

Поддержка механизма write combining в адаптере высокоскоростной коммуникационной сети (как устройстве PCI Express)

К. А. Курочкин, Е. Л. Сыромятников

ОАО «НИЦЭВТ»

При взаимодействии с хост-системой подключенного по интерфейсу PCI Express устройства (каковым, в частности, является адаптер высокоскоростной коммуникационной сети «Ангара») может использоваться два основных режима: Programmed Input/Output (PIO) и Direct Memory Access (DMA).

В режиме PIO передача данных в устройство PCI Express осуществляется посредством выполнения ЦП команд записи по адресам, ассоциированным с одним из регионов устройства PCI Express. В режиме DMA ЦП передает DMA-контроллеру базовый адрес и размер блока данных, который необходимо передать, после чего устройство самостоятельно посылает запросы чтения этого участка памяти без участия ЦП.

Адаптер высокоскоростной коммуникационной сети «Ангара» посылает сетевые пакеты, записываемые хост-системой в кольцевые FIFO-буфера («инжекционные буфера»), расположенные в одном из регионов устройства PCI Express.

В случае, если пакеты содержат малое количество данных, то эффективнее оказывается записывать непосредственно передаваемые данные, а не указатель на них для чтения их адаптером (то есть, использовать режим PIO, а не DMA), так как накладные расходы, связанные с необходимостью предварительной передачи базового адреса и размера участка памяти и последующего его чтения будут существенно выше затрат на запись самих данных.

Для повышения эффективности работы с периферийными устройствами в режиме PIO в современных ЦП был введен механизм write combining, обеспечивающий агрегацию команд записи в регион устройства PCI Express по последовательным адресам. Однако подобная агрегация может приводить к нарушению последовательности записей в инжекционный буфер. Таким образом, на стороне адаптера стоит задача производить переупорядочивание записей, восстанавливая семантику FIFO.

При дизайне блока приёма трафика PCI Express важно учесть вопрос масштабируемости — необходимо иметь возможность поддерживать большое количество одновременно инжектирующих процессов; аппаратным решением этой задачи (помимо возможных программных, ведущих к падению некоторых характеристик производительности и дополнительным накладным расходам) является поддержка большого количества инжекционных буферов. В свою очередь, большое количество независимых инжекционных буферов достаточной ёмкости (размер инжекционного буфера ограничен снизу необходимостью покрывать задержку на обновление кредитной информации FIFO-буфера на хост-системе) ведёт к большому количеству необходимых аппаратных ресурсов адаптера и низкой эффективности этих ресурсов.

Исходя из этих требований, для адаптера высокоскоростной коммуникационной сети «Ангара» был предложен механизм объединения инжекционных буферов, позволяющий динамически распределять для инжектирующих потоков свободные буфера, тем самым эффективно используя аппаратные ресурсы при любом (в пределах аппаратно поддерживаемого) количестве параллельно выполняющихся процессов. Каждый из буферов поддерживает слабое упорядочивание данных.

Концепция данного механизма была проработана и реализована на RTL-модели. Были получены оценки потребления аппаратных ресурсов для нового механизма и механизма, реализованного в текущей версии адаптера.