

Аннотация:

В рамках работ были проведены численные эксперименты в программном пакете обработки и интерпретации микросейсмических данных FracLoc с использованием оборудования Информационно-вычислительного центра Новосибирского государственного университета. В процессе работы были выявлены наиболее ресурсоемкие модули, проведены работы по оптимизации вычислительных алгоритмов в модулях FL Adapt, FL ATime, а также разработана многопоточная версия вычислительного ядра модуля FL Syn 3D.

Тема работы:

Разработка программно-аппаратного комплекса для мониторинга подземных газовых хранилищ

Состав коллектива:

Кречетов Антон Вячеславович, сотрудник компании Акционерное Общество «Технологии обратных задач», директор подразделения в г. Новосибирск, krechetov.av@tinp.ru, руководитель.

Гапеев Денис Николаевич, сотрудник компании Акционерное Общество «Технологии обратных задач», заместитель директор по научной работе, DG@tinp.ru, руководитель.

Корнеев Владимир Иванович, сотрудник компании Акционерное Общество «Технологии обратных задач», консультант.

Артюхов Алексей Андреевич, сотрудник компании Акционерное Общество «Технологии обратных задач», младший инженер-программист, программист. Студент НГУ: кафедра параллельных вычислений ФИТ, 3-ий курс 15202, работа не является курсовой/дипломной.

Авдеев Александр Васильевич, заведующий лабораторией высокопроизводительных вычислительных систем Факультета информационных технологий Новосибирского государственного университета.

Сергей Владимирович Гололобов, кандидат физико-математических наук, ассистент, доцент, старший преподаватель Новосибирского государственного университета.

Научное содержание работы:

Постановка задачи.

Современные алгоритмы и методики обработки геофизических материалов включают в себя процедуры, которые требуют значительных вычислительных затрат. Планируется провести ряд тестов для определения необходимых вычислительных ресурсов достаточных для потоковой обработки данных микросейсмического подземных газовых хранилищ (ПХГ). На настоящем этапе работ из задачи разработки макета программно-аппаратного комплекса постоянно действующего мониторинга ПХГ, в части микросейсмического мониторинга была выделена актуальная подзадача адаптации имеющегося программного обеспечения для потоковой обработки большого объема геофизических данных.

Подробное описание работы, включая используемые алгоритмы.

Численные эксперименты проводилась в программном пакете обработки и интерпретации микросейсмических данных FracLoc. В процессе работы были выявлены наиболее ресурсоемкие модули:

1. *FL Adapt* – модуль фильтрации техногенных квазигармонических помех.

2. *FL Impulse Filter* – модуль фильтрации локальных по времени и пространству помех.
3. *FL ATime* – модуль выделения микросейсмических событий.
4. *FL Syn 3D* – модуль решения задачи распространения лучей в сейсмической среде.

Были проведены работы по оптимизации вычислительных алгоритмов в модулях *FL Adapt*, *FL ATime*, а также разработана многопоточная версия вычислительного ядра модуля *FL Syn 3D*.

Модуль *FL Adapt* предназначен для фильтрации микросейсмических записи от техногенных квазигармонических помех. Алгоритм обрабатывает трассы независимо друг от друга, при этом каждая трасса разбивается на перекрывающиеся части данных одинаковой длины, к которым применяются сглаживающие окна (например, окна Ханна, Хемминга, Бартлетта или Блэкмана—Наталла). Для каждого участка данных решается задача определения амплитуды и фазы помехи:

$$\min_{A, \phi_0} \|f(t) - A \cdot \sin(\omega t - \phi_0)\|.$$

Получив параметры помехи в каждом отрезке данных, строится общая для трассы квазигармоническая помеха и вычитается из первичных данных.

Анализ алгоритма вычисления квазигармонических помех, проведенных при помощи процедур профилирования программ, позволил выявить множественное вычисление стандартных функций вида $\sin(nx)$, $\cos(nx)$, где n – целое число. Применяв формулы синус/косинус суммы, был реализован и внедрен рекурсивных алгоритм расчета значений $\sin(nx)$, $\cos(nx)$ через ранее посчитанные значения $\sin(x)$, $\cos(x)$, $\sin((n-1)x)$, $\cos((n-1)x)$. Кроме того, анализ алгоритма позволил выявить возможность однократного расчета сглаживающих окон и использования сохраненных значений в последующих вычислениях.

Модуль *FL ATime* предназначен для выделения микросейсмических событий. Алгоритм основан на многократных решениях прямой задачи (определение времен вступлений и амплитуд прямой волны).

Вычисления в модуле *ATime* делятся на две стадии: стадия расчёта функции, по значениям которой принимается решение о локации микросейсмического события, и стадия выделения (определения параметров) событий.

Каждое значение определяющей функции соответствует предполагаемому микросейсмическому событию, и чем больше значение – тем больше вероятность, что событие достоверно. Определив параметры события, можно вычесть волновой отклик данного события из микросейсмических данных, что позволяет компенсировать влияние микросейсма. Процедура компенсации изменяет определяющую функцию только в области влияния микросейсмического события. Следовательно, после компенсации микросейсма необходим пересчет определяющей функции в области его влияния.

На второй стадии происходит поиск максимумов определяющей функции, компенсация микросейсмиков и обновление определяющей функции.

Анализ алгоритма позволил выявить повторные вычисления прямой задачи непосредственно во время выделения событий, что приводило к избыточным расчетам. В целях оптимизации, было принято решение о выделении расчёта прямой задачи в отдельный модуль *Prepare 1D* и сохранение результатов моделирования в виде полиномиальных аппроксимаций, а также реализация в *ATime* работы с аппроксимирующими функциями.

В *Prepare 1D* аппроксимация выполняется на трёхмерной сетке. Сетка строится из отдельных параллелепипедов, каждый из которых представляет собой отдельный слой в

одномерной сейсмической модели. Далее, в процессе аппроксимации, вычисляется разность между значением функции и значением текущей аппроксимации для этой функции на текущей сетке. Если точность аппроксимации на каком-либо элементе сетки не соответствует необходимому значению, то происходит его дальнейшее адаптивное дробление. Аппроксимация продолжается до тех пор, пока в каждом элементе сетки не будет получена нужная точность.

Модуль *FL Syn 3D* предназначен для решения задачи распространения лучей в сейсмической среде. Алгоритм основан на решении трёхмерного уравнения Эйконала методом Fast Iterative Method (FIM).

Анализ алгоритма позволил выявить возможности создания параллельной версии программы, что и было выполнено в процессе работ.

Модуль *FL Impulse Filter* предназначен для фильтрации локальных по времени и пространству помех. Алгоритм модуля состоит из следующих шагов:

1. Расчет оконного преобразования Фурье, амплитудного и фазового спектра.
2. Расчет медианы найденных спектров, определение порогового значения.
3. Уменьшение anomalно больших спектральных значений. Если значение на амплитудном спектре больше чем пороговое, то значение уменьшается до порогового.
4. Расчет обратного оконного преобразования Фурье.

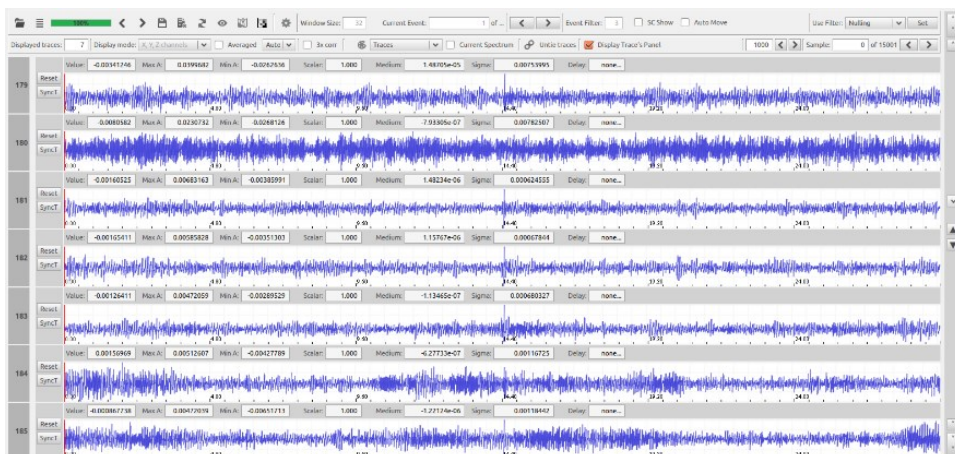
Работы по данному модулю были отнесены к следующему этапу оптимизации алгоритмов для вычислений на GPU (2019 г.).

Полученные результаты.

FL Adapt

Внедрение рекурсивного алгоритма расчета значений $\sin(nx)$, $\cos(nx)$ через ранее посчитанные значения $\sin(x)$, $\cos(x)$, $\sin((n-1)x)$, $\cos((n-1)x)$ позволило ускорить вычисления в 3.9 раза, а реализация однократного расчета сглаживающих окон и использования сохраненных значений в последующих вычислениях ускорило работу модуля в 2.4 раза дополнительно. Общий выигрыш производительности составил порядок.

Пример применения фильтра представлен на рис. 1 (длина окна 2 секунды, удалено 50 квазигармонических помех).



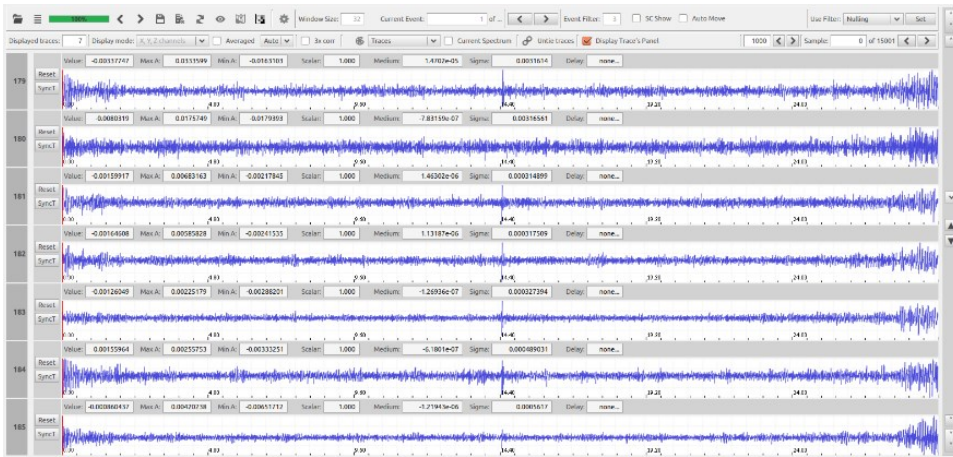


Рис. 1 – Данные до фильтрации (сверху) и данные после фильтрации (снизу)

FL Syn 3D

Разработка параллельной версии программы проводилась с использованием средств OpenMP. На рис. 2 представлен график роста производительности в зависимости от количества ядер. Реальный выигрыш в производительности параллельной версии расчетного модуля не достиг теоретических оценок.

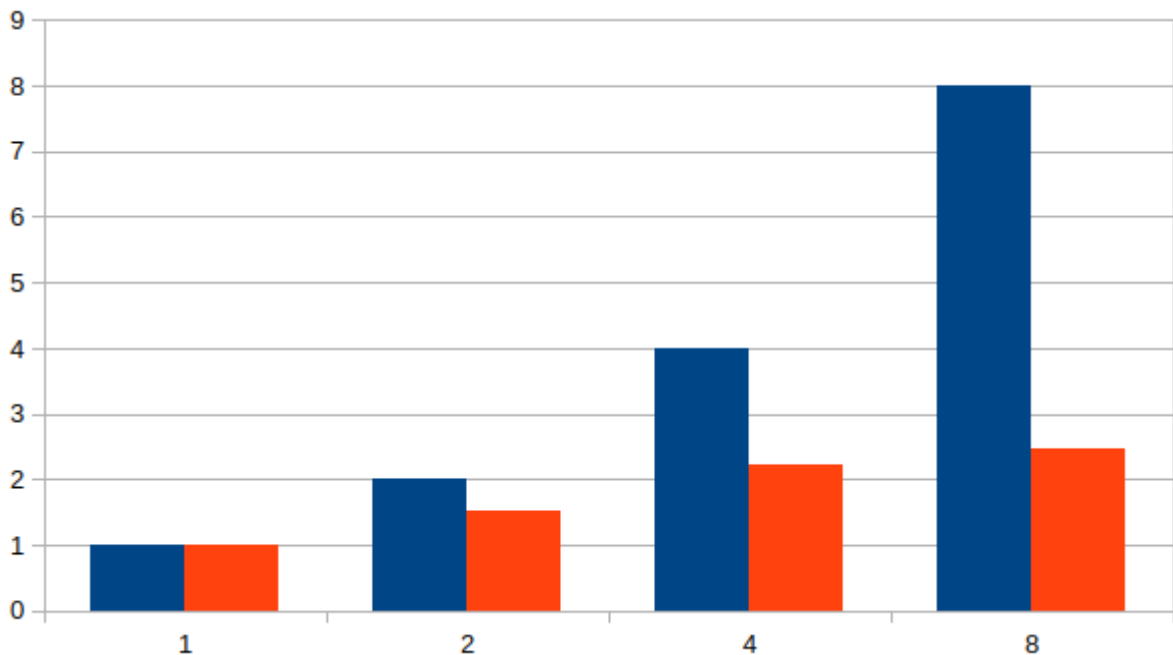


Рис. 2 – Рост производительности в зависимости от количества ядер (синий – теоретический, красный – фактический)

В рамках работ были проведены тесты на точность вычисления алгоритмов, используемых при моделировании распространения сейсмических волн. Точность расчета оценивалась в сравнении с одномерной задачей моделирования.

Численные эксперименты проводились на наклонной горизонтально-слоистой модели среды с размещением источника на глубине 250 м и областью расчетов 1000x1000x500 м. Слои имеют наклон 20 град по ординате и поворот -40 град по азимуту (рис. 3). В каждом слое задана своя скорость. Расчёт начинался с шагом в 15 м по всем направлениям и далее

уменьшался до 1 м. Решение фиксировалось в приёмниках, расположенных на поверхности (рис. 3).

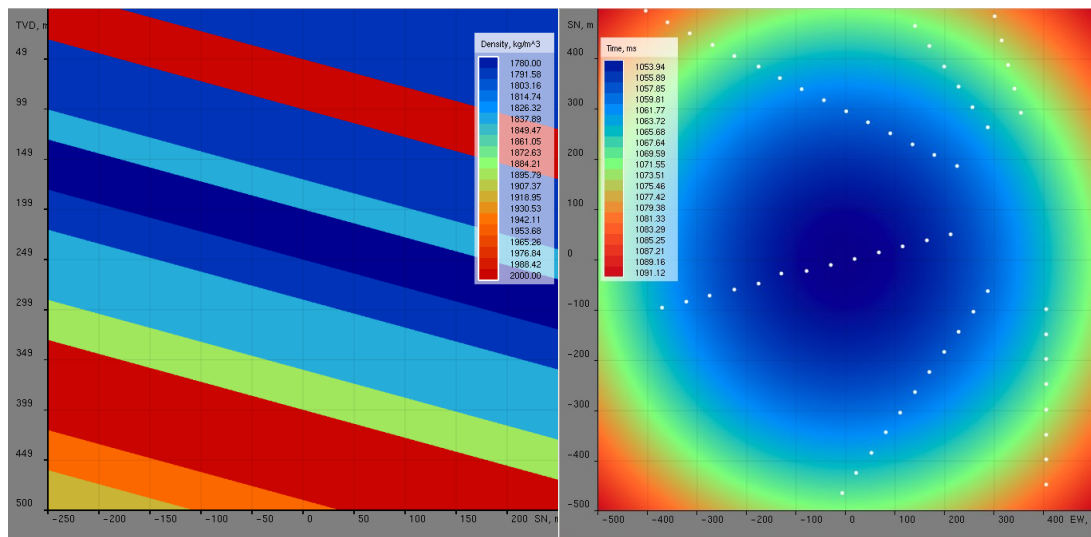


Рис. 3 – Наклонная горизонтально-слоистая модель (слева) и микросейсмическая антенна (справа).

Для оценки погрешности в модуле расчета распространения лучей в сейсмической среде используется подход, основанный на решении серии задач на сетках с разным пространственным шагом. Как видно на рис. 4, красная кривая (разность одномерного и трехмерного решателя) хорошо согласуется с желтой кривой – оценкой погрешности, построенной только по данным трехмерного решателя. Таким образом, по достижению необходимой точности расчетов модуль останавливает расчет и выдает полученный результат.

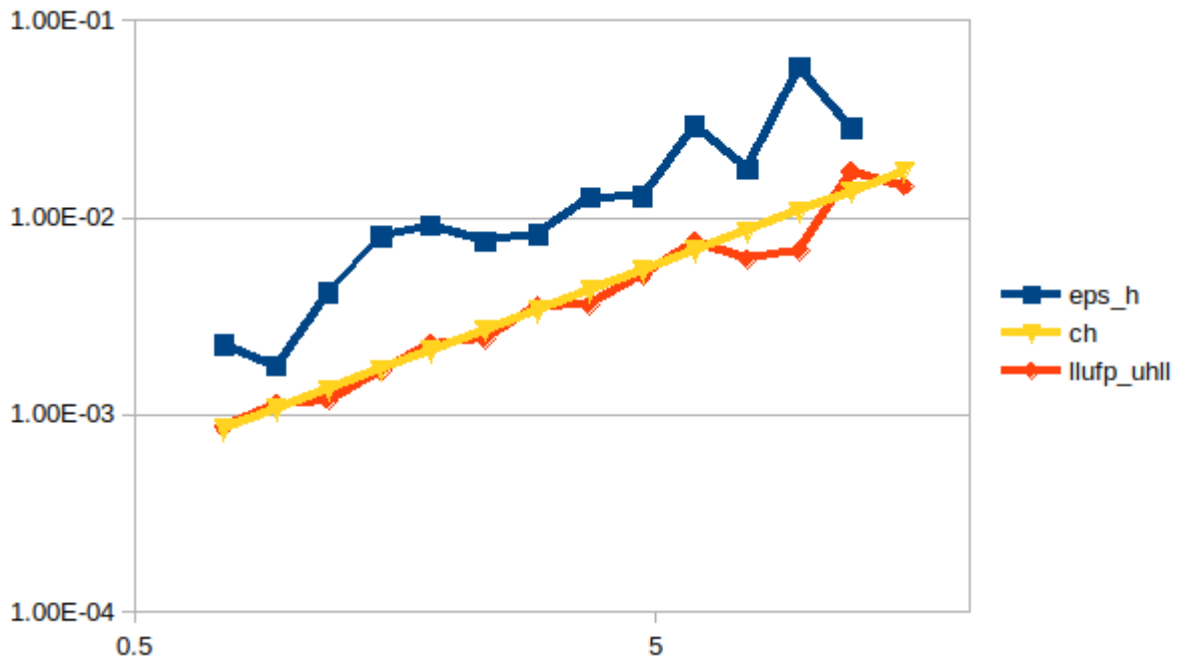


Рис. 4 – Точность расчета времен вступления в зависимости от шага сетка. Красная кривая – разница между одномерным и трехмерным решателем, желтая – оценка погрешности трехмерного решателя, построенная по 14 экспериментам, синяя – оценка погрешности (по 2 экспериментам).

FL Prepare 1D

Процесс аппроксимации сигнала для определённой антенны является ресурсоемким, из-за того, что для каждого приёмника в антенне необходимо рассчитывать свою аппроксимирующую функцию (приёмников может быть несколько сотен/тысяч). Также время расчетов сильно зависит от размеров области поиска микросейсмических событий и необходимой точности аппроксимации времен вступлений и амплитудных параметров. Вычисления аппроксимирующих функций для разных приёмников осуществляются параллельно на узлах вычислителя. Т.к. в процессе расчёта нет необходимости в дополнительных обменах данными между узлами, и задачи сильно различаются по времени расчетов, то выбранный нами подход эффективен в случае динамической балансировки загрузки вычислительных узлов.

Разработана динамическая балансировка нагрузки между узлами, выполняющими аппроксимацию, для этого среди вычислительных узлов выделяется один – управляющий, который распределяет задачи между узлами, отправляя им новые области, в которых необходимо производить вычисления. Благодаря такому подходу удалось снизить время неполной загрузки вычислительных узлов (ожидание решения последней задачи) с 90 до 20 минут на реальных данных.

FL ATime

Реализация работы с аппроксимированными функциями позволила сократить время расчета первой стадии (вычисление определяющей функции) для тестовой задачи со 140 минут до 40 минут, т.е. в 3.5 раза. Оптимизация использования средств MPI и OpenMP для второй стадии расчетов (выделение событий) позволила сократить время выделения одного события на 25%.

Эффект от использования кластера в достижении целей работы:

На настоящем этапе работ из задачи разработки макета программно-аппаратного комплекса постоянно действующего мониторинга ПХГ, в части микросейсмического мониторинга была выделена актуальная подзадача адаптации имеющегося программного обеспечения для потоковой обработки большого объема геофизических данных. Ключевым параметром для оценки результатов адаптации программного обеспечения была избрана оценка скорости обработки специально подготовленных блоков данных наиболее ресурсоемкими модулями обрабатывающего пакета реализованных с применением технологий и подходов параллельных вычислений до и после оптимизации логической и вычислительной частей модулей. Проведение серии экспериментов стало возможным исключительно с применением оборудования и при активной поддержке специалистов Информационно-вычислительного центра Новосибирского государственного университета.

Коллектив исследователей на предстоящий 2019 год ставит перед собой задачи оценки возможности адаптации программного обеспечения в разрезе подхода постоянно действующего мониторинга с применением вычислителей на GPU. Успешная оптимизация вычислительных блоков при условии сокращения времени обработки данных в перспективе позволит вплотную подойти к возможностям по сокращению массогабаритных параметров полевого программно-аппаратного комплекса постоянно действующего мониторинга ПХГ.