

Построение эффективного тракта обработки сигналов с использованием технологии CUDA

Александр Тумачек, Андрей Ермаков (Москва)

В статье приводится сравнение алгоритмов ЦОС на платформах ADSP-TS201s и Nvidia GT240. Отмечено преимущество технологии CUDA при большой размерности матриц.

Известно, что системы связи характеризуются пропускной способностью и качеством принимаемых сигналов. При возрастающих требованиях к пропускной способности каналов и всё более плотном заполнении эфира важно эффективно задействовать частотно-временной ресурс. Например, системы связи с TDMA и CDMA 20 лет назад применялись весьма ограниченно, сейчас это массовый рынок оборудования с множеством сервисов и приложений.

Системы подвижной связи, такие как сотовые сети, изначально были предназначены для передачи голоса, но были приспособлены для передачи больших объёмов информации. Сотовые сети и Интернет охватывают сотни миллионов пользователей по всему миру. Возможно, в будущем мы увидим новое поколение сетей связи, а Интернет

преобразуется в симбиоз гибридных разноуровневых систем. Соединение связанных сетей – от микро- до макроячеек, от частной телефонной сети до Интернета, от низкоорбитальных спутниковых комплексов – до геостационарных платформ – станет реальностью. Инфраструктурные проекты такого масштаба возможно организовать только при снятии существующих ограничений, увеличивая мощность и эффективность связного оборудования [1].

С появлением высокоскоростных и широкополосных систем аналого-цифрового преобразования возникает задача анализа и обработки массивов информации, полученной из эфира в реальном масштабе времени. С расширением полосы приёма возрастает нагрузка на тракт обработки сигналов и, следовательно, на подсистемы, реализующие их обработку, в том числе на демодулятор, где происходит получение оценок сигнала и восстановление энергии. Демодулятор является критичным к производительности вычислительной платформы.

Для получения оценок сигнала может применяться многоканальный пространственно-временной фильтр с оптимальными коэффициентами, полученными решением уравнения Винера Хопфа [2]. Основная математическая нагрузка при получении оценок ложится на узел расчёта оптимального вектора весовых коэффициентов (ВВК) на каждом шаге. Помимо этого, выполняется анализ спектра выборочной корреляционной матрицы сигнала путём расчёта QR-разложения и накопления собственных векторов и собственных значений.

Для расчёта ВВК применяется решение системы линейных алгебраических уравнений (СЛАУ) для квадратной выборочной корреляционной матрицы сигналов и шумов (ВКМ) размерностью $(N \times N_{\text{trackt}}) \times (N \times N_{\text{trackt}})$, где $N \times N_{\text{trackt}}$ – длина линии задержки ли-

нейного корректора, умноженная на число трактов системы. Решение основано на процедуре LU-разложения и вычислении коэффициентов из линейных уравнений для треугольных матриц. Для получения СЛАУ используются операции перемножения вектора на матрицу, внешнее и внутреннее произведение векторов, вычитание и сложение матриц. При обработке широкополосных сигналов большая часть времени уходит на выполнение этих действий.

Для повышения скорости вычислений авторами были использованы массивно-параллельные процессоры и технология CUDA (Computer Unify Design Architecture) [3]. Для тестирования и проверки тракта обработки демодулятора использовалось окружение Matlab 2010b с поддержкой GPU. Для расчёта LU-разложения и решения СЛАУ применялась следующая процедура:

```
% code analog [L,U] = lu(RxxSum);
% LU implementation on CUDA
dRxxSum = GPUArray(RxxSum);
%set GPU Matri(NxNtrackt)x(NxNtrackt)
%dRxxSum on device
[dL,dU,dpiv] = lu(dRxxSum,'vector');
L = gather(dL);
U = gather(dU);
y = L\f;
Wopt = U\y;
```

Итоговая реализация алгоритмов решения СЛАУ и вычисления QR-разложения была выполнена в виде динамических библиотек на C++ с использованием CUDA и ADSP-TS201s.

Представим некоторые результаты сравнения производительности GPU и ADSP-TS201s. На рисунке 1 показан график времени вычисления QR-разложения для GPU. В таблице 1 приведены значения времени вычисления QR-разложения для ключевых размерностей на платформах ЦОС-ADSP и GPU.

Можно заметить, что система с GPU показывает наибольшую эффективность при размерности задачи больше числа вычислительных ядер; в наших экспериментах это 128. В сравнении с



Рис. 1. QR-разложение выборочных корреляционных матриц сигнала

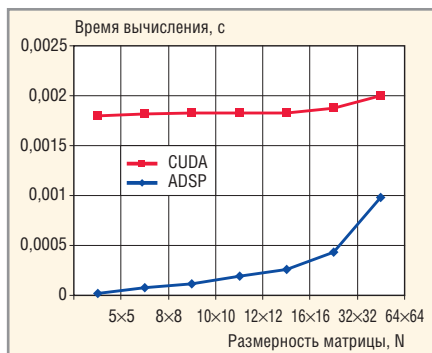


Рис. 2. QR-разложение выборочных корреляционных матриц сигнала на платформе CUDA и ADSP TS-201s

Таблица 1. Значения времени вычисления QR-разложения для ключевых размерностей на платформах ЦОС-ADSP и GPU

Размерность	Количество тактов ADSP-TS201S	Время выполнения, с, ADSP-TS201S	Время выполнения, с, NVIDIA GT 240
5 × 5	10 985	0,000017576	0,001802
8 × 8	45 685	0,000073096	0,001816
10 × 10	74 853	0,000119764	0,001825
12 × 12	118 300	0,000189280	0,001828
16 × 16	161 747	0,0002588	0,00183
32 × 32	205 194	0,00042831	0,00187
64 × 64	248 641	0,00097828	0,002

Таблица 2. Время QR-разложения при больших размерностях матриц

Размерность	Время выполнения, с, GT 240
64 × 64	0,0020364
80 × 80	0,0042650
100 × 100	0,0038563
128 × 128	0,0043692
140 × 140	0,0053865
160 × 160	0,0058847
180 × 180	0,0060568
256 × 256	0,0080498
512 × 512	0,1038184
1024 × 1024	0,0738037
2048 × 2048	0,2699768

GPU, система демодулятора на ADSP-TS201S эффективна на небольших размерностях (см. рис. 2). При использовании ADSP-TS201s для расчёта QR-разло-

жения на больших матрицах возникают проблемы недостатка памяти и устойчивости работы алгоритма для матриц плохой обусловленности. Таблица 2 со-

держит данные о времени QR-разложения при больших размерностях матриц.

Таким образом, в ходе работы по сравнению платформ ЦОС ADSP-TS201s и GT 240 Nvidia было установлено, что для расширения возможностей широкополосного приёма и обработки сигналов, для задач больших размерностей целесообразно использовать архитектуру CUDA Nvidia. Задачи меньших размерностей (для $n < 64$) по-прежнему могут решаться на платформе ADSP-TS-201s, которая обеспечивают максимум производительности на сегодняшний день. Реализация связанного оборудования на этих платформах в обоих случаях отвечает требованиям к вычислительной мощности. Для работы в широком диапазоне размерностей рекомендуется использование обеих вычислительных платформ.

ЛИТЕРАТУРА

1. Lee H.-N. Adaptive Diversity Combining, Equalization and Sequence Decoding for Time-Varying Dispersive Channels. Los Angeles, Univ. of California. 1999. P. 3–4.
2. Коуэн К.Ф.Н., Грант П.М. Адаптивные фильтры. Мир, 1988. С. 31–35.
3. www.nvidia.com.

