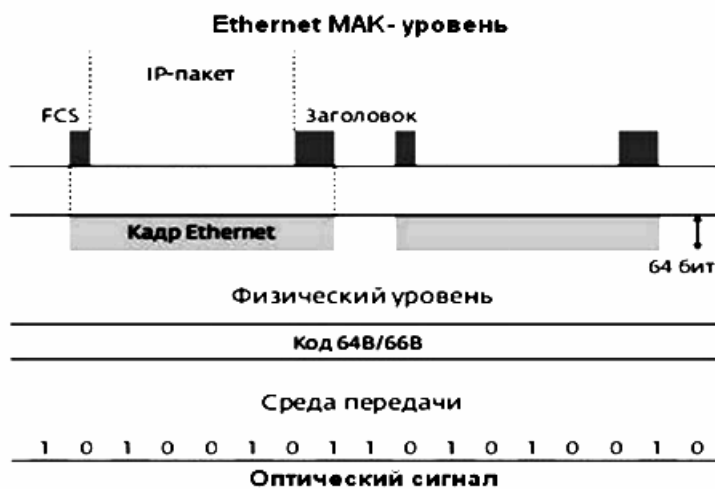


Рис. 3. Фізичний і MAC-рівні стандарту 100 GbE

Поскольку последовательный поток битов распределяется по параллельным индивидуальным потокам, то для восстановления исходного потока в приемнике между параллельными потоками должна сохраняться временная синхронизация. Учитывая скорости и расстояния передачи данных, предусмотренные стандартом 100 GbE, физические рассогласования потоков неизбежны. Для компенсации временных сдвигов используются выравнивающие маркеры, представляющие собой DC-сбалансированный поток, состоящий из байтов  $\{M_0, M_1, M_2, \text{VIP}_3, M_4, M_5, M_6, \text{VIP}_7\}$ , причем  $M_4, M_5, M_6$  являются побайтно обратные  $M_0, M_1$  и  $M_2$  байтам. Каждая полоса для подуровня PCS передачи данных (0, 1 ... 19) имеет единственное байтное  $M_0, M_1, M_2$  кодирование, позволяющее получить и расшифровать номер соответствующей полосы подуровня PCS (рис. 5). Измеряя временной сдвиг поступления кодовых блоков, приемник восстанавливает синхронность параллельных индивидуальных потоков. С прекращением получения данных подуровнем PCS маркеры выравнивания удаляются.

Наличие периодического маркера выравнивания позволяет приемнику успешно функционировать при значительных перекосах сигналов между параллельными каналами передачи данных. После кодирования и синхронизации на подуровне PCS данные поступают на физический подуровень PMA, выполняющий функции тестирования передачи данных: генерацию тестовых последовательностей; формирование петли обратной связи данных для тестирования и т.п.

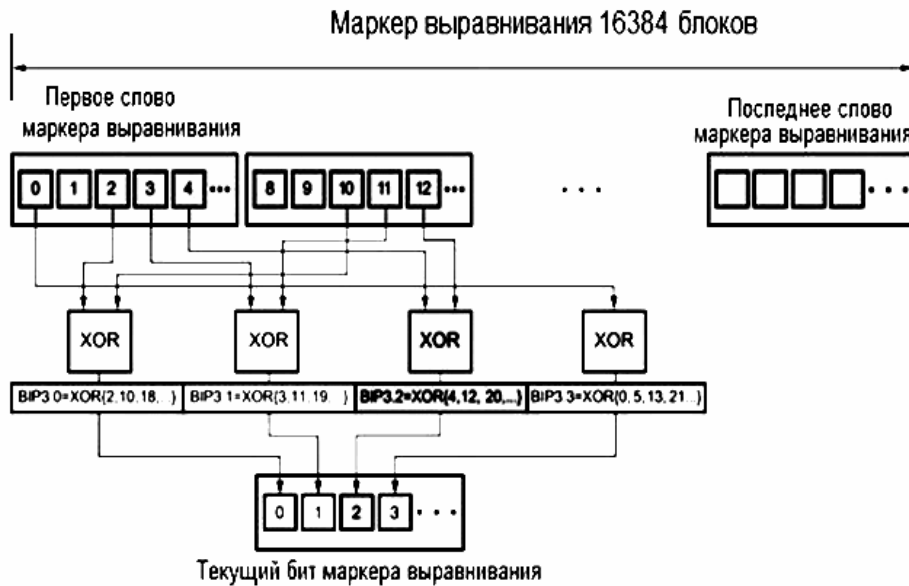
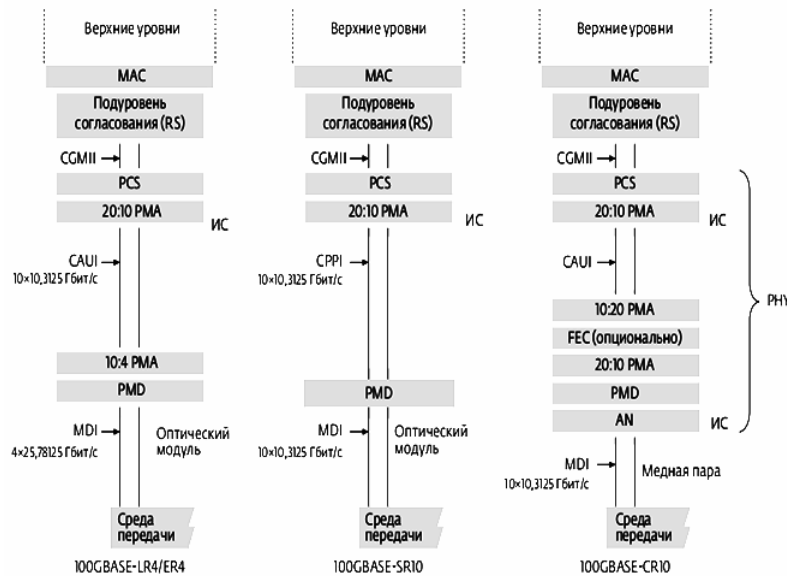


Рис. 4. Включение маркера выравнивания подуровнем PCS

К группе протоколов, используемых для реализации интерфейсов физического уровня, основанной на методе блочного кодирования данных кодом 64В/66В и использующей спецификацию PMA, относится протокол 100GBASE-R физического подуровня PCS. Протокол осуществляет кодирование/декодирование данных, полученных от интерфейса (переданных на интерфейс) CGMII в виде кода в последовательной форме, а также распределение данных на PMA по нескольким параллельным потокам. В качестве примера на рис. 5 приведены варианты архитектуры 100 Гбит Ethernet для протоколов 100GBASE-ER4, 100GBASE-LR4, 100GBASE-SR10 и 100GBASE-CR10 систем передачи сигналов.



*Рис. 5. Варианты архитектуры 100 Гбит Ethernet для протоколов 100GBASE-ER4, 100GBASE-LR4, 100GBASE-SR10 и 100GBASE-CR10*

Схемы содержат CPPI – параллельный физический интерфейс, сервисный интерфейс CAUI (Gb per second Attachment Unit Interface) – 100 Гбит модуль интерфейса подключения; а также логический интерфейс CGMII.

В настоящее время для одновременной передачи данных со скоростью 100 Гбит и более по нескольким каналам используется высокоскоростной последовательный интерфейс на основе меди. Для компенсации ухудшения качества сигнала применяют его выравнивание на сторонах передачи и приема. Подобные решения эффективны при небольших длинах медных кабелей, что в принципе недопустимо при реализации технологии облачных вычислений.



Реализацию стандарта IEEE Std 802.3ba GbE с учетом требований к широте полосы пропускания передаваемых сигналов на большие расстояния в сетевых устройствах можно обеспечить высокопроизводительными сетевыми решениями на основе оптических интерфейсов. Известные оптические стандарты включают модули форматов: SFP+, CFP (C form-factor pluggable) и CFP2 [9]. Модуль CFP2 (рис. 6) обладает пропускной способностью 100 Гбит/с, как и модуль CFP, но имеет в два раза меньшие габаритные размеры и энергопотребление, а также меньшую стоимость.

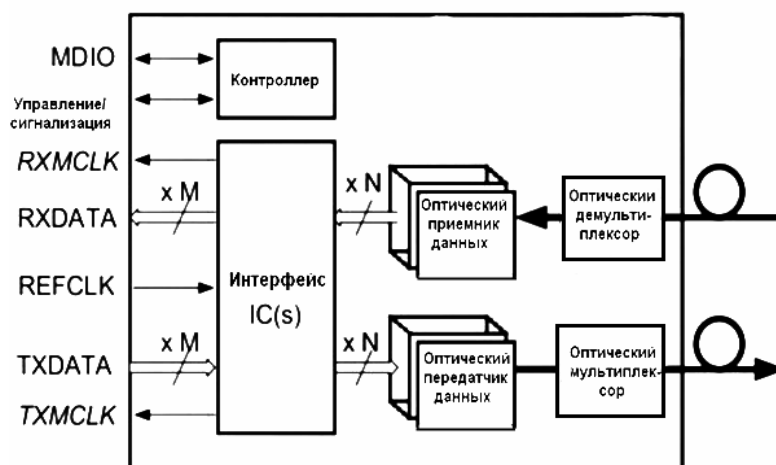


Рис. 6. Блок-схема оптического модуля формата CFP2

Технология облачных вычислений нуждается в применении современных широкополосных быстродействующих сетевых инфраструктур, способных обеспечить функционирование логического MAC – и физического уровней. Кроме того, реализация облачных вычислений требует преодоления ряда аппаратных проблем, обусловленных ограничением функций подсистемы хранения данных, коммутаторов, маршрутизаторов и систем приема/передачи. Успешное преодоление перечисленных проблем возможно при использовании специализированных устройств на ПЛИС (программируемых логических интегральных схем), как реконфигурируемых вычислительных сред, обладающих широким набором средств для обработки высокоскоростных потоков ввода/вывода данных со скоростью 100 Гбит/с и более. Устройства на ПЛИС для облачных вычислений обладают преимуществами в сравнении с другими программируемыми устройствами в производительности и энергопотреблении. Они также обеспечивают эффективную защиту информации, перекладывая эту ответственность с программ на аппаратуру.

Для отображения алгоритмов вычислений при прошивке ПЛИС необходимо компилировать модули с динамическими структурами данных, рекурсиями и неоднородным мелкодисперсным параллелизмом.

Один из возможных выходов – это оверлейная архитектура (VESPA, VIPERS и VEGAS). Другой путь – распространение специализированного языка. Для реализации параллельных вычислений в конфигурациях на ПЛИС с массовым параллелизмом предлагается открытый кросс-платформенный стандарт OpenCL (Open Computing Language) [10] (рис.7). Устройства OpenCL логически поделены на вычислительные узлы, состоящие из обрабатывающих элементов. Стандарт OpenCL охватывает иерархию уровней памяти, показанную на рис. 8. Центральным элементом модели платформы OpenCL является хост – первичное устройство, управляющее OpenCL вычислениями и осуществляющее взаимодействие с пользователем.

Использование стандарта OpenCL позволяет описывать параллельные алгоритмы вычислений при их реализации на ПЛИС на более высоком уровне абстракции и с меньшим временем проектирования, чем при использовании языков описания аппаратных средств более низкого уровня VHDL или Verilog.

В устройствах для обработки высокоскоростных потоков данных 100 Гбит/с могут использоваться гетерогенные 3D-матрицы ПЛИС Virtex-7 H580T, состоящие из матриц кремния SSI (Stacked Silicon Interconnect) [11]. Устройством ПЛИС Virtex-7 H580T предусматриваются дополнительные возможности отладки функций оптических транспортных сетей OTN, логического MAC-уровня, Interlaken IP и исключается необходимость использования специальных кодируемых элементов и устройств ASSP (Application Specific Standard Product). ПЛИС Virtex-7 H580T поддерживает: протокол SFI-S с одиннадцатью полосами на 11.2 Гбит/с и до семидесяти двух 13,1 Гбит/с SerDes (Serializer/Deserializer); протокол SFI-S с пятью полосами 28 Гбит/с (одна полоса на устранение перекося) и до шестнадцати 28 Гбит/с SerDes. Гетерогенная архитектура ПЛИС, реализующая до шестнадцати трансиверов, обеспечивает скорость 28 Гбит/с для оптического модуля формата CFP2. На скорости 100 Гбит/с предусмотрены дальний (LR – до 10 км) и сверхдальний (ER – до 40 км) режимы работы оптического модуля CFP2. Применяя двух-и/или четырех ядерный блок кодирования можно подключить к Virtex-7 H580T до восьми 4x25/28 Гбит/с модулей CFP2. ПЛИС с поддержкой масштабируемого интерфейса SerDes могут быть использованы для реализации двухпортового блока кодирования с расширенными функциями тестирования и отладки. На рис. 9 показано совместное подключение двухпортового блока кодирования (с расширением двух 100 Гбит портов) на основе Virtex-7 H580T и оптического модуля CFP2.

Блок кодирования обеспечивает поддержку: 10x10 Гбит/с CAUI интерфейса для последовательного интерфейса связи 4x25 Гбит/с перевод 11,2 Гбит/с десяти полосного SFI-S интерфейса в 28 Гбит/с четырехполосный SFI-S интерфейс, с устранением перекося полосы; физического перевода интерфейса с OTL 4.10 на OTL 4.4.

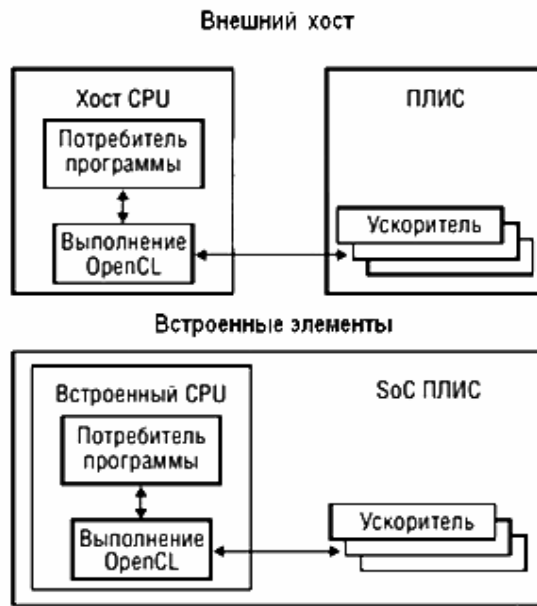


Рис. 7. Использование OpenCL на ПЛИС

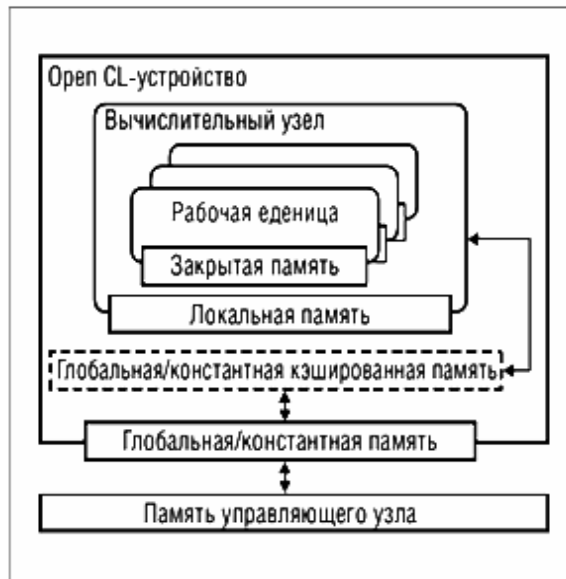


Рис. 8. Иерархия памяти OpenCL

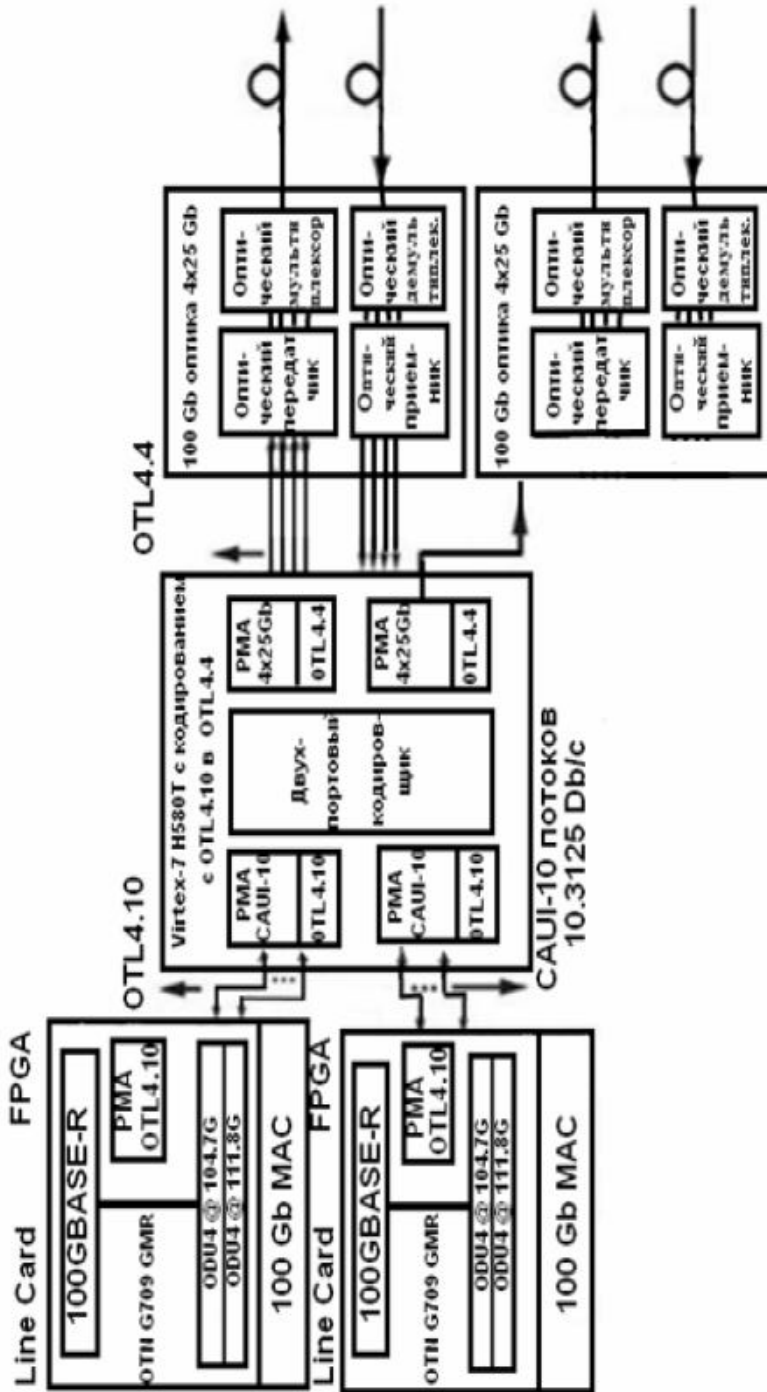


Рис. 9. Блок-схема 100 Гбит оптического модуля и ПЛИС

В целях повышения плотности 100 Гбит портов блока кодирования осуществляется подключение ASIC/FPGA/ASSP 100 Гбит Ethernet с использованием функций стандартов MAC или OTN. На одной ПЛИС Virtex-7 H580T можно создавать 100 Гбит/с OTN транспондеры, содержащие несколько модулей CFP2. Для поддержки 100 Гбит/с связи, обнаружения и исправления ошибок при совместном функционировании ПЛИС Virtex-7 H580T и оптических модулей CFP в соответствии с протоколом OIF SFI-S 1.0 используется процессор FEC на 4-20 полос масштабируемого интерфейса.

Достоинства от использования блока кодирования в Virtex-7 H580T заключаются еще в том, что он позволяет осуществлять большие объемы тестирования, отладки и контроля ПЛИС. Для этого в ПЛИС реализуется встроенная модель генератора PRB на 13,1 Гбит/с или 28 Гбит/с для SerDes, позволяющая осуществить проверку физического подслоя PCS в различных режимах работы системы передачи данных. Кроме того, Virtex-7 H580T, имея значительное число блоков оперативной памяти, может в течении нескольких миллисекунд обеспечить проверку потоков получаемых данных различных длин. Блок кодирования также обеспечивает имитацию перекосов распространения сигнала. Для минимизации дрожания сигналов трансиверов в ПЛИС используется синхронизация с фазовой автоподстройкой частоты. Для компенсации потерь и поддержания целостности передаваемого сигнала в трансиверах реализуется передача предсказаний и автоматическая адаптация в линейном эквалайзере приемника.

Проведенные исследования 28 Гбит/с трансивера на Virtex-7 H580T показали, что полученная для него глазковая диаграмма (рис. 10) отражает низкий джиттер и высокое качество передаваемого сигнала.

Глазковая диаграмма (рис. 10) представляет собой открытый глаз без чрезмерного выравнивания.

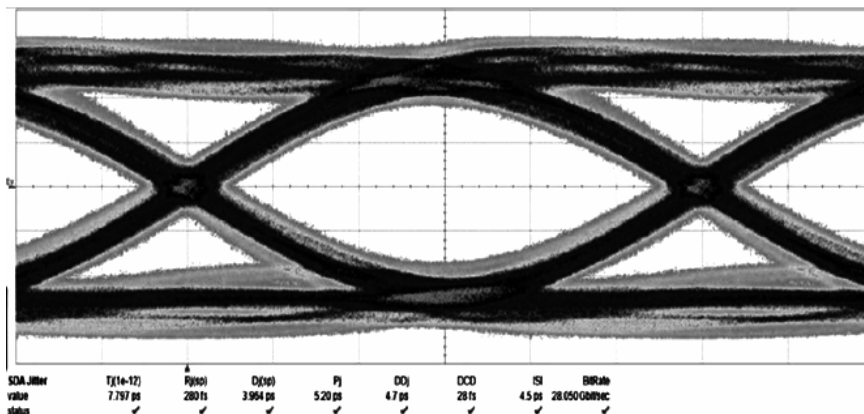


Рис. 10. Глазковая диаграмма 28 Гбит/с трансивера на Virtex-7 H580T

**Выводы.** При реализации параллельных вычислений в конфигурациях на ПЛИС с массовым параллелизмом может быть применен кросс-платформенный стандарт OpenCL. Инфраструктурные решения для технологии облачных вычислений, позволяющие обеспечить реализацию стандартов IEEE Std 802.3ba GbE и ИТ-услуг ISO 20000, эффективно строятся на трансиверах 28 Гбит/с на ПЛИС Virtex-7 H580T с низкими фазовыми шумами и оптических модулях CFP2. Подобные инфраструктурные решения для технологии облачных вычислений обеспечивают высокое качество передаваемого широкополосного высокоскоростного сигнала на расстояния до 40 км.

#### СПИСОК ЛІТЕРАТУРИ

1. Риз Дж. *Облачные вычисления*. – СПб.: БХВ-Петербург. – 2011. – 288 с.
2. Фингар П. *Dot.Cloud: облачные вычисления – бизнес-платформа XXI века*. – М.: Акваринная Книга, 2011. – 256 с.
3. Hewitt C. *ORGs for Scalable, Robust, Privacy-Friendly Client Cloud Computing / Massachusetts Institute of Technology*. – 2008. – Vol. 12. – №. 5. – P. 96-99.
4. Zhou S. *Understanding the Evolution Dynamics of Internet Topology / Physical Review E*. – 2006. – Vol. 74. – № 1. – P. 016124–1–016124–11.
5. *IEEE 802.3ba-2010. IEEE Standard for Information Technology. Amendment 4: Media Access Control Parameters, Physical Layers and Management Parameters for 40 Gb/s and 100 Gb/s Operation – 2010*. – 433 p.
6. [http://www.livebusiness.ru/tags/SAAS\\_platformy](http://www.livebusiness.ru/tags/SAAS_platformy)
7. D'Ambrosia J. *100 Gigabit Ethernet and Beyond / IEEE Communications Magazin*. – 2010. – Vol. 48. – № 3. – P. 6-13.
8. Toyoda H., Ono G., Nishimura Sh. *100GbE PHY and MAC Layer Implementation / IEEE Communications Magazin*. – 2010. – Vol. 50. – №. 3. – P. 41-47.
9. *CFP MSA Hardware Specification, Revision*. – 2010. – 53 p.
10. Вычужанин В.В. *Использование стандарта программирования OpenCL для организации вычислений на ПЛИС / Современная электроника*. – 2012. – № 2. – С. 70-74.
11. [www.xilinx.com](http://www.xilinx.com)

Стаття надійшла до редакції 02.04.2013

**Рецензент** – доктор технічних наук, професор, завідувачий кафедрою «Електротехніка та електрообладнання суден» Одеського національного морського університету **В.О. Яровенко**