

DOI: <https://doi.org/10.15407/rpra23.04.258>

УДК 524.7

PACS number: 95.85.Bh

В. А. САМОДУРОВ^{1,2}, С. Ю. ПАВЛОВ¹, В. А. ТЮРИН¹,
 А. Ю. ЗАЙЦЕВ³, Е. А. ИСАЕВ^{1,2}, А. С. ПОЗАНЕНКО⁴,
 С. В. ЛОГВИНЕНКО², В. В. ОРЕШКО², Д. В. ДУМСКИЙ^{1,2},
 Ю. А. БЕЛЯЦКИЙ², Д. В. ПЕРВУХИН¹, Т. И. ГРОХЛИНА^{3,1}

¹ Национальный исследовательский университет “Высшая школа экономики”,
 ул. Мясницкая, 20, г. Москва, 101000, Россия

E-mail: vsamodurov@hse.ru, sergeypavlov.yu@gmail.com; slava555-tyurin@mail.ru;
 eisaev@hse.ru; dumsky@prao.ru; dpervuhin@hse.ru

² Пушинская радиоастрономическая обсерватория АКЦ ФИАН,
 г. Пушкино, Московская обл., 142290, Россия

E-mail: lsv@prao.ru; oreshko@prao.ru; jury@prao.ru

³ Институт математических проблем биологии РАН,
 ул. проф. Виткевича, 1, г. Пушкино, Московская обл., 142290, Россия

E-mail: sasha-z@psn.ru; grokhlina@mail.ru

⁴ Институт космических исследований РАН,
 ул. Профсоюзная, 84/32, г. Москва, 117997, Россия

E-mail: apozenan@iki.rssi.ru

РАЗРАБОТКА МЕТОДИКИ ОБРАБОТКИ ДАННЫХ КРУГЛОСУТОЧНОГО ОБЗОРА БСА ФИАН НА ЧАСТОТЕ 110 МГц МЕТОДАМИ ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ ВЫЧИСЛЕНИЙ

Предмет и цель работы: Цель работы – разработка работоспособной методики быстрой обработки больших массивов радиоастрономических данных для поиска откликов на внегалактические транзиентные события, которые априори должны иметь большие дисперсионные задержки ($DM \sim 100 \div 2000$ пк · см⁻³). В качестве тестового образца использованы записи 20-суточных непрерывных наблюдений приполярной зоны небесной сферы площадью около 8 квадратных градусов и отдельные файлы с повторяющимся быстрым радиовсплеском FRB 121102.

Методы и методология: Радиотелескоп БСА ФИАН имеет многолучевую диаграмму, способную круглосуточно наблюдать в 96 лучах в диапазоне склонений от -8° до $+42^\circ$ в частотном диапазоне $109 \div 111.5$ МГц. Число частотных полос – от 6 до 32, постоянная времени изменяется в пределах от 0.1 до 0.0125 с. В режиме записи 32 частотных полос с постоянной времени 0.0125 с каждый час записывается 3.4 ГБ данных, в сутки – 87 ГБ, в год – 32 ТБ. На август 2018 г. накоплено около 120 ТБ данных. Обработка такого объема данных для ряда научных задач настоятельно требует использования технологий высокопроизводительных вычислений.

Результаты: В работе предлагаются два способа обработки данных: с использованием графических процессоров (CPU+GPU, расчет на языке программирования C/C++ с использованием OpenCL) и при помощи кластерных вычислений (с использованием интерфейса передачи сообщений MPI на многопроцессорных узлах). Для отработки методики используются как обработка пульсарных данных (использование графических ускорителей позволяет увеличить скорость обработки на 2–3 порядка), так и поиск откликов на внегалактические транзиентные события, которые априори должны иметь большие дисперсионные задержки. В качестве примеров таких событий можно назвать быстрые радиовсплески (FRB), отклики на гамма-всплески (GRB) и, наконец, возможные отклики на гравитационные события, зафиксированные детекторами LIGO. После обработки приполярной зоны было обнаружено 697 кандидатов в импульсные события с высокими мерами дисперсии. Примерно половина из них вызвана техногенными помехами, оставшаяся часть порождена как обычными межпланетными мерцаниями радиоисточников, так и возможными искомыми FRB. Разделить эти два класса событий пока затруднительно. Обработка данных наблюдений в зоне всплеска FRB 121102 также показывает наличие кандидатов в события. Результаты требуют более тщательного анализа.

Заключение: После окончательной отработки методика высокопроизводительных вычислений позволит обрабатывать данные наблюдений с использованием многолучевой диаграммы БСА ФИАН сразу же после их записи.

Ключевые слова: радионаблюдения, большие данные, методика обработки

© В. А. Самодуров, С. Ю. Павлов, В. А. Тюрин,
 А. Ю. Зайцев, Е. А. Исаев, А. С. Позаненко,
 С. В. Логвиненко, В. В. Орешко, Д. В. Думский,
 Ю. А. Беляцкий, Д. В. Первухин, Т. И. Грохлина, 2018

1. Введение

Пушинская радиоастрономическая обсерватория АКЦ ФИАН располагает одним из наиболее высокочувствительных радиотелескопов в мире – БСА (Большая сканирующая антенна), работающим в диапазоне 109–111.5 МГц. БСА ФИАН является радиотелескопом меридианного типа и представляет собой эквидистантную фазированную антенную решетку, состоящую из 16384 волновых диполей, расположенных на площадке 384×187 м (геометрическая площадь – более 70 тыс. м², эффективная – около 45 тыс. м²). В своем диапазоне БСА является самым чувствительным телескопом в мире (и одним из наиболее чувствительных в мире в метровом диапазоне длин волн). Системная эквивалентная плотность потока (SEFD) радиотелескопа равна 34 Ян [1] в зените при минимальной температуре фона, что в несколько раз лучше, чем у радиотелескопа LOFAR [2] (SEFD = 60000 Ян на частоте 80 МГц и 3500 Ян на частоте 120 МГц для отдельной станции; с учетом полного числа станций, 24+14, SEFD = 1600 и 90 Ян на частотах 80 и 120 МГц соответственно).

БСА проектировался (начало наблюдений в 1974 г.) с возможностью формирования на нем многолучевой диаграммы вдоль линии Юг–Север. В 2010–2011 гг., помимо модернизации элементов приемной системы, была разработана и создана новая диаграммообразующая система из 128 лучей.

Начиная с 6 июля 2012 г. на радиотелескопе БСА ФИАН ведется круглосуточный мониторинг, покрывающий за сутки значительную часть небесной сферы (с 6 июля 2012 г. ежесуточно наблюдалось около 2 ср в полосах склонений общей шириной 25°, а с 1 апреля 2013 г. ежесуточно покрывается уже 5.08 ср или 0.4 поверхности всей сферы в полосе склонений 50°). С июня 2014 г. ведутся наблюдения одновременно в двух модах: стандартной моде наблюдений (постоянная времени – 0.1 с; режим регистрации данных – в 6 полосах по 415 кГц – запущен в июле 2012 г., в год регистрируется 848 ГБ) и “быстрой” моде (постоянная времени – 12.5 мс; режим регистрации данных – в 32 полосах по 78 кГц – работает с июня 2014 г., в год регистрируется 32 ТБ).

Данные используются при исследовании импульсных явлений. Одной из научных задач при обра-

ботке полученных данных является поиск откликов на внегалактические события, которые априори должны иметь большие дисперсионные задержки (мера дисперсии $DM \sim 100 \div 2000$ пк · см⁻³). Такие события включают быстрые радиовсплески (FRB) [3], обнаруженные до сих пор только на частотах 1 ГГц и выше, послесвечение соседних космических гамма-всплесков (GRB) и возможные электромагнитные отклики на события, связанные с гравитационными волнами, зарегистрированными в экспериментах LIGO/Virgo [4].

Обработка такого значительного объема данных для ряда научных задач требует использования высокопроизводительных вычислений. После отработки создаваемая методика высокопроизводительных вычислений позволит обрабатывать данные, полученные с помощью многолучевой диаграммы БСА, сразу же после их записи.

В работе обсуждаются два способа обработки данных:

1) при помощи кластерных вычислений с применением интерфейса передачи сообщений MPI на многопроцессорных узлах (скорость вычислений увеличивается от нескольких десятков раз при использовании простых кластеров до миллионов раз при использовании наиболее современных суперкомпьютеров);

2) с применением графических процессоров CPU+GPU с расчетом на языке программирования C/C++ с использованием OpenCL (скорость обработки увеличивается примерно на два–три порядка).

2. Обработка данных БСА методами кластерных вычислений: методика и результаты

В качестве базовой задачи для отработки методики обработки данных при помощи кластерных вычислений использован поиск откликов на внегалактические транзиентные события, которые априори должны иметь большие дисперсионные задержки ($DM \sim 100 \div 2000$ пк · см⁻³): FRB (на частотах ~1 ГГц), отклики на GRB, возможные отклики на гравитационные события, зафиксированные детекторами LIGO. Первые попытки поиска данных такого рода на БСА ФИАН уже сделаны [5–7], но они выполнялись пока только для грубых данных (6 полос, 10 раз в секунду), что оставляет место для сомнений в достоверности результатов.

Итогом первой части работы является проработка методики и создание программного обеспечения (использован язык программирования C++), работающего на компьютерных кластерах под управлением технологии MPI, позволяющего производить массовые вычисления больших объемов данных и находить кандидатов во внеземные объекты с высокими мерами дисперсий.

На основе подробного обзора существующей литературы можно сделать вывод, что данные БСА вполне пригодны для поиска обсуждаемого явления, но его особенности (расплывание импульсов на низких частотах с временем расплывания до нескольких секунд) крайне осложняют выработку эффективной методики обработки данных и поиска событий.

В качестве первичной основы методики взят поиск корреляции между данными в разных частотных каналах. Поскольку сигнал от далеких источников радиоизлучения при прохождении сквозь межзвездную и межгалактическую среду замедляется тем больше, чем ниже частота наблюдений, импульсный сигнал на более низких частотах приходит с запаздыванием от долей секунды до нескольких секунд. Это позволяет организовать проверку корреляционных коэффициентов между соседними частотами на отрезке

в несколько секунд (нами выбраны 16-секундные отрезки, поскольку это позволяет анализировать данные вплоть до $DM \sim 1000$ пк · см⁻³).

Из литературы определена частота возможных событий FRB в приполярной области размером несколько квадратных градусов. По разным литературным источникам можно заключить, что она может быть в диапазоне от одного события раз в несколько дней до нескольких событий в час.

Результаты обработки данных наблюдений небесной полярной зоны при помощи разработанного алгоритма следующие. Объем обработанных данных – 480 файлов 20-суточных наблюдений объемом почти 300 ГБ. Обнаружено более 2 тыс. событий [8], но значительная их часть зарегистрирована в момент подачи калибровочных сигналов, после очистки данных от этих явно ложных событий остается 697 событий с значениями меры дисперсии вплоть до $DM = 650$ пк · см⁻³. Найдено, что часть событий вызвана техногенными помехами (рис. 1), часть – мерцающими источниками (рис. 2).

На рис. 1 в 11-м луче видна четкая хорошая корреляция данных с задержкой импульса при понижении частоты. К сожалению, вероятность того, что это – техногенная помеха, значительна (частота 109.27 МГц демонстрирует частые

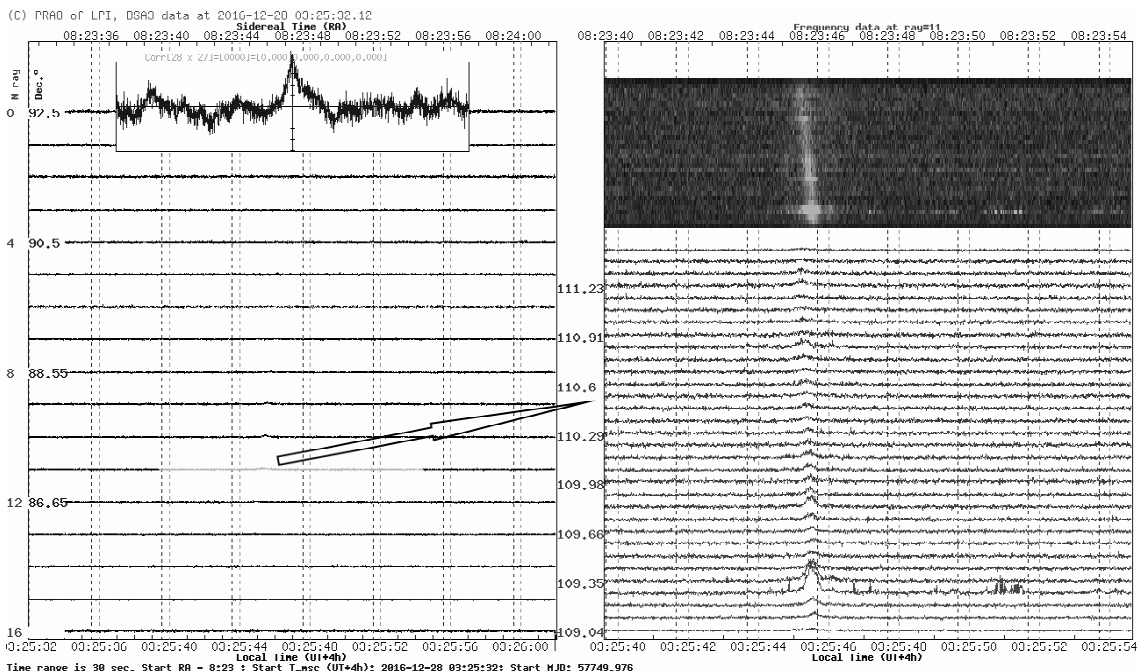


Рис. 1. Образец полярных данных (16 лучей с шагом 0.5°), полученный 03:25:32 LT 28.12.2016 г. В 11-м луче видна четкая хорошая корреляция данных с задержкой импульса при понижении частоты (хорошо видно в верхней части правой панели)

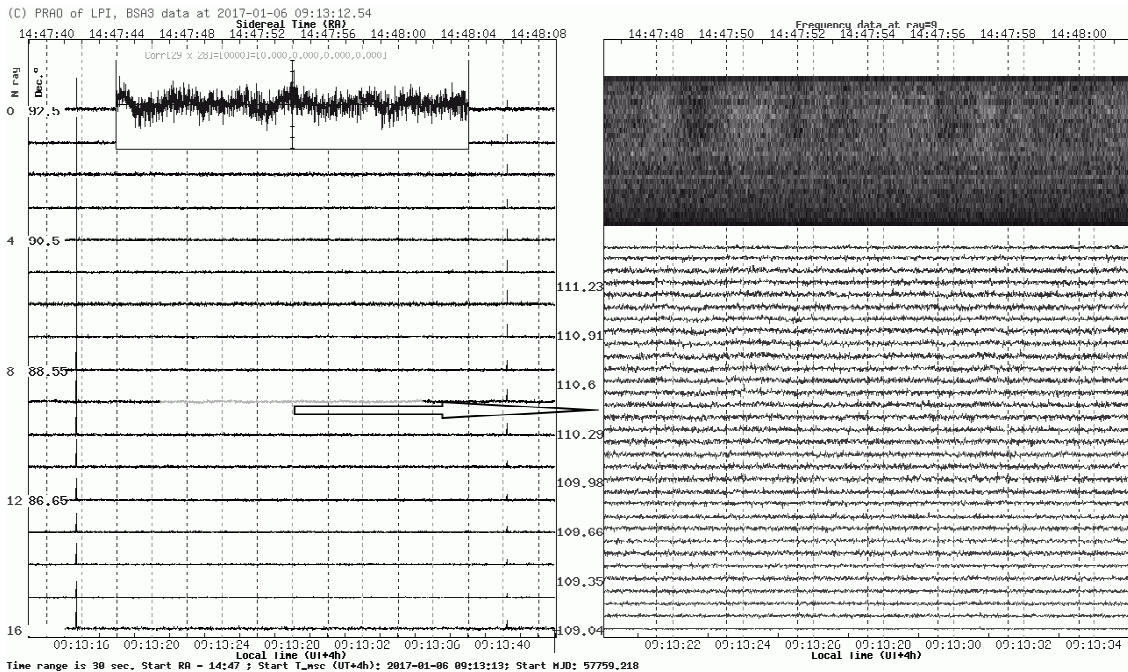


Рис. 2. Образец полярных данных, полученный в 09:13:12 LT 06.01.2017 г., – типичный образец корреляции данных наблюдений межпланетных мерцаний радиоисточников. Коэффициент корреляции слабый, сдвиги данных по времени на более низких частотах порождены случайными мерцаниями

всплески, что наводит на мысль о техногенной помехе).

Однако среди найденных событий могут находиться и искомые FRB. Вопрос требует дальнейшего изучения.

В настоящее время ведется обработка данных с целью выявления возможных событий, связанных с повторяющимся всплеском FRB 121102 [9]. Некоторые зафиксированные события кажутся подходящими для дальнейшего анализа (рис. 3). Так, на рис. 3 координаты $RA = 05^h 32^m 09^s$ и $\delta = +33^\circ 05'$ в наших данных совпадают с объектом FRB 121102, причем присутствует плавный колоколообразный сдвиг на корреляциях 10-секундных отрезков данных на разных частотах (врезка слева внизу показывает коэффициент корреляции Пирсона в зависимости от сдвига данных на двух разных частотах). Корреляционный максимум сдвинут на 1–2 с вправо от центра врезки (который соответствует нулевому сдвигу во времени данных на более низкой частоте от более высокой). Картина примерно соответствует ожидаемому явлению. Прямой корреляционный расчет с использованием наших данных дает $DM = (440 \pm 200)$ пк · см⁻³. Из каталога FRB [10] для данного объекта – $DM = (557 \pm 2)$ пк · см⁻³.

3. Обработка данных БСА на графических ускорителях с использованием технологии OpenCL

Во второй части работы для обработки данных разработан комплекс программ на графических ускорителях персональных компьютеров на основе технологии OpenCL. Реализован программный комплекс, состоящий из алгоритмов фурье-анализа и алгоритмов агрегирующей статистики данных (поточный анализ данных с целью нахождения среднемедианных значений, дисперсий, стандартных отклонений и т. п.). Этот комплекс значительно ускоряет вычисления (на 2–3 порядка) при анализе астрономических данных. Выполненная работа впервые позволит поставить анализ данных на поток – фактически в режиме онлайн. Решение с использованием графических ускорителей легко масштабируется, данный комплекс программ можно будет использовать на любом персональном компьютере с графическими картами для потоковой обработки радиоастрономических данных, а после его доработки – для любых больших массивов, равномерных во времени.

В результате проведенной работы показано, что графические ускорители вместе с технологией

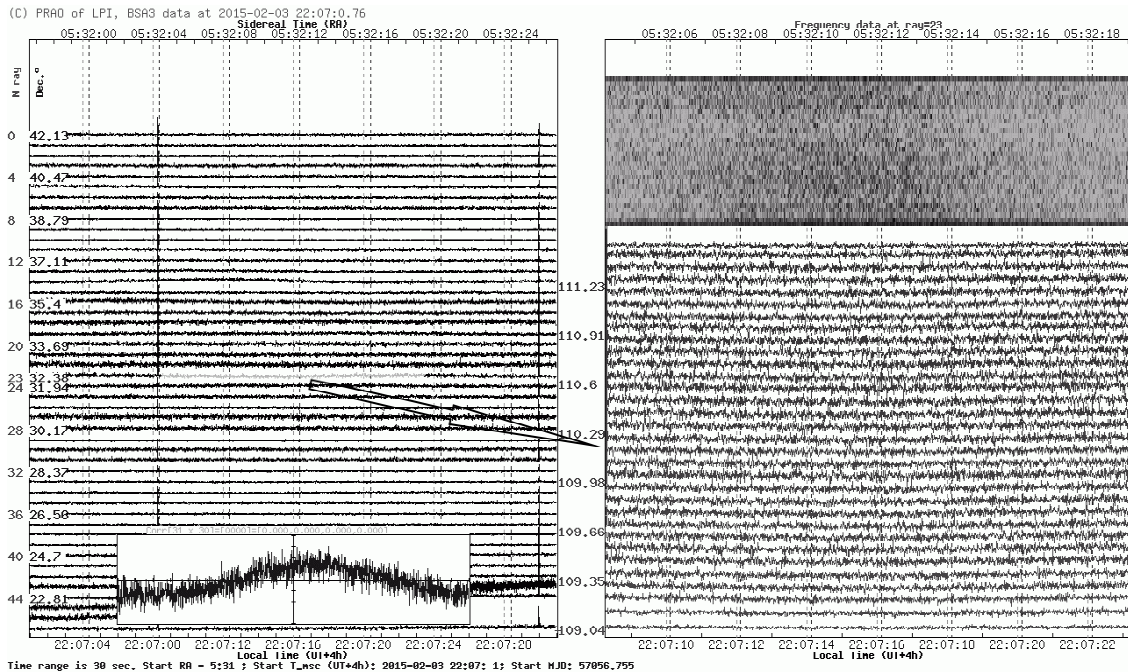


Рис. 3. Пример возможного события от повторяющегося всплеска FRB 121102, зафиксированный 03.02.2015 г. (пояснения в тексте)

OpenCL полностью состоятельны в вопросах анализа астрономических данных и оказались способными конкурировать с суперкомпьютерами в вопросах производительности и цены. С увеличением мощности графических ускорителей появляются все новые возможности для решения задач радиоастрономии.

В рамках второй части работы были решены следующие задачи:

- проанализированы основные технологии вычислений при помощи графических ускорителей;
- выбрана наиболее подходящая технология для анализа имеющегося массива астрономических данных в конкретных условиях – OpenCL;
- проанализированы особенности работы с памятью и ядрами процессоров в графических ускорителях;
- на основе выделенных особенностей разработана архитектура решения для прикладной задачи – фурье-анализа, агрегирования спектрограмм и вычисления метрик;
- с применением разработанной архитектуры выполнено решение с использованием технологии OpenCL и языков C/C++;
- во время выполнения решения были экспериментально проверены некоторые особенности работы с графическими ускорителями;

- решение оптимизировано с использованием директив препроцессора;
- решение проверено на тестовых данных на двух разных видеокартах Nvidia – GTX 1080TI и GTX 950M, различающихся мощностью в 5–6 раз;
- проведен сравнительный анализ времени выполнения двух разных решений проблемы обработки стыков данных при их кусочно-последовательной обработке и выбрано наилучшее решение;
- проведен сравнительный анализ времени выполнения решения на двух разных по мощности графических ускорителях;
- проведен сравнительный анализ времени выполнения решения на графических ускорителях и с использованием других способов обработки данных.

4. Заключение

В настоящей работе была разработана методика поиска излучения объектов внеземного происхождения, основанная на корреляционном анализе радиочастотного спектра с использованием небольших компьютерных кластеров. С помощью этой методики можно выделять слабые импульсные сигналы с мерой дисперсии $DM > 100 \text{ пк} \cdot \text{см}^{-3}$. Она была проверена на данных 20-дневного

наблюдения приполярной области небесной сферы (и отдельных файлов с повторяющимся всплеском FRB 121102). В итоге было обнаружено 697 событий, примерно половина из которых вызвана техногенными помехами, оставшаяся часть порождена как обычными межпланетными мерцаниями радиоисточников, так и возможными искомыми FRB. Разделить эти два класса событий пока затруднительно.

Вторая часть работы посвящена разработке методики потоковой обработки больших данных на графических ускорителях обычных персональных компьютеров. В процессе выполнения работы были созданы решения для типовых задач, которые значительно ускоряют вычисления (на 2–3 порядка) при анализе астрономических данных. Это фурье-анализ поступающих данных и потоковый анализ данных с целью нахождения средне медианных значений, дисперсий, стандартных отклонений и т. п. для отрезков данных в несколько секунд.

Выполненная работа позволит поставить анализ данных на поток и работать значительно быстрее, чем регистрируются данные, фактически в режиме онлайн.

СПИСОК ЛИТЕРАТУРЫ

- Орешко В. В. Радиотелескопы ПРАО – состояние и перспективы. (2014). URL: http://www.prao.ru/conf/rrc2014/docs/22092014/06_Oreshko.pdf (дата обращения 27.09.2018)
- Van Haarlem M. P., Wise M. W., Gunst A. W., Heald G., McKean J. P., Hessels J. W. T., de Bruyn A. G., Nijboer R., Swinbank J., Fallows R., and 191 coauthors. LOFAR: The LOw-Frequency Array. *Astron. Astrophys.* 2013. Vol. 556. id. A2. DOI: 10.1051/0004-6361/201220873
- Lorimer D. R., Bailes M., McLaughlin M. A., Narkevic D. J., and Crawford F. A Bright Millisecond Radio Burst of Extragalactic Origin. *Science*. 2007. Vol. 318, Is. 5851. P. 777–780. DOI: 10.1126/science.1147532
- Pozanenko A. S., Barkov M. V., Minaev P. Yu., Volnova A. A., Mazaeva E. D., Moskvitin A. S., Krugov M. A., Samodurov V. A., Loznikov V. M., and Lyutikov M. GRB 170817A Associated with GW170817: Multi-frequency Observations and Modeling of Prompt Gamma-Ray Emission. *Astrophys. J. Lett.* 2018. Vol. 852, No. 2. id. L30. DOI: 10.3847/2041-8213/aaa2f6
- Федорова В. А. FRBs: поиск вариаций низкочастотного радиоизлучения из соответствующих участков неба. *Всероссийская конференция “Астрофизика высоких энергий сегодня и завтра”: Архив докладов.* (2016). URL: <http://hea.iki.rssi.ru/heaconf/hea/2016/hea/talk/63/> (дата обращения 30.08.2018)
- Федорова В. А., Родин А. Е., Дагкесаманский Р. Д., Самодуров В. А., Логвиненко С. В. Обнаружение быстрых радиовсплесков на радиотелескопе БСА ФИАН. *XXXV Всероссийская конференция “Актуальные проблемы внегалактической астрономии”. Тезисы докладов.* (24–27 апреля 2018 г., Пушино). Пушино, Россия, 2018. с. 39.
- Родин А. Е., Федорова В. А., Самодуров В. А., Логвиненко С. В. Обнаружение быстрых радиовсплесков на радиотелескопе БСА ФИАН. *Астрономический циркуляр.* 2018. № 1641. URL: <http://comet.sai.msu.ru/~gmg/AC/AC1641.pdf> (дата обращения 27.09.2018)
- Павлов С. Ю., Тюрин В. А., Самодуров В. А. Обработка данных круглосуточного обзора БСА ФИАН (110 МГц) методами высокопроизводительных вычислений. *XXXV Всероссийская конференция “Актуальные проблемы внегалактической астрономии”. Тезисы докладов.* (24–27 апреля 2018 г., Пушино). Пушино, Россия, 2018. с. 43
- Scholz P., Spitler L. G., Hessels J. W. T., Chatterjee S., Cordes J. M., Kaspi V. M., Wharton R. S., Bassa C. G., Bogdanov S., Camilo F., Crawford F., Deneva J., van Leeuwen J., Lynch R., Madsen E. C., McLaughlin M. A., Mickaliger M., Parent E., Patel C., Ransom S. M., Seymour A., Stairs I. H., Stappers B. W., and Tendulkar S. P. The Repeating Fast Radio Burst FRB 121102: Multi-wavelength Observations and Additional Bursts. *Astrophys. J.* 2016. Vol. 833, Is. 2. id. 177. DOI: 10.3847/1538-4357/833/2/177
- FRB Catalogue. (2018). URL: <http://frbcat.org/> (дата обращения 22.09.2018)

REFERENCES

- ORESHKO, V. V., 2014. *The state and prospects for LPI PRAO radiotelescopes.* (in Russian). [online]. [viewed 27.09.2018] Available from: http://www.prao.ru/conf/rrc2014/docs/22092014/06_Oreshko.pdf
- VAN HAARLEM, M. P., WISE, M. W., GUNST, A. W., HEALD, G., MCKEAN, J. P., HESSELS, J. W. T., DE BRUYN, A. G., NIJBOER, R., SWINBANK, J., FALLOWS, R., and 191 coauthors, 2013. LOFAR: The LOw-Frequency Array. *Astron. Astrophys.* vol. 556, id. A2. DOI: 10.1051/0004-6361/201220873
- LORIMER, D. R., BAILES, M., MCLAUGHLIN, M. A., NARKEVIC, D. J. and CRAWFORD, F. A., 2007. Bright Millisecond Radio Burst of Extragalactic Origin. *Science*. vol. 318, is. 5851, pp. 777–780. DOI: 10.1126/science.1147532
- POZANENKO, A. S., BARKOV, M. V., MINAEV, P. YU., VOLNOVA, A. A., MAZAEVA, E. D., MOSKVITIN, A. S., KRUGOV, M. A., SAMODUROV, V. A., LOZNIKOV, V. M. and LYUTIKOV, M., 2018. GRB 170817A Associated with GW170817: Multi-frequency Observations and Modeling of Prompt Gamma-Ray Emission. *Astrophys. J. Lett.* vol. 852, no. 2, id. L30. DOI: 10.3847/2041-8213/aaa2f6
- FEDOROVA, V. A., 2016. FRBs: search for variations of low-frequency radio emission from the corresponding areas of the sky. In: *“High Energy Astrophysics Today and Tomorrow” Conference Proceedings* [online]. Moscow, Russia, December 20–23. Available from: <http://hea.iki.rssi.ru/heaconf/hea/2016/hea/talk/63/>
- FEDOROVA, V. A., RODIN, A. E., SAMODUROV, V. A. and LOGVINENKO, S. V., 2018. Detection of fast radio bursts by BSA LPI radio telescope. *Modern Problems for*

- Extragalactic Astronomy: Proceedings of XXXV All-Russian Conference.* Pushchino, Russia, Pushchino Radio Astronomy Observatory, April 24–27, 2018. p. 39.
7. RODIN, A. E., FEDOROVA, V. A., SAMODUROV, V. A. and LOGVINENKO, S. V., 2018. Detection of Fast Radio Bursts at the LPA LPI Radio Telescope. *Astron. Tsirkular.* no. 1641. (in Russian). [online]. [viewed 27.09.2018]. Available from: <http://comet.sai.msu.ru/~gmr/AC/AC1641.pdf>
 8. PAVLOV, S. YU., TYURIN, V. A. and SAMODUROV, V. A., 2018. Data processing for daily survey (110 MHz) by BSA LPI using high-performance computing methods. *Modern Problems for Extragalactic Astronomy: Proceedings of XXXV All-Russian Conference.* Pushchino, Russia, Pushchino Radio Astronomy Observatory, April 24–27, 2018. p. 43
 9. SCHOLZ, P., SPITLER, L. G., HESSELS, J. W. T., CHATTERJEE, S., CORDES, J. M., KASPI, V. M., WHARTON, R. S., BASSA, C. G., BOGDANOV, S., CAMILO, F., CRAWFORD, F., DENEVA, J., VAN LEEUWEN, J., LYNCH, R., MADSEN, E. C., MCLAUGHLIN, M. A., MICKALIGER, M., PARENT, E., PATEL, C., RANSOM, S. M., SEYMOUR, A., STAIRS, I. H., STAPPERS, B. W. and TENDULKAR, S. P., 2016. The Repeating Fast Radio Burst FRB 121102: Multi-wavelength Observations and Additional Bursts. *Astrophys. J.* vol. 833, is. 2, id. 177. DOI: 10.3847/1538-4357/833/2/177
 10. FRB TEAM, 2018. *FRB Catalogue* [online]. [viewed 22.09.2018]. Available from: <http://frbcat.org/>

V. A. Samodurov^{1,2}, S. Yu. Pavlov¹, V. A. Tyurin¹,
A. Yu. Zajcev³, E. A. Isaev^{1,2}, A. S. Pozanenko⁴,
S. V. Logvinenko², V. V. Oreshko², D. V. Dumskij^{1,2},
Yu. A. Belyatskij², D. V. Pervuhin¹, T. I. Grokhlina^{3,1}

¹National Research University Higher School of Economics,
20, Myasnitskaya St., Moscow, 101000, Russia

²Pushchino Radio Astronomy Observatory
of the P. N. Lebedev Physics Institute Astro Space Center,
Pushchino, Moscow region, 142290, Russia

³Institute of Mathematical Problems of Biology, RAS,
1, Professor Vitkevich St., Pushchino,
Moscow region, 142290, Russia

⁴Space Research Institute, RAS,
84/32, Profsoyuznaya St., Moscow, 117997, Russia

DEVELOPMENT OF DATA PROCESSING TECHNIQUES FOR DAILY SURVEY AT 110 MHz BY BSA LPI USING HIGH-PERFORMANCE COMPUTING METHODS

Purpose: The goal of the work is to develop a workable technique for fast processing of large arrays of radio astronomy data to search for responses to extragalactic transient events, which a priori should have large dispersion ($DM \sim 100 \div 2000 \text{ pc} \cdot \text{cm}^{-3}$). Records of 20-day continuous observations of the sky sphere circumpolar zone with an area of about 8 sq deg and separate files with repeating fast radio burst FRB 121102 were used as a test sample of the technique.

Design/methodology/approach: The BSA (Big Scanning Antenna) LPI radio telescope has a multi-beam diagram and is capable of recording daily in the frequency range of 109–111.5 MHz in 96 beams in the declination range from -8° to $+42^\circ$. The num-

ber of frequency bands varies within 6 to 32 for the time constant varying within 0.1 to 0.0125 s, respectively. In the mode of recording 32 frequency bands with the time constant 0.0125 s, 3.4 GB of data are recorded every hour, 87 GB – daily, and 32 TB – per year. To end of August 2018, about 120 TB of data have been accumulated. Processing this amount of data for a number of scientific problems urgently requires the use of high-performance computing technologies.

Findings: The paper proposes two ways to process data: using graphics processors (CPU+GPU, calculation in the C/C++ programming language using OpenCL) and using cluster computing (by using the message passing interface MPI on multiprocessor nodes). To refine the methodology, we use both the processing of pulsar data (using graphic accelerators allows us to increase the processing speed by 2–3 orders of magnitude) and search for responses to extragalactic transient events, which a priori should have large dispersion delays. Examples of such events include fast radio bursts (FRB), responses to gamma radio bursts (GRB) and, finally, possible responses to gravitational events recorded by the LIGO detectors. After processing the polar zone, 697 candidates for pulsed events with high dispersion measures were detected. About half of them is caused by man-made interference, the rest being caused both by the usual interplanetary scintillations of radio sources, and also by the possible sought-for FRBs. It is still difficult to separate these two classes of events. Processing the data observed in the FRB 121102 zone also indicates the presence of event candidates. The results require more thorough analysis.

Conclusions: After finalizing the high-performance computing methodology, it will allow us to process data observed with the BSA LPI multipath diagram immediately after they are recorded.

Key words: radio observation, big data, processing technique

V. A. Самодуров^{1,2}, С. Ю. Павлов¹, В. А. Тюрин¹,
А. Ю. Зайцев³, Е. А. Исаев^{1,2}, А. С. Позаненко⁴,
С. В. Логвиненко², В. В. Орешко², Д. В. Думський^{1,2},
Ю. А. Беляцький², Д. В. Первухін¹, Т. І. Грохліна^{3,1}

¹Національний дослідницький університет
“Вища школа економіки”,
вул. М’ясиницька, 20, м. Москва, 101000, Росія

²Пушинська радіоастрономічна обсерваторія АКЦ ФІАН,
м. Пушино, Московська обл., 142290, Росія

³Інститут математичних проблем біології РАН,
вул. проф. Віткевича, 1, м. Пушино,
Московська обл., 142290, Росія

⁴Інститут космічних досліджень РАН,
вул. Профсоюзна, 84/32, м. Москва, 117997, Росія

РОЗРОБКА МЕТОДИКИ ОБРОБКИ ДАНИХ ЦІЛОДОВОГО ОГЛЯДУ БСА ФІАН НА ЧАСТОТІ 110 МГЦ МЕТОДАМИ ВИСОКОПРОДУКТИВНИХ ОБЧИСЛЕНЬ

Предмет та мета роботи: Мета роботи – розробка працездатної методики швидкої обробки великих масивів радіоастрономічних даних для пошуку відгуків на позагалактичні транзйентні події, які априорі повинні мати великі дисперсійні затримки ($DM \sim 100 \div 2000 \text{ пк} \cdot \text{см}^{-3}$). У якості тестового

зразка використано записи 20-добових безперервних спостережень приполярної зони небесної сфери площею близько 8 квадратних градусів та окремі файли з повторюваним швидким радіосплеском FRB 21102.

Методи і методологія: Радіотелескоп БСА ФІАН має багатопроменеву діаграму, здатну цілодобово спостерігати у 96 променях в діапазоні схилення від -8° до $+42^\circ$ в частотному діапазоні 109 ± 111.5 МГц. Число частотних смуг – від 6 до 32, постійна часу змінюється в межах від 0.1 до 0.0125 с. У режимі запису 32 частотних смуг з постійною часу 0.0125 с щогодини записується 3.4 ГБ даних, за добу – 87 ГБ, за рік – 32 ТБ. На серпень 2018 р. накопичено близько 120 ТБ даних. Обробка такого обсягу даних для низки наукових завдань вкрай вимагає використання технологій високопродуктивних обчислень.

Результати: Пропонується два способи обробки даних: з використанням графічних процесорів (CPU+GPU, обчислення мовою програмування C/C++ з використанням OpenCL) та за допомогою кластерних обчислень (з використанням інтерфейсу передачі повідомлень MPI на багатопроцесорних вузлах). Для відпрацювання методики використовуються як обробка пульсарних даних (використання графічних прискорювачів дозволяє підвищити швидкість об-

робки на 2–3 порядки), так і пошук відгуків на позагалактичні транзйентні події, які априорі повинні мати великі дисперсійні затримки. Прикладами таких подій можна назвати швидкі радіосплески (FRB), відгуки на гамма-сплески (GRB) і, нарешті, можливі відгуки на гравітаційні події, зафіксовані детекторами LIGO. Після обробки приполярної зони було виявлено 697 кандидатів в імпульсні події з високими мірами дисперсії. Приблизно половина з них викликана техногенними перешкодами, решта породжена як звичайними міжпланетними мерехтіннями радіоджерел, так і можливими шуканими FRB. Розділити ці два класи подій наразі важко. Обробка даних спостережень у зоні сплеску FRB 121102 також показує наявність кандидатів у події. Результати вимагають більш ретельного аналізу.

Висновок: Після остаточного відпрацювання методика високопродуктивних обчислень дозволить обробляти дані спостережень з використанням багатопроменевої діаграми БСА ФІАН відразу ж після їх запису.

Ключові слова: радіоспостереження, великі дані, методика обробки

Стаття постуила в редакцію 10.10.2018