

РОССИЙСКАЯ ФЕДЕРАЦИЯ



(19) **RU** **2 604 985** ⁽¹¹⁾ **C2** ⁽¹³⁾

(51) МПК

G06F 17/50 (2006.01)

ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(21)(22) Заявка: [2014145373/08](#), 11.11.2014

(24) Дата начала отсчета срока действия патента:
11.11.2014

Приоритет(ы):

(22) Дата подачи заявки: 11.11.2014

(43) Дата публикации заявки: 27.05.2016 Бюл.
№ [15](#)

(45) Опубликовано: [20.12.2016](#) Бюл. № [35](#)

(56) Список документов, цитированных в отчете о
поиске: US 7979814 B1, 12.07.2011. US
2006/0242618 A1, 26.10.2006. US
2005/0166174 A1, 28.07.2005. US 7594095 B1,
22.09.2009. RU 2412469 C2, 20.02.2011.

Адрес для переписки:

394026, г. Воронеж, Московский просп., 14,
ФГБОУ ВО "ВГТУ", патентный отдел

(72) Автор(ы):

**Башкиров Алексей Викторович (RU),
Климов Александр Иванович (RU),
Науменко Юрий Сергеевич (RU)**

(73) Патентообладатель(и):

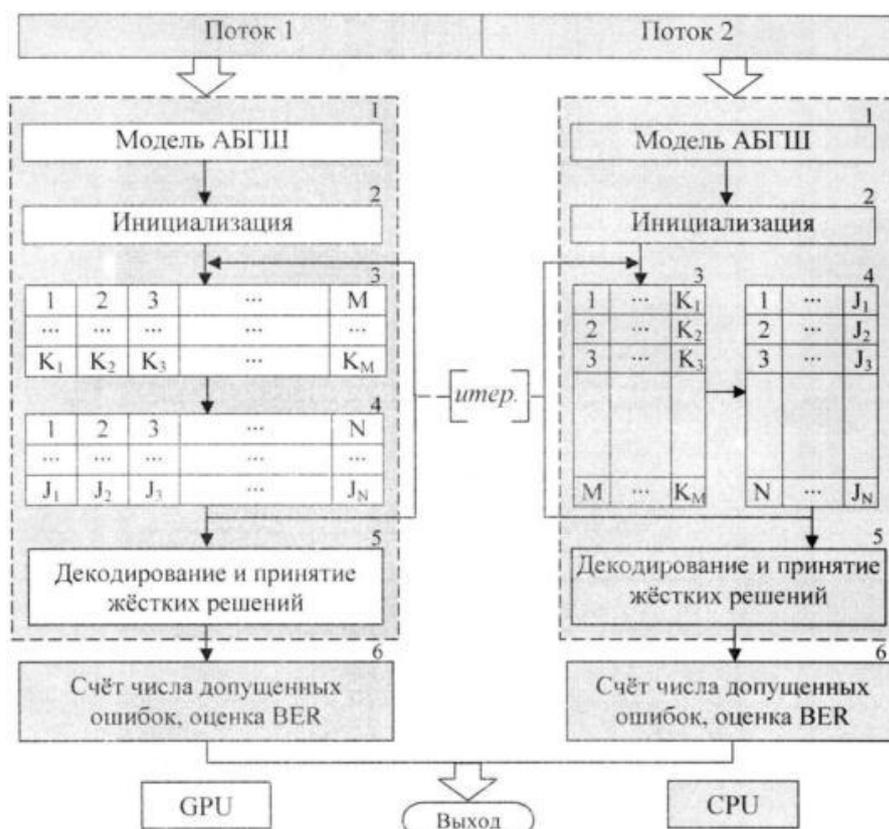
**Федеральное государственное бюджетное
образовательное учреждение высшего
образования "Воронежский
государственный технический
университет" (ФГБОУ ВО "ВГТУ",
ВГТУ) (RU)**

(54) СПОСОБ ОРГАНИЗАЦИИ ВЫЧИСЛЕНИЙ НА ГРАФИЧЕСКИХ ПРОЦЕССОРАХ ДЛЯ
МОДЕЛИРОВАНИЯ ПОМЕХОУСТОЙЧИВОСТИ НИЗКОПЛОТНОСТНЫХ КОДЕКОВ

(57) Реферат:

Изобретение относится к автоматизированному проектированию, технике моделирования и проверки кодов и может быть использовано при цифровом моделировании характеристик помехоустойчивых низкоплотностных кодеков в однопроцессорных гетерогенных компьютерных системах малой производительности. Техническим результатом применения способа является увеличение производительности вычислений, производимых в процессе моделирования. Такой результат достигается благодаря тому, что хостовой частью выполняют подготовку и передачу данных GPU части, затем производят предварительную оценку значений производительности вычислений на CPU (T_{CPU}) и GPU (T_{GPU}) в отдельности друг от друга, затем общее число моделируемых точек разбивают в соотношении $Q/Z = T_{CPU}/T_{GPU}$, после чего часть точек Q моделируют основным потоком вычислений на GPU, синхронизируемых с помощью CPU, и запускают дополнительный поток вычислений на CPU для моделирования части

точек Z , по завершении моделирования осуществляют передачу результатов вычислений хостовой части. 2 ил.



Обобщённая схема организации моделирования

Фиг. 1

Изобретение относится к автоматизированному проектированию, технике моделирования и проверки кодов и может быть использовано при цифровом моделировании характеристик помехоустойчивых низкоплотных кодеков в однопроцессорных гетерогенных компьютерных системах малой производительности для ускорения производимых вычислений.

Техническим результатом является увеличение производительности вычислений, производимых в процессе моделирования.

Известен способ моделирования, описанный в патенте США «Lithographic simulations using graphical processing units», US 2006/0242618 A1, МПК G06F 17/50, опубл. 26.10.2006. Способ заключается в инициализации хостовой части основным вычислителем - CPU (англ. central processing unit - центральный процессор), инициализации данных дополнительного вычислителя - GPU (англ. graphics processing unit - графический процессор), передачи данных от хостовой части к GPU, запуске вычислений на GPU, синхронизируемых хостовой частью и передаче результатов вычислений от GPU обратно хостовой части.

Недостатком такого подхода является неполное использование доступных вычислительных мощностей CPU.

Наиболее близким по своей технической сущности к заявленному способу является способ, описанный в патенте США «Model implementation on GPU», US 7979814 B1, МПК G06F 17/50, опубл. 12.07.2011. Способ заключается в подготовке и передаче хостовой частью данных GPU части, запуске основного потока для CPU и старте потока вычислений на GPU, передаче результатов вычислений хостовой частью по завершении моделирования.

Недостатком способа является неполная загрузка основного вычислителя гетерогенной системы вследствие того, что он реализует лишь инициализацию и сопровождение (синхронизацию) вычислений GPU части. В отношении вычислений в гетерогенных системах низкой производительности, такая схема организации моделирования становится схемой с неполным использованием потенциала гетерогенной системы в условиях, когда выигрыш в скорости вычислений на GPU невелик, а гетерогенная система содержит в качестве основного вычислителя лишь однопроцессорное устройство.

Сущность изобретения

Решаемой технической задачей изобретения является увеличение производительности вычислений, выполняемых в процессе моделирования.

Сущность изобретения поясняется приведенными далее чертежами в отношении моделирования и проверки низкоплотностных кодов (N, J, K), где N - длина кода, J - количество единиц в столбце, а K - количество единиц в строке проверочной матрицы кода. На фиг. 1 представлена архитектура программной реализации предлагаемых решений, содержащая в своем составе: 1 - блок моделирования канала с аддитивным белым гауссовским шумом (АБГШ); 2 - блок инициализирующих процедур; 3 и 4 - блоки условной архитектуры итеративного декодера по итеративному алгоритму распространения доверия; 5 - блок финального декодирования и принятия жестких решений, 6 - блок оценки BER (англ. Bit Error Rate - вероятность битовой ошибки). Блоки 3 и 4 в обоих потоках отвечают за передачу сообщений от проверочных вершин к кодовым и обратно, однако в первом потоке, соответствующем вычислениям на GPU, блоки выполняются параллельно, а в потоке, соответствующем вычислениям на CPU, блоки выполняются последовательно.

Предлагаемый способ увеличения производительности заключается в следующем.

1. Сначала производят предварительную оценку производительности вычислений на CPU и GPU в отдельности друг от друга в соответствии с процедурой, псевдокод которой представлен в листинге 1.

Листинг 1

Псевдокод процедуры оценки производительности CPU и GPU

```
1. Инициализация GPU.
2. Число симуляций numsim=ceil(200000.0/N).
3. Старт таймера t1.
4. while(sim < numsim; sim=sim+1;)
  {
  4.1. Модель АБГШ.
  4.2. Декодирование на GPU в 10 итераций.
  }
5. Стоп таймера t1; TGPU=t1.
6. Старт таймера t2.
7. while(sim < numsim; sim=sim+1;)
  {
  7.1. Модель АБГШ.
  7.2. Декодирование на CPU в 10 итераций.
  }
8. Стоп таймера t2; TCPU=t2.
9. Коэффициент разбиения Q = (TCPU/TGPU)*1.2;
10. Принимается Z = 1.
```

Эмпирически было установлено, что моделирование декодирования 200 000 бит достаточно для приблизительной оценки производительности вычислений на CPU и GPU и при этом не так велико, чтобы загрузить гетерогенную систему на значительное время (более 10 сек). Домножение на коэффициент 1, 2 является поправочным; коэффициент вычислен также эмпирически. Функция $\text{ceil}(x)$ в листинге 1 возвращает ближайшее целое к значению x , округленное вверх.

2. Затем осуществляют разбиение задачи в соотношении $\frac{Q}{Z}$, причем:

$$\frac{Q}{Z} = \frac{T_{\text{CPU}}}{T_{\text{GPU}}},$$

где T_{CPU} и T_{GPU} - время, затраченное на расчеты центральным и графическим процессором в процессе выполнения вышеописанной процедуры соответственно. В отношении моделирования низкоплотностных кодеков разделить задачу оказывается возможным по точкам значений SNR, так как обычно их симулируется достаточное количество. Разбиение вычислительной задачи осуществляют следующим образом: Вводят вспомогательный коэффициент

$$\psi = 1 + \frac{1}{Q};$$

Вычисляют общее число точек моделирования:

$$\text{Max}_{\text{sim}} = \text{floor} \left(\frac{\text{SNR}_{\text{final}} - \text{SNR}_{\text{init}}}{\text{SNR}_{\text{incr}}} \right) + 1;$$

Вычисляют границы интервалов моделирования:

$$\text{SNR}_Q = \text{SNR}_{\text{init}} + \text{SNR}_{\text{incr}} \left(\text{floor} \left(\frac{\text{Max}_{\text{sim}}}{\psi} \right) - 1 \right);$$

$$\text{SNR}_Z = \text{SNR}_{\text{init}} + \text{SNR}_{\text{incr}} \left(\text{floor} \left(\frac{\text{Max}_{\text{sim}}}{\psi} \right) \right);$$

Здесь SNR_{init} - минимальное значение сигнал/шум; $\text{SNR}_{\text{final}}$ - максимальное значение сигнал/шум; SNR_{incr} - величина инкремента значения сигнал/шум.

3. После этого осуществляют организацию многопоточных вычислений, где моделирование части точек - Q производят основным потоком на вспомогательном вычислителе (GPU) с синхронизацией хостовой (CPU) частью, а моделирование части точек Z осуществляют дополнительным потоком вычислений на CPU. Основной поток (GPU, синхронизируемый CPU) моделирует точки SNR_1 : $\text{SNR}_{\text{init}} \leq \text{SNR}_1 \leq \text{SNR}_Q$.

Дополнительный поток (CPU) моделирует точки SNR_2 : $\text{SNR}_Z \leq \text{SNR}_2 \leq \text{SNR}_{\text{final}}$.

Достижимое повышение производительности подтверждается программно полученными данными вычислений, представленными на фиг. 2 в виде графика временного выигрыша, в зависимости от длины кода, при моделировании в гетерогенной системе посредством открытого стандарта реализации техники GPGPU (англ. General-purpose graphics processing units - вычислений общего назначения на графических процессорах) OpenCL (англ. Open Computing Language - открытый язык вычислений) характеристик группы кодов (N, 3, 6) при 10 точках SNR (англ. signal-to-noise ratio - отношение сигнал/шум).

Таким образом, при малых значениях длины (N=96) выигрыш достигает 80% и снижается до 60% при N=204 и 41% при N=273. Затем наблюдается установившийся участок со средним значением выигрыша 21% при длине кода от N=504 до N=3000. Говоря об актуальности и практической важности полученных результатов, упомянутые выше значения длины кода рекомендованы следующими современными техническими стандартами в области телекоммуникаций: 802.11 - WiFi

(беспроводные локальные и городские сети, N:648-1944); 802.16 - Mobile WiMAX
(местные и городские беспроводные сети, N:576-2304); 802.22 - WRAN
(беспроводные региональные сети, N:384-2304).

Область применимости способа ограничена условием $Q < \text{Max}_{sim}$. Стоит отметить, что принципиально способ реализуем и в гетерогенных системах с многопроцессорным CPU, с учетом полной загрузки всех ядер CPU.

Патенты

1. Патент США «Lithographic simulations using graphical processing units», US 2006/0242618 A1, МПК G06F 17/50, опубл. 26.10.2006.

2. Патент США «Model implementation on GPU», US 7979814 B1, МПК G06F 17/50, опубл. 12.07.2011.