

ПАРАЛЛЕЛЬНАЯ ОБРАБОТКА ДАННЫХ МНОГОКОМПОНЕНТНЫХ СЕЙСМИЧЕСКИХ НАБЛЮДЕНИЙ

Ключевые слова: *параллельное программирование, обработка сейсмических данных, оптимизация кластерных вычислений, большие наборы данных.*

ВВЕДЕНИЕ

С развитием новых технологий в сейсморазведке возросли объемы результатов исследований. Как известно, многие данные сейсмических наблюдений являются неточными, а иногда и искаженными за счет некоторых помех (шумов). Для их устранения использование обычного ПК нецелесообразно в силу его малой вычислительной способности. Выполнение расчетов на современных многоядерных ПК также не дает выигрыша во времени. Поэтому для обработки таких данных используют другие параллельные вычислители, среди которых наиболее эффективны:

- грид-системы (grid);
- системы для облачных вычислений (cloud computing);
- кластерные вычислительные комплексы;
- вычислители с использованием ускорителей обработки графики (GPU).

Задача параллельной обработки многокомпонентных сейсмических наблюдений в первую очередь связана с обработкой больших объемов данных. Ввиду недостаточной пропускной способности сети нецелесообразно использование грид-систем и систем для облачных вычислений для распараллеливания. Применение вычислителей с GPU возможно, но недостаточный объем памяти на некоторых типах устройств ограничивает такое решение.

Для проведения расчетов задачи задействован кластерный комплекс Инпарком-256, разработанный Институтом кибернетики им. В.М. Глушкова совместно с ГНПО «Электронмаш».

ОБРАБОТКА ДАННЫХ СЕЙСМОРАЗВЕДКИ

Типичная схема сейсмического исследования земной коры следующая [1]. На поверхности вдоль планового сейсмического профиля размещают приемники (геофоны), способные регистрировать свои перемещения в пространстве. Вдоль того же профиля помещают источник сейсмического сигнала — определенное количество вибраторов-машин, синхронно ударяющих об землю тяжелыми молотами. В результате воздействия источника в земной коре возбуждается акустическая волна, которая со скоростью звука распространяется вниз и в стороны от источника. На сейсмической границе, т.е. границе перепада скорости звука в породе, происходит преломление и отражение волн. Отраженная волна возвращается в обратном направлении, достигает поверхности и регистрируется геофонами. В каждом геофоне записывается сейсмическая трасса как зависимость от времени зарегистрированной скорости вертикального смещения прибора. Совокупность сейсмических трасс образует сейсмограмму. После регистрации сейсмоисточник и приемники перемещают вдоль профиля и повторяют операцию.

Для получения модели некоторого района сейсмические исследования проводят вдоль сети профилей. К настоящему времени преобладающе используется 3D-сейсморазведка, когда волну регистрируют приемники, расположенные вдоль нескольких параллельных линий приема, как правило, ортогонально к профилям возбуждения. В результате таких сейсмических наблюдений формируется сетка, которую регулярно сохраняют для соблюдения четкой последовательности результатов. Также широко используют многокомпонентную сейсморазведку 3С, в которой геофон отдельно регистрирует скорости смещения вдоль трех пространственных осей, и 9С, в которой трехкомпонентный прием сочетается с поочередным возбуждением трех направленных волн в каждой точке.

Сейсмограммы дают важную информацию о структурном строении земной коры. В частности, нефть и газ находятся в пористых породах-коллекторах (песчаниках, трещиноватых известняках), ограниченных плотными покрывками. Значительная разница плотности и соответственно скорости звука между коллектором и плотной покрывкой создает резкие границы, отображаемые на сейсмограммах высокими амплитудами сигнала.

ПОВЕРХНОСТНЫЕ ВОЛНЫ

В записях отраженных волн, полученных на суше, обычно встречаются поверхностные волны, генерируемые сейсмическим источником и распространяемые по всей поверхности [2, 3]. Волны, энергия колебаний которых сосредоточена в узком интервале глубин близко к поверхности, называют поверхностными. Среди них можно выделить волны Релея (ground roll), Лява, прямые звуковые. Часто их путают с волнами, отражаемыми на неглубоких границах; такую разновидность волн также можно условно назвать поверхностными.

Поверхностные волны, которые получены в результате сейсморазведки и в отличие от классических волн Релея содержат разные моды, иногда называются псевдорелеевскими волнами. Волны Релея — это колебания, эллиптически поляризованные в вертикальной плоскости, их можно выразить суперпозицией взаимосвязанных смещений (продольного и вертикально поляризованного поперечного). Фаза вертикальной составляющей отличается на четверть периода от фазы горизонтальной составляющей, что и приводит к эллиптическому характеру перемещения частиц среды в направлении, противоположном направлению распространения волн Релея [2–4]. Поскольку эти волны практически не несут информации о глубинных свойствах геологической среды, в традиционной сейсморазведке они воспринимаются как серьезная преграда и должны быть устранены.

Обычно поверхностные волны характеризуются относительно высокими амплитудами и поэтому существенно маскируют объемно отраженные волны, которые представляют сигнальную часть записи в традиционных методах поиска нефти и газа. Теоретические наблюдения показывают [4, 5], что поверхностные источники сейсмических колебаний типа вертикальной силы, самыми распространенными среди которых являются вибраторы, генерируют приблизительно 65% энергии в виде поверхностных волн Релея и только 8% в виде продольных волн. Вследствие неоднородной структуры волн Релея их амплитуда довольно быстро уменьшается экспоненциально в глубину. Поэтому энергия на глубине, равной длине волны, на порядок меньше, чем на поверхности полупространства. Таким образом, они распространяются по поверхности, практически не проникая в глубину.

В связи с квазичилиндрической формой плотность энергии в волнах Релея по мере их удаления от источника уменьшается пропорционально расстоянию R , так что геометрически расстояние пропорционально \sqrt{R} . Такой набор параметров меньше, чем в объемных волнах, расстояние которых пропорционально длине R

в однородном пространстве. Если отсутствует группировка приемников и источников поверхностных волн в некоторых секторах сейсмических записей, то такие волны могут существенно превышать по амплитуде отраженные волны и их относительная роль в общей энергии волнового процесса, который регистрируется на земной поверхности, может усиливаться с удалением от источника. В такой ситуации выходные сейсмограммы и остальные сейсмические изображения могут иметь низкое качество, не удовлетворяющее условиям дальнейшей геологической интерпретации.

Поскольку упругие свойства среды меняются по глубине, это больше влияет на скорость распространения длинных (низкочастотных) волн, чем коротких (высокочастотных). Поэтому при резком увеличении скорости дифференциации ближе к поверхности земли поверхностные волны имеют значительную дисперсию скоростей, в отличие от объемных сейсмических волн.

Эти свойства поверхностных волн используются для их выявления и улучшения качества результирующего сейсмического материала. Если поверхностные волны достаточно сильные, то при традиционных однокомпонентных наблюдениях их начинают нейтрализовать путем группировки источников и приемников, несмотря на то, что такая процедура уменьшает горизонтальное и вертикальное разрешение записи, а также искажает ее динамические качества.

Наиболее популярные процедуры обработки данных:

- миграция сейсмограмм;
- коррекция изображений по наклону отражающихся границ (DMO — dip move out);
- коррекция статических поправок;
- AVO-анализ;
- инверсия сейсмограмм;
- вычитание кратных волн на основе рядов Борна;
- автоматическое регулирование усиления.

МЕТОД ПОДАВЛЕНИЯ ПОВЕРХНОСТНЫХ ВОЛН РЕЛЕЕВСКОГО ТИПА

Данный метод осуществляется в два этапа:

- 1) обнаружение помехи;
- 2) оценка помехи и ее вычитание.

Обычно оба этапа выполняются в скользящем вдоль записи пространственно-временном окне, формируя матрицу D объемом $N \times 3$, где N — размер скользящего окна, 3 — количество компонентов (X, Y, Z).

При выборе размера окна необходимо учитывать следующие условия:

— окно должно содержать только одну волну, что предполагает отсутствие интерференции;

— в пределах окна должно иметь место достаточно высокое соотношение сигнал–помеха (шум);

— окно должно быть достаточно длинным (статистически представленным), чтобы четко разделить регулярный компонент записи и случайную помеху (шум).

Текущая матрица D подвергается сингулярному разложению, позволяющему представить ее в виде суммы трех матриц единичного ранга, известных как собственные значения (eigen values), или главные компоненты:

$$D = \sum_{i=1}^3 E_i = \sum_{i=1}^3 \sigma_i u_i v_i^T, \quad (1)$$

где u_i, v_i — соответственно левые и правые сингулярные векторы, σ_i — сингулярные числа, T — операция транспонирования. Сингулярные числа предполагается упорядочить по убыванию: $\sigma_1 > \sigma_2 > \sigma_3$.

Для обнаружения участков, искаженных релеевскими волнами, в работе [6] использован параметр $e = (\sigma_1 - \sigma_3)(\sigma_2 - \sigma_3)$. С одной стороны, он увеличивается при одновременном росте двух первых сингулярных чисел, с другой — при относительном увеличении только второго сингулярного числа.

После определения параметра e во всех точках обрабатываемой сейсмограммы его значение упорядочивается по возрастанию. В результате дальнейшего анализа полученной кривой выбирается пороговое значение e_g , превышение которого параметром e или равенство ему расцениваются как признак наличия поверхностной волны в текущем окне. Порог e_g выбирается в точке, где кривая начинает резко возрастать, реагируя на наличие поверхностных волн. Иногда для получения масштабированного графика используют $\log_{10} e$. Для выявления участков поверхностных волн следует создать сейсмограмму, заполненную значениями e .

Для выбора порога e_g используется поляризационная фильтрация по формуле

$$F = (1 - g)D + g(D - E_1 - E_2), \quad (2)$$

где g — индикатор наличия поверхностных волн, если $e \geq e_g$, то $g = 1$, если $e < e_g$, то $g = 0$; D — исходные трехкомпонентные данные; E_1, E_2 — изображения, профильтрованные низкочастотным фильтром.

ОБОСНОВАНИЕ РАСПАРАЛЛЕЛИВАНИЯ И ВЫБОР ПЛАТФОРМЫ

Во многих случаях нецелесообразно приобретение, установка, содержание и обслуживание специальной компьютерной техники для ускорения всего нескольких программ. Интуитивная оценка потенциального выигрыша в производительности часто завышена. Согласно закону Амдала при идеальном распараллеливании доли вычислений η на n процессорных элементах (ПЭ) ускорение составит

$$S_n = \frac{1}{(1 - \eta) + \eta / n} \leq \frac{1}{1 - \eta}. \quad (3)$$

Ускорение вычисления в результате идеального распараллеливания 95% расчетов на 20 ПЭ составляет всего $1 / (0,05 + 0,95 / 20) = 10,26$. На практике эффективность параллельных вычислений оказывается еще ниже. При запуске на современном ПК, оснащенный двумя двухъядерными процессорами, четырех независимых экземпляров последовательной программы обработки волн релеевского типа в многокомпонентных сейсмических наблюдениях достигнуто ускорение всего в полтора раза по сравнению с последовательным выполнением. Для двух экземпляров программы выигрыш времени составил 25%. Распараллеливание также связано с накладными расходами на коммуникацию и синхронизацию, дополнительно замедляющими параллельное решение одной задачи по сравнению с решением нескольких независимых задач на том же вычислителе.

В 1988 г. Густафсон и Барсис определили, что во многих случаях время непараллельных вычислений программы мало зависит от объема ее параллельной части. В частности, это происходит, если непараллельную часть программы составляют только подготовка вычислений и обработка результатов. Тогда в качестве параметра удобно взять долю этой непараллельной части не в исходном, а в распараллеленном расчете:

$$\sigma_n = \frac{1 - \eta}{\eta / n + 1 - \eta}. \quad (4)$$

Преобразовав формулу (3) с учетом (4), Густафсон и Барсис получили следующее выражение:

$$S_n \leq n + (1-n)\sigma_n \Rightarrow \sigma_n \leq \frac{n - S_n}{n-1} \quad (5)$$

(закон Густафсона–Барсиса).

Формула (5) дает возможность оценить предельную долю последовательной части программы для достижения заданного ускорения. Например, для 19-кратного ускорения на 21 ПЭ доля последовательных расчетов на каждом ПЭ не должна превышать $(21-19)/(21-1)=10\%$.

Основной вывод из закона Густафсона–Барсиса: распараллеливание эффективно при длительном времени параллельных вычислений, т.е. для больших задач. Распараллеливание на многих процессорах эффективно только для задач сверхразмерности.

Для задач обработки данных сейсморазведки решающим фактором является чрезвычайно большие объемы и высокая коммерческая стоимость решения. В настоящее время типовой объем однокомпонентных 3D-сейсмограмм составляет 50–500 Gb на месторождение.

В целом кроссплатформная реализация в среде MPI обеспечивает оптимальное соотношение стоимость/эффективность для всех перечисленных задач.

АЛГОРИТМ ПАРАЛЛЕЛЬНОЙ ОБРАБОТКИ ПОВЕРХНОСТНЫХ ВОЛН РЕЛЕЕВСКОГО ТИПА В МНОГОКОМПОНЕНТНЫХ СЕЙСМИЧЕСКИХ НАБЛЮДЕНИЯХ

Алгоритм имеет следующую структуру.

1. Хост читает три компонента из разных файлов, распределяет некоторое количество трасс между процессами-клонами.

2. Процессы-клоны получают три компонента и объединяют их в одну матрицу.

3. Процессы-клоны осуществляют низкочастотную фильтрацию каждой трассы.

4. Процессы-клоны определяют места с наиболее выраженными признаками наличия поверхностных волн.

5. Процессы-клоны осуществляют вычитание из трассы участков с наиболее выраженными признаками наличия поверхностных волн и возвращают результат обработки хосту.

6. Хост записывает результат обработки в хранилище.

Для выполнения программы использовались две сейсмические записи. Каждая запись состоит из трех компонентов, которые записаны в отдельных файлах. Длина трассы — 3 000 элементов. В первой записи количество трасс — 664, во второй — 86 984. Размер первой записи — 243 Mb, второй — 3 Gb.

На рис 1, 2 в виде диаграмм приведены результаты

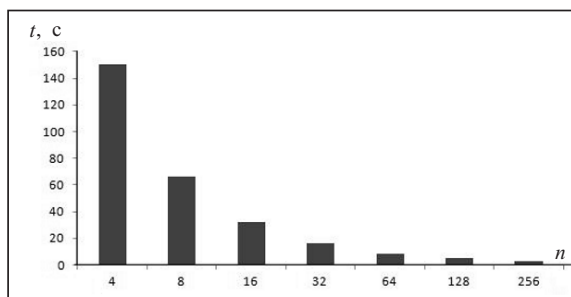


Рис 1. Диаграмма времени выполнения программы на первой сейсмической записи

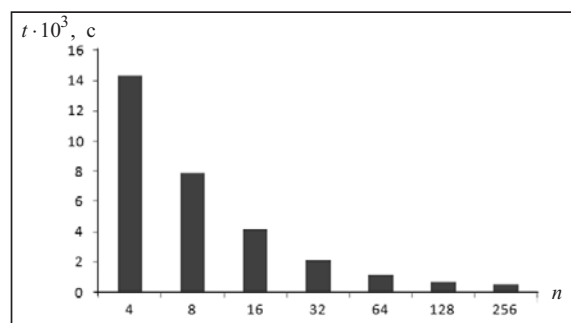


Рис 2. Диаграмма времени выполнения программы на второй сейсмической записи

выполнения алгоритма на разных сейсмических записях средствами кластерного комплекса Инпарк-256 в среде MPI. С увеличением числа n процессоров уменьшается время $t(c)$ выполнения программы.

Отметим, что время выполнения после запуска четырех копий последовательной программы с разными наборами данных на ПК с двухъядерным процессором вычисления в полтора раза меньше, чем для параллельного варианта той же программы на четырех процессорах кластера. Такой эффект объясняется реализацией многопоточности в современных операционных системах.

ЗАКЛЮЧЕНИЕ

В работе описана реализация параллельного алгоритма обработки данных поверхностных волн релеевского типа в многокомпонентных сейсмических наблюдениях. Результаты экспериментов показали, что при увеличении количества процессоров время выполнения программы уменьшается. Таким образом, существует возможность обработки сейсмических наблюдений большего объема за более короткий период.

При разработке программы для ускорения выполнения задач за счет уменьшения нагрузки на файловую систему использованы:

- локальные папки для временных файлов отдельных процессов;
- индексные файлы для ускорения навигации в больших файлах.

СПИСОК ЛИТЕРАТУРЫ

1. Marmalevskiy N., Roganov Y., Kostyukevych F., Roganov V. Duplex wave migration and interferometry for imaging onshore data without angle limitations // 70th EAGE Conf. & Exhibition (Rome, Italy, 9–12 June, 2008). — N P273 (<http://www.tesseral-geo.com/publications/Marmalevskiy%20-%20Duplex%20Wave%20Migration%20and%20Interferometry.zip>).
2. Шериф Р., Гелдарт Л. Сейсморазведка. — М.: Мир, 1987. — 448 с.
3. Сейсморазведка. Справочник геофизика / Под ред. В.П. Номоконова. — М.: Недра, 1990. — Кн. 1. — 336 с.
4. Гольдин С.В. Линейные преобразования сейсмических сигналов. — М.: Недра, 1974. — 352 с.
5. Miller C.F., Pursey H. The field and radiation impedance of mechanical radiators on the free surface of a semi-infinite isotropic solid // Proc. Roy. Soc. A. — 1954. — **233**, N 3. — P. 521–541.
6. Jin S., Ronen S. Ground roll detection and attenuation by 3C Polarization analysis // 67th EAGE Conf: Extend Abstr. — 2005. — Paper B020.
7. Тяпкин Ю.К., Тяпкина Е.Ю. Поляризация фильтрация поверхностных волн-помех на многокомпонентных сейсмических записях. Одноканальный метод // Геофиз. журн. — 2008. — **30**, № 3. — С. 14–26.
8. Тульчинский В.Г., Тульчинский П.Г. Применение суперкомпьютеров СКИТ для разработки и выполнения параллельных геофизических программ // Компьютерная математика. — 2009. — № 5. — С. 10–17.

Поступила 17.05.2010