

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ М. В. ЛОМОНОСОВА
ФИЗИЧЕСКИЙ ФАКУЛЬТЕТ

ЛЕБЕДЕВ Владислав Владимирович

**ОГРАНИЧЕНИЕ ФУНКЦИИ СВЕТИМОСТИ
БЫСТРЫХ РАДИОВСПЛЕСКОВ ПО ДАННЫМ
СИМЕ И ИССЛЕДОВАНИЕ ПОПУЛЯЦИИ
СЛАБЫХ ВСПЛЕСКОВ**

Кафедра физики частиц и космологии

КУРСОВАЯ РАБОТА 2 КУРСА

Подпись студента: _____

Научный руководитель:
доктор физико-
математических наук,
профессор РАН, член-
корреспондент РАН
Рубцов Григорий Игоревич

Подпись научного руководите-
ля: _____

Москва – 2026

Аннотация

В работе исследуется функция светимости быстрых радиовсплесков. Строится модель, описывающая яркие внегалактические всплески из каталога CHIME/FRB и низкосветимый галактический FRB 200428. С помощью программы `frbpoppy` моделируется внутренняя популяция всплесков и применяется модель регистрации CHIME/FRB, после чего результат сравнивается с реальными данными. Яркая часть хорошо описывается функцией Шехтера с показателем $\alpha \approx -0.2$ в распределении $\psi(L) = dN/d \log_{10} L$ и характерной светимостью $L^* = 10^{45}$ эрг/с. Для слабых всплесков ниже $L_{\text{br}} = 3 \cdot 10^{40}$ эрг/с допустим степенной закон с показателем $\beta \approx -0.6$. Модель предсказывает, что среди галактических событий наиболее вероятны для регистрации слабые по светимости всплески.

Содержание

1	Введение	2
2	Данные и моделирование	3
2.1	Программа <code>frbpoppy</code>	3
2.2	Реальные данные	3
2.3	Функция светимости	4
3	Статистическое сравнение моделей	5
3.1	Принцип сравнения	5
3.2	Нормировка данных	6
4	Результаты	6
4.1	Яркие внегалактические всплески	6
4.2	Слабые галактические всплески	7
4.3	Итоговая функция светимости	8
4.4	Вероятности регистрации по диапазонам светимости	10
5	Сравнение с литературой	10
6	Заключение	12
A	Приложение: статистические критерии сравнения	14
A.1	Одномерные распределения S/N и DM	14
A.2	Двумерное сравнение S/N и DM	17

1 Введение

Быстрые радиовсплески представляют собой короткие радиоимпульсы длительностью порядка миллисекунд [1]. Значительная часть зарегистрированных всплесков имеет внегалактическое происхождение, однако открытие галактического события FRB 200428, связанного с магнитаром SGR 1935+2154, показало, что по крайней мере часть популяции быстрых радиовсплесков может проявляться при значительно меньших светимостях, чем типичные космологические события [2, 3].

В данной работе исследуется вопрос: можно ли описать яркие внегалактические однократные всплески и слабые галактические всплески в рамках одной популяции источников с единой непрерывной функцией светимости? Для ответа на этот вопрос недостаточно рассматривать только одно галактическое событие. Сначала необходимо ограничить функцию светимости по большой выборке внегалактических событий, а затем проверить, какие слабые события такая функция предсказывает для окрестности Млечного Пути.

Под функцией светимости в работе понимается распределение числа всплесков по характерной радиосветимости L - изотропно-эквивалентной радиосветимости всплеска, интегрированной по рабочей радиополосе и имеющей размерность эрг/с. Космическая популяция всплесков описывается функцией светимости, но телескопы регистрируют только часть этой популяции.

Цель работы: построить модель функции светимости быстрых радиовсплесков, совместимую с данными каталога CHIME/FRB, и оценить, какие слабые галактические всплески такая модель допускает.

Общий принцип работы: меняя параметры функции светимости, смоделировать космическую популяцию ярких всплесков → применить к этой популяции модель регистрации телескопом → сравнить смоделированную выборку с реальными данными → выбрать функцию светимости, которая лучше всего описывает данные → экстраполировать лучшие модели на участок слабых всплесков / отдельно провести моделирование для слабых всплесков → проверить согласованность наблюдений с моделями, выбрать согласующиеся модели и построить их предсказания.

2 Данные и моделирование

2.1 Программа `frbpoppy`

Для моделирования космической популяции всплесков использовалась программа `frbpoppy` [4], которая позволяет задавать функцию светимости и другие параметры популяции, а также моделировать регистрацию всплесков телескопами, в том числе CHIME. Программа задаёт космическую популяцию из N_{src} источников с различными параметрами распределения. Затем она вычисляет, какие из этих источников будут зарегистрированы данным телескопом и их количество N_{det} , учитывая его чувствительность и другие характеристики. В результате получается смоделированный массив всплесков, который может быть сравнен с реальными данными.

Важно учитывать, что N_{src} - это размер статистической выборки Монте-Карло, который нельзя интерпретировать как физическое количество источников. В связи с этим физическая нормировка вводилась через *эффективность наблюдения* $\varepsilon = \frac{N_{\text{det}}}{N_{\text{src}}}$, которая показывает, какую часть космической популяции регистрирует телескоп.

2.2 Реальные данные

В работе использовались данные каталога CHIME/FRB [5]. После очистки данных и отбора однократных всплесков в итоговый набор вошло $N_{\text{obs}} = 2672$ ярких однократных всплеска.

Для сравнения модельных наблюдательных популяций с реальными данными использовались следующие параметры: отношение сигнал/шум S/N , характеризующее значимость регистрации; мера дисперсии DM , связанная с количеством свободных электронов вдоль луча зрения; совместное двумерное распределение S/N и DM .

Использование нескольких наблюдаемых величин необходимо, потому что одна величина не отделяет свойства источников от свойств обзора. Например, распределение по S/N чувствительно к порогам регистрации, а распределение по DM связано с расстояниями и вкладом межгалактической среды.

2.3 Функция светимости

Функция светимости задавалась как

$$\varphi(L) = \frac{dN}{dL} \quad (1)$$

Однако её удобнее представлять в виде $\psi(L) = \frac{dN}{d\log_{10} L}$. Такой вид удобен, потому что светимости быстрых радиовсплесков меняются на много порядков, и логарифмическая шкала позволяет наглядно представлять распределение. В дальнейшем параметры функции светимости будут относиться к $\psi(L)$. Соответствующие параметры со штрихами будут относиться к $\varphi(L)$, и между ними будет выполняться соотношение $x' = x - 1$, поскольку $\psi(L) = L \cdot \ln 10 \cdot \varphi(L)$. В работе рассматривалось три вида функций светимости:

1. Степенная функция

$$\psi(L) \propto L^p \quad (2)$$

2. Функция Шехтера [6]

$$\varphi(L) \propto L^{\alpha'} \exp\left(-\frac{L}{L^*}\right) \quad (3)$$

В логарифмической форме это соответствует $\psi(L) \propto L^\alpha \exp\left(-\frac{L}{L^*}\right)$. Параметр L^* назовём *характерной светимостью*, который задаёт светимость, выше которой функция резко подавляется экспонентой. Эта функция позволяет естественно описать популяцию ярких всплесков, в которой не должно быть всплесков с аномально большой светимостью.

3. Составная функция

$$\psi(L) \propto \begin{cases} L^\alpha \exp\left(-\frac{L}{L^*}\right), & L_{\text{br}} \leq L \leq L_{\text{max}} \\ L_{\text{br}}^\alpha \exp\left(-\frac{L_{\text{br}}}{L^*}\right) \left(\frac{L}{L_{\text{br}}}\right)^\beta, & L_{\text{min}} \leq L < L_{\text{br}} \end{cases} \quad (4)$$

Эта функция представляет собой комбинацию функции Шехтера для

ярких всплесков (с показателем α) и степенной функции для слабых всплесков (с показателем β). Точка L_{br} называется *точкой излома*. Множитель перед функцией для слабых всплесков обеспечивает непрерывность функции светимости в точке излома.

3 Статистическое сравнение моделей

3.1 Принцип сравнения

Модель функции светимости нельзя сравнивать с каталогом СНМЕ напрямую по светимости, потому что для большинства одиночных событий каталога нет точных расстояний. Поэтому сравнение проводилось после моделирования регистрации: сначала строилась внутренняя популяция всплесков, затем к ней применялась модель наблюдения СНМЕ, и уже зарегистрированные модельные события сравнивались с реальным каталогом.

В работе использовались наблюдаемые величины, которые непосредственно содержатся в каталоге: отношение сигнал/шум S/N , мера дисперсии DM и флюенс F . Для каждой величины строились распределения по интервалам и сравнивались нормированные формы распределений, поскольку общее число модельных событий зависит от размера моделирования и последующей нормировки.

Для выбора модели использовалась совокупность критериев. Это сделано потому, что разные статистические меры чувствительны к разным типам несовпадения: общий сдвиг распределения, локальные расхождения по интервалам, отличие совместного распределения $S/N-DM$ и так далее. Основной двумерный контроль выполнялся по плоскости $S/N-DM$, так как модель должна воспроизводить как отдельные распределения, так и связь между наблюдаемой значимостью события и мерой дисперсии.

В результате каждая модель получала единый численный критерий качества. Наибольшую роль в нём играло двумерное сравнение $S/N-DM$, потому что оно наиболее сильно ограничивает наблюдаемую структуру каталога. Одномерные распределения S/N и DM использовались как отдельные проверки формы, а флюенс входил как дополнительный контроль.

Подробные определения статистических критериев, нормировок и весов приведены в приложении. Все модели ранжировались по одному и тому же правилу: чем меньше балл, тем лучше зарегистрированная модельная популяция воспроизводит наблюдаемые распределения CHIME.

3.2 Нормировка данных

После того как для модели найдена эффективность обнаружения ε , можно оценить внутреннюю частоту ярких событий. Если за время наблюдения T зарегистрировано N событий, то внутренняя частота событий за это время будет $R = \frac{N}{\varepsilon T}$. Для перехода к частоте в одной галактике типа Млечного Пути использовалась плотность таких галактик $n_{\text{MW}} = 0.01 \text{Mpc}^{-3}$ [7]. Это число используется как шкала перевода объёмной частоты в галактическую.

4 Результаты

4.1 Яркие внегалактические всплески

Наилучшей моделью, описывающей как яркие, так и слабые радиовсплески, оказалась составная функция светимости.

Сравнение ярких внегалактических событий показывает, что функция Шехтера лучше описывает наблюдаемые распределения CHIME/FRB, чем простая степенная функция без экспоненциального подавления на больших светимостях. Наилучшая область по показателю яркой части соответствует диапазону $-0.3 \lesssim \alpha \lesssim -0.1$ с характерной светимостью $L^* \simeq 10^{45}$ эрг/с.

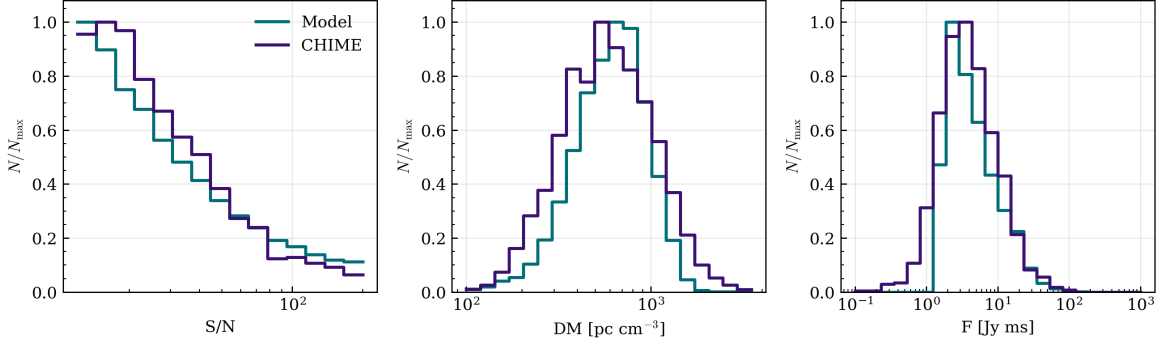


Рис. 1: Сравнение зарегистрированных модельных ярких внегалактических всплесков с каталогом CHIME по отношению сигнал/шум S/N , мере дисперсии DM и флюенсу F для функции Шехтера с показателем $\alpha = -0.2$ и характерной светимостью $L^* = 10^{45}$ эрг/с.

4.2 Слабые галактические всплески

После ограничения участка ярких всплесков был рассмотрен участок более слабых всплесков. Наилучшей моделью оказалась степенная функция после точки излома $L_{br} = 3 \cdot 10^{40}$ эрг/с с допустимым диапазоном степенного показателя слабой части $-0.7 \lesssim \beta \lesssim -0.5$. Также допускается показатель $\beta = -0.4$, однако в таком случае слабые всплески становятся достаточно редкими событиями: вероятность регистрации хотя бы одного всплеска из диапазона 10^{36} – 10^{38} эрг/с равна $\approx 4.0\%$.

Для диапазона светимостей, соответствующего FRB 200428, использовался интервал $10^{36} \leq L \leq 10^{38}$ эрг/с. Такой порядок величины согласуется с оценками радиосветимости FRB 200428: средняя радиосветимость оценивается около $7.5 \cdot 10^{37}$ эрг/с [3, 8]. При $\beta = -0.6$ для этого диапазона была получена галактическая частота всплесков $R = 6.25 \cdot 10^{-2}$ год $^{-1}$, а эффективность регистрации для таких событий $\varepsilon = 0.348$.

За T лет среднее число зарегистрированных событий равно $\lambda(T) = R\varepsilon T$. Для интервала 10^{36} – 10^{38} эрг/с за 10 лет получается

$$\lambda_{10} = 6.25 \cdot 10^{-2} \cdot 0.348 \cdot 10 \approx 0.218.$$

Вероятность зарегистрировать хотя бы одно такое событие по пуассонов-

скому распределению равна

$$P(N \geq 1) = 1 - e^{-\lambda_{10}} \approx 0.196.$$

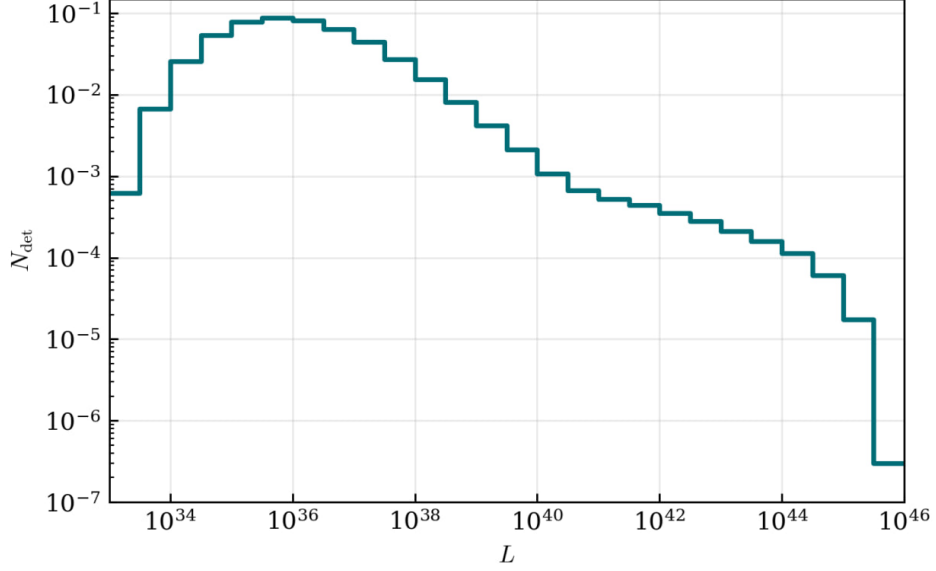


Рис. 2: Модельное распределение зарегистрированных галактических событий по светимости для итоговой модели с $\beta = -0.6$.

4.3 Итоговая функция светимости

Наилучшая модель функции светимости, описывающая как яркие внегалактические, так и слабые галактические всплески, имеет вид:

$$\psi(L) \propto \begin{cases} L^{-0.2} \exp\left(-\frac{L}{L^*}\right), & L_{\text{br}} \leq L \leq 10^{46} \text{ эрг/с} \\ L_{\text{br}}^{-0.2} \exp\left(-\frac{L_{\text{br}}}{L^*}\right) \left(\frac{L}{L_{\text{br}}}\right)^{-0.6}, & 10^{35} \text{ эрг/с} \leq L < L_{\text{br}} \end{cases} \quad (5)$$

где $L_{\text{br}} = 3 \cdot 10^{40}$ эрг/с и $L^* = 10^{45}$ эрг/с. Множитель перед степенной функцией для слабых всплесков обеспечивает непрерывность функции светимости в точке излома. Эта функция хорошо описывает наблюдаемые распределения ярких внегалактических всплесков и допускает слабые галактические всплески с частотой, согласующейся с наблюдениями.

Распределению $\varphi(L) = dN/dL$ соответствуют $\alpha' = \alpha - 1 = -1.2$ для участка ярких всплесков и $\beta' = \beta - 1 = -1.6$ для участка слабых всплесков.

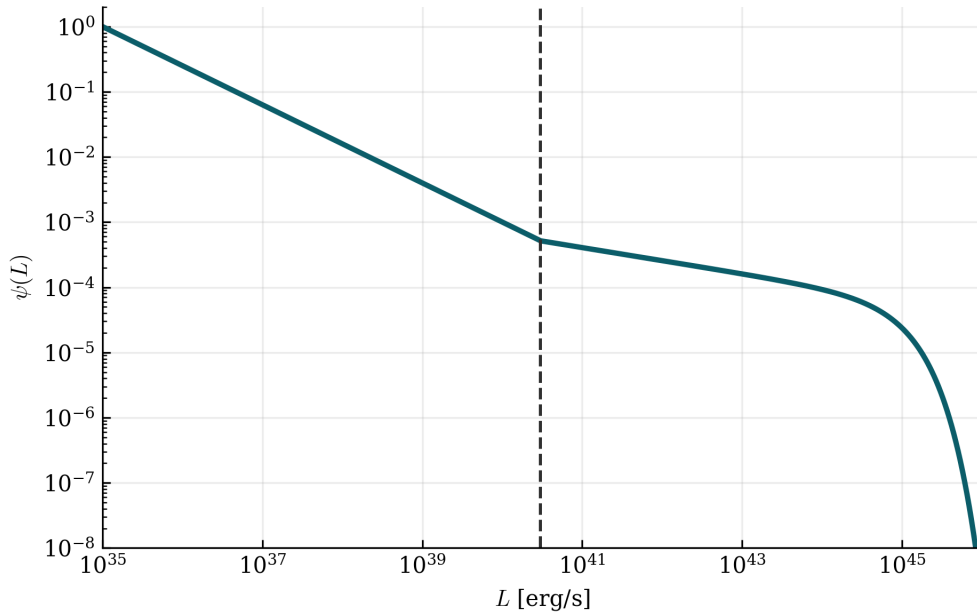


Рис. 3: График нормированной итоговой функции светимости $\psi(L)$ в логарифмическом масштабе: $\alpha = -0.2$, $L^* = 10^{45}$ эрг/с, $L_{\text{br}} = 3 \cdot 10^{40}$ эрг/с, $\beta = -0.6$.

Данная функция не является единственной возможной моделью, однако на степенные показатели α и β наложены достаточно строгие ограничения. Иные α не позволят воспроизвести реальные данные достаточно хорошо, а иные β будут приводить к слишком высокой или слишком низкой частоте слабых галактических всплесков, что будет противоречить наблюдениям.

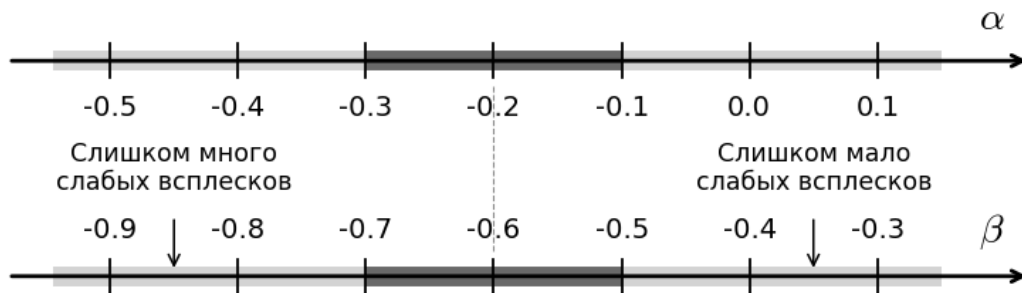


Рис. 4: Допустимые диапазоны показателей α и β для функции светимости отмечены тёмно-серым цветом, недопустимые – светло-серым.

4.4 Вероятности регистрации по диапазонам светимости

Для итоговой модели были вычислены ожидаемые числа зарегистрированных галактических событий в нескольких диапазонах светимости. В таблице указаны: внутренняя частота событий R , частота зарегистрированных событий R_{det} , ожидаемое число регистраций за 10 лет λ_{10} , вероятность хотя бы одной регистрации за 10 лет P_{10} и время t_1 , за которое в среднем ожидается одна регистрация.

Таблица 1: Ожидаемая регистрация галактических событий.

Диапазон L , эрг/с	R , год $^{-1}$	R_{det} , год $^{-1}$	λ_{10}	P_{10}	t_1 , лет
$10^{35}\text{--}10^{36}$	$1.99 \cdot 10^{-1}$	$1.65 \cdot 10^{-2}$	0.165	0.152	60.5
$10^{36}\text{--}10^{38}$	$6.25 \cdot 10^{-2}$	$2.18 \cdot 10^{-2}$	0.218	0.196	46.0
$10^{38}\text{--}10^{40}$	$3.95 \cdot 10^{-3}$	$2.99 \cdot 10^{-3}$	0.0299	0.0295	334
$10^{35}\text{--}10^{40}$	$2.65 \cdot 10^{-1}$	$4.13 \cdot 10^{-2}$	0.413	0.338	24.2
$10^{40}\text{--}10^{42}$	$3.36 \cdot 10^{-4}$	$2.71 \cdot 10^{-4}$	$2.71 \cdot 10^{-3}$	$2.71 \cdot 10^{-3}$	3685
$10^{42}\text{--}10^{46}$	$1.45 \cdot 10^{-4}$	$1.20 \cdot 10^{-4}$	$1.20 \cdot 10^{-3}$	$1.20 \cdot 10^{-3}$	8358

Из таблицы видно, что распределение ожидаемых регистраций по светимости не является монотонным. Максимум зарегистрированного распределения приходится на область порядка $10^{35}\text{--}10^{36}$ эрг/с. Для более широкого диапазона $10^{36} \leq L \leq 10^{38}$ эрг/с наибольшая вероятность регистрации за 10 лет получается $P_{10} = 0.196$. Таким образом, основной вклад в ожидаемые галактические регистрации даёт область $10^{35}\text{--}10^{38}$ эрг/с, тогда как вклад событий со светимостью выше 10^{38} эрг/с быстро уменьшается. Для объединённого диапазона $10^{35}\text{--}10^{40}$ эрг/с получено $R_{\text{det}} = 4.13 \cdot 10^{-2}$ год $^{-1}$, что соответствует среднему времени ожидания одной регистрации $t_1 \simeq 24.2$ года.

5 Сравнение с литературой

Полученные выводы согласуются с рядом результатов, известных из литературы.

1. Предпочтение функции Шехтера для участка ярких всплесков согласуется с работами по функциям светимости и энергии быстрых радиовсплесков. В работе Luo et al. [9] функция Шехтера использовалась для описания светимостей быстрых радиовсплесков, причём характерная светимость L^* получалась порядка 10^{45} эрг/с. В работе Hashimoto et al. [10], основанной на первом каталоге CHIME/FRB, также обсуждаются формы функций энергии, близкие к функции Шехтера. В нашей работе характерная светимость также равна $L^* = 10^{45}$ эрг/с.
2. Полученный диапазон показателя α для яркой части $-0.3 \lesssim \alpha \lesssim -0.1$ согласуется с результатами Luo et al. [9], где для функции Шехтера был получен показатель $\alpha' = -1.79_{-0.35}^{+0.31}$, что в логарифмической шкале соответствует $\alpha = -0.79_{-0.35}^{+0.31}$. С учётом неопределённостей на уровне 2σ данные диапазоны оказываются совместимыми.
3. Необходимость излома светимости согласуется с результатом Chen et al. [11], где по первому каталогу CHIME/FRB была получена ломаная степенная функция с изломом $L_b \simeq 1.33 \cdot 10^{41}$ эрг/с. В нашей модели излом находится при $L_{br} = 3 \cdot 10^{40}$ эрг/с, то есть в близкой области по порядку величины.
4. Сама идея связи галактического события FRB 200428 с низкой светимостью с общей популяцией быстрых радиовсплесков соответствует работам, появившимся после его регистрации [2, 3, 12]. Отличие данной работы состоит в том, что связь проверяется не только качественно, но и через единую составную функцию светимости и оценку вероятности регистрации галактических событий разных светимостей.
5. Использование моделирования наблюдаемой популяции через `frbpoppy` согласуется с методологией популяционного синтеза быстрых радиовсплесков [4]. В данной работе этот подход применяется к задаче совместного описания яркой внегалактической и слабой галактической областей.

6 Заключение

В работе построена составная функция светимости быстрых радиовсплесков, совместимая с наблюдаемыми распределениями каталога CHIME/FRB и допускающая слабые галактические события. Участок ярких всплесков описывается функцией Шехтера с показателями $-0.3 \lesssim \alpha \lesssim -0.1$ в распределении $\psi(L) = dN/d \log_{10} L$ и характерной светимостью $L^* = 10^{45}$ эрг/с.

Для участка слабых всплесков со светимостью ниже $L_{\text{br}} = 3 \cdot 10^{40}$ эрг/с получен допустимый диапазон степенного показателя $-0.7 \lesssim \beta \lesssim -0.5$. В качестве итогового представителя этого диапазона используется $\beta = -0.6$. При таком продолжении вероятность регистрации события в диапазоне светимостей FRB 200428, 10^{36} – 10^{38} эрг/с, за 10 лет составляет около 20%, а для получения среднего числа регистраций, равного единице за 10 лет, требуется увеличить эффективную долю регистрируемых событий примерно в 4.6 раза.

Новым результатом является совместная проверка яркой внегалактической и слабой галактической областей в рамках одной функции светимости после учёта регистрации CHIME/FRB. Получено, что область малых светимостей для функции распределения не является произвольной: слишком крутое продолжение приводит к избыточной частоте слабых событий, а слишком пологое – к слишком малой вероятности регистрации галактических событий диапазона FRB 200428.

Практическое следствие модели состоит в том, что для галактических всплесков наиболее перспективным для регистрации оказывается диапазон $10^{36} \lesssim L \lesssim 10^{38}$ эрг/с. Это связано с компромиссом между внутренней частотой событий и эффективностью их обнаружения.

Список литературы

- [1] E. Petroff, J. W. T. Hessels, and D. R. Lorimer. Fast radio bursts at the dawn of the 2020s. *Astron. Astrophys. Rev.*, 30(1):2, 2022. INSPIRE HEP; doi:10.1007/s00159-022-00139-w.
- [2] B. C. Andersen et al. A bright millisecond-duration radio burst from

- a Galactic magnetar. *Nature*, 587(7832):54–58, 2020. INSPIRE HEP; doi:10.1038/s41586-020-2863-y.
- [3] Christopher D. Bochenek, Vikram Ravi, Konstantin V. Belov, Gregg Hallinan, Jonathon Kocz, Shri R. Kulkarni, and Dan L. McKenna. A fast radio burst associated with a Galactic magnetar. *Nature*, 587(7832):59–62, 2020. INSPIRE HEP; doi:10.1038/s41586-020-2872-x.
- [4] D. W. Gardenier, J. van Leeuwen, L. Connor, and E. Petroff. Synthesising the intrinsic FRB population using frbpoppy. *Astron. Astrophys.*, 632:A125, 2019. INSPIRE HEP; doi:10.1051/0004-6361/201936404.
- [5] Thomas Abbott et al. The Second CHIME/FRB Catalog of Fast Radio Bursts. *Astrophys. J. Suppl.*, 283(1):34, 2026. INSPIRE HEP; doi:10.3847/1538-4365/ae3828.
- [6] P. Schechter. An analytic expression for the luminosity function for galaxies. *Astrophys. J.*, 203:297–306, 1976. INSPIRE HEP; doi:10.1086/154079.
- [7] J. Abadie et al. Predictions for the Rates of Compact Binary Coalescences Observable by Ground-based Gravitational-wave Detectors. *Class. Quant. Grav.*, 27:173001, 2010. INSPIRE HEP; doi:10.1088/0264-9381/27/17/173001.
- [8] Yun-Wei Yu, Yuan-Chuan Zou, Zi-Gao Dai, and Wen-Fei Yu. The confrontation of the shock-powered synchrotron maser model with the Galactic FRB 200428. *Mon. Not. Roy. Astron. Soc.*, 500(2):2704–2710, 2021. doi:10.1093/mnras/staa3374.
- [9] Rui Luo, Yunpeng Men, Kejia Lee, Weiyang Wang, D. R. Lorimer, and Bing Zhang. On the FRB luminosity function – II. Event rate density. *Mon. Not. Roy. Astron. Soc.*, 494(1):665–679, 2020. INSPIRE HEP; doi:10.1093/mnras/staa704.
- [10] Tetsuya Hashimoto et al. Energy functions of fast radio bursts derived from the first CHIME/FRB catalogue. *Mon. Not. Roy. Astron. Soc.*, 511(2):1961–1976, 2022. INSPIRE HEP; doi:10.1093/mnras/stac065.

- [11] J. H. Chen, X. D. Jia, X. F. Dong, and F. Y. Wang. The Formation Rate and Luminosity Function of Fast Radio Bursts. *Astrophys. J. Lett.*, 973(2):L54, 2024. INSPIRE HEP; doi:10.3847/2041-8213/ad7b39.
- [12] Wenbin Lu, Pawan Kumar, and Bing Zhang. A unified picture of Galactic and cosmological fast radio bursts. *Mon. Not. Roy. Astron. Soc.*, 498(1):1397–1405, 2020. INSPIRE HEP; doi:10.1093/mnras/staa2450.

А Приложение: статистические критерии сравнения

А.1 Одномерные распределения S/N и DM

В работе сравнивались популяции всплесков, зарегистрированных модельным наблюдением, с реальными данными каталога CHIME/FRB. Сравнение выполнялось по непосредственно измеряемым параметрам каталога: отношению сигнал/шум S/N и мере дисперсии DM.

Сравниваемая величина x делится на интервалы

$$[x_0; x_1), [x_1; x_2), \dots, [x_{m-1}; x_m)$$

Для каждого интервала i вычисляются n_i^{obs} – число реальных событий CHIME в интервале i и n_i^{mod} – число зарегистрированных модельных событий в интервале i . Сравнение модели с данными выполняется на основе этих чисел, используя различные статистические критерии. Чтобы сравнивать именно форму распределений, а не абсолютное число событий, эти числа нормировались:

$$p_i^{\text{obs}} = \frac{n_i^{\text{obs}}}{\sum_{j=1}^m n_j^{\text{obs}}}, \quad p_i^{\text{mod}} = \frac{n_i^{\text{mod}}}{\sum_{j=1}^m n_j^{\text{mod}}}$$

После такой нормировки p_i и q_i можно понимать как вероятности попасть в i -й интервал. Например, если в некотором интервале по DM находится 20% событий каталога и 18% зарегистрированных модельных событий, то в этом интервале модель хорошо воспроизводит данные. Если в каталоге там 20%, а в модели 5%, то модель недодает событий в этой области.

Пусть есть набор данных x_1, x_2, \dots, x_n . Тогда *эмпирическая функция распределения* определяется как:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I(x_i \leq x) \quad (6)$$

где $I(x_i \leq x)$ равна единице, если условие выполняется, и нулю, если не выполняется.

В работе использовались несколько статистических критериев для сравнения одномерных распределений:

1. Расстояние Колмогорова-Смирнова

Статистика Колмогорова-Смирнова сравнивает две эмпирические функции распределения:

$$D_{KS} = \sup_x |F_{\text{obs}}(x) - F_{\text{mod}}(x)| \quad (7)$$

Здесь $F_{\text{obs}}(x)$ – эмпирическая функция распределения для каталога СНМЕ, а $F_{\text{mod}}(x)$ – эмпирическая функция распределения для зарегистрированной модельной выборки. Если D_{KS} мало, то реальная и модельная выборки накапливаются похожим образом. Если D_{KS} велико, значит, в какой-то области одна выборка дает заметно больше событий, чем другая. Но у этого критерия есть ограничение: он показывает максимальное накопленное расхождение, но не всегда хорошо описывает локальные различия формы в отдельных интервалах. Поэтому одного критерия Колмогорова-Смирнова недостаточно.

2. Расстояние Йенсена-Шеннона

Сначала вводится среднее распределение

$$m_i = \frac{p_i + q_i}{2}$$

Затем вычисляется

$$D_{JS} = \left[\frac{1}{2} \sum_i p_i \log_2 \frac{p_i}{m_i} + \frac{1}{2} \sum_i q_i \log_2 \frac{q_i}{m_i} \right]^{1/2}. \quad (8)$$

Если в каком-то интервале $p_i = 0$, соответствующее слагаемое для p_i считается равным нулю; аналогично для $q_i = 0$.

Это расстояние показывает, насколько две вероятностные гистограммы отличаются по форме. Оно симметрично: расстояние от модели до данных такое же, как от данных до модели.

3. Расстояние полной вариации

Полное вариационное расстояние определяется как

$$D_{\text{TV}} = \frac{1}{2} \sum_i |p_i - q_i| \quad (9)$$

Полная вариация учитывает все локальные несовпадения по всей гистограмме.

4. Расстояние Вассерштейна

Расстояние Вассерштейна использует такое понятие, как *квантиль*. Квантиль Q уровня u – это значение, ниже которого лежит доля u от всех данных. Например, медиана – это квантиль уровня 0.5, которая делит данные на две равные части. Расстояние Вассерштейна между двумя распределениями p и q определяется как

$$D_W = \langle |Q_{\text{obs}}(u) - Q_{\text{mod}}(u)| \rangle_u \quad (10)$$

Поскольку S/N и DM не имеют одинаковой размерности, для сравнения их распределений использовалось нормированное расстояние Вассерштейна, делённое на диапазон $x_{\text{max}} - x_{\text{min}}$:

$$D'_W = \frac{\langle |Q_{\text{obs}}(u) - Q_{\text{mod}}(u)| \rangle_u}{x_{\text{max}} - x_{\text{min}}}$$

Расстояние Вассерштейна можно понимать как среднюю величину сдвига, необходимого для превращения одного распределения в другое. Этот критерий учитывает, насколько далеко распределение сдвинуто. Если модельная гистограмма похожа на реальную, но немного смещена вправо или влево, расстояние Вассерштейна это заметит.

5. Смещение медиан

Даже если две гистограммы в целом похожи, модель может быть немного сдвинута относительно данных. Для контроля такого систематического сдвига использовалась величина

$$D_{\text{med}} = \frac{|\text{med}_{\text{obs}} - \text{med}_{\text{mod}}|}{x_{\text{max}} - x_{\text{min}}} \quad (11)$$

Для каждой одномерной величины строился объединённый критерий:

$$S_{1D} = 0.35D_{\text{KS}} + 0.25D_{\text{JS}} + 0.15D_{\text{TV}} \\ + 0.20D'_{\text{W}} + 0.05D_{\text{med}}$$

Сумма коэффициентов равна единице. Чем меньше S_{1D} , тем лучше модель воспроизводит распределение данной наблюдаемой величины. Расстояние Колмогорова-Смирнова получило самый большой вес, потому что оно проверяет накопленную форму распределения без зависимости от интервалов; расстояние Йенсена-Шеннона проверяет форму нормированной гистограммы; расстояние Вассерштейна важно для контроля смещений; полная вариация и смещение медиан являются дополнительными критериями.

А.2 Двумерное сравнение S/N и DM

Одномерные распределения не полностью описывают данные. Модель может отдельно хорошо воспроизвести распределение по S/N и отдельно хорошо воспроизвести распределение по DM , но при этом неверно описать их совместное распределение. Поэтому дополнительно сравнивалась двумерная гистограмма в плоскости S/N - DM . Для каждой ячейки считались h_{ij}^{obs} и h_{ij}^{mod} – число событий в ячейке i, j для реальных данных и модели соответственно. Здесь i нумерует интервалы по S/N , а j – интервалы по DM . Эти числа также нормировались:

$$r_{ij}^{\text{obs}} = \frac{h_{ij}^{\text{obs}}}{\sum_{i,j} h_{ij}^{\text{obs}}}, \quad r_{ij}^{\text{mod}} = \frac{h_{ij}^{\text{mod}}}{\sum_{i,j} h_{ij}^{\text{mod}}}$$

Затем вычислялось ожидаемое число событий, если бы данные подчинялись модельному распределению: $\mu_{ij} = N_{\text{obs}} r_{ij}^{\text{mod}}$ После этого вычислялось

пуассоновское отклонение P :

$$D_P = \frac{2}{N_{\text{obs}}} \sum_{ij} \left[\mu_{ij} - h_{ij}^{\text{obs}} + h_{ij}^{\text{obs}} \ln \frac{h_{ij}^{\text{obs}}}{\mu_{ij}} \right] \quad (12)$$

Если в ячейке $h_{ij}^{\text{obs}} = 0$, логарифмическое слагаемое для этой ячейки считается нулевым. Это отклонение отвечает на вопрос: насколько сильно наблюдаемая двумерная картина отличается от той, которую предсказывает модель?

Кроме пуассоновского отклонения для двумерной гистограммы также рассматривались расстояния Йенсена–Шеннона и полное вариационное расстояние, примененные уже к двумерным вероятностным массивам. Дополнительно контролировалось положение центра распределения на плоскости $\log_{10} S/N - \log_{10} DM$.

Двумерный критерий имеет вид

$$S_{2D} = 0.60D_P + 0.20D_{JS} + 0.10D_{TV} + 0.10D_{\text{cen}}$$

Здесь D_{cen} – нормированное расстояние между центрами двумерных распределений.

После вычисления одномерных и двумерных баллов использовался общий критерий:

$$S_{\text{obs}} = \frac{1.5S_{2D}(S/N, DM) + 0.8S_{1D}(S/N) + 0.8S_{1D}(DM) + 0.35S_{1D}(F)}{1.5 + 0.8 + 0.8 + 0.35}$$

Этот критерий использовался для ранжирования моделей между собой. Чем меньше S_{obs} , тем лучше модель воспроизводит данные. Веса подобраны так, чтобы учитывать приоритет и информативность каждого критерия.