

УДК 681.3

И.А. Мартынов, Э.И. Ватутин

evatutin@rambler.ru

Юго-Западный государственный университет, Курск

ИЗМЕРЕНИЕ РЕАЛЬНОЙ ПРОПУСКНОЙ СПОСОБНОСТИ ШИНЫ PCI EXPRESS С ИСПОЛЬЗОВАНИЕМ ВИДЕОКАРТ С ПОДДЕРЖКОЙ ТЕХНОЛОГИИ CUDA В КАЧЕСТВЕ ПЕРИФЕРИЙНЫХ УСТРОЙСТВ

В работе предложена методика и приведены результаты измерения реальной пропускной способности шины PCI Express современных компьютеров путем измерения времени передачи блоков информации фиксированного размера.

При проектировании специализированных вычислительных устройств-акселераторов, работающих в составе аппаратно-программных комплексов, общей стратегией сокращения времени выполнения поставленной задачи является перенос наиболее трудоемких операций с программного уровня на аппаратный [1–3]. При этом, в соответствии с законом Амдала, выигрыш во времени обработки (ускорение) определяется как

$$\eta = \frac{1}{\alpha + \frac{1-\alpha}{p} + \beta(p)}, \text{ где } \alpha - \text{ в данном случае доля вычислений, выполня-$$

емая программной частью, p – ожидаемое ускорение от использования аппаратной части (в простейшем случае – число параллельно работающих вычислителей), $\beta(p)$ – накладные расходы, по большей части связанные с передачей исходных данных из оперативной памяти в память специализированного устройства и возвратом результата. С ростом значения слагаемого $\beta(p)$ эффективность использования аппаратно-программного комплекса падает, что накладывает ограничение на значение p , которое в каждом конкретном случае имеет оптимальное значение p^* , отклонение от которого нецелесообразно, т.к. приводит к снижению итогового выигрыша η либо не приводит к его существенному увеличению при росте затрат на аппаратную составляющую. Снижение значения слагаемого $\beta(p)$ можно добиться путем уменьшения трафика между программной и аппаратной частью, что возможно далеко не всегда и определяется спецификой задачи, либо путем повышения пропускной способности интерфейса, что имеет вполне определенные ограничения физической природы.

На сегодняшний день широко распространенным интерфейсом взаимодействия с периферийными устройствами является шина PCI Express, и многие акселераторные средства (GPU AMD/ATI и nVidia, Intel MIC, NeuroMatrix, FPGA и пр.) сопровождаются его поддержкой. В данной работе произведено измерение его реальной пропускной способности в зависимости от аппаратной конфигурации машины (в первую очередь от типа про-

цессора и связанного с ним поколения стандарта PCI Express и поддерживаемого типа оперативной памяти). Для этого в качестве периферийных устройств были выбраны видеокарты NVidia с поддержкой технологии CUDA. В ходе тестирования осуществлялось выделение блоков фиксированного размера (100 МБ) в оперативной памяти (Host) и в глобальной памяти видеокарты (Device) с последующим поочередным копированием блока информации. В используемой схеме время копирования информации в/из GPU лимитировано именно интерфейсом PCI Express, т.к. пропускные способности оперативной памяти и глобальной памяти GPU существенно выше (см. таблицу).

Таблица. Результаты тестирования пропускной способности, ГБ/с

GPU	CPU	CPU→CPU	CPU→GPU	GPU→CPU	GPU→GPU
GTX 970	i7 4770	8,7–8,9	5,0–5,2	5,0–5,4	65,1–65,9
GTX 770	i7 4770	5,3–6,2	4,2–5,5	4,3–6,4	81,5–81,9
GTX 760	i5 3470	8,4	4,7	4,8	66,5
GTX 660	i7 920	3,5–3,8	2,8–3,0	3,0–3,2	53,2–54,0
GTX 660	E6750	1,6	2,0	1,4	53,6
GTX 450	E6300	1,4	1,8	1,4	22,8
GTX 450 Eco	i3 2130	5,1	4,9	5,3	8,8
GTX 450 Eco	i5 4590	4,2–4,3	2,0–2,1	1,9	8,8–9,1
GTX 450 Eco	i3 4160	4,2–4,3	2,0–2,1	1,9	8,8–9,1
GTS 250	A4400+	1,2	0,9	0,9	11,6
K20Xm	E5-2650	4,8	2,8	2,5	83,6
C1060	X5570	4,5	4,3	3,2	36,2

Полученные результаты позволяют сделать вывод о том, что имеет место сильная зависимость от используемого поколения вычислительной техники. Они могут быть использованы при теоретической оценке ускорения в ходе проектирования специализированных устройств-акселераторов.

1. Ватутин Э.И. Проектирование логических мультиконтроллеров. Синтез разбиений параллельных граф-схем алгоритмов. Saarbrücken: Lambert Academic Publishing, 2011 г. 292 с.
2. Ватутин Э.И., Зотов И.В., Титов В.С. Акселератор для быстрого преобразования конструктивных подмножеств вершин параллельных алгоритмов // Параллельные вычисления и задачи управления (РАСО'10). М.: ИПУ РАН, 2010. С. 1301–1366.
3. Martynov I.A., Vatutin E.I., Titov V.S. Hardware oriented classification of binary relations of graph-schemes of parallel algorithms // Eighth World Conference on Intelligent Systems for Industrial Automation (WCIS – 2014). Tashkent, 2014. PP. 70–73.