

DOI: 10.55648/1998-6920-2024-18-2-57-68

УДК  
621.3.049.779

# Принципы организации программно-аналитической системы для параллельной обработки сейсмических данных

А. Ю. Выродов, В. А. Перепёлкин, М. С. Хайретдинов, А. В. Хрыпченко

Институт вычислительной математики и математической геофизики СО РАН

*Аннотация:* Прогресс в развитии современных информационных технологий непосредственно связан с применением ресурсоемких приложений в наукоемких исследованиях, а также в промышленных прикладных задачах. В настоящее время остро стоит проблема анализа больших объемов геофизических данных и повышения производительности систем для их исследования. Один из путей решения данной проблемы заключается в применении многопроцессорных ЭВМ и многомашинных вычислительных комплексов, способных производить параллельную, в том числе распределенную, обработку данных. В работе представлены описание и реализация вычислительной модели для параллельной обработки сейсмических данных на базе системы LuNA для автоматического конструирования параллельных программ.

*Ключевые слова:* сеймотрасса, быстрое преобразование Фурье, взаимокорреляционная функция, параллельная обработка, система LuNA, активные знания, сейсмические данные.

*Для цитирования:* Выродов А. Ю., Перепёлкин В. А., Хайретдинов М. С., Хрыпченко А. В. Принципы организации программно-аналитической системы для параллельной обработки сейсмических данных // Вестник СибГУТИ. 2024. Т. 18, № 2. С. 57–68.  
<https://doi.org/10.55648/1998-6920-2024-18-2-57-68>.



Контент доступен под лицензией  
Creative Commons Attribution 4.0  
License

© Выродов А. Ю., Перепёлкин В. А.,  
Хайретдинов М. С., Хрыпченко А. В., 2024

Статья поступила в редакцию 11.12.2023;  
принята к публикации 22.01.2024.

## 1. Введение

Построение эффективной системы управления вычислительными ресурсами современных систем обработки геофизических данных требует разработки новых методов, не зависящих от размеров системы и позволяющих автоматизировать процессы управления с учетом текущего состояния системы в целом и ее отдельных компонентов. Предметной областью является разработка системы. Для этого нужно описать сценарии и частичный порядок действий, связанных с организацией вычислительного процесса, специфицировать характеристики объектов данных, подаваемых на вход вычислительным программам, и результирующих объектов данных.

Теория синтеза параллельных программ и систем на вычислительных моделях [1], получившая развитие в концепции активных знаний [2], предлагает подход к автоматизации конструирования параллельных программ на основе формального описания предметной области,

в которой проводится моделирование, включая описание накопленных в этой области программных модулей. В качестве базового математического аппарата для такой формализации используются вычислительные модели. Это позволяет, с одной стороны, автоматически переиспользовать накопленные программные модули для решения новых задач в предметной области и, с другой стороны, автоматически решать формально поставленные в предметной области задачи с учетом предъявляемых нефункциональных требований [3]. В частности, в настоящей работе основным нефункциональным требованием является ускорение вычислений за счёт параллельной работы имеющихся модулей обработки сейсмических сигналов. В работе используется поддерживающая указанный подход система LuNA [4] для автоматического конструирования параллельной программы.

## 2. Требования, предъявляемые к системе

В техническом задании к разрабатываемой системе определяются:

- требования к системе в целом;
- требования к функциям (задачам), выполняемым системой;
- требования к видам обеспечения [5].

Примером реализации параллельной обработки данных является организация многоканальной свертки сейсмических сигналов с числом каналов до 100. Задачей представленной работы является организация программных средств многоканальной параллельной свертки, исследование производительности и визуализация результатов обработки. За основу реализации выбран язык Python и система LuNA. Для этого используется взаимокорреляционная функция, именуемая дальше ВКФ, и быстрое преобразование Фурье, именуемое дальше БПФ. Стоит также ввести термин «секционирование» – это процесс разбиения на отрезки во временном сигнале, необходимый для вычисления свертки в реальном времени.

Подробно алгоритм взаимокорреляционной функции описывается в книге Бендат Дж., Пирсол А. «Измерение и анализ случайных процессов» [11].

Расчет ВКФ с секционированием:

$$\hat{R}_{xy}(rh) = \sum_{n=1}^{511} x_n y_{n+r} + \sum_{n=512}^{1023} x_n y_{n+r} + \dots + \sum_{n=512(M-1)}^{512M-1} x_n y_{n+r}, \quad (1)$$

где  $M$  – число секций. Размер секции – 512 элементов. Время расчета взаимокорреляционной функции:

$$T = T_0 \cdot m \cdot N_{оп}, \quad (2)$$

$T_0$  – время выполнения элементарной операции,  $N_{оп}$  – число отсчетов опорного сигнала [6].

В результате алгоритм должен быть реализован в параллельном виде в многоканальном режиме.

Задача реализации на практике имеет временные требования, обуславливающие необходимость применения параллельных вычислений: согласно техническим требованиям время вычислений не должно превышать 1 с.

Вычисление этой операции в реальном масштабе времени, т.е. в темпе поступления данных при априорно неизвестных временах прихода сейсмических и акустических колебаний, возможно лишь путем секционирования длинных временных последовательностей. К тому же использование алгоритма быстрого преобразования Фурье (БПФ) для обработки секционированных данных обеспечивает возможности высокопроизводительных вычислений.

Пусть  $Y$  – искомая коррелотрасса. При секционировании оценка ее вычисляется в итерационном виде:

$$Y^{i+1} = f(Y^i, X_{i+1}), \quad (3)$$

где  $Y^i$  – приближение  $Y$  на  $i$ -м шаге,  $X_{i+1}$  – текущие отсчеты данных.

Секционированные входные данные по мере поступления накапливаются во вспомогательных буферах и после заполнения очередного буфера обрабатываются при помощи БПФ. Таким образом, для того чтобы иметь возможность обрабатывать сигналы в режиме реального времени с использованием БПФ, надо представить аperiodическую корреляционную свертку двух массивов большого размера как сумму циклических обычных сверток частей (секций) этих массивов. Причем это разбиение надо сделать таким образом, чтобы трудоемкость обработки одной секции зависела только от величины этой секции и не зависела от множества всех входных отсчетов размера  $N$ . Это аналогично тому, что имеет место в цифровой фильтрации, когда размер опорной последовательности, представляющей дискретный аналог импульсной функции фильтра, несоизмеримо меньше размеров входных и выходных массивов, которые можно считать бесконечными. Напротив, в задачах обработки вибросейсмических сигналов входной  $X$  и опорный  $S$  массивы имеют одинаково большие размерности, а результат их свертки  $Y$  представляется сравнительно небольшим числом отсчетов  $M$  ( $M \ll N$ ). Это обстоятельство лежит в основе выбора подхода к разбиению массивов  $X$  и  $S$ . С учетом этого алгоритм свертки обоих массивов может быть представлен в виде

$$Y_m = \sum_{n=0}^{N-1} X_n \cdot S_{n-m} = \sum_{l=0}^{\frac{N}{M}-1} \sum_{n=0}^{M-1} X_{M \cdot l + n} \cdot S_{M \cdot l + n - m} = \sum_{l=0}^{\frac{N}{M}-1} \sum_{n=0}^{2L-1} A_n^l \cdot B_{\langle m-n \rangle_{2M}}^l, \quad (4)$$

где  $M$  – размер одной секции входного массива (подразумевается, что  $N$  делится на  $M$  нацело, в противном случае массив расширяется путем дописывания нулей).  $A$  и  $B$  – массивы, состоящие из  $2L$  отсчетов и определяемые следующим образом:

$$A_i^l = \begin{cases} X_{m \cdot l + i}, & \text{если } 0 \leq i \leq M \\ 0, & \text{если } M \leq i \leq 2M \end{cases}, \quad (5)$$

$$B_i^l = S_{M \cdot (l-1) + \langle M-i \rangle_{2M}}, \quad (6)$$

$$0 \leq i \leq 2M. \quad (7)$$

В соответствии с теоремой о свертке для дискретного случая циклическая свертка двух массивов равна обратному дискретному преобразованию Фурье (ОДПФ) от произведения прямых ДПФ этих массивов, следовательно:

$$Y_m = \frac{1}{N} \sum_{l=0}^{\frac{N}{M}-1} \sum_{k=0}^{2M-1} \left( \sum_{n=0}^{2M-1} A_n^l \cdot W^{-n \cdot k} \right) \cdot \left( \sum_{l=0}^{2M-1} B_n^l \cdot W^{-l \cdot k} \right) \cdot W^{k \cdot m}, \quad (8)$$

где  $W = e^{j \cdot \frac{2\pi}{2M}}$

На рис. 1 приведена схема, иллюстрирующая порядок действий, производимых при обработке  $i$ -ой секции входной последовательности  $X$ . Значок \* означает БПФ от массивов и умножение спектра массива  $A$  на спектр, комплексно сопряженный со спектром массива  $B$ . Такие действия производятся для каждой секции входной последовательности. По завершении работы вычисляется коррелограмма путем выполнения обратного БПФ от спектра  $Y$ .

Трудоемкость используемого алгоритма БПФ (двухточечного, рассчитанного на вещественные данные) составляет  $\frac{N}{2} \cdot (3 \cdot \log_2 N - 5) + 4$  операций сложения и  $N \cdot (\log_2 N - 3) + 4$  операций умножения, т.е. в сумме:  $\frac{N}{2} \cdot (5 \cdot \log_2 N - 11) + 8$ .

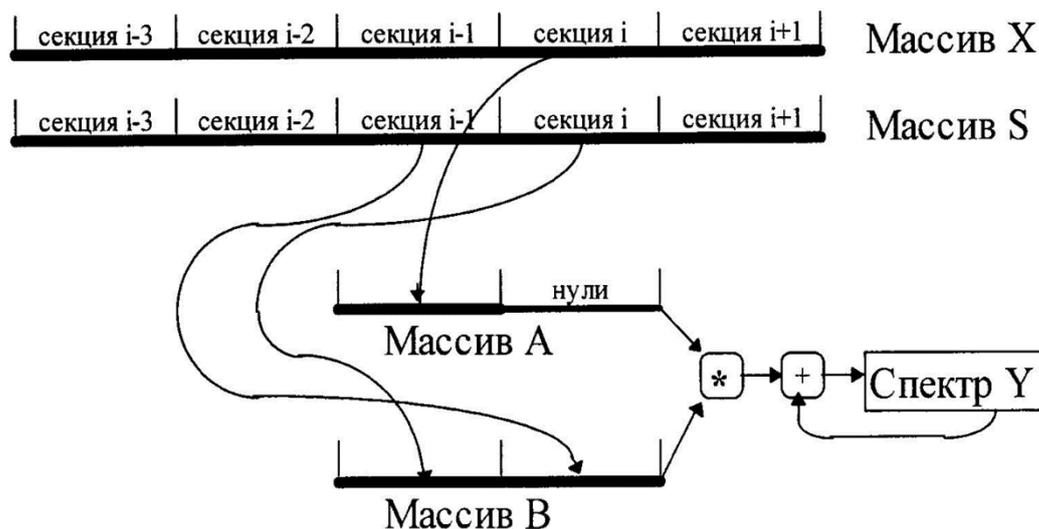


Рис. 1. Схема порядка действий, производимых при обработке  $i$ -ой секции входной последовательности  $X$

### 3. Характеристики формата представления данных

Данные каждого сеанса представлены в виде отдельных файлов трасс. Каждый файл состоит из заголовка фиксированной длины и последовательно размещенных отсчетов одного канала.

Таблица 1. Представления файла коррелотрассы

Заголовок ( $N$ байт)
Отсчет 1 ( $M$ байт)
Отсчет 2 ( $M$ байт)
...
Отсчет $K$ ( $M$ байт)

Таблица 2. Представления заголовка файла формата РС

Смещ.	Длина	Тип и имя	Значение	Комментарии
+00	2	int ID	ID='PC'	Идентификатор формата
+02	1	char Ver	Ver=1	Номер версии
+03	1	char SubVer	SubVer=1	Номер подверсии
+04	2	int HLen	Hlen=42	Длина заголовка в байтах
+06	4	long Datetime		Дата и примерное время начала в секундах с 1 января 1970 г.
+10	4	long Duration		Длительность сеанса в секундах
+14	8	double Scale		Коэффициент пересчета в физическую величину
+22	1	char Year		Год начала сеанса

Смещ.	Длина	Тип и имя	Значение	Комментарии
+23	1	char Month		Месяц начала сеанса
+24	1	char Day		День начала сеанса
+25	1	char Hour		Час начала сеанса
+26	1	char Minute		Минута начала сеанса
+27	1	char Second		Секунда начала сеанса
+28	4	long MicroSec		Поправка в микросекундах к времени начала сеанса
+32	2	int SamplRate		Частота дискретизации в Гц
+34	4	long SamplNum		Число отсчетов в трассе
+38	2	int SamplType		Длина и тип отсчета: 0002h – int – знаковое 16 бит 0102h – unsigned – беззнаковое 0004h – long – двойное 32 бит 1004h – float – плавающая точка 1008h – double – двойное
+40	1	char TrNum		Номер трассы в исходном файле
+41	1	char Reserved	00	Зарезервировано, д.б. ноль

Формат отсчета задан полем SamplType. Типы int, unsigned, long, float и double подчиняются стандартам типов, принятым в Borland C++, v.3.1. Порядок байт – младший байт по младшему адресу. Стандарт чисел с плавающей точкой – IEEE. Стандарт отрицательных целых чисел – обратный дополнительный код [7].

Используемые форматы сейсмической записи – Seg-D, Seg-Y, Seg-B, Seg-A, Cray Seg-Y, CGG Geovector, PC.

#### 4. Исходные данные экспериментов

Используются данные в виде записей сейсмических колебаний от вибрационного источника ЦВ-40 на удалении 50 км. Амплитуда возмущающей силы источника составляет 40 тс. Вибрационный источник ЦВ-40 излучает одновременно длительные сейсмические колебания в земле и акустические – в атмосфере.

Время начала сеансов излучения составляет 15, 17, 21 и 23 часа (по Гринвичу), полосы частот излучения  $\Delta f$  – 6.25–9.57 Гц, длительность  $T = 2850$  с. [6].

Полученные записи  $X(t)$  сворачиваются с опорным сигналом  $Y(t)$  в соответствии с алгоритмом (1) с применением процедуры БПФ. Опорный сигнал синтезируется в точке приема.

В рамках решаемой задачи подготовлены тестовые данные в количестве 100 сеймотрасс и опорного сигнала. Размер файла сейсмограммы – 1.34 Мб. На рис. 2 представлен эксперимент с последовательной обработкой сеймотрасс для нахождения взаимокорреляционной свертки. В результате получаем время обработки более 4 с.

```
16_11011500_14.U8831Vk02 - OK A=9.38 SNR=27.0
16_11011500_15.U8831Vk00 - OK A=19.64 SNR=28.0
16_11011500_15.U8831Vk01 - OK A=17.48 SNR=21.6
16_11011500_15.U8831Vk02 - OK A=9.38 SNR=27.0
16_11011500_16.U8831Vk00 - OK A=19.64 SNR=28.0
16_11011500_16.U8831Vk01 - OK A=17.48 SNR=21.6
16_11011500_16.U8831Vk02 - OK A=9.38 SNR=27.0
processing time: 4.14s
```

Рис. 2. Время обработки 100 сейсмотрасс

Чтобы уменьшить временные затраты на обработку сигналов, решено применить параллельные вычисления, используя систему LuNA для автоматического конструирования параллельной программы.

## 5. Принцип параллелизма в системе LuNA

Проанализируем задачу с точки зрения параллельной обработки данных. Обработка сейсмотрасс для каждого канала может осуществляться независимо, в т.ч. параллельно. В идеале для  $N$  каналов это даёт возможность ускорить вычисления за счёт параллельной обработки в  $N$  раз. На практике эта величина будет меньше вследствие отсутствия достаточного количества вычислительных ядер для одновременной обработки всех каналов, одновременного доступа параллельных процессов обработки к общей файловой системе (локальной или сетевой), а также накладных расходов операционной системы на создание требуемого количества процессов. Предварительное исследование производительности последовательной программы показало, что для удовлетворения требования по времени обработки в 1 секунду требуется достичь ускорения в несколько раз (около 5 раз). Таким образом, наличие около 100 каналов создаёт достаточное количество задач для достижения требуемой производительности. В качестве аппаратного обеспечения может быть использован параллельный вычислитель с общей или распределенной памятью с 5 и более ядрами.

В качестве подхода к реализации параллельной обработки сейсмотрасс предлагается применить подход активных знаний [8]. Этот подход целесообразно применять в случаях, когда, во-первых, некоторый программный модуль используется многократно для решения различных задач и, во-вторых, когда одна и та же задача должна решаться в различных условиях – на вычислителях с разной конфигурацией, для разных классов входных данных, с разными нефункциональными требованиями (по времени выполнения, по энергоэффективности, по надёжности и т.п.). Ручное программирование потребовало бы переработки программы под изменяющиеся условия, поэтому целесообразно применение методов автоматизации конструирования программ.

Применение системы LuNA для автоматического конструирования параллельной программы на основе концепции активных знаний подразумевает формальное описание предметной области в виде вычислительной модели [9]. По существу, это означает описание множества переменных – величин предметной области, а также множества операций, задающих возможность вычислять одни величины из других. Каждой операции ставится в соответствие вычислительный модуль (например, последовательная программа или процедура) с указанием его нефункциональных свойств. Далее ставится задача в виде перечня входных и выходных переменных, а система автоматически конструирует требуемую программу, выводя частично упорядоченное множество операций, обеспечивающих вычисление выходных переменных из входных.

Применительно к исследуемой задаче вычислительная модель будет содержать множество из 100 информационно независимых операций. По этому описанию система LuNA автоматически построит параллельную программу, обеспечивающую параллельную обработку данных.

## 6. Вычислительная модель проектируемой системы

Для того чтобы использовать систему LuNA, необходимо было описать вычислительную модель, представленную в виде двудольного графа операций и переменных.

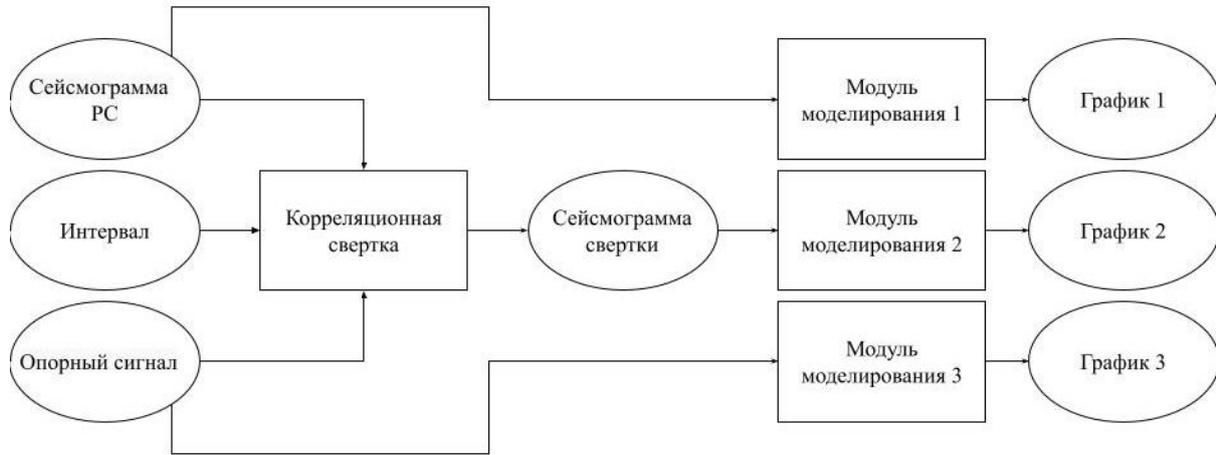


Рис. 3. Сценарий вычислительной модели

На рис. 3 представлен сценарий вычислительной модели, описывающий нахождение корреляционной свертки и построение графика. Овалами обозначены данные, прямоугольниками – операции.

Совокупность сценариев описывается вычислительной моделью и позволяет пользователям, задавая входные и выходные данные, получать результат, избавляя пользователя от рутинной, не связанной с сутью решаемых задач [10].

Таким образом, описав вычислительную модель, получаем файл, представленный ниже (код содержит незначительные упрощения). Файл содержит описание всех входных и выходных переменных, а также операций, производимых над ними.

```

#!/usr/bin/env cm
{
"variables": ["files", "opora", "path", "t1", "t2", "f", "step", "f1", "f2", "duration", "spec_one_output", "get_file_output",
"spec_one_output", "corr_output", "draw_output", "draw_plot_output", "RHO_output", "FMax_output"],
"operations": {
"spec": {
"inputs": ["get_file_output"],
"outputs": ["spec_one_output"],
"type": "external_cmd",
"external_cmd": "python ./Python/spec.py"
},
"draw": {
"inputs": ["spec_one_output"],
"outputs": ["draw_output"],
"type": "external_cmd",
"external_cmd": "python ./Python/draw.py"
},
"draw_plot": {
"inputs": ["get_file_output"],
"outputs": ["draw_plot_output"],
}
}
}

```

```

    "type": "external_cmd",
    "external_cmd": "python ./Python/draw_plot.py"
  },
  "get_file": {
    "inputs": ["path", "t1", "t2", "f", "step", "f1", "f2"],
    "outputs": ["get_file_output"],
    "type": "external_cmd",
    "external_cmd": "python ./Python/get_file.py"
  },
  "calc_rho": {
    "inputs": ["spec_one_output"],
    "outputs": ["RHO_output"],
    "type": "external_cmd",
    "external_cmd": "python ./Python/calc_rho.py"
  },
  "calc_fmax": {
    "inputs": ["get_file_output", "spec_one_output"],
    "outputs": ["FMax_output"],
    "type": "external_cmd",
    "external_cmd": "python ./Python/calc_fmax.py"
  },
  "corr": {
    "inputs": ["files", "opora", "duration"],
    "outputs": ["corr_output"],
    "type": "external_cmd",
    "external_cmd": "./corr/corr.exe"
  }
}
}
}

```

Рис. 4. Вычислительная модель

## 7. Краткая характеристика результатов корреляционной обработки

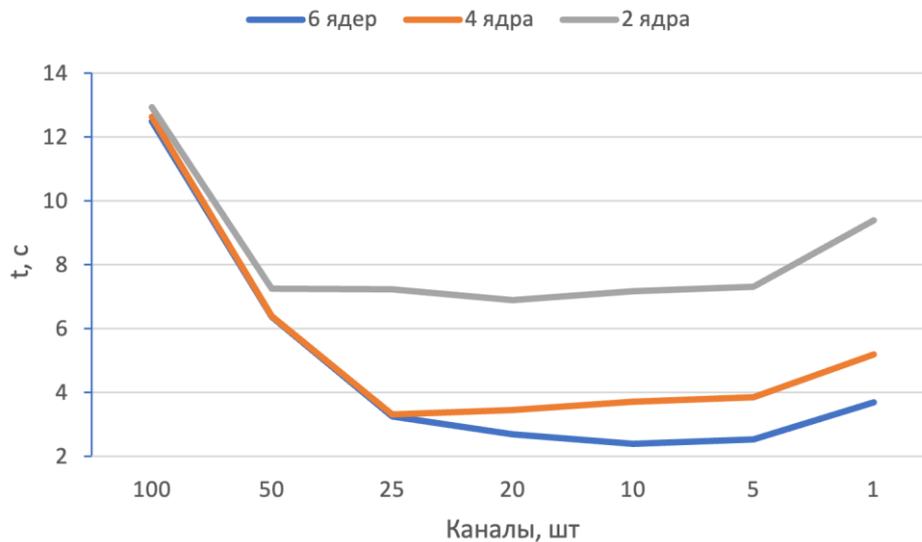


Рис. 5. Зависимость времени обработки от количества ядер процессора

Тестирование выполнялось на компьютере MacBook Pro с процессором M1 Pro, 12 Гб оперативной памяти и частотой процессора 3.2 ГГц. В ходе экспериментов параллельно запускалось разное количество фрагментов. Каждый фрагмент содержал равное количество каналов (1 фрагмент – 100 каналов, 2 фрагмента – 50 каналов в каждом и т.д.). Из рис. 5 видно, что для 6-ядерного процессора оптимальным является параллельный запуск 10 фрагментов по 10 каналов в каждом. Из графика видно, что ускорение близко к линейному (около 2 раз на 2 ядрах, около 4 раз на 4 ядрах и около 6 раз на 6 ядрах), поэтому и на большем числе ядер можно ожидать ускорение в соответствующее число раз.

Как следует из анализа полученных результатов обработки в виде коррелотрасс, параллельная обработка дала прирост производительности и уменьшение времени обработки сигналов. Время обработки уменьшилось до 2.4 с (рис. 5). Для улучшения результатов необходимо провести эксперимент на 8-ядерном процессоре.

## 8. Заключение

Организация программно-аналитической системы для параллельной обработки сейсмических данных является важным шагом в развитии сейсмологии и геологии. Внедрение такой системы позволит значительно улучшить скорость и точность обработки данных, сократить время и затраты на исследования, а также повысить эффективность принятия решений в области геологического исследования. В ходе работы был проведен корреляционный анализ данных вибрационного зондирования наземной и атмосферной трасс распространения геоакустических колебаний.

В результате применения параллельных вычислений на основе системы LuNA было уменьшено время многоканальной обработки сейсмотрасс в 4 раза. Кроме того, программно-аналитическая система способна обеспечить параллельную обработку больших объемов данных, что особенно актуально в контексте современных глобальных вызовов, таких как изменение климата и использование природных ресурсов. Благодаря использованию современных технологий и надежным принципам организации данная система обеспечивает возможность более эффективного изучения сейсмических явлений и разработки новых методов прогнозирования и предотвращения природных катаклизмов. Однако для полного использования потенциала такой системы необходимо дальнейшее исследование и развитие методов анализа данных, а также реализация принципа параллельных вычислений для поточной свертки, применяемой в решении задач, связанных с обработкой сейсмических данных в системе LuNA.

## Литература

1. A foreign function library for Python// Python Documentation [Электронный ресурс]. URL: <https://docs.python.org/2/library/ctypes.html> (дата обращения: 2023).
2. *Kholodkov K. I., Aleshin I. M.* Exact calculation of a posteriori probability distribution with distributed computing systems // *Computer Research and Modeling*. 2015. P. 539–542.
3. *Malyshkin V.* Active Knowledge, LuNA and Literacy for Oncoming Centuries // *Programming Languages with Applications to Biology and Security: Essays Dedicated to Pierpaolo Degano on the Occasion of His 65th Birthday // Lecture Notes in Computer Science*. Springer Cham. 2015. V. 9465. P. 292–303. DOI 10.1007/978-3-319-25527-9\_19.
4. *Malyshkin V. E., Perepelkin V. A.* LuNA Fragmented Programming System, Main Functions and Peculiarities of Run-Time Subsystem // *Parallel Computing Technologies*. 2011. LNCS 6873. P. 53–61.
5. Matplotlib: Python plotting – Matplotlib 3.2.1 documentation // *Matplotlib Documentation* [Электронный ресурс]. URL: <https://matplotlib.org/> (дата обращения: 2023).

6. *Караваяев Д. А., Якименко А. А., Булавина Н. А.* Технология моделирования полного сейсмического поля на высокопроизводительных вычислительных системах // Труды XIII Международной научно-технической конференции «Актуальные проблемы электронного приборостроения», 3–6 октября 2016, Новосибирск. С. 41–45.
7. Python interface to Tcl/Tk // Python Documentation [Электронный ресурс]. URL: <https://docs.python.org/3/library/tkinter.html> (дата обращения: 2023).
8. The sound amplifying forest with emphasis on sounds from wind turbines | Elis Johansson // Department of Civil and Environmental Engineering. Division of Applied Acoustics. Chalmers University of Technology. Sweden. 2010. 97 p.
9. *Марапулец Ю. В., Сенкевич Ю. И., Луковенкова О. О., Солодчук А. А., Ларионов И. А., Мищенко М. А., Малкин Е. И., Щербина А. О., Гапеев М. И.* Комплексный анализ акустических и электромагнитных сигналов для оценки уровня сейсмической опасности: монография. Владивосток: Дальнаука, 2020. 120 с.
10. *Вальковский В. А., Малышкин В. Э.* Синтез параллельных программ и систем на вычислительных моделях. Новосибирск: Наука, 1988. 129 с.
11. *Bendat J. S., Piersol A. G.* Random data: analysis and measurement procedures. New York (N.Y.) Wiley-Interscience, 1971. 88 p.

#### **Выродов Алексей Юрьевич**

аспирант, Институт вычислительной математики и математической геофизики СО РАН, e-mail: g47711@gmail.com, ORCID ID: 0009-0002-3399-3603.

#### **Перепёлкин Владислав Александрович**

к.т.н., Институт вычислительной математики и математической геофизики СО РАН, e-mail: perepelkin@ssd.sgcc.ru, ORCID ID: 0000-0002-6998-4525.

#### **Хайретдинов Марат Саматович**

д.т.н., г.н.с., профессор, Институт вычислительной математики и математической геофизики СО РАН (ИВМиМГ СО РАН, 630090, Новосибирск, просп. Ак. Лаврентьева, 6), тел. +7 383 330 87 43, e-mail: marat@opg.sgcc.ru, ORCID ID: 0000-0002-3250-521X.

#### **Хрыпченко Алёна Валерьевна**

аспирант, Институт вычислительной математики и математической геофизики СО РАН, e-mail: ahrypchenko@yandex.ru, ORCID ID: 0009-0002-7627-6276.

*Авторы прочитали и одобрили окончательный вариант рукописи.*

*Авторы заявляют об отсутствии конфликта интересов.*

*Вклад соавторов: Каждый автор внес равную долю участия как во все этапы проводимого теоретического исследования, так и при написании разделов данной статьи.*

## Principles of Organizing a Software-analytical System for Parallel Processing of Seismic Data

Alexey Y. Vyrodov, Vladislav A. Perepelkin, Marat S. Khayretdinov, Alena V. Khrypchenko

Institute of Computational Mathematics and Mathematical Geophysics SB RAS

**Abstract:** Progress in the development of modern information technologies is directly related to the use of resource-intensive applications in science-intensive research, as well as in industrial applications. Currently, there is an acute problem of analyzing large volumes of geophysical data and increasing the productivity of systems for their study. One of the ways to solve this problem is to use multiprocessor computers and multi-machine computing systems capable of performing parallel, including distributed data processing. The paper presents a description and implementation of a computational model for parallel processing of seismic data based on the LuNA system for the automatic construction of parallel programs.

**Keywords:** seismic trace, fast Fourier transform, cross-correlation function, parallel processing, LuNA system, active knowledge, seismic data.

**For citation:** Vyrodov A. Yu., Perepelkin V. A., Khayretdinov M. S., Khrypchenko A. V. Principles of organizing a software-analytical system for parallel processing of seismic data (in Russian). *Vestnik SibGUTI*, 2024, vol. 18, no. 2, pp. 57-68. <https://doi.org/10.55648/1998-6920-2024-18-2-57-68>.



Content is available under the license  
Creative Commons Attribution 4.0  
License

© Vyrodov A. Yu., Perepelkin V. A.,  
Khayretdinov M. S., Khrypchenko A. V., 2024

The article was submitted: 11.12.2023;  
accepted for publication 22.01.2024.

### References

1. *A foreign function library for Python*. *Python Documentation*, available at: <https://docs.python.org/2/library/ctypes.html> (accessed: 2023).
2. Kholodkov K. I., Aleshin I. M., Exact calculation of a posteriori probability distribution with distributed computing systems. *Computer Research and Modeling*, 2015, vol. 7, pp. 539-542.
3. Malyshkin V. Active Knowledge, LuNA and Literacy for Oncoming Centuries. *Programming Languages with Applications to Biology and Security: Essays Dedicated to Pierpaolo Degano on the Occasion of His 65th Birthday. Lecture Notes in Computer Science*, Springer Cham, 2015, vol. 9465, pp. 292-303, DOI 10.1007/978-3-319-25527-9\_19.
4. Malyshkin V. E., Perepelkin V. A. 2011. LuNA Fragmented Programming System, Main Functions and Peculiarities of Run-Time Subsystem. *Parallel Computing Technologies*, LNCS 6873, pp. 53-61.
5. Matplotlib: Python plotting – Matplotlib 3.2.1 documentation. *Matplotlib Documentation*, available at: <https://matplotlib.org/> (accessed: 2023).
6. Karavaev D. A., Yakimenko A. A., Bulavina N. A. *Technology for modeling the full seismic field on high-performance computing systems*, 2016.
7. Python interface to Tcl/Tk. *Python Documentation*, available at: <https://docs.python.org/3/library/tkinter.html> (accessed: 2023).
8. The sound amplifying forest with emphasis on sounds from wind turbines| Elis Johansson. *Department of Civil and Environmental Engineering. Division of Applied Acoustics*, Chalmers University of Technology, Sweden, 2010, 97 p.
9. Marapulets Yu. V., Senkevich Yu. I., Lukovenkova O. O., Solodchuk A. A., Larionov I. A., Mishchenko M. A., Malkin E. I., Shcherbina A. O., Gapeev M. I. *Comprehensive analysis of acoustic and electromagnetic signals to assess the level of seismic hazard*, 2020.

10. Valkovsky V. A., Malyshkin V. E. *Synthesis of parallel programs and systems using computational models*. — Novosibirsk: Science. Sib. department, 1988, 129 p.
11. Julius S. Bendat, Allan G. Piersol. *Random data: analysis and measurement procedures*. New York (N.Y.) Wiley-Interscience, 1971, 88 p.

**Alexey Yu. Vyrodov**

Postgraduate student, Institute of Computational Mathematics and Mathematical Geophysics SB RAS, e-mail: g47711@gmail.com, ORCID ID: 0009-0002-3399-3603.

**Vladislav A. Perepelkin**

Ph.D., Institute of Computational Mathematics and Mathematical Geophysics SB RAS, e-mail: perepelkin@ssd.sccc.ru, ORCID ID: 0000-0002-6998-4525.

**Marat S. Khairtdinov**

Doctor of Technical Sciences, Chief Researcher, Professor, Institute of Computational Mathematics and Mathematical Geophysics SB RAS (ICM&MG SB RAS, 630090, Novosibirsk, Academician Lavrentiev Avenue, 6), tel. +7 383 330 87 43, e-mail: marat@opg.sccc.ru, ORCID ID: 0000-0002-3250-521X.

**Alena V. Khrypchenko**

Postgraduate student, Institute of Computational Mathematics and Mathematical Geophysics SB RAS, e-mail: ahrypchenko@yandex.ru, ORCID ID: 0009-0002-7627-6276.