

О разработке интерконнекта на активных оптоволоконных кабелях и программируемых логических интегральных схемах

С. М. Абрамов, И. А. Адамович, С. А. Блохин,
А. В. Елистратов, Л. Я. Карачинский, Ю. А. Климов,
И. И. Новиков, А. Ю. Пономарев, С. С. Ранцев, И. А. Фохт,
А. Ю. Хренов, А. Б. Шворин, Ю. В. Шевчук

ИПС им. А. К. Айламазяна РАН

27 сентября 2013

На пути к эксафлопсу

Вопросы, требующие решения:

- Вычислители (процессоры, ускорители)
- **Коммуникационная сеть**
- Системы хранения данных
- Энергопотребление
- Охлаждение
- Программное обеспечение



Коммуникационная сеть

Варианты решений:

- Широко используется универсальное коммерчески доступное решение InfiniBand (в Top500: 41% систем, 28% суммарной производительности)
- Для топовых и специализированных систем используются заказные сети (Cray, IBM, Fujitsu) (в Top500: 15% систем, 58% суммарной производительности)



Российские разработчики:

- ОАО «НИЦЭВТ»: **Ангара**
- ИПМ им. М. В. Келдыша РАН и НИИ «Квант»:
МВС-Экспресс
- РФЯЦ-ВНИИЭФ
- ИПС им. А. К. Айламазяна РАН: **СКИФ-Аврора**

Зачем своя сеть

InfiniBand — хорошая универсальная коммерчески доступная коммуникационная сеть. Но есть причины заниматься своими сетями.

- Бывают ситуации, в которых InfiniBand показывает слабые результаты. Например, темп выдачи сообщений (message rate)
- Чтобы построить эксафлопс, нужно обладать компетенцией во всех вопросах, в частности, в коммуникационных сетях



Разрабатываемый интерконнект

Характеристики интерконнекта

- Бескоммутаторный интерконнект (direct network)
- Четыре QSFP+ разъема на адаптере
- Скорость передачи данных по кабелю — 56 Гбит/с (4×14 Гбит/с)
- Соединение с узлом — PCI-Express Gen3 x8 (64 Гбит/с)
- ПЛИС фирмы Altera Stratix V серии GX (специализирована для интенсивных коммуникаций)



ПЛИС в интерконнектах

- ПЛИС по сравнению со СБИС:
 - ▶ цикл разработки — существенно короче
 - ▶ эффективность — сравнима
 - ▶ стоимость — разработка дешевле, цена экземпляра выше
 - ▶ гибкость — ПЛИС может использоваться как средство прототипирования (с переносом в СБИС)
- Примеры использования ПЛИС в интерконнектах
 - ▶ Ангара (ОАО «НИЦЭВТ»)
 - ▶ СКИФ-Аврора (ИПС им. А. К. Айламазяна РАН, Eurotech)
 - ▶ Extoll

Оптические межсоединения

- Рост вычислительной мощности требует повышения пропускной способности межсоединений
- Медные кабели не обладают достаточно высокими характеристиками
- Искажения уменьшают, «программируя» начальную форму импульса. При малых диаметрах потерь избежать невозможно
- 25 Гбит/с по меди — не далее 5 м.
Bert Jan Offrein, IBM Silicon Photonics Workshop, Munich, 23 May 2011
- Выход: оптика, если нужна высокая скорость передачи на значительные расстояния

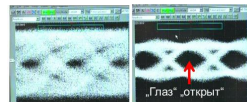
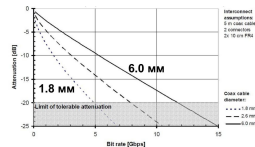


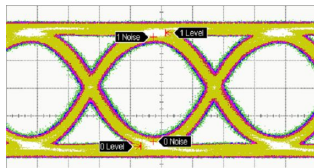
Диаграмма «глаза» медного кабеля. *IBM (Ch. Berger et al, Proc. SPIE, vol. 4942, pp.222-235)*



Потери в кабеле длиной 5м для кабелей с различными диаметрами

Разрабатываемый активный оптический кабель

- Разрабатывается АОК с разъемами QSFP+ со скоростью передачи 56 Гбит/с (4×14), совместимый с InfiniBand FDR
- Работы ведут:
 - ▶ Выходцы из ФТИ им. А. Ф. Иоффе, Санкт-Петербург
 - ▶ ООО «Коннектор Оптикс», Санкт-Петербург — кристаллы ВИЛ и ФД
 - ▶ ЗАО «Центр ВОСПИ», Москва — разработка АОК

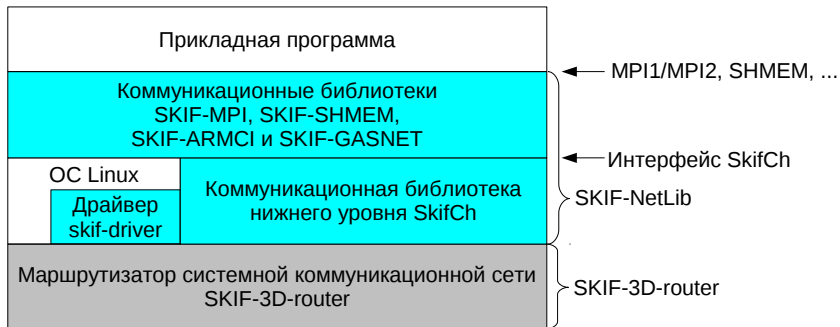


Глазковая диаграмма одного канала собранной оптической линии связи на скорости передачи данных 12.5 Гбит/с

Аппаратно-программное обеспечение

- Аппаратная часть
 - ▶ Прошивка ПЛИС, включающая маршрутизатор
 - ▶ SkifCh — интерфейс взаимодействия с процессором
 - ▶ Каналы передачи данных по АОК (с контролем ошибок)
- Драйвер ядра Linux для работы с платой как с PCIe-устройством
- Системные программы и утилиты для настройки и управления интерконнектом
- Прикладные параллельные библиотеки и системы: MPI, SHMEM, Co-Array Fortran, UPC, GASNet

Структура аппаратно-программного обеспечения



Заключение

- Будет завершена разработка отечественного интерконнекта на скорости передачи данных 56 Гбит/с (4×14 Гбит/с)
- Первые опытные образцы планируется продемонстрировать на НСКФ (26–29 ноября 2013 г., г. Переславль-Залесский)
- Запланировано дальнейшее развитие аппаратного и программного обеспечения
- В дальнейшем планируется разработка интерконнекта с пропускной способностью 100 Гбит/с (4×25 Гбит/с)

Работа выполняется в рамках государственного контракта с Министерством промышленности и торговли Российской Федерации № 12411.1006899.11.105.

Спасибо за внимание!

Вопросы?